A Classification Method of Unknown Malicious Websites Using Address Features of each Network Address Class

Shihori Kanazawa*, Yoshitaka Nakamura**, Hiroshi Inamura**, and Osamu Takahashi**

* Graduate School of Systems Information Science, Future University Hakodate, Japan
** School of Systems Information Science, Future University Hakodate, Japan
{g2116011, y-nakamr, inamura, osamu}@ fun.ac.jp

Abstract - Recently, cyber-attacks through web sites such as Drive-by download attacks or phishing attacks increase rapidly. The attackers acquire personal information of users illegally by these attacks and economically damages to the users. The conventional detection method of malicious Web site uses a blacklist, and characteristics of the domain name. Since the domain name can be changed relatively easily, it is inappropriate for the method of detecting malicious Web sites to using domain names. In this paper, we propose a method of classifying Web sites as benign or malicious by using only a part of the network address, in order to reduce the classification cost. And, we evaluated the proposed method by cross-validation. As a result of evaluation, high classification accuracy was provided in IP address Class A, and we could confirm the effectiveness of the proposed distinction method.

Keywords: cyber-attack, Drive-by download, malicious Web site, network address, IP address of Class

1. INTRODUCTION

In recent years, the threat of attacks by viruses and malwares on the Internet are increasing year by year. Among them, attacks using Web sites are increasing rapidly. According to "2017 Information Security Ten major threats" announced by Information-technology Promotion Agency (IPA) in March 2017, the first place, the fourth place, and the sixth place are attacks against Web sites[1]. As an example of attack, cyber-attacks through Web sites such as Drive-by download attacks or phishing attacks increasing particularly rapidly. The attackers acquire personal information of users illegally by these attacks and inflicts economical damage[2]. Figure 1 shows the number of incidents occurring and damage amount data from Ref.[2] by National Police Agency. The number of the illegal acquisitions of the personal information that occurred by 2012 was 50 cases. However, the number of occurrences of the incidents has increased to 1400 cases until 2015. In order to prevent such damage, it is necessary to take measures to prevent users from accessing malicious Web sites.

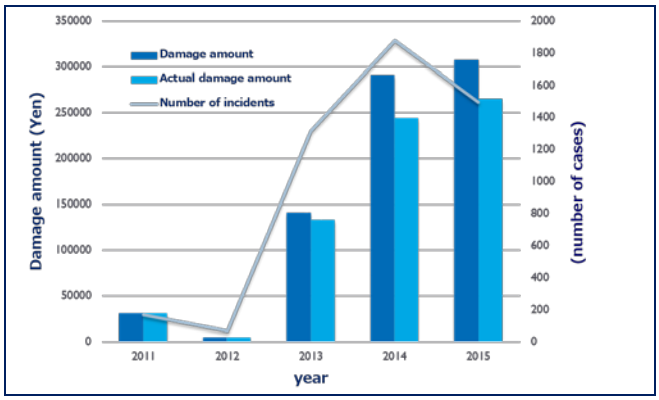


Figure 1: The number of incidents, damage amount, and actual damage amount

There are some methods to block access if the visited Web site is a malicious Web site to prevent users from accessing those sites. The first method is to use the Web reputation system[3]. The Web reputation system has a function of blocking malicious Web sites. If the domain name of the connection destination or the Web site is judged to malicious, the web reputation system blocks web access. In this way, the Web reputation system prevents damage caused by malicious programs and phishing. However, the Web reputation system can block access only to Web sites that have been confirmed to have carried out fraudulent acts such as virus distribution and phishing scams. The second method is to use Intrusion Prevention System. IPS supports sophisticated and advanced security threats such as bot attacks and DoS attacks that are considered to be difficult to protect only by general firewalls and anti-virus software. IPS examines the contents and behaviors of communication packets, and blocks web access if IPS detects communication that can be considered malicious. However, IPS can only detect known suspicious packets included in Web access communication.

