Attack Detection Approach by Packet Analysis Using Online Learning with Kernel Method and Correlation Change Method

Ayahiko Niimi, Koki Takahata
School of Systems Information Science
Future University Hakodate
2-116 Kamedanakano, Hakodate
Hokkaido 041-8655, Japan

Abstract

Recently, information systems are used in schools and companies and have become essential to work. However, cyber-attacks, such as stealing confidential information, stopping systems and tampering with information, pose risks. Thus, an anomaly detection and misuse detection based on machine learning and statistical methods for network monitoring is used as countermeasures against cyber-attacks. In this paper, we propose two methods to attack detection. One is an attack detection method using an online learning method. The other is an attack detection method using a structural change detection method. If abnormal traffic is monitored and discovered quickly, we can implement countermeasures before confidential information is stolen and serves are stopped. First, in this research, we propose a system using an online learning method that applies the kernel method to the intrusion detection problem. The outline of the proposed method and the learning algorithm are described. To verify of our proposed method, we conducted an experiment and discussed the results. Next, the proposed structural change detection method attempts to use structural changes to detect cyber-attacks. In addition, we propose an anomaly detection method to detect collapsed correlation via an attack on a network by structural change detection, where HTTP-DNS and syn-ack pairs are used as attributes. We conducted an experiment to evaluate the proposed structural change detection method. As a result, security can be reinforced relatively to availability and confidentiality.

1. Introduction

Information systems are recently used in companies and schools, and are indispensable to conduct business, however, there is a risk of reciving cyber-attacks such as stealing information from information systems, stopping system functions, and tampering with the content. There is a risk of receiving. Therefore, taking measures against such cyber-attacks is important to maintain the availability, confidentiality and integrity of the information system.

As a countermeasure against cyber-attacks, a method is used to monitor whether the communication is unusual on the network. Therefore, if you can monitor unusual communications and discover them early, you can take measures before the information is stolen or services are stopped. There are monitoring tools for network monitoring and network protocols for monitoring. In monitoring tools, the traffic that passes through the server is comprehensively provided, including application layer and management. The protocol can aggregate traffic passing through network switches and router ports and can transfer aggregated data to the monitoring server; thus, there are network switches and routers that implement these protocols. Therefore, by using these monitoring tools and protocols, it is possible to set up an environment that automatically monitors network aggregate information.

Use the collected data to perform pattern matching with existing unauthorized communication, set thresholds for traffic volume, and manually set alerts based on past experience such as alerting when thresholds are exceeded. However, such a method is very expensive to operate because it is necessary to change the setting every time malware or cyber-attack occurs by a specialist. Therefore, the machine learning method and the statistical method can reduce the cost by acquiring the feature value mechanically by calculation, so that the machine learning and statistical method for network traffic packets are used. Numerous cyber-attack detection methods have been studied. There are two major types of cyber-attack detection methods using machine learning and statistical methods.

- (Method 1): A method of binary classification of attack communication and normal communication
- (Method 2): An anomaly detection method that defines normal state from normal communication and detects from the departure from the normal state

Method 1: consists of both past attack traffic and normal traffic. By creating a classification model with both data as supervised data, we extract feature quantities of past attack traffic and detect anomalies by classifying new traffic similar to the feature quantities.

Method 2: models a normal state from the traffic that is usually performed and defines how much it deviates from the model as an outlier. After that, if the threshold is exceeded, the new outliers of traffic are calculated and detected as abnormal. Here it is assumed that the attack traffic is different from the normal traffic and deviates from the normal model.

Comparing these two methods, Method 1 needs to prepare attack data, and labeling of teacher data requires specialized knowledge, for example, while the cost is high, Method 2 usually assumes that the traffic being performed is normal. Assume that this has the advantage of an easy collection of data. However, Method 2 only determines that all data apart from the normal state is abnormal, and the feature value cannot be obtained from the attack traffic. It will be necessary.

In addition, compared to methods using pattern matching and thresholds, cyber-attack detection methods using machine learning and statistical methods have the problem that the frequency of alerts due to false detection increases, so they should be used in combination with methods based on heuristics it is considered to be practical.

