Master Thesis

Research Toward Cyber Attack Prediction using Social Data Analysis

(ソーシャルデータ解析に基づいたサイバー攻撃予測に関する研究)

Author: Munkhdorj BAATARSUREN
Supervisor: Yuji SEKIYA (Professor)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF ENGINEERING
AT
THE UNIVERSITY OF TOKYO

February, 2016

Abstract

Nowadays, the number and severity of cyber attacks are constantly growing, so the detection and countermeasure of cyber attack are essential. The most common methods used in cyber attack detection are signature scan and anomaly detection. In the case of applying these approaches, a countermeasure against an upcoming cyber attack is made only if a signature of cyber attack or an anomaly is detected. That means cyber defense systems encounter cyber attacks with no preparation, so except the detection functionality, the prediction of cyber attack is necessary.

On the other hand, intended cyber attacks targeting on certain a organization or individual accounts for the huge number of all the cyber incidents. Also, some researchers implied that an intended cyber attack is mostly caused by particular social events. Therefore, I came up with the idea of predicting cyber attack by analyzing social activities.

To evolve the existing cyber attack detection mechanism, the functionality of cyber attack prediction has to be realized. Accordingly, I decided to verify the possibility of the prediction of cyber attack prediction based on social data analysis. I defined the *prediction of cyber attack* as the combination of the prediction of attack motivation, opportunity, and timing. The challenges for achieving the goal are disclosure of useful social data, development of dataset creation method, development of prediction algorithm, and practical evaluation. In this thesis, I attempt to discover an useful social data for the prediction of cyber attack motivation and opportunity. I also, attempted to promote the dataset creation method for the further analysis.

I collected five types of social data including an archive of cyber incidents, the news articles related to cyber attack victims, security vulnerability feeds, the tweets posted by some hacktivists, and the tweets posted by common people in Japan. All of the collected data will also be open for the people who use the data in purpose of research.

For the prediction of cyber attack motivation, the news articles were used as the dataset. I label the dataset based on the changes in the number of articles published before an attack. In the evaluation experiment, few types of features and classification algorithms were used. As a result, using Artificial Neural Networks and the core keywords extracted from the news articles directly correlated to a cyber attack or the news articles not correlated to cyber attack brought better precision/recall. I also investigated the number and mood of the two types of twitter feeds, but no useful pattern for the prediction of cyber attack motivation was found.

For the prediction of cyber attack opportunity, the security vulnerability feeds were used as the dataset. I referenced Exploit Database which is the public database storing the information on security vulnerability exploits and labeled the dataset as four classes, webapps, dos, remote, and local. In the evaluation experiment, few types of features as basic security metrics and the core keywords selected from the textual description of the vulnerability feeds. The precision/recall of the prediction result was better when using the core keywords as the feature and Artificial Neural Networks as the prediction algorithm.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Background

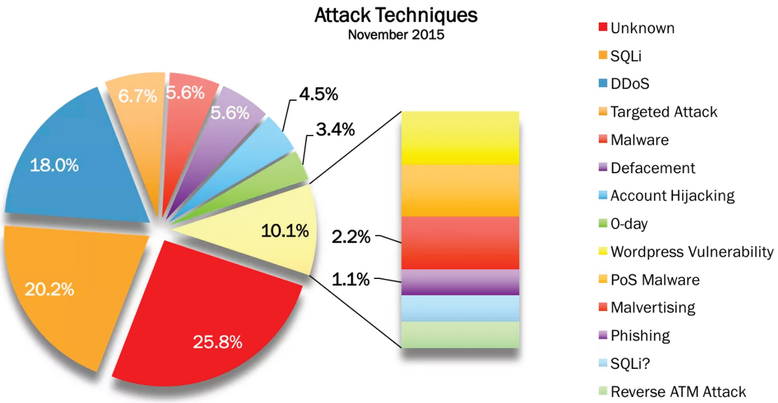


Figure 1.1 Attack techniques mainly used in 2015 [1]

In recent years, an increase in the damage caused by cyber attacks has become a severe social problem. The existing cyber attack defense systems, or, IDSs (Intrusion Detection System) [2] are deeply dependent on the security experts and the accuracy of detection functionality. Detection techniques and countermeasures are being improved in accordance with the attack methods that have continuously been discovered one after another.

However, there are limitations in the existing systems, that means it is no longer able to detect all of attacks with a high degree of accuracy. In many cases of cyber incident, system administrators did not notice even that they have been attacked. So when the attack is detected, the damage caused by the attack will be at a severe level. As depicted in Figure 1.1, according to the web site called Hackmageddon [1] that aggregates statistical information on cyber attacks, about 26% of the victims of the cyber incidents occured in 2015 did not notice that they were being attacked. Thus, improving the accuracy of the detection of cyber attack has become an important issue.

As illustrated in the left half of Figure 1.2, the existing cyber attack defense systems can be divided into two types by the mechanism used to detect intrusion [3], [4]. The first is signature-based systems that operate by scanning the signature collected and registered by security experts or organizations. The second is anomaly-based systems which detect abnormal operations by analyzing the information collected from the target system. In the both mechanisms, a countermeasure is taken against an attack only if a signature of intrusion or an abnormal operation is found. In other words, no preventive action is carried out unless a sign of cyber attack is detected. Because, only real-time parameters such as network traffic and system stats are monitored, and forecasting cyber attack by using those parameters is nearly impossible.

On the other hand, the number of cyber attacks against certain organizations or individuals is keep rising day by day, and most of them are intended attacks. According to *2016 Internet Security Threat Report* [5], the number of zero-day vulnerabilities, security holes in software used for intended cyber attack, discovered more than doubled to 54, a 125 percent increase from the year before. Or put another way, a new zero-day vulnerability was found every week (on average) in 2015.

From the above, predicting cyber attack is essential in cyber security these days. I hope that the intended cyber attacks can be predicted and, this study focuses on the possibility of cyber attack prediction. I define *prediction* as the combination of the following three.

1. Motivation[1]
2. Opportunity[2]
3. Timing[3]

This time, I introduce the experimental results for 1 and 2.

There are various types of cyber attack [6], and in the subsequent chapters and sections, I simply classify cyber attack as below. Depending on the purpose and motivation, cyber attack can be categorized as cyber crime, hacktivism, cyber espionage, and cyber war. By the techniques used in intrusion, cyber attack is categorized as SQLi (SQL Injection), DDoS (Distributed Denial of Service), Targeted attack, Account hijacking, etc.

1.2 Existing Problems and Research Motivation

The conventional cyber defense approach is reactive, that is, a countermeasure is carried out from the moment a sign of cyber attack is detected. In this way of defense, it comes that defense systems face cyber attacks with no preparation, and it is a severe problem. To solve this problem, it is necessary to the existing systems to be pre-activated. In addition to the countermeasure and detection, a functionality of prediction is essential.

Although there are several studies such as [7], [8] that focus on the possibility of cyber attack prediction, their proposed methods are closer to early detection than prediction. Because, the data to be analyzed are network traffic and system parameters, and as mentioned above, predicting cyber attack by monitoring such data does not function until an attack begins.

1.3 Research Goal and Contribution

In the real world, there is weather forecast, and thanks to that we can avoid bad weather conditions such as heavy rain and strong wind. However, there is nothing as weather forecast in cyber world,

[1] Predicting the probability of cyber attack targeting on a certain objective

[2] Determining the highly probable type of cyber attack

[3] Estimating the timing of cyber attack

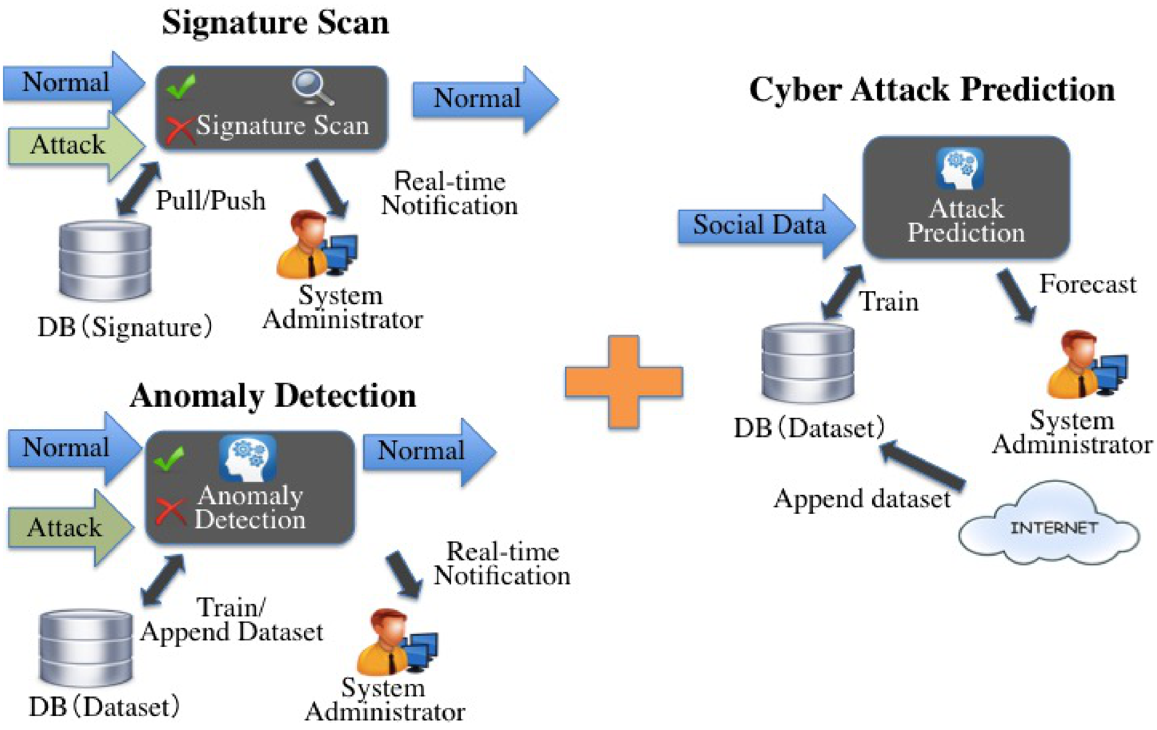


Figure 1.2 The envisioned cyber attack detection

and to bring such functionality to cyber world is one of the motivations for this study. Also, the hope of predicting intended cyber attacks from social activities motivated this study. After a survey, it was found that some researchers had been confirmed that social actions are causing cyber attack [9], [10]. Thus, I decided to focus on the possibility of cyber attack prediction based on social activity analysis.

As depicted in Figure 1.2, the cyber defense systems have to include the functionality of cyber attack prediction in addition to the existing detection techniques. I see that there are four stages or challenges to achieve this goal as follows.

1. Discovery and collection of useful data
2. Creation and validation of dataset
3. Validation of prediction method
4. Evaluation and practical implementation

In this thesis, I aim to verify the possibility of predicting cyber attack from social activities and propose the preliminary methods for the first three steps.

The main contributions of this research to cyber security are adopting the idea of predicting cyber attack by analyzing social activities and provision of the preliminary methods for dataset creation and prediction. I hope that the achievements introduced in this thesis will lead to the ultimate goal of realizing a practical cyber attack prediction system.

1.4 Organization of Thesis

This thesis consists of six chapters. In Chapter 2, some of the related works, totally five, are introduced. The first is a study on the correlation between cyber attack and social activities. The second and third are the examples related to predicting real world events by analyzing social data as news articles and tweets. The fourth and fifth works are related to exploiting security vulnerabilities for cyber attack. Chapter 3 gives the background knowledge for some techniques used in cyber attack prediction. In Chapter 4, the preliminary ideas for dataset creation and prediction methods are presented. In Chapter 5, the social data to be analyzed in this research and the collecting mechanism are described. In Chapter 6, the results of the experiment for validating the dataset and prediction methods are presented. The experiments were conducted with three phases, some of the preliminary experiments, the experiment for predicting cyber attack motivation and experiment for predicting highly possible type of cyber attack. Finally, Chapter 7 concludes the experiment results and discusses the current issues and future works for this research.

Chapter 2

RELATED WORKS

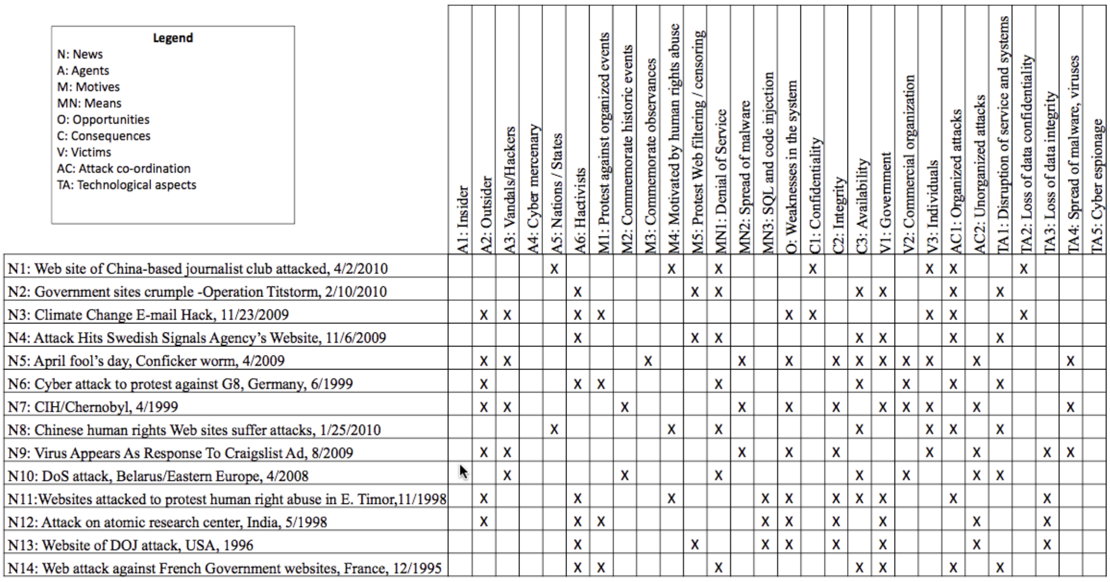
In this chapter, totally five works are reviewed as related works. I chose those works from the view point of the correlation between cyber attack and social events, and predicting real world events by inspecting the public insights.

2.1 Cyber Attack and Social Activities

It has been emphasized in some publications [9], [10] that a cyber attack is backed by certain social actions. In these studies, the correlation between cyber attacks and social actions is investigated by analyzing news articles.