Since the above two methods use known information such as information of suspicious packets included in known malicious Web sites, there is an advantage that the detection rate of known malicious Web sites is relatively high. However, these methods have drawbacks that can not be detected unknown malicious Web sites. Therefore, it is also unknown whether these methods can obtain sufficient accuracy. In order to solve such a problem, it is necessary to consider the detection conditions that enables detection including unknown malicious Web sites. Also, it is necessary to classify unknown Web sites into benign Web

sites and malicious Web sites. Therefore, we propose a method to detect and classify an unknown malicious Web sites.

This paper is constructed as follows. Section 2 mentions about researches related to our study. Section 3 mentions the requirements of the proposed method and our plan of the experiment. Section 4 mentions experimental results. Finally, this paper concludes, and discusses on future work in Section 5.

2. RELATED WORK

In recent years, some system has been developed to prevent the user from accessing to a malicious Web site. For example, one of them is Web reputation systems [3,4] for detecting a malicious Web site. This system utilizes a list of known malicious Web sites. The list of known malicious Web sites is called the blacklist. However, this system using the blacklist can not deal with unknown malicious Web sites.

There are some other researches to detect malicious Web sites including researches based on the features of URLs, researches based on the features of domain names, and researches based on the features of IP addresses. Each type of research is described in detail below.

First, we describe the research based on the feature of URLs. J. Ma et al., and K. Tanaka proposed a supervised learning approach for classifying URLs as normal or malicious based on the lexical structure of URLs [5]. This system can classify malicious domains and normal domains by the features that can be extracted from the DNS communication. Next, we describe the research based on the feature of domain names. I. Ryu et al. uses the features of the domain names as the detection condition [6]. A malicious domain name has a feature that the length of the domain name is 10 characters or more and alphanumeric characters are mixed. The malicious domain name has these features because it is often generated automatically using Fast-Flux attack method[7]. Fast-Flux uses computers infected with bots (botnet) to distribute viruses or guidance information for phishing attacks. As another feature, malicious domain names have many features that mix alphanumeric characters. L. Bilge et al. proposed a system that employs DNS analysis techniques to detect domains that are involved in malicious activity [8,9]. This system can classify malicious domains and normal domains by the features that can be extracted from the DNS communication. Malicious domain names are likely to be accessed by a client infected with malware. These methods are effective for detection of known Web sites, because these methods use the blacklist of domain names and URLs. However, these methods can not be easily maintained an up-to-date blacklist, because domain names can be easily and continuously generated or changed. As research based on IP addresses, Chiba et al. proposed a method of utilizing the feature of malicious IP addresses [10,11]. This research classifies malicious IP addresses and normal IP addresses by the feature of malicious IP addresses, because Cyber Attack is prone to use particular IP addresses[10,11,12]. However, this method has high cost because it is necessary to translate an IP address into all bit strings.

From these studies, it can be said that the approach using blacklists tends to fail for versatile domain-based detection. And the approach of using domain names to classify malicious Web sites can be said to be difficult because domain name can be changed is easily. In this research, we propose a new method to detect a malicious Web site effectively by reducing avoidance from blacklist. This method uses only the domain name for detection in order to expand the detection range of the malicious Web sites. We also propose a method to classify Web sites at low cost by using a part of IP address in order to reduce the cost.

3. A METHOD FOR CLASSIFYING WEB SITES BY USING IP ADDRESS

3.1 Purpose of the study

In this paper, we propose a method to detect unknown malicious Web sites by increasing detection conditions using domain names of such Web sites. In the proposed method, the following two features of the malicious IP address are used to classify the Web site. It is known that a cyber-attack is prone to use particular IP addresses [10,11,12]. However, in order to use this IP address distribution feature, it is inefficient to classify malicious Web sites by comparing all the bit strings of IP addresses for each access communication. Figure 2 shows the usage distribution of malicious IP addresses.

