Computers, Materials & Continua
DOI:10.32604/cmc.2021.015172
Article

Tech Science Press

Network Log-Based SSH Brute-Force Attack Detection Model

Jeonghoon Park[1], Jinsu Kim[1], B. B. Gupta[2] and Namje Park[1,*]

[1]Department of Convergence Information Security, Graduate School, Jeju National University, Jeju, 63243, Korea
[2]Department of Computer Engineering, National Institute of Technology Kurukshetra, Kurukshetra, 136119, India
[*]Corresponding Author: Namje Park. Email: namjepark@jejunu.ac.kr
Received: 09 November 2020; Accepted: 15 February 2021

Abstract: The rapid advancement of IT technology has enabled the quick discovery, sharing and collection of quality information, but has also increased cyberattacks at a fast pace at the same time. There exists no means to block these cyberattacks completely, and all security policies need to consider the possibility of external attacks. Therefore, it is crucial to reduce external attacks through preventative measures. In general, since routers located in the upper part of a firewall can hardly be protected by security systems, they are exposed to numerous unblocked cyberattacks. Routers block unnecessary services and accept necessary ones while taking appropriate measures to reduce vulnerability, block unauthorized access, and generate relevant logs. Most logs created through unauthorized access are caused by SSH brute-force attacks, and therefore IP data of the attack can be collected through the logs. This paper proposes a model to detect SSH brute-force attacks through their logs, collect their IP address, and control access from that IP address. In this paper, we present a model that extracts and fragments the specific data required from the packets of collected routers in order to detect indiscriminate SSH input attacks. To do so, the model multiplies a user's access records in each packet by weights and adds them to the blacklist according to a final calculated result value. In addition, the model can specify the internal IP of an attack attempt and defend against the first 29 destination IP addresses attempting the attack.

Keywords: SSH brute-force attack; ELK Stack; IT infra; log; access control

1 Introduction

In recent years, IT technology has been making rapid progress at an unprecedented pace. This has enabled us to quickly find, collect, and share quality data; but it has also brought to us a growing number of cyberattacks that snatch and forge data during data communication [1–4]. Cyberattacks come in a variety of forms. In a general attack, data are first collected from the target, and this is followed by invasion into a system based on the collected data [5–8]. Continuous intrusions into the system and attacks against other systems connected to the target system take place [9–11]. Numerous security systems are being installed on IT infrastructure to decrease the number of cyberattacks to an allowable level and protect important data assets through intensified

security. Security systems are built redundantly with multiple devices, such as IPS (Intrusion Preventing System), to more comprehensively analyze the traffic authorized by the firewall, which serves the most basic role in security [12–15]. Security systems are installed in the user's PC with security software in case security accidents occur [16–19]. Despite the extensive security systems of IT infrastructure, Internet routers located close to the Internet are often not protected by a separate device. Routers connect different networks, and designate and transfers optimal paths of packets [20–22]. Through ICMP or IP Scanning, router or network information can be easily identified, based on which malicious attacks may occur [23–25]. Some cyberattacks, such as worms, viruses, and distributed denial of service (DDos), are mainly aimed to destroy the systems on a network while other cyberattacks pretend to be normal packets and persistently attempt intrusion into a certain system [26–29]. In particular, attack systems that evolve along with increasingly enhanced security systems are becoming more intelligent, using machine learning and deep learning to carry out more sophisticated attacks [30–33]. Deep learning (DL) is one of the technologies most frequently applied to defend against these network attacks. By adding the concept of learning to existing methods, research continues to be carried out to collect and defend a variety of information, such as the record of packet flows or access, and the usage patterns of access requesters. In particular, blacklists organized for collected IPs are generally based on collected logs, and studies have been conducted on reducing incorrect blacklist settings and controlling access through learning IP data for blacklists [34–38]. This research proposes a model to detect SSH brute-force attacks against the Internet router through logs, collect their IP address, and control access from that IP address.

2 Related Research

2.1 SSH (Secure Shell)

Known as one of the most important remote access protocols, Secure Shell (SSH) logs into another computer through communications or executes commands to a system, copies files, and performs various functions [28,29]. In existing application programs and protocols without an encryption process, such as Telnet using plain-text communication, there is a high possibility that account information and execution commands are extorted. In contrast, SSH ensures safe management and confidentiality of communications by supporting remote system communications through strong authentication and encryption [30].

In general, network systems should be directly connected using a console in the initial setting, but they are used by supporting remote access through SSH. While the user who remotely accesses the system allowing SSH service is typically a manager, it cannot be assumed that all access requests come from managers. Thus, permission of remote access requests from all origins may expose a system to threats, and it is important to allow manager IPs to access SSH only remotely and establish appropriate policies for the start and end points of the firewall [31].

2.2 Brute-Force Attack

In brute-force attacks, the attacker submits all possible values as account inputs and attempts to access the system's account information. Methods employed in brute-force attacks are divided into dictionary attack methods, which try all strings in a pre-arranged listing, and random sequence methods, which try all possible string combinations in a sequence [32]. A common way to defend against brute-force attacks is to control access when wrong passwords are entered more than a certain number of times. In the case of servers, the threshold for login failure is set through the account lock policy, and access to the account is restricted when the threshold is exceeded [33].

Additionally, the complexity of the password, password usage duration, and minimum password length are set in the password policy to protect the system and its account from being snatched due to a simple password. While network systems do not block access following login failures exceeding the threshold, they do prevent unauthorized access through the ACL. Thus, Internet routers have logs on many attempts of unauthorized external access, and these can serve as a valuable reference for the access control policy.

