DDoS Attack Forecasting System Architecture Using Honeynet

Dongwoo Kwon and James Won-Ki Hong
Division of IT Convergence Engineering
Pohang University of Science and Technology (POSTECH)
Pohang, Republic of Korea
{dwkwon, jwkhong}@postech.ac.kr

Hongtaek Ju
Department of Computer Engineering
Keimyung University
Daegu, Republic of Korea
juht@kmu.ac.kr

Abstract— **This paper proposes a proactive security system to forecast Distributed Denial of Service (DDoS) attacks. A reactive system focused on detection after network attacks occur has difficulties responding rapidly to massive distributed attacks, such as DDoS. By forecasting the attack, we can take active countermeasures such as strengthening the power of the security devices and it would also enable us to plan a recovery procedure and countermeasures beforehand, providing a more rapid response. In this paper, we discuss previous studies related to intrusion forecasting, define the concept of intrusion forecasting and propose the Internet Intrusion Forecasting System Architecture. To obtain intrusion factors for DDoS attack forecasts, Honeynet was deployed and we analyze Hflow data gathered from Honeynet.**

Keywords-DDoS; Honeynet; Intrusion Forecasting; Proactive Network Security

I. INTRODUCTION

Most of the today's network security systems are reactive; they detect the intrusion of the network first and then manage it. Examples of such reactive systems are Firewall and Intrusion Detection/Prevention System (IDPS). However, reactive security systems are limited. The limitation becomes obvious when we consider the DDoS attack. DDoS attacks concentrate a tremendous amount of traffic on the targeted system and stop it using a Botnet that consists of a large number of Zombie PCs that are already infected by malicious Bot agents. It is difficult to distinguish between normal traffic and the attack traffic due to the nature of the distributed attack. In addition, reactive security systems tend to be easily disabled by DDoS attacks, because by the time the attack is detected, a large number of attack packets are already inbounded.

We need a proactive security system, which predicts and measures potential attacks, to overcome the limitations of reactive security systems that focus on detection. Intrusion forecasting calculates the possibility of a potential attack, enabling us to plan a specific countermeasure when the probability of an attack is high. By predicting DDoS attacks, we can take active countermeasures, such as strengthening the security system temporally or inbounding appropriate additional security devices by consulting a specialized organization. In addition, it would enable us to plan a recovery procedure and plan countermeasures beforehand, providing a more rapid response. For example, we can strengthen the security of the system for which an attack is suspected, or even duplicate the suspected system to replace the original after an attack. In the long term, it creates a more secure and effective system, because it provides a specific and reasonable budget for the security system.

The previous studies regarding intrusion forecasting lack certain details. Most of the prediction methods merely depend on preceding attack trends [1-5]. They do not provide a specific forecasting of the exact type, time and target of the attack. Although these studies are meaningful, we need more specific and concrete forecasts for proactive forecasting systems to be effective. There are also the studies that predict the propagation of attacks and predict the next stage of attacks based on information from network scanning and present attacks [6-8]. However, these studies shorten the detection time but they do not guarantee sufficient time to provide a countermeasure to the attack.

In this paper, we present our ongoing study for Internet intrusion forecasting. We discuss the related studies of the forecasting the Internet intrusion in Section 2. For the comparison with detection, we define the concept of the intrusion forecasting and design general system architecture to predict the Internet intrusion in Section 3. We have deployed the Honeynet as a way to collect raw data and analyze Hflow data gathered from the Honeynet in Section 4. Finally, we finish this paper with concluding remarks and future works in Section 5.

II. RELATED WORK

Hideshima et al. [1] developed a system called STARMINE. This system visualizes the attack traffic in a 3-dimensional graph using spatial, temporal and logical analysis. This study provided the basis to understand the characteristics of attack traffic and to predict intrusion. The forecast depends on the judgment of the individual who interprets the graph.

Schechter [9] proposed a forecasting method to predict the probability of Internet intrusion using a regression model. The study merely provided a theoretical approach using an econometric model of the intruder and the victim rather than presenting experiments and quantifiable results. However, the study is worth noting because it emphasizes the possibility of specifically forecasting attacks, rather than merely predicting increases or decreases in attack frequency.

