Traffic and Overhead Analysis of Applied Pre-filtering ACL Firewall on HPC Service Network

Jae-Kook Lee, Taeyoung Hong, and Guohua Li

Abstract: **In an high-performance computing (HPC), supercomputing service environment, the security of infrastructure nodes that are points of contact for researchers is very important. We have applied various security devices such as anti-DDoS, IPS, firewall, web application firewall, and etc. on an HPC service network to provide more secure supercomputing services. Firewalls are a common and essential element of network security devices with the ability to block network traffic according to pre-defined rules. With the increasing demands for services, cyberattacks, as well as overheads on firewall policies have also increased. To reduce this overhead, in our previous research, we analyzed dropped packets log and performed a method on the firewall as Abnormal IP that can detect and deny anomalous IPs in real-time. As the number of abnormal IPs increased, the performance of the firewall significantly deteriorated. To solve this problem, we applied access control list (ACL) at the front-end of the firewall to perform pre-filtering, thereby improving the performance of the firewall on the HPC service network. This research is expected to contribute as a preliminary study in the HPC field by deriving pre-filtering ACL to reduce the CPU load of firewall server by showing the result of about 21.5% improvement in performance.**

Index Terms: **Network performance, network security, traffic analysis, traffic overhead.**

I. INTRODUCTION

SUPERCOMPUTERS are used to perform computationally intensive research in the areas of molecular modeling, physical simulation, decryption, geophysical modeling, automotive and aerospace design, and financial modeling. Most computational scientists working in these areas become the users of supercomputers [1]. As the number of users increases, the demand for computational resources grows. In addition, various infrastructure nodes, such as login nodes, web servers, and data mover (DM) nodes, are required when supercomputer users apply for resources or request necessary technical support. As these infrastructure nodes are always at the risk of cyberattacks, network security at supercomputing centers is important. Most supercomputing centers operate network security systems, such as anti distributed denial-of-service (anti-DDoS) systems, intrusion prevention systems (IPS), and firewalls, to protect resources from cyberattacks [2], [3].

We also have deployed anti-DDoS, IPS, and firewall devices and built high available architecture to operate these systems continuously without intervention or failure. In accordance with increasing abnormal behavior every year, we applied additionally solutions to protect infrastructure nodes against cyberattacks. Our proposed solutions can detect and prevent cyberattacks based on the analysis of access logs of infrastructure nodes and event logs of firewall and IPS devices in our HPC service environment [5]. But, as cyberattacks increase, pre-defined rules of firewalls grew with serious overhead. This overhead causes the following problems: (1) Delay when storing and synchronizing rules in the duplication configuration for the firewall; (2) considerable time utilization or increase in the failure rate to restore backup settings.

To solve these problems, in this paper we first analyzed the firewall event logs to trace traffic events. Next, we filtered IP packets by applying network access control methods to the firewall to reduce the overhead caused by the increase in the number of firewall policies. This study is expected to contribute as a preliminary study in the HPC field by applying pre-filtering ACL method to reduce the CPU load of firewall server with the improvement performance result.

The remainder of this paper is structured as follows: In Section II we examine several related works. In Section III, we explain current network security limitations based on the analysis of firewall events. In Section IV, we describe the process of applying a pre-filtering ACL. Section V presents the evaluation of results for the proposed pre-filtering ACL. Finally, in Section VI, we discuss our conclusions and future works.

II. RELATED WORK

Cyberattacks in [4] can be classified into seven types, namely denial of service, eavesdropping, spoofing, user to root (U2R), abuse login, application based attack, and unsecured or suspicious access. To defend against these attacks, we have deployed our network security infrastructure as shown in Table 1 on Korea institute of science and technology information (KISTI) supercomputing service network. For denial of service, we installed anti-DDoS system in network-based IDS type with applied signature-based technique. For eavesdropping, we used one time password (OTP) and encryption protocols (SSH, SFTP, and HTTPS). For spoofing, we used both anti-DDoS and IPS in network-based intrusion detection system (IDS) type with threshold metric or signature-based techniques. For U2R and abuse login, we installed threat management system in host-based IDS type with techniques in papers [5], [6]. For application-based attack, we deployed web

Manuscript received August 20, 2019; revised January 13, 2021; approved for publication by B. Byunghoon Kang, Division III Editor, March 15, 2021.

This research has been performed as a subproject of Project No. K-21-L02-C01-S01 (The National Supercomputing Infrastructure Construction and Service) supported by the Korea Institute of Science and Technology Information (KISTI).

J.-K. Lee, T. Hong, and G. Li are with the Department of Supercomputing Infrastructure Center, KISTI, email: {jklee, tyhong, ghlee}@kisti.re.kr.

G. Li is the corresponding author.

Digital Object Identifier: 10.23919/JCN.2021.000011

application firewall in network-based IDS type with signature-based technique. For unsecured or suspicious access, we built a firewall in network-based IDS with rule-based packet filtering.

Among these cyberattack solutions, firewall which plays an important role can be classified into five types; these are packet filtering firewall, circuit-level gateway, stateful inspection firewall, application-level gateway, and next-generation firewall according to these papers [7], [8]. We have summarized the advantages and disadvantages of these firewall types in Table 2 based on HPC environment. One of the advantages of packet filtering firewall is that it is efficient at processing packets and suitable for high-speed network environment. One of the most typical disadvantages of packet filtering firewall is that it is difficult to securely configure. The circuit-level gateway provides privacy for data passing in/out of private network which is more secure than packet filtering. However, it requires modification to network protocol attack as one of its most typical disadvantages. One of advantages of the stateful inspection firewall is its capability to block many types of denial of service attacks and IPS although it has high processing overhead. Application-level gateway is capable of detecting and blocking the types of attack even not visible at the OSI model network or transport layers but it still has high processing overhead and it is complex to configure and maintain. Next-generation firewall provides traditional firewall combined with IDS/IPS, and in the same way, it needs complex designed architecture and requires high front-end investment in resources to acquire, configure and deploy. Among these firewall types, packet filtering firewall is suitable for high-performance infrastructure environment because it is the best in terms of performance while with security issues [9], [10].

