International Journal of Networked and Distributed Computing, Vol. 2, No. 1 (January 2014), 45-53

An IDS Visualization System for Anomalous Warning Events

Satoshi Kimura [1] **Hiroyuki Inaba** [2]

[1] *Computer Science, Kyoto Institute of Technology,*
6068585 , Kyoto
E-mail: kimura08@sec.is.kit.ac.jp

[2] *Computer Science, Kyoto Institute of Technology,*
6068585 , Kyoto
E-mail: inaba@kit.ac.jp

Received 18 February 2013
Accepted 19 November 2013

Abstract

Intrusion Detection System(IDS) has received attention to deal with the illegal access to the network. However, IDS has a critical problem which outputs a tremendous number of logs. Analyzing these logs apply a large amount of load to a network manager. In this paper, we propose a novel visualization system of IDS considering order relation of IP addresses that emphasize the anomalous warning events based on past tendency.

Keywords: IDS, Visualization, Log analysis, IP address

1. Introduction

Recently, as Internet is developing explosively, illegal access to the network is increasing. To deal with this serious problem, the necessity of Intrusion Detection System(IDS) is increasing. IDS is the notifying system of network manager to inspect symptoms of the illegal access. IDS also enables us to early detect threatening attack to the computers and to deal with its attacks. However, there exists a problem of IDS. While IDS generally outputs the text warning logs, the amount of these logs are tremendous especially for large scale network. Analyzing these logs apply a large amount of load to a network manager. To overcome this problem, there exist several methods for analyzing logs [1–3] and some visualization methods for the logs [4–7].

In this paper, we propose a novel visualization system of IDS considering order relation of IP addresses that emphasize the anomalous warning events based on past tendency. Visualization on emphasizing anomalous warning events enable network manager to reduce the analyzing load than conventional methods.

We utilize Snort [8, 9] as IDS. Snort is free and open source software that is generally classified as a signature match IDS.

2. Related Work

Research on analyzing network packets and its visualization have been studied intensely for many years. In this chapter, we introduce several works.

2.1. Works on log analysis

We introduce several works on log analysis of IDS.

There exists a study on log analysis based on tendency of IDS alert events [1]. They have shown that the appropriate intensive terms and an alert threshold can be automatically determined by coefficient of variance. A part of our study on log analysis is based on this method.

In [2], they pay attention to the frequency of detections of events in an unit time, and they newly introduce three parameters. Those parameters are frequency of detections of events in an unit time, arrival time and event vast length of each events. As a result, the frequency of detections can be represented by Poisson distribution, and arrival time and event vast length can be represented by exponential distribution. Therefore, they confirm that each event randomly detected at a glance can be represented by theoretical statistical distribution. They also say that the detected events out of these theoretical statistical distributions are regarded as anomalous events.

In [3], a log analysis method of paying attention to variation of detected events and its frequency in IDS logs are proposed. They plan to detect anomalous events from the number of detected events. To do their aim, they use frequency analysis method which pays attention to the variation of the number of detections, and a ratio analysis method which uses the average and a standard deviation for each event. As a result, they could specify anomalous warning events of tremendously increasing events even though the number of detections is small.

2.2. *Work on visualization*

In the following, we will explain Hashing Alert Matrix [4], HeiankyoView [5] and IDS RainStorm [6] as log visualization system.

Hashing Alert Matrix [4] is the visualization method of IDS data which use a hash function. In the method, IP address of each event is not plotted as the raw value, but it is plotted as the hashed value. By using the hashed value, it can use a limited drawing area efficiently. However, the hierarchical structure of IP address is disappeared.

HeiankyoView [5] is the visualization method of network intrusion detection data which pay attention to hierarchical structure of IP address. Since it can arrange all detected data on twodimensional plain smartly, we can grasp the situation of detections at a glance.

IDS RainStorm [6] is the visualization method of IDS data which is specialized in displaying class B of IPv4. It also has the property that we can grasp relation of source and destination IP address along time axis.