Takemori et al. [6] proposed the Security Operation Center (SOC) framework for the cooperation between ISPs to forecast new attacks. This framework performs statistically automated and manual forecasts using Bayesian network. It quickly detects abnormal events in a high-speed network and selects a target by predicting the type and quantity of the attack. Although the purpose of the prediction is not to prepare for the attack, this study is valuable because it predicts the attack in a spatial rather than a temporal context.

Sindhu et al. [7] used Neuro-genetic algorithm to predict attacks within a short time. The purpose of this study is to predict and block attacks just before they occur to improve the effectiveness of IPS. Nanda et al. [8] predicted attacks by using graph theory. This study proposed a model that uses system vulnerability to predict the progression of attacks. This study also attempts to shorten the time of intrusion detection.

Ishida et al. [2] proposed a forecasting method using Bayesian inference, which calculates the increase or decrease of the probability of the next attack based on the number of attacks observed previously. Zhang et al. [3] numerically expressed the present security situation, and used time-series analysis to forecast the variation of the security situation due to time. Both works, by forecasting the increase or the decrease of intrusions, serve as a valuable foundation for the field of intrusion forecasting.

Kim et al. [4] proposed a method to forecast the increase or decrease of the Bot agents by month that uses Hidden Markov Model (HMM). This study argues HMM is a superior forecasting method for predicting attacks than time-series analysis, because time-series analysis does not precisely represent the hidden characteristics of attacks. Kim et al. [5] proposed the framework of an intrusion forecasting system that is more accurate by using two algorithms, rather than just one algorithm. This study proved that accuracy is particularly high when they used Markov chain and time-series analysis. These two studies worked to improve the accuracy of forecasting based on the increase or decrease of attacks. Our study is focused on informative forecasts by providing us with identifying the type, time and target of attacks rather than merely forecasting the increase or decrease of attacks.

III. INTRUSION FORECASTING

A. *The Concept of Intrusion Forecasting*

Intrusion forecasting predicts the attack possibilities by analyzing previous statistics from various environments and the present network situation, just as weather forecasting does. It predicts the type, time and target of the attack. The acquisition of administration authority, using software vulnerability, is an example of a type of direct attack. In contrast, DDoS attack, which exhausts network resources, is a type of indirect attack. The time of the attack is the timeframe that has a high attack probability and the attack targets are specific targets, such as web, database and e-mail servers, installed inside the network.

Intrusion factors are necessary to determine the probability of an attack. There is the exterior and interior intrusion factor. The exterior intrusion factor occurs outside the target network and represents the risk imposed by the external environment. Conversely, the interior intrusion factor originates within the target network, and demonstrates the reason this particular network is vulnerable to an attack.

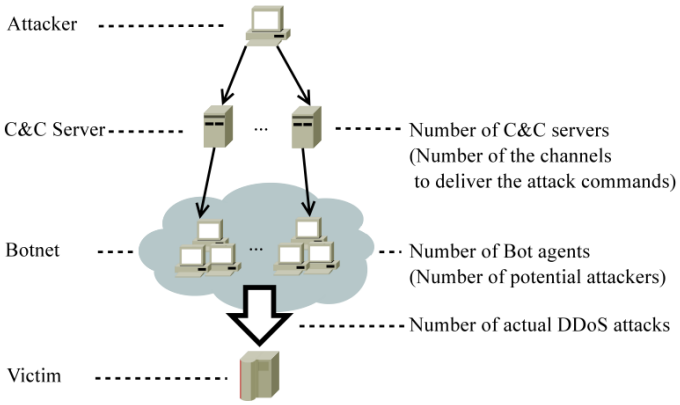

Figure 1. Exterior Intrusion Factors for DDoS Attack

Fig. 1 shows examples of exterior intrusion factors for DDoS attack. The international number of malicious Bot agents, the number of Command and Control (C&C) servers, and the number of actual DDoS attacks are the intrusion factors. The number of the malicious Bot agents represents the number of potential attackers that can directly attack the target. C&C servers are the channels that deliver the attack commands. The larger the number of C&C servers, the longer it takes to find and block each one. Both the number of Bot agents and C&C servers can increase the severity of the impact of the attack. The risk imposed by the outside environment increases as the numbers of the malicious Bot agents, the C&C server and the actual size of the DDoS attack increase.