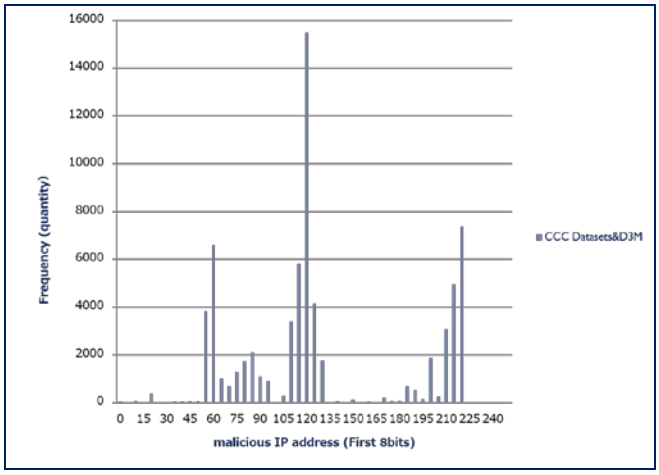


Figure 2: The usage distribution of malicious IP addresses

As shown in Figure 2, specific IP addresses are frequently used. Also, IP addresses can be classified into three IP address classes. Therefore, in this paper, we propose a method to reduce the cost of classifying benign or malicious Web sites using only the network address part according to the IP address class of each IP address.

3.2 Approach

Figure 3 shows the outline of the proposed system using the approach to lower classification cost.

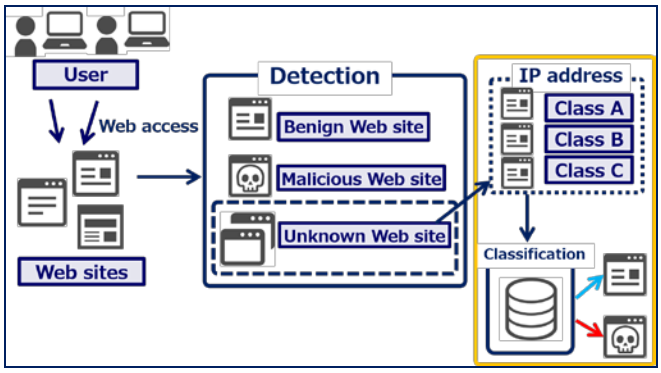


Figure 3: An approach to lowering the classification cost

Web sites accessed by clients can be classified into three categories: benign Web sites, malicious Web sites, and unknown Web sites. It is important how to deal with these unknown Web sites. The unknown Web sites are very likely to be malicious Web sites. In this approach, we will classify unknown Web sites to benign / malicious at low cost by using IP addresses of only unknown Web sites.

3.3 Proposed method

The proposed system consists of two methods. One is a method of detecting unknown Web sites by removing known malicious and benign Web sites. The other method is to classify whether an unknown Web site is malignant or benign. Details of the detection method are described in Section 3.4, and details of the classification method are explained in Section 3.5. Figure 4 shows the overview of the whole proposed system.

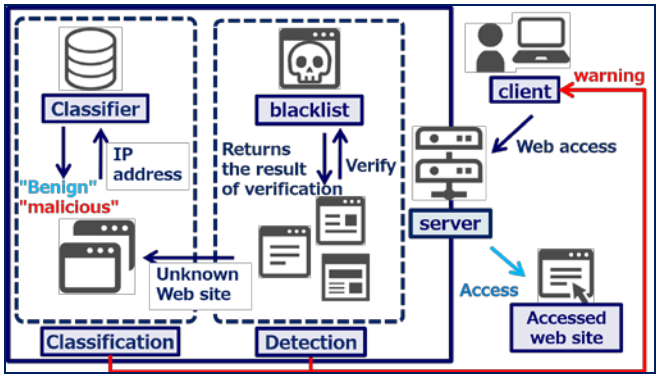


Figure 4: The overview of the proposed method

The proposed system consists of a detection unit and a classification unit. In the detection unit, unknown malicious Web sites that do not exist in the known blacklist are taken out. First, when the DNS server is accessed by the client, the detection unit checks the accessed Web site against the blacklist of known malicious Web Sites. If the Web site is detected to be unknown Web sites, the detection unit sends the IP address of the Web site to the classification unit to use for classification. The classification unit uses the features of this IP address to classify whether the Web site is benign or malicious. If the classification result is benign, this system allows the client to access the Web site. If the classification result is malicious, this system alerts the client, and update the blacklist in order to keep the classifier up to date.