The basic elements of the firewall security policy are known as rules, of which there are usually hundreds [4]. Meanwhile, the number of web hacking incidents has increased since the advent of web services. Due to the characteristics of web services, there are limitations in applying policies to a specific IP or port with firewall devices. Most security centers provide network security services by operating both an IDS and a firewall to cover these limitations [7]. Accurate deployment of firewall policies has a significant impact on the effectiveness of the entire security system. These policies should be established for the types of services of infrastructure resources. We have studied the types of firewalls and design principles [7], which helped us to set up a firewall policy and make rules for our infrastructure. The infrastructure resources are largely divided into the un-trust zone, trust zone, and demilitarized zone (DMZ). Our firewall policy was established by separating these zones. We manage IP addresses by creating an object name, managing objects by creating a group name, and operating a firewall rule with a configurable time. Each rule has its priority, and the higher-priority rules are applied earlier than lower-priority rules.

Firewall policies are generally established based on a black list or white list. A firewall with black-list-based rules drops traffic first according to the IP addresses in the list and passes others (Table 3). In contrast, a firewall with white-list-based rules passes traffic first according to IP addresses in the list and drops others (Table 4). Often, if there are many destination IPs to block, a black-list-based ACL should be used, otherwise a white-list-based ACL is advised [11]–[13]. We apply a black-list-based ACL because most supercomputing service users need access to our HPC resources anytime and anywhere.

The fatal disadvantage of a black-list-based ACL is that as the number of cyberattacks increases, and thus, the number of blacklisted IPs increases, which leads to the degradation of firewall performance [12]. Therefore, we add anti-DDoS security device to distinguish DDoS attacks from cyberattacks, which can reduce the traffic load on the firewall. However, there is still an increasing number of attacks and traffic loads that need to be dropped in the firewall. This problem is serious, and further aspects must be considered to reduce the traffic load on the firewall. Another problem with the IPS system is that if a normal IP is detected as a cyberattack, it will not only reduce the efficiency of the IPS system but also inconvenience normal service users. To increase the efficiency of IPS, it is necessary to continuously analyze the detected attack logs and reduce the threshold for the policy to minimize the false positive rate. This threshold can be adjusted with three variables: the number of acknowledgements of attacks, the duration of acknowledgments of attacks, and the time limit. If an attack is certain, the attack is prevented by applying the defense function of the IPS system [14], [15].

In some papers, various detection methods have been proposed based on analyzing traffic from network devices to overcome these problems. In [16], an efficient autonomous architecture which is able to collect traffic data from network devices and detect a service network attack by an automated AI-based module is proposed. In addition, some proposed algorithms in papers [17] and [18] include the rule generation to detect cyberattacks to reduce false positive rate. Studies that classifying and predicting cyberattack behaviors are outstanding challenged and actively ongoing. There are risks in these studies that are not immediately applicable to the service network. After being verified as a stable solution, there are various compatibility problems to apply to actual network devices [19].

Since we have already built various network security devices as well as firewalls, so we decided to use packet filtering firewall which has the highest performance that is suitable for our HPC environment. We also complete a duplex configuration to ensure high availability (HA) of services. Most supercomputing users not only send HPC jobs by accessing the login nodes, but also transfer files by accessing the DM nodes. In addition, technical support services are provided through access to web servers. The nodes that provide these services are vulnerable to cyberattacks and are therefore only accessible through the security devices we construct [20]. To reduce the load of infrastructure nodes with the login and DM nodes, we implement the technology of dropping attacks by analyzing the access and event log data of firewall and IPS systems [5], [21]. Therefore, as the number of cyberattacks rises, the overhead of the firewall increases. This causes delays in policy synchronization during the backup setting time and increases the failure rate of restoring data from the backup set. To solve this problem, we analyzed the firewall event as follows, and applied the existing pre-filtering method to the supercomputing service network environment and performed a performance comparison analysis.

Table 1. Cyberattacks and protection methods.

Cyberattcks	Installed systems (KISTI)	System types	Applied techniques
Denial of service	Anti-DDoS system	Network based IDS	Signature based IDS
Eavesdropping	OTP, Encryption Protocol (SSH, SFTP, HTTPS)	N/A	N/A
Spoofing	Anti-DDoS and IPS	Network based IDS	Threshold metric and signature based IDS
User to root (U2R)	Threat management system	Host based IDS	[5] [6]
Abuse login			
Application based attack	Web application firewall	Network based IDS	Signature based IDS
Unsecured or suspicious access	Firewall	Network based IDS	Rule based packet filtering

Table 2. Advantages and disadvantages about types of firewalls on HPC environment.

Types of firewalls	Advantages	Disadvantages
Packet filtering firewall	Efficient at processing packets and suitable for high-speed network environment	Difficult to securely configure
Circuit-level gateway	Providing privacy for data passing in/out of private network	Requiring modification to network protocol attack
Stateful inspection firewall	Capable of blocking many types of denial of service attacks and IPS	High processing overhead
Application-level gateway	Capable of detecting and blocking types of attack not visible at the open systems interconnection (OSI) model network or transport layers	High processing overhead and complex to configure and maintain
Next-generation firewall	Providing traditional firewall combined with IDS/IPS	Complex designed architecture and requiring high front-end investment of resource to acquire, configure and deploy these complex systems

Table 3. Black list based rules.

No.	Src. IP	Src. port	Dst. IP	Dst. port	Policy
1	Abnormal IP	Any	Serviced IP	Serviced port	DROP
			...		
N	Any	Any	Any	Any	PASS

Table 4. White list based rules.

