EFFECTIVENESS AND WEAKNESS OF QUANTIFIED/AUTOMATED ANOMALY BASED IDS

HidemaTanaka

National Defense Academy of Japan
Hashirimizu 1-10-20 Yokosuka, Kanagawa Japan 239-8686.

ABSTRACT

We shall discuss new problems of quantification/automation of anomaly-based Intrusion Detection System(IDS). We shall analyze effectiveness and weakness using our proposal method as an example, and derive new attack scenario. Development of anomaly-based IDS is necessary for correspondence to a high network attack, however, we shall show that it makes new different problems at the same time. In this paper, we shall discuss some attack scenario which makes invalidate our detection. As the result, we conclude that it is difficult to prevent such attacks technically, and security requirements for operation side become serious.

KEYWORDS

Anomaly-based intrusion detection system, Automated IDS, Discrete Fourier Transform, Spectrum analysis, Kyoto2006+ dataset

1. INTRODUCTION

Since network service became public common infrastructure, the problem of attack detection continues [1]. The techniques of attack grow complex and frequency of attack keeps on increasing. An outcome of detection against sub-specific attack methods by signature-based intrusion detection (IDS) becomes clear, however, it is always late to the offensive of brand new attack methods (Zero-day attacks) [12]. It is necessary to block communication off at the time when a doubtful sign was found. Therefore, development of anomaly-based IDS is expected. However, almost of all of them take statistical detection method, Zero-day attack was detected after it was already spread. Although the methods using machine learning schemes are focused, there are no effective schemes. This is because the number of input parameters is limited as compared with signature-based IDS. Therefore essentially, proactive counter measures are not achieved.

In this paper, we shall discuss a problem with quantification and automation of anomaly based IDS. Since the techniques of quantification and automation are research level, unfortunately, there are no practical methods. And recent developing techniques using machine learning and deep learning, are also statically methods. They also take statistical strategy, it is not made the target of this research. Therefore, we take our proposal method [15]. Our method quantifies the difference between ordinal and malicious sessions using the entropy. By calculating a value of entropy of target session, it is possible to judge ordinal or malicious automatically. On the other hand, when the malicious session has an almost same value of entropy of ordinal session, it always fails detection. Based on these facts, we success the automation of anomaly based IDS, on the other hand, it generates a new type of security problems. We shall discuss this new type of security problem and countermeasures and show that security requirements for operation side becomes

DOI: 10.5121/ijnsa.2017.9601

serious. As a result, correspondence to new sophisticated network attacks makes different new problems.

2. INTRUSION DETECTION SYSTEMS AND THEIR WEAKNESS

Intrusion detection system (IDS) is categorized into two types mainly; signature-based and anomaly based. Signature based IDS detects malicious packets by comparing with "signature" which is a database generated by analysis of known attacks. As freeware signature-based IDS, Snort [23], Bro [20], Swatch [24] and Logsurfer [21] are well known. Among them, Snort is the most typical and has high detection rate. Bro supports customize parameters with simple scripts. Swatch and Logsurfer use also sys-log to detect malicious packets. In these other ones, pattern matchings applied [2], [3], [5], effective signature generation methods [9] and other techniques are developed. These types of IDS can judge recent sessions which are almost known attack methods and are already analyzed. However, unknown attacks (Zero-day attacks) cannot be detected completely, the security measure only based on signature-based IDS is insufficient. Furthermore, there is no effect on encrypted malicious communication which is increasing recently. Especially in the case of Drive by Download type attacks, it only depends on the judgment by IP address of C&C server.

In anomaly-based IDS, normal behaviour is defined to distinguish ordinal session from malicious. Therefore, it may have an advantage in the detection of Zero-day attacks. There are many methods; for example, Wang method [17], Imai method [10], Sato method [11] and Enkhbold method [6] are proposed in recent years. Wang method is unsupervised using Mahalanobis distance. Imai and Sato methods are also unsupervised using cluster analysis. Enkhbold method is our previous work. We show details in section 3. Zhoumethod[18] resembles Enkhbold method and is also the technique which focuses on frequency and spectrum analysis, however, there is a difference in the used change number of characteristics. Zhou focuses only in "time interval" but Enkhbold focuses in "time interval and payload". There are also many other proposal methods [4], [8] and [16]. And recently some techniques using machine learning or deep learning, however, almost of anomaly-based IDS are statistical analysis method and they cannot detect malicious communication in real time. In particular, the number of valid feature parameters of the machine learning anomaly-based IDS decreases with respect to the encrypted malicious communication. Therefore, the detection rate drops markedly. An automatic operation of machine learning anomaly based IDS is easy to realize, but for these reasons, we have excluded from our subject.These anomaly type IDSs are difficult to define "normal behavior", in general, they have a non-negligible false positive rate. As a result, they are also difficult to operate in real detection operations and an effective method is not established yet.