Sharma et al., [9] divided social action into four categories of SPEC (Social, Political, Economic, and Cultural) and collected the news articles on cyber attack incidents and analyzed them by using FCA (Formal Concept Analysis) [11]. FCA is a technique for finding and expressing the association of the background and concept of a particular phenomenon in mathematical way. As depicted in Figure 2.1, the basic data format in FCA is a binary table as K = (G, M, J), G is object set, M is set of properties, and J is the association of G and M. In this case, G is a news article on a cyber attack, M is a set of cyber attack factors such as victim, geographical location, timing, duration, correlation of attacks, and the technique used in the attack.

Several things were made clear as the conclusion of the FCA analysis [9]. For instance, *when government agencies are victims, the attacker will be hacktivists with probability of 75%,* and more generally *for cyber attacks where the social motive is to commemorate historic events, agents are vandals and hackers, the attack co-ordination is unorganized and the victims are commercial organizations*. The conclusions of Sharma et al., was one of the motivations for the idea of inspecting social events to predict cyber attack in this thesis.



Legend
N: News
A: Agents
M: Motives
MN: Means
O: Opportunities
C: Consequences
V: Victims
AC: Attack co-ordination
TA: Technological aspects

	A1: Insider	A2: Outsider	A3: Vandals/Hackers	A4: Cyber mercenary	A5: Nations / States	A6: Hactivists	M1: Protest against organized events	M2: Commemorate historic events	M3: Commemorate observances	M4: Motivated by human rights abuse	M5: Protest Web filtering / censoring	MN1: Denial of Service	MN2: Spread of malware	MN3: SQL and code injection	O: Weaknesses in the system	C1: Confidentiality	C2: Integrity	C3: Availability	V1: Government	V2: Commercial organization	V3: Individuals	AC1: Organized attacks	AC2: Unorganized attacks	TA1: Disruption of service and systems	TA2: Loss of data confidentiality	TA3: Loss of data integrity	TA4: Spread of malware, viruses	TA5: Cyber espionage
N1: Web site of China-based journalist club attacked, 4/2/2010					X					X		X				X					X	X			X			
N2: Government sites crumple -Operation Titstorm, 2/10/2010						X					X	X						X	X			X		X				
N3: Climate Change E-mail Hack, 11/23/2009		X	X			X	X								X	X					X	X			X			
N4: Attack Hits Swedish Signals Agency's Website, 11/6/2009						X					X	X						X	X			X		X				
N5: April fool's day, Conficker worm, 4/2009		X	X						X				X		X		X	X	X	X	X		X				X	
N6: Cyber attack to protest against G8, Germany, 6/1999		X				X	X					X						X		X		X		X				
N7: CIH/Chernobyl, 4/1999		X	X					X					X		X		X		X	X	X		X				X	
N8: Chinese human rights Web sites suffer attacks, 1/25/2010					X					X		X						X			X	X		X				
N9: Virus Appears As Response To Craigslist Ad, 8/2009		X	X										X		X		X				X		X			X	X	
N10: DoS attack, Belarus/Eastern Europe, 4/2008			X					X				X						X		X			X	X				
N11:Websites attacked to protest human right abuse in E. Timor,11/1998		X				X				X				X	X		X	X	X			X				X		
N12: Attack on atomic research center, India, 5/1998		X				X	X							X	X		X		X				X			X		
N13: Website of DOJ attack, USA, 1996						X					X			X	X		X		X				X			X		
N14: Web attack against French Government websites, France, 12/1995						X	X					X						X	X			X		X				

Figure 2.1 Sample FCA table (vertical: News/social events, horizontal: Cyber attack factors) [9]

2.2 News Article Analysis for Stock Price Prediction

In this section, a study of some researchers who attempted to acquire the social insight from financial news articles to predict stock price is introduced as an example of predicting real world events by analyzing social data.

Shynkevich et al., [12] proposed a method to forecast upcoming trends in the capital markets by analyzing financial news articles. Unlike other researchers, Shynkevich et al., used the multiple kernel learning technique to effectively combine information extracted from stock-specific and subindustry-specific news articles for prediction of an upcoming price movement. Multiple kernel learning is a method which manage with the issue of a kernel choice [13]. This technique decreases the risk of wrong selection of a kernel to some extent by adopting a set of kernels and for each kernel determined its weight such that every prediction are built on weighted aggregate of various kernels.

As feature for training models, they used unique keywords appeared in three or more articles. They chose 500 unique keywords in descending order of the computed Chi-square value [12] which is defined as Equation 2.1.

$$\chi^2 = \sum_{j=1}^{4} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{2.1}$$

Where O_{i1} is the observed frequency of the feature i within the set of positive messages, E_{i1} is the expected frequency, and j indicates four possible outcomes when a feature occurred within positive messages ($j = 1$); occurred within negative messages ($j = 2$); did not occur within positive messages ($j = 3$); did not occur within negative messages ($j = 4$).

News articles were divided into these two categories based on their relevance to a targeted stock and analyzed by separate kernels. Also, each article was transformed into a feature vector of 500 elements, and the elements were represented by the $TF * IDF$ value. TF (Term Frequency) is the frequency of the keyword in the article. IDF (Inverse Document Frequency) $= \log_n N$ where N is the total number of articles and, n is the total number of articles where the term appeared. The experimental results showed that utilizing two categories of news improved the prediction accuracy in comparison with methods based on a single news category. The best accuracy for predicting stock price was 81.6% when using multiple kernel learning.

2.3 Twitter Analysis for Stock Price Prediction

In this section, a study of some researchers who attempted to predict stock price by analyzing tweets is introduced as another example of predicting real world events based on social insight.

Bollen et al., [14] assumed that emotions could profoundly affect individual behavior and decision-making, and attempted to predict trends in stock markets by inspecting social mood. They investigated whether measurements of collective mood states derived from large-scale Twitter feeds were correlated to the value of the Dow Jones Industrial Average (DJIA) over time. They analyzed the text content of daily Twitter feeds by using two types of sentiments, negative and positive, and six states of mood (Calm, Alert, Sure, Vital, Kind, and Happy). Bollen et al., cross-validated the resulting mood time series by comparing their ability to detect the public 's response to the presidential election and Thanksgiving day in 2008.

They trained an Artificial Neural Network on the changes of sentiments and mood correlated to the changes of stock price. Their experimental results showed that the accuracy of DJIA predictions could be significantly improved by the inclusion of specific public mood dimensions but not others. Also, the best accuracy of predicting the daily up and down changes in the closing values of the DJIA was 87.6%. The result of work done by Bollen et al., is not directly related to this thesis, but I adopt the idea of predicting real world events by analyzing the public mood state derived from Twitter feeds.

2.4 Security Vulnerability Feeds Analysis

Security vulnerability feed is considered as a factor used for cyber attack. There are some researchers who have conducted a practical experiment to determine the timing of cyber attack after the public release of a vulnerability feed.

Holm et al., [15] conducted a statistical analysis of how 18 security estimation metrics based on CVSS [16] data correlate with the time-to-compromise (TTC) of 34 successful attacks. The empirical data originates from an international cyber defense exercise involving over 100 participants and were collected by studying network traffic logs, attacker logs, observer logs, and network vulnerabilities. They used the vulnerability exposure model proposed by Boyer et al., [17], which is expressed as Equation 2.2.

$$T = \sum_{i=1}^{N} (t - T_i) \tag{2.2}$$

N is the number of open known vulnerabilities that apply to a system, T_i is the discovery date of vulnerability i, t is current date, and T is the total number of vulnerability days.

The practical exercise was conducted in the environment depicted in Figure 2.2 including a

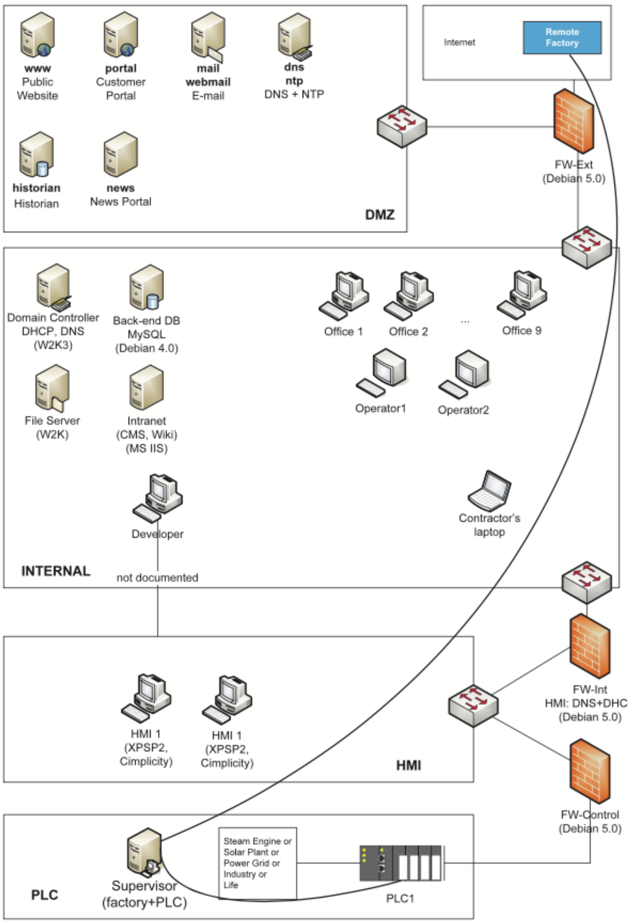


Figure 2.2 The overall network architecture used during the practical exercise [15]

16-person red team (i.e., attackers), six blue teams of 6-10 people per team (i.e., defenders), a white team (i.e., game management), a green team (i.e., technical infrastructure management) and one observer per team. The participants of the exercise primarily included computer security specialists and computer security researchers originating from various northern European governments, military, the private sector, and academic institutions.

Though the consequences of their experiment were not directly connected to the prediction of cyber attack, it became clear that if the attackers found there was a vulnerability in the targeting system, they could abuse it in few hours, not days. The result of the experiment was one of the motivations of the idea of making use of vulnerability feeds to predict highly possible cyber attack type.

On the other hand, Sabottke et al., [18] and Bozorgi et al., [19] attempted to predict whether a security vulnerability is exploited or not. Their goal was an early warning for security vulnerability exploits based on social data analysis. As the dataset for the prediction/classification model, they used twitter feeds combining with the vulnerability feeds. To label the dataset, they referenced Exploit Database [20], Microsoft Security Advisories [21], and Symantec [22]. Bozorgi et al., evaluated their model which predicts whether and how soon individual vulnerabilities are likely to be exploited, and they achieved 87.5% as the best accuracy. Meanwhile Sabottke et al., evaluated few types of features such as pure twitter, pure vulnerability, and combined. Also, they implied that the precision of the result brought by the pure twitter was higher. This work motivated my research, and some of the useful ideas were applied to my research.

Chapter 3

BACKGROUND KNOWLEDGE

Before proceeding to the further steps, some useful concepts and techniques adopted in the dataset creation and prediction methods have to be introduced. This chapter gives a brief introduction of those concepts and techniques.

3.1 K-Means Clustering

K-means is one of the widely used unsupervised clustering algorithms, which was firstly introduced about half a century ago. The basic concept of K-means is collecting geometrically closer nodes in the same cluster. The mathematical definition of the algorithm is as follows [23].

Let $X = \{x_i\}, i = 1, ..., n$ be the set of n d-dimensional points to be clustered into a set of K clusters, $C = \{c_k, k = 1, ..., K\}$. K-means algorithm finds a partition such that the squared error between the empirical mean of a cluster and the points in the cluster is minimized. Let μ_k be the mean of cluster c_k. The squared error between μ_k and the points in cluster c_k is defined as Equation 3.1.

$$J(c_k) = \sum_{x_i \in c_k} ||x_i - \mu_k||^2 \tag{3.1}$$

Also, the goal of the algorithm is to minimize the sum of the squared error over all K clusters and can be expressed as Equation 3.2.

$$J(c_k) = \sum_{k=1}^{K} \sum_{x_i \in c_k} ||x_i - \mu_k||^2 \tag{3.2}$$

The common procedures in the K-means implementations are as follows.

1. Select an initial/random partition with K clusters, and repeat step 2 and 3 until the movement of the member nodes among the clusters gets stable.

2. Generate a new partition by moving each member to its closest cluster center.

3. Recalculate cluster centers.

3.2 Expressing The Mood of Textual Statement

Mood analysis of textual statements and comments is one of the basic techniques used for knowing the public insight against a certain target. For example, the well-known types of mood are anger, happiness, sadness, and surprise. However, the basic two types of mood, *positive* and *negative*, are most commonly used in text analysis.

I define the term *sentiment score* as the metric for expressing how the mood of a statement is severe. The sentiment score s_i of the statement i is defined in $[-1.0; 1.0]$, that is -1.0 and 1.0 are respectively the extreme values for negative and positive mood. Also, in general, a lexicon dictionary annotated by human, where each lexicon labeled with the category of mood is needed to conduct sentiment analysis. I made good use of the dictionary created by Mohammad et al., [24]. The dictionary has the following types of mood: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. In the current stage of this study, I decided to use the moods mentioned above after converting them to the basic two types of mood, positive and negative.

In this thesis, a naive algorithm for calculating the sentiment score of a statement (e.g., sentence) introduced by Grimmer et al., [25] is used. The algorithm could be expressed as the following equation.

$$T = \sum_{m=1}^{M} \frac{S_m Wim}{N_i} \tag{3.3}$$

Where $T([-1; 1])$ is the sentiment score of a target text, and if $T > 0$, the sentiment of the text will be positive else it will be negative. M, S_m, W_{im}, and N_i are respectively the vocabulary size of the dictionary, the sentiment score of the word came out in the text, and the number of all the words in the dictionary that came out in the text.

3.3 LINE: Large-scale Information Network Embedding

LINE is an embedding technique that converts a large scale information network into a low-dimensional vector space [26]. In other words, if a population expressed as a directed or undirected graph is given, the LINE returns another population where each node is presented as an unique numerical vector. I then adopt the LINE for converting a keyword network to a numerical vector population.

The main idea behind the LINE is as follows [26]. First of all, there are important concepts of *firs-order proximity* and *second-order proximity* in information network. The first-order proximity is defined as the proximity the local pairwise proximity between two vertices, and the second-order proximity between a pair of vertices (u, v) in a network is defined as the similarity between their neighborhood network structures [26]. The LINE is trained so as to preserve the first-order proximity and second-order proximity separately and then concatenate the embeddings trained by the two methods for each vertex.