The interior intrusion factors, such as the value of the attack target and network security level, determine which networks attackers choose to target. The value of the attack target represents the network's importance and the network security level represents the cost of the attack. A target with high value and a low level of network security is a high-risk target.

B. *Internet Intrusion Forecasting System Architecture*

We designed the architecture of the Internet Intrusion Forecasting System by conducting analysis from physical, technical and informational perspectives. Fig. 2 shows the architecture of the proposed intrusion forecasting system. This system consists of four modules: Data Provider, Data Collection Module, Intrusion Forecasting Module, and Forecasting Post-processing Module.

The Data Provider offers the raw data to predict the attack. It consists of the central security agency, the Honeynet data analysis system, network security devices, and network operators. The central security agency and the Honeynet data analysis system provide exterior intrusion factors while the network operator provides interior intrusion factors. The central security agency is Internet security organization such as Computer Emergency Response Team (CERT). The central security agency provides various exterior intrusion factors that are difficult to measure. The Honeynet data analysis system directly observes and analyzes the attacking traffic and

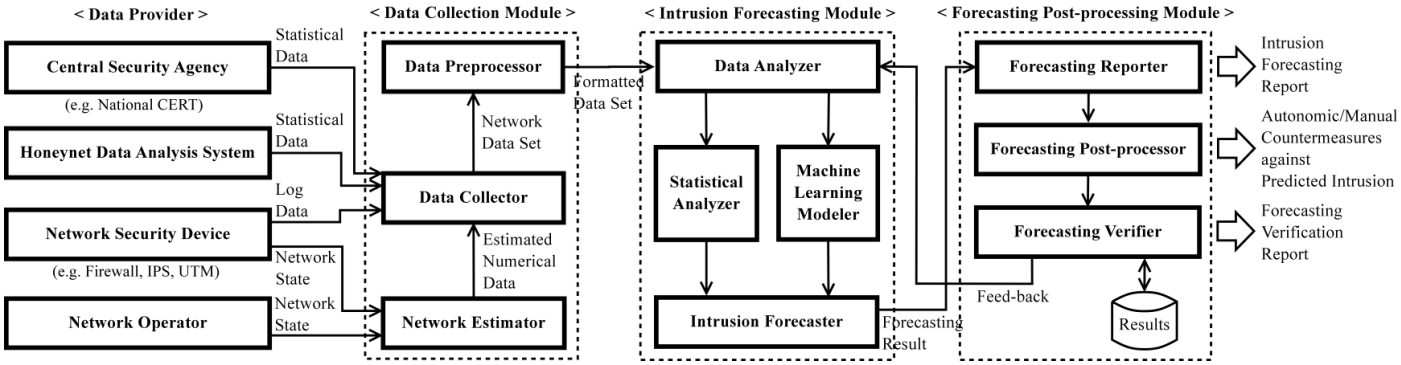

Figure 2. Internet Intrusion Forecasting System Architecture

intrusive activity through multiple Honeypots. The network security devices are reactive security devices, such as Firewall, IPS and Unified Threat Management (UTM). These provide the data, such as security logs and traffic data, which used for estimating the target's network security level. Finally, network operators provide the direct interior intrusion factors, such as the importance of the targeted network.

The Data Collection Module collects the data from the Data Provider. This module consists of the Data Collector, Network Estimator and Data Preprocessor. The Data Collector collects quantitative data from the central security agency, the Honeynet data analysis system and network security devices. The Network Estimator collects qualitative data, which is difficult to express in objective numbers, from the network security devices and network operators. After estimating and converting the data into quantitative data, it sends them to the Data Collector. The Data Preprocessor converts the data into an analysis-friendly form and sends them to the Intrusion Forecasting Module.

The Intrusion Forecasting Module is the central module of the proposed architecture. It consists of the Data Analyzer, Statistical Analyzer, Machine Learning Modeler, and Intrusion Forecaster. The Data Analyzer determines possible correlations among the pre-processed data and sends them to the appropriate analyzer by observing the distribution and the forecast type. The Statistical Analyzer analyzes the data using statistical models, such as time-series, regression and structural equation model analysis. The Machine Learning Modeler uses machine learning algorithms that exclude statistical approaches. Finally, the Intrusion Forecaster creates the intrusion forecast by integrating the algorithm-categorized results. It uses several appropriate algorithms simultaneously to provide higher accuracy according to the experimental result of Kim et al [5].

