KIPS Trans. Comp. and Comm. Sys.
Vol.9, No.5 pp.113~120 pISSN: 2287-5891

https://doi.org/10.3745/KTCCS.2020.9.5.113

Analysis of Traffic and Attack Frequency in the NURION Supercomputing Service Network

Jae-Kook Lee† · Sung-Jun Kim†† · Taeyoung Hong†††

ABSTRACT

KISTI(Korea Institute of Science and Technology Information) provides HPC(High Performance Computing) service to users of university, institute, government, affiliated organization, company and so on. The NURION, supercomputer that launched its official service on Jan. 1, 2019, is the fifth supercomputer established by the KISTI. The NURION has 25.7 petaflops computation performance. Understanding how supercomputing services are used and how researchers are using is critical to system operators and managers. It is central to monitor and analysis network traffic. In this paper, we briefly introduce the NURION system and supercomputing service network with security configuration. And we describe the monitoring system that checks the status of supercomputing services in real time. We analyze inbound/outbound traffics and abnormal (attack) IP addresses data that are collected in the NURION supercomputing service network for 11 months (from January to November 1919) using time series and correlation analysis method.

Keywords : NURION Supercomputer, Network Traffic, Frequency Analysis, Correlation Analysis, Time Series Analysis

누리온 슈퍼컴퓨팅서비스 네트워크에서 트래픽 및 공격 빈도 분석

이 재 국† · 김 성 준†† · 홍 태 영†††

요 약

한국과학기술정보연구원은 대용량 데이터를 초고속으로 생산·처리·활용할 수 있는 국가슈퍼컴퓨팅시스템을 구축·운영하여 사용자(대학, 연구소, 정부 및 산하기관, 기업체 등)에게 HPC(High Performance Computing) 서비스를 제공하고 있다. 2019년 1월 1일 공식 서비스를 개시한 국가슈퍼컴퓨터 누리온은 한국과학기술정보연구원에서 5번째로 구축한 시스템으로 이론성능 25.7 페타플롭스를 갖는다. 시스템 운영자나 사용자의 관점에서 슈퍼컴퓨터의 사용 방법과 운영 방식을 이해하는 것은 매우 중요하다. 이를 이해하는 작업은 네트워크 트래픽을 모니터링하고 분석하는 것에서 시작된다. 본 논문에서는 누리온 시스템과 슈퍼컴퓨팅서비스 네트워크 및 보안 구성에 대하여 간략히 소개한다. 그리고 슈퍼컴퓨팅서비스 현황을 실시간으로 확인하기 위한 모니터링 체계를 기술하고 서비스를 시작하고 11개월(2019년 1월~11월) 동안 수집된 슈퍼컴퓨팅서비스 네트워크의 인바운드 및 아웃바운드 트래픽과 비정상행위(공격) 탐지 IP 개수에 대한 시계열 및 상관관계 분석을 수행한다.

키워드 : 슈퍼컴퓨터 누리온, 네트워크 트래픽, 빈도 분석, 상관관계 분석, 시계열 분석

1. 서 론

한국과학기술정보연구원(이하 KISTI) 국가슈퍼컴퓨팅본부에서는 2019년 1월 1일부터 5번째로 구축한 국가슈퍼컴퓨팅시스템인 누리온을 운영하고 있다[1]. 누리온은 전 세계에서 14번째로 빠른 컴퓨터(2019년 11월 기준)로 이론 성능 25.7 페타플롭스(PFlops, 1015) 연산 능력을 갖고 있다(2018년 6월 TOP 500 최초 등록 순위는 11위). KISTI는 1988년 1호기 국가슈퍼컴퓨팅시스템을 시작으로 1993년 2호기, 2001년 3호기, 2008년 4호기를 구축·운영하였다[2]. 현재 5호기 누리온은 150여개 국내외 연구소, 정부산하기관, 대학 및 기업 등에서 열유체, 대기환경 등과 같은 계산과학분야 및 전산구조분야 등 전통적인 HPC(High Performance Computing) 서비스를 제공하고 있을 뿐만 아니라 기계학습, 인공지능, 빅 데이터 분석 등 다양한 분야로 활용 범위를 넓혀가고 있다. 누리온은 국가과학기술연구망인 KREONET에 10Gbps 대역폭 네트워크에 연결하여 서비스하고 있다. 누리온 서비스 네

※ 본 연구는 2019년도 한국과학기술정보연구원(KISTI) 주요사업 과제로 진행한 것입니다.
※ 이 논문은 2019년 한국정보처리학회 추계학술발표대회에서 "누리온 슈퍼컴퓨팅서비스 네트워크에서 트래픽 및 공격 빈도 분석" 제목으로 발표된 논문을 확장한 것임.
† 정 회 원 : 한국과학기술정보연구원 슈퍼컴퓨팅인프라센터 연구원
†† 정 회 원 : 한국과학기술정보연구원 슈퍼컴퓨팅인프라센터 선임연구원
††† 정 회 원 : 한국과학기술정보연구원 슈퍼컴퓨팅인프라센터 센터장
Manuscript Received : December 31, 2019
Accepted : January 16, 2020
* Corresponding Author : Sung-Jun Kim(sjkim@kisti.re.kr)