3. Visualization system considering order relation of IP addresses that emphasize the anomalous warning events

In this chapter, we propose a new method of visualization system considering order relation of IP addresses that emphasize the anomalous warning events by using the improved version of our proposed method for log analysis [1], and UnEqual Scaling Alert Matrix [7].

3.1. *Log analysis method using the coefficient of variance*

We have studied the log analysis method using the coefficient of variance without relying on a network manager subjectivity [1]. First, we compare the number of detections N_1 during the most recent term T_1 with the number of detections N_2 during the previous term T_2 for each event. The term T_1, T_2 are called intensive terms. If a certain ratio F is greater than a certain threshold T_h, then a warning message is outputted. We have defined a ratio of the number of detection F as following formula.

$$F = \frac{(N_1 - N_2)}{N_2} \tag{1}$$

The length of the intensive term is dynamically decided by the past three months log data. The conceptual image of the method is shown in Fig.1.

Published by Atlantis Press

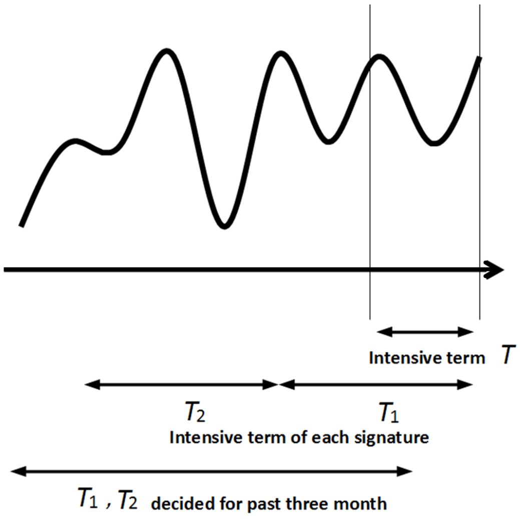

Intensive term T
T_2
T_1
Intensive term of each signature
T_1 , T_2 decided for past three month

Figure 1: Conceptual image of deciding intensive term.

To decide the length of the intensive term, we first calculate a standard deviation for each event. Since an average number of detection is considerably different for each warning event, a standard deviation itself is not suitable for deciding the intensive term. Therefore, we use a coefficient of variance which is a value of standard deviation divided by average number of detection. It is suited for a characteristic measure for deciding the intensive term which is not dependent on the number of each warning event. We decide coefficient of variance $C_v(T)$ for intensive term T as following formula.

$$C_v(T) = \frac{S_d(T)}{\overline{X}(T)} \tag{2}$$

In the formula, $S_d(T)$ is a standard deviation, and $\overline{X(T)}$ is an average number of detections. Generally, it is expected that a variation of coefficient of variance decrease with increasing intensive term.

We define a rate of change $R_d(T)$ associated with intensive term T as following formula.

$$R_d(T) = \left| \frac{C_v(T+1) - C_v(T)}{C_v(T)} \right| \tag{3}$$

The intensive term T is decided such that the rates $R_d(T)$ *reach* R_{th}, because $C_v(T)$ gradually decrease with increasing T like Fig.2. Where R_{th} is a threshold value, 0.01 is used in [1].

Since a value of coefficient of variance indicates relative dispersion of each warning event, it is expected that a large threshold value T_h is required for a large coefficient of variance. Therefore, we define the threshold T_h as following formula.

$$T_h = \delta \times C_v(T) \tag{4}$$

The parameter δ is a constant value which does not depend on a kind of warning events. We also define $N_A(\delta)$ as the number of anomalous warning events exceeding T_h.

As a result of experience, the value of coefficient of variance converges in a certain value. By increasing δ every 0.1 used in Eq.(4), we could confirm that the number of the detections exceeding threshold value steadily decreased.