The Forecasting Post-processing Module consists of the Forecasting Reporter, Forecasting Post-processor and Forecasting Verifier. The Forecasting Reporter generates a year/month/week forecast intrusion report from the provided result and sends it to the central security agency and the network operator. The Forecasting Post-processor conducts autonomic countermeasures against predicted intrusion or provides appropriate guidelines for network operators. The Forecasting Verifier verifies the result by comparing it to the actual attack after the predicted time. Additionally, it improves the forecasting algorithm by sending feedback to the Intrusion Forecaster to provide higher accuracy by adjusting and applying data parameter values.

IV. HONEYNET DATA ANALYSIS

To gather raw data for intrusion factors, we have deployed the internal Honeynet and the external Honeynet of our university. The Honeynet provides more detail data such as system logs than statistical data from security organization. Using the Honeynet also has the advantage of collecting valuable raw data. It holds a meaningful correlation, since the location of the installed Honeynet and the attacked network are close by. We use Hflow to integrate and store various types of data such as network flows, IPS logs, and data regarding intrusive activity captured by Sebek. As a first step to forecast DDoS attacks, we analyzed Hflow data gathered from the Honeynets. This analysis used data collected between December 24th, 2010 and August 23th, 2011.

A. Inbound Traffic Analysis

Most of the total inbound traffic is TCP packets (99.8%) and the others are UDP (0.18%) and ICMP (0.02%) packets. TCP packets consisted of port 22 for SSH (92.8%), port 21 for FTP (4.45%), Etc. (2.75%). The result of analyzing sampled the flows on port 22 presented SSH brute-force attacks that were performed by the attacker to obtain administrator authority. Fig. 3 shows the number of inbound packets entering from the internal and external Honeynets of our university. In the external Honeynet, the attack occurs intensively between March 10th and March 16th of 2011. The peak of the attack occurred on March 15th. The attack succeeded and the attacker acquired administrator authority on March 16th. In the internal Honeynet, the attack succeeded on June 9th.

To trace intrusive activity, we analyzed the data of Process, Command, Process_to_com, and Sys_open tables in Hflow database. When the attacker successfully connected as an administrator to the Honeypots, the public key of the attacker was added to the file for known hosts and SSH settings were changed for host-based authentication. This enables an attacker to connect to the systems as an administrator without any authentication. Finally, the script file for brute-force SSH attacks and the binary files of the Bot agent were executed in the background. TCP packets on port 21 were used to perform brute-force attacks for acquiring upload permission. Most of

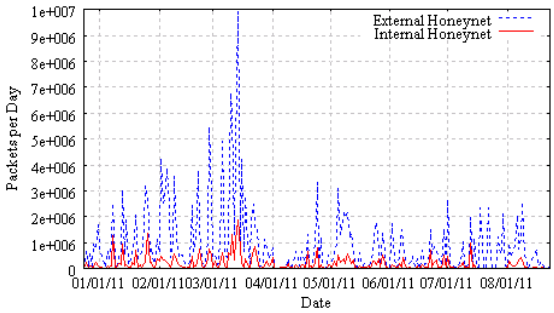

Figure 3. Number of Inbound Packets

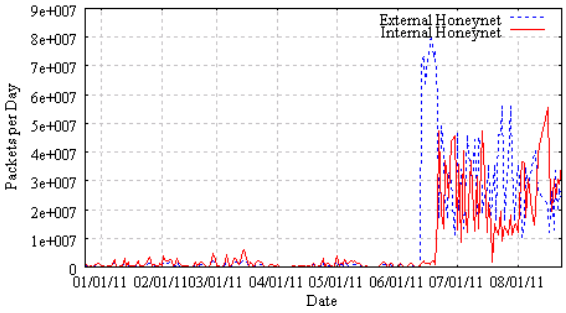

Figure 4. Number of Outbound Packets

the packets on the other ports were used to scan ports that operate vulnerable services.