Their advantage and disadvantage points are summarized in Table 1. In this paper, we will discuss disadvantages of anomaly-based IDSs as follows.

- **W1:** high false positive
- **W2:** difficulty for automation detection
- **W3:** necessity for observation of the whole session

W1 is the main topic of anomaly-based IDSs. Almost research works relate to this weakness for improvements. W2 is also an important topic. Since many methods are based on visual identification and human decisions, automation is impossible. In addition, there is also significant fact that methods based on such human decisions have quite a few false positive than automated schemes, as a result, it becomes effective improvement and solution of W1. W3 suggests that almost anomaly based IDS cannot judge not to observe the whole communication. This is

necessary for more precise statics information, too. Therefore, only an exposed detection is possible but real-time detection is infeasible.

Table 1. Advantage and Disadvantage

	Advantage	Disadvantage
Signature based	・known attack method can be detected completely	・Zero-day attack method is impossible to detect ・inapplicable to encrypted sessions
Anomaly based	・possibility to detect Zero-day attack ・applicable to encrypted sessions	**subject of this paper**

3. QUANTIFIED ANOMALY BASED INTRUSION DETECTION SYSTEM

3.1. Outline

There are some existing methods for quantified anomaly-based intrusion detection systems, in this paper, we focus on our proposal method [15] because it can realize real-time and automated detection. The basic technique of this method is based on Enkhbold method [6]. Our method is based on discrete Fourier transformation (DFT) and calculation of entropy.

In the followings, we define session the total communication set between one client and servers. Our method has the assumption that the behavior of almost ordinary session seems random because the demands of the user are various. On the other hand, since malicious sessions have a certain purpose, their behavior will have some trends. These differences appear in the spectrum analysis of time changes of payload size in the session. Main procedure is as follows (Figure 1).

- **Step-1:** Make discrete waveform using time changes of payload size. Note that positive value is the payload from the clients, and negative value is from the server.
- **Step-2:** Perform DFT to the waveform and derive spectrum.
- **Step-3:** Calculate “entropy” using the standard spectrum and scheme based on Shannon-Hartley theory.
- **Step-4:** Look up the detection table to judge.

The standard spectrum in Step-3, is defined using only ordinary sessions which could be confirmed certainly. The derivation of ordinary spectrum is the same procedure shown above and we define the standard spectrum as the average of them. Figure 2 shows the resultant of Step-2. Only from the figure, we can easily judge ordinal or malicious comparing with the standard spectrum.

Then we calculate ordinary entropy for each ordinary spectrum using the standard spectrum. Also, we calculate malicious entropy for each malicious session (note that malicious session is confirmed to be malicious certainly) using the standard spectrum. The detailed derivation schemes of the spectrum and calculation of entropy is described in section 3.2. As the results, we can derive the detection table (see Table 3, an example table of Kyoto 2006+). Since our method is based on the assumption mentioned above, the value of entropy becomes large when it is malicious session. Therefore, the table shows the threshold of the value of entropy between ordinal and malicious with probability.

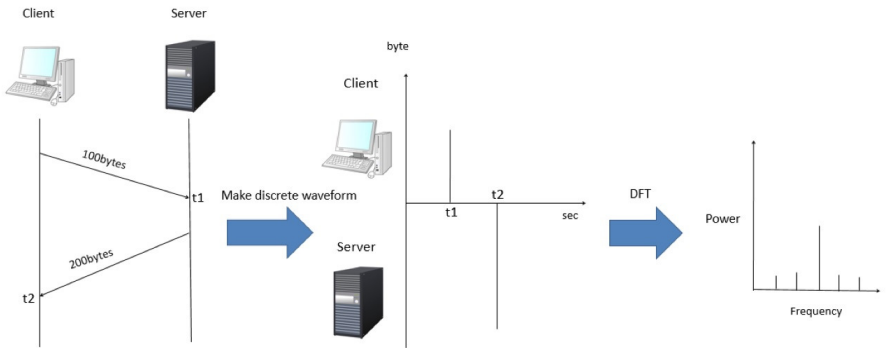

Figure 1. Outline of previous method [15]