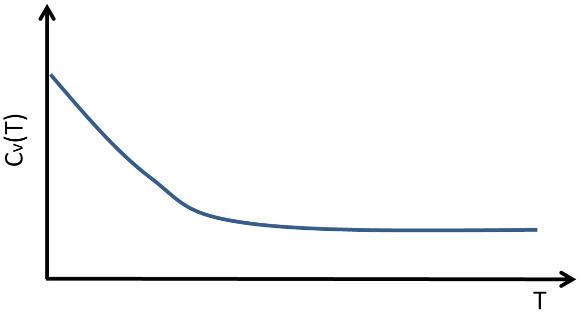


Figure 2: Conceptual image of relationship between $Cv(T)$ and T

A best of features in [1] is that the threshold of each signature is dynamically determined by coefficient of variance without a network manager's subjectivity. However, it should be noted that the method make an assumption that the number of detections for deciding intensive term T is stable.

Meanwhile, if the value of N_1 is lower than N_2 in Eq.(1), we may also define the situation as anomalous one. But we consider this case as non-anomalous warning event in this study.

3.2. Methods of improved UnEqual Scaling Alert Matrix

In order to grasp the information of source and destination IP address of each warning event at single display, let us prepare the two-dimensional plain having a horizontal axis as source IP address and a vertical axis as destination IP address.

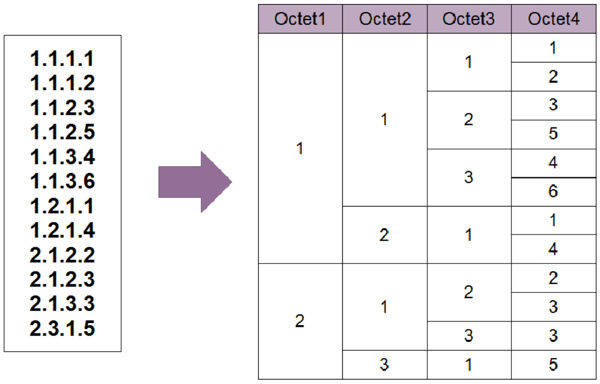
However, this method has the fatal problem. We can not grasp the situation of warning events correctly because many different points are plotted on the same point of a monitor having general resolution about 1024 pixel × 1024 pixel. To overcome this problem, the authors have proposed UnEqual Scaling Alert Matrix [7]. In the method, all 32bit IP address is unequally plotted according to the ratio of the number of source and destination IP block. An outline of this method is as following.

1. The data of logs for certain fixed interval are sorted in ascending order on source or destination IP address.

2. The sorted data are classified by the first octet of IP address. After that they are hierarchically classified by the second octet and the third octet in the same way(Fig.3).

3. A plotted area of display is divided based on the ratio of each classified octet. Therefore, an IP block which is observed more frequently have a wider plotted area(Fig.4).

It is noted that the order of IP address is preserved in this method. Therefore, a network manager can easily grasp the situations of several attacks such as network scan which is observed under some IP blocks.



1.1.1.1
1.1.1.2
1.1.2.3
1.1.2.5
1.1.3.4
1.1.3.6
1.2.1.1
1.2.1.4
2.1.2.2
2.1.2.3
2.1.3.3
2.3.1.5

Octet1	Octet2	Octet3	Octet4
1	1	1	1
			2
		2	3
			5
		3	4
			6
	2	1	1
			4
2	1	2	2
			3
		3	3
	3	1	5

Figure 3: Classification of observed IP address.

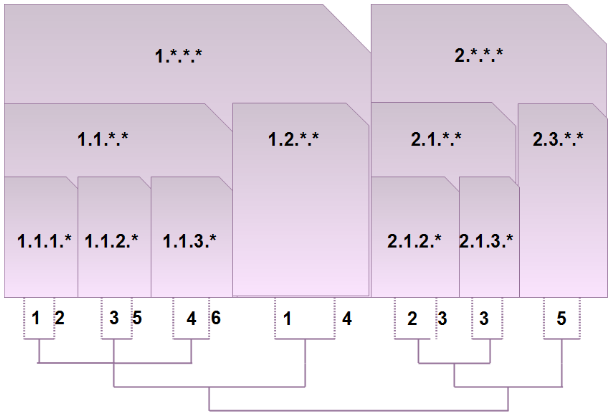


Figure 4: Unequal scaling based on the ratio of each classified octet.