트워크는 크게 사용자가 사용할 계정을 신청하고 국가슈퍼컴퓨팅서비스를 사용하면서 발생하는 다양한 질의·응답 등의 서비스를 제공받기 위한 웹 서비스 포트(HTTP/HTTPS)와 사용자가 터미널로 접속하여 연산 작업을 제출할 수 있는 서비스 포트(SSH), 연산을 위해 필요한 데이터를 업로드하거나 연산 결과를 다운로드 할 수 있는 파일 전송 서비스 포트(FTP/SFTP)로 구성된다[3]. 사용자는 이러한 서비스 포트를 이용하여 원격에서 슈퍼컴퓨팅 자원을 이용하여 연구를 진행하게 된다. 원격에서 사용자들의 접속이 가능한 누리온을 보다 안전하게 이용할 수 있도록 슈퍼컴퓨팅서비스 네트워크에는 DDoS 공격 대응 시스템 및 침입차단시스템, 방화벽, 웹 방화벽 등 다양한 보안 장비들이 운영되고 있으며 시스템 레벨에서도 다중인증 및 서버보안 솔루션을 적용하여 시스템을 보호하고 있다. 또한 슈퍼컴퓨팅종합상황실에서 시스템 현황 및 보안 모니터링을 24×7 수행하여 실시간으로 이상 징후를 탐지하고 차단한다[3-5]. 슈퍼컴퓨터를 운영하고 관리하는 입장에서 네트워크 트래픽을 모니터링하고 분석하는 작업은 시스템이 잘 사용되고 있는지, 외부의 공격으로부터 안전하게 보호되고 있는지를 확인하는데 기본이 된다[6-8].

본 논문에서는 국가 슈퍼컴퓨팅시스템인 슈퍼컴퓨터 5호기 누리온을 간략히 소개하고 누리온이 유·무상 서비스를 시작한 2019년 1월 1일부터 11개월 동안 발생한 인바운드 및 아웃바운드 트래픽과 비정상행위 탐지 결과 데이터를 이용하여 시계열 및 상관관계 분석을 수행하고 그 특징을 확인한다.

본 논문의 구성은 다음과 같다. 2장에서는 누리온 시스템 및 서비스 네트워크 환경을 기술하고, 사용자 작업 모니터링 시스템 및 비정상행위를 탐지하는 시스템에 대하여 설명한다. 3장에서는 슈퍼컴퓨팅서비스를 제공하면서 외부 네트워크와 내부 네트워크의 가장 앞단에 위치한 DDoS 대응 시스템 단에서 수집된 트래픽 데이터와 비정상적으로 슈퍼컴퓨터의 접속을 시도하는 공격지 IP 탐지 결과 데이터를 이용하여 시계열 및 상관관계 분석을 진행하고 그 특징을 확인한다. 끝으로 4장에서는 결론을 맺고 향후 분석결과를 활용한 연구주제를 제시한다.

2. 누리온 서비스 및 모니터링

2.1 누리온 및 서비스 네트워크

누리온은 KISTI에서 5번째로 구축한 국가슈퍼컴퓨팅시스템으로 이론 성능 25.7 페타플롭스를 갖는 리눅스 기반의 초병렬 클러스터 시스템이다. 누리온은 병렬 프로그램이 수행되는 계산노드 및 계산노드와 스토리지 시스템을 초고속으로 연결해주는 인터커넥트 네트워크, 사용자 데이터 및 계산 결과가 저장되는 스토리지 및 사용자가 원격에서 접속할 수 있는 인프라 노드 등으로 구성되어 있다. 계산노드는 인텔 매니코어 기반의 나이츠랜딩 노드(Compute Nodes-KNL) 8,305대와 제온 서버 프로세서인 스카이레이크 노드(Compute Nodes-SKL) 132대로 구성되어 있다. 인터커넥트 네트워크는 100Gbps급의 인텔 OPA(Omni-Path Architecture)이며 스토리지는 사용자 홈디렉토리(/home01), 어플리케이션 디렉토리(/apps) 및 사용자 작업 디렉토리(/scratch) 21 페타바이트 이상의 러스터(Lustre) 병렬파일시스템과 고속의 파일 입출력을 제공하기 위한 800테라바이트 이상의 버스트버퍼(/scratch_ ime)로 구성되어 있다. 또한 시스템의 중요 데이터 백업 및 사용자의 데이터 장기 보관을 위해서 10 페타바이트 이상의 테이프 스토리지(Tape Storage)도 갖추고 있다. Fig. 1은 누리온 시스템 구성도이다.

누리온은 국가과학기술연구망인 KREONET에 연결되어 사용자가 인터넷을 이용하여 원격에서도 슈퍼컴퓨팅서비스를 이용할 수 있다. Fig. 2는 누리온 국가슈퍼컴퓨팅서비스 네트워크 및 보안 개념도를 나타낸다. 비 신뢰구간에 있는 사용자는 인프라 노드인 로그인 노드(Login Nodes)와 데이터

