

ScienceDMZ 기반 초고속 빅데이터 전송을 위한 시스템 과 네트워크 통합 성능 모니터링 환경 개발

김동학*, 문정훈(교신저자)**, 이상권**, 박종선**, 김병서***

*홍익대학교 전자전산학과

**한국과학기술정보연구원

***홍익대학교 소프트웨어융합학과

e-mail : *rlaehdgkr7@naver.com

**{jihmoon, sglee, japark}@kisti.re.kr

***jsnbs@hongik.ac.kr

Developing an integrated System and Network performance monitoring environment for High-speed Big data transfer on ScienceDMZ technology

Dong-Hak Kim*, Jeong-Hoon Moon**, Sang-gwon Lee**, Jong-sun Park**, Byung-Seo Kim****

*Dept. of Electronic Computer Science, Hongik University

**Korea Institute of Science and Technology Information

***Dept. of Software and Communications Engineering, Hongik University

요 약

본 논문은 최근 데이터 집약형 과학분야 연구개발의 비약적인 발전과 관측, 실험, 분석 장비들의 고도화에 따라 생산되는 과학데이터의 빅데이터화, 고부가가치화 등으로 연구 패러다임의 변화가 빅데이터 중심으로 가속되고 있다. 이러한 과학 빅데이터는 ExaByte 급의 대용량으로서 한 곳에서 관리되기보다는 전 세계적으로 분산되어 관리 운영되고 있다. 응용연구자들은 이러한 과학 빅데이터에 대한 초고속 전송/저장/공유에 대한 요구가 높아지고 있으며, 이러한 문제의 해결을 위해 ScienceDMZ 기반의 다양한 고속전송환경이 구축 개발되고 있다. 따라서 본 논문에서는 장거리 빅데이터 전송을 위한 ScienceDMZ의 핵심기술인 DTN(Data Transfer Node)을 통한 빅데이터의 장거리 전송 시 고대역 네트워크 환경과 시스템 성능에 대한 통합 모니터링 환경을 구축 개발하였다.

1. 서론

본 논문에서는 최근 들어 비약적으로 발전하고 있는 빅데이터 분야의 고속전송을 위한 기술개발로서 데이터 집약형 과학분야의 빅데이터는 관측, 실험, 분석 장비의 발전으로 데이터의 양이 폭증하고 있다. 따라서 현재의 네트워크 환경과 전송 기술로서는 여러 가지 제약들로 인해 주어진 대역폭을 최대한 활용하는 고속전송이 어려운 상황이다. 이러한 전송 환경의 개선과 주어진 대역폭 대비 최적의 전송을 위해 미국의 ESnet 에서 제안된 ScienceDMZ 기반의 전송 환경의 구축을 통해 과학 빅데이터의 초고속 전송, 공유 및 저장을 한다^[1]. 과학 빅데이터의 장거리 전송을 위해서는 고대역의 네트워크 환경 뿐만 아니라 네트워크 성능을 급격히 저하시키는 Packet-Loss의 발생을 최소화함으로써 빅데이터의 초고속 전송이 가능하게 되며, 또한 빅데이터의 전송을 위한 전송 가속 시

스템인 DTN의 고속화/최적화 튜닝을 통해 이루어진다. 따라서 전송 가속 시스템인 DTN과 고대역 네트워크에 대한 통합 성능 모니터링 체계는 과학 빅데이터의 효과적 전송, 저장, 공유를 위해 매우 중요하며, 이러한 시스템과 네트워크에 대한 통합형 성능 모니터링 기술은 향후 대륙간 100G 네트워크 기반의 초고속성능 DTN의 구축 개발로 Peta-Scale DTN 구축과 전송을 가능케한다.

따라서 본 논문에서는 네트워크 성능 모니터링을 위해 미국의 ESnet에서 개발한 PerfSONAR를 통해 모니터링 하며, 시스템을 IPMI를 통해 모니터링한다. 또한 PerfSONAR와 IPMI를 통한 실시간 모니터링 결과들은 Grafana를 통해 가시화 시킴으로써 효과적인 통합 성능 모니터링 환경을 구현한다. 특히 PerfSONAR는 Active/Passive Measurement를 동시에 할 수 있으며, 패킷 카운터를 통해 현재 측정 구간의 네트워크 저하 요소인 Packet-Loss의 발생에 따른 적

절한 대응이 가능하며, 또한 IPMI 를 통해 DTN 노드의 CPU, 메모리, 하드디스크, 네트워크 사용 등을 모니터링 함으로써 효과적인 노드 관리가 가능하다. 따라서 본론을 통해 DTN 과 네트워크 환경의 관리를 위해 PerfSONAR 와 Grafana 의 통합 구축과 연동 DB 의 개발, 가시화 환경을 개발 구축하고, 향후 연구방향으로 IPMI 를 통한 시스템 모니터링과 PerfSONAR 결과를 통합한 ScienceDMZ 기반의 통합형 모니터링 환경을 구축 개발하고, ScienceDMZ 확장에 따른 Peta-Scale 전송 환경과 HPC 기반의 ScienceDMZ 환경에 대한 통합형 성능 모니터링 환경에 대한 연구 개발을 한다.