In order to emphasize anomalous warning events, they are plotted by more remarkable points than other normal events. The anomalous warning events are decided by the log analysis method using coefficient of variance denoted in section 3.1. In our implementation, the anomalous warning events are plotted by 5 × 5 dots larger red points and the normal warning events are plotted by 3 × 3 dots small points. Furthermore, priority 1's warning events are plotted in black color, and priority 2's warning events are plotted in green color. Also priority 3's are plotted in blue color. Where, priority 1-3 are attack risk level which are set for each Snort's signature. By this color classification, it is expected that a network manager can pay attention to each warning event appropriately. In order to grasp temporal change of detected warning events, we also implement animation mode which can continuously draw the situation every one hour.

Figure 5 shows a display image of the visualization system considering order relation of IP addresses that emphasize the anomalous warning events. Figure 6 is a magnified image of upper left area of Fig.5 for the sake of reader's detailed observation.

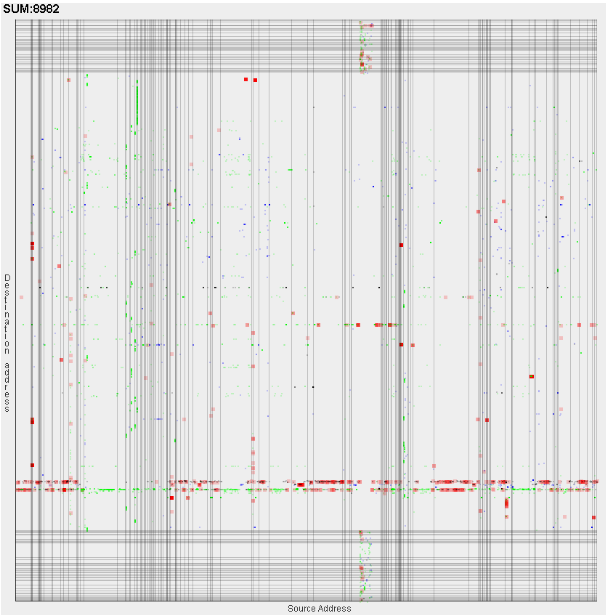


Figure 5: UnEqual Scaling Alert Matrix which emphasize anomalous warning events.

An area of drawing region of Fig.5 is 800 × 800 dots, and a number of detections which are plotted is 8982. To easily grasp hierarchical structure of IP address, the scale of the first octet of source and destination IP address is drawn. When a network manager click to the plotting point that he pay attention to, the detailed information of the warning events is displayed on a terminal. When two or more plotting points overlap each other, all detailed information for each warning event are displayed on a terminal (Fig.7).

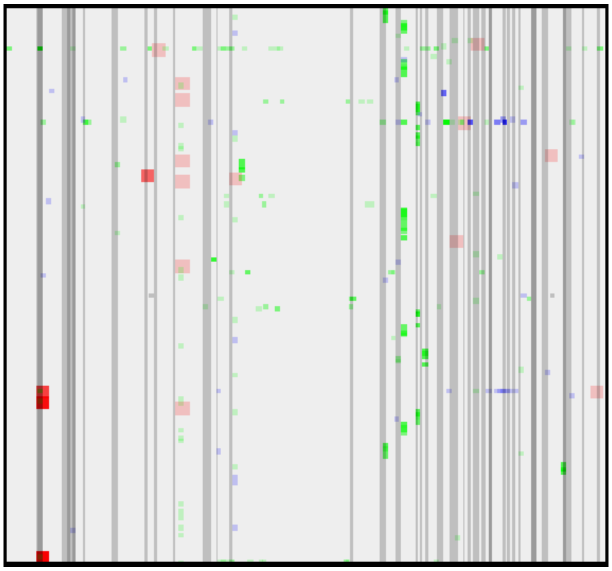
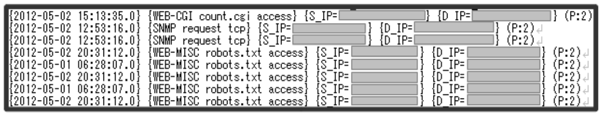


