**Regular Paper**

# Classification Method of Unknown Web Sites Based on Distribution Information of Malicious IP addresses

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*** - In recent years, cyber-attacks via 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 inflicts economical damage to the users. Therefore, it is important to detect malicious Web sites which cause economic damage. The conventional detection method of known malicious Web sites uses blacklists. And the detection method of unknown malicious Web sites uses the features of the domain name. Since it is relatively easy to change the domain name, it is inappropriate for the method of detecting malicious Web sites to using domain names. On the other hand, there is a characteristic that it is difficult to change the IP address once it is set. In this paper, we propose a method to classify Web sites as malicious or benign by using only a part of the network address, to reduce the classification cost. And, we analyzed features of the network address part of the IP address class to classify unknown Web sites. We evaluated our proposed classification method by cross-validation. As a result of evaluation, high classification accuracy was provided in IP address Class A. From this result, we could confirm the effectiveness of the proposed method. Also, as a result of changing the acquisition method of features of IP address based on the hypothesis that temporal change exists in the features of malicious IP address, we could confirm improvement of classification accuracy.

***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 or malwares on the Internet are increasing year by year. Among them, attacks via Web sites are increasing rapidly. According to "10 Major Security Threats 2017" [1] announced by Information-technology Promotion Agency, Japan (IPA), Attacks on Web sites appear in 1st, 4th, and 6th place in March 2017. As an example of attack, cyber-attacks via Web sites such as Drive-by download attacks or phishing attacks increase rapidly. Drive-by download attacks make users download malicious programs such as viruses or malwares. Phishing attacks navigate to fake Web sites. Attackers acquire personal information of users illegally by these attacks and inflicts economical damage. Figure 1 shows the number of incident occurrence and damage amount data from Ref. [2] by National Police Agency, Japan. In 2012,

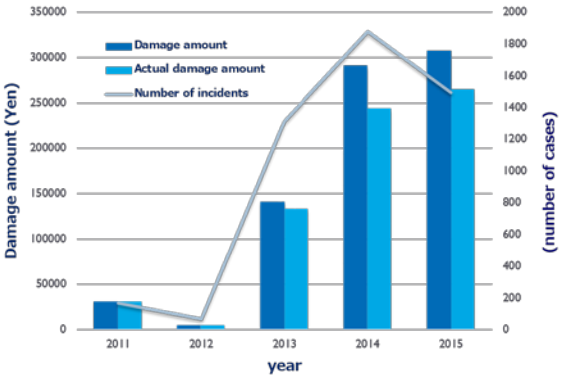



Figure 1: The number of incidents, damage amount, and actual damage amount

the total number of illegal acquisitions of personal information was only 50 cases for a year. However, the number of incident occurrence has increased to 1,400 cases for a year until 2015. In order to prevent such damage, it is necessary to take measures to prevent users from accessing malicious Web sites.

There are some methods to prevent users from accessing malicious Web sites. One method is to use Web reputation system [3]. The Web reputation system has a function to block malicious Web sites. If the connected domain name is judged to be malicious, the system blocks the access. In this way, the Web reputation system prevents damage caused by malicious programs or phishing. However, the Web reputation system can block only access to Web sites clearly made malicious activity such as virus distribution and phishing scams. Another method is to use the Intrusion Prevention System (IPS) [4]. IPS prevents sophisticated and advanced security threats such as bot attacks and DoS attacks that are 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 as malicious. There are two types of detection mechanisms in IPS. The first mechanism uses the signature, that is attack patterns collected in database(DB) in advance. If the access data in the communication matches the pattern in the DB, the data is regarded as malicious and the communication is blocked. The second is the anomaly detection mechanism. The mechanism uses a whitelist of benign operations. If the access is not in the whitelist, the anomaly detection mechanism regards the access as malicious and blocks the communication. Therefore, IPS can only detect known suspicious packets included in Web ac-

cess 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 cannot be detected unknown malicious Web sites. Therefore, it is also unknown whether these methods can obtain sufficient accuracy. In order to solve such problems, it is necessary to consider the detection conditions that enables to detect malicious Web sites including unknown 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 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. Section 4 mentions our plan of the experiment and 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 malicious Web sites. For example, one of them is Web reputation system [3],[5] which detects malicious Web sites. This system uses a list of known malicious Web sites. The list of known malicious Web sites is called the blacklist. However, this system using the blacklist cannot 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 introduce the researches based on the feature of URLs. Ma's group, and Tanaka's group proposed a supervised learning approaches for classifying URLs as malicious or benign based on the lexical structure of URLs respectively [6],[10]. These approaches can classify malicious domains and benign domains by the features that can be extracted from the DNS communication. Next, research is based on the feature of domain names. Ryu's group use the features of the domain names as the detection condition [7]. A malicious domain name has a feature that the length of the domain name is 10 or more characters and alphanumeric characters are mixed. The malicious domain name has these features because it is often generated automatically using Fast-Flux attack method [8]. Fast-Flux uses computers infected with bots (botnet) to distribute viruses or guidance information for phishing attacks. Bilge's group proposed a system that employs DNS analysis techniques to detect domains that are involved in malicious activity [9]-[10]. This system can classify malicious domains and benign domains by the features that can be extracted from the DNS communication. Bilge's group analyzed the DNS communication log over several months using these features and show that malicious domains can be detected with high accuracy [9]. These researches are effective for detection of known Web sites, because these researches use the blacklist of domain names and URLs. However, these researches cannot maintain the blacklist up to date easily, because domain names of Web site can be easily and continuously changed.