2.3 Related Research Trend

Controlling accessing IPs by a blacklist prepared through security log analysis is useful to defend against random attacks. This section provides various study cases in which attackers are specified through a blacklist.

Dooyong mentioned that a blacklist of IP addresses is an important element for IT system protection. While various aspects of data and their operation records must be thoroughly reviewed for the blacklisting of IP addresses, the majority of current IP blacklists rely on security monitoring by skilled experts in specific domains. To solve this problem, he designed and established a blacklist model adopting machine learning (ML), and arranged and sorted out data through logistic regression analysis and a random forest; this resulted in incorrectly set blacklists being decreased by 90% [34].

Meanwhile, Dmytro warned that while blacklists of IP addresses and domains exist as components for various security systems, it is difficult for them to be applied to certain risks such as zero-day attacks. He estimated the accuracy of the model's prediction by crossing IPs in the blacklists with a blacklist dataset containing about 270,200 unique blacklisted IPs. His method was more effective than another recent blacklist prediction method [35].

Meanwhile, Rick emphasized the risk of brute-force attacks against web applications and their damage, and noted that the detection of attacks was based on the server's log file analysis, host-based intrusion detection system, or firewall, indicating several relevant problems. Further, he investigated the feasibility of a network-based monitoring approach that detects brute-force attacks against web applications in an encrypted environment and their damage. Afterward, he analyzed brute-force attacks through histograms on data packet payload sizes based on IP Flow Information Export (IPFIX) [36].

3 Structure of the Proposed SSH Brute-Force Attack Detection Model

3.1 Setting a SSH Brute-Force Attack Detection Model

This research designed an access control policy to detect IPs approaching a system maliciously through the logs generated from a router, one of the devices comprising IT infrastructure, and prepared a blacklist of IPs with higher risk levels to protect the system from malicious access.

A blacklist refers to a specific problematic IP group and to the technology applied to block specific IPs. Fig. 1 shows the proposed SSH brute-force attack detection model.

To identify IP groups suspected of accessing the system through a brute-force attack, we needed to obtain SSH brute-force attack logs from the router, and fragment and analyze their content. The IPs that had a high frequency of attacks were listed to be used in the blocking policy of routers and other security equipment. The detection model proposed in this paper aims to collect logs, fragment the collected logs to analyze high-risk access IPs, detect SSH brute-force attacks, and defend against SSH brute-force attacks by managing the detected attackers' IPs using blacklist techniques. If IPs that attempt malicious access can be obtained continuously, they can

be used to defend against future brute-force, thereby enhancing the security of IT infrastructure. In addition, the proposed method can block malicious attacks while preventing channels from concealing their IP address.

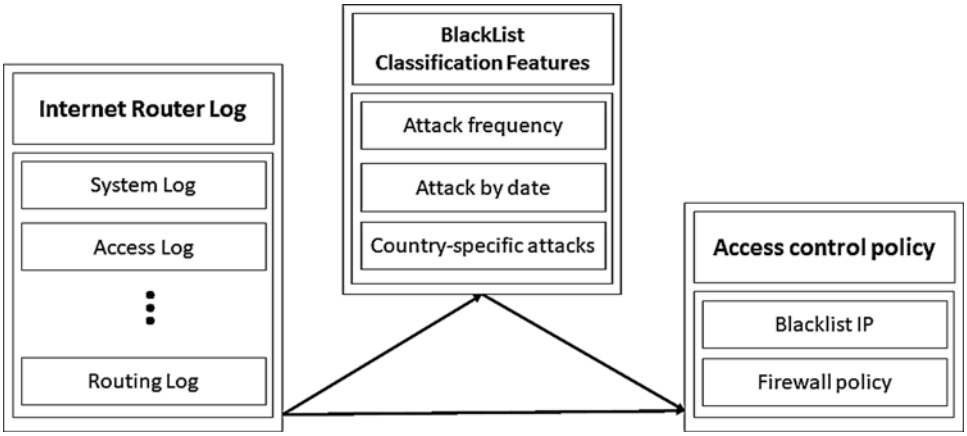

Figure 1: SSH brute-force attack detection model

3.2 Research Hypothesis

To develop an access control policy based on logs from routers, this study established hypotheses based on the research model and verified the implementation of the proposed SSH brute-force attack detection model by testing each hypothesis. The research hypotheses for the proposed model are as follows:

(1) Identification of continuous and malicious access to the Internet router
(2) Specification and detection of unauthorized access
(3) Fragmentation of unauthorized attack logs

First, a hypothesis was established to verify whether there is continuous and malicious access to the router and attacks against the target system. Then, to collect attack logs by identifying unauthorized access from logs on attacks against the router, it was assumed that unauthorized access can be specified and detected by confirming the points of an excessively huge number of access through SSH. The last hypothesis assumes that data to be used for a blacklist can be extracted by removing repeated messages from repeated logs and fragmenting necessary data only.

3.2.1 Identification of Continuous and Malicious Access to the Internet Router

The research subject is a router applied without a separate system, since it is connected to a network before security equipment is installed in a fail-safe way. Routers are equipped with no separate network or access point to ensure connection with the external network. Thus, if there is no change in the network, fewer logs will be generated, and the administrator's access will be recorded as logs. Tab. 1 shows the number of logs generated on the router over a year, separated by months.