Figure 6: Magnified image of upper left area of Fig.5.



```
{2012-05-02 15:13:35.0} {WEB-CGI count.cgi access} {S_IP=[redacted]} {D IP=[redacted]} (P:2)
{2012-05-02 12:53:16.0} {SNMP request tcp} {S_IP=[redacted]} {D_IP=[redacted]} (P:2)
{2012-05-02 12:53:16.0} {SNMP request tcp} {S_IP=[redacted]} {D IP=[redacted]} (P:2)
{2012-05-02 20:31:12.0} {WEB-MISC robots.txt access} {S_IP=[redacted]} {D_IP=[redacted]} (P:2)
{2012-05-01 06:28:07.0} {WEB-MISC robots.txt access} {S_IP=[redacted]} {D_IP=[redacted]} (P:2)
{2012-05-02 20:31:12.0} {WEB-MISC robots.txt access} {S_IP=[redacted]} {D_IP=[redacted]} (P:2)
{2012-05-01 06:28:07.0} {WEB-MISC robots.txt access} {S_IP=[redacted]} {D_IP=[redacted]} (P:2)
{2012-05-02 20:31:12.0} {WEB-MISC robots.txt access} {S_IP=[redacted]} {D_IP=[redacted]} (P:2)
```

Figure 7: Detailed information on terminal.

4. Evaluation of proposed methods

First, we show experiment condition for evaluating the proposed methods. Secondly, we show the experimental result about the number of anomalous warning events, and show the display image of the visualization system which emphasize anomalous warning events.

4.1. Experiment condition

The proposed visualization system is implemented by Java programming language. The detected data for evaluation is obtained from our campus network(IPv4 B-class network). The term of getting data is from 2012/5/1 through 2012/5/7. The term for deciding intensive term T is three months for 2012/2/1 through 2012/4/30. These three months are learning term to decide intensive term T for each

Published by Atlantis Press

Table 1: Detailed information for the major target signatures.

Signature No.	$\overline{X(1)}$	Intensive term T	$C_v(T)$	Number of detections				
				$\delta = 1.1$	$\delta = 1.2$	$\delta = 1.3$	$\delta = 1.4$	$\delta = 1.5$
7	3.8	10	0.182	68	56	42	29	23
8	32.4	11	0.161	577	542	385	358	307
11	5.7	7	1.558	639	631	611	601	601
22	2.2	15	1.279	449	447	424	409	408

signatures. We use the discrete value of Intensive term T per hour. Although the detected data is obtained from IPv4 B-class network in this experiment, our proposed method is also available for other class network.

4.2. *Evaluation of the number of anomalous warning events*

Fig.8 is the relationship between the number of anomalous warning events and the parameter δ defined in Eq.(4).

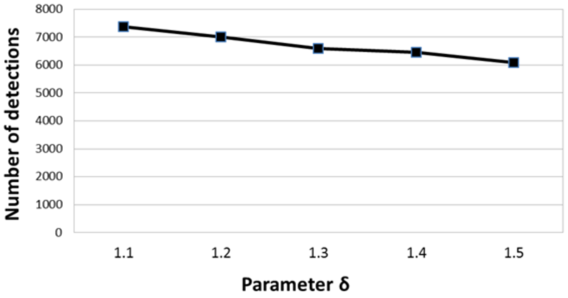


Figure 8: Number of anomalous warning events.

In Fig.8, we can see that the number of detections decreases with increasing the parameter δ. On deciding anomalous warning events, we previously exclude the event signatures detected less than one per hour on the average. Because it is considered that the rare event signatures are meaningless in statistics. If the rare events are detected, we can consider the events anomalous. After excluding the rare events, we call the remaining signatures target signatures. There exist 25 target signatures in our experiment.

We show some examples of target signatures in Table.1. It is considered that these target signatures are suitable as comparable subject. Because we pay attention to the following points.