The following expressions are the algorithm for embedding a graph by using the first-order proximity as an example of the mathematical definition for the procedure described above. The joint probability between vertex v_i and v_j is expressed as Equation 3.4.

$$p_1(v_i, v_j) = \frac{1}{1 + exp(-\vec{u}_i^T \cdot \vec{u}_j)} \tag{3.4}$$

Where $\vec{v_i} \in R^d$ is the low-dimensional vector representation of vertex v_i. Equation 3.4 defines a distribution $p(\cdot, \cdot)$ over the space $V \times V$ and its empirical probability can be defined as $\hat{p}(i, j) = \frac{w_{ij}}{W}$, where $W = \sum_{(i,j) \in E} w_{ij}$. To preserve the first-order proximity, a straight forward way is to minimize the objective function 3.5.

$$O_1 = d(\hat{p}_1(\cdot, \cdot), p_1(\cdot, \cdot)) \tag{3.5}$$

Where $d(\cdot, \cdot)$ is the distance between two distributions. Jian et al., chose to minimize the KL-divergence of two probability distributions. Replacing $d(\cdot, \cdot)$ with KL-divergence omitting some some constants, Equation 3.5 is written as:

$$O_1 = -\sum_{(i,j) \in E} w_{ij} log p_1(v_i, v_j) \tag{3.6}$$

Note that the first-order proximity is only applicable for undirected graphs, not for directed graphs. By finding the $\{\vec{u}_i\}_i = 1..|V|$ that minimize the objective in Equation 3.6, we can represent every vertex in the d-dimensional space [26].

3.4 Artificial Neural Networks

3.4.1 Feedforward Neural Networks

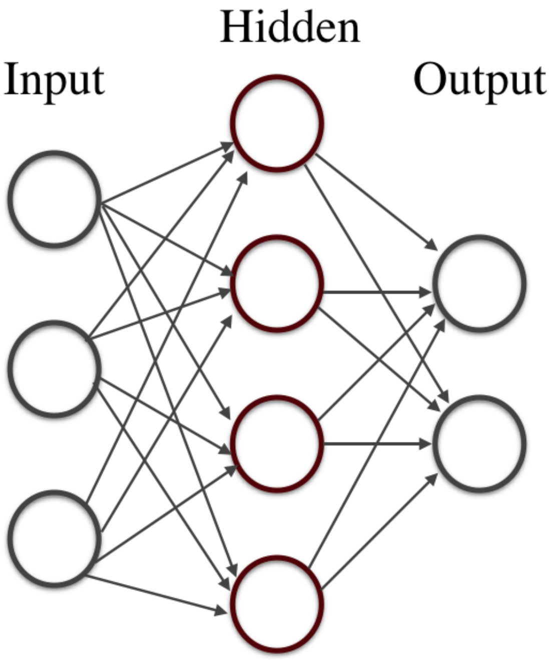


Figure 3.1 Simple feedforward neural network

As an example illustrated in Figure 3.1, feedforward neural networks, or multilayer perceptrons, are the quintessential deep learning models [27]. Feedforward networks consist of three types of neuron layers including input layer, hidden layers, and output layer. The goal of a feedforward network is approximate some function f. A feedforward network defines a mapping $y = f(x, \theta)$ and learns the value of the parameters θ that result in the best function approximation [27].

Each neuron in a layer has input and output gates, and forwards the signal entered through the input gate to each neuron in the following layer after performing a linear transformation ($y = wx + b$) on the signal. Where w is called weight, and b is called bias. On the other hand, a neuron in the following layer receives the nonlinear weighted sum of the output of each neuron in the preceding layer, which is expressed as Equation 3.7.

$$f(x) = K(\sum_i (w_i x_i + b_i)) \tag{3.7}$$

Where K is some non-linear function as hyperbolic tangent and commonly called activation function.

The learning process is simply to find the optimal value of θ (w_i and b_i) for the input and output signals. Also there are two basic concepts for finding the optimal values that have to mentioned. The first is cost function, or loss function, and the second is backpropagation. The cost function defines the difference between the output of a neural network and the expected output, and simply saying, decreasing the value of the cost function is the learning process. The backpropagation is the method for tuning the parameters based on the output signal of the neural network.

3.4.2 Convolutional Neural Networks

CNN (Convolutional Neural Network) [28], [29] is a powerful class of ANN (Artificial Neural Networks) and well-known in image recognition. It differs from the other ANN classes by the concepts of convolution and filtering. The advantages are autonomous feature selection, dimension reduction, and applicability to the input of various length. Also the main characteristic of CNN is grabbing the most essential feature of an image or a sentence. The operation principle of CNN is briefly described by using the CNN model for sentence classification depicted in Figure 3.2.

A convolution operation involves a filter w, or simply a low-dimensional weight, which is applied to a window of h words to produce a new feature c. A new feature c_i can be expressed as Equation 3.8 [28].

$$c_i = f(w \cdot x_{i:i+h-1} + b) \tag{3.8}$$

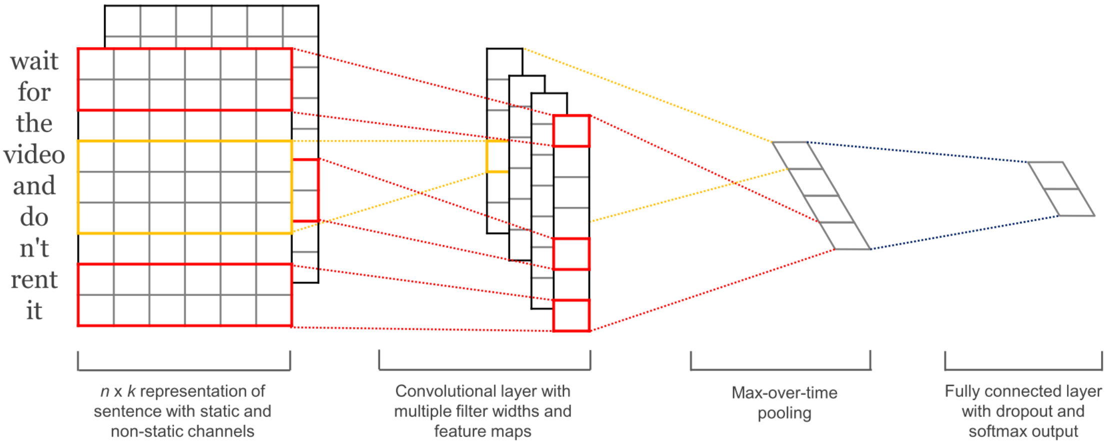

wait
for
the
video
and
do
n't
rent
it
n x *k* representation of sentence with static and non-static channels
Convolutional layer with multiple filter widths and feature maps
Max-over-time pooling
Fully connected layer with dropout and softmax output

Figure 3.2 CNN model for sentence classification [28]

Where x is a word vector, $x_{i:i+h-1}$ is a window of words, b is a bias, and f is an activation function. This filter is applied to each possible window of words in the sentence to produce a feature map $c = [c_1, c_2, ..., c_{n-h+1}]$. After filtering, a max-overtime pooling operation [30] is applied over the feature map and take the maximum value $\hat{c} = max\{c\}$ as the feature corresponding to this particular filter. The idea is to capture the most important feature with the highest value for each feature map. This pooling scheme naturally deals with variable sentence lengths [28].

3.5 Word2vec

Word2vec is a word embedding technique introduced by Mikolov Tomas et al., [31]. Word2vec is known as a way to convert words to numerical vectors keeping the semantic relations of the words.

For instance, *king* - *woman* would be *queen*. The concept of word2vec is training neural networks so as output semantically close words for a input word by the means of applying CBOW (Continues Bag of Words) which is a technique to train an ANN so as to predict an unknown keyword from the surrounding words, Skip-gram which is a technique for training an ANN so as to predict the highly possible neighbors from a word, etc [31].

However, training the neural networks on big vocabulary is a challenging task. So Mikolov Tomas et al., divided the dataset into correlated groups by making a good use of Hamming code in order to update the parameters of overlapping groups with a single step.

Chapter 4

IDEAS FOR DATASET CREATION AND PREDICTION

In this chapter, I introduce the novel ideas for predicting cyber attack from the social insights and the methods to create dataset.

4.1 Ideas for Cyber Attack Prediction

In this section, I introduce the ideas for predicting cyber attack from social feeds. I hypothesize two things here, and the first is that the cyber attack motivation targeting on a certain objective can be predicted by analyzing the social feeds for the target objective. The second is that there is an opportunity to predict the highly possible cyber attack type against a certain objective by analyzing some domain specific social feeds.

4.1.1 Prediction of Cyber Attack Motivation

As Sharma et al., [9] do, I assume that a news article is a reliable source to know the social insight against a certain objective. Besides news articles, as Bollen et al., [14] do, I hope that twitter feeds also can be a good source for knowing the social insight. I hypothesize that some changes appear in the social insight against a cyber attack victim just before an attack. The idea for prediction procedure is as follows:

1. Collect the archive of the cyber incidents occurred in the past.

2. Collect the social feeds (news articles and tweets) related to each victim in the archive.

3. Detect some pattern by comparing the social feeds against a victim, which released just before an attack with the feeds released in other period [1].

4. Train a machine learning model on the pattern detected in the previous step.

5. Input daily social feeds to the trained model and check whether the pattern appears in the daily feeds.

In the step 5, if the pattern is detected in a daily feed, the prediction result will be positive. That means there is a motivation of cyber attack against the target objective.

4.1.2 Prediction of Cyber Attack Opportunity

As Holm et al., [15] did, I assume that not only security professionals and system administrators, but also cyber attackers utilize security vulnerability feeds for their activities. I define the security vulnerability feeds as some publicly known information security vulnerabilities and exposures such as CVE [32]. The security vulnerability feeds are a type of domain specific public feeds and include more technical aspects than the usual social data such as news articles. Therefore, I hypothesize that the security vulnerability feeds can be a source for predicting cyber attack opportunity, or the highly possible attack type, against a certain objective. The concept for prediction procedure is as follows:

1. Collect the archive of the cyber incidents performed by using the security vulnerability feeds [2].

2. Collect the vulnerability feeds exploited in the incidents.

3. Detect some patterns from the vulnerability feeds.

4. Train a machine learning model on the pattern detected in the previous step.

[1] Briefly: Dataset creation.

[2] Exploits of security vulnerability feeds.

5. Input daily vulnerability feeds to the trained model and check which pattern appears in the daily feeds.

In the step 5, the pattern detected in a daily feed will be the prediction result.

4.2 Ideas for Dataset Creation

Generally, dataset creation process includes data preprocessing, feature selection, and labeling. For the feature selection, each experiment conducted in the further steps applies different approaches, and the detailed description of the approaches is given in Chapter 5. The data preprocessing and idea for dataset labeling are described in this section.

In this study, the data preprocessing means basic text cleansing process including trimming, lowercasing, and exception of unknown symbols and characters. Language filtering is also done in this step, and only the text written in English is passed to the further steps.

For the prediction of cyber attack motivation, the basic idea for dataset labeling is dividing the dataset into two sets (the set directly correlated to cyber attack and the set not correlated to cyber attack). I assume that some changes (e.g., increase in the number of feeds) appear in the social feeds against a certain objective just before a cyber attack. So the idea for dataset labeling is annotating the social feeds published within the period where the changes occurred as directly correlated with the attack. Thus, those feeds are labeled as positive for cyber attack motivation. On the other hand, the feeds published in the period where any significant changes are noticed are labeled as negative for cyber attack motivation.

Also, another idea for labeling the social feeds as twitter feeds, the factor for dividing dataset could be the average sentiment score of the tweets posted before an attack. The sentiment score is a metric ($[-1.0; 1.0]$) used to express the mood (positive: 1.0, negative: -1.0) of a textual statement.

The vulnerability feeds have some exploitability and impact metrics, and those metrics are used in the further experiments as the feature for prediction. The dataset was labeled by referencing Exploit Database [20] which is a CVE compliant archive of public exploits and corresponding vulnerable software, developed for use by penetration testers and vulnerability researchers.

In this thesis, the following labels which are most frequently appeared in the Exploit Database are used. Here following a label are the cyber attack types for to the label.

1. webapps: SQLi, Malicious code injection

2. dos: DoS, DDoS

3. local: Targeted attack, Malware, Account hijacking

4. remote: Arbitrary code execution

Chapter 5

DATA COLLECTION METHODOLOGY

In this chapter, the social data used in the further experiment and the collection methodology are introduced. There are totally four sections, and the first three sections contain the detailed description of the social data and the collection methodology. The latest section summarizes this chapter.

5.1 Cyber Attack History and News Articles

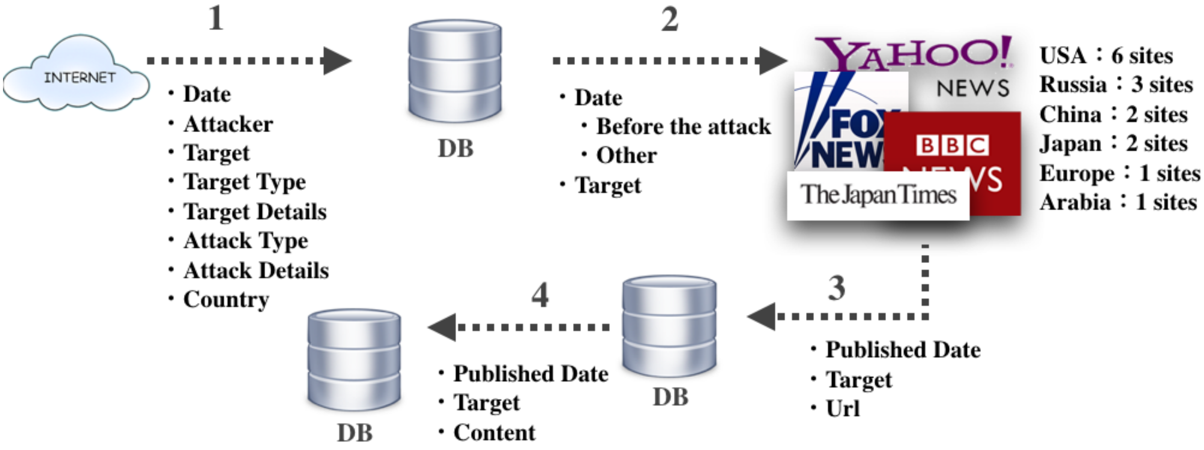


Figure 5.1 Current system for collecting cyber attack history and news articles

In Figure 5.1, the operation flow for collecting the cyber incident archive and cyber attack victim related news articles is shown, and it is roughly divided into four steps as follows. Collecting cyber attack history, searching the URLs of the related news articles for each cyber attack victim, storing the URLs, and parsing the text content of the URLs.