Table 1: Number of internet router logs generated

Classification	19/4	19/5	19/6	19/7	19/8	19/9	19/10	19/11	19/12	20/1	20/2	20/3
R1	6187	4856	9988	6864	3780	5763	8592	15464	16912	52218	18141	6776
R2	6226	4387	11160	6055	3339	6235	8731	15318	16942	50312	18454	6966

Fig. 2 shows the numbers of logs generated in the target system, and the figures in the table cannot be obtained only with system logs. In addition, the host name was changed to router-1 to show the logs for the Internet router and to infer ownership of the equipment. Since IP addresses are sensitive data, we have reduced the likelihood of privacy exposure by replacing them with *. Routers guarantee the availability of systems between enterprises and institutions, and they do not receive frequent maintenance. There is no special work except for changes to routing tables, ISP and circuits, and transfer equipment. However, there must be a reason for the huge number of logs generated, and most logs generated are repeated ones.

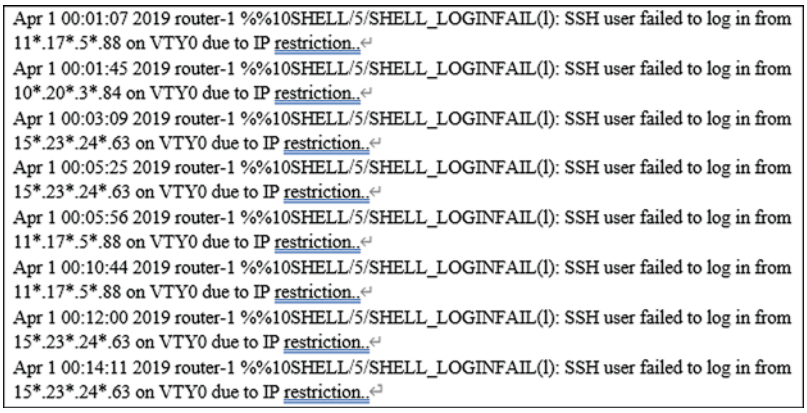

Figure 2: Sample of internet router logs

The analysis results of the Internet router logs produced are as follows: "SSH user failed to login from" means that SSH access was attempted but failed, and whoever requested access is displayed via IP information. "on VTY0 due to IP restriction" means that access failure occurred due to IP access restriction; in other words, the SSH user of the starting point IP information was not able to log into VTY0 due to IP restriction, where VTY refers to a virtual terminal line to access the router interface.

Fig. 3 shows the one-year analysis results of all logs generated from the two routers used in the experiment, system logs generated through normal access, logs generated by unauthorized access attempts, and the access rate of unauthorized access logs to all logs. In fact, most of the logs are from unauthorized access, which indicates that there are continued attempts to access the routers. Vulnerabilities of Internet routers, including account management, log management, and function management, are improved through the inspection of the vulnerabilities of important information infrastructure. As for access to VTY by the administrator, only registered persons designated by the ACL (Access-List) are allowed, and other unauthorized access is blocked to prevent unidentified access. In other words, brute-force attacks that are attempted countless times

generate too many logs on unauthorized access. A comparison of unauthorized access logs with all types of logs, including normal attempts, is shown in Fig. 3.

Classification	Internet router-1				Internet router-2			
	No. of all logs	System logs	Unauthorized access	Unauthorized access rate	No. of all logs	System logs	Unauthorized access	Unauthorized access rate
2019/04	6187	37	6150	99.40%	6226	33	6193	99.47%
2019/05	4856	186	4670	96.17%	4387	131	4256	97.01%
2019/06	9988	46	9942	99.44%	11160	30	11130	99.73%
2019/07	6864	164	6700	97.61%	6055	66	5989	98.91%
2019/08	3780	562	3218	85.13%	3339	30	3309	99.10%
2019/09	5763	30	5733	99.48%	6235	29	6206	99.53%
2019/10	8592	32	8560	99.63%	8731	30	8701	99.66%
2019/11	15464	45	15419	99.71%	15318	30	15288	99.80%
2019/12	16912	29	16883	99.83%	16942	29	16913	99.83%
2020/01	52218	30	52118	99.94%	50312	30	50282	99.94%
2020/02	18141	6	18135	99.97%	18454	38	18416	99.79%
2020/03	6776	86	6690	98.73%	6966	35	6931	99.50%

Figure 3: Analysis of internet router attack logs

3.2.2 Specification and Detection of Unauthorized Access

Unauthorized access to the Internet router is normal access through SSH but is not authorized access on the target system. The Internet router only allows SSH access through the registered administrator IP, blocks all other accesses, and generates relevant logs. These logs take up a large portion of the entire logs, and unauthorized SSH access attempts are a result of brute-force attack attempts to identify the account information of the target system's Internet router.

Fig. 4 shows the top 10 IPs that attempted unauthorized access to the target systems used for research as of April 2019. Out of a total of 6,187 logs, 3,171 logs were generated by IP "A," taking up 51.25%. The figures indicate that it is difficult to consider that normal access is attempted, or values are entered manually. In addition, access was retried over very short intervals, and the regular frequencies and intervals reveal that the access was attempted using an automated tool.

Fig. 5 shows, in ascending order, 1.2 million attempts extracted by one of the top 10 IPs at unauthorized access to IPsenf determine the frequency. Automatic access by scripts or programs is also suspected as these IPs have attempted to connect sequentially for several days and with very a short time intervals (less than one second). In addition, the above picture shows that there are 12 attempts to invade the area on the same day, depending on the time zone.