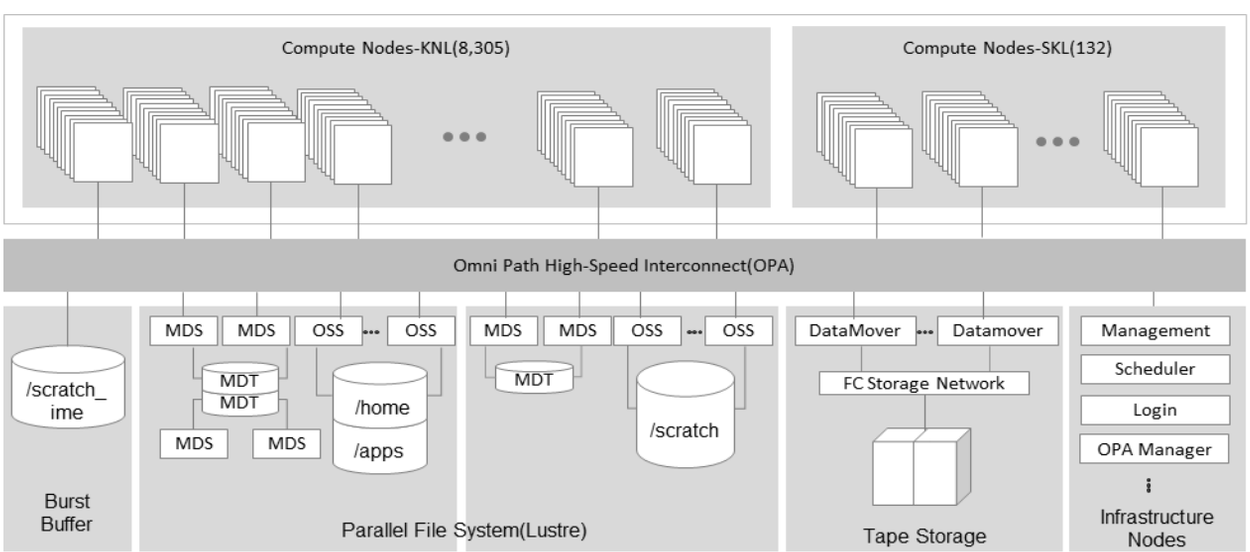

Fig. 1. System Architecture of the NURION Supercomputer

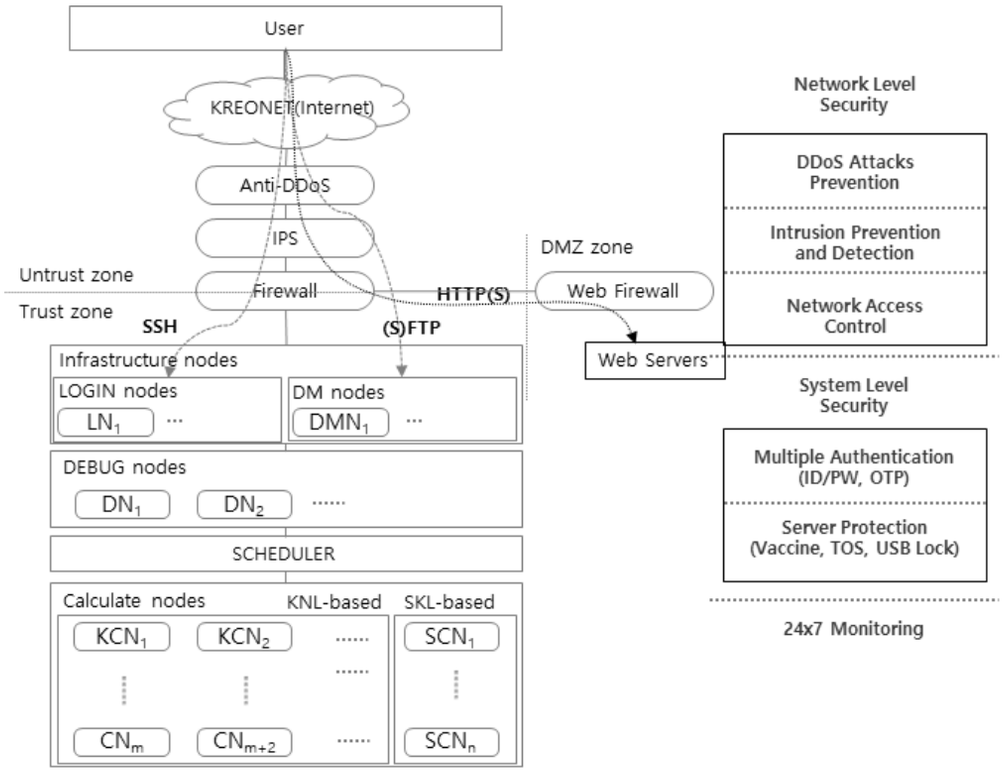

Fig. 2. Conceptual Diagram on Network and Security of the NURION Supercomputing Service

무버 노드(Datamover Nodes)에만 지정된 서비스 포트를 이용하여 접근할 수 있다. 서비스 거부 공격(DoS/DDoS)이나 해킹, 비인가 접근 등 다양한 사이버 공격으로 부터 누리온을 보호하기 위하여 네트워크 레벨에서는 DDoS 공격 대응 시스템, 침입차단시스템(IPS), 방화벽 및 웹 방화벽 시스템 등을 운영하고 있으며 시스템 레벨에서는 일회용패스워드(OTP) 및 사용자 행위 감시, 백신 등을 통하여 보안을 강화하고 있다. 각 레벨에서 운영 중인 시스템 및 인프라 노드에서 발생되는 이벤트 및 로그는 실시간으로 수집 시스템에 수집되어 24×7 모니터링하고 있으며 실시간으로 분석하여 비정상행위를 탐지하고 공격지 IP를 차단하는 탐지시스템[3, 4](이하 슈퍼컴퓨터 위협 관리 시스템(Supercomputer Threat Monitoring System))을 활용하여 안전하게 누리온 자원을 이용할 수 있도록 한다. 슈퍼컴퓨터 위협 관리 시스템은 서버단에서 수집된 실패 로그와 네트워크 단에서 수집된 차단(DROP) 로그를 기반으로 임계치 이상 접속을 실패하거나 접근 권한이 없는데도 지속적으로 접근을 시도하는 비정상행위를 유발하는 공격지 IP를 실시간으로 차단한다.

