International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 20 (2017) pp. 9909-9913
© Research India Publications. http://www.ripublication.com

Detection of DDoS Attack on the Client Side Using Support Vector Machine

Donghoon Kim * and Ki Young Lee**

**Department of Information and Telecommunication Engineering, Incheon National University, Incheon, 22012, Korea.*

*** Department of Information and Telecommunication Engineering, Incheon National University, Incheon, 22012, Korea.*

Corresponding author: Ki Young Lee, Professor.

Abstract

Nowadays, cyber-attack attempts have been increasing and DDoS (Denial of Service) attacks are one of the major threat in computer networks. It attempts server's resources unavailable and generates massive traffic on the network. These attacks are evolving very quickly in scope and complexity. Also, it is not possible to prevent the network congestion even if these attacks are detected on the server side, unless client as act a zombie blocks their attack packets. To detect DDoS attack early and invalidate the attack itself, this paper proposes a method to detect DDoS attack on the client side using machine learning algorithm

Keywords: DDoS Attack, Attributes Selection, SVM, Cyber Security, Machine Learning

INTRODUCTION

Nowadays, cyber-attack attempts have been increasing and DDoS (Denial of Service) attacks are one of the major threat in computer networks. In the past, DDoS attacks have forced server's resources to be used or making the network itself congested. However, recent scope of the attack has been extended to make zombie PC's hard drive unavailable for removing their cyber foot prints. And there are many side effects because of the attack. It generates huge traffic on the network, so many of the general users could suffer from low latency. Furthermore, the users who do not know that their PC is performed an attack may suffer financially by the billing. Therefore, not only servers need to detect the attack, but also clients should recognize whether the attack is performing by their own. It is possible to reduce network traffic and mitigate success of the attack to server.

Many research studies continue to figure out the solution for DDoS attack. The traditional method for detecting DDoS attacks has been used the pattern matching. However, the methods of DDoS attacks are getting complicated and variant to avoid the pattern matching algorithms. In recent years, various types of machine learning methods have been introduced into detection systems to detect such complex, disordered, and changing attacks. However, machine learning algorithm can flexibly adapt to various variants of attacks.

The remainder of this paper is organized as follows. In Section 2, we review publications in this research. In Section 3, we propose attributes for classification and propose a method for detecting DDoS attacks on the client side using SVM (support vector machine). Then we evaluate performance of our method and we make a conclusion and suggest future works.

RELATED WORKS

Many of detection method for DDoS attack using machine learning such as Bayesian network, SVM and other algorithms. The machine learning algorithms use predefined attributes for classification and generate a model using a large amount of data set composed of this information. Then apply the new data to this model for classification. Most of studies on DDoS detection using machine learning have studied the improvement of performance by decreasing the false positive rate, increasing the detection rate and/or decreasing computational complexity using 41 attributes and data set defined in KDD99 [1, 2].

Table 1. Basic features of individual TCP connections.

Attribute	Description
duration	length (number of seconds) of the connection
protocol_type	type of the protocol.
service	network service on the destination
src_bytes	number of data bytes from source to destination
dst_bytes	number of data bytes from destination to source
flag	normal or error status of the connection
land	1 if connection is from/to the same host/port; 0 otherwise
wrong_fragment	number of wrong fragments
urgent	number of urgent packets

Table 2. Content features within a connection suggested by domain knowledge.

Attribute	Description
hot	number of hot indicators
num_failed_logins	number of failed login attempts
logged_in	1 if successfully logged in; 0 otherwise
num_compromised	number of compromised conditions
root_shell	1 if root shell is obtained; 0 otherwise
su_attempted	1 if su root command attempted; 0 otherwise
num_root	number of root accesses
num_file_creations	number of file creation operations
num_shells	number of shell prompts
num_access_files	number of operations on access control files
num_outbound_cmds	number of outbound commands in an ftp session
is_hot_login	1 if the login belongs to the hot list; 0 otherwise
is_guest_login	1 if the login is a guest login; 0 otherwise

Table 3. Traffic features computed using a two-second time window.