- Coefficient of variance $C_v(T)$
- A ratio of anomalous warning events decreasing with a value of parameter δ
- Average detections
- Intensive term T

$\overline{X(1)}$ in Table.1 means the number of average detections per hour for the target signatures. $C_v(T)$ is the value of coefficient of variance for the intensive term T decided in Eq.(3).

The event name correspond to signature number in Table.1 are described in Table.2. We define the decreasing ratio $D_r(\delta)$ in the following formula to easily grasp a ratio of the number of anomalous warning events.

$$D_r(\delta) = 100 \cdot \frac{N_A(\delta)}{N_A(1.1)} \tag{5}$$

The decreasing ratio $D_r(\delta)$ for signature number 7, 8, 11, 22 in Table.1 are shown in Fig.9.

Table 2: Name of the major target signatures.

Name of signature
7:ATTACK-RESPONSES 403 Forbidden
8:WEB-MISC robots.txt access
11:DNS SPOOF query response with TTL of 1 min. and no authority
22:SNMP request tcp

Published by Atlantis Press

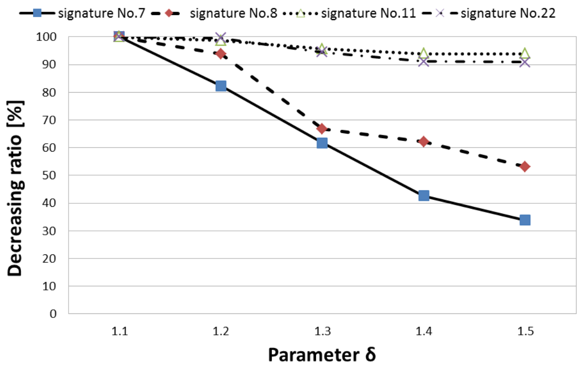


Figure 9: Decreasing ratio $D_r(\delta)$ of signature No.7, 8, 11, 22.

In Fig.9, the number of anomalous warning event having small value of coefficient of variance such as signature number 7, 8 tremendously decreases with increasing the parameter δ. In contrast, the number of anomalous warning event having large value of coefficient of variance such as signature number 11, 22 hardly decreases with increasing the parameter δ. It is considered that the signatures having small value of coefficient of variance are detected stably, and the signatures having large value of coefficient of variance are detected unstably for the past three months.

We show the time variance of the number of detections for signature number 8 and 11 every 1 hour in Fig.10 and 12. We also show a ratio F defined by Eq.(1) for signature number 8 and 11 in Fig.11 and 13.

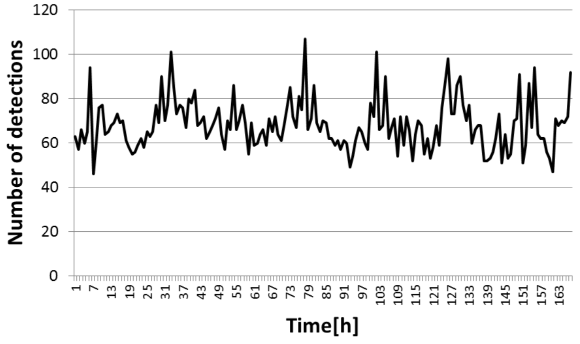


Figure 10: Number of detections for signature number 8.

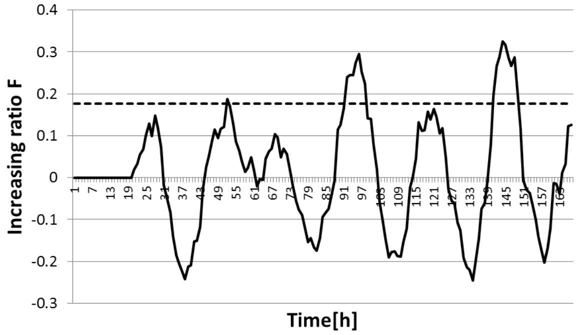


Figure 11: Increasing ratio F of the number of detections for signature number 8.