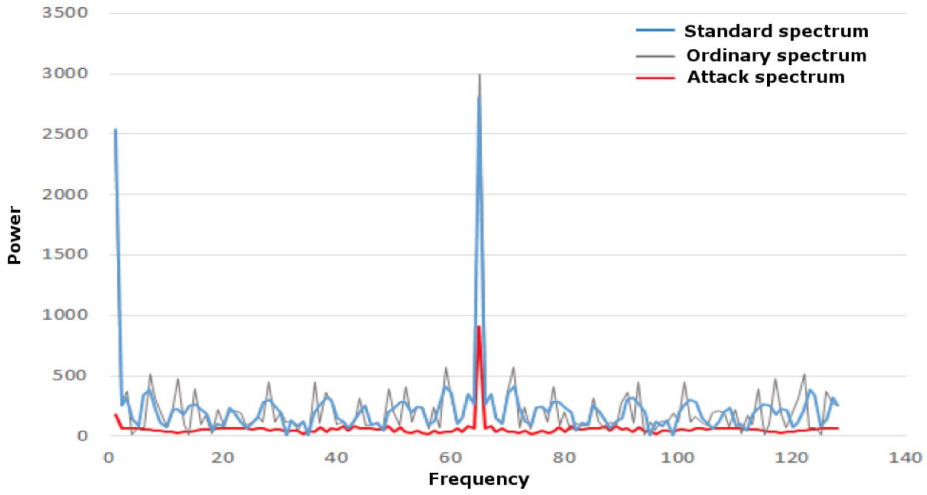

Figure 2. Example of spectrum analysis

Our prototype method [14] is not quantified but is based on visual identification for detecting. Therefore, it is quite primitive but it is powerful for detecting. This method classifies the communication situation into following three types.

- **Type-1.** One client - one server - with various payload size session (O-O-V)
- **Type-2.** One client - one server - with fixed payload size session (O-O-F)
- **Type-3.** One client - some servers - with various payloadsize session (O-M-V)
- **Type-4.**One client - some servers - with fixed payload size session (O-M-F)
- **Type-5.** Some clients - one server - with various payload size session (M-O-V)
- **Type-6.** Some clients - one server - with fixed payload size session (M-O-F)

Note that the type of “Some clients - some servers” can be regarded as a combination of “One client - one server”, we omitted the case. Type-4 “One client - some servers - with fixed payload size session” has not be seen as ordinary session yet in our network observation. From our observation experience, this type is almost systematic/collusion port scanner XSS which does not use advanced techniques. Therefore, we can detect such attacks and concluded that this case is the malicious session, we omitted the classification. And we have not yet foundType-6 “Some clients - one server - with various payload size session” in our observations. However, access by mail

server and proxy server in local network will correspond to this type. Since access to the internal server is not targeted at this time, the detailed analysis is necessary but it is excluded in this paper. Table 2 shows an example rate of the classified session (2008/1/10, 2008/1/20 and 2008/1/30).The number and type of attack vary greatly from day to day, but following this classification method, the rate as of 2008 and the trend of current (2016) do not change much. Therefore, the efficiency of detection table improves on leaps and bounds by applying this classification method to proposal method.

Table2. Example Rate of classified session per-day (Kyoto2006+)

	2008/1/10		2008/1/20		2008/1/30	
	Ordinary session	Attack session	Ordinary session	Attack session	Ordinary session	Attack session
O-O-F (Number of sessions)	12.0% (1694)	2.8% (398)	7.8% (1375)	8.5% (1492)	9.7% (1492)	2.6% (407)
O-O-V (Number of sessions)	51.6% (7255)	1.9% (266)	33.6% (5898)	8.5% (1496)	44.9% (6917)	2.7% (408)
O-M-F (Number of sessions)	0.0% (0)	0.0% (0)	2.6% (464)	2.8% (491)	0.0% (0)	5.8% (890)
O-M-V (Number of sessions)	29.7% (4177)	0.0% (0)	33.2% (5816)	3.0% (504)	28.8% (4428)	5.5% (852)
M-O-F (Number of sessions)	0.0% (0)	2.0% (278)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
M-O-V (Number of sessions)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)

3.2. Entropy and detection table

Our method [14] [15] uses discrete Fourier transform (DFT). Note that there are some mistakes and typos in [15], so we correct them in this paper.In addition, our method does not follow the strict definition, but we are paying attention to experimentally confirmed advantage.