No.	Src. IP	Src. port	Dst. IP	Dst. port	Policy
1	Normal IP	Any	Serviced IP	Serviced port	PASS
			...		
N	Any	Any	Any	Any	DROP

III. ANALYSIS OF FIREWALL EVENTS

Table 5 shows rules of the firewall on the KISTI supercomputing service network. We apply a hybrid rule with black list and white-list-based ACLs. We allow access to the supercomputing service only from the IP specified, but we allow arbitrary access to login nodes so that users can submit jobs and check their job statuses from anywhere, such as the $M-1$ rule in Table 5. To provide users with the convenience of using a supercomputer, we also allow access to DM nodes in a trust zone and web servers in a DMZ from anywhere, such as the *M* and *M + 1* rules in Table 5. Login nodes, DM nodes, and web servers have overloaded because they also process abnormal traffic. Therefore, we prioritize the unconditional dropping of abnormal IPs as the first rule. Access port 22 was opened for the login nodes, ports 80 and 443 were opened for the web servers, and ports 20, 21, and 22 were opened for the DM nodes. Packets that do not match any rules of the firewall are dropped by the last rule.

Before checking the effectiveness of the firewall, in 2017, we detected 235,737 abnormal IPs using the proposed method in previous studies [5]. In these studies, we used the failed access log and dropped packet events to detect abnormal IPs and add them to the abnormal IP list on the firewall in real time. Fig. 1 shows the abnormal IPs detected, ranked by the country and the cumulative data from these countries. It can be seen that the cu-

Table 5. Filtering rules of the firewall.

No.	Src. IP	Src. Port	Dst. IP	Dst. Port	Policy
1	Abnormal IP	Any	Any	Any	DROP
			...		
$M-1$	Any	Any	Login nodes	22	PASS
M	Any	Any	Web servers	80 /443	PASS
$M+1$	Any	Any	DM nodes	20 /21	PASS
			...		
N	Any	Any	Any	Any	DROP

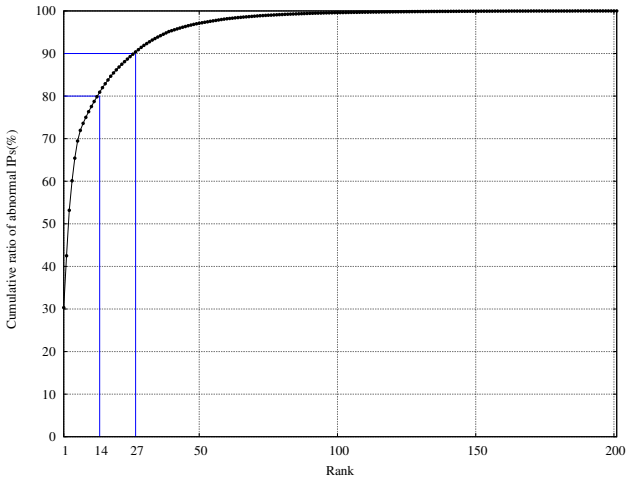

Fig. 1. The cumulative ratio of abnormal IPs.

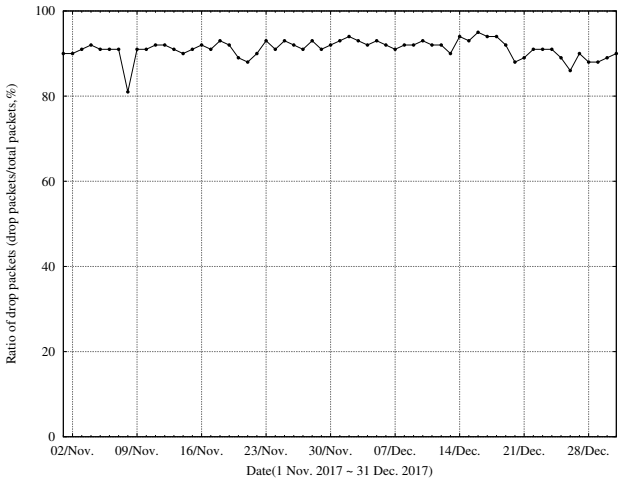

Fig. 2. Ratio of dropped packets out of total inbound packets by the firewall.

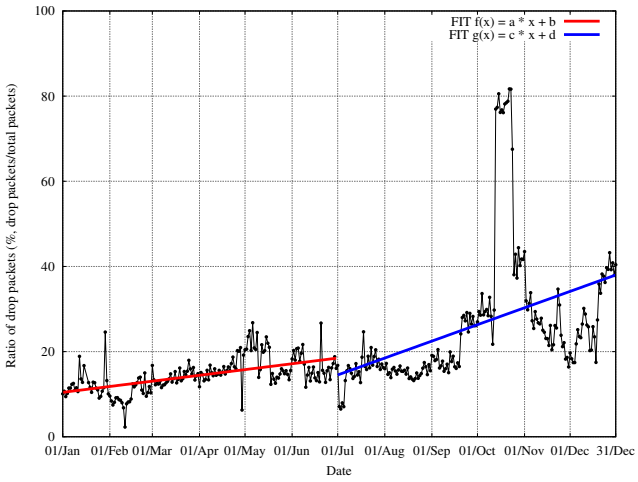

Fig. 3. Ratio of dropped packets out of total dropped packets by the drop rule of the firewall.

mulative ratio of country 1 among the abnormal IPs was 30%. The result of the top 14 countries was more than 80%. In addition, the cumulative data from 27 countries were over 90% of the total abnormal IPs. If all traffic from a specific country is applied to the ACL, the load on the firewall can be reduced. However, blocking in certain regions is limited because a supercomputing user who has been approved for use should be able to access remotely anytime and anywhere, even with an unspecified IP. For this reason, we conducted a firewall traffic analysis.

To check the effectiveness of the firewall, we analyzed firewall events on the KISTI supercomputing service network in 2017. Fig. 2 shows the line graph of the ratio of the dropped packets within all inbound packets to our infrastructure nodes from November 1 to December 31, 2017. We can confirm that the ratio of the dropped packets is approximately 90% during these days except for November 8, 2017 (system maintenance day). This means that only about 10% of the traffic is normal and the other 90% of the attempted IPs are abnormal. It can be seen that the firewall has excessive load due to abnormal traffic. This not only impacts the performance of the firewall, but also significantly reduces the performance of the entire security system. From an operational standpoint, it is inefficient because the detection and addition of abnormal IPs in real-time control systems require operating personnel.