2. 본론

2-1 실험 데이터 수집 방법

2-1-1 Active / Passive Measurement

빅 데이터 전송과 이에 대한 올바른 성능 측정을 통해 빅데이터 전송에 대한 성능 보장과 실시간 대응은 향후 ScienceDMZ 기반 빅데이터의 초고속 전송에 중요한 성능 모니터링 환경이다.

현 인터넷 망에서는 다양한 네트워크 측정 방법들이 연구에 사용되고 있다. 이는 각기 다른 목적을 위해 고안된 방법들로서 대표적으로 Active Measurement 와 Passive Measurement 가 있으며, PerfSONAR 는 이 두 방식을 사용하여 네트워크 성능을 측정한다.

Active Measurement 는 테스트 패킷이나 트래픽을 부가적으로 발생시켜서 네트워크의 성능을 관찰하는 방식으로써 단방향 지연시간(One-way Delay time), 지연 시간 편차(Delay variation), 단방향 패킷 손실(One-way Packet Loss), 패킷 손실 패턴(Packet loss Pattern)이 이에 속한다. PerfSONAR 의 단방향 지연시간 측정 도구인 owping, ping 연결 테스트에 사용되는 ping, traceroute 테스트에 사용되는 traceroute 가 이 방식을 사용하여 측정한다.^[2] Passive Measurement 는 부가적인 트래픽 발생 없이 네트워크에서 흐르는 트래픽을 수집하여 성능을 측정하는 방식으로써 망의 사용량 변화, 트래픽의 처리량(Throughput), 트래픽 특성을 분석할 때 사용된다.^[3] 따라서 처리량을 측정하는 도구인 iperf 와 iperf3 가 Passive Measurement 로 사용된다.^[4]

2-1-2 perfSONAR 에서 네트워크 성능 테스트 구성

<표 1> perfSONAR 에서 설정한 테스트 세부 내용

Test Types	Throughput	Ping	Traceroute
Time Between tests	6 Hours	5 Min	10 Min
Test Duration	20 sec	·	·
Packets per tests	·	10	·
Packet Size (Bytes)	·	·	40

<표 1> 예선 perfSONAR 에서 네트워크 성능을 측정하기위해 구성한 테스트들의 세부 내용을 보여주고 있다. 실험 데이터를 수집하기 위해 다섯 대의 서버

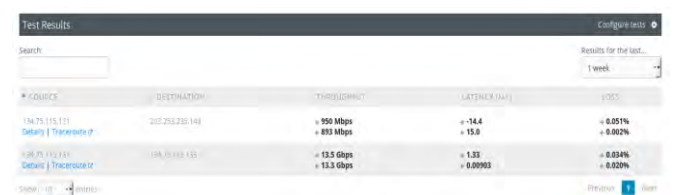
에 네트워크 성능(Throughput) 측정 도구 iperf, iperf3 와 단방향 지연 측정 도구 owping, powerstream 을 이용하여 데이터를 수집한다. Iperf3 을 이용한 성능 측정 간격은 TCP 프로토콜을 사용하여 6 시간 간격으로 1 번의 명령어를 수행하여 측정하며 owping 과 powerstream 은 일정 주기의 네트워크 트래픽을 수집하여 지연이 발생했을 시에 측정한다. 또한, Ping 과 Traceroute 는 부가적인 트래픽을 발생시켜 각각 5 분, 10 분 간격으로 측정한다. 이와 같은 방법으로 구성된 데이터를 바탕으로 분석을 수행한다.

2-2 실험 및 분석 결과

본 실험은 모두 Cent OS 7 운영체제 기반으로 된 Dell PowerEdge R730 모델 3 대(이하 DTN)와 Dell PowerEdge R720 모델 2 대(이하 Lustre)를 사용하였으며 IPMI tool 을 사용하여 원격으로 장비를 제어하며 수행하였다. DTN 은 Mellanox Technologies Ethernet 40Gb/s, Lustre 는 Ethernet 1Gb/s 의 유선 네트워크 망으로 구축되었다. 따라서 본 실험은 총 세 가지 사례로 성능 차이를 관찰하였다. 아래 그림은 Lustre File System 간에 네트워크 성능을 관찰하였을 때 수집된 결과이다.