The main purpose of this paper is to detect the change of traffic behavior in the network by machine learning method and statistical method. First, we propose a system using an online learning method that applies the kernel method to the intrusion detection problem as a method using the machine learning method. Next, we propose a network anomaly detection method using the structure detection method. In this network anomaly detection method using the structure detection method, we aimed at the analysis of the traffic that breaks the correlation, the analysis of the attack, and the confirmation whether it could be detected by the structure change detection.

This paper is organized as follows. This section describes the research background and purpose. The proposal of an anomaly detection method using online learning with kernel method is described in Section 2. Section 3 describes the structural change detection related to this research. The proposal of an anomaly detection method using structural change detection is described in Section 4. Section 5 describes the future prospects of this research and the summary of research in addition to the discussion of experiments.

2. Online Learning with Kernel Method

Here, we propose a system that uses an online learning method that applies the kernel method to the intrusion detection problem. The outline of the proposed method and the learning algorithm are described. The Figure 1 shows an overview of the system using online learning.

The input data is divided into learning packet data and unknown packet data. As packet data for learning, packet data detected by pattern matching and attack data obtained on another network are assumed. It is assumed that packet data in the network that performs intrusion detection is used for normal data. For unknown data, packet data input on the intranet using the detection system is used. In practice, while updating the model by learning the model (1 in Figure 1), the operation is performed by identifying unknown data (2 in Figure 1).

The flow of the experimental method is described. Figures 2 and 3 show the learning stage and the classification stage. For each packet of learning data and unknown data, the feature is extracted from the packet. A process to extract features from both training data and unknown data is performed.

- First, learning is performed using the training data (1 in the Figure). To perform two-class classification, the attack packet and the normal packet are labeled, and learning is performed for each session together with the extracted features.

- Next, give unknown data (2 in the Figure). It identifies the attack data or the normal data using the model learned in the learning phase for the unknown data. Judgment is made as to whether the attack has come to the network by identifying.

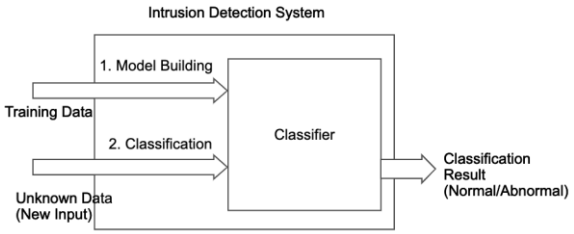

Figure 1. Overall of Proposed System

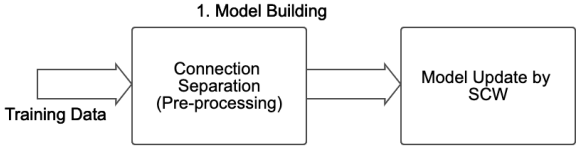

Figure 2. Flow of the Proposed Method (Model Building)

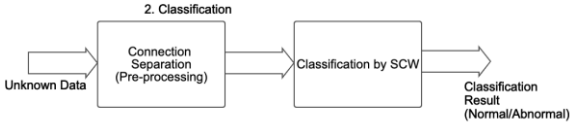

Figure 3. Flow of the Proposed Method (Classification)

In the experiment, SVM (Support Vector Machine) and SCW (Soft-Confidence Weighted learning) [1] were compared. We used the data extracted from CCC DATAset2011[2] for the attack data, and used the packets obtained on the campus network for the normal data, performed cross-validation and calculated the accuracy. From the experimental results, SCW can be classified with the same accuracy, precision, and recall as SVM, and thus it is considered that sufficient accuracy can be obtained.

3. Structural Change Detection

This chapter describes the structural change detection method that is related to research. The problem setting of structural change detection and three methods to solve it are described.