We analyzed firewall events in more detail by rule to determine the causes of packet drops. As shown in Fig. 2 , the ratio of the dropped packets are given out of total inbound packets, and Fig. 3 shows the ratio of the dropped packets by the first filtering rule of the firewall (IP deny rule in Table 5) out of total dropped packets from January 1 to December 31, 2017. Our cyberthreat detection system [6] automatically detects and adds to this rule's assigned IPs for abnormal attacks.

As shown in Fig. 3, the daily percentage of dropped IPs by the first firewall rule increased linearly over the course of a year. There are also shorter periods that show exponential growth in the number of attacks (In Fig. 3 , during October) that could not be predicted. From this analysis, we can look forward to reducing the traffic processing load of the firewall without transferring these packets to the firewall directly. In Fig. 3, the red and blue lines show the fitted trend line with a linear equation, (1).

To obtain the values of a and c, we use the fitting algorithm of gnuplot, which is a portable command-line driven graphing utility for Linux [22] . The coefficient a and c for the drop packets ratio are found to be $a = 5.15e-7$ and $c = 1.49e-6$ by fitting. In this graph, the slope value c is greater than a. This means that the number of dropped packets in the second half of the year increases faster than in the first half. If we continually operate without any action, the load on the firewall server increases exponentially.

$$\begin{cases} F(x) = ax + b \\ G(x) = cx + d \end{cases} \tag{1}$$

The two boxes below show gnuplot fitting logs using (1). The first box shows fitting logs of the function *F(x)* and the second shows fitting logs of the function *G(x)*. We used initial coefficient values $a = 10e-10$ and $c = 10e-10$.

```
iter      chisq       delta/lim  lambda         a               b
   0 2.8641106008e+04   0.00e+00  1.27e+00    1.000000e-09   1.000000e+00
   1 2.9224368177e+03  -8.80e+05  1.27e-01    6.881838e-09   4.182704e+00
   2 2.8054456203e+03  -4.17e+03  1.27e-02    3.985782e-08  -4.495393e+01
   3 2.0050344845e+03  -3.99e+04  1.27e-03    4.550188e-07  -6.639947e+02
   4 1.9921311412e+03  -6.48e+02  1.27e-04    5.147609e-07  -7.530753e+02
   5 1.9921311144e+03  -1.34e-03  1.27e-05    5.148470e-07  -7.532037e+02
iter      chisq       delta/lim  lambda         a               b

After 5 iterations the fit converged.
Final sum of squares of residuals : 1992.13
Real change during last iteration : -1.34126e-08

Degrees of freedom (FIT_NDF) : 177
The rms of residuals  (FIT_STDFIT) = sqrt(WSSR/ndf) : 3.35484
The variance of residuals (reduced chisquare) = WSSR/ndf   : 11.255

Final set of parameters          Asymptotic Standard Error
=======================          ==========================
a = 5.14847e-07                  +/- 5.588e-08    (10.85\%)
b = -753.204                     +/- 83.32        (11.06\%)

Correlation matrix of the fit parameters: a and b
a       1.000
b       -1.000  1.000
```

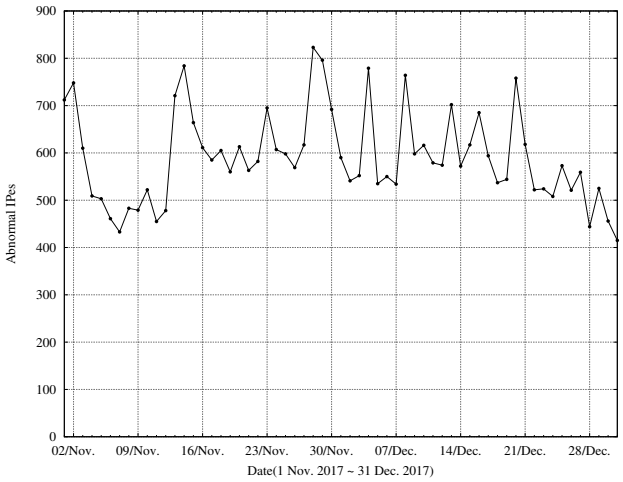

Fig. 4. Number of abnormal IPs are detected by our cyber threat detection system.