3.4 The method of detection unknown Web sites

Figure 5 shows details of the detection unit which detect unknown Web sites.

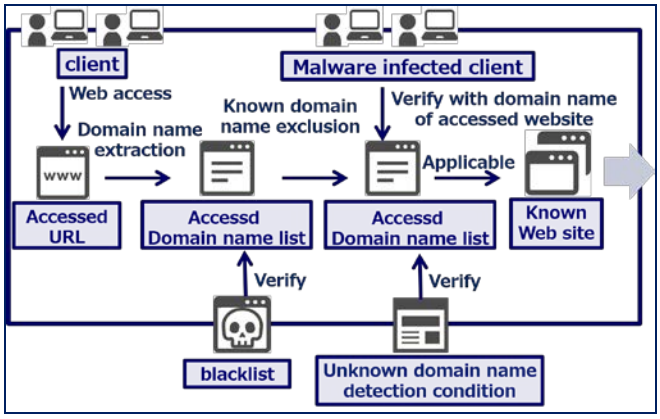


Figure 5: Details of detection method of unknown Web site

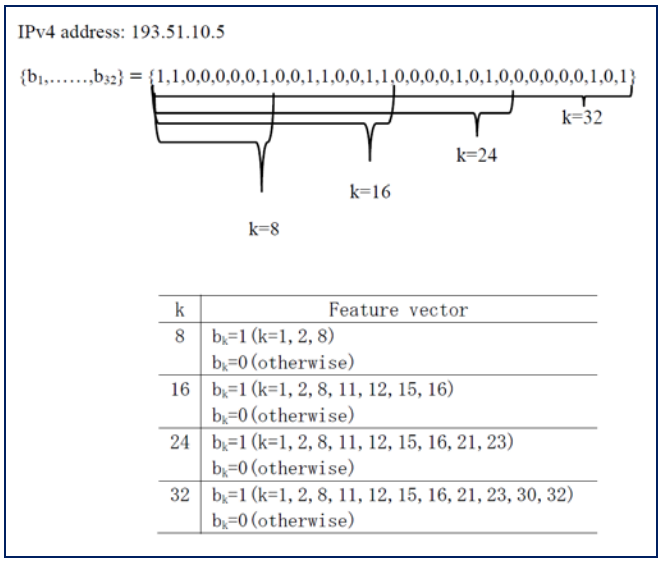
This detection unit uses domain names to detect unknown Web sites. The domain names are obtained from Web site URLs. The domain name is checked against the blacklist in order to exclude the known malicious domain. If the domain name of the Web site is not on the blacklist, the domain name is checked against the detection conditions based on the domain name. Unknown domain name detection condition has features of known malicious domain name. Also, in order to select an unknown Web site, the proposed system uses the domain name accessed by the client infected with malware (malware infected client).

3.5 The method of classifying malicious Web sites

The classification unit of the proposed system consists of two phases. The phase 1 is the construction of a classifier for generation of feature vectors using the training datasets (3.5.1). The phase 2 is the classification of the test dataset using constructed classifier (3.5.2). The training datasets are consisted by known malicious and benign Web sites. Test datasets are unknown Web sites.