Since unauthorized access by IPs to the Internet router can be specified through the number of access attempts or access log messages and accounts for most of the logs, these logs can be used as security logs. Furthermore, these access attempts fall under SSH brute-force attacks and their IP information can be collected through logs.

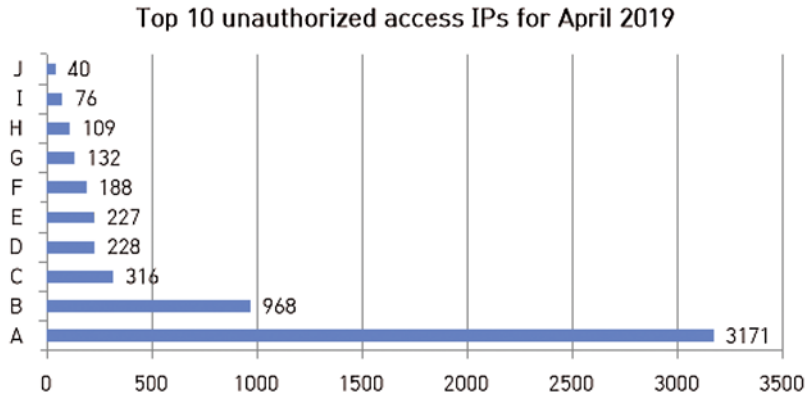

Figure 4: Number of unauthorized access by IPs

No.	Day	Time	Day	Time	Day	Time	Day	Time	Day	Time
1	9	4:40:00	10	12:18:37	11	23:03:41	12	15:39:57	15	11:10:37
2	9	4:40:01	10	12:18:37	11	23:03:41	12	15:39:57	15	11:10:38
3	9	4:40:02	10	12:18:38	11	23:03:42	12	15:39:57	15	11:10:39
4	9	4:40:02	10	12:18:38	11	23:03:42	12	15:39:57	15	11:10:39
5	9	4:40:03	10	12:18:38	11	23:03:42	12	15:39:57	15	11:10:39
6	9	4:40:03	10	12:18:38	11	23:03:42	12	15:39:58	15	11:10:39
7	9	4:40:03	10	12:18:39	11	23:03:42	12	15:39:58	15	11:10:39
8	9	4:40:04	10	12:18:42	11	23:03:42	12	15:39:58	15	11:10:40
9	9	4:40:04	10	12:18:44	11	23:03:43	12	15:39:58	15	11:10:40
10	9	4:40:04	10	12:18:44	11	23:03:43	12	15:39:58	15	11:10:40
11	9	4:40:04	10	12:18:45	11	23:03:43	12	15:39:59	15	11:10:41
12	9	4:40:05	10	12:18:45	11	23:03:43	12	15:39:59	15	11:10:41
13	9	10:33:08	11	2:28:40	12	7:24:22	13	0:18:41	15	11:20:38
14	9	10:33:08	11	2:28:42	12	7:24:22	13	0:18:41	15	11:20:38
15	9	10:33:08	11	2:28:43	12	7:24:23	13	0:18:41	15	11:20:38
16	9	10:33:08	11	2:28:44	12	7:24:24	13	0:18:41	15	11:20:38
17	9	10:33:09	11	2:28:44	12	7:24:24	13	0:18:41	15	11:20:39
18	9	10:33:09	11	2:28:44	12	7:24:24	13	0:18:41	15	11:20:39
19	9	10:33:09	11	2:28:44	12	7:24:24	13	0:18:42	15	11:20:40
20	9	10:33:09	11	2:28:45	12	7:24:25	13	0:18:42	15	11:20:41
21	9	10:33:10	11	2:28:45	12	7:24:25	13	0:18:42	15	11:20:42
22	9	10:33:10	11	2:28:45	12	7:24:25	13	0:18:42	15	11:20:42
23	9	10:33:10	11	2:28:45	12	7:24:26	13	0:18:42	15	11:20:42
24	9	10:33:12	11	2:28:45	12	7:24:26	13	0:18:42	15	11:20:43

Figure 5: IP access attempt information sample

3.2.3 Fragmentation of Logs Associated with Unauthorized Access

Logs of SSH brute-force attacks collected in the Internet router are shown as follows:

> "Apr 1 00:01:07 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 11*.17*.5*.88 on VTY0 due to IP restriction."

The above log can be analyzed as follows: First, "Apr 1 00:01:07 2019" refers to the month, day, time, and year, respectively; "router-1" is a hostname; and the log shows SSH access failure analyzed previously. This log can be rearranged as a construction consisting of month, day, time, year, hostname, and message. The log is generated with the hostname and time information (month, day, time, and year), accurately verifying the content of the log. In other words, the log reveals when and to which equipment the SSH access has been attempted. The part excluding the IP is repeated in the message, whose log length can be reduced and the construction can be fragmented by deleting the repeated part. The above log can be reduced to "Apr 1 00:01:07 2019 router-1 11*.17*.5*.88", and it can be saved in various file formats. Tab. 2 shows the fragmentation of a log of SSH brute-force attacks. The log includes month, day, time, year, hostname, and IP data.