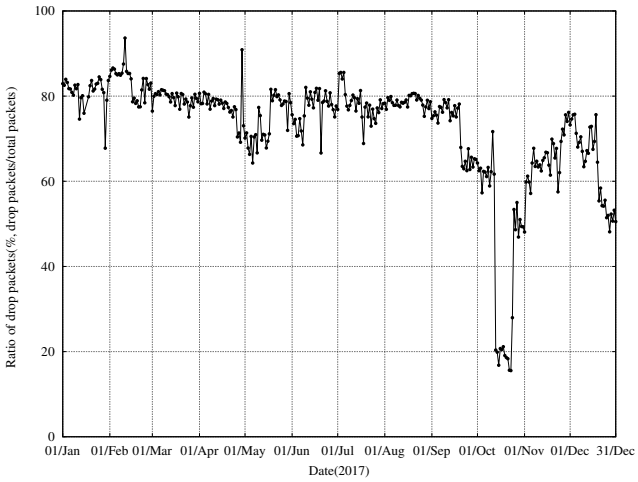

Fig. 5. Ratio of dropped packets by non-matching of firewall rules out of total dropped packets.

```
iter    chisq           delta/lim lambda        c               d
   0 1.4988044409e+05   0.00e+00  1.28e+00    1.000000e-09   1.000000e+00
   1 4.5958972899e+04  -2.26e+05  1.28e-01    1.298616e-08   6.639141e+00
   2 4.4861374086e+04  -2.45e+03  1.28e-02    1.124974e-07  -1.432383e+02
   3 3.7665212620e+04  -1.91e+04  1.28e-03    1.318868e-06  -1.960976e+03
   4 3.7557811350e+04  -2.86e+02  1.28e-04    1.485118e-06  -2.211478e+03
   5 3.7557811146e+04  -5.43e-04  1.28e-05    1.485347e-06  -2.211824e+03
iter    chisq           delta/lim lambda        c               d

After 5 iterations the fit converged.
Final sum of squares of residuals : 37557.8
Real change during last iteration : -5.43091e-09

Degrees of freedom    (FIT_NDF)  : 182
The rms of residuals  (FIT_STDFIT) = sqrt(WSSR/ndf)  : 14.3653
The variance of residuals (reduced chisquare) = WSSR/ndf : 206.362

Final set of parameters         Asymptotic Standard Error
=======================         ==========================
c = 1.48535e-06                 +/- 2.308e-07    (15.54\%)
d = -2211.82                    +/- 347.7        (15.72\%)

Correlation matrix of the fit parameters:  c and d
c       1.000
d       -1.000  1.000
```

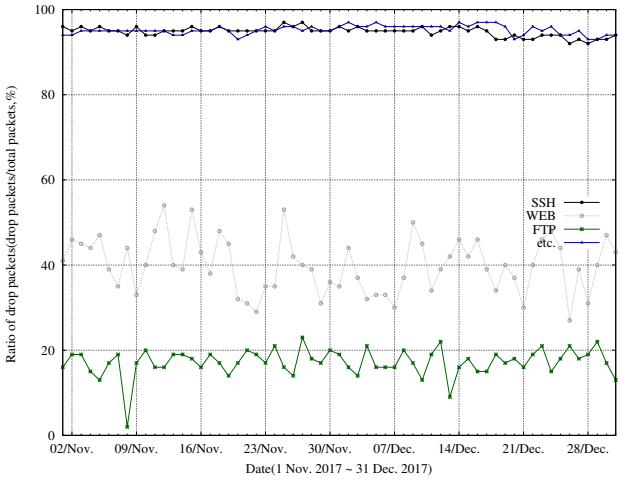

Fig. 6. Ratio of dropped packets for each service(SSH, HTTP/HTTPS, FTP) on the firewall.

Fig. 4 shows the statistics of anomalous IPs from November 1, 2017 to December 31, 2017. As shown in the graph, hundreds of abnormal IPs occurred continuously each day. These have led to a continuous increase in the number of IPs in the filtering rule and show a trend in which the percentage of packets blocked by the rule is increasing, as shown in Fig. 3. However, these results show that the overhead of the firewall grew significantly owing to an increase in the number of incoming packets for abnormal IPs or ports according to the firewall policy.

Fig. 5 shows the ratio of the packets dropped by the last filtering rule of the firewall policy portrayed in Table 5. In this case, there were no matched rules on the firewall. Compared with Fig. 3, the graph trend is almost the opposite. This means that if more IPs that are dropped by the first rule, fewer IPs are dropped by the last rule. We will explore solutions for reducing the network traffic load on this firewall through an analysis that affects the dropped packet rates.

Finally, we analyzed the ratio of dropped packets for each service by extracting them from the firewall events. Fig. 6 shows the percentage of dropped packets for services. The green line shows the ratio of dropped packets out of total file transfer protocol (FTP) packets. The drop ratio of FTP service was approximately 22%, which was detected as the smallest packet drop ratio among all services. The drop ratio of HTTP/HTTPS for using the web service (the gray line in Fig. 6) was approximately 50%, which was twice that of the FTP service. Finally, the services with the highest packet drop ratio of almost 95% were SSH or other service ports with remote access. Most supercomputing service users access login nodes via SSH and submit HPC jobs with schedulers. In addition, attacks are often attempted via the SSH protocol. Services classified as etc. refer to HPC application services that use other ports except SSH, HTTP/HTTPS, and FTP service ports.

Table 6. Firewall orientation.

No.	Orientation
1	DMZ zone -> trust zone
2	DMZ -> un-trust zone
3	trust zone -> DMZ zone
4	trust zone -> un-trust zone
5	un-trust zone -> DMZ zone
6	un-trust zone -> trust zone

IV. PRE-FILTERING ACL

To solve the previously mentioned problems by applying our policy to reduce the overhead of the firewall, a physical network security environment was constructed in the national supercomputing center on KISTI. It currently provides supercomputing services based on the 5th national supercomputer "Nurion," which has a processing speed of 25.7 petaflops. In this section, we describe the scenario in which an IP filtering policy is applied in this service environment, and then detail the ACL policy.

A. HPC Service Environment

Our service environment targets HPC services, and focuses on the requirements of HPC users who use our resources. Therefore, increased security is essential to provide paid services with a billing system that is accessible anytime, anywhere, and pays only for the use of resources [23], [24]. Most HPC users are researchers in the field of computational science, and most of them execute parallel jobs to analyze data. Recently, there has been a growing demand for big data applications, and even more demand for artificial intelligence on the execution of models such as deep learning. To meet the needs of these users, we are not only trying to satisfy the convenience of users but also to ensure the security of these data. Based on these considerations, we constructed a network security system suitable for the service environment, as shown in Fig. 7.

Users can access infrastructure nodes located in the trust zone through network security devices, such as anti-DDoS, IPS, and fire–wall in the trust zone (Fig. 7). All the security devices are in the duplex configuration, which are highly available in an active-standby architecture. This architecture is automatically switched to the device in a standby state when the active device fails or the network lines are cut off to maintain the continuity of service. The infrastructure nodes for the supercomputing service consist of login nodes, DM nodes, web servers, and compute nodes on which jobs are running. We operate separately an un-trust zone, trust zone, and DMZ. Login nodes and DM nodes were placed in the trust zone and web servers were placed within the DMZ. As shown in Fig. 7, we create and apply an ACL policy in front of the firewall devices separating the trust zone and DMZ from the un-trust zone.