3.1. Definition of Structural Change Detection

Structural change detection refers to detecting changes in structural relationships among variables of input data. In this paper, structural changes are replaced with correlation changes. As a known sample, consider sample D of multivariate data with the number of data N and dimension number M. Also, as an unknown sample, consider a sample D' of multivariate data with the number of data N' and the number of dimensions M.

$$D = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_N\} \quad (1)$$
$$\boldsymbol{x}_i = \{x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(M)}\}$$
$$D' = \{\boldsymbol{x}'_1, \boldsymbol{x}'_2, \ldots, \boldsymbol{x}'_i, \ldots, \boldsymbol{x}'_{N'}\} \quad (2)$$
$$\boldsymbol{x}'_i = \{x'^{(1)}_i, x'^{(2)}_i, \ldots, x'^{(M)}_i\}$$

At this time, the correlation change is realized by obtaining the difference of the correlation obtained from each of the two samples D and D'. Using such a method, there is an advantage that the abnormal part can be identified from the change of the correlation which shows not only the change of a single variable but also the relation between variables. In this paper, we show how to represent correlation under the assumption of normal distribution $\mathcal{N}$ (formula (3)). Assuming that each data x is normally distributed, the covariance matrix Σand the average μ are calculated. In addition, to focus only on the correlation between the variables, not the average, the Z-score is calculated by standardizing the average to 0.

$$\mathcal{N}(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \quad (3)$$

However, covariance, which is an element of the covariance matrix Σ , does not show a direct correlation between two variables and may appear to be high due to the influence of other variables. Therefore, by using direct correlation, correlations other than two variables are eliminated. Thus, a direct correlation can be obtained by using the inverse of the covariance matrix. The inverse of the covariance matrix is called the precision matrix, which is denoted by Λ.

$$\Lambda \equiv \Sigma^{-1} \quad (4)$$

3.2. A Method by Experience Distribution (Maximum Likelihood Estimation)

Assume we have a normal distribution and the precision matrix is the inverse of the covariance matrix, the maximum likelihood estimation method uses the covariance as the empirical distribution (6) as the pair of each variable. We can calculate the precision matrix by finding the covariance matrix and finding its inverse.

$$\Sigma = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \cdots & \Sigma_{1,M} \\ \Sigma_{2,1} & \Sigma_{2,2} & \cdots & \Sigma_{2,M} \\ \vdots & & \ddots & \vdots \\ \Sigma_{M,1} & \Sigma_{M,2} & \cdots & \Sigma_{M,M} \end{pmatrix} \quad (5)$$

$$\Sigma_{j,k} = \sum_{i}^{N} \frac{(x_i^{(j)} - \mu^{(j)})(x_i^{(k)} - \mu^{(k)})}{N} \quad (6)$$

The empirical distribution method can be calculated easily, but it cannot be calculated if the matrix is not a regular matrix and each element of the precision matrix is not completely zero. Hence, there is no correlation in any element and thus, there is a problem that it is unclear whether there is an element for which there is a correlation. To solve that, there are other methods, such as a method to obtain precision matrix Λ by graphical lasso and density ratio estimation.

3.3. Graphical Lasso

The empirical distribution method was the maximum likelihood estimation method. On the other hand, graphical lasso[3] is not a method of maximum likelihood estimation, but is a method that is used to obtain the accuracy matrix by performing maximum a posteriori estimation from the prior distribution. To calculate by adding L1 regularization term by using Laplace distribution for prior distribution, the matrix can be returned as a sparse result, and it can be

calculated by the empirical distribution that can be calculated even if the matrix is not regular. There is an advantage that it is easy to determine the presence or absence of correlation to solve the following problem.

If we write the formula of the precision matrix Λ^*, we want to estimate as a log-likelihood maximization problem, which can express it as the logarithm of the product of likelihood and Laplace distribution $p(\Lambda)$.

$$\Lambda^\star = \arg\max_{\Lambda} \left\{ \ln p(\Lambda) \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{x}^{(n)}|\boldsymbol{0}, \Lambda) \right\} \quad (7)$$

Since it is a maximization problem, ignoring the irrelevant constants and organizing them as a function, the matrix Λ^* obtained by this equation becomes the precision matrix.

$$\Lambda^\star = \arg\max_{\Lambda} \ f(\Lambda; \boldsymbol{S}, \rho) \quad (8)$$

$$f(\Lambda; \boldsymbol{S}, \rho) \equiv \ln\det\Lambda - (\boldsymbol{S}\Lambda) - \rho\|\Lambda\|_1 \quad (9)$$

where S is the empirical covariance matrix and det is the determinant.