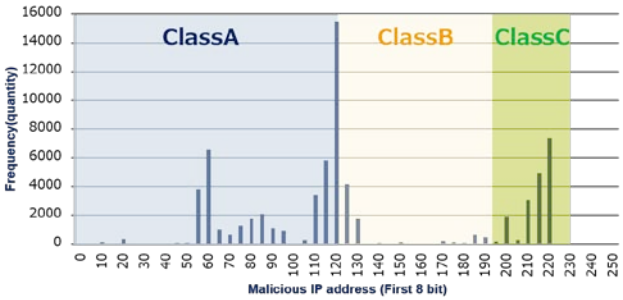



Figure 1: Usage distribution of IP addresses

As research based on IP addresses, Chiba's group proposed a method of utilizing the feature of malicious IP addresses [11]-[12]. This method classifies malicious IP addresses and benign IP addresses by the feature of malicious IP addresses, because cyber-attack is prone to use particular IP addresses [11]-[13]. This method has an advantage that the classification is difficult to be avoided, since publicly registered IP addresses cannot be easily changed. However, this method has two problems. First, classification range of IP addresses is narrow, because the IP addresses that can acquire features are limited. Second, classification cost is high, because high-dimensional feature set is required to represent the features of IP addresses. From these researches, it can be said that domain-based detection approaches using blacklists tend to fail. And it can be also said that it is difficult to use domain names to classify malicious Web sites since domain names can be easily changed. To solve these problems, we propose a new method to detect a malicious Web site effectively by reducing avoidance from the blacklist. This method uses only the domain name for detection to expand the detection range of the malicious Web sites. We also propose a method to classify Web sites by using a part of IP address to reduce the classification cost.

## 3 CLASSIFICATION METHOD OF MALICIOUS WEB SITES

### 3.1 Approach

In this research, we propose a system to classify malicious Web sites using domain name features and IP address features together. In the detection using the domain name feature, we solve the drawback of the blacklist type detection by extending the detection condition using multiple detection conditions used in existing methods together. In the classification using the features of the IP address, we aim to maintain high classification accuracy while reducing classification cost by acquiring the Web site’s distribution features in all ranges of the IP address and expressing each feature with small amount of information.

According to Ref. [14], the distribution of malicious IP addresses is as shown in Fig.2. In Ref. [11], authors described that cyber-attacks tend to use specific IP address

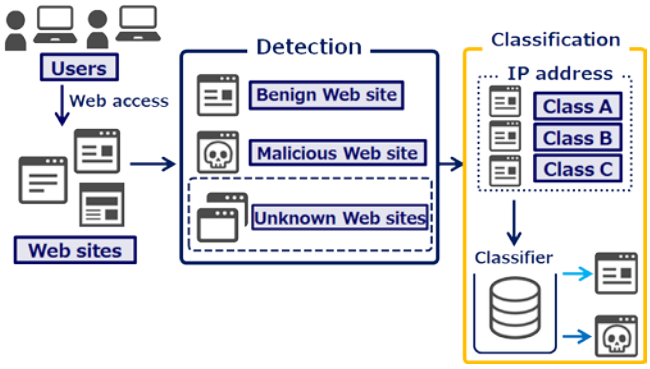



Figure 2: Approach to reduce the classification cost

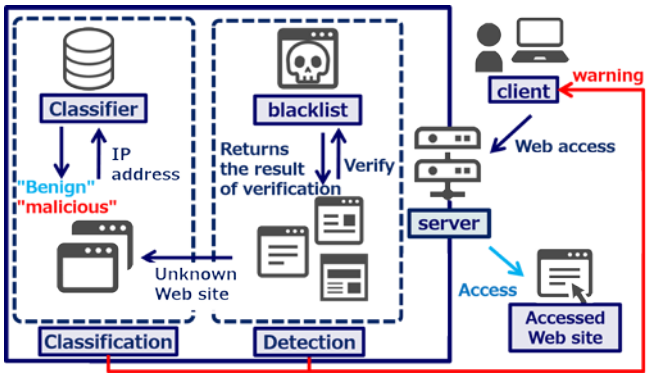



Figure 3: Overview of the proposed method

groups, and the usage frequency of malicious IP addresses also has different features for each IP address class. Using features of this usage frequency, we obtain the features of the range of all IP addresses and use it for classification. Also, based on the features of the network address space, only the network address part of each IP address is used for classification after binary conversion of the IP address. With this approach, it is possible to reduce the number of dimensions of feature vectors required for classification and enable high precision and low-cost classification.

## 3.2 Proposed Method

Figure 3 shows the outline of the proposed method using the approach to reduce classification cost. Web sites accessed by clients can be classified into three categories: malicious Web sites, benign Web sites, and unknown Web sites. It is important how to deal with these unknown Web sites. In this approach, we will classify unknown Web sites to malicious or benign at low cost by using IP addresses of only unknown Web sites.

Figure 4 shows the overview of the whole proposed method. The proposed method consists of two units. One is a unit to detect unknown Web sites by removing known malicious and benign Web sites. Another unit is to classify whether an unknown Web site is malicious or benign.

Details of the detection unit are described in Section 3.3, and details of the classification unit are described in Section 3.4.