3.5.1 Construction of classifier using training datasets

It is necessary for construction of classifier to generate feature vectors. Feature vectors have various features of the training datasets. And the number of feature vectors is referred to as the dimension number. In this paper, we propose a method of classifying Web sites as benign or malicious by using a low number of dimensions, in order to reduce the classification cost. Figure 7 shows the method to generate feature vectors.



k	Feature vector
8	b_k=1 (k=1, 2, 8) b_k=0 (otherwise)
16	b_k=1 (k=1, 2, 8, 11, 12, 15, 16) b_k=0 (otherwise)
24	b_k=1 (k=1, 2, 8, 11, 12, 15, 16, 21, 23) b_k=0 (otherwise)
32	b_k=1 (k=1, 2, 8, 11, 12, 15, 16, 21, 23, 30, 32) b_k=0 (otherwise)

Figure 6: The example of generating feature vectors

First, the IP address of the training datasets is translated into bit strings. Secondly, all bit strings are represented as k-dimensional vector {b_1,.....,b_k}. Three types of feature vectors are generated according to the IP address class such as Class A, Class B, and Class C. A vector of 8 dimensions in the case of Class A, a vector of 16 dimensions in the case of Class B, and a vector of 24 dimensions in the case of Class C is generated. Finally, generated feature vectors of malicious are labeled "1", and generated feature vectors of benign are labeled "0". Table 1 shows that an example of labeling feature vectors of the training datasets.

Table 1: Example of labeling feature vectors of training datasets

IP address	Feature vector	Label
193.51.10.5	1,1,0,0,0,0,0,1,0,0,1,1,0,0,1,1	1
10.10.10.10	0,0,0,0,1,0,1,0,0,0,0,0,0,1,0,1	1
203.4.12.89	1,1,0,0,1,0,1,1,0,0,0,0,0,1,0,0	0
…	…	…

The classifier used in this paper is a Support Vector Machine (SVM), which is one of pattern identification methods. Reference[5] written by J. Ma et al. clearly indicated that malicious Web sites are detected with high accuracy by using SVM. The proposed system constructs three classifiers based on feature vectors described above.

3.5.2 Classification of the test datasets using constructed classifier

The test datasets are classified by the constructed classifiers in the phase 1. Figure 8 shows that classification has three steps.

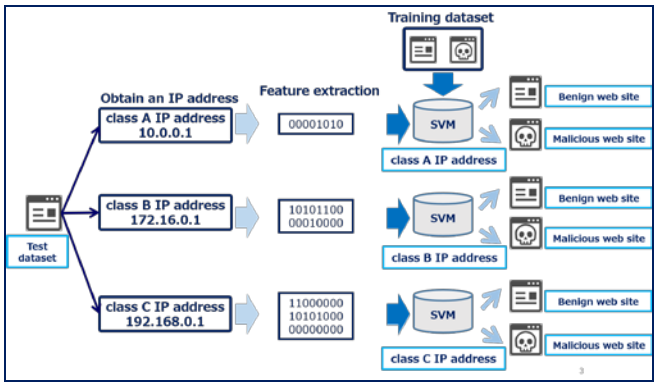


Figure 7: The process of the classification

First, the classification section obtains the IP address from the communication log of the test data set. A feature vector is generated from this obtained IP address. Finally, the feature vectors of the test dataset are classified as benign or malicious by the classifier constructed in the phase 1.

3.6 EVALUATION

The aim of the proposed system is to classify unknown Web sites as malicious or benign. The system may show that IP address of Class A achieves high accuracy with the first 8 bits. IP address of Class B shows high accuracy with the first 16 bits. IP address of Class C shows high accuracy with the first 24 bits.

We conduct an experiment to confirm the effectiveness of the classifier. We evaluated the constructed classifier at three points of accuracy, precision, recall rate. These three points are called evaluation indices.