To solve the maximization problem, the gradient method is repeated using derivative values of $f(\Lambda; S, \rho)$

Using the two-sample precision matrices Λ and Λ' obtained by graphical lasso, calculate the outliers of another sample by assuming that one sample is normal data. Use Kullback-Leibler divergence (formula (10)) with two samples as outliers.

$$\mathrm{KL}(p||p') = \int_{-\infty}^{\infty} p(\boldsymbol{x}) \ln \frac{p(\boldsymbol{x})}{p'(\boldsymbol{x}')} \mathrm{d}\boldsymbol{x} \quad (10)$$

By fitting a normal distribution as the empirical distribution, it can be calculated as follows. S denotes the covariance matrix of sample D.

$$a_i = \frac{1}{2} \ln \frac{\Lambda_{i,i}}{\Lambda'_{i,i}} - \frac{1}{2} \left\{ \frac{[\Lambda S \Lambda]_{i,i}}{\Lambda_{i,i}} - \frac{[\Lambda' S \Lambda']_{i,i}}{\Lambda'_{i,i}} \right\} \quad (11)$$

3.4. Density Ratio Estimation Method

When a structural change is to be determined by empirical distribution and graphical lasso, it is necessary to determine the precision matrix of each of the two samples and calculate their difference $\Theta \equiv \Lambda - \Lambda'$ but the density ratio estimation is more accurate. In [4], a method has been proposed which is applied to the estimation of matrix difference Θ.

Density ratio estimation refers to a method that directly estimates the ratio of two probability density functions. Let the two samples be $D = \{x_1, \ldots x_N\}$ and $D' = \{x'_1, \ldots x'_N\}$, their probability density functions are defined as $p_D(x)$, and $p_{D'}(x)$, respectively. Here, when the ratio of the probability density functions is defined as in equation (12), an estimation method for directly obtaining the density ratio r(x) is used to estimate the density ratio.

Usually, when calculating the ratio of the probability density functions of two samples, first, from each sample, the probability density functions $p_D(x)$ and $p_{D'}(x)$ are estimated and from the two estimated probability density functions, the ratio r(x) is obtained. When this method is adopted, there is a problem that the ratio changes significantly, especially when an error occurs with the value of denominator close to zero. Therefore, direct estimation of the ratio has the advantage that no numerical change occurs when calculating such a ratio.

$$r(\boldsymbol{x}) = \frac{p_D(\boldsymbol{x})}{p_{D'}(\boldsymbol{x})} \quad (12)$$

When the density ratio estimation method is applied to the precision matrix, the ratio of the normal distribution (formula (13) is proportional to the difference of the precision matrix (formula (14)). Estimate directly the ratio of the normal distribution using

$$r(\boldsymbol{x}) = \frac{\mathcal{N}(\boldsymbol{x})}{\mathcal{N}'(\boldsymbol{x})} \quad (13)$$

$$r(\boldsymbol{x}) \propto \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathsf{T}}(\Lambda - \Lambda')\boldsymbol{x}\right) = \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathsf{T}}\Theta\boldsymbol{x}\right) \quad (14)$$

In the density ratio estimation, the density ratio is modeled and equation (12) is transformed to a modified equation, equation (15), to the problem of minimizing the distance between the left side and the right side Find the density ratio r(x).

$$r(\boldsymbol{x})p(\boldsymbol{x}) = p'(\boldsymbol{x}) \quad (15)$$

To find the precision matrix, consider a density ratio model such as equation (16) with parameters Θ and sample data x'.