5.1.1 Cyber Attack History

Table 5.1 Schema of the table for cyber attack history

Column	Id	Date	Attacker	Target	TC	TD	AD	AT	AP	Country
Example	798	2015-08-10	Del..	www.av..	News	Avionews	Del..	SQLi	CC	Italy

TC: Target Category, TD: Target Details, AD: Attack Details, AT: Attack Type, AP: Attack Purpose, CC: Cyber Crime

First of all, an archive of cyber incidents has to be created to proceed to the further steps. Because, without the archive, the correlated news articles for a cyber incident cannot be collected, and the dataset creation and validation test cannot be performed. The archive of cyber incidents is collected from some cyber attack news sites including [33], [34], [35] and [36], and stored in the table with the schema depicted in Table 5.1.

After that, several factors as the date for each incident, attacker, target, target category (government agencies, educational institutions, individuals, etc.), detailed description of the target, attack type (DDoS, SQL injection, etc.), attack purpose (cyber crime, hachtivism, etc.), other details of the attack, country where the incident was occurred are extracted. Though the history of cyber incidents are collected automatically, the extraction of these factors is done manually.

5.1.2 News Articles

First, the URLs of the related news articles for each cyber attack victim are acquired. To collect the URLs, Google Custom Search API [37], a query API provided by Google, is used. The search terms are the official name of the victims, and the search sources are fixed to 15 official English language news sites. The news medias consist of six American medias (Yahoo News, The Guardian, New York Times, Fox News, Washington Post, and NBC News), three Russian medias

Table 5.2 Schema of the table for the news articles

Column name	Unique id	Incident id	Source	URL	Content	Published date
Example	220	14	bbc	www.bbc.com/news/..	All nine bank of ..	2015-01-21

(PravdaReport, RT, and The Moscow Times), two Chinese medias (Chinadaily and Shanghai Daily), two Japanese medias (NHK World and Japan Times), one European media (BBC News), and Al Jazeera from the Arabic world. Also the articles published within two months before the day of a cyber attack are the target of this step.

After collecting the URLs, the text content of each URL is fetched and preprocessed. I implemented an especial HTML parser for each news media site by using a python library, Beautiful Soup 4.4.0 [38], in order to fetch only the body of the main article. After fetching the text contents, each of them is gone through a simple preprocessing step including trimming spaces and newlines, excluding unknown symbols, converting to lowercase, etc. The text contents are stored in the table with the schema depicted in Table 5.2.

5.2 Security Vulnerability Feeds

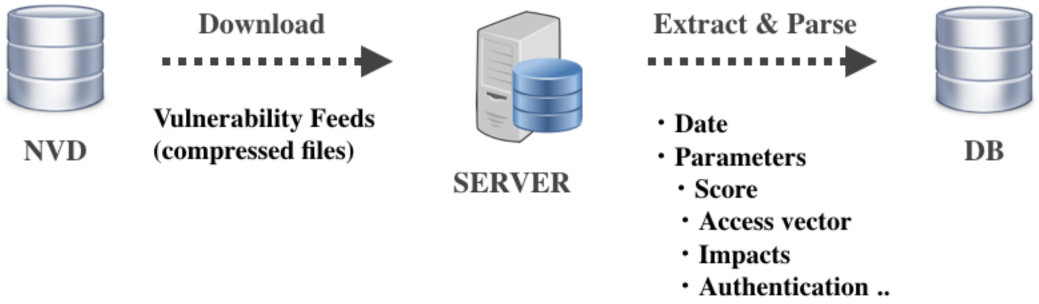


Figure 5.2 Working flow for collecting security vulnerability feeds

For security vulnerability feeds, as depicted in Figure 5.2, the system accesses to National Vulnerability Database (NVD) [39] and downloads the older and newly uploaded feeds. NVD

Table 5.3 Schema of the table for vulnerability feeds

Column	Id	SL	PD	UD	CVSS-S	AV	AC	AUTH	CI	II	AI	Source	Details
Example	CVE-..	oracle..	2015..	2015..	3.5	NET	MED	SIN	NONE	PAR	NONE	CON	..

SL: Software List, PD: Published Date, UD: Updated Date, CVSS-S: CVSS Score, AV: Access Vector, AC: Access Complexity, AUTH: Authentication, CI: Confidentiality Impact, II: Integrity Impact, AI: Availability Impact, NET: Network, MED: Medium, SIN: Single, PAR: Partial, CON: Confirm

is the U.S. government repository of standards based vulnerability management data represented using the Security Content Automation Protocol (SCAP) [40]. NVD includes databases of security checklists, security related software flaws, misconfigurations, product names, and impact metrics.

Each security vulnerability feed provided by NVD includes a number of fields. However, only the useful fields for cyber attack are stored in the database table with the schema depicted in Table 5.3. The chosen fields are published/updated date, CVSS score [41] which is a security severity score estimated by FIRST [42], access vector (Remote, Network, and Local), access complexity (High, Medium, and Low), authentication (Multiple, Single, and None), confidentiality impact (Complete, Partial, and None), integrity impact (Complete, Partial, and None), availability impact (Complete, Partial, and None), source, and details in text.

5.3 Twitter Feeds

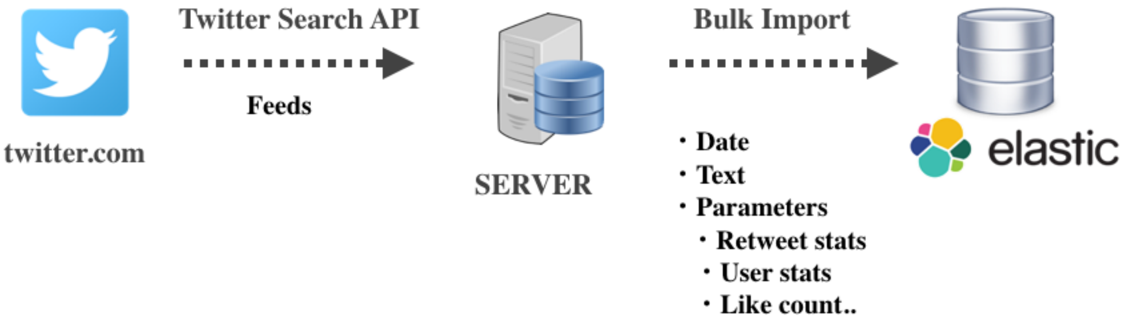


Figure 5.3 Working flow for collecting twitter feeds

As a source for knowing the public insight, I use twitter feeds in the further experiment. In this section, two kinds of twitter feeds and the collection methodology are introduced. As depicted in Figure 5.3, all of the collected twitter feeds are stored in an Elasticsearch server [43], which is one of the emerging open source search engine software that realizes fast yet robust search and aggregation on a huge amount of textual data. Also, the data format used in Elasticsearch is *json* as same as that of the twitter feeds. Because of these facts, I deployed the Elasticsearch server to store the twitter feeds.

5.3.1 Twitter Feeds of Hacktivists

I came up with an idea of analyzing the twitter feeds of hacktivist groups and began to collect the twitter feeds in March of 2016. The twitter.com provides a free API [44] for searching a certain user's tweets posted in the last week. I chose 90 twitter accounts based on the number of the followers ($>$ 1000), account description, and the content of the posts, which were divided into four well-known hacktivist clans including Anonymous, Zyklon, LizardSquad, and Dr.SHA6H. The twitter feeds of the 90 accounts have being collected in real time since March 3 of 2016 and stored in the Elasticsearch server shown in Figure 5.3.

5.3.2 Twitter Feeds from Japan

The free search API provided by the twitter.com allows us to get the tweets posted at oldest a week ago, so getting older tweets is difficult and costs much. However, thanks to the provision by professor Eiji ARAMAKI [45] from NAIST (Nara Institute of Science and Technology), a certain amount of twitter feeds posted in Japan in the past was able to be collected. The rough range of the published date of the twitter feeds is from October 2008 to April 2010, and the written language is mainly Japanese. Also, each twitter feed consists of three fields, the textual content of a tweet, user who posted the tweet, and date of post. As I did with the tweets of the hacktivist groups, the tweets provided by professor Eiji ARAMAKI were stored in the Elasticsearch server.

5.4 Summary

For the experiment, I collected four kinds of social data including the cyber incident archive, the news articles related to cyber attack victims, the tweets posted by hacktivist groups, and the tweets posted in Japan from October of 2008 to April of 2010. Also, the prototype systems for collecting, storing, and processing all of the collected social data were implemented. The total amount and other details of the collected social data are shown in Table 5.4.

Table 5.4 Summary of collected social data

Name	Source	Amount	Collecting Method	Date Range
CIA (World)	Internet	1406 cases	GCS API	2015/01/01 - 2016/10/30
CIA (Japan)	Internet	145 cases	Manually	2008/10/30 - 2010/04/30 -
NA	15 medias	23658	GCS API	2014/12/01 - 2016/04/30
SVF	NVD	77354	Manually	1988/10/01 - 2016/10/03
Tweets (HGs)	twitter.com	291596	Twitter API	2016/03/06 -
Tweets (Japan)	Prof. ARAMAKI	158246380		2008/10/25 - 2010/4/30

CIA: Cyber Incident Archive, GCS API: Google Custom Search API, NA: News Article, SVF: Security Vulnerability Feed

Chapter 6

EXPERIMENT

In this chapter, the detailed description of the experiments conducted to verify the ideas introduced in Chapter 4. At first, the result of the preliminary experiments and the conclusions drawn from the result are presented. After that, the result of the subsequent experiments is presented.

6.1 Preliminary Experiment

In this section, the detailed descriptions of three preliminary experiments, two for predicting cyber attack motivation and one for predicting cyber attack opportunity, are presented. During the experiments for predicting cyber attack motivation, the news articles related to the victims of cyber attack are classified by SVM and CNN. In the experiments for predicting cyber attack opportunity, the security vulnerability feeds are classified by SVM.

6.1.1 Prediction of Cyber Attack Motivation using News Articles 1: SVM

At first, I attempted to predict both of the probability of cyber attack occurrence for a certain target and the highly probable type of attack simultaneously by using SVM. A Support Vector Machine (SVM) [46], one of the widely used supervised machine learning algorithms, is a discriminative classifier formally defined by a separating hyperplane. That means a labeled training data is given, and the algorithm outputs an optimal hyperplane which categorizes new examples. When tuning a SVM classifier, we need to choose the best kernel function and find out an opti-

Table 6.1 Prediction result of the experiment using SVM

	TA	AH	DDoS	SQLi
Vector A	ACC: 62.4% PRE: 0.49 REC: 0.63	ACC: 60.7% PRE: 0.43 REC: 0.63	ACC: 70.3% PRE: 0.41 REC: 0.77	ACC: 59.1% PRE: 0.52 REC: 0.59
Vector B	ACC: 56.0% PRE: 0.59 REC: 0.56	ACC: 64.0% PRE: 0.48 REC: 0.65	ACC: 62.0% PRE: 0.26 REC: 0.80	ACC: 56.8% PRE: 0.42 REC: 0.58

TA: Targeted attack, AH: Account hijacking, ACC: Accuracy, PRE: Precision, REC: Recall

mal value of the parameters as γ, C and *penalty* for the task that we are trying to solve. In this experiment, LIBSVM [47], an open source SVM library was utilized.

I estimated the sentiment score for 6915 news articles related to the victims of the cyber attacks carried out in the first half of 2015 by means of the algorithm introduced by Grimmer et al., [25]. Also, I made good use of the dictionary created by Mohammad et al., [24]. The kernel function of the SVM classifier was Radial Basis Function expressed as $K = exp(-\gamma|u - v|^2)$.

For dataset, all of the articles were converted to 1000 dimensional vectors. I extract the 1000 keywords mostly appeared in all of the 6915 articles, and based on those keywords, two types of feature vectors called *vector A* and *B* are created. The vector A can be expressed as $A = [W_1, W_2, .., W_{1000}]$ where W_i is the number of occurrence of the i^{th} word in the core 1000 keywords. On the other hand, the vector B ($[V_1 \cdot S_1, V_2 \cdot S_2, .., V_{1000} \cdot S_{1000}]$) was created based on the positive and negative 500 words mostly appeared in all of the 6915 articles. Where V_j was the number of occurrence of the j^{th} word in the positive and negative 500 keywords mentioned above. S_i is the sentiment score for V_i.

To label the dataset, I aggregated all the 6915 articles including 6514 positive, 329 negative, and 72 neutral articles. As a result, I found that the weekly average rate of increase in the number of news articles were respectively 42% for total, 35% for positive, and 33% for negative articles. From the result, he assumed that the articles published within the period from a week before the attack was carried out to the day of attack were directly correlated with the attack, so the vectors for those articles were labeled as positive and the others as negative.

The result of the experiment is depicted in Table 6.1. I implemented a SVM classifier for each attack type and trained each classifier with the 80% of the articles related to the attack type. The remaining 20% was used for test, and the average *accuracy* was defined as $\frac{m}{n}$ (m: number of correct outputs, n: size of test-set).

6.1.2 Prediction of Cyber Attack Motivation using News Articles 2: CNN

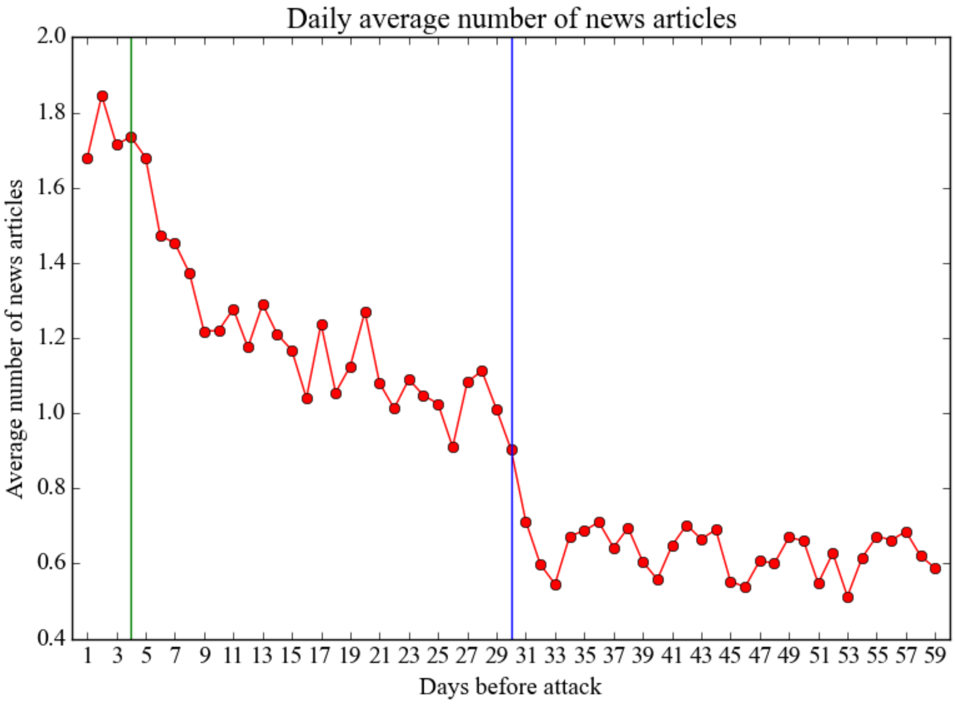


Figure 6.1 Daily average number of news articles published before a cyber attack

Subsequently, I attempted to improve the accuracy of cyber attack motivation for a certain target by applying CNN. He implemented a CNN model which was basically identical to Kim Yoon's model built for sentence classification [28]. According to Kim Yoon, this model brought a good accuracy for several datasets consisted of long sentences. The model has three layers of input layer, one convolutional layer with three max-over-time filters, and one output layer activated by ReLU (Rectified Linear Unit). The window sizes of the filters are respectively three, four and five.