Let $f(x)$ be a discrete wave form derived in Step-1, and T be a session time from start to end ($0 \leq x \leq T$). Since the value of T is not a fixed value, when we perform DFT to any waveforms, each resultant spectrum has various frequency range. Note that the expression "frequency" is not correct, but we use it for the convenience. Then we cannot use the standard spectrum to calculate its entropy, we normalize each session with $1/T$. In this process, we set 10^{-m} of minimum scale, then each discrete waveform has $N = 10^m$points. As the result, we perform DFT to $f(x)$ as follows.

$$|F(k)| = \sum_{n=0}^{N-1} (f(n) \times W_{\text{han}}(n))\, e^{\frac{-i2\pi kn}{N}} \qquad (1)$$

$$W_{\text{han}}(n) = 0.5 - 0.5\cos\frac{2\pi n}{N-1} \qquad (2)$$

where $|F(k)|$,$(0 \leq k \leq N- 1)$, is the power of the spectrum. And where $W_{\text{han}}(\cdot)$ denote Hanning window function. The analysis shown in [14] concludes Hanning window function is optimal for this scheme [7]. As already shown above, the standard spectrum is the average spectrum among ordinary spectrums. Therefore, it is necessary to collect many numbers of ordinary sessions as much as possible for the successful detection. The distribution of ordinary entropy value is derived by using these results. Therefore, we can distinguish malicious one from the distribution. From Shannon-Hartley theorem, the entropy is calculated as follows.

$$I = \int_{f_1}^{f_2} \log_2 \left(1 + \frac{S(f)}{N(f)}\right) df \quad (3)$$

where $S(f)$ denotes the power of the signal at frequency f and $N(f)$ denotes one of noise.

Note that our derived spectrums do not assume AWGN (Additive White Gaussian Noise) and they are independent spectra each other. Therefore, we cannot use Shannon-Hartley theorem directly. However, from following reasons, we propose a quantification method applying modified Shannon-Hartley theorem.

As a method for quantitatively evaluating two spectra as shown in figure 2, we examined the followings.

(Trial-a) Spectra wave correlation

This is considered a typical solution. But, since the spectrum power is ignored, it was not an effective method. Especially, it was effective only in the case where it is visible by the human eye as shown in Figure 1.

(Trial-b) Spectral power difference and their average

The range that the evaluation value can take is large, the false detection rate becomes high. It is effective only in case of fixed payload size session (Type-2,4,6).Our assumption is that "request from normal the user is various, and resultant payload size will be various".This result means that we can detect when the contradiction is made to the assumption. This trial is not adequate.

(Trial-c) The area of the part delimited by two spectra

The detection rate was the highest among our trials. This is an improvement of (Trial-b).At first, we tried the difference total. But in the same case of (Trial-b), the range that the evaluation value can take becomes large.

Table 3. Detection table (Kyoto 2006+)

E	Prob. of ordinal	Prob. of malicious	Ratio
$0 \leqq E < 47$	81.6%	18.4%	73.0%
$47 \leqq E < 76$	19.6%	80.4%	23.8%
$76 \leqq E < 78$	54.1%	45.9%	0.3%
$78 \leqq E$	20.0%	80.0%	2.9%

We tried several ideas based on these results. A (Trial-c) seems to be a naïve Shannon-Hartley theorem. It was experimentally confirmed that when the logarithm is taken with respect to the spectrum ratio, the detection rate further improved. From these results, we modify Shannon-Hartley theorem and derive following.

$$E = \sum_{n=0}^{N-1} \log_2 \left(\frac{\min\{S_s(n), S_t(n)\}}{\max\{S_s(n), S_t(n)\}}\right) \times \Delta f \quad (4)$$

where $S_s(n)$ and $S_f(n)$ denote the standard spectrum power and a target spectrum power at point n. And where Δf denotes the unit frequency scale which is defined as follows.

$$\Delta f = \frac{f_s}{N-1} \tag{5}$$

where f_s denotes the sampling rate for a real network environment. In our case, we determine it by the average of a total number of sessions per unit time. We call the evaluation value E as entropy. As mentioned above, eq. (4) does not assume AWGN, which is incorrect from the definition of entropy. However, from the assumptions in our prototype and easy to see in Figure 1, the entropy of the ordinary session with few uncertain elements will become small, and the entropy of the attack session with large uncertainty will become large. This expectation is similar to the characteristic of entropy. Therefore, we call the evaluation value E as entropy for the convenience.