The detection unit detects an unknown Web site by excluding malicious Web sites from the detection target using the blacklist of the domain name and transmits the IP

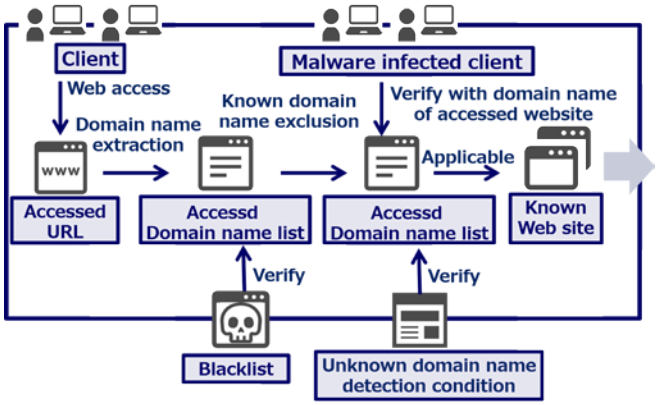



Figure 4: Details of detection method of unknown Web site

address of the corresponding Web site to the classification unit.

The classification unit classifies whether the unknown Web site is a malicious or benign Web site by using the features of the IP address transmitted from the detection unit. If the classification result is a benign Web site, the proposed method allows the client to access the Web site. On the other hand, if the classification result is a malicious Website, the proposed method interrupts communication by warning the client Also, the proposed method adds relevant malicious Web sites to the blacklist and keeps it up-to-date.

## 3.3 The Detection Method of Unknown Web Sites

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

In the detection unit, unknown Web sites are detected using the features of the domain name. The domain name can be obtained from the URL of the Web site. First, when communication from a client to a certain Web site passes through the DNS server, the detection unit collates the domain name of the destination against the blacklist and excludes known malicious domains. If the domain name of the destination does not exist in the blacklist, the domain name is compared with the detection condition based on the features of the domain name such as the length of the domain name, character type, and so on. The Web sites that do not satisfy these conditions are defined as unknown Web sites and are determined as classification targets in the classification unit.

Also, the proposed method uses the domain name accessed by the clients infected with malware to check unknown Web sites. Figure 6 shows details of detection method of malware infected clients. Generally, malwares attempt clients to access numerous malicious Web sites to expand infection of malware. Therefore, malicious Web sites are likely to be accessed simultaneously from multiple malware infected clients [10]. Tanaka's group proposed the method to search clients accessing Web sites with malicious domain.

The clients detected by this method are called malware infected clients. By applying this method, it is possible to detect malicious Web sites with malicious domains that many clients are forced to access. In our proposed method, we use

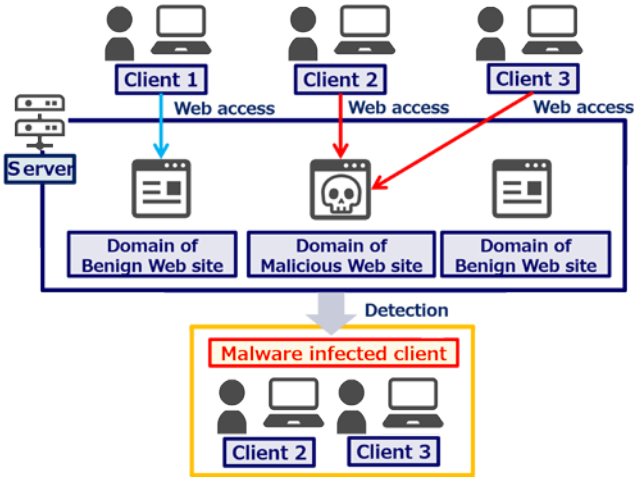
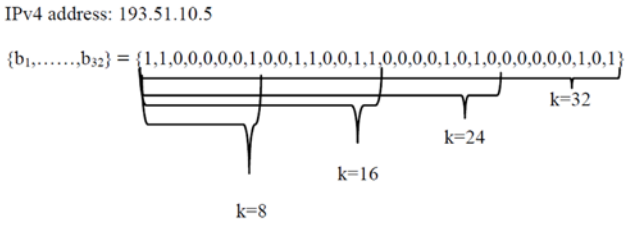



Figure 5: Detection method of malware infected clients




| k | Feature vector |
|---|---|
| 8 | $b_k$=1 (k=1, 2, 8)<br>$b_k$=0 (otherwise) |
| 16 | $b_k$=1 (k=1, 2, 8, 11, 12, 15, 16)<br>$b_k$=0 (otherwise) |
| 24 | $b_k$=1 (k=1, 2, 8, 11, 12, 15, 16, 21, 23)<br>$b_k$=0 (otherwise) |
| 32 | $b_k$=1 (k=1, 2, 8, 11, 12, 15, 16, 21, 23, 30, 32)<br>$b_k$=0 (otherwise) |

Figure 6: Examples of generating feature vectors

this malware infected clients to regard some Web sites as the known malicious Web sites. Finally, the IP address of Web sites which do not match the blacklist and the domain name detection conditions are used for classification and send to the classification unit.

## 3.4 The Classification Method of Malicious Web Sites

The classification unit is composed of two phases. First, the unit generates feature vectors from data of a blacklist in which known malicious Web sites are accumulated and data of a whitelist in which known benign Web sites are accumulated and uses these vectors as training data to construct classifiers (Section 3.4.1). Next, the unit classify unknown Web sites using the constructed classifiers (Section 3.4.2).

### 3.4.1 Construction of Classifier Using Training Datasets

In order to construct classifier, it is necessary to generate feature vectors. These feature vectors contain various features of the training datasets. Our method vectorizes the features of known malicious Web sites and benign Web sites and generate classifiers for classification using these feature vectors. The number of elements of the feature vector is called the dimension number. Since the number of dimensions of these feature vectors influences the cost of classification, we use the feature vectors with reduced dimension.