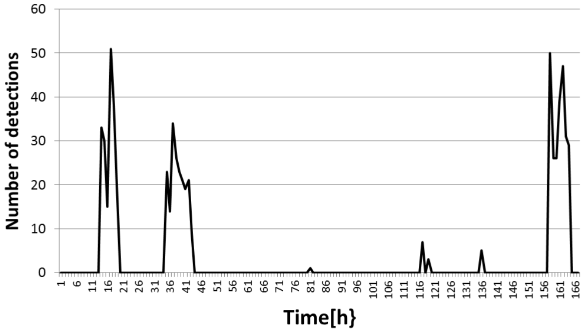


Figure 12: Number of detections for signature number 11.

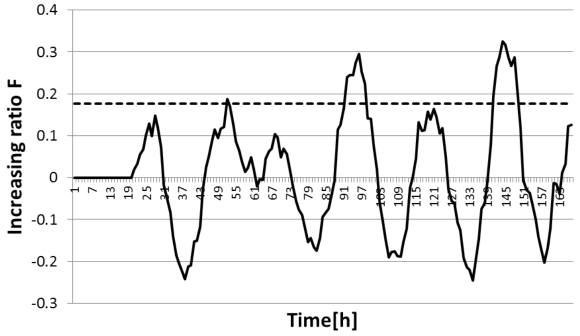


Figure 13: Increasing ratio F of the number of detections for signature number 11.

A broken line in Fig.11, Fig.13 indicates the threshold value T_h defined by Eq.(4) where the parameter δ is set to 1.1.

In Fig.10, we can see that the warning event(No.8) is continuously detected (Note that

No.8's $C_v(T)$ is small). On the other hand, the warning event(No.11) is not continuously detected (Note that No.11's $C_v(T)$ is large). We have confirmed that the other target signatures tend to the same property. These observations show the validity of our previous consideration.

In Fig.11, each part where a solid line exceeds a broken line indicates the anomalous situation. Since the time variance of the number of signature number 8 in Fig.10 is stable, the diagram in Fig.11 changes slowly. On the other hand, since the time variance of the number of signature number 11 in Fig.12 is unstable, the diagram in Fig.13 changes rapidly.

Finally, we show the original UnEqual Scaling Alert Matrix visualization system in Fig.14 and the improved version of the method in Fig.15.

In Fig.14, anomalous warning events are not emphasized and all warning events are plotted with equal shape. On the other hand, in Fig.15, anomalous warning events are emphasized. It is expected that the emphasizing plots pay attention to the anomalous warning events to a network manager.

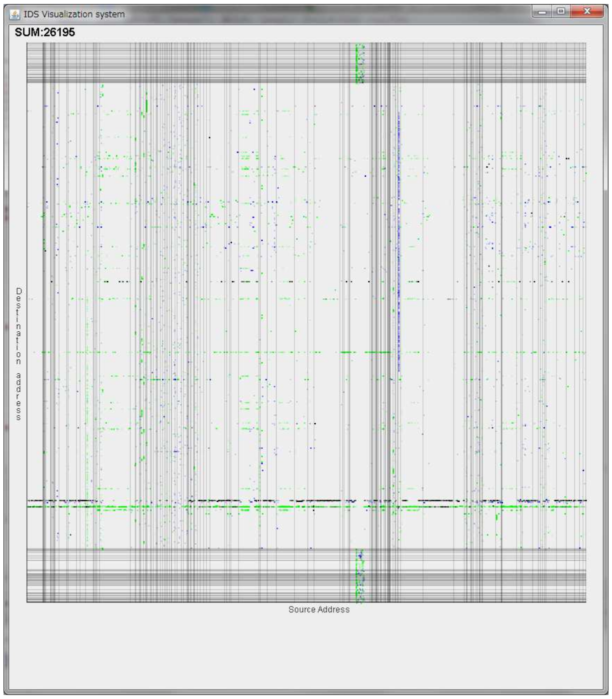


Figure 14: UnEqual Scaling visualization system without emphasizing anomalous events.