2.2 사용자 작업 모니터링

동시에 여러 사용자가 접속하여 사용할 수 있는 슈퍼컴퓨터 누리온은 사용자가 작성한 대규모 계산과학 병렬 프로그램을 관리하기 위하여 작업관리 프로그램(이하 스케줄러)을 이용한다. 스케줄러는 작업을 수행할 수 있는 계산 자원의 상태를 파악하여 사용자가 작업을 제출(Submit)하면 유휴 계산 노드에 이를 할당하고 계산을 수행하게 된다. 누리온에서 수행되는 사용자 작업을 모니터링하기 위하여 작업을 제출한 사용자 정보는 계정관리 데이터베이스에 저장된 정보를 기반으로 추출하며, 할당된 유휴 계산노드의 상태 및 사용자별 작업 정보는 누리온의 작업 스케줄러인 PBSPro에서 추출한다[9]. Fig. 3은 사용자 작업 모니터링을 위한 시스템(이하 작업 모니터링 시스템)의 구조를 도식화 한 것이며 Table 1은 각 구성요소와 역할을 나타낸다. 생성모듈(Generator Module)은 누

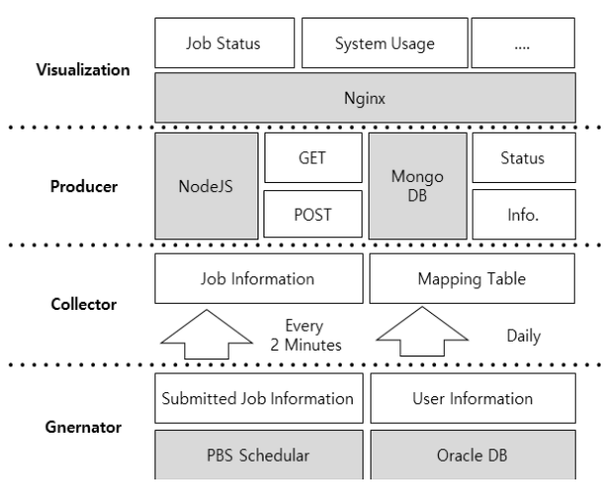

Fig. 3. Architecture of Job Monitoring System

Table 1. System Monitoring Modules and Roles

Modules	Role
Generator Module	Search user information and generate the system and job's status information
Collector Module	Cluster and classify of job information
Producer Module	Select user and job information from database and transfer that to visualization module
Visualization Module	Visualize the status of job and usage of system

리온 시스템의 활용 상태 및 사용자 작업 정보를 PBSPro에서 주기적(2분)으로 추출하며, 사용자 정보를 계정관리 데이터베이스에서 매일 1회 추출한다. 수집모듈(Collector Module)은 생성모듈에서 전달된 작업 및 사용자 정보를 분류하고 저장하는 역할을 하며, 분류된 정보는 전통적인 관계형 데이터베이스보다 더 융통성 있는 데이터 모델을 사용하고, 데이터 저장 및 검색에 특환된 메커니즘을 제공하는 NoSQL 데이터베이스에 저장하여 보관한다. 처리모듈(Producer Module)은 데이터베이스에서 정보를 추출하고 가시화 모듈에서 요구하는 형태로 정보를 가공하여 전달하는 역할을 수행한다. 가시화 모듈(Visualization Module)은 사용자가 웹 인터페이스를 통하여 사용자 작업 및 시스템 현황을 실시간으로 확인할 수 있도록 정보를 표출한다.

3. 서비스 트래픽 및 공격 데이터 분석

트래픽 데이터는 슈퍼컴퓨팅서비스 네트워크로 유입되는 DDoS 공격 대응 장비 단에서 수집 하였으며, 사용자 작업에 관련된 데이터는 사용자 작업 모니터링 시스템에서 수집하였다. 그리고 비정상행위를 수행한 공격자 IP 주소 데이터는 슈퍼컴퓨터 위협 관리 시스템에서 수집하였다.

3.1 네트워크 트래픽 분석

슈퍼컴퓨팅서비스 네트워크의 트래픽을 분석하기 위하여 시계열 그래프를 이용하였다. Fig. 4의 상단에 있는 그래프는 2019년도 1월 1일부터 11월 30일까지 11개월간의 인바운드 트래픽(pps)을 시계열로 나타낸 것이다(일별 누락이 있는 부분은 시스템의 정기점검 및 긴급점검으로 해당 일을 제외했기 때문이다). 일별 인바운드 및 아웃바운드 트래픽을 각 각 X_i, Y_i라고 하고 전체 날짜를 N이라고 할 때, 1은 11개월간 인바운드 및 아웃바운드 트래픽의 평균($\overline{X}=\sum_{i=1}^{N} X_i \div N=1$, $\overline{Y}=\sum_{i=1}^{N} Y_i \div N=1$)을 의미한다. 시계열 분석을 통하여 누리온 시스템이 서비스를 시작하고 3월부터 4월 정기점검 전까지 일별 인바운드 트래픽이 평균보다 높이 나타났다가 5월부터 일별 트래픽이 평균 이하로 줄어든 것을 확인할 수 있다. 그리고 9월에 시작하여 지속적으로 평균을 상회하는 날이 많이 나타나는 것을 확인 할 수 있다. 난다. Fig. 4 하단에 있는 그래프는 동일 기간 동안 아웃바운드 트래픽을 시계열로 나타낸 것이다. 트래픽의 평균을 1로 하였을 때 인바운드 트래픽의 분산($S_X^2=\frac{1}{N}\sum_{i=1}^{N}(X_i-\overline{X})^2$)은 0.58이고, 표준편차($S_X$)는 0.76이다. 아웃바운드 트래픽의 경우 분산($S_Y^2=\frac{1}{N}\sum_{i=1}^{N}(Y_i-\overline{Y})^2$)은 1.84이고, 표준편차($S_Y$)는 1.36이다. 인바운드 트래픽 보다 아웃바운드 트래픽의 분산과 표준편차가 크게 나타난다. 이는 아웃바운드 트래픽이 평균대비 편차가 심한 것을 확인할 수 있다. 이는 평상시에는 아웃바운드 트래픽이 적은데 비해 사용자들이 슈퍼컴퓨터를 이용하여 계산을 수행하고 결과 파일을 외부 네트워크로 전송할 때 아웃바운드 트래픽이 증가하였기 때문으로 이해된다.