Table 1: Examples of labeling feature vectors of training datasets

| IP address | Feature vector | Label |
|---|---|---|
| 193.51.10.5 | 1,1,0,0,0,0,0,0,0,0,1,1,0,0,1,1 | 1 |
| 10.10.10.10 | 0,0,0,0,1,0,1,0 | 1 |
| 203.4.12.89 | 1,1,0,0,1,0,1,1,0,0,0,0,0,1,0,0,0,0<br>,0,1,1,0,0 | 0 |
| … | … | … |

Figure 7 shows a feature vector generation method. First, the target IP address is converted into bit string. Next, all bit strings are represented as k-dimensional vector $\{b_1,.....,b_k\}$. 3 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 are generated for each IP address. Finally, generated feature vectors of malicious are labeled as “1”, and generated feature vectors of benign are labeled as “0” in this system. Table 1 shows examples of labeling the feature vector of the training data set. The classifier used in this method is Support Vector Machine (SVM), which is one of pattern identification methods. Since we examined the method with the goal of improving classification accuracy for related research, we use SVM to keep the same condition. In addition, Ref. [6] clearly indicated that malicious Web sites are detected with high accuracy by using SVM. In addition, when the features of the IP addresses are frequently updated, an approach such as on-line learning may be suitable. Online SVM is one of such identification method. Since it is possible to follow the time change of learning data, on-line learning approach may be effective in the proposed method which learns data change on time. However, since the observational data set we are using is released every year, the frequency of re-learning by updating the data is presumed to be several times a year at most, as long as the data update frequency is this level. Therefore, we judged that batch type SVM algorithm has sufficient performance at present, and the proposed method classifies using the batch type SVM.

The proposed method constructs three classifiers based on feature vectors described above.

### 3.4.2 Classification of Test Datasets Using Constructed Classifier

An unknown IP address sent from the detection unit are classified by the constructed classifiers described in Section 3.4.1. Figure 8 shows that classification has three steps. First, the IP address passed from the detection unit is classified for

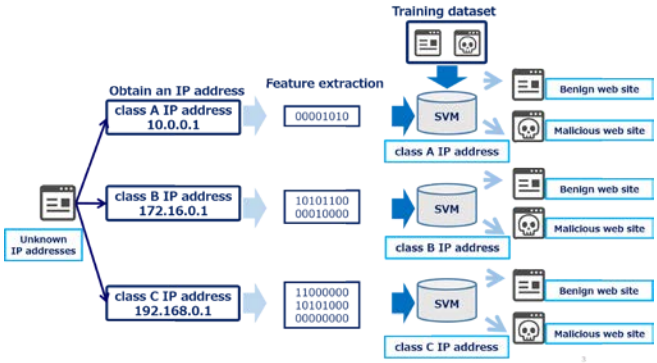



Figure 7: Process of the classification

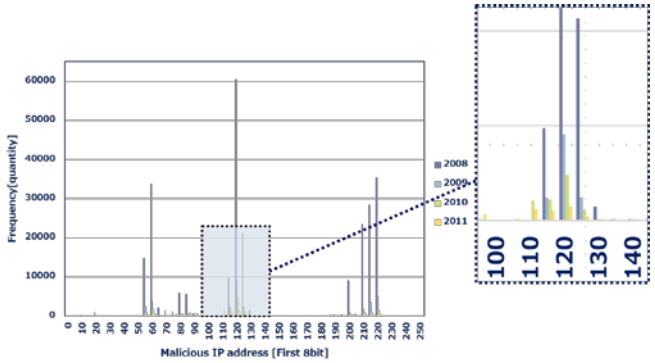



Figure 8: Usage frequency of malicious IP addresses (Enlargement of the first 8 bits of IP address around 120) (2008-2011)

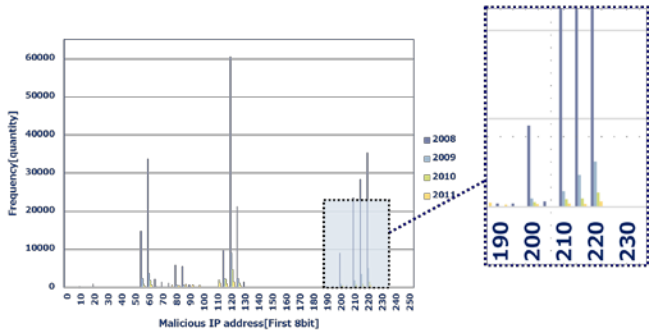



Figure 9: Usage frequency of malicious IP addresses (Enlargement of the first 8 bits of IP address 190-230) (2008-2011)

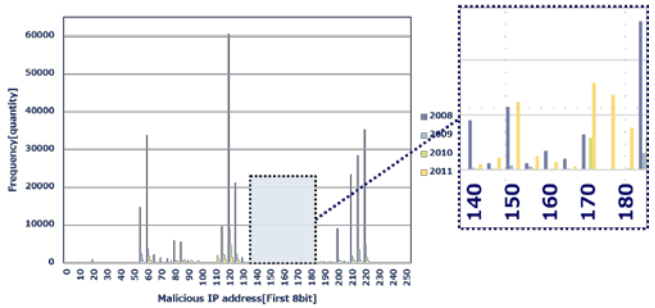



Figure 10: Usage frequency of malicious IP addresses (Enlargement of the first 8 bits of IP address 140-180) (2008-2011)

each IP address class in the classification unit. Next, feature vector is generated from this obtained unknown IP address. This feature vector is classified as malicious or benign by the classifiers.

Then, the classifier is updated by using IP addresses already known as malicious or benign and IP addresses for which classification has been completed as training data. As a result, the features of the malicious IP address distribution included in the classifier can maintain the latest state, and the classification accuracy can also be improved for unknown Web sites.