As a dataset, I used 19600 news articles related to the victims of the cyber attacks carried out in 2015. For this time, feature selection was not an issue, but converting the articles into numerical vectors was the issue. As Kim Yoon did, I utilized pre-trained word2vec [31]. The 300 dimensional word2vec [48] trained on 100 billion words from Google news was used in two ways. The one is setting the word2vec as static and update the other parameters, and the second is updating word2vec as well.

To label the dataset, I aggregated all of the 19600 articles. The result is depicted in Figure 6.1, where the x-axis is the days before an attack and the y-axis is the daily average number of news articles on cyber attack victims. From the result, it was found that the daily average number of the articles on cyber attack victims was increase in high rate within the four days before the attack. That is, I assumed that the articles published within this period (to the left from the green line) were directly correlated with the attack, and labeled the vectors for those articles as positive and the articles published in stable period (to the right of the green line) as negative. The articles were divided into sentences, and the CNN model, a sentence classifier, was trained on the 80% of the dataset and tested on the remaining 20%. In the experiment, I calculated the accuracy of prediction in the same way adopted in the previous experiment where a SVM classifier was applied. The result is depicted in Table 6.2.

Table 6.2 Prediction result of the experiment using CNN

Type of word2vec	Result
static word2vec	ACC: 69.8% PRE: 0.42 REC: 0.71
non-static word2vec	ACC: 73.3% PRE: 0.49 REC: 0.73

ACC: Accuracy, PRE: Precision, REC: Recall

6.1.3 Prediction of Cyber Attack Motivation using Twitter Feeds 1: Tweets from Japan

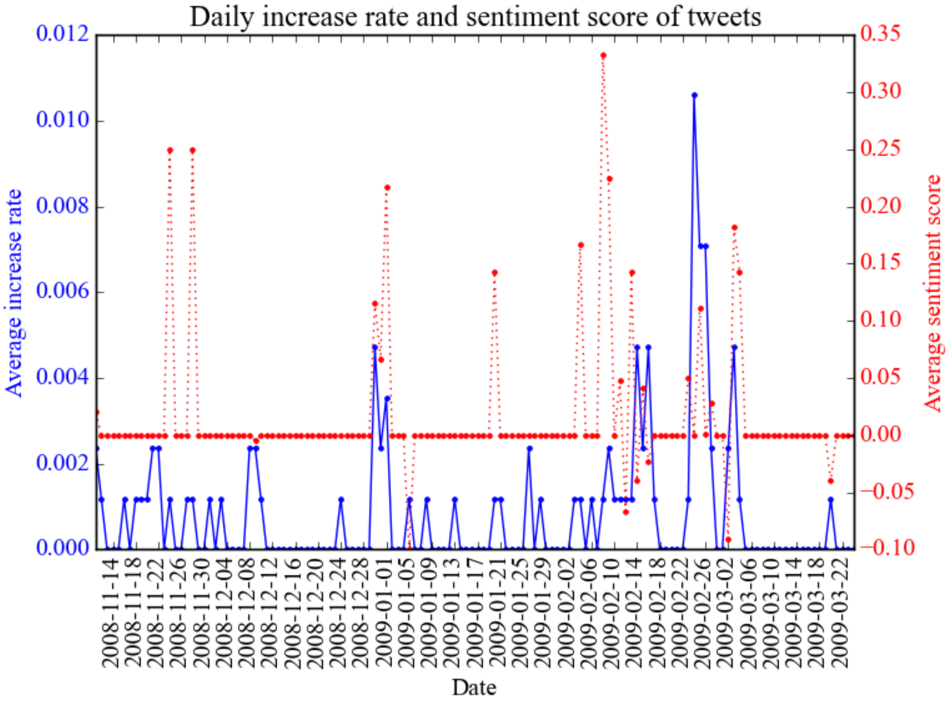


Figure 6.2 Daily increase rate in the number and sentiment score of the tweets against GENO (First)

As mentioned in Chapter 4, the twitter feeds posted in Japan within October 2008 to April 2010 was provided by professor Eiji ARAMAK [45]. Also, I collected the cyber incidents occurred in Japan within October 2008 to April 2010. In the experiment, 34 cases out of the incidents were used, and the social insight against the victims posted before and after an attack was analyzed.

First, the increase rate in the number of the twitter feeds (=daily number of tweets / the daily maximum number of tweets) was calculated. Second, the daily sentiment score of the tweets was calculated. To calculate the sentiment score, I used the sentiment dictionary created by Takamura

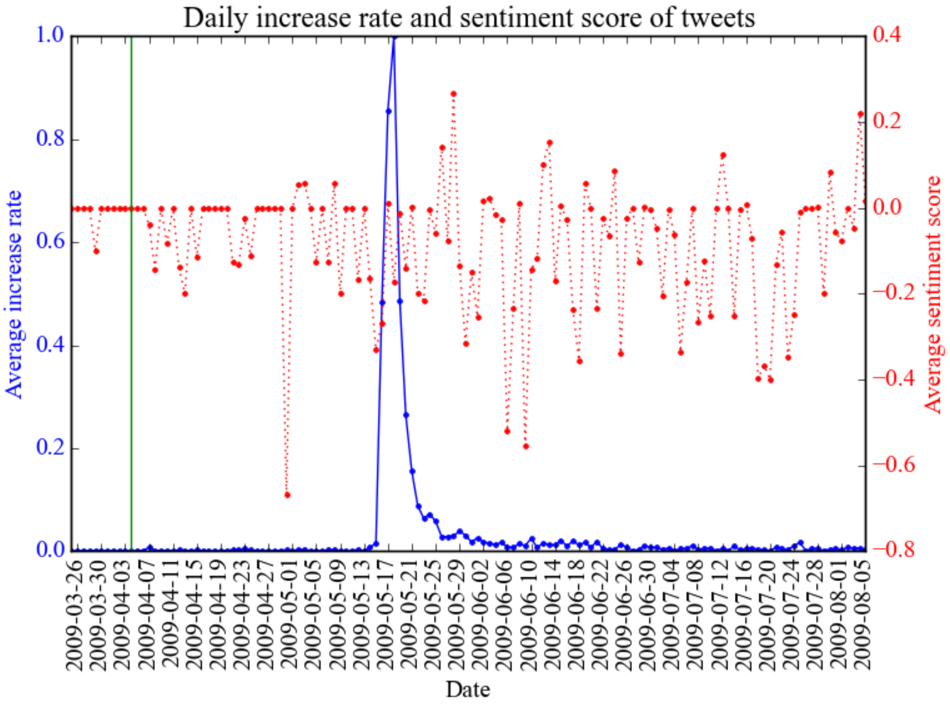


Figure 6.3 Daily increase rate in the number and sentiment score of the tweets against GENO (Second)

et al., [49]. In Figure 6.2 and 6.3, the result for GENO which is a well-known e-commerce site is depicted as an example. Where the green line indicates the date of a cyber attack.

From the result above, there were no significant changes in the twitter feeds before an attack. Although it was not directly related to prediction of cyber attack, it was found that the public statements on twitter against the victim of a cyber attack became negative, and the frequency of statements increased after the attack compared to before the attack. The results for the other 33 victims were similar to that of GENO.

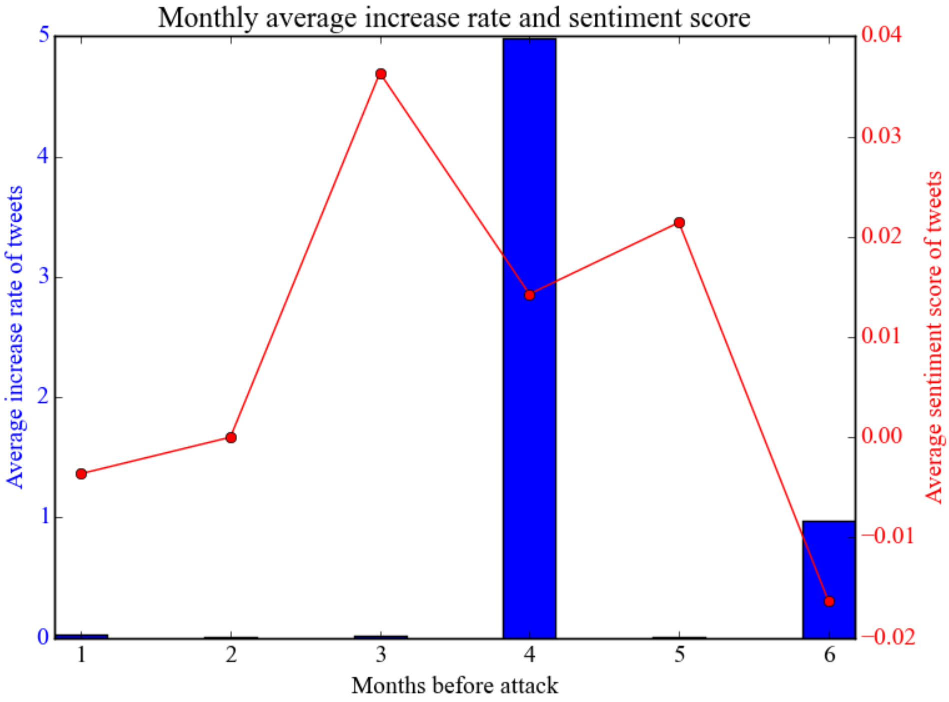


Figure 6.4 Monthly average increase rate in the number and sentiment score of the tweets by hacktivists (Partial)

6.1.4 Prediction of Cyber Attack Motivation using Twitter Feeds 2: Tweets of Hacktivists

As introduced in Chapter 4, since March 2016, I began to collect the tweets of 90 twitter accounts which belong to four well-known hacktivist clans including Anonymous, Zyklon, LizardSquad, and Dr.SHA6H. In the experiment, 35 cases out of the incidents occurred since March 2016 were used, and the twitter posts of the hacktivists for the victims of the incidents.

First, the monthly average increase rate in the number of the twitter feeds (=the monthly average number of tweets / the overall average number of tweets) was calculated. The reason for aggregating monthly average instead of daily average is that the daily number of tweets were fewer than the that of the previous experiment. Second, the monthly average sentiment score

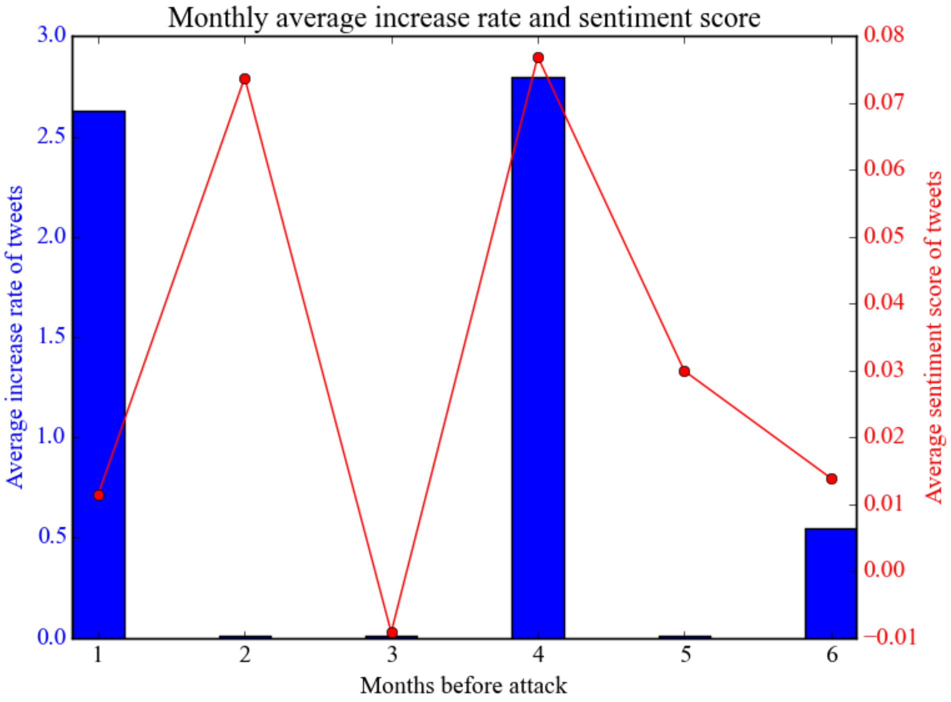


Figure 6.5 Monthly average increase rate in the number and sentiment score of the tweets by hacktivists (Overall)

of the tweets was calculated. In Figure 6.4 and 6.5, the results of the experiment are depicted. Figure 6.4 is the result for some victims about whom the hacktivists frequently tweeted. On the other hand, Figure 6.5 is the result for all the 35 victims. Where x-axis presents the months before an attack. The blue bar and the red line are respectively the monthly average number and sentiment score of tweets.

For the result depicted in Figure 6.4, there were no significant changes in the twitter feeds before an attack. However, as depicted in Figure 6.5, the number of tweets increased in the month just before the attack.

6.1.5 Prediction of Cyber Attack Opportunity using Security Vulnerability Analysis

Table 6.3 Features and metrics used in the experiment

Metric	OL	CVSS score	AV	AC	AUTH	CI	II	AI
Example	OS	7.5	NET	LOW	NONE	PART	PART	PART
	APP	2.3	LOCAL	LOW	MULTI	NONE	NONE	NONE
	OS	6.6	AN	MED	SIN	PART	PART	PART
	HW	9.2	NET	HIGH	NONE	COMP	COMP	COMP

OL: Operation Level, AV: Access Vector, AC: Access Complexity, AUTH: Authentication, CI: Confidentiality Impact, II: Integrity Impact, AI: Availability Impact, APP: Application, AN: Adjacent Network, NET: Network, PART: Partial, MULT: Multiple, COMP: Complete, SIN: Single, OS: Operating System, HW: Hardware

To predict the highly possible attack type, I used security vulnerability feeds downloaded from NVD. The vulnerability feeds have some exploitability and impact metrics, and those metrics were used as the feature. I adopted C-SVM (C-Support Vector Machine) as a multi-class classifier where C ($[0.0, 1.0]$) is the penalty parameter. In the experiment, various types of kernel functions as *rbf, polynomial, linear,* and *sigmoid* were adopted.