As the result, we can get the value of entropy for a target session. Then we judge it whether ordinal or malicious by using its entropy. The detection using only entropy distribution, the experience of the staff is necessary and there is no objectivity. Therefore, we use judgement method based on probability. Let $PO(E)$ be the probability of ordinal session when the value of entropy is equal to E. And let $PM(E)$ be the probability of malicious session.

$$PO(E) = \frac{\#N_E}{\#N_E + \#N_M} \tag{6}$$

$$PM(E) = \frac{\#N_M}{\#N_E + \#N_M} \tag{7}$$

Let Q be the threshold for successful detection probability and we search for the range of E which satisfies followings.

$$\sum_{E \in \widetilde{E_S}} PO(E) \leq Q \quad \text{or} \sum_{E \in \widetilde{E_S}} PM(E) \leq Q \tag{8}$$

where

$$\widetilde{E_S} = \{E | E(m) \leq E < E(j)\} \tag{9}$$

Note that $E(m)$ denotes the$m$-thvalue of E. The table where these results are gathered is defined as the detection table (see Table 2). When the result of detection is confirmed true, the detection table is updated successively.

3.3. Effectiveness

We demonstrated our method using Kyoto 2006+ dataset which is open to the public and made by actual traffic data during Nov. 2006 to Aug. 2009 [13] [19]. Because observation of session needs various time, and the processing of normalization is also needed, the real-time detection cannot be done. Therefore, we solve these problems by introducing shortening time into observation of sessions. From experiments using Kyoto 2006+, we can confirm that 2 seconds of observation is sufficient in the case of Kyoto 2006+ to detect malicious sessions [15]. As the result, in addition, we can skip the normalization processing because the frequency range is unified. As already shown in Table 2 is also such result and we can operate real-time detection using the table. From the table, we can confirm that our assumption “low entropy session is ordinal but the case of high entropy will be malicious” is almost true.

In addition, we found that changes in the amount of data are intense according to user demand in HTTP protocol especially. On the other hand, because of many attack sessions, for example, XSS attack, use the characteristic amount of data only at the start of the session, we can detect them easily by our method. Therefore, we can expect that our proposal scheme is effective especially on detection in http protocols. From these facts, our proposal method has good effectiveness when used in following conditions.

1) Observe HTTP protocol
2) 2 second shortening time
3) Type-3 communication situation (see section 3.1)

Note that our proposal method needs many numbers of sessions by which ordinal or malicious is checked clearly. And the resultant detection table and range of value of entropy are depended on the situation of the communication environment in each organization. In the case of Kyoto 2006+ dataset, ordinal or malicious is already defined clearly to each session. Thus, we can derive a correct detection table for Kyoto 2006+ case, however, scrupulous attention is necessary for the collection of these session data in actual practical use.

4. COMPARISON WITH OTHER METHODS (ADVANTAGEOUS POINTS)

For the case of Kyoto 2006+, our method has 81.7% of successful detection, 18.3% of the false positive and 0.0% of false negative. When we limit to "HTTP and Type-3" communication, it has 98.7% of successful detection. All misjudge in this condition, occurs in the case of $74 \leq E \leq 76$, however, these cases have only 0.3% of ratio. In the same way, Sato [11] evaluated their proposal anomaly-based IDS using Kyoto 2006+. Comparing with Sato method, we have quite higher detection rate, especially in http communication. Note that, because Sato method uses human decisions, comparison by the same conditions is impossible. Detailed comparisons with other methods are shown in [14] and [15], and our advantage in detection rate could be confirmed.

Since our method uses only detection table, we do not need huge memory space for the database. In addition, we need only DFT calculation, we do not need high performance CPUs. Therefore, we can conclude our method is high-cost performance method. As already described, since automatic and real-time detection is possible by our method, when combining with some IDSs, we can expect to function as proactive detection [15]. As the results, the advantage in operation can be confirmed.

Unfortunately, we have not detected Zero-day attack yet. But we believe that our methods will be succesfull. Likewise, validity for encrypted sessions has not been confirmed. However, since the proposal method uses only the time interval and the size of the payload, it can be applied to encrypted sessions and expected to achieve the same high detection rate.