The denominator is present to normalize the model value to be $0 \le r_\Theta(x) \le 1$.

$$r_\Theta(\boldsymbol{x}) = \frac{\exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathsf{T}}\Theta\boldsymbol{x}\right)}{\int p'(\boldsymbol{x}') \exp\left(-\frac{1}{2}\boldsymbol{x}'^{\mathsf{T}}\Theta\boldsymbol{x}'\right) \mathrm{d}\boldsymbol{x}'} \quad (16)$$

Here, since r(x) p (x) and p'(x) are probability density functions, the Kullback-Leibler divergence can be used to find the distance of the probability density function. Use (Expression (10)) to minimize

the distance between two functions. If we approximate the integral calculation to an empirical distribution and apply r(x) p (x), p'(x) to the Kullback-Leibler divergence, the minimization formula is given as 17), where R is a parameter that adjusts the sparsity.

$$\min_{\Theta} \frac{1}{2N}\sum_{n=1}^{N} \boldsymbol{x}_n^{\mathrm{T}}\Theta\boldsymbol{x}_n + \ln\frac{1}{N}\sum_{n'=1}^{N'}\exp\left(-\frac{1}{2}\boldsymbol{x}_{n'}'^{\mathrm{T}}\Theta\boldsymbol{x}_{n'}'\right) \quad (17)$$
$$\text{subject to } ||\Theta||_1 \leq R$$

In graphical lasso, Kullback-Leibler divergence can be used to calculate the degree of anomaly between two samples by the calculated precision matrix Λ and covariance matrix S, but density ratio estimation is used to calculate differences directly. The degree of anomaly cannot be calculated with the same definition.

4. A Proposed Method by Structural Change Detection

This section outlines the method proposed in this paper and describes its conFigureuration, attribute extraction, and its application to time-series data.

4.1. Proposal of Anomaly Detection Method by Structural Change Detection

Anomaly detection research is being conducted to automatically detect anomalies from the subject being monitored. In many methods of anomaly detection, the normal state is modeled using the probability distribution from the data considered to be normal and how far the newly input data is from normal is calculated as the anomaly score value. It realizes to detect by making abnormal when exceeding the defined abnormal score value.

A method called structural change detection has been proposed as one of the anomaly detection researches. Structural change is a method for extracting the correlation between variables inherent in data, and by calculating the breakdown of the correlation as an abnormality score by defining a normal state from the correlation between the variables, it is possible to see what abnormality detection can do.

In this paper, we apply such anomaly detection to the field of network security and confirm that change in communication behavior in the internal network is detected using structural change detection. Among the attacks on the system and the operation of malware, there are some that perform communication inside the network for the purpose of stealing confidential information, spreading infection, sending DDoS attacks as a springboard, sending SPAM mail, etc. It is the purpose of this paper to assume the traffic that has a correlation from the communication in such internal network and to confirm whether abnormal detection can be done by correlation change detection.

4.1. Proposed Method

In this section, we describe an anomaly detection method used in detecting the structural change by graphical lasso and density ratio estimation.

In this paper, we consider structural change as a change in correlation and propose an anomaly detection method by applying the technique of detecting changes in correlation to detecting that the correlation is broken from correlated variables included in time-series data of network traffic.

The conFigureuration of the proposed method is simply shown as a network of computers connected by switching hubs as shown in Figure 4. Switching hubs have devices that implement SNMP and NetFlow, which are protocols for aggregating communication contents and transferring them to the monitoring server and by using these devices, switching from traffic performed by hosts operating inside the network is performed. It is assumed that traffic passing through the hub is collected, aggregated, and analyzed by the monitoring server.

The monitoring server performs anomaly detection in real time, using the transferred aggregate data as input. Anomaly detection is performed by the method of calculating structure change detection by two kinds of algorithm of graphic lasso and density ratio estimation. Herein, we propose to detect the following correlations by structural change detection from the data.