In order to realize these proposed methods, it is necessary to set the dataset used as training data in a state suitable for classification. Therefore, we analyzed the usage distribution of IP addresses class based on the data on malicious IP address usage frequency. The classifier pattern to be considered from the analysis result is the following three.

1) Classifiers using blacklist itself (General blacklist)
2) Classifiers using the features of IP addresses on blacklist without assuming temporal change.
3) Classifiers re-learned the temporal changes on the features of IP addresses on blacklist

In case of 3), it is necessary to analyze the IP address used as the training data to confirm whether it can cope with the temporal changes of the IP address included in the blacklist.

## 3.5 Temporal Change on IP address Distribution

We analyzed the usage status of each IP address class to check whether the distribution of malicious IP addresses changes over time. Figure 9 shows the usage frequency of malicious IP addresses from 2008 to 2011, when the number of IP addresses collected from CCC datasets [15] which is a bot observation data group containing malware is relatively large. At the part where the first 8 bits of the IP address are around 120, we can see that the usage frequency has decreased from 2008 to 2011. Also, Fig.9 shows at the part where the first 8 bits of the IP address are around 110.The malicious IP address is not used from 2008 to 2009. However, we can see that the usage frequency has increased from 2010. Also, Figure 10 shows at the part where the first 8 bits of the IP address are from 200 to 220. We can see that the usage frequency has decreased from 2008 to 2011.

In addition, Fig. 11 shows at the part where the first 8 bits of the IP address are from 170 to 180. We can see that the usage frequency has increased from 2010 to 2011. Therefore, it can be said that the temporal change of IP address distribution features. In addition, Fig.12 shows at the part where the first 8 bits of the IP address are from 170 to 180. we can see that the usage frequency has increased from 2010 to 2011. It can be said that the temporal change of IP address distribution features. From this result, the distribution of malicious IP addresses has changed every year in each address class. By extracting the features of each IP address class by each year and using the feature for classification, highly accurate classification corresponding to temporal changes in IP address usage can be realized. Using these features, the training data used for constructing the SVM

classifier in the classification unit is changed every year in the proposed system.

# 4 EVALUATION

## 4.1 Outline

In order to confirm the effectiveness of the proposed method, we evaluated the classifier constructed in phase 1 of the classification unit by three evaluation items such as accuracy, precision rate, and recall rate. In this paper, we define True Positive (TP) as a number that can classify malicious IP address correctly as malicious IP address, and True Negative (TN) as a number that can classify benign IP address correctly as benign IP address. And we also define False Positive (FP) as a number that can classify benign IP address incorrectly as malicious IP address, and False Negative (FN) as a number that can classify malicious IP address incorrectly as benign IP address. Accuracy (1), precision rate (2), recall rate (3) are calculated by the following formulas.

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

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

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

The dataset of benign and malicious are obtained from Malware Workshop (MWS) Datasets [14]. As malicious data, we use the CCC dataset (2008-2011) which is a bot observation data group containing malware specimens, and D3M (2010-2015) which is Web infectious malware data, from MWS dataset. As benign data, we use 50,000 data of Alexa Top Global Sites [15] (2016) and data created based on NCD in MWS Cup (2014) which is a white dataset. For the experiments, we extract the ratio of malicious IP address to benign IP address to be 8: 2, 5: 5, 2: 8 respectively from these datasets. Training datasets and test datasets are randomly extracted from malicious and benign data respectively.

In the previous section we confirmed the temporal change of the distribution of the malicious IP addresses. In order to evaluate the influence of this temporal change on classification accuracy, we compare the accuracy of the three types classifiers by using the first 8, 16, 24, 32 bits of each IP address.

## 4.2 Evaluation Experiment 1

In evaluation experiment 1, we compare the classification using blacklist and the classification of the proposed method to confirm the classification performance of each classifiers. We created 5 types of blacklists using malicious IP addresses that can be obtained from the datasets [14] for classification using blacklists. Each blacklist consists of 17000 IP addresses. We also created 5 types of test datasets for classification of the proposed method using malicious IP addresses and benign IP addresses that can be obtained from the datasets [14]. This test datasets consists of 20000 IP addresses, and the ratio of malicious IP address to benign IP address is 8:2 in each test datasets. The overview of evaluation experiment 1 is shown in Fig.12. We conducted experiments to compare the results of classification of two patterns using the same input test datasets. Pattern 1 is classification using blacklist, and pattern 2 is classification using the proposed method. In experiment pattern 1, 32-bit IP address is used as blacklist data. The classifier classifies by comparing the IP address of the input data with the IP address of the blacklist, and outputs the classification result. We calculate the average classification accuracy of 5 input test datasets. For experiment pattern 2, we use the same dataset as experiment pattern 1.

We construct classifiers of the proposed method using the dataset as training data. When test dataset is input, the classifiers of the proposed method classify according to the IP address class of the input IP address and outputs the classification result. We calculate the average classification accuracy by 5-fold cross validation. Table 2 shows the number of IP addresses used in evaluation experiment 1.

Table 3 shows the classification results of classification of the experiment pattern 1. On average, the accuracy was 18.762%, the precision rate was 5.268%, the recall rate was 100%, and accuracy and precision rate are very low.

Table 4 shows the average classification result of the experiment pattern 1. From this result, accuracies and precision rates were more than 80%, and recall rates were more than 90%. Compared to the classification result using blacklist, high values were obtained.