This paper defines the correct classification of an actual malicious IP address into a malicious category as a true positive (TP), the incorrect classification of actual benign IP addresses into a malicious category as a false positive (FP), the incorrect classification of actual malicious IP addresses into a benign category as false negative (FN), and the correct classification of actual benign IP address into a benign category as a true negative (TN). Accuracy (1), precision rate (2) and recall rate (3) are calculated as follows:

$$Accuracy = (TP + TN)/(TP+ TN + FP+FN) \cdots(1)$$

$$Precision = TP/(TP+ FP) \quad \cdots(2)$$

$$Recall = TP/(TP+ FN) \quad \cdots(3)$$

The dataset of benign and malicious are obtained from Malware Workshop Datasets [13]. The dataset is created with the ratio of malicious IP addresses and benign IP addresses of 8:2, 5:5, 2:8. The malicious training datasets was created from CCC datasets (2008 - 2011) and D3M (2010 - 2015). The benign training datasets were created from Alexa's top sites 50,000 (2016) [14] and NCD in MWS Cup (2014). IP addresses of Class A have 49164 malicious IP addresses and 40667 benign IP addresses. IP addresses of Class B have 3523 malicious IP addresses and 10735 benign IP addresses. IP addresses of Class C have 75000 malicious IP addresses and 14288 benign IP addresses. The

classification experiments are conducted with 5-way cross validation.

4. Result

We carry out the experimental evaluation in four ways as defined in Figure 7. We defined the classification using first 8 bits for Case 1, first 16 bits for Case 2, first 24 bits for Case 3, first 32 bits for Case 4. Table 2 shows the experimental result of classifying the IP address belonging to the IP address of Class A.

Table 2: Experimental result of IP address of Class A

	Accuracy	Precision	Recall
Case1(k=8)	84.06079	89.86656	90.25327
Case2(k=16)	83.74358	89.79705	89.89365
Case3(k=24)	83.76079	89.8188	89.89058
Case4(k=32)	83.87882	89.95386	89.8875

The recall rate achieved the highest value in Case 1. Accuracy and Recall are the highest in Case 1, while Precision is Case 4 the highest.

Table 3 shows the experimental result of classifying the IP address belonging to the IP address of Class B.

Table 3: Experimental result of IP address of Class B

	Accuracy	Precision	Recall
Case1(k=8)	83.54143	87.85521	92.15756
Case2(k=16)	81.69694	88.16298	89.07026
Case3(k=24)	82.94552	88.61024	90.27679
Case4(k=32)	83.25766	88.76131	90.5252

The recall rate achieved the highest value in Case 1. The accuracy also achieved the highest value at Case 1. The precision rate achieved the highest value at Case 4. The accuracy, precision rate and recall rate of Case 2 were lower than those of Case 1, Case 3, and Case 4.

Table 4 shows the experimental result of classifying the IP address belonging to IP address of Class C.

Table 4: Experimental result of IP address of Class C

	Accuracy	Precision	Recall
Case1(k=8)	81.78191	87.19137	90.52493
Case2(k=16)	81.20101	86.9819	89.965
Case3(k=24)	81.243	87.00101	90.0
Case4(k=32)	81.222	86.56467	90.58618

The precision rate achieved the highest value in Case 1. The accuracy also achieved the highest value in Case 1. The recall rate achieved the highest value in Case 4. The accuracy, precision rate and recall rate of Case 3 were lower than those of Case 1, Case 2, and Case 4.

4.2 Discussion

The IP address belonging to Class A IP address seems to obtain the best result because the IP address frequently used for malicious activities and the IP address not used is clearly classified. In addition, Class A is considered to have sufficient information suitable for classification of a malicious Web site because it has a large number of distributions to users. Therefore, classification using network address is effective on Class A IP address. The Class B IP address is not used for many malicious activities, so few features is found in these addresses. Also, in the case of the Class C IP address, there is a possibility that the specific IP address has lowered three evaluation indices. In addition, because Class C is easiest to distribute to users among all IP address classes, malicious IP address is likely to be easily changed. Therefore, it is necessary to analyze the IP address which reduces evaluation indices of classification by IP address of Class.

We compared the usage status of IP addresses that were identified by mistake and IP addresses used for malicious activities. Figure 9 shows the result of comparing the usage status of IP addresses used for malicious activities with the IP addresses belonging to Class B IP address that were incorrectly classified.