5. ANALYSIS OF WEAKNESS

5.1. Forgery of entropy

In this section, we shall discuss the secure operation of our proposal method. First, we assume the adversary who exists outside of local network but knows the detail information of detection table. Then he/she can make camouflaged malicious session which has an ordinal value of entropy using some cheat tools such as ostinato [22]. It is very easy to make such packets using Wireshark [25] and ostinato, then the attack succeeds clearly (needless to simulate). In other words, our proposal method misjudges malicious session which has ordinal entropy value. This is quite big weak point and is the problem with our proposal method. We cannot find out any

countermeasures without keeping the detailed detection table as the confidential information. Practical use of network security is outsourced by many organizations, necessity of confirmation of staff's trust is important.

Next, we assume the adversary who exists in internal of the local network. The adversary can make plausible sessions which are not appropriate but valid. Note that since these are not real malicious, our proposal method judges them ordinal and takes their entropy values into ordinal probability. It is possible for the adversary to injection false results into the detection table by repeating such plausible sessions. Such sessions can be generated easily using Wireshark and ostinato. Therefore, this case is also weak point for our proposal method. The countermeasure of this problem is not also easy, and it will be done only to forbid the use of application which enables packet control.

From these analyses, we can conclude that the inner conviction or interference from the internal of the local network, is serious for our proposal method.In addition, the above malicious acts are also effective against IDS based on machine learning. At the same time, the fact that forgery of entropy is possible also implies that it is impossible to test the detection rate by malware simulation. There are several malware simulations, for example, Zero-day simulation is only manipulating the transmission timing and payload. This is exactly the same as the attack described above, only the position of the attacker and the defender are different, it is not useful for testing or improving the detection rate. In this way, only the actual attack example is effective, so it is important to keep the detecting table setting confidence.

5.2. Skip at time shortening

As shown in Section 3.3, our proposal method is applied a time shortening function to be more effective operation and detection. In fact, in the case of XSS, characteristics which can be judged as malicious communication can be found at the beginning of the session. However, if adversary gets the information concerning to the information how long session to observe, it will be easily skipped. Or, if adversary knows the timing of detection beforehand, there is also a possibility that the attack method which shifted this timing is easy to develop. Therefore, we need to set various observation timings. As the result, the costs concerning to detections and processing become large, however, it will not become a serious problem of operation because our proposal method needs only numerical analysis and does not require database search.

On the other hand, the information concerning to timing of observation becomes confidential. As already pointed out in above section, staff's trust problem will also become important here. In addition, an effective setting of detection cannot also be opened. Therefore, a third person cannot verify the effectiveness, and there is a possibility which becomes disincentive of technological development. In the future, it is expected that not only malware distribution but also ordinary communication will increase encrypted communication, so it becomes increasingly difficult to detect, and our proposal method can contribute to an improvement of attacks.

5.3. Attacking tool for encrypted session

Related to Section 5.1, our proposal method can be applied to the generation of effective attack communication, especially in encrypted communication. Needless to say, the signature-based IDS is not useful at all for encrypted communication. Similarly, since the number of feature parameters decreases, the detection rate of machine learning type IDS also decreases. In conclusion, our proposal method is almost the only detection method for encrypted communication. However, as described above, it enables encrypted malware communication

which cannot be detected by applying in reverse. There is no countermeasure at this moment besides strict operation to prevent this problem.

6. CONCLUSION

In this paper, we have introduced quantified anomaly-based intrusion detection system and showed the effectiveness especially from the point of view of real time and automatic detection. And our proposal method contributes this theme and it will help staff who is the beginner or well trained. On the other hand, we pointed out that another problem causes. In particular, we claimed that staff's trust problem is serious. In addition, Information sharing and an open discussion will be in the difficult situation. The inner conviction is a big problem already, and we can expect this problem to become more serious. On the other hand, a countermeasure to this problem is immature.

In fact, such problem on the practical operation also occurs in other anomaly-based IDSs based on human decisions. It is because the discovery of malicious session or Zero-day attack is based on staff's trust. The growth of complex and frequent network attacks develops a problem of hard detection and high cost for security. Automation is mentioned by one of the solutions, however, we show that it will also make another problem of staff's trust and hardness of operations. Especially, the inner crime becomes big issues recently. We should be careful that a solution for new problem also generates different new problems.

REFERENCES

[1] G.Bruneau. The history and evaluation of intrusion detection. SANS Institute Reading Room, https://www.sans.org/reading-room/whitepapers/detection/history-evolution-intrusion-detection-344, 2001.