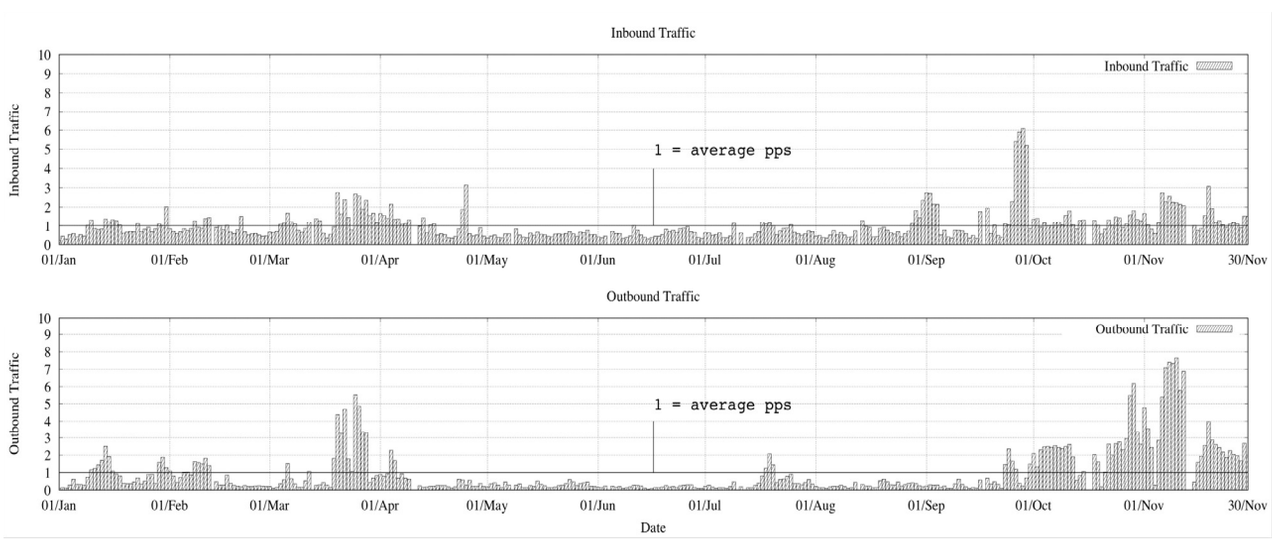

Fig. 4. Time Series Graph of Lnbound and Outbound Traffic(pps) in the NURION Service Networks

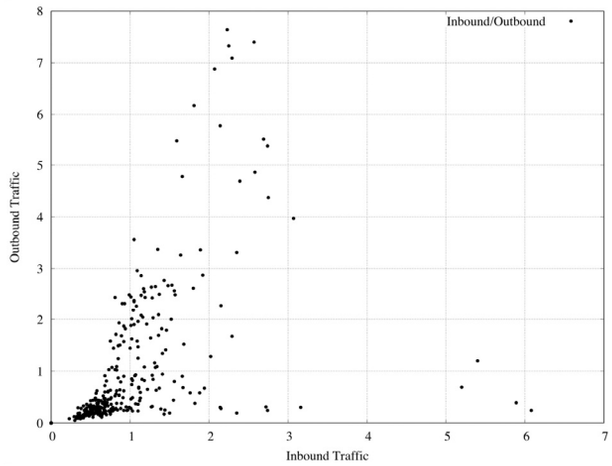

Fig. 5. Correlation Graph Between Lnbound and Outbound Traffic

Fig. 5는 일별 인바운드 트래픽과 아웃바운드 트래픽의 상관관계를 분석하기 위하여 그래프로 나타낸 것이다. 그래프에서 보는 것과 같이 인바운드와 아웃바운드 트래픽은 우 상향 관계로 인바운드 트래픽이 많으면 아웃바운드 트래픽이 많이 나타난다. 상관관계의 정도를 확인하기 위하여 보편적으로 사용되고 있는 피어슨 상관계수(coefficient of correlation) [10, 11]를 Equation (1)을 이용하여 계산하면 0.47 이다. 즉 양의 상관관계를 갖는 것을 확인 할 수 있다.

$$C_{XY} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{S_X S_Y} \quad (1)$$

다만 아웃바운드 트래픽이 인바운드 트래픽보다 편차가 큰 이유는 위에서 기술하였듯이 슈퍼컴퓨터를 이용하여 연산을 수행하고 그 결과를 사용자의 로컬 시스템으로 전송함에 따른 것으로 해석된다.

3.2 사용자 작업 및 공격지 IP 주소 분석

Fig. 6 상단의 그래프는 누리온에 제출된 계산과학 작업의 개수를 시계열 그래프로 나타낸 것이다. 1은 11개월간 제출된 작업의 일일 평균 개수를 나타낸다. 하단의 그래프는 누리온 시스템의 계산노드 사용률을 시계열로 나타낸 것으로 1은 누리온에 있는 나이츠랜딩 노드와 스카이레이크 노드 8,437 계산노드 모두가 사용되는 것을 의미한다.