Attribute	Description
count	number of connections to the same host as the current connection in the past two seconds
serror_rate	% of connections that have SYN errors
rerror_rate	% of connections that have REJ errors
same_srv_rate	% of connections to the same service
diff_srv_rate	% of connections to different services
srv_count	number of connections to the same service as the current connection in the past two seconds
srv_serror_rate	% of connections that have SYN errors
srv_rerror_rate	% of connections that have REJ errors
srv_diff_host_rate	% of connections to different hosts

However, W. Wang et al. pointed out that some of the attributes may be redundant or even act like noise which decrease detection performance [3]. They ranked for 41 attributes through information gain and chi-square test. Then they showed that the best performance was obtained by using only 9 attributes through their experiment.

S. Umarani et al. used PCA (principle component analysis) to reduce dimension to remove redundant and noise form attributes [4].

However, many of the attributes defined in KDD99 are composed of information that can be obtained from the server side and it contains information about various cyber-attacks such as unauthorized access from a remote machine, port scanning or buffer overflow, as well as information about DDoS as shown in the table 1-3, so it is not appropriate to use on the client side for detecting DDoS attacks. Therefore, it is one of the main issue to select appropriate attributes for detecting attacks from huge massive traffic.

Other studies are focused on increasing performance with various machine learning algorithms used in detection of DDoS attacks including decision tree, Naïve Bayes, artificial neural networks and SVM [5-10]. Nowadays, SVM has become an extremely popular algorithm for classification and future estimate problems because Naïve Bayes and artificial neural networks depends on the number of data set and attributes. These algorithms require huge amount of data set to minimize empirical error. Therefore, in the absence of enough training set, a significant drop in performance may occurs. On the other hand, SVM requires relatively less data set than above techniques, so it performs very well without enough training set especially in binary classification. However, complex data transformations and resulting boundary plane are very difficult to interpret.

PROPOSED ALGORITHM

In this section, we propose attributes and a method for detecting DDoS attack on the client side. Detection performance in machine learning is dependent on the number and definition of valid attributes, and the number of training sets for learning. Since each machine learning has advantages and disadvantages, it is important to select the appropriate technique depending on the application. For binary classification, we defined attributes given in Table 4 and choose SVM as machine learning algorithm

Table 4. Attributes for detecting DDoS attack on the client side.

Attribute	Description
connection_duration	elapsed time time since connection
header_size	ip packet header size
cumulative_size	Cumulative size of transmitted packets
connected_port	the number of connected port to the same host
service	Network service on the destination
flags	normal or error status of the connection
protocol	type of the protocol

Connection_duration and cumulative_size can provide information to detect DDoS attack since most DDoS attacks continuously transmit massive packets to a victim server, the client side must keep an eye on the size of the accumulated packets over the time of the connection.

Also, DDoS attacks such as SYN flood can be suspected when many ports are assigned to the same server. In addition, IP header information including header_size, service, flags and protocol are used for classifying normal and abnormal status.

We use SVM method to create a model for normal and abnormal binary classification with a limited amount of training set for detecting DDoS attack on the client.

The experiment was carried out in the following manner. All outgoing packets including normal and attack packets from a client were captured over three days for data collection. We installed a Tomcat web server on another PC, and the client performed DDoS attack to the server using DDoS attack tools, stacheldraht and tfn (tribe flood network) at some specific times. Figure 1 shows our DDoS attack scenario.