[2] M.A.Alia, A.A.Hnaif, H.K.Al-Anie, K.A.Maria, A.M.Manasrah and M.I. Sarwar. A novel header matching algorithm for intrusion detection systems International Journal of Network Security and Its Applications. vol.3, No.4, 2011.

[3] A.V.Aho and M.J.Corasick. Efficient string matching: An aid to bibliographic search. Communications of the ACM, vol.18(6), pp.333-340, 1975.

[4] P.Barford, J.Kline, D.Plonka and A.Ron. A signal analysis of network traffic anomalies. In Proceedings of Internet Measurement Workshop, pp.71-82, 2002.

[5] B.Commentz-Walter. A string matching algorithm fast on the average. InProceedings of International Colloquium on Automata, Languages and Programming (ICALP), pp.118- 132, 1979.

[6] E.Chimedtseren, K.Iwai, H.Tanaka and T.Kurokawa. Intrusion detection system using Discrete Fourier Transform. In Proceedings of Seventh IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA2014)pp.1-5, 2014.

[7] F.J.Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. In Proceedings of the IEEE, vol.66, no.1, pp.51-83, 1978.

[8] S.S.Kim, A.L.Narasimha Reddy and M.Vannucci. Detecting traffic anomalies through aggregate analysis of packet header data. In Networking 2004 Springer Lecture Notes in Computer Science 3042, pp.1057-1059, 2004.

[9] C.Kreibich and J.Crowcroft. Honeycomb: Creating intrusion detection signatures using honeypots. ACM SIGCOMM Computer Communication Review vol.34(1), pp.51-56, 2004.

[10] K.Imai,S.AokiandT.Miyamoto.Anomaly detection based on clustering of network traffic characteristics considering results go signature base IDS evaluation. ICISS Technical Report vol.489, no.114, pp.7-12, 2015.

[11] M.Sato, H.Yamaki and H.Takakura. Unknown attacks detection using feature extraction from anomaly-based ids alerts. In Applications and the Internet (SAINT), 2012 IEEE/IPSJ 12th International Symposium on IEEE, 2012 pp. 273-277, 2012.

[12] K.Skinner and A.Valdes. Adaptive model based monitoring for cyber-attack detection In Recent Advances in Intrusion Detection 2000 Springer Lecture Notes in Computer Science 1907, pp.80-92, 2000.

[13] J.Song, H.Takakura, Y.Okabe, M.Eto, D.Inoue, and K.Nakao. Statistical analysis of honey pot data and building of Kyoto 2006+ dataset for NIDS evaluation. In Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security ACM, 2011, pp.29-36, 2011.

[14] Y.Tsuge and H.Tanaka. Intrusion detection system using discrete Fourier transform with window function International Journal of Network Security & Its Applications (IJNSA) Vol.8, No.2, pp.23-34, 2016.

[15] Y.Tsuge and H.Tanaka. Intrusion detection system with spectrum quantification analysis International Journal of Cyber-Security and Digital Forensics (IJCSDF) 5(4), pp.197- 207, 2016.

[16] S.K.Wagh, V.K.Pachghare and S.R.Kolhe. Survey: Learning techniques for intrusion detection system. International Journal of Advance Foundation and Research in Computer vol.1, issue 2, pp.21-28, 2014.

[17] K.Wang and S.J.Stolfo. Anomalous payload based network intrusion detection. In International Workshop on Recent Advances in Intrusion Detection (RAID) Springer Lecture Notes in Computer Science 3224, pp.203-222, 2004.

[18] M.Zhou and S.D.Lang. A frequency-based approach to intrusion detection. In Proceedings of the Workshop on Network Security Threats andCountermeasures 2003.

[19] Traffic Data from Kyoto University's Honeypots. http://www.takakura.com/Kyoto data/.

[20] The bro network security monitor. https://www.bro.org/.

[21] Logsurfer. https://www.cert.dfn.de/eng/logsurf/.

[22] OSTINATO Network Traffic Generator and Analyzer. http://ostinato.org.

[23] Snort. https://www.snort.org/.

[24] Swatch. https://www.swatch.sourceforge.net/.

[25] wireshark. https://www.wireshark.org/download.html.

Author

Hidema Tanka is an associate professor of National Defense Academy Japan. His main research area is analysis of cryptographic algorithm, code theory, information security and cyber warfare. and its domestic laws.