Table 2: Fragmentation of log data of SSH brute-force attacks

Month	Day	Time	Year	Hostname	IP
Apr	1	0:01:07	2019	Router-1	11*.17*.5*.88
Apr	1	0:01:45	2019	Router-1	10*.20*.3*.84
Apr	1	0:03:09	2019	Router-1	15*.23*.24*.63
Apr	1	0:05:25	2019	Router-1	15*.23*.24*.63
Apr	1	0:05:56	2019	Router-1	11*.17*.5*.88
Apr	1	0:10:44	2019	Router-1	11*.17*.5*.88
Apr	1	0:12:00	2019	Router-1	15*.23*.24*.63
Apr	1	0:14:11	2019	Router-1	15*.23*.24*.63
Apr	1	0:15:33	2019	Router-1	11*.17*.5*.88

The fragmented data of the log can be exported into a CSV file, through which the number of access attempts by time, day, year, and IP can be easily identified. Moreover, by adding a country field and registering the country code associated with each IP, the country from which attacks were attempted can be specified.

3.3 Design of the Proposed SSH Brute-Force Attack Detection Model

We study a blacklist access control policy against SSH brute-force attacks by analyzing source logs from the Internet router, extracting logs of SSH brute-force attacks only, and fragmenting them. This requires the extraction of logs related to SSH brute-force attacks caused by unauthorized access from source logs. Source logs are arranged in a text file, and a function is applied to classify them into two types: SSH brute-force attack logs, and other logs. Then, repeated strings are eliminated from the SSH brute-force attack logs and only the necessary data are left, which are then exported to a CSV file. IPs to be included in the blacklist are identified and used for the ACL of the Internet router and the blocking policy of security equipment.

3.3.1 Processing of Source Logs

A log document composed of one row and as many columns as the number of logs generated can be utilized to detect SSH brute-force attacks. To do this, a filter is used to fragment the log content. To process logs, the initial log information is first classified by applying two functions: (1) the FIND function, which returns the location of a starting point of the cell that includes the finding value; and (2) the ISNUMBER function, which returns the resulting value according to the result of a formula. The FIND function is used to determine the starting point of a message in a log, while the ISNUMBER function is applied to express TRUE or FALSE when a message related to SSH brute-force attacks is detected. These enable the extraction of logs only of brute-force attacks. Strings indicating SSH brute-force attacks are "SSH user failed to log in from" and "on VTY0 due to IP restriction." Only when the ISNUMBER function is applied to the two strings and they are found on a log is the result TRUE. Fig. 6 outlines log processing.

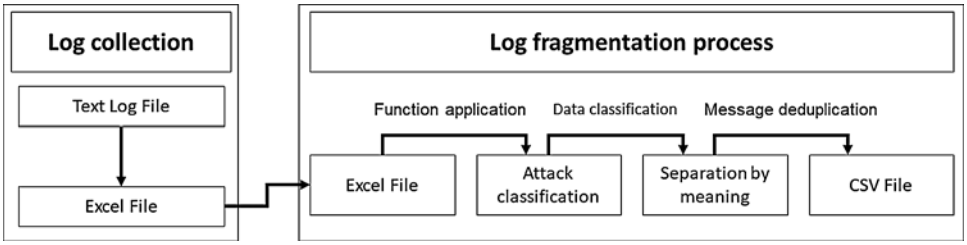

Figure 6: Log processing steps

Once the raw log information is determined as either TRUE or FALSE through the aforementioned functions, the logs related to SSH brute-force attacks are extracted through filtering. At least 3,000 and up to 5,000 logs are generated from the Internet router in a month. Extracting the system logs from a large number of logs can be cumbersome as only specific logs must be extracted. However, when extracting only the logs not related to SSH brute-force attacks, the number of logs extracted is reduced exponentially; this allows users to identify system logs more easily.

Tab. 3 shows an example of the results that can be obtained when each function is applied to the raw log. This confirms that the log returns TRUE if it includes "SSH user failed to log in from" and "on VTY0 due to IP restriction."

Table 3: Results of applying the functions

	Raw log information
FALSE	Apr 23 12:13:48 2019 router-1 %%10SSH/4/TrapLogoff(t): 1.3.6.1.4.1.25506.2.22.1.3.0.4 SSH user logoff trap information
TRUE	Apr 23 12:38:26 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 5*.21*.12*.66 on VTY0 due to IP restriction.
TRUE	Apr 23 12:50:28 2019 router-1 %%10SHELL/5/SHELL_LOGINFAIL(l): SSH user failed to log in from 6*.24*.20*.20 on VTY0 due to IP restriction.

3.3.2 *Fragmentation of Logs*

The fragmentation process for the extracted logs of SSH brute-force attacks is simple because the pattern of all logs is set and the row can be classified by meaning. In addition, log information can be fragmented by eliminating the two phrases referring to the SSH brute-force attacks, "SSH user failed to log in from" and "on VTY0 due to IP restriction" and extracting the IP data. Fig. 7 illustrates the process of extraction and fragmentation of log information through log analysis.

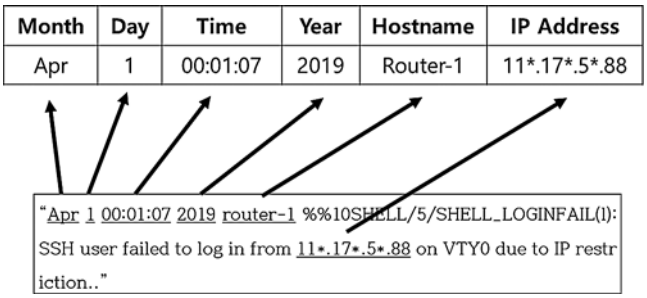

Figure 7: CVS file example

3.3.3 *Log Analysis*

Information for analysis undergo classification and fragmentation, and be exported to a CSV file. Log analysis is performed by dividing the log content into month, day, time, year, hostname, and IP address. Entered logs are sequentially sorted by a certain column, which is divided into sub-columns for the next entry. Fig. 8 shows the breakdown of a log and a summary of analysis results.