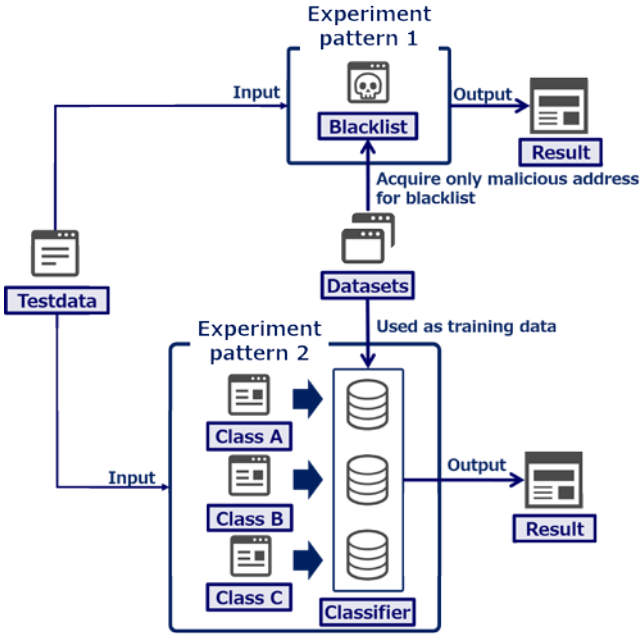



Figure 11: Overview of evaluation experiment 1

Table 2: Number of IP addresses used in experiment 1

| | Malicious IP addresses | Benign IP addresses |
|---|---|---|
| Blacklist | 322,687 | |
| Whitelist | | 538,156 |

Table 3: Detection result using blacklist

| | Accuracy | Precision | Recall |
|---|---|---|---|
| Testdata1 | 18.500 | 5.111 | 100 |
| Testdata2 | 18.545 | 5.185 | 100 |
| Testdata3 | 18.605 | 5.305 | 100 |
| Testdata4 | 19.335 | 5.372 | 100 |
| Testdata5 | 18.825 | 5.368 | 100 |
| Average | 18.762 | 5.268 | 100 |

Table 4: Average of classification results of the proposed method

| | Accuracy | Precision | Recall |
|---|---|---|---|
| Testdata1 | 81.945 | 86.543 | 91.983 |
| Testdata2 | 81.670 | 86.385 | 91.774 |
| Testdata3 | 81.520 | 86.177 | 91.747 |
| Testdata4 | 81.782 | 86.743 | 91.744 |
| Testdata5 | 81.735 | 86.442 | 91.751 |
| Average | 81.730 | 86.458 | 91.800 |

Table 5: Number of IP addresses used in experiment 2

| | Malicious IP addresses | Benign IP addresses |
|---|---|---|
| IP addresses of Class A | 49,164 | 40,667 |
| IP addresses of Class B | 3,523 | 10,735 |
| IP addresses of Class C | 75,000 | 14,288 |

Table 6: Result of IP address of Class A

| | Accuracy | Precision | Recall |
|---|---|---|---|
| Case1(k=8) | 84.060 | 90.253 | 89.866 |
| Case2(k=16) | 83.743 | 89.893 | 89.797 |
| Case3(k=24) | 83.760 | 89.890 | 89.818 |
| Case4(k=32) | 83.878 | 89.887 | 89.953 |

Table 7:Result of IP address of Class B

| | Accuracy | Precision | Recall |
|---|---|---|---|
| Case1(k=8) | 83.541 | 92.157 | 87.855 |
| Case2(k=16) | 81.696 | 89.070 | 88.162 |
| Case3(k=24) | 82.945 | 90.276 | 88.610 |
| Case4(k=32) | 83.257 | 90.525 | 88.761 |

Table 8: Result of IP address of Class C (8:2)

| | Accuracy | Precision | Recall |
|---|---|---|---|
| Case1(k=8) | 81.781 | 90.524 | 87.191 |
| Case2(k=16) | 81.201 | 89.965 | 86.981 |
| Case3(k=24) | 81.243 | 90.000 | 87.001 |
| Case4(k=32) | 81.222 | 90.586 | 86.564 |

Table 9: Result of IP address of Class C (5:5)

| | Accuracy | Precision | Recall |
|---|---|---|---|
| Case1(k=8) | 67.252 | 73.824 | 65.248 |
| Case2(k=16) | 67.252 | 73.824 | 65.248 |
| Case3(k=24) | 67.280 | 73.712 | 65.310 |
| Case4(k=32) | 66.034 | 69.050 | 65.122 |

## 4.3 Evaluation Experiment 2

In evaluation experiment 2, we experimented in detail to examine whether classification into IP address class is appropriate. We confirm the effectiveness of the classification using features of each IP address class. For each IP address corresponding to the IP address classes A, B, and C, we evaluated the classification accuracies using the first 8 bits, 16 bits, and 24 bits part of the IP address which is the network address part of each address class as feature.

The evaluation procedure is as follows. First, test datasets are input to the proposed method with the classifier generated by the training data, and classification result is acquired. Next, the classification accuracy is calculated by 5-fold cross validation. Finally, the classification accuracy for each feature vector used for input is compared and evaluated. The number of IP addresses used in evaluation experiment 2 is shown in Tables 5.

Evaluation experiment 2 is conducted in 4 cases. For the number of first bits of the IP address used for classification, the classification results were evaluated by 8 bits for Case 1, 16 bits for Case 2, 24 bits for Case 3, and 32 bits for Case 4.

Table 6 shows the classification results for IP address class A. The accuracy and the precision rate achieved the highest value in Case 1 using only the network address part of IP address class A as the feature vector. On the other hand, the recall rate achieved the highest value in Case 4. Although the accuracies of the Ref. [11] were from 74.7% to 75.1%, the accuracy of the proposed method was 84%.