In addition, we build the firewall orientation (Table 6) with these three zones and operate them by adding firewall policies. According to these policies, we create groups with defined objects to add firewall rules. For each rule, we can enable or disable policies, make PASS or DROP policies, and enable or disable log collections.

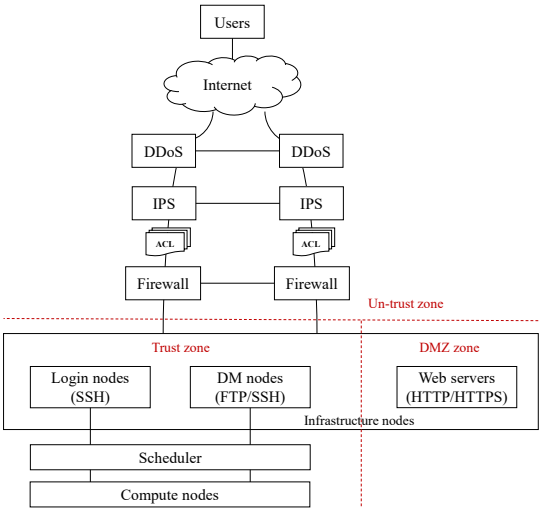

Fig. 7. Service environment of network security construction.

Table 7. Filtering rules of ACL.

No.	Src. IP	Src. port	Dst. IP	Dst. port	Policy
1	Any	Any	Assigned IP	Assigned port	PASS
			...		
N	Any	Any	Any	Any	DROP

B. Method

The ratio of the packets dropped by the firewall is explained in Section III. We analyzed the average daily number of packets and dropped packets obtained from the firewall logs as counted data for eight weeks. In addition, we selected the total number of packets and number of dropped packets for each service, such as SSH, HTTP/HTTPS, and TCP. We confirmed that the number of abnormal IPs is constantly rising. As such, this policy reduces the number of packets processed by the server that adds abnormal IPs to the first filtering rule of the firewall, but it increases the firewall load. We propose a method to add ACL rules in front of the firewall as shown in Fig. 7. The added ACL rules are listed in Table 7. Assigned IPs as the destination IPs are considered operating infrastructure nodes. Assigned ports as the destination ports are considered the service ports of infrastructure nodes with login nodes, DM nodes, and web servers.

An evaluation of the addition of this policy to reduce the traffic overhead of the firewall is detailed in Section V. To ensure service continuity, we applied the ACL policy according to the following principles (Fig. 8). At the front-end, the anti-DDoS and IPS systems detect and drop attacks in real time. After adopting new policies through log analysis results at the back-end, the accuracy of intrusion detection must be improved.

V. EVALUATION

To reduce the overhead of the firewall, we added and applied an ACL policy in front of the firewall that allows access only to the infrastructure nodes in a service. After the application, the results were evaluated according to the following three tests resulting from our previous analysis of firewall event log data. We took all days and applied method-based ACLs to the service. We applied our method-based ACLs and then analyzed changes

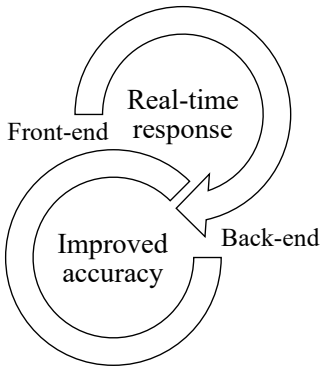

Fig. 8. Principles of network security policy.

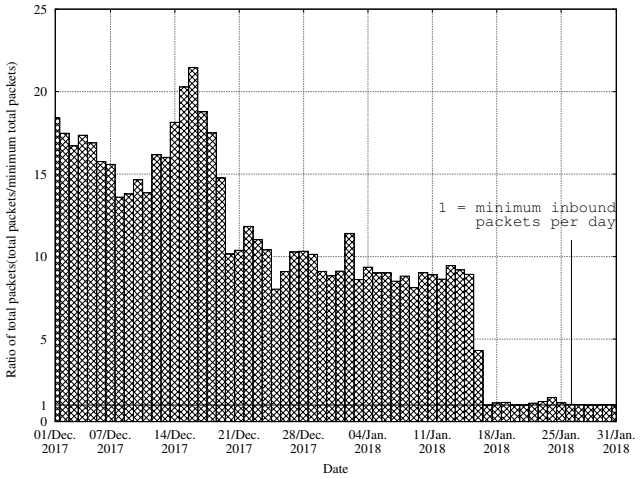

Fig. 9. Traffic volume to be handled by the firewall.

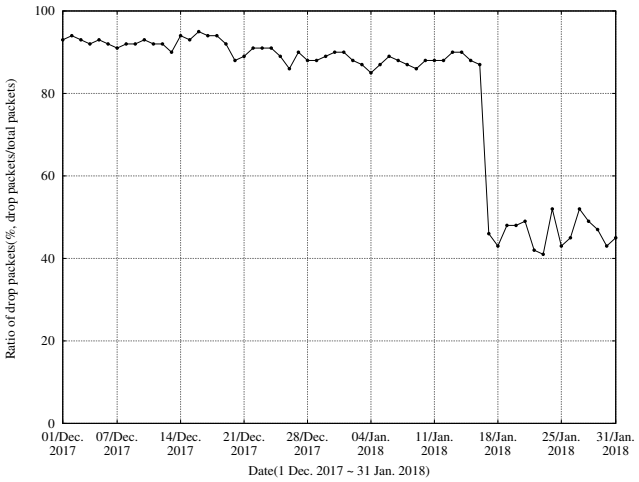

Fig. 10. Ratio of the dropped packets with pre-filtering ACL.