Month	Day	Time	Year	Hostname	IP Address	Count
Apr	1	00:01:07	2019	Router-1	11*.17*.5*.88	x

Figure 8: Fragmented log

If a is a log fragmented as above, the month field extracts the Apr column and generates a log counter, and then the Apr column is divided by day to obtain the number of logs generated per day. The count increases by 1 when the same IP is found among the logs per month, and allows us to determine the number of access requests made over an entire month. Among the classified logs, the IP address of 11*.17*.588, was extracted and found to account for 51.87% of the total logs. In addition, collected IP addresses can be used to identify access requests. Most importantly, the number of access attempts can be identified by accessing IPs and related maliciousness. When the counts are sorted by day for the IPs with high risk of multiple intrusions, attack patterns by date can also be analyzed.

While SSH brute-force attacks occurred with high frequency, the analysis of attacking patterns by date confirmed that the attacks took place over a few days rather than in a single day. These attacks are more malicious than those occurring in a single day and must be blocked since they occur continuously.

3.3.4 IP Blacklist

IPs carrying out SSH brute-force attacks can all be assumed to be malicious hosts; however, designating and blocking all attacking IPs as blacklisted IPs may be an ineffective policy. Some of the attacking IPs can be a one-time attack and have no intention to attack anymore. Thus, the frequency of attack is an important factor to consider in proving the intention of an attack. The number of attacks per day is also a significant factor that determines the continuity of attacks. A time-based classification can provide more details, such as whether the attacks occur during business hours or over holidays, and the country from which the attack originates can be determined by the IP. Tab. 4 lists the characteristics of malicious hosts on the Internet router, and attack strength, which is the blocking frequency by IP, is the most important indicator in selecting blacklisted IPs.

Table 4: Characteristics for selecting blacklisted IPs

Division	Characteristic
IP identification	SSH brute-force attack
Attack strength	Blocking frequency by IP
Attack timing	Number of days of detection
Attack position	Country/International

When selecting blacklisted IPs, those with one-time attacks should be excluded. A total of 419 IP addresses were involved in the attacks against the router. The IP with the highest frequency carried out 3,171 attacks, demonstrating its intention of attack. For ambiguous one-time attacks, it was difficult to determine whether 405 IPs, i.e., 96% of the total IPs, had any intent to attack further since they carried out less than 10 attacks; thus, these were not classified as blacklisted IPs. This made the analysis faster. Other parameters can be calculated to establish the allowable criteria to determine blacklisted IPs. Three characteristics of attack are defined as x, and whether an IP is classified as blacklisted is y. Formula 1 is used to identify a blacklisted IP.

$$y(x) = x1 * x2 * x3 \tag{1}$$

Here, x1 is the frequency of attacks. n attacks ($0 < n$) are multiplied by 0.01, and thus 100 or more attacks are represented as 1. x1, which indicates attack frequency, can be calculated as follows:

$$x1 = n * 0.01 \tag{2}$$

x2 denotes the number of days detected, and is m ($0 < m \leq 31$). m is multiplied by 0.1, and thus it weighs 0.1 per day. x2 can be calculated as follows:

$$x2 = m * 0.1 \tag{3}$$

x3 denotes the location of attack, which is either a specific country or international. The weighted values are 1 and 1.1 for country and international, respectively. x3 can be calculated as follows:

$$x3 = \begin{cases} if) & Country = 1 \\ if) & International = 1.1 \end{cases} \tag{4}$$

To illustrate the use of these formulas, we assumed that 100 SSH brute-force attacks were made against the Internet router per day within a country. The relevant formula is (100*0.01)*(1*0.1)*1, and the value of 0.1 is returned. For an IP to be determined as a blacklisted IP, it must meet the requirements of at least 100 attacks, 2 or more detection days, and an international location. In the Tab. 5, having a value of 0.22 or higher means the IP will be blacklisted IP; when the value is between 0.1 and 0.22, the IP is classified as "suspect." Tab. 5 shows the process and results of determining blacklisted IPs.

Table 5: Example of determining blacklisted IPs

IP	Attack frequency (x1 ∗ 0.01)	Detection days (x2 ∗ 0.1)	Location (Country 1/ International 1.1)	Blacklist rating	Blacklist status
a	1	0.2	1.1	0.22	Blacklist
b	1.32	0.1	1	0.132	Suspect
c	8	0.1	1	0.8	Blacklist
d	0.9	0.7	1.1	0.69	Blacklist

3.3.5 Verification of the Proposed SSH Brute-Force Attack Detection Model

This research aims to identify SSH brute-force attacks against an Internet router through logs, select blacklisted IPs based on the analysis of attack characteristics, and establish an access control policy using the blacklisted IPs. Three research hypotheses on continuous and malicious access to the Internet router, specification of unauthorized access, and the fragmentation of logs of unauthorized access were set up and verified to establish access control policies based on logs.

Regarding the first hypothesis that there is continuous and malicious access to the Internet router, we revealed that more than 90% of logs were of unauthorized access. Attacks by unauthorized access generated logs that had the same message pattern, including "SSH user failed to log in from [attacker's IP address] on VTY0 due to IP restriction." By extracting relevant logs, unauthorized access was identified (which is associated with the second hypothesis) and the SSH brute-force attacks were detected. Lastly, the repeated message content was eliminated, and logs were fragmented by month, day, time, and year, which confirmed that the third hypothesis of log fragmentation was true. We digitized the characteristics of fragmented attack logs to detect blacklisted IPs and finally proposed a blacklisted IP identification model that can determine whether an IP should be blacklisted.