- Correlation between HTTP communication and DNS communication
- Correlation between syn and synack in the TCP header of TCP communication

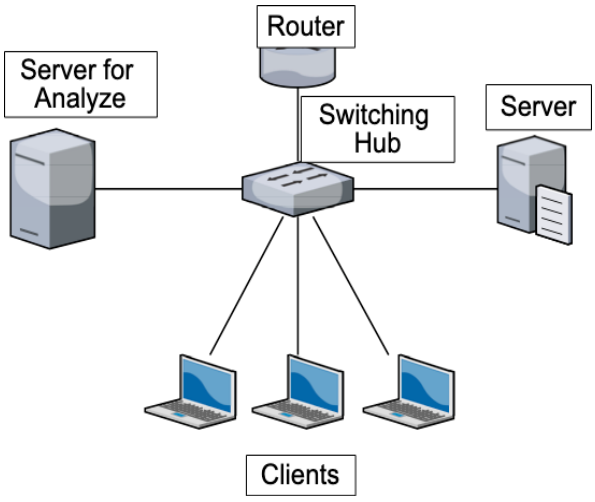

Figure 4. Overview of the Proposed Method

Since HTTP communication and DNS communication query the IP address to DNS when accessing hosts in a specific domain by HTTP, it is expected that there is a correlation between the two communications. We assumed that the correlation would be broken when performing tunneling attack using DNS and communication to multiple domains such as DomainFlux.

With regard to syn and synack included in the TCP header, since two communications are originally communicated as a set of request and response, there is a correlation in the frequency of the connection including the flag every fixed time. In this case, it is assumed that the correlation is broken when an attack is made such as SynFlooding, in which a specific amount of traffic is biased. However, for ack, not only the three way handshake but also the ack flag is sent each time the reception of data is confirmed, so it is counted on synack instead of ack alone, and syn is limited to not only synack alone. We have represented these in a table.

We consider applying correlation change of two samples to time-series data. As shown in Figure 5, the current time is t_0, the time window size s earlier than the current time is t_{-1}, and let the time window size s for the previous time be t_{-2}. The samples from t_0 to t_{-1} are D_1, t_{-1} to t_{-2} are D_2, and the two samples D_1. By replacing it with the problem of finding the difference of the correlation obtained from D_2, we can find the correlation change of the time series data.

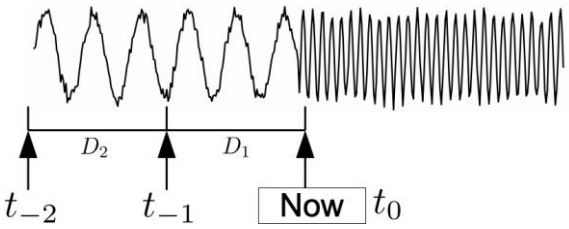

Figure 5. Time Window

In both algorithms, when the difference between time series data exceeds a certain value, an anomaly is detected, and an alert is issued to detect an anomaly. Also, with graphical lassos, the degree of anomaly can be defined by the expected value of the two-sample Kullback-Leibler divergence (formula 11). The degree of anomaly can be calculated using the precision and covariance matrices obtained from each sample. However, since the density ratio estimation method directly calculates the difference of the precision matrix, it cannot be calculated by equation 11. Therefore, in the method of density ratio estimation, the difference of correlation is treated as anomalous degree and in the method of the graphical lasso, the degree of an anomaly by correlation difference and the expected value of Kullback-Leibler divergence is calculated.

A virtual environment is created on a physical machine, normal communication and attack communication are simultaneously generated in the virtual environment, and abnormal values are checked when attack communication occurs. The experimental environment is shown in table 1 and Figure 6. Kernel-based virtual machine (KVM) [5] was used for the virtual machine and Open vSwitch [6] was used for the virtual network.