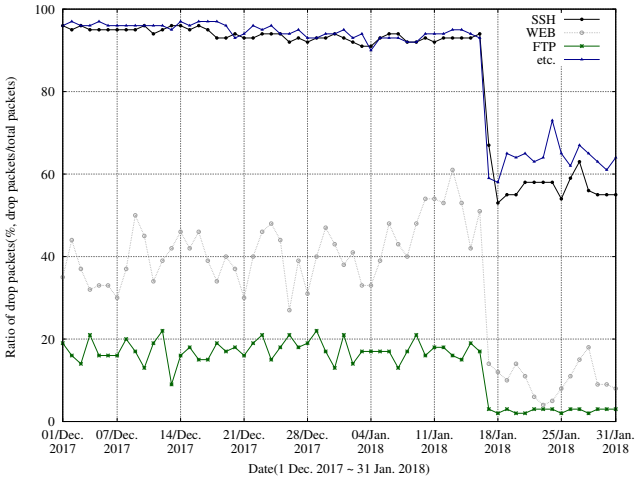

Fig. 11. Ratio of the dropped packets for all services with pre-filtering ACL.

in the amount of inbound traffic. Because of the privacy of the data, we have drawn the trend by statistical and computational methods, as shown in Fig. 9. This figure shows the amount of inbound traffic handled by the firewall from December 1, 2017 to January 31, 2018. We defined 1 volume unit as the minimum inbound average packets per day. Fig. 9 shows that inbound traffic to the firewall is reduced to 5% of the maximum after implementing ACL rules in front of the firewall on January 16, 2018. Therefore, we could reduce the traffic load on the firewall by about 20 times.

Fig. 10 shows the ratio of the dropped packets before and after applying the ACL policy. As of January 16, 2018, our drop ratio was reduced from approximately 90% to less than 50%. By adding a method-based ACL to the front of the firewall policy based on the blacklist, we reduced the traffic load of firewall servers by almost 40%. Although our method is simple, it has efficient results in the network security operation. Fig. 11 shows that the dropped packet ratio for all services is reduced.

The dropped packet ratio for the FTP service was reduced from 22% to 3%, and that for the web service was reduced from 50% to 11%. The dropped packet ratio for the SSH service was reduced from almost 90% to 60%, while those for other remote services were reduced from 90% to 65%. The number of blocked IPs with abnormal access shows a sharp decrease after application of the ACL policy, as shown in Fig. 12. In the graph, the number of abnormal IPs decreased from about 600 per day to about 50 per day, a 92% reduction of traffic attacks on firewall servers.

We measured the CPU load on the firewall server for 40 h before and 40 h after applying the pre-filtering ACL. The first 40 h, represented by the black data points in Fig. 13, show the CPU load before applying the pre-filtering ACL. The next 40 h in blue data points show the CPU load after applying the pre-filtering ACL. In Fig. 13, the CPU load values are obtained by setting the average CPU load value over the 80 h of data collection to 1. It can be seen that the majority of the blue points are lower than the black points, which means the load of the server was reduced after the pre-filtering ACL was applied. The CPU load was reduced by an average of 21.5%.

As a result, from an operational perspective, the application of pre-filtering ACL to our security system has resulted in a reduction of traffic load on the firewall servers for all packet types such as SSH, HTTP/HTTPS, FTP, and other HPC applications. Firewall server performance has improved by approxi-

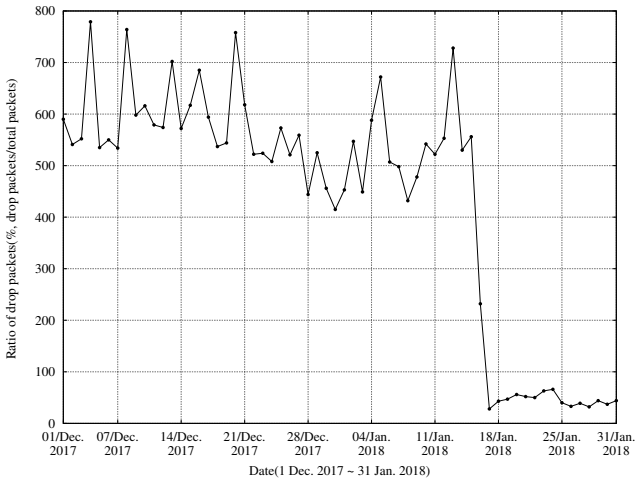

Fig. 12. Number of abnormal IPs with pre-filtering ACL.

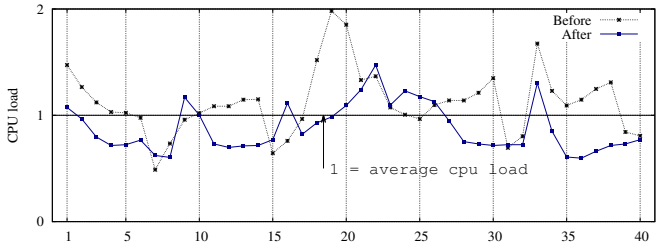

Fig. 13. The CPU load changes with pre-filtering ACL.

mately 21.5% in the obtained results.

VI. CONCLUSION AND FUTURE WORK

In this paper, we evaluated the ratio of dropped packets in firewall events to be almost 90%. More than 95% of the packets attempting to access to login nodes in the HPC environment through the SSH service port were dropped. As the number of packets dropped by the firewall increases, not only the overhead of event-based solution which detects abnormal traffic using these drop packets increases, but also the number of pre-defined rules applied to the firewall increases. And this overhead causes problems: (1) Delay when storing and synchronizing rules in the duplication configuration for the firewall; (2) considerable time utilization or increase in the failure rate to restore backup settings. To solve these problems, we applied a pre-filtering ACL policy with the assigned IPs to the firewall. The results show that the ratio of the dropped packets is reduced from 90% to less than 50%. In addition, the number of daily abnormal IPs decreased from 600 to 50. Finally, by applying a pre-filtering ACL, we found that the CPU load was accurately reduced by 21.5%. All analysis results show our main contribution of this paper is the reduction of the firewall's traffic overhead by applying the pre-filtering policy.

Although we have reduced traffic overhead on the firewall from an operational standpoint, effective automatic management method of existing firewall rules needs to be improved. In addition, our supercomputing service environment has limitations in applying various development components directly because of the stability priority. In the future, we plan to introduce a machine learning solution for intelligent firewall analysis-based traffic handling and build an intelligent automated security system.