그래프에 보는 것과 같이 서비스를 개시하고 3개월간 지속적으로 제출된 작업의 개수가 증가한 것을 확인할 수 있다. 이 기간 시스템 사용율도 증가한 것을 확인할 수 있다. 4월에 갑자기 제출된 작업의 개수가 증가한 이유는 공공연구기관에서 서비스 이용을 신규로 신청하고 작은 작업 다수를 전용으로 사용하는 노드에 제출하였기 때문이다. 시스템 사용율이 급증하지 않은 이유는 제출된 작업의 규모가 크지 않아 적은 노드에서 실행되었기 때문을 확인할 수 있다. 이 기간 Fig. 4의 트래픽을 보면 인바운드나 아웃바운드 트래픽에 크게 변화가 없는 것을 확인할 수 있다. 즉 사용자가 접속을 유지하면서 작업을 제출하였기 때문으로 이해된다. 7월부터는 평균 이상 지속적으로 작업이 제출되는 것을 확인할 수 있다. 이를 확인하기 위하여 제출된 계산과학 작업의 누적평균(Cumulative Average)를 시계열 그래프로 나타내면 Fig. 7과 같다. 5월부터 소폭 감소한 작업의 개수가 7월부터 다시 증가한 것을 확인할 수 있다. 그래프에서 1월 1일 제출된 전체 작업의 개수를 1로 하였다. 제출된 작업의 누적평균이 14배 이상 증가한 것을 확인할 수 있다. 이를 통하여 사용자들이 시스템을 잘 활용하고 있음을 확인할 수 있다. 또한 Fig. 6 하단의 시스템 사용률 그래프에서 11월 정기점검(데이터 누락 부분)을 전후로 제출된 작업의 개수는 평균 수준을 유지하는데 시스템 사용률은 0.9 이상을 나타낸다. 이는 사용자들의 작업 규모가 확대되어 시스템 전체 노드에 전개하여 계산과학 작업이 수행되기 때문으로 거대 문제 해결을 위한 슈퍼컴퓨터의 사용