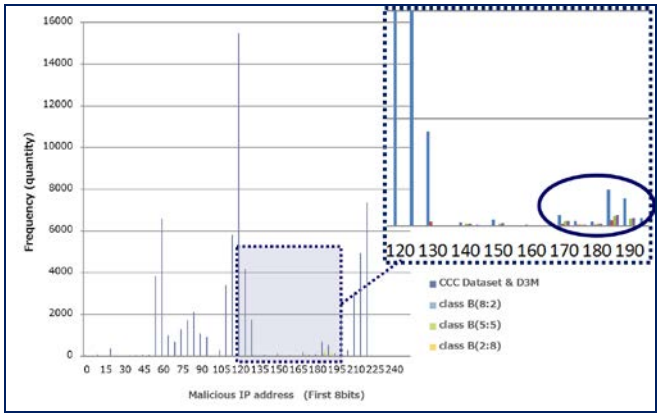


Figure 8: The result of comparing the usage status of IP addresses (Class B)

The IP address whose leading 8 bits are around 120 is often used for malicious activities, so the features of malicious activities are considered to exist. However, the incorrectly classified IP addresses tend to be concentrated on areas where the features of malicious activities do not exist.

Figure 10 shows the result of comparing the usage status of IP addresses used for malicious activity with the IP addresses belonging to Class C IP address that were incorrectly classified.

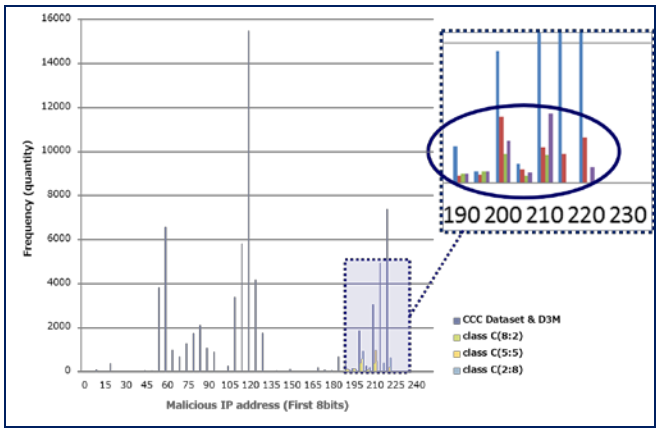


Figure 9: The result of comparing the usage status of IP addresses (Class C)

Since IP addresses belonging to Class C IP address are partially used for malicious activities, it is considered that there are the features of malicious activities. However, the IP address incorrectly classified tends to be concentrated on area where there are the features of malicious activities do not exist.

4.3 Limitation

There may be some limitations in this research. The evaluation indices of Class B and Class C IP address is low in the current training datasets. It is probably due to the amount of Class B and Class C IP address in the datasets. The number of IP addresses belonging to Class B IP address is very small. On the other hand, the number of IP addresses belonging to Class C IP address is very large. Therefore, it is necessary to devise selection of the IP address of the training dataset.

4.4 Future Work

Further research should be done to classify malicious Web sites. First, it is necessary to carefully select IP addresses of training datasets belonging to Class C and Class B IP address. The evaluation indices may be improved by this selection. Secondly, it is necessary to analyze trends of datasets by year. The features of Class C may change from year to year. Thirdly, it is necessary to consider the way to return to the benign IP address in the method classifying by IP addresses. Malicious IP addresses are used for malicious activities for a while. However, when malicious IP addresses are used frequently, malicious activities become to use the other IP addresses for avoiding detection. The IP address that is no longer used may be used by someone else. Finally, it is necessary to confirm that the efficiency of the classification is improved by the proposed system.