(그림 1) Lustre 간에 수집된 성능 측정 결과 그래프

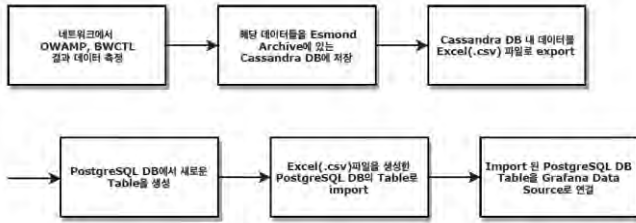


(그림 2) DTN 간에 수집된 네트워크 성능 측정 결과

과란색 실선으로 표현된 데이터가 특정 장비에서 송신했을 때의 처리량이며 과란색 점선으로 표현된 데이터는 수신했을 때의 처리량임을 그림 1 을 통해서 확인할 수 있다. 따라서 Lustre 간 네트워크 평균 처리량(Throughput)은 980Mb/s 로 측정되고 있으며 Packet Loss 가 0.05% 미만으로 거의 발생하지 않는 것을 확인할 수 있으며 DTN 간에 네트워크 평균 처리량은 송수신 때 13Gb/s 로 측정되고 이 또한 Packet Loss 가 0.05% 미만으로 측정되는 사항을 perfSONAR 테스트 세부 결과 화면인 그림 2 를 통해 확인할 수 있다.

2-3 결과데이터들을 그래프로 가시화하기 위한 절차

PerfSONAR 의 Archive software 인 Esmond 내부에는 두 개의 데이터베이스가 존재한다. 새로운 정보를 추가하는데 사용되는 사용자 이름, API Key, Public IP 주소 등과 같은 메타데이터를 저장하는 PostgreSQL 데이터베이스와 측정 결과 데이터를 Cassandra 데이터베이스가 있다^[5]. 2-2 과정을 통해 얻은 실험 측정 데이터들은 모두 Cassandra 데이터베이스에 저장되며 오픈소스인 Grafana 를 사용하여 실시간 네트워크 성능을 표현하기 위해선 이 데이터베이스에 접근^[6]해야 한다.



(그림 3) Grafana 의 Data Source 연결 절차

그림 3 은 Cassandra 데이터베이스에서 Grafana 의 Data Source 로 연결하기 위한 절차를 상세하게 보여주고 있다. 오픈 소스 Grafana 의 Data Source 로 연결할 수 있는 데이터베이스는 매우 한정적^[7]이다. 따라서 일부 데이터베이스만 연결이 가능하기 때문에 Cassandra 데이터베이스에서 Grafana 의 Data Source 로 연결할 수 있는 데이터베이스로 측정된 데이터를 옮기는 과정을 거치게 된다. 이는 Grafana 의 궁극적인 목표인 실시간 모니터링의 장점을 저해할 위험이 발생하지만 이 또한 측정된 데이터들이 저장되어 있는 Cassandra 데이터베이스의 keyspace 를 주기적으로 Excel 파일(.csv)로 export 하고 해당 파일을 PostgreSQL 데이터베이스의 Table 로 import 하는 Python 프로그램을 통해 그 위험성을 최소화시킬 수 있었다.

2-4 측정된 데이터를 Excel(.csv)파일로 추출하는 과정

```

export_excel.py
1 import sqlite3
2 import os
3 conn = sqlite3.connect('tput.db')
4 c=conn.cursor()
5
6 export_ln=""
7
8 for row in c.execute('SELECT * FROM tput_data order by id desc'):
9     for m in range(1,9):
10         export_ln=export_ln+str(row[m])+","
11
12     export_ln=export_ln+"\n"
13
14 f=open("tput.csv","w")
15 f.write(export_ln)
16 f.close
17
18 c.close()
19
20 print(".csv file export done.")
    
```

(그림 4) SQLite3 기반으로 만들어진 데이터베이스 파일을 Excel(.csv)파일로 export 하는 Python 코드

본 실험에서는 네트워크에서 수집된 데이터와 같이 빅데이터를 지원할 수 있는 Cassandra 데이터베이스^[8]를 사용하였다. 하지만 Cassandra 데이터베이스와 같이 특정 데이터베이스에서만 구동되는 것이 아닌 빅데이터를 지원하는 다양한 데이터베이스에서도 호환되어 사용될 수 있는 이식성을 고려하여 그림 4 와 같이 SQLite3 기반으로 측정된 데이터를 Excel(.csv)파일로 추출할 수 있는 Python 코드를 개발하였다.

위 Python 코드를 통해 Excel 파일로 추출된 데이터들을 사용하여 Grafana 에서 Data Source 로 지원하는 다른 데이터베이스로 Import 시킬 수 있으며 특정 데이터베이스에만 적용시켜 결과 데이터를 추출할 수 있는 한계점을 개선하였다.