Attack frequency, number of detection days, and country from which the attack originated were used to determine blacklisted IPs. The weighted values of 0.01, 0.1, and 1 (country) or 1.1 (international) were set for attack frequency, detection days, and location, respectively. IPs with 100 or more attacks and 2 or more detection days and an overseas origin were classified as blacklisted IPs. Tab. 6 shows the number of access control policies set according to the number of blacklisted IPs with malicious intention by month.

We imposed a firewall to block blacklisted IPs attempting to access the internal traffic of destination. For 46 days from May when the firewall policy was established through June, a total

of 10,147 IPs were blocked by the firewall through the CIDR blocking a single IP or processing the bands of multiple IPs. We found that 29 destination IPs were Chinese IPs with access requests.

Table 6: Establishment of access control policies in accordance with blacklist

Month	No. of policies	Month	No. of policies	Month	No. of policies
2019/04	7	2019/08	4	2019/12	22
2019/05	24	2019/09	14	2020/01	27
2019/06	18	2019/10	19	2020/02	33
2019/07	11	2019/11	18	2020/03	19

Tab. 7 lists the number of times each internal IP was blocked by the firewall. IPs "A" and "B" showed that more than 90% of the entire blocking was normal traffic, but they were also blocked since their destination address was in the same band as that of a blacklisted IP. This may cause further inconvenience by requiring the CIDR processing of internal IPs.

Table 7: Logs blocked by firewall

Internal IP	Number of intercepts	Blocking specific gravity (%)	Internal IP	Number of policies	Relative height (%)
A IP	5,466	53.87	E IP	45	0.44
B IP	4,236	41.75	F IP	30	0.30
C IP	285	2.81	G IP	30	0.30
D IP	54	0.53	H IP	1	0.01

4 Conclusions

The advancement of infrastructure has increased the necessity to collect a significant amount of personal data for user convenience, driving IT infrastructure managers to encounter many important information assets and systems. However, there is no standardized method for analyzing the data generated during system management, which thus tends to rely on the manager's experience. While skilled managers can solve a problem based on their experience, most managers take a long time to determine the meaning of each log and the cause of the error from a huge number of logs generated by a system. Systems designed to manage important information assets require timely action against errors and appropriate prevention, but the general process alone is not enough to solve the error when a quick action is required in a certain case.

This research analyzed logs generated in a router for one year and studied methods for protecting the system through the detection and access control of SSH brute-force attacks against the target system. We confirmed that many of the logs were generated due to SSH brute-force attacks, and then proposed a blacklisted IP determination model by fragmenting and examining the logs. This approach can prevent continuous attacks by detecting access to specific internal IPs through routers that are not normally protected by security devices. The approach can also prevent unauthorized access to internal IPs and attack site IPs by creating a blacklist based on the risk, thereby preventing infection to other systems. However, this method aims to prevent future attacks by analyzing the attacks that have already occurred, and thus it does not have the capability to respond to real-time attacks.

While the logs that are currently generated must be determined faster than the analysis of logs on previous attacks to respond to real-time attacks, in practice, it is difficult for a manager to immediately identify and determine them. To prevent potential attacks, it is necessary to apply a real-time log determination using machine learning in the future by analyzing past logs while conducting real-time analysis on current logs.

Funding Statement: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5C2A04083374).

Conflicts of Interest: We declare that we have no conflicts of interest to report regarding the present study.

References

[1] J. Kim, "The effective monitoring and controlling method through malicious domain-IP traceback," M.S. thesis, Korea University, Korea, 2019.
[2] J. Kim, N. Park, G. Kim and S. Jin, "CCTV video processing metadata security scheme using character order preserving-transformation in the emerging multimedia," *Electronics*, vol. 8, no. 4, pp. 1–15, 2019.
[3] T. Kim, "A study on the efficient security monitoring by using open source ELG stack," M.S. thesis, Sungkyunkwan University, Korea, 2020.
[4] N. Park, "The core competencies of SEL-based innovative creativity education," *Pure and Applied Mathematics*, vol. 118, no. 19, pp. 837–849, 2018.
[5] N. Park, J. Park and H. Kim, "Inter-authentication and session key sharing procedure for secure M2M/IoT environment," *Information*, vol. 18, no. 1, pp. 261–266, 2015.
[6] J. Noh, "Study on precursor phenomena analysis of cyber terrorism and IP traceback," M.S. thesis, Hoseo University, Korea, 2015.
[7] B. Moon, "IoT malware analysis using IoT honeypot," M.S. thesis, Sungkyunkwan University, Korea, 2018.
[8] K. Bhushan and B. B. Gupta, "Distributed denial of service (DDoS) attack mitigation in software defined network (SDN)-based cloud computing environment," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 5, pp. 1985–1997, 2019.
[9] A. Dahiya and B. B. Gupta, "Multi attribute auction based incentivized solution against DDoS attacks," *Computers & Security*, vol. 92, no. 11, 12, pp. 101763, 2020.
[10] D. Lee and N. Park, "Geocasting-based synchronization of almanac on the maritime cloud for distributed smart surveillance," *Supercomputing*, vol. 73, no. 3, pp. 1103–1118, 2017.
[11] S. Moon, "Design and implementation of a network log analysis and prediction system based on the bigdata," Ph.D. dissertation, Wonkwang University, Korea, 2018.
[12] J. Kim and N. Park, "Blockchain-based data-preserving AI learning environment model for AI cybersecurity systems in IoT service environments," *Applied Sciences*, vol. 10, no. 14, pp. 1–12, 2020.
[13] N. Park, B. Kim and J. Kim, "A mechanism of masking identification information regarding moving objects Recorded on visual surveillance systems by differentially implementing access permission," *Electronics*, vol. 8, no. 7, pp. 1–17, 2019.
[14] Y. Yoo, "A study on the collection and utilization of IP addresses used in cybercrime," M.S. thesis, Korea University, Korea, 2015.
[15] J. Lee, S. Kim, J. Woo and C. Park, "Analysis and response of SSH brute force attacks in multi-user computing environment," *Korea Information Processing Society, Transactions on Computer and Communication Systems*, vol. 4, no. 6, pp. 205–212, 2015.
[16] D. Lee and N. Park, "Blockchain based privacy preserving multimedia intelligent video surveillance using secure merkle tree," *Multimedia Tools and Applications*, vol. 76, no. 19, pp. 1–18, 2020.
[17] A. Tewari and B. B. Gupta, "Security, privacy and trust of different layers in internet of things (IoTs) framework," *Future Generation Computer Systems*, vol. 180, no. 4, pp. 909–920, 2020.