5. CONCLUSION

In this paper, we propose a method of detecting unknown Web sites and classifying Web sites as benign or malicious by using only a part of the network address, in order to reduce the cost of the classification. And, we evaluate the accuracy of the proposed classification for each IP address Class by cross-validation. As a result of evaluation, high classification precision was provided in Class A IP address. From this result, we confirm the effectiveness of the proposed classification method. But, there may be some limitations in this system. First, the evaluation indices of Class B and Class C IP address are relatively low in the current training datasets. Secondly, the IP addresses classified as malicious cannot return to benign IP addresses.

In the future, further research should be done to investigate the feature of malicious Web sites. First, it is necessary to carefully select IP addresses of the training datasets belonging to Class B and Class C IP address. This selection of IP addresses may improve the evaluation indices. Secondly, it is necessary to analyze trends of datasets by year. The features of Class C IP address may change from year to year. Thirdly, it is necessary to consider the way to return to the benign IP address in the method classifying by IP addresses. Finally, it is necessary to confirm that the efficiency is improved by the proposed system.

REFERENCES

[1] Information-technology Promotion Agency, Japan, "10 Major Security Threats" <https://www.ipa.go.jp/files/000058504.pdf> [Accessed May 31, 2017] (*in Japanese*).

[2] National Police Agency, "Public information of the National Police Agency: About the occurrence situation of illegal remittance offenses related to Internet banking in Heisei 26, " <https://www.npa.go.jp/cyber/pdf/H270212_banking.pdf> [Accessed January 17, 2017] (*in Japanese*).

[3] TREND MICRO Incorporated, "Web reputation," <http://www.trendmicro.co.jp/why-trendmicro/spn/features/web/index.html> [Accessed May 31, 2017] *(in Japanese)*.

[4] M. A. Rajab, L. Ballard, N. Jagpal, P. Mavrommatis, D. Nojiri, N. Provos, and L. Schmidt, "Trends in circumventing web-malware detection," Google, Google Technical Report, 2011.

[5] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: learning to detect malicious web sites from suspicious urls, " Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'09), pp. 1245–1254, 2009.

[6] I. Ryu, "Detection method of malicious site by DNS information," Master's thesis of Waseda University 2012.

[7] Hitachi Solutions, Ltd. "Information security blog," <http://securityblog.jp/words/2898.html> [Accessed January 17, 2017] *(in Japanese)*.

[8] L. Bilge, E. Kirda, C. Kruegel, and M. Balduzzi, "Exposure Finding Malicious Domains Using Passive DNS Analysis," Proceedings of the 18th Annual

Network & Distributed System Security Symposium (NDSS Symposium 2011), 2011.

[9] K. Tanaka, A. Nagao, and M. Morii, "Extracting Malicious Website from DNS Log-Analysis Method and Anonymity-," Proceedings of the Computer Security Symposium 2013(CSS2013), pp.132-138, 2013 *(in Japanese)*.

[10] D. Chiba, K. Tobe, T. Mori, and S. Goto, "Detecting Malicious Websites by Learning IP Address Features," Proceedings of the IEEE/IPSJ 12th International Symposium on Applications and the Internet(SAINT2012), pp.29-39, 2012.

[11] D. Chiba, T. Mori, and S. Goto, "Deciding priority crawling in searching for malicious websites," Proceedings of the Computer Security Symposium 2012(CSS2012), pp.805-812, 2012 *(in Japanese)*.

[12] D. Chiba, T. Yagi, M. Akiyama, and T.Mori, "Correlation Analysis Between IP Addresses Used in Variety of Attacks," Proceedings of the Computer Security Symposium 2011(CSS2011), pp.185-190, 2013 *(in Japanese)*.

[13] M. Kamizono, M. Akiyama, T. Kasama, J. Murakami, M. Hatada, and M. Terada, "Datasets for Anti-Malware Research ~MWS Datasets 2015~," he Special Interest Group Technical Reports of IPSJ Vol.2015-CSEC-70, No.6, pp. 1-8, 2015. (*in Japanese*).

[14] Alexa Internet, Inc., "The top 500 sites on the web," <http://www.alexa.com/topsites> [Accessed January 17, 2017].