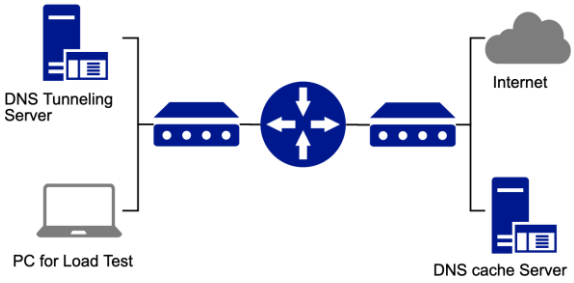

Figure 6. Experimental Environment

As normal communication, the client PC accessed an external web server. We used a DNS cache server that exists in another network for name resolution of a web server. From the client PC, Apache JMeter[7], which was originally used for server load testing, is implemented by continuing a certain amount of HTTP communication. As for the domain name to be sent, HTTP was randomly sent from the famous 1 million domain list [8] provided by Cisco, and a DNS cache server was used in the process. For attack communication, HTTP communication was performed on DNS using the DNS tunneling tool, iodine. At this time, to assume communication to the external network, the network was divided by routers. These communications were captured using tcpdump [9] as analysis data. The captured data is shown in Figure 3 in [10]. Details of the experimental results are given in [10]. The horizontal axis shows the time, and the vertical axis shows the frequency of each second. Also, Figure 5 in [10] shows Z scores calculated assuming normal distribution of frequencies respectively. By calculating the z-score, it can be seen that DNS communication rises every 80 seconds.

Table 1. Experimental Environment(Software)

Purpose	Software
Virtual Machine	KVM
Virtual Network	Open vSwitch
Client OS	Ubuntu 18.04
Router OS	Vyos
HTTP communication	Apache JMeter
DNS tunneling tool	iodine

We use graphical lasso and density ratio estimation to detect the structural change of such

analyzed data and confirm that we can extract attack communication. The difference in the correlation by graphical lasso is shown in Figure 6 in [10] and the difference in the correlation by density ratio estimation is shown in Figure 7 in [10]. The horizontal axis indicates the time and the vertical axis is a graph in which the difference between two samples of the correlation value calculated by each algorithm is plotted for each pair of two variables.

In the graphical lasso method, it is clear that the correlation difference changes in the communication when an attack occurs only between the communication of DNS request and response. Based on density ratio estimation, in this method, although there is a change in the correlation difference among all variables, it can be seen that the change in the correlation difference at the time of attack occurrence is large. Thus, an anomaly can be detected because the difference in correlation is large.

The Figure 8 in [10] shows the change of outliers for each variable using the graphical lasso calculated by the definition of outliers. From the result of the graphical lasso, we observe that the variable of the HTTP response is largely changed other than the time of attack. It has been confirmed that DNS tunneling can be detected by the change of outliers of graphical lasso and the correlation change of density ratio estimation.

In this experiment, we used HTTP requests, HTTP responses, DNS requests, and DNS responses as variables to test whether anomalies can be detected by graphical lasso and density ratio estimation. From the results, DNS requests and DNS response outliers increase at the time of attack when compared by variables, it is confirmed that DNS tunneling can be realized by focusing attention on these variables.

We assumed that there is a correlation between HTTP communication and DNS communication as a premise, and a method using the fact that the correlation is broken between HTTP communication and DNS communication when a change occurs only in DNS communication like DNS tunneling. Although the correlation change is performed only by DNS autocorrelation, it is considered that the outliers increase.

One of the causes is that compared to the time when the correlation in DNS tunneling is broken, the correlation between HTTP communication and DNS communication in normal communication is not so different as detectable. It is considered that the assumption that there is a correlation due to the influence of other protocols such as IMAP was not achieved. Therefore, the change in the frequency of DNS communication is a major factor in the detection of DNS tunneling, and as a future task, it is necessary to confirm how much traffic volume can be detected in DNS tunneling. Also, as a DNS related technology in recent years, DNS over HTTPS, which is a protocol that performs name resolution by performing DNS communication on HTTPS, DNS flooding that stops the function of DNS, dynamically generates a domain and blacklists the domain. There are regular protocols and attack methods that are different from conventional DNS such as DomainFlux, which is a method to avoid this. An attack method can be detected as an attack as well as DNS tunneling, but as for DNS over HTTPS, if all name resolution is performed with DNS over HTTPS, DNS communication is completely lost, so use correlation change. However, there is a problem that the method cannot be used.

8. Conclusions and Future Works

The main purpose of this paper is to detect the change of traffic behavior in the network by machine learning method and statistical method. First, we propose a system using an online learning method that applies the kernel method to the intrusion detection problem as a method using the machine learning method. Next, we propose a network anomaly detection method using the structure detection method. In this network anomaly detection method using the structure detection method, we aimed at the analysis of the traffic that breaks the correlation, the analysis of the attack, and the confirmation whether it could be detected by the structure change detection.