B. Outbound Traffic Analysis

Fig. 4 shows the number of outbound packets from the Honeynets. In the Honeypots, outbound packets are not generally created except for packets by operating system, such as DNS queries, because no application services have been operated. In the external Honeynet of our university, however, massive traffic occurred dramatically on June 11th and 19th of 2011 after the successful attack on March 16th. 99.98% of this traffic consisted of TCP packets. The destination port ratio of traffic was 80.86% for port 22 (SSH), 19.09% for port 21 (FTP), 0.05% for miscellaneous traffic. All of the traffic on port 21 and 22 was bound for IP addresses owned by a large ISP in Japan to perform sequential port scanning, which determines what systems operate SSH and/or FTP services. In the internal Honeynet, the number of outbound packets increased rapidly on June 17th. These packets were also used to scan port 21 and 22 of the Honeypots.

After finishing port scanning in the external Honeynet, the number of outbound packets again increased sharply. This traffic was used to perform SSH brute-force attacks on the ISP in Japan. The attacker attempted to perform brute-force attacks for distributing Bot agents using FTP service and create many Zombie PCs via SSH service. Attackers attempt SSH brute-force attacks to achieve their various purposes. We could analyze the attacker's purpose that spread malicious Bot agents by collecting and analyzing various kinds of data from the Honeynets. Based on such collected data, we estimate the intrusion factors, such as the number of malicious Bot agents, to forecast DDoS attacks.

V. Concluding Remarks And Future Work

This paper proposed a proactive forecasting system to overcome the limitations of the reactive systems. It also discussed previous studies regarding intrusion forecasting and defined the concept of intrusion forecasting. Additionally, it proposed the architecture for intrusion forecasting system. The Honeynets were deployed in our university to collect the raw data necessary to forecast DDoS attacks and we analyzed Hflow data gathered from the Honeynets as a first step to estimate intrusion factors. As the forecasting method, we have chosen regression analysis based on the result of the previous study that suggested its efficacy for the specific analysis [9] and use it as a single algorithm from the Statistical Analyzer of the Intrusion Forecasting Module. In the future, several forecasting methods with regression analysis should be considered to improve the accuracy of forecasts.

Acknowledgement

This research was supported by World Class University program funded by the Ministry of Education, Science and Technology through the National Research Foundation of Korea (R31-10100), Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0020518), and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012R1A1A2006331).

References

[1] Y. Hideshima and H. Koike, "STARMINE: A visualization system for cyber attacks," 2006 Asian-Pacific Symposium on Information Visualization, pp. 131-138, February 2006.

[2] C. Ishida, Y. Arakawa, I. Sasase, and K. Takemori, "Forecast techniques for predicting increase or decrease of attacks using bayesian inference," 2005 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, pp. 450–453, August 2005.

[3] Y. Zhang, X. Tan, and H. Xi, "A novel approach to network security situation awareness based on multi-perspective analysis," 2007 International Conference on Computational Intelligence and Security, pp. 768–772, December 2007.

[4] D.-H. Kim, T. Lee, S.-O.D. Jung, H.-J. Lee, and H.P. In, "Cyber threat trend analysis model using HMM," 2007 International Symposium on Information Assurance and Security, pp. 177-182, August 2007.

[5] S.-H. Kim, S.-J. Shin, H.-W. Kim, K.-H. Kwon, and Y.-G. Han, "Hybrid intrusion forecasting framework for early warning system," IEICE TRANS. INF. & SYST., vol. E91-D, no. 5, pp. 1234–1241, May 2008.

[6] K. Takemori, Y. Miyake, C. Ishida, and I. Sasase, "A SOC framework for ISP federation and attack forecast by learning propagation patterns," 2007 IEEE Intelligence and Security Informatics, pp. 172–179, May 2007.

[7] S.S.S. Sindhu, S. Geetha, S.S. Sivanath, and A. Kannan, "A neuro-genetic ensemble short term forecasting framework for anomaly intrusion prediction," 2006 International Conference on Advanced Computing and Communications, pp. 187–190, December 2006.

[8] S. Nanda and N. Deo, "A highly scalable model for network attack identification and path prediction," 2007 IEEE Southeast Conference, pp. 663-668, March 2007.

[9] S.E. Schechter, "Toward econometric models of the security risk from remote attacks," IEEE Security & Privacy, vol. 3, issue 1, pp. 40–44, January-February 2005.