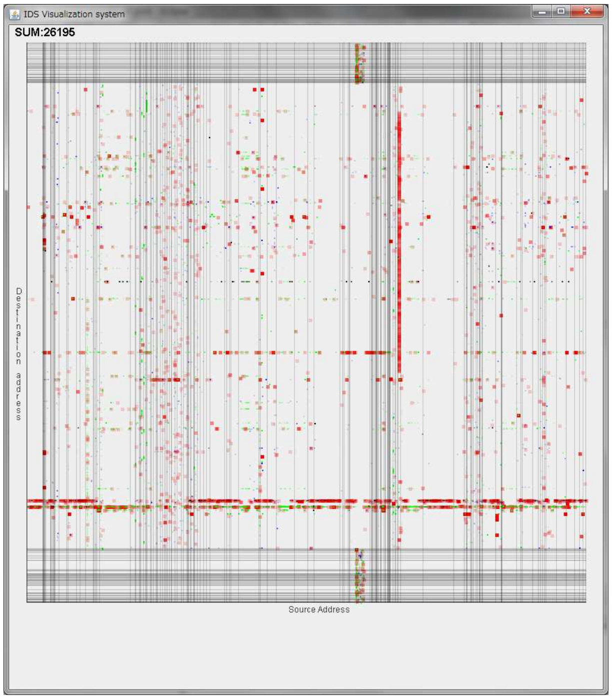


Figure 15: Unequal Scaling visualization system with emphasizing anomalous events.

5. Conclusion and Future Work

In this paper, we propose IDS visualization system which emphasizes the anomalous warning events based on past tendency. The system is intended for reducing a load of a network manager. In the proposed method, we introduce a new algorithm for deciding anomalous warning events. Since the anomalous warning events are plotted as emphasized shape on UnEqual Scaling Alert Matrix, it is easy for a network manager to pay attention the events.

Our proposed method can be applied to IPv4. Even though IPv4 will be replaced by IPv6 in the near future, it is difficult for our proposed method to adapt IPv6 due to extensive address space. Development of an improved method to adapt IPv6 remains as the future work.

Acknowledgments

We thank the anonymous reviewers for their helpful feedback.

References

1. T. Toda, H. Inaba. ”A Study on Log Analysis Based on Tendency of IDS Alert Events”(in Japanese), IEICE Technical Report, SITE2010-7, pp.7-12, Jun. 2010.
2. K. Takemori, Y. Miyake, T. Tanaka, I. Sasase. ”Modeling Techniques about Statistical Theory of Attack Events”(in Japanese), Technical Report of IEICE, vol.103, no.691 pp.20-27, Mar. 2004.
3. K. Takemori, Y. Miyake, K. Nakao, F. Sugaya, I. Sasase. ”A Support System for Analyzing IDS Log Applied to Security Operation Center”(in Japanese), IEICE Trans. A, vol.J87-A, no.6, pp.816-825, Jun. 2004.
4. L. Li, H. Inaba, K. Wakasugi. ”Notes on 2D Visualization Method for IDS that can Distinguish Individual Warning Event”(in Japanese), IIEEJ Journal, vol.40, no.2 pp.369-376, 2011.
5. T. Itoh, H. Takakura, and K. Koyamada. ”Hierarchical visualization of network intrusion detection data”, IEEE Computer Graphics Applications, vol.26, no.2 pp.40-47, March/April. 2006.
6. I.R.V.I. Alarms. ”IDS RainStorm: Visualizing IDS Alarms”,In Proc. IEEE Workshop on Visualization for Computer Security, pp.1-10, Oct. 2005.
7. S. Mizoguchi, H. Inaba. ”Proposal of 3D Visualization Method for IDS Considering Order Relation of IP addresses”(in Japanese), IEICE Technical Report, vol.111, no.125, pp.19-24, July. 2011.
8. ”Snort”, http://www.snort.org/
9. M. Roesch.”Snort: Lightweight Intrusion Detection for Networks”,LISA ’99 Proceedings of the 13th USENIX conference on System administration, pp.229-238, Nov. 1999.