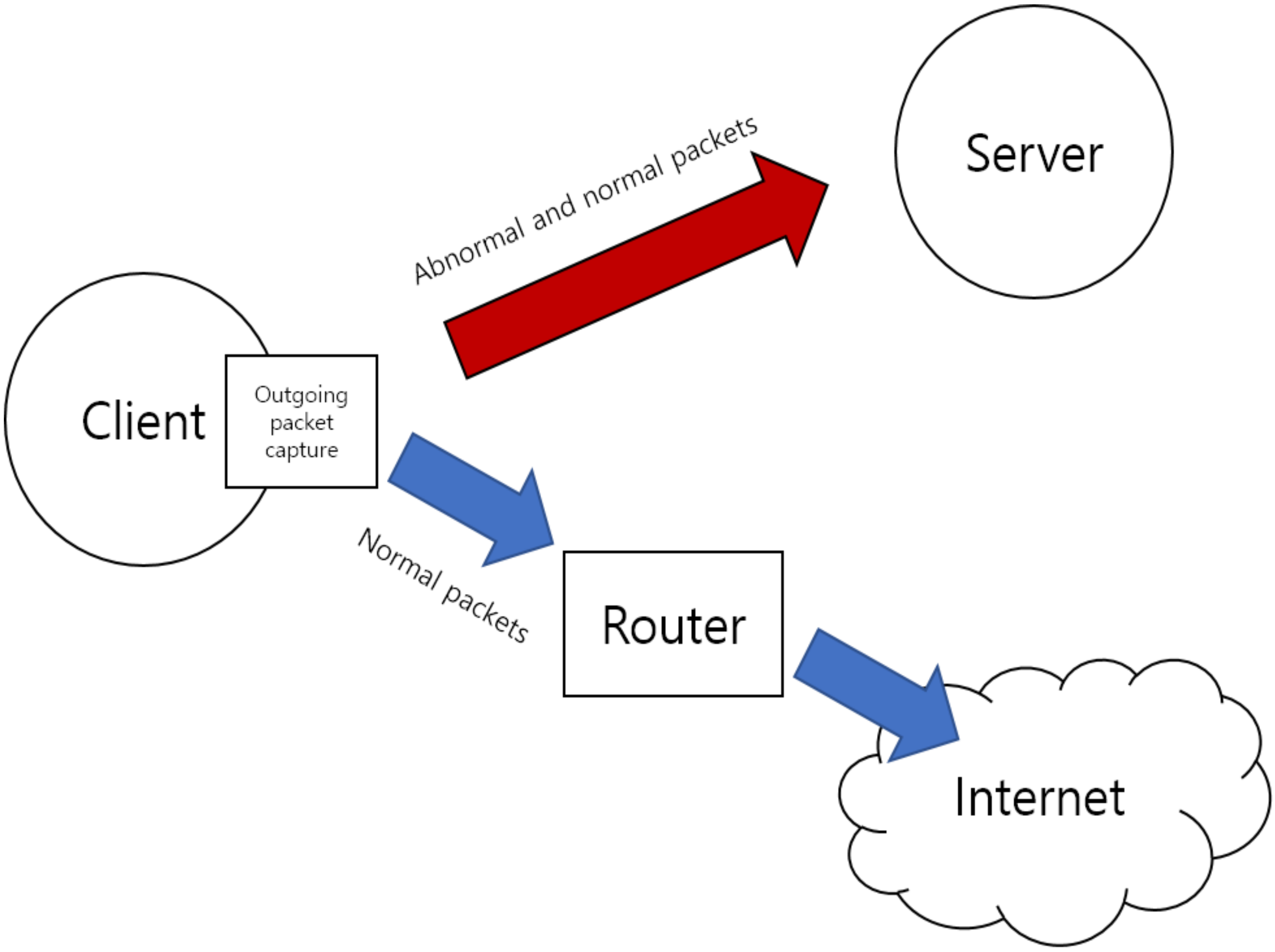

Figure1. A DDoS attack scenario for collecting data.

Afterwards, the captured packet information is converted into the values for the attribute defined above and marked with normal and attack packets. We randomly selected 100,000 datasets from the total data, used 70,000 and 30,000 of them training set and test set respectively. Of the 100,000 data sets, only about 15% of the data was marked as attack. The method used for DDoS attack is shown in Table 5

Table 5. Attack name and the number of data used for the attack

Attack Name	The number of data for training set	The number of data for test set
stacheldraht_ICMP	1774	713
stacheldraht_UDP	1962	802
stacheldraht_TCP_SYN	1483	689
tfn_TCP	1502	643
tfn_UDP	1621	832
tfn_TCP_SYN	2054	894

RESULTS AND DISCUSSION

We generate a model to detect DDoS attack using SVM and Bayesian networks with 10,396 training sets and then performed performance analysis using test sets. The results are shown in Table 6 and Table 7.