This time, the metrics depicted in Table 6.3 were used as the feature vector. Except the CVSS score [41] which is a severity score estimated by FIRST [42], I quantified all the metrics according to the value. The metrics except CVSS score were divided into three levels by referencing [41]. In the experiment, 2055 security vulnerability feeds registered within 2006/01/01-2016/07/19 were used as dataset. The dataset was labeled by referencing Exploit Database [20] which is a CVE compliant archive of public exploits and corresponding vulnerable software, developed for use by penetration testers and vulnerability researchers. The most common four labels of *webapps*, *dos*, *local*, and *remote* were used in the experiment.

The C-SVM classifier was trained on the 80% of the dataset and tested it with the remaining

Table 6.4 Result of highly possible attack type prediction

Kernel	Linear	Sigmoid	RBF	Polynomial
Accuracy	75.2%	76.8%	83.5%	83.8%

20% of the dataset. The accuracy was calculated in the same way adopted in the preceding experiments. The result is depicted in Table 6.4.

6.2 Additional Experiment

In this section, the detailed description of some additional experiments is given. From the results of the preliminary experiment, the following conclusions were made, and some additional experiments according to those conclusions were conducted.

First, from the results of the experiment for predicting cyber attack motivation using news articles, I concluded that using the mostly appeared keywords as a feature was not bad idea, and the CNN model was not always necessary. Because, the difference between the accuracy of the prediction using SVM and CNN was not significantly big even utilizing word2vec. I decided to use keyword based features and feedforward neural networks in the additional experiment. Also, in the preliminary experiment, all of the news articles were used together as a dataset, but in the further experiment, the news articles would be used separately by victim.

Second, from the results of twitter feeds analysis, I concluded that twitter feeds could not be a reliable source as news articles in the current stage. Because, any significant changes appeared in the tweets against a cyber attack victim, which posted before the attack. Although there was a change in the twitter statements of some hacktivists a month before an attack, the amount was insufficient to use as a feature.

Third, from the results of cyber attack opportunity prediction, I concluded that the security vulnerability feeds could be used for predicting the highly possible cyber attack. I considered trying more features with bigger volume as the textual description of a vulnerability feed to improve the accuracy of prediction. Also I supposed that applying feedforward neural networks instead of SVM could also be a improvement.

6.2.1 Additional Experiment on Cyber Attack Motivation Prediction

In the additional experiment, the news articles related to the victims of cyber attack were used separately by victim. That is, the motivation of cyber attack against an objective was predicted by using the model trained on only the news articles related to the objective. The experiment was conducted with multiple patterns, various number of hidden layers, of feedforward neural networks. The patterns have various number of hidden layers (1, 2, 3, and 4), a common activation function Relu ($f(x) = max(0, x)$), a common optimizer of Adam optimizer [50] which is one of the emerging optimizer functions in artificial neural networks, and a common loss function of mean squared error. Also the number of neurons in a hidden layer was constantly 128.

Dataset creation was basically same as that of the preliminary experiment, but the method for dataset labeling was renewed. In the preliminary experiment, I aggregated all the news articles and determined the threshold dates for labeling dataset. For this time, I applied an autonomous method for determining the threshold dates. At the beginning, all of the related news articles for each victim were converted into numerical vector based on the mostly appeared 1000 keywords in those articles. The element V_i of a feature vector is the frequency of appearance for the i^{th} term in the core 1000 keywords. After that, the numerical vectors were divided into two clusters by using K-Means, and for each cluster, the average date difference between the attack date and the published date of the article D was calculated. If let D_1 be the largest D, and let D_0 be the smallest D, the thresholds for positive and negative labels would be respectively D_0 and D_1. That means, an article j was labeled as positive if the date difference D_j was smaller than D_0 and labeled as negative if the D was greater than D_1.

For feature selection, I used five types of features (*ON, OFF, ON-OFF, OFF-ON,* and *JACCARD*). Where ON and OFF are respectively the 1000 dimensional feature vectors based on the mostly appeared 1000 keywords extracted from the articles labeled as positive and those labeled as negative. On the other hand, ON-OFF and OFF-ON are respectively the 1000 dimensional feature vectors based on the 1000 keywords extracted from the words in the positive articles - words in the negative articles and words in the negative articles - words in the positive articles.

The JACCARD feature for an article was created by averaging the 32 dimensional LINE vector of all the words in the article. To express each words in an article as a LINE vector, I built a keyword network for each victim and input that to the LINE engine introduced in Chapter 3. The LINE engine is supposed to take a weighted graph as an input, so I use the Jaccard Similarity

Table 6.5 Result of the additional experiment for cyber attack motivation prediction

L	ON	OFF	ON-OFF	OFF-ON	JACCARD
4	ACC: 66.9% PRE: 0.29 REC: 0.37	ACC: 70.9% PRE: 0.22 REC: 0.38	ACC: 92.3% PRE: 0.71 REC: 0.99	ACC: 84.4% PRE: 0.97 REC: 0.67	ACC: 71.9% PRE: 0.33 REC: 0.46
3	ACC: 69.7% PRE: 0.29 REC: 0.37	ACC: 70.0% PRE: 0.22 REC: 0.38	ACC: 79.4% PRE: 0.71 REC: 0.99	ACC: 85.3% PRE: 0.97 REC: 0.67	ACC: 74.5% PRE: 0.39 REC: 0.62
2	ACC: 74.7% PRE: 0.45 REC: 0.46	ACC: 70.2% PRE: 0.23 REC: 0.22	ACC: 82.8% PRE: 0.38 REC: 0.75	ACC: 91.5% PRE: 1.00 REC: 0.79	ACC: 74.0% PRE: 0.31 REC: 0.44
1	ACC: 70.4% PRE: 0.11 REC: 0.24	ACC: 77.2% PRE: 0.51 REC: 0.65	ACC: 82.1% PRE: 0.48 REC: 0.74	ACC: 79.4% PRE: 1.00 REC: 0.63	ACC: 78.1% PRE: 0.60 REC: 0.66

L: Number of hidden layers, ACC: Accuracy, PRE: Precision, REC: Recall

(J) between every pair of keywords as the weight. The Jaccard Similarity between the keywords A and B is expressed as Equation 6.1.

$$J(A,B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{6.1}$$

Here $|A|$, $|B|$, and $|A| + |B| - |A \cap B|$ are respectively the number of articles where A appeared, the number of articles where B appeared, and the number of articles where A and B appeared simultaneously.

In the experiment, I focused on five incidents and the average prediction accuracy is shown in Table 6.5. The cases are defacement of Uber's home page (2015-02-27), account hijacking on Obama (2015-02-10), DDoS on France government offices (2015-01-15), DDoS on ISIS web services (2015-02-09), and DDoS on Indiana's administration offices (2015-03-28). The average number of the related news articles for each case was 180. The 80% and 20% of the news articles were respectively used for training and testing.

6.2.2 Additional Experiment on Cyber Attack Opportunity Prediction

The experiment was conducted in the same way as the preliminary experiment. The only changes were using various types of features and feedforward neural networks instead of SVM. The neural network models were exactly same as that of the additional experiment of predicting cyber attack motivation.

The method for dataset creation was basically same as that of the preliminary experiment. However, two new features were added to the existing feature. The first is called *JACCARD* and extracted from the textual description of a vulnerability feed. As done in the previous experiment, the 32 dimensional LINE vector of each word appeared in all of the vulnerability feeds was calculated. The weighted graph given to the LINE engine was the Jaccard Similarity based keyword network. The Jaccard Similarity between keyword A and B was calculated by Equation 6.1. Where $|A|$, $|B|$, and $|A| + |B| - |A \cap B|$ are respectively the number of vulnerability feeds where A appeared, the number of feeds where B appeared, and the number of feeds where A and B appeared simultaneously. The LINE vector for a vulnerability feed was determined as the vector average of the LINE vector for all the words in the textual description of the vulnerability feed.

The second is called *KEYWORD* and extracted from the textual description of a vulnerability feed as well. The core 1000 keywords which mostly appeared in the textual description of all the vulnerability feeds was chosen. The i^{th} element of the KEYWORD vector for a vulnerability feed is the occurrence frequency of the i^{th} word in the core 1000 keywords. The third is called *BASIC8* which is the same feature used in the preliminary experiment.

The result of the experiment is shown in Table 6.6. The experiment was identical to multi-class classification problem, and the classes where class1: webapps, class2: dos, class3: local, and class3: remote. Precision and recall are the methods mostly used for evaluating binary classification. In this experiment, I calculated the precision and recall for each class by dividing the prediction result into two new classes of the target class and others.

Table 6.6 Result of the additional experiment for cyber attack opportunity prediction

L	BASIC8	KEYWORD	JACCARD
4	ACC: 82.2%	ACC: 85.9%	ACC: 87.6%
	PRE1: 0.97	PRE1: 0.97	PRE1: 0.99
	REC1: 0.89	REC1: 0.93	REC1: 0.95
	PRE2: 0.48	PRE2: 0.73	PRE2: 0.64
	REC2: 0.79	REC2: 0.73	REC2: 0.78
	PRE3: 0.60	PRE3: 0.64	PRE3: 0.72
	REC3: 0.42	REC3: 0.75	REC3: 0.68
	PRE4: 0.37	PRE4: 0.52	PRE4: 0.55
	REC4: 0.63	REC4: 0.57	REC4: 0.62
3	ACC: 79.9%	ACC: 88.8%	ACC: 85.4%
	PRE1: 0.97	PRE1: 0.96	PRE1: 0.95
	REC1: 0.88	REC1: 0.97	REC1: 0.93
	PRE2: 0.35	PRE2: 0.76	PRE2: 0.61
	REC2: 0.77	REC2: 0.70	REC2: 0.74
	PRE3: 0.48	PRE3: 0.72	PRE3: 0.87
	REC3: 0.63	REC3: 0.76	REC3: 0.70
	PRE4: 0.45	PRE4: 0.55	PRE4: 0.54
	REC4: 0.41	REC4: 0.56	REC4: 0.59
2	ACC: 82.7%	ACC: 87.6%	ACC: 87.3%
	PRE1: 0.96	PRE1: 0.96	PRE1: 0.96
	REC1: 0.91	REC1: 0.94	REC1: 0.94
	PRE2: 0.45	PRE2: 0.66	PRE2: 0.75
	REC2: 0.77	REC2: 0.79	REC2: 0.80
	PRE3: 0.65	PRE3: 0.79	PRE3: 0.69
	REC3: 0.58	REC3: 0.75	REC3: 0.83
	PRE4: 0.42	PRE4: 0.64	PRE4: 0.61
	REC4: 0.50	REC4: 0.64	REC4: 0.58
1	ACC: 79.8%	ACC: 87.3%	ACC: 87.8%
	PRE1: 0.96	PRE1: 0.97	PRE1: 0.97
	REC1: 0.87	REC1: 0.95	REC1: 0.95
	PRE2: 0.44	PRE2: 0.59	PRE2: 0.67
	REC2: 0.87	REC2: 0.70	REC2: 0.76
	PRE3: 0.64	PRE3: 0.61	PRE3: 0.64
	REC3: 0.64	REC3: 0.62	REC3: 0.64
	PRE4: 0.40	PRE4: 0.62	PRE4: 0.49
	REC4: 0.41	REC4: 0.57	REC4: 0.61

L: Number of hidden layers, ACC: Accuracy, PRE*N*: Precision for class *N*, REC*N*: Recall for class *N*

6.3 Discussion

Prediction of cyber attack motivation using news articles. It is concluded that the news articles related to the cyber attack victims can be used as dataset. For the dataset creation, as depicted in Figure 6.6, using the feature ON-OFF and OFF-ON brought better accuracy and recall. [1] In the case of this study, recall is more impactful factor than precision. That is, predicting all of the upcoming cyber attacks without any loss (with better recall) is more essential than predicting cyber attack motivation with less noise (with better precision), so I focus on the recall of the experiment result.

From the result of the preliminary and additional experiment, it can be implied that feature selection is more influential than prediction algorithm. Also, it can be implied that using core keywords, ON-OFF (core words only appeared in the ON articles) and OFF-ON (core words only appeared in the OFF articles), as the feature is a better way to express the characteristics of the news articles than using all of the words.

I aggregated the daily average occurrence of the top 20 keywords mostly appeared in the news articles before an attack for some victims, and the results are depicted in Figure 6.7, 6.8, 6.10, 6.11, and 6.9. From the results, it seems that the top 20 keywords are represent a certain topic, and the topic causes the attack. I hypothesize that Figure 6.7, 6.8, 6.10, 6.11, and 6.9 are respectively correlated to the Paris Attack, the business success of Uber, hostage taking of ISIS, Obama's comments on Ukraine crisis, and law changes in Indiana state.

Twitter feeds. As far as the twitter feeds that I have collected are used, no remarkable pattern appears in the number and mood before an attack. One of the possible reason is that the amount of the twitter feeds is insufficient. Also, for the case of this study, twitter feeds can be implied as less reliable information source than news articles for knowing the public insights. Because, in general, news articles are edited by professional journalists and include emerging yet reliable facts.

[1] Here, precision refers the fraction of retrieved instances that are relevant, while recall refers the fraction of relevant instances that are retrieved. $Recall = \frac{TP}{TP+FN}$, $Precision = \frac{TP}{TP+FP}$, here *TP* is the number of true positive classifications, *FN* is the number of false negatives, and *FP* is the total number of false positives identified by the classifier.