Table 7 shows the classification results for IP address class B. The accuracy and the precision rate achieved the highest value in Case 1, and the recall rate achieved the highest value in Case 4. Therefore, in Case 2 where only the network address part of IP address class B is used as the feature vector, the highest value cannot be achieved. The accuracies of the proposed method were 81.6% to 83.5% which were lower than the accuracies of the Ref. [11] (84.6% to 86.2%).

Table 8 shows the classification results for IP address class C. The accuracy and the recall rate achieved the highest value in Case 1, and the precision rate achieved the highest value in Case 4. The accuracy, the recall rate, and the precision rate of Case 3 where only the network address part of IP address class C are used as the feature vector were lower than those of Case 1, 2 and 4. Also the accuracies of the proposed method were 81.2% to 81.7% which were lower than the accuracies of the Ref. [11] (85.1% to 88.5%).

Table 9 shows the classification results when the ratio of malicious IP address to benign IP address is set to 5: 5 in IP address class C. The accuracy and recall rate achieved the

highest value in Case 3 using only the network address part of IP address class C as the feature vector. The precision rate achieved the highest value in Case 1 and 2. Also, in Case 4, the accuracy, the precision rate and the recall rate achieved the lowest values.

## 4.4 Evaluation Experiment 3

In evaluation experiment 3, we made classification based on training data using features of IP address every year. And we evaluate the effectiveness of the training data in each year according to the accuracy obtained every year. The number of IP addresses used in evaluation experiment 3 is shown in Tables 10, 11, and 12.

Figure 13 shows the results with IP addresses corresponding to IP address class A. The accuracy of classification varies depending on the year.

Figure 14 shows the results with IP addresses corresponding to IP address class B. While the classification accuracy of Ref. [11] is from 84.6% to 86.2%, the classification accuracy of the proposed method exceeded 90% from 2008 to 2009.

Figure 15 shows the results with IP addresses corresponding to IP address class C. While the classification accuracy of Ref. [11] is from 85.1% to 88.5%, the classification accuracy of the proposed method exceeded 85% in 2008 and 2011.

Table 10: Number of IP addresses of Class A used in experiment 3

| | Malicious IP addresses | Benign IP addresses |
|---|---|---|
| 2008 | 159,103 | 40,897 |
| 2009 | 22,537 | 40,897 |
| 2010 | 14,369 | 40,897 |
| 2011 | 9,538 | 40,897 |

Table 11: Number of IP addresses of Class B used in experiment 3

| | Malicious IP addresses | Benign IP addresses |
|---|---|---|
| 2008 | 200 | 10,735 |
| 2009 | 22,537 | 10,735 |
| 2010 | 14,369 | 10,735 |
| 2011 | 945 | 10,735 |

Table 12: Number of IP addresses of Class C used in experiment 3

| | Malicious IP addresses | Benign IP addresses |
|---|---|---|
| 2008 | 97,688 | 14,288 |
| 2009 | 11,619 | 14,288 |
| 2010 | 4,073 | 14,288 |
| 2011 | 1,854 | 14,288 |

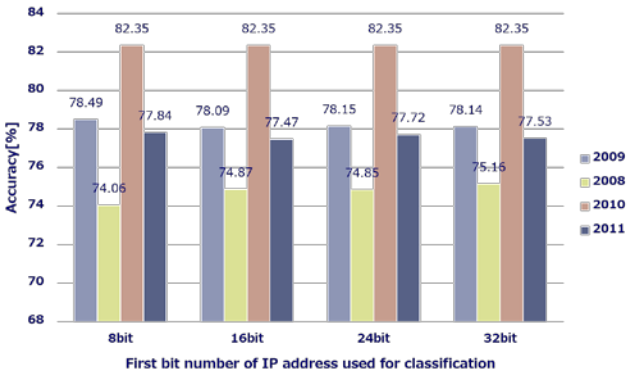



Figure 12: The result of evaluation experiment 3 in IP address of Class A

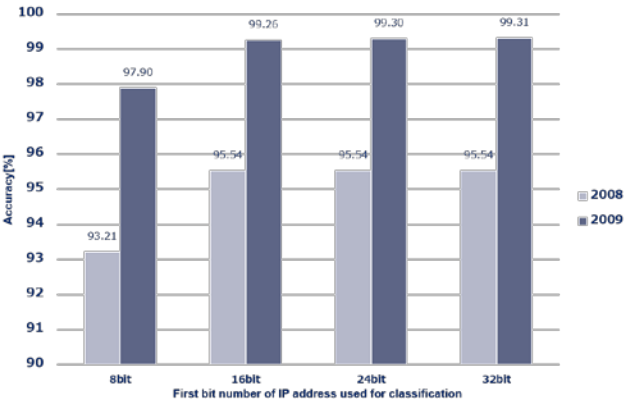



Figure 13: The result of evaluation experiment 3 in IP address of Class B

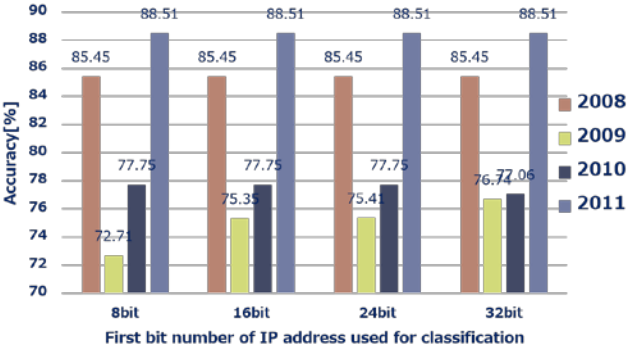



Figure 14: The result of evaluation experiment 3 in IP address of Class C

## 4.5 Discussion

From the results of evaluation experiment 1, since the accuracy showed low values, the classification of malicious Web sites using blacklist has low generalization ability and has no ability to deal with unknown Web sites. From the result of evaluation experiment 2, the proposed method is effective for unknown malicious Web sites with the address of IP address class A, because the sufficient accuracy can be maintained by classification using only the network address part. On the other hand, as for the addresses of IP address class B and IP address class C, overall accuracy is lower