Table 6. Accuracy of SVM and Bayesian networks with proposed attributes

Method	Detection Rate (%)	False Positive Rate (%)
SVM	99.94	0.143
Bayesian Networks	98.57	0.106

Table 7. Computational time evaluation for SVM and Bayesian networks with proposed attributes

Method	Training (s)	Test (s)
SVM	0.54	0.15
Bayesian Networks	1.43	0.32

From the results in the above two tables, SVM performs better than the Bayesian network in accuracy and computational time because the total number of data sets and the number of attributes are not large. Thus, SVM provides good performance even with a small number of data sets. However, the attack methods used in this experiment are not sophisticated or complicated enough as they are applied in practice, and the number of attributes is small and these attributes may not key attributes in detecting DDoS attacks as arbitrarily selected. Therefore, the reliability of this experiment may be lacked and it is necessary to design the experiment to increase the reliability in the future. Also, it may be reasonable to obtain the raw data packets of the various clients used in the actual DDoS attack and to use it for the performance evaluation. Furthermore, the key attributes extraction, such as client's network usage pattern or other network-related attributes, applicable to the client to increase the DDoS attack detection performance through the various studies will be additionally needed

CONCLUSION

In this paper, we have proposed some attributes and a method for detecting variant of DDoS attack on the client side using SVM. However, the attack methods used in this experiment are not sophisticated or complicated enough as they are applied in practice, and the number of attributes is small and these attributes are arbitrarily selected. Since the number of attributes and effectiveness for detection are very important for machine learning, it is necessary to use user's network usage pattern or other network-related attributes for practical applications. Also, obtaining the raw data packets of the various clients used in the actual DDoS attack is needed to reliable performance evaluation.

Furthermore, future research will be needed to optimize performance using dimension reduction such as Information Gain, Chi-square statistics and PCA (principal component analysis).

ACKNOWLEDGMENTS

This work was supported by the Incheon National University Research Grant in 2014.

REFERENCES

[1] W. Lee, S. Stolfo, K. Mok, A Data Mining Framework for Building Intrusion Detection Models, Proceedings of the 1999 IEEE Symposium on Security and Privacy, (1999), 120-132.

[2] KDD Cup 1999 data: http://kdd.ics.uci.edu/databases/kdd99/kddcup99.html

[3] Wei Wang, Sylvain Gombault, Efficient Detection of DDoS attacks with Important Attributs, Third International Conference on Risks and Security of Internet and Systems: CriSIS'2008, (2008), 61-67.

[4] S. Umarani, D. Sharmila, Predicting Application Layer DDoS Attacks Using Machine Learning Algorithms, International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol 8, No 10, (2014), 1907-1912.

[5] Taeshik Shon, Yongdae Kim, Cheolwon Lee, Jongsub Moon, A machine learning framework for network anomaly detection using SVM an GA, Information Assurance Workshop, (2005), 176-183.

[6] S. Seufert, D. O'Brein, Machine Learning for Automatic Defence Against Distributed Denial of Service Attacks, Communications, (2007), 1217-1222.

[7] T. Subbulakshmi, K. BalaKrishnan, S. Mercy Shalinie, D. AnandKumar, V. GanapathiSubramanian, K. Kannathal, Detection of DDoS attacks using Enhanced Support Vector Machines with real time generated dataset, Advanced Computing (ICoAC), (2011), 17-22.

[8] A. Ramamoorthi, T. Subbulakshmi, S. Mercy Shalinie, Real time detection and classification of DDoS attacks using enhanced SVM with string kernels, Recent Trends in Information Technology (ICRTIT), (2011), 91-96.

[9] Sujay Apale, Rupesh Kamble, Manoj Ghodekar, Hitesh Nemade, Rina Waghmode, Defense Mechanism for DDoS Attack Through Machine Learning, International Journal of Research in Engineering and Technology, Vol. 03, Issue 10, (2014), 291-294.

[10] Niharika Sharma, Amit Mahajan, Vibhakar Mansotra, Machine Learning Techniques Used in Detection of DOS Attacks: A Literature Review, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 6, Issue 3, (2016), 100-105.