Table 6.7 Top 20 keywords in the tweets posted by hacktivists against some of the cyber attack victims

Victim	Minecraft	Google CEO	Donald Trump	Wikileaks	Discord
keywords	stream	DuckDuckGo	want	Julian	discord
	urharmless	skills	will	successfully	thing
	hidden	website	backlash	anonymous	excuse
	time	YourAnonNews	outbuillied	detention	crisis
	minecraft	palestine	scare	wikileaks	KINGSA7AN
	least	new	worse	accept	misters
	on!	censored	YourAnonNews	government	unban
	mathewhutchison	spying	AnonyPress	Sullivan	please
	run	palestine	die	Rudolf	MaybeJord666
	nyxeira1	urharmless	JSavoly	western	womanbeater
	league	google	AnonyPress	e-mail	backdoored
	gay	JEWISH	Chinese	Brazil	hacking
	like	lack	elected	Clinton	thank
	service	urharmless	witness	online	love
	FuxNet	autism	trump	failed	RT
	videos	RT	anonymous	financed	migrant
	anon	hack	clinton	embed	invite
	EisMC2	anonymous	assassins	Ryan	internet
	RT	fake	black	bank	urharmless
	starting	machine	anti-trump	engine	sorry

I aggregated the top 20 keywords mostly appeared in the tweets posted by hacktivists against some of the cyber attack victims. The aggregation results for Minecraft, Google CEO, Donald Trump, Wikileaks, and Discord are depicted in 6.7. Although, the amount of the twitter feeds is insufficient for telling a concrete conclusion, it is obvious that the hacktivists tweets something related to the victim of a cyber attack before the attack.

Prediction of cyber attack opportunity using CVE feeds. Using high-dimensional features such as KEYWORD and JACCARD brought better prediction/classification results. A possible reason for that is that using the high-dimensional features avoids reduplication of the samples in a dataset. As depicted in Figure 6.12, the prediction result for the attack type *webapps* is best. A possible reason for that is that the number of samples labeled as webapps possesses greater part of the dataset.

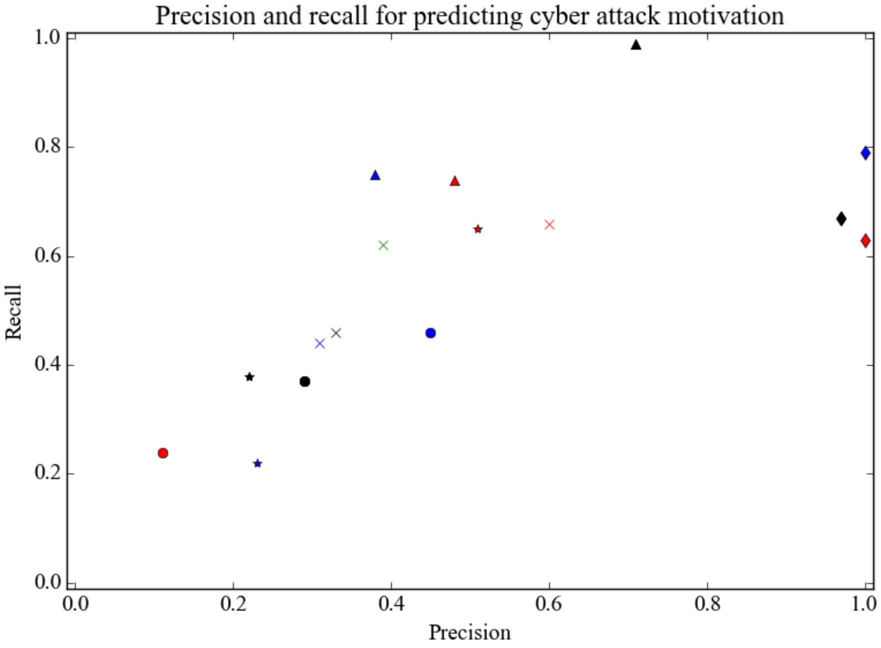


Figure 6.6 Precision and recall plot for the result of predicting cyber attack motivation

Red: One hidden layer, Blue: Two hidden layers, Green: Three hidden layers, Black: Four hidden layers, Circle: ON, Star: OFF, Triangle: ON-OFF, Diamond: OFF-ON, Cross: JACCARD

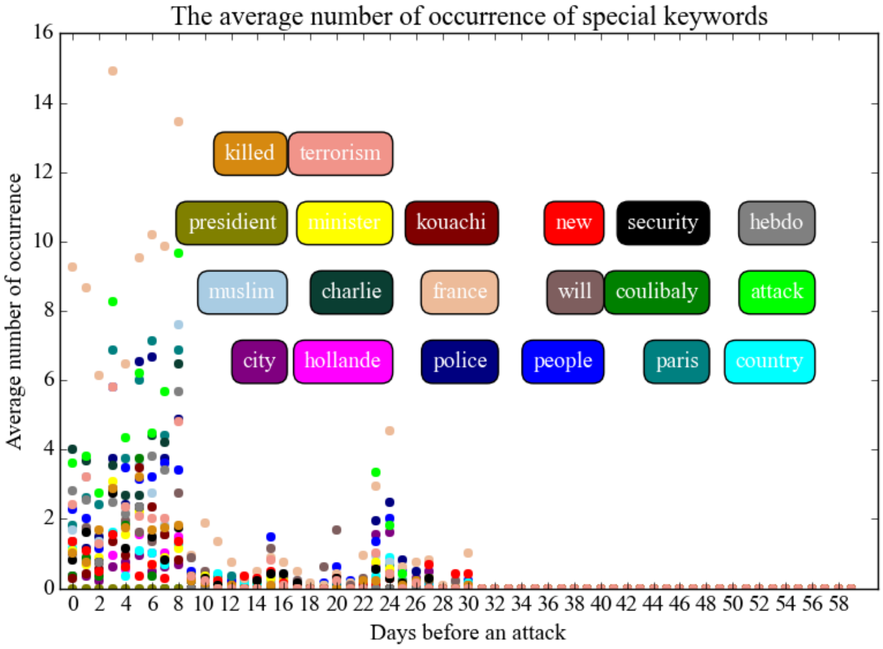


Figure 6.7 The average occurrence of the keywords appeared mostly before an attack (Target: French government agencies)

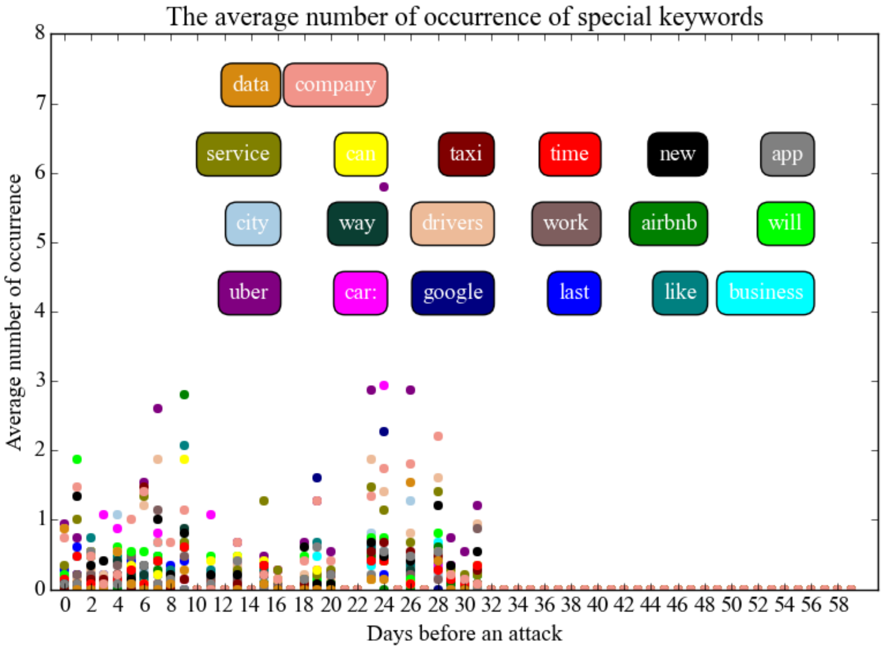


Figure 6.8 The average occurrence of the keywords appeared mostly before an attack (Target: Uber)

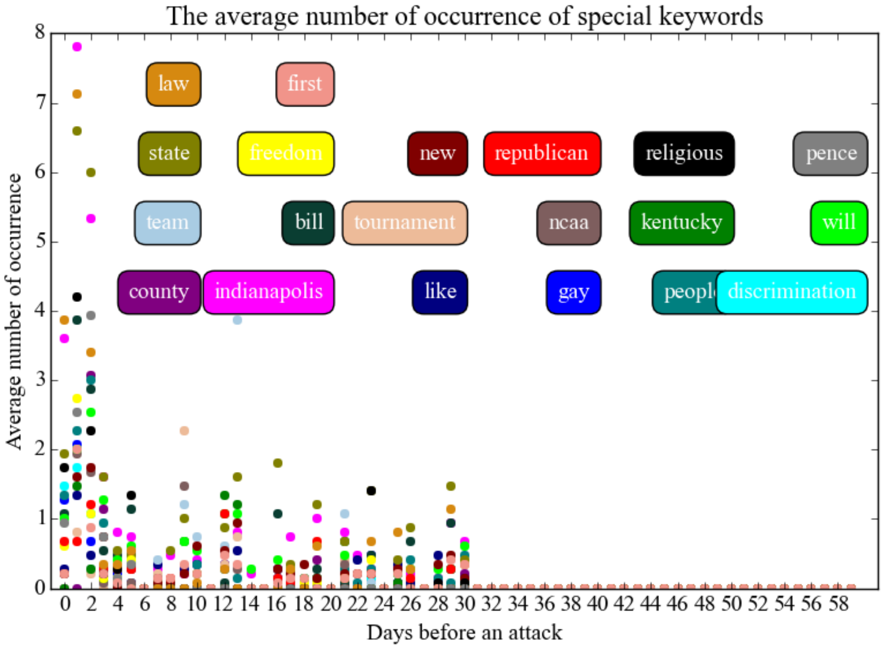


Figure 6.9 The average occurrence of the keywords appeared mostly before an attack (Target: Indiana government agencies)

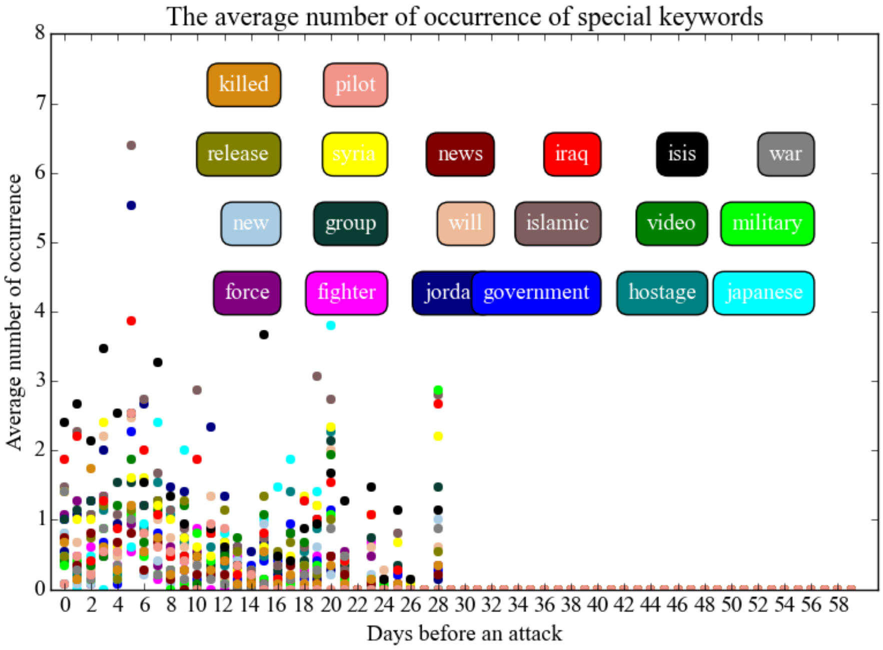


Figure 6.10 The average occurrence of the keywords appeared mostly before an attack (Target: ISIS)

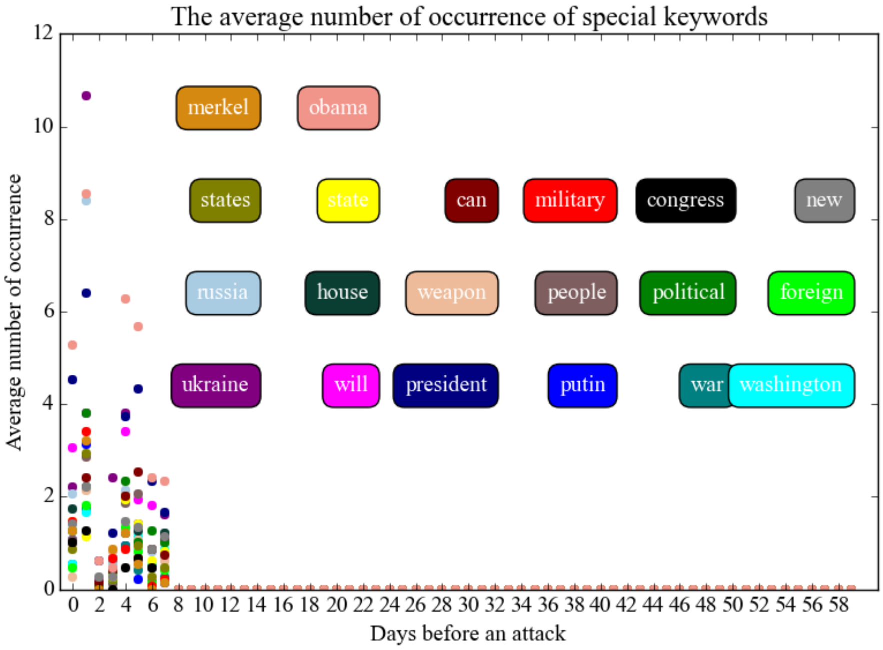


Figure 6.11 The average occurrence of the keywords appeared mostly before an attack (Target: Barack Obama)

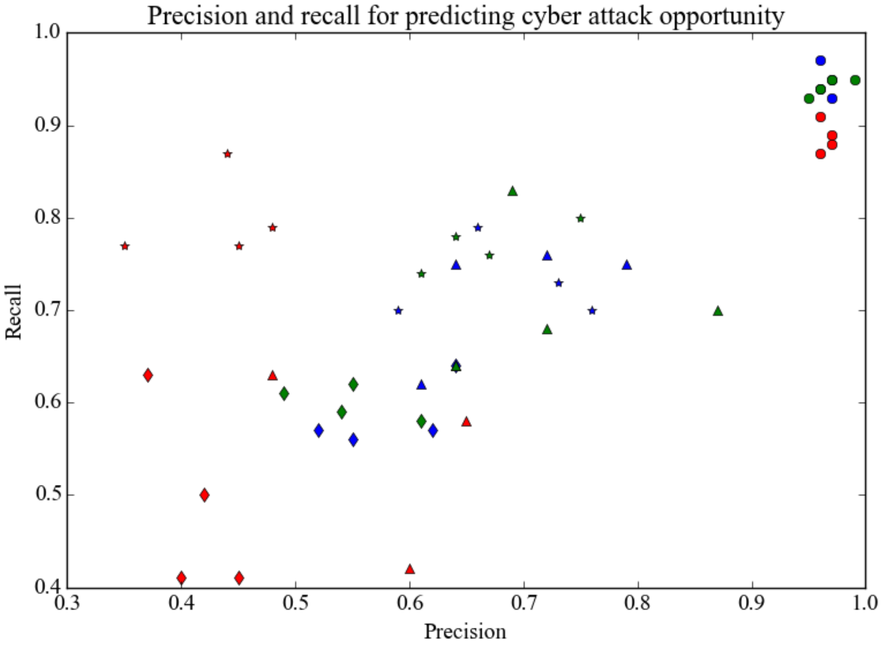


Figure 6.12 Precision and recall plot for the result of predicting cyber attack motivation

Red: BASIC8, Blue: KEYWORD, Green: JACCARD, Circle: Webapps, Star: Dos, Triangle: Local, Diamond: Remote

Chapter 7

SUMMARY

At the beginning of this chapter, the overall conclusion of this thesis is given. After that, the works have not been done in this study are clarified.