[18] A. Quyoom, A. A. Mir and A. Sarwar, "Security attacks and challenges of VANETs: A literature survey," *Journal of Multimedia Information System*, vol. 7, no. 1, pp. 45–54, 2020.

[19] D. Lee and N. Park, "Electronic identity information hiding methods using a secret sharing scheme in multimedia-centric internet of things environment," *Personal and Ubiquitous Computing*, vol. 22, no. 1, pp. 3–10, 2018.

[20] D. Lim, "A study on the new method improving network security utilizing PF," M.S. thesis, Dongguk University, Korea, 2009.

[21] D. Jeon, "Black eye: IP blacklisting via threat feature extraction and machine learning from security logs," M.A. thesis, Kyungpook National University, Korea, 2019.

[22] J. Kim and N. Park, "Lightweight knowledge-based authentication model for intelligent closed circuit television in mobile personal computing," *Personal and Ubiquitous Computing*, vol. 38, no. 1, pp. 1–9, 2019.

[23] N. Park, H. Hu and Q. Jin, "Security and privacy mechanisms for sensor middleware and application in internet of things (IoT)," *International Journal of Distributed Sensor Networks*, vol. 12, no. 1, pp. 1–3, 2016.

[24] S. Joo, "A study on the security improvement of network infrastructure using the information protection management system (ISMS)," M.S. thesis, Konkuk University, Korea, 2018.

[25] J. Choi, "Attack type revealed by direct analysis of actual large-capacity attack log," M.S. thesis, Seoul National University, Korea, 2015.

[26] D. Lee, N. Park, G. Kim and S. Jin, "De-identification of metering data for smart grid personal security in intelligent CCTV-based P2P cloud computing environment," *Peer-to-Peer Networking and Applications*, vol. 11, no. 6, pp. 1299–1308, 2018.

[27] S. Han, "A study on enhance information security through the improvement of network infrastructure security management," M.S. thesis, Dongguk University, Korea, 2016.

[28] J. Hyun, "Security operation implementation through big data analysis by using open source ELK Stack," M.S. thesis, Korea University, Korea, 2018.

[29] N. Park and H. Bang, "Mobile middleware platform for secure vessel traffic system in IoT service environment," *Security and Communication Networks*, vol. 9, no. 6, pp. 500–512, 2014.

[30] S. Hong, "A study on the detection and visibility of cyber threats on window systems using sysmon and ELK stack," M.S. thesis, Dongguk University, Korea, 2020.

[31] J. Lee, S. Kim, C. Park, T. Hong and H. Chae, "Heavy-tailed distribution of the SSH brute-force attack duration in a multi-user environment," *Journal of the Korean Physical Society*, vol. 69, no. 2, pp. 253–258, 2016.

[32] N. Park, J. Kwak, S. Kim, D. Won and H. Kim, "WIPI mobile platform with secure service for mobile RFID network environment," in *Proc. APWeb*, Harbin, China, pp. 741–748, 2015.

[33] N. Park and N. Kang, "Mutual authentication scheme in secure internet of things technology for comfortable lifestyle," *Sensors*, vol. 16, no. 1, pp. 1–16, 2015.

[34] M. M. Najafabadi, T. M. Khoshgoftaar, C. Calvert and C. Kemp, "Detection of SSH brute force attacks using aggregated netflow data," in *Proc. ICMLA*, Miami, FL, USA, pp. 283–288, 2015.

[35] M. M. Najafabadi, T. M. Khoshgoftaar, C. Kemp, N. Seliya and R. Zuech, "Machine learning for detecting brute force attacks at the network level," in *Proc. BIBE*, Boca Raton, FL, USA, pp. 379–385, 2014.

[36] D. Likhomanov and V. Poliukh, "Predicting malicious hosts by blacklisted IPv4 address density estimation," in *Proc. DESSERT*, Kyiv, Ukraine, pp. 741–748, 2015.

[37] R. Hofstede, M. Jonker, A. Sperotto and A. Pras, "Flow-based web application brute-force attack and compromise detection," *Network and Systems Management*, vol. 25, no. 4, pp. 735–758, 2017.

[38] J. Park, "Detection and Access Control of SSH Brute-Force Attack Using Internet Router Logs," M.S. thesis, Jeju National University, Korea, 2020.