2-5 Grafana 를 사용한 실시간 네트워크 성능 모니터링

2-4 과정을 통하여 Excel 파일로 추출된 데이터를 사용자가 원하는 데이터베이스에 Import 시키고 이 데이터베이스와 Grafana 의 Data Source 간에 연결이 완료되면 각각의 metrics 들을 생성해야 그래프로 표현할 수 있다. 본 실험에선 PostgreSQL 데이터베이스에 Excel(.csv)파일을 특정 테이블로 import 시켜 구성하였다.

```

SELECT
  $__time(times),
  tput_data as Busan_tput
FROM
  public.tput_data
WHERE
  $__timeFilter(times)
  AND
  host_ad='ps.daeg.kreonet.net'
  AND
  source_ad='ps.busn.kreonet.net'
    
```

(그림 5) Grafana 의 Metric 으로 생성하기 위한 PostgreSQL 의 SQL 구문

Grafana 의 그래프로 표현하기 위해 각각의 metrics 들은 그림 5 와 같은 SQL 구문을 사용하여 표현하였으며 PostgreSQL 데이터베이스에서 실제로 원하는 데이터를 불러올 수 있는 SQL 구문 문법을 사용하여야 정상적으로 그래프를 표현할 수 있다.



(그림 6) Grafana 로 표현된 도시별 KREONET 의 실시간 네트워크 성능 그래프

그림 6 은 KREONET 을 기준으로 국내 주요 도시별 로 존재하는 KREONET 지역망 센터들 간 연결되어 있는 네트워크 망에서 수집된 데이터들을 기반으로 앞서 설명한 그림 5 와 같이 구성된 metrics 들을 하나의 Grafana 그래프에 표현한 실시간 네트워크 성능 그래프이다. 그래프가 다소 각진 면이 있지만 이는 전혀 문제되는 사항이 아니며 사용자가 원하는 대로 그래프의 metrics 들을 부드럽게 표현할 수 있다. 단위는 Gb/s 로 평균 6~8 Gb/s 의 네트워크 처리량이 측정 되는 것을 확인할 수 있다.



(그림 7) ScienceDMZ 기반 글로벌 DTN 성능 측정을 위한 Maddash 테이블

이러한 PerfSONAR 를 통한 ScienceDMZ 환경의 고대역 빅데이터 전송에 대한 전송 성능 모니터링은 향후 그림 7 과 같은 글로벌 ScienceDMZ 환경에서의 대규모 DTN 들간 전송성능과 시스템성능의 효과적인 모니터링을 수행한다.

3. 결론

본 논문에서는 최근 급증하는 과학 빅데이터의 초고속 전송을 위한 아키텍처인 ScienceDMZ 의 핵심 기능인 DTN 의 전송성능 및 고대역 네트워크 환경의 효과적인 성능 모니터링을 위해 PerfSONAR 와 IPMI 를 통한 모니터링 결과를 오픈소스인 Grafana 를 통한 가시화를 구현하였고, 이에 따라 통합형 모니터링 구축을 위한 데이터베이스의 상호연동을 위한 추출, 변환, 삽입 모듈을 개발하였고, 향후 대륙간 100G 네트워크 환경에서의 DTN 을 통한 Peta-Scale 전송 환경에 대한 통합형 모니터링과 세계 각국의 HPC 기반의 ScienceDMZ 들 간의 다양한 전송 환경에 대한 통합형 성능 모니터링에 대한 연구 개발이 필요하다.

ACKNOWLEDGEMENT

본 논문은 2018 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2018R1A2B6002399 호)

참고문헌

- [1] E.Dart et al.. “The ScienceDMZ: A Network Design Pattern for Data-Intensive Science”. Supercomputing Conference 2013. Nov. 17-21. 2013. Denver. CO. USA.
- [2] J. H. Jeong, S. Y. Lee, Y. J. Kim “Methodology and Systems for Internet Traffic Measurement”, Electronics and telecommunications trends, Vol 16. No. 5. p33-35. Oct, 2001.
- [3] J. S. Park, S. H. Yoon, S. S. Yoon “Flow-Based Internet Traffic Measurement Technologies”, Electronics and telecommunications trends, Vol 19. No. 6. p93-104. Dec. 2004.
- [4] Multiple tools of perfSONAR, http://docs.perfsonar.net/info_about.html
- [5] Managing Measurement Archive Data, http://docs.perfsonar.net/multi_ma_backups.html
- [6] How to see a Cassandra Database Keyspace, <http://corona.sdk.tistory.com/851>
- [7] List available as Grafana Data Source, <http://docs.grafana.org/features/datasources/>
- [8] Cassandra takes advantage of Big Table and Dynamo, http://www.kdata.or.kr/info/info_04_view.html?field=&keyword=&type=techreport&page=38&dbnum=176031&mode=detail&type=techreport