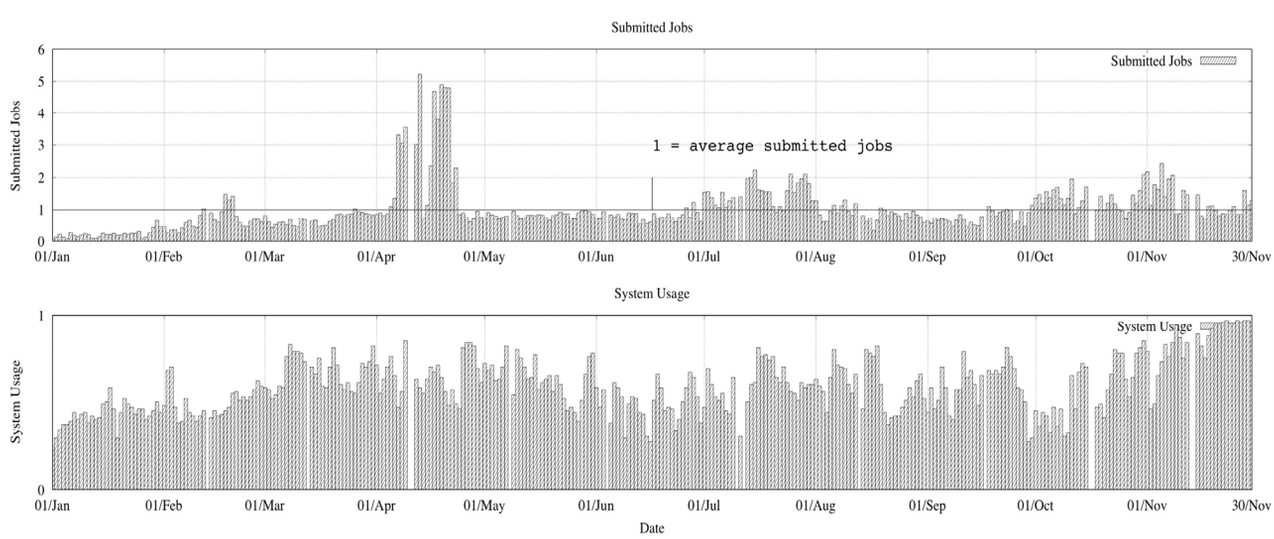

Fig. 6. Time Series Graph of Submitted Jobs and System Usage in the NURION

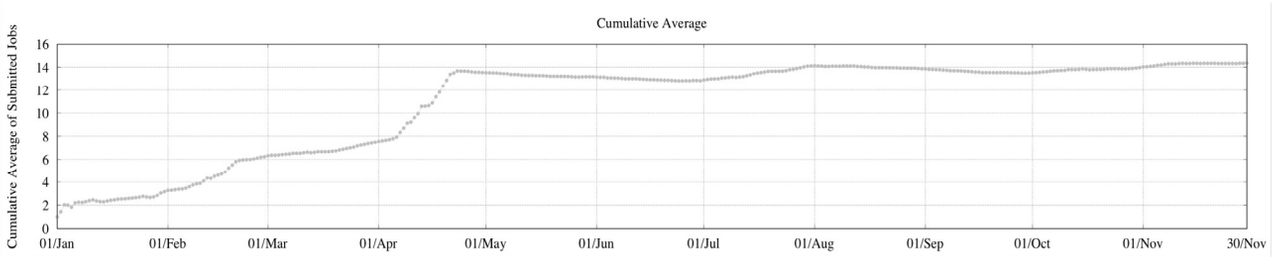

Fig. 7. Cumulative Average Graph of Submitted Jobs

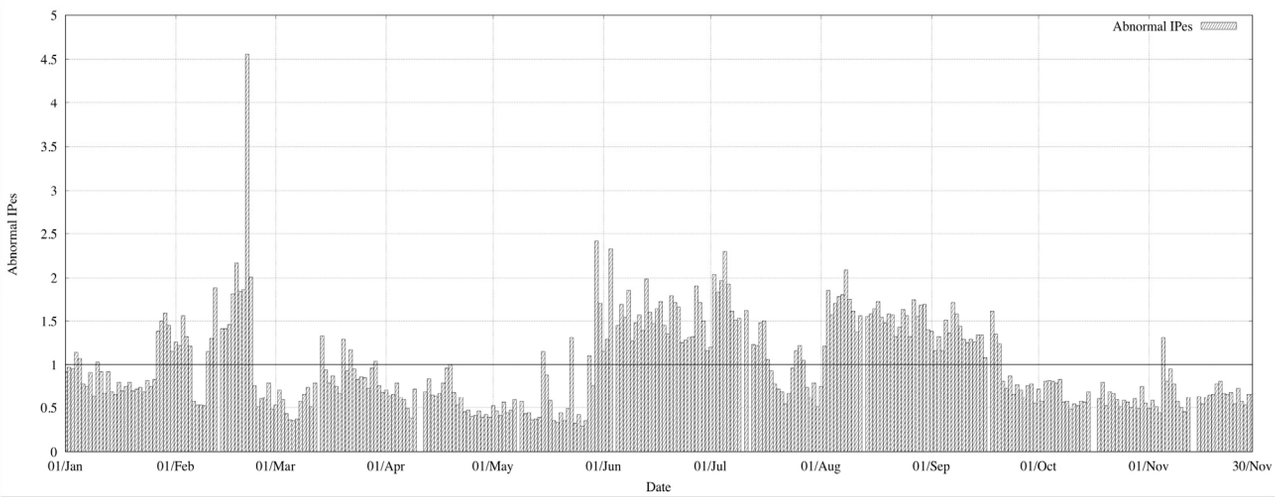

Fig. 8. Detected Attack IP Addresses

목적에 더 부합한다고 할 수 있다.

Fig. 8은 2019년도 1월부터 11월까지 11개월간 누리온 시스템을 대상으로 비정상행위를 수행하여 슈퍼컴퓨터 위협관리 시스템을 통해 탐지/차단된 공격지 소스 IP의 개수(동일한 IP는 한 건으로 처리)를 일별로 나타낸 것으로 기존 그래프와 같이 11개월의 평균을 1로 나타내었다. 서비스를 개시하고 한 달 정도 뒤인 2월에 많은 공격지 IP가 탐지된 것을 확인할 수 있다. 공격자들이 새로운 서비스가 오픈되어 많은 공격을 시도 했던 것으로 추정된다. 그러나 한 가지 특이 사항은 Fig. 4에서 인바운드 트래픽이나 아웃바운드 트래픽이 평균보다 적었던 6월부터 9월까지 탐지된 공격지 소스 IP의 개수가 평균을 상회하는 것을 확인할 수 있다. 그리고 10월 이후에도 트래픽이 평균이상으로 유입되지만 탐지된 공격지 소스 IP의 개수는 평균 이하로 오히려 적은 것을 확인할 수 있다.

인바운드 트래픽과 공격지 소스 IP 개수와의 상관관계 분석을 위하여 그래프로 나타내면 Fig. 9와 같다. 공격지 IP 개수의 평균을 $\overline{Z}$라 하고 표준편차를 S_Z라 할 때 인바운드 트래픽과 공격지 IP 개수와의 상관관계를 나타내는 피어슨 상관계수 C_{XZ}는 Equation (2)와 같다.

$$C_{XZ} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Z_i - \overline{Z})}{S_X S_Z} \quad (2)$$

인바운드 트래픽이 많으면 공격지 소스 IP의 개수도 증가할 것으로 예상되었으나 Fig. 9와 같이 상관관계를 찾기가 힘들다. 피어슨 상관계수도 -0.16으로 상관관계를 명확히 정의하기가 쉽지 않음을 확인할 수 있다. 이는 최근 공격이 정상 트래픽과 구별이 쉽지 않으며 대량의 공격보다는 정상상태와 유사한 형태의 공격이 발생하고 있음을 유추할 수 있다.

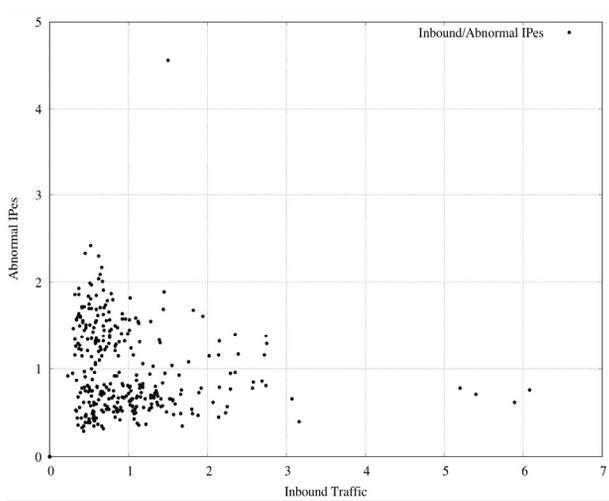

Fig. 9. Correlation Graph between Inbound Traffic and Detected Attack IP Addresses

4. 결론 및 향후 연구 과제

KISTI는 국가슈퍼컴퓨팅시스템인 누리온을 통하여 HPC 서비스를 제공하고 있다. 향후 누리온은 기계학습 및 인공지능, 빅데이터 분석 분야로 확대할 예정이며 사용자의 접근성을 용이하게 하게 위하여 클라우드 서비스와 연계할 예정이다. 네트워크 트래픽을 모니터링하고 분석하는 것은 시스템을 운영하고 관리하는 관점에서 시스템이 안전하고 안정적으로 운영되고 있는지 확인할 수 있는 기초가 자료가 될 뿐만 아니라 향후 계획하고 있는 서비스가 미치는 영향을 확인할 수 있는 기본 데이터가 된다.