First, it is proposed to use online learning using the kernel method among the machine learning methods as a detection method. As a result, when new case data is obtained, if the case data can be learned by additionally inputting the model to the models learned so far, the burden of learning including past data can be reduced. SCW was used as the online learning method, and the kernel method was applied.

In the experiment, SVM and SCW were compared. We use data extracted from CCC DATAset2011 for attack data, the accuracy was calculated by performing cross-validation using packets obtained on the campus network for normal data. From the experimental results, SCW can be classified with the same accuracy, precision, and recall as SVM, and thus it is considered that sufficient accuracy can be obtained. The future perspective is computational cost. Since the kernel method uses a model based on the inner product with past data, when the number of data is increased, a large computational cost is required. Therefore, it is necessary to simplify the calculation by integrating the past data and not saving the data.

The Gaussian kernel is used as the kernel function. The idea is to define a kernel function specialized for online learning. Prospects for improvement include future design of kernel functions.

Second, we proposed a method to detect network anomalies by changing the correlation. We implemented the proposed method by graphical lasso and the method of detecting the change of correlation

by density ratio estimation and confirmed whether it could detect SynFlooding attack and DNS tunneling attack through experiments for actual communication. As a result, we confirmed the characteristics of graphical lasso and density ratio estimation and also detected the problem of network anomaly detection by detecting changes in the time-series correlation. Graphical lasso and density ratio estimation, which are algorithms used in the proposed method, can detect changes in correlation among variables in multiple variables, so it can confirm whether changes in multiple variables are detected. In that case, it is thought that it becomes an issue to examine which value is adopted as an abnormal value. Since it is necessary to consider which value of the anomaly score is considered as an anomaly, we confirmed in this paper whether it could be detected by visualizing it in a graph, but from the viewpoint of automatic detection, we think that it is necessary to set a threshold and to issue an alert by probability statistics.

Moreover, correlation change detection can be used to extract the relationship between numbers, and it is considered to be one of the many network anomaly detections in many network security by understanding whether the relationship between variables is broken.

9. References

[1] Hoi, S. C. H., Wang, J. and Zhao, P.: Exact Soft Confidence-Weighted Learning., ICML, icml.cc / Omnipress (2012).

[2] Mitsuhiro Hatada, Mitsuaki Akiyama, Takahiro Matsuki, Takahiro Kasama, "Empowering Anti-malware Research in Japan by Sharing the MWS Datasets," IPSJ Journal of Information Processing, Vol.23, No.5, pp.579-588, Sep. 2015.

[3] J. Friedman, T. Hastie, and R. Tibshirani, “Sparse inverse covarianceestimation with the graphical lasso,” Biostatistics, vol. 9, no. 3, pp. 432–441, Jul. 2008. [Online]. Available: http://dx.doi.org/10.1093/biostatistics/kxm045

[4] S. Liu, J. A. Quinn, M. U. Gutmann, and M. Sugiyama, “Direct learning of sparse changes in markov networks by density ratio estimation,” in Machine Learning and Knowledge Discovery in Databases, H. Blockeel, K. Kersting, S. Nijssen, and F. Zelezny, Eds., Sep. 2013, pp. 596–611.

[5] KVM. KVM. https://www.linux-kvm.org/page/Main_Page. (Access Date:2019/01/08).

[6] O. vSwitch. Open vSwitch. https://www.openvswitch.org/. (Access Date:2019/01/08).

[7] Apache. Apache JMeter - Apache JMeter™. https://jmeter.apache.org/. (Access Date: 2019/01/08).

[8] Cisco. Cisco Popularity List. http://s3-us-west-1.amazonaws.com/umbrella-static/index.html. (Access Date: 2019/01/08).

[9] Tcpdump. Tcpdump/Libpcap public repository. http://www.tcpdump.org. (Access Date: 2019/01/08).

[10] Ayahiko Niimi: Data Anonymization Using Imbalance-d Data for Deep Learning. World Congress on Internet Security (WorldCIS-2018), Church College, University of Cambridge, Cambridge, UK, 5pages, 2018.

10. Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP17K00310.