REFERENCES

[1] G. Xu, H. Ibeid, X. Jiang, V. Svilan, and Z. Bian, "Simulation-based performance prediction of HPC applications: A case study of HPL," in *Proc. IEEE/ACM HUST/ProTools*, 2020, pp. 81–88.

[2] S. Peisert, "Security in high-performance computing environments," *ACM Commun. ACM*, vol. 60, no. 9, pp. 72–80, Sept. 2017.

[3] K. Fan, H. Yang, and A. Xu, "Analysis of power network behavior security analysis technology," *MATEC Web Conferences*, vol. 246, p. 03017, 2018.

[4] V. S, and M. Sylviaa , "Intrusion detection system – a study," *Int. J. Security Privacy Trust Manage.*, vol. 4, no. 1, pp. 31–44, Feb. 2015.

[5] J. K. Lee, S. J. Kim, C. Y. Park, and T. Hong, "Heavy-tailed distribution of the SSH brute-force attack duration in a multi-user environment," *J. Korean Phys. Soc.*, vol. 69, no. 2, pp.253–258, July 2016.

[6] J. K. Lee, S. J. Kim, J. Woo, and C. Y. Park, "Analysis and response of SSH brute force attacks in multi-user computing environment," *KIPS Trans. Comp. Commun. Syst.*, vol. 4, no. 6, pp.205–212, June 2015.

[7] V. Redya, S. Chatrapati, and Kamalesh, "Paper on types of firewall and design principles," *Int. J. Sci. Research*, vol. 5, no. 5, pp. 1583–1590, May 2016.

[8] F. A. B Hamid Ali, "A study of technology in firewall system," in *Proc. IEEE ISBEIA*, 2011.

[9] N. Stojanovski, M. Gusev, "Analysis of identity based firewall systems," *ICEST*, vol. 1, pp. 357–362, 2010.

[10] R. Oppliger, "Internet security: Firewalls and beyond," *ACM Commun. ACM*, vol. 40, no. 5, pp. 92–102, May 1997.

[11] H. Hamed, A. EI-Atawy, and E. AI-Shaer, "On dynamic optimization of packet matching in high speed firewalls," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 10, pp. 1817–1830, Oct. 2006.

[12] A. Khoumsi, M. Erradi, and W. Krombi, "A formal basis for the design and analysis of firewall security policies," *J. King Saud University Comput. Inf. Sci.*, vol. 30, no. 1, pp.51–66, Jan. 2018.

[13] K. Mindo, C. Sogomo, and N.M. Karie, "Analysis of network and firewall security policies in dynamic and heterogeneous networks," *Int. J. Advanced Research Comput. Sci. Software Eng.*, vol.6, no. 4, Apr. 2016.

[14] C. Y. Ho *et al.*, "False positives and negatives from real traffic with intrusion detection/prevention systems," *Int. J. Future Comput. Commun.*, vol.1, no.2, Aug. 2012.

[15] M. Kumar, M.H.Hanumanthappa, and T.V.Suresh Kumar, "Intrusion detection system-false positive alert reduction technique," *ACEEE Int. J. Network Security*, vol. 2, no. 3, pp. 37–40, July 2011.

[16] J. Fesl *et al.*, "Towards HPC-based autonomous cyber security system," in *Proc. IEEE ACIT*, 2019.

[17] S. Lysenko, K. Bobrovnikova, R. Shchuka, and O. Savenko, "A cyber-attacks detection technique based on evolutionary algorithms," in *Proc. IEEE DESSERT*, 2020.

[18] G. Kbar and A. Alazab, "A comprehensive protection method for securing the organization's network against cyberattacks," in *Proc. IEEE CCC*, 2019.

[19] I. Perry *et al.*, "Differentiating and predicting cyberattack behaviors using LSTM," in *Proc. IEEE DSC*, 2018.

[20] C. Xue, X. Wu, T. Qin, H. Bi, and Z. H. Wang, "Improvement of ACL assemply algorithm," in *Proc. IEEE ICSRS*, 2017.

[21] A. Kumar S, A. Bandyopadhyay, B. H, I. Singhania, and K. Shah, "Analysis of network traffic and security through log aggregation," *Int. J. Comp. Sci. Inf. Security*, vol. 16, no. 6, pp. 124–131, June 2018.

[22] P. K. Janert, "Gnuplot in Action," *Manning Publications Company*, Mar. 2016.

[23] R. Bulusu, P. Jain, P. Pawar, M. Afzal, and A. Wandhekar, "Addressing security aspects for HPC infrastructure," in *Proc. IEEE ICICT*, 2018.

[24] M. Chung, W. Ahn, B. Min, J. Seo, and J. Moon, "An analytical method for developing appropriate protection profiles of instrumentation and control system for nuclear power plants," *J. Supercomputing*, vol. 74, no. 3, pp. 1378–1393, Mar. 2018.

Jae-Kook Lee received his B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from the Chungnam National University, South Korea, in 2002, 2004 and 2012, respectively. He is also working as a Researcher in the Supercomputing Center at the Korea Institute of Science and Technology Information (KISTI), South Korea from 2013. His research interests are in HPC system, and network security.

Taeyoung Hong received his B.S. and M.S. degrees in Physics from the Sungkyunkwan University, South Korea, in 1999 and 2002, respectively. He is also working as a Director in the Supercomputing Center at the Korea Institute of Science and Technology Information (KISTI), South Korea from 2003. His research interests are in HPC system operation, and parallel file system.

Guohua Li received her Ph.D. in Interdisciplinary IT from Konkuk University in 2018. She completed her M.S. degree from Konkuk University in 2013. She participated in the development of container-based HPC cloud service platform as a Co-Researcher from 2016 to 2017. She currently is a Post-Doctoral Researcher for the Korea Institute of Science and Technology Information (KISTI). Her research interests include HPC, network security, and cloud computing.