본 논문에서는 국가 슈퍼컴퓨팅시스템인 누리온이 유·무상 서비스를 시작하고 11개월 동안 DDoS 대응 시스템 및 슈퍼컴퓨터 위협 관리 시스템, 사용자 작업 모니터링 시스템에서 수집한 네트워크 트래픽 데이터, 공격지 IP 주소 탐지·차단 데이터, 사용자 제출 작업 데이터, 시스템 사용률 데이터를 이용하여 시계열 분석과 상관관계 분석을 수행하였다. 분석을 통하여 인바운드 및 아웃바운드 트래픽은 양의 상관관계를 갖고 있으며 상관계수도 0.47로 나타난 것을 확인하였다. 사용자 작업 및 시스템 사용률 분석을 통하여 시스템이 보다 안정적으로 운영되고 있으며 작업의 크기도 점점 확대되고 있음을 확인하였다. 또한 트래픽 데이터와 공격지 IP 주조 탐지/차단 데이터 분석을 통하여 최근의 공격들이 시스템의 가용률을 낮추기 위한 목적의 공격보다는 실제 서비스 트래픽과 구분하기 어려운 표적 공격(Cyber Target Attack)의 형태를 취하고 있음을 유추할 수 있었다. 향후 트래픽의 헤더 정보 및 이상행위탐지 이벤트 정보, 사용자 작업 정보 등에서 특징(feature)을 추출하고 이것을 기반으로 딥러닝 모델을 적용하여 보고 추가적인 이상행위 특징을 탐지하는 연구로 확대할 계획이다.

References

[1] KISTI National Supercomputing Center [Internet], https://www.ksc.re.kr

[2] Bu Young Ahn, Ji Hoon Jang, Sun Il Ahn, Myung Il Kim, Noo Ri On, Jong Hyun Hong, and Sik Lee, "Study of High Performance Computing Activation Strategy," *International Journal of Multimedia and Ubiquitous Engineering*, Vol.9, No.6, pp.59-66, 2014.

[3] Jae-Kook Lee, Sung-Jun Kim, Chan Yeol Park, Taeyoung Hong, and Huiseung 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, Jul. 2016.

[4] Jae-Kook Lee, Sung-Jun Kim, and Taeyoung Hong, "Brute-force Attacks Analysis against SSH in HPC Multi-user Service Environment," *Indian Journal of Science and Technology*, Vol.9, No.24, pp.1-4, Jun. 2016.

[5] Jae-Kook Lee, Sung-Jun Kim, Joon Woo, and Chan Yeol Park, "Analysis and Response of SSH Brute Force Attacks in Multi-user Computing Environment," *KIPS Transactions on Computer and Communication Systems*, Vol.4, No.6, pp.205-212, Jun. 2015.

[6] Alessandro D'Alconzo, Idilio Drago and Andrea Morichetta, "A Survey on Big Data for Network Traffic Monitoring and Analysis," *IEEE Transactions on Network and Service Management*, Vol.16, No.3, pp.800-813, Sep. 2019.

[7] A. Callado, Carlos Kamienski, Geza Szabo, Balazs Peter Gero, Judith Kelner, Stenio Fernandes, and Djamel Sadok, "A survey on Internet traffic identification," *IEEE Commun. Surveys Trts.*, Vol.11, No.3, pp.37-52, 2009.

[8] I. Drago, M. Mellia, and M. Crovella, "Studying interdomain routing over long timescales," *Proceedings of the 2013 Conference on Internet Measurement Conference*, Oct. 2013.

[9] Sung-Jun Kim and Taeyoung Hong, "Implementation supercomputer system dashboard using RESTful API," *The KIPS Fall Conference 2019*, Vol.26, No.2, Nov. 2019.

[10] Donghwoon Kwon, Hyunjoo Kim, Jinoh Kim, Sang C. Suh, Ikkyun Kim, and Kuinam J. Kim, "A Survey of Deep Learning-based Network Anomaly Detection," *Cluster Computing Journal*, Vol.22, No.1, pp.949-961, 2019.

[11] Mukrimah Nawir, Amiza Amir, Naimah Yaakob, and Ong Bi Lynn, "Effective and efficient network anomaly detection system using machine learning algorithm," *Bulletine of Electrical Engineering and Informatics*, Vol.8, No.1, pp.46-51, Mar. 2019.

[12] R. Artusi, P. Verderio, and E. Marubini, "Bravais-Pearson and Spearman correlation coefficients: meaning, test of hypothesis and confidence interval," *The International Journal of Biological Markers*, Vol.17, No.2, pp.148-151, 2002.

[13] Douglas G. Bonett and Thomas A. Wright, "Sample size requirements for estimating pearson, kendall and spearman correlations," *Psychometrika*, Vol.65, No.1, pp.23-28, Mar. 2000

[14] Thuy T.T. Nguyen, and Grenville Armitage, "A survey of techniques for internet traffic classification using machine learning," *IEEE Communications Surveys & Tutorials*, Vol.10, No.4, pp.56-76, 2008.

이 재 국

https://orcid.org/0000-0002-6159-3124

e-mail : jklee@kisti.re.kr

2002년 충남대학교 컴퓨터과학과(학사)

2004년 충남대학교 컴퓨터과학과(석사)

2012년 충남대학교 컴퓨터공학과(박사)

2010년 ~ 2013년 한국인터넷진흥원 책임연구원

2013년 ~ 현 재 한국과학기술정보연구원 슈퍼컴퓨팅인프라센터 연구원

관심분야 : 시스템 및 네트워크 보안, HPC 시스템

김 성 준

https://orcid.org/0000-0002-7423-4542
e-mail : sjkim@kisti.re.kr
2000년 한남대학교 컴퓨터공학과(학사)
2002년 한남대학교 컴퓨터공학과(석사)
2013년 충남대학교 컴퓨터공학과(박사수료)
2004년 ~ 현 재 한국과학기술정보연구원 슈퍼컴퓨팅인프라센터 선임연구원
관심분야 : 로그 분석, 시스템 모니터링

홍 태 영

https://orcid.org/0000-0002-1841-4485
e-mail : tyhong@kisti.re.kr
1999년 성균관대학교 물리학과(학사)
2002년 성균관대학교 물리학과(석사)
2004년 ~ 현 재 한국과학기술정보연구원 슈퍼컴퓨팅인프라센터 센터장
관심분야 : HPC 시스템, 병렬파일시스템