than those of IP address class A. There are two possible causes. First, since the number of IP addresses of the training data set is extremely small, there is a possibility that the feature may not clearly appear. Second, there is a possibility that time changes may occur in the features of malicious IP addresses. Finally, we describe the experimental results of evaluation experiment 3. From section 3.6, it was found that time changes were observed in the features of malicious IP addresses.

Based on these analysis, it is possible to improve classification accuracy by considering temporal change of features of malicious IP addresses. In Ref. [11], the classifiers are constructed with all bits of IP address as features. However, in the proposed method, it is shown that the same accuracy can be obtained by using the features of the temporal change of the appearing IP address distribution even with the classifiers using only the network part of each IP address. From these results, it is difficult to maintain accuracy by the method that only accumulates and uses the malicious IP address data group (blacklist). However, by changing the state of training data for each IP address class and each year, it is possible to maintain the features of malicious IP address and to maintain classification accuracy. Therefore, it is possible to deal with unknown Web sites by finding the range of network address group with variation of features among IP address classes.

## 5 CONCLUSION

In this paper, we proposed 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. As a result of evaluation experiment about classifiers using the features of IP addresses on blacklist without assuming temporal change, high classification accuracy was provided in IP address Class A, and we confirm the effectiveness of the proposed classification method. However, from the results of evaluation experiment about classifiers re-learned the temporal changes on the features of IP addresses, it is shown that the same accuracy can be obtained by using the features of the temporal changes of the appearing IP address distribution even with the classifiers using only the network part of each IP address.

In the future, further research should be done to investigate the feature of malicious Web sites to improve the accuracy of classification. For more effective classification, further analysis of the data in each IP address class. On the other hand, although significant difference in features is observed for each IP address class in the current usage status of IP address, considering the address distribution by CIDR, there is a possibility of further improving the classification accuracy. However, although there are many CIDR IP addresses by 24-bit prefix corresponding to IP address class C in the investigation result by Ref. [16], the influence on the classification accuracy is unknown in the case of using addresses with other prefix lengths. Also in the classification method using the IP address, it is necessary to study how to classify it as a benign IP address when a new benign Web site is constructed for the IP address once classified to be malicious. In addition, since we examined the method with the goal of improving discrimination accuracy for related research, we use SVM to keep the same condition. In addition, although the proposed method uses SVM according to related research at present, in the future it is necessary to examine concrete classifier selection and design.

## 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] R. Farmer, and B. Glass, "Building Web Reputation Systems, " Yahoo! Press, (2010).

[4] "What is IDS/IPS? | JUNIPER NETWORKS," <https://www.juniper.net/us/en/products-services/what-is/ids-ips/> [Accessed Oct 19, 2017]

[5] 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).

[6] 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).

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

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

[9] 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).

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

[11] D. Chiba, K. Tobe, T. Mori, and S. Goto, "Analyzing Spatial Structure of IP Addresses for Detecting Malicious Web sites," Journal of Information Processing Vol.21, No.3, pp.539-550, (2013).

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

[13] D. Chiba, T. Yagi, M. Akiyama, and T. Mori, "Correlation Analysis Between IP Addresses Used in Varie-

ty of Attacks," Proceedings of the Computer Security Symposium 2011(CSS2011), pp.185-190, (2013) *(in Japanese)*.
[14] Y. Takata, M. Terada, J. Murakami, T. Kasama, K. Yoshioka, and M. Hatada, "Datasets for Anti-Malware Research ~MWS Datasets 2016~," IPSJ SIG Technical Report, Vol.2016-CSEC-74, No.17, pp. 1-8, (2016). (*in Japanese*).
[15] Alexa Internet, Inc., "The top 500 sites on the web," <http://www.alexa.com/topsites> [Accessed January 17, 2017].
[16] University of Oregon Route Views Project, <http://www.routeviews.org/> [Accessed December 27, 2017].

(Received October 20, 2017)
(Revised December 27, 2017)

**Shihori Kanazawa** received B.E. degree in systems information science from Future University Hakodate, Japan in 2016. She is a graduate student of Future University Hakodate. Her research interests include Web security using the detection of malicious Web sites. She is a student member of IPSJ.

**Yoshitaka Nakamura** received B.E., M.S., and Ph.D. degrees from Osaka University in 2002, 2004 and 2007, respectively. He is currently an associate professor at the School of Systems Information Science, Future University Hakodate. His research interest includes information security and ubiquitous computing. He is a member of IEEE, IEICE, and IPSJ.

**Hiroshi Inamura** He is a professor of School of Systems Information Science, Future University Hakodate, since 2016. His current research interests include mobile computing, system software for smart devices, mobile/sensor network, and their security. He was an executive research engineer in NTT docomo, Inc. He received B.E., M.E. and D.E. degree in Keio University, Japan. He is a member of IPSJ, IEICE, ACM and IEEE.

**Osamu Takahashi** received M.E. degree from Hokkaido University in 1975. He is currently a professor at the School of Systems Information Science, Future University Hakodate. His research interest includes ad-hoc network, network security, and mobile computing. He is a member of IEEE, IEICE.