7.1 Conclusion

In this thesis, I attempted to explore the social data used for cyber attack prediction and promote the dataset creation methods. I define cyber attack prediction as the combination of the prediction of three factors, the motivation, opportunity (highly possible type of cyber attack), and occurrence timing of cyber attack. This thesis focused on the disclosure of useful social data and dataset creation methods for the prediction of cyber attack motivation and opportunity.

Social data for cyber attack prediction. I collected five types of social data including the archive of the cyber incidents occurred since 2015 to 2016. Except the archive, I collected the news articles related to the victims of the cyber incidents from well-known 15 medias including BBC, FoxNews, and NHK. Plus, I collected a certain amount of the two types of twitter feeds, the tweets posted by 90 hacktivist accounts around the world and tweets posted by common accounts in Japan. Security vulnerability feeds, or CVE feeds, were also collected from NVD.

Prediction of cyber attack motivation. Using the news articles related to the cyber attack victims as the dataset brought noticeable results. I divided the news articles into two parts, train set and test set, with the ratio of 80%/20%, and labeled all of the articles as positive (including a motivation of cyber attack) and negative (no correlation with cyber attack).

To label the dataset, I aggregated all of the news articles and investigated how the number of news articles on the target of a cyber attack increases before the attack. The result was interesting that the average number of news articles increased dramatically during the week before the attack. Consequently, I labeled the articles published in the period of the high rate increase as positive and labeled the articles published in the period of low/stable rate of increase as negative. To determine the threshold dates for the positive and negative line, I adopted two methods of dividing the dataset manually and dividing the dataset by means of K-Means clustering.

In the preliminary and additional experiment, I used both of the conventional machine learning algorithm as SVM and Artificial Neural Networks as CNN. Also, the experiments were conducted with few types of features including whole sentence and core keywords. As a result, using the core keywords extracted from only the *positive* or *negative* articles brought better accuracy and recall than the others.

I also investigated the number and mood of the tweets posted before a cyber attack. For the tweets of the 90 hacktivist accounts, it was found that the average number of the posts on the targets of the cyber attack increased during the first and fourth month before the attack. However, there was no noticeable movement in the mood of the tweets. On the other hand, for the twitter feeds posted in Japan, no obvious changes were shown in the number and mood of the tweets before an attack. On the contrary, the number of the tweets increased and the mood of the tweets went negative after the attack.

Prediction of cyber attack opportunity. I came up with the idea of analyzing CVE feeds and used the CVE feeds as the dataset for predicting the highly possible cyber attack type (cyber attack opportunity). To label the dataset, I used the Exploit Database and adopted few types of features including the BASIC8 and KEYWORD. As a result, using the higher dimensional features of KEYWORD and JACCARD brought better results.

7.2 Future Work

In this study, I focused on the social data preparation and dataset creation for the prediction of cyber attack motivation and opportunity. I haven't challenged the prediction of the timing of cyber attack occurrence, and it is remained as a future work. There are some researchers who have attempted to accurately predict the timing of cyber attack by analyzing CVE feeds, but no convincing result was acquired. Thus, disclosure of other types of social data that can be used for accurate prediction of cyber attack timing is necessary.

For the prediction of cyber attack motivation, all of the evaluation experiment were conducted on the data for the cyber incidents occurred in the past. I was not able to predict real-world cyber attack and check the results. Because, in general, the targets of cyber attack do not often reveal the information for the attack immediately. To evaluate the dataset creation and prediction methods that I have promoted on real-world cyber attacks is remained as another future work.

I collected and analyzed two types of twitter feeds. However, no remarkable pattern that could be used for the prediction of cyber attack motivation was found. I concluded that a possible reason was that the amount of the twitter feeds was insufficient. So, increasing the amount of the tweets and verifying whether twitter feeds can be used for cyber attack opportunity prediction are also in the remaining challenges.

For the prediction of cyber attack opportunity using CVE feeds, as same as the prediction of cyber attack motivation using news articles, the evaluation was conducted on the data for the cyber incidents occurred in the past. So, to evaluate the dataset creation and prediction methods on real-world cyber attacks is still a challenging task. Plus, the evaluation experiment showed higher precision yet recall for the cyber attack type *webapps*. An imbalance in the dataset is likely the reason for that, and in particular, the class webapps accounts for 48% of the dataset. Accordingly, increasing the amount of the other types/classes in the dataset to eliminate the imbalance is needed to be done in the future.

ACKNOWLEDGEMENT

Sincere gratitude is hereby extended to the generous people, without whose great supports this thesis would not have been possible.

Mostly, I would like to express my deepest appreciation to my supervisor Professor Sekiya Yuji at The University of Tokyo. His magnificent guidance and inspiring advices powered by the profound knowledge always yielded motivation for research. I am also grateful to Professor Sekiya's provision of a good environment to proceed this research.

I sincerely appreciate to Professor Kudoh Tomohiro, Profesor Nakayama Masaya, Professor Sato Hiroyuki, Professor Ogawa Takefumi, Lecturer Nakamura Fumitaka, and Lecturer Taenaka Yuzo of CNL (Computer Network Laboratory) at The University of Tokyo, for their appropriate guidance and great advices.

I want to show my genuine gratitude to the counsellor and editor of this thesis Lecturer Miyamoto Daisuke of CNL at The University of Tokyo, for his persistent encouragement and seasonable advices. He was always close to this research, spent plenty of whose precious time to promote and accomplish this thesis.

I am also profoundly grateful to Professor Aramaki Eiji at NAIST (Nara Institute of Science and Technology) for his great advices and donation of a precious data. Without his magnificent contribution, this research would not have been successful.

I would like to mention PhD candidate Niijima Arinobu and Seino Yoshihiro of CNL at The University of Tokyo, for their invaluable comments and attentiveness to this research.

My special gratitude is given to Graduate Students of CNL at The University of Tokyo, for their consciousness and assists. This research is partially associated with their studies, would not have been proceeded without their significant advices.

Eventually, I would heartily like to thank my colleagues at The University of Tokyo, colleagues from Data Artist Inc., and anonymous people who always gave me encouragement and precious advices.

PUBLICATIONS

Journals and Books

1. Baatarsuren Munkhdorj, Sekiya Yuji. “Cyber Attack Prediction using Social Data Analysis”, High Speed Networks (Jan.2017, to appear).

Domestic Conferences

1. バータルスレン ムンフドルジ, 関谷 勇司．“ニュース記事解析に基づいたサイバー攻撃予測”, 信学技法, vol. 115, no. 484, IN2015-136, pp. 165-170, 2016年3月.

International Conferences (Reviewed)

1. Baatarsuren Munkhdorj, Sekiya Yuji. “Cyber Attack Prediction using Social Data Analysis”, International Conference on Data Compression, Communication, Processing and Security (CCPS2016)(Mar.20XX, to appear).

REFERENCES

[1] P. Passeri, "November 2015 cyber attacks statistics," http://www.hackmageddon.com/2015/12/11/november-2015-cyber-attacks-statistics/.

[2] S. G. Kene and D. P. Theng, "A review on intrusion detection techniques for cloud computing and security challenges," in *Electronics and Communication Systems (ICECS), 2015 2nd International Conference*. IEEE, 2015, pp. 227–232.

[3] H. D. Nguyen, S. Gutta, and Q. Cheng, "An active distributed approach for cyber attack detection," in *Signals, Systems and Computers (ASILOMAR)*. IEEE, 2010, pp. 1540–1544.

[4] A. Amin, S. Anwar, A. Adnan, M. A. Khan, and Z. Iqbal, "Classification of cyber attacks based on rough set theory," in *Anti-Cybercrime (ICACC), 2015 First International Conference*. IEEE, 2015, pp. 1–6.

[5] "2016 internet security threat report," https://www.symantec.com/security-center/threat-report.

[6] R. Koch, B. Stelte, and M. Golling, "Attack trends in present computer networks," in *Cyber Conflict (CYCON), 2012 4th International Conference*. IEEE, 2012, pp. 1–12.

[7] Z. Zhan, M. Xu, and S. Xu, "Predicting cyber attack rates with extreme values," in *IEEE Transactions on Information Forensics and Security 10.8*. IEEE, 2015, pp. 1666–1677.

[8] T. R. Pillai, S. Palaniappan, A. Abdullah, and H. M. Imran, "Predictive modeling for intrusions in communication systems using GARMA and ARMA models," in *Information Tech-*

nology: Towards New Smart World (NSITNSW), 2015 5th National Symposium. IEEE, 2015, pp. 1–6.

[9] A. C. Sharma, R. A. Gandhi, W. Mahoney, W. Sousan, and Q. Zhu, “Building a social dimensional threat model from current and historic events of cyber attacks,” in *Social Computing (SocialCom), 2010 IEEE Second International Conference*. IEEE, 2010, pp. 981–986.

[10] R. Gandhi, A. Sharma, W. Mahoney, W. Sousan, Q. Zhu, and P. Laplante, “Dimensions of cyber-attacks: cultural, social, economic, and political,” in *Technology and Society Magazine*. IEEE, 2011, pp. 28–38.

[11] B. Wormuth and P. Becker, “Introduction to formal concept analysis,” in *2nd International Conference of Formal Concept Analysis February*, 2004.

[12] Y. Shynkevich, T. McGinnity, S. Coleman, and A. Belatreche, “Stock price prediction based on stock-specific and sub-industry-specific news articles,” in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1–8.

[13] A. K. Sirohi, P. K. Mahato, and V. Attar, “Multiple kernel learning for stock price direction prediction,” in *Advances in Engineering and Technology Research (ICAETR), 2014 International Conference*. IEEE, 2014, pp. 1–4.

[14] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” in *Journal of Computational Science 2.1*. Elsevier, 2011, pp. 1–8.

[15] H. Holm, M. Ekstedt, and D. Andersson, “Empirical analysis of system-level vulnerability metrics through actual attacks,” in *Transactions on Dependable and Secure Computing 9.6*. IEEE, 2012, pp. 825–837.

[16] “Common vulnerability scoring system, v3 development update,” https://www.first.org/cvss.

[17] W. Boyer and M. McQueen, "Ideal based cyber security technical metrics for control systems," in *International Workshop on Critical Information Infrastructures Security*. Springer Berlin Heidelberg, 2007, pp. 246–260.

[18] C. Sabottke, O. Suciu, and T. Dumitra, "Vulnerability disclosure in the age of social media: exploiting twitter for predicting real-world exploits," in *24th USENIX Security Symposium (USENIX Security 15)*, 2015, pp. 1041–1056.

[19] M. Bozorgi, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond heuristics: learning to classify vulnerabilities and predict exploits," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 105–114.

[20] "Offensive security exploit database archive," https://www.exploit-db.com/.

[21] "Microsoft security advisories," https://technet.microsoft.com/en-us/security/advisories.aspx.

[22] "Vulnerabilities & exploits," https://www.symantec.com/connect/topics/security/vulnerabilities-exploits.

[23] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern recognition letters*, vol. 31, no. 8, pp. 651–666, 2010.

[24] S. M. Mohammad and P. D. Turney, "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon," in *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. Association for Computational Linguistics, 2010, pp. 26–34.

[25] J. Grimmer and B. M. Stewart, "Text as data: The promise and pitfalls of automatic content analysis methods for political texts," in *Political Analysis*. SPM-PMSAPSA, 2013, p. mps028.

[26] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," in *Proceedings of the 24th International Conference on World Wide Web*. ACM, 2015, pp. 1067–1077.

[27] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," 2016, book in preparation for MIT Press. [Online]. Available: http://www.deeplearningbook.org

[28] K. Yoon, "Convolutional neural networks for sentence classification," in *arXiv preprint arXiv:1408.5882*, 2014.

[29] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[30] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, no. Aug, pp. 2493–2537, 2011.

[31] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems*, 2013, pp. 3111–3119.

[32] "Common vulnerabilities and exposures," https://cve.mitre.org/.

[33] "Hackmageddon information security timelines and statistics," http://www.hackmageddon.com/.

[34] "abc news: Cyber attack news," http://abcnews.go.com/topics/news/cyber-attack.htm.

[35] "Entrepreneur: Cyber attacks," http://www.entrepreneur.com/topic/cyber-attacks.

[36] "Ndtv: Cyber attacks news," http://www.ndtv.com/topic/cyber-attacks/news.

[37] "Google custom search engine," https://cse.google.jp.

[38] "Beautiful soup documentation," https://www.crummy.com/software/BeautifulSoup/bs4/doc/.

[39] "National vulnerability database," https://nvd.nist.gov/.

[40] "The security content automation protocol (scap)," https://scap.nist.gov/.

[41] "Common vulnerability scoring system v3.0: Specification document," https://www.first.org/cvss/specification-document.

[42] "First is the global forum for incident response and security teams," https://www.first.org/.

[43] R. Kuć and M. Rogozinski, *ElasticSearch server*. Packt Publishing Ltd, 2016.

[44] "Twitter developer documentation," https://dev.twitter.com/rest/reference/get/search/tweets.

[45] "Social computing," http://isw3.naist.jp/Contents/Research/mi-08-en.html.

[46] T. Joachims, "Introduction to support vector machines," 2002.

[47] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," in *ACM Transactions on Intelligent Systems and Technology (TIST) 2*. ACM, 2011, p. 27.

[48] "word2vec," https://code.google.com/archive/p/word2vec/.

[49] H. Takamura, T. Inui, and M. Okumura, "Extracting semantic orientations of words using spin model," in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2005, pp. 133–140.

[50] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.