

# Data Center Network Characteristic Analysis through a GSDC Case Study

Jin Kim, Ilyeon Yeo, and Kujin Cho

**Abstract**— With the birth of “big data” concept, there are lots of effort to use the data efficiently. The first condition to use those data is gathering of small data in one spot. We call the data spot (or data pool) as data center. Also in the data center, there are a million and one computing facilities in a data center. Those computing devices connected each other through network. In the field of network management, it is interesting question how network traffic is flowed in data center network. In the research, we try to understand data center network traffic flow through GSDC data center case study. For that, we have gathered in/out bound traffic in whole network devices. The GSDC serves computing service and data storage service for various users.

**Research Keywords**—data center network, network management, network traffic treend, gsdc.

## 1 INTRODUCTION

In ancient times, the computer is used to solve difficult problem using required data. Entire system is made for special purpose thing. Nowadays, however, it is more useful that finding of meaningful in data pool. The concept of big data is exactly same as it. A lot of data is stored in data center and a lot of computing units use them. Normally, data center network topology is tree structure[1]. So there are two different traffic patterns which are north-south traffic and east-west traffic. In the data center, north-south traffic means the connection between server and client – the connection might use external/open network -, and east-west traffic means server and server – some batch process and distributed system have to connect each other - . That means, in a data center, several traffic patterns could be mixed in a data center network. In this research, we try to know GSDC network characters because GSDC supports six experiments using batch and grid system.

Nowadays, revolutionary network technologies are rising such as SDN, NFV, etc. The basic idea of these technologies is to control the traffic path. It is easy if system admin know the traffic pattern in their network. It is the starting point of the research.

The structure of paper is as follows: section 1 is instruction; section 2 describes research data set and the analysis result. Finally conclusion and future work are in section 3.

## 2 DATA CENTER NETOWKR CHARACTERISTIC ANALYSIS

### 2.1 Global Science experimental Data hub Center

GSDC is the special purposed data center which supports basic sciences such as nuclear physic and bio, etc. That makes the most remarkable character as compared with other data center in the network aspect. We tried to gather the network in/out bound traffic for 2 years. The network topology of GSDC is spine-leaf structure with 2 backbone spine switch and leaf switch is located on the top-of-rack (TOR).

The bandwidth between spine and leaf switch is 80Gbps and VRRP is enable on the spine switch with 80Gbps. The connection to storage is same. We use the term internal traffic what goes to storage side, and the term of external traffic what goes to other side such as user connection or traffic between servers. The target is uplink traffic volume and every day we have gathered them for two years.

### 2.2 characteristic analysis

In this research, the main focus of analysis is to look at north-south traffic and east-west traffic

- F.A. Jin Kim is with the Korea Institute of Science and Technology Information. E-mail: jkim@kisti.re.kr
- S.B. Ilyeon Yeo is with the Korea Institute of Science and Technology Information. E-mail: ilyeon9@kisti.re.kr
- T.C. Kujin Cho (corresponding author) is with the Korea Institute of Science and Technology Information. E-mail: kcho@kisti.re.kr

movement. GSDC supports seventh experiments; ALICE, CMS, LIGO, HCP, TEM, BIO, RENO. It can be divided in two different groups; grid-using-batch, non-grid-using-batch because L3 analysis is not applied, it is hard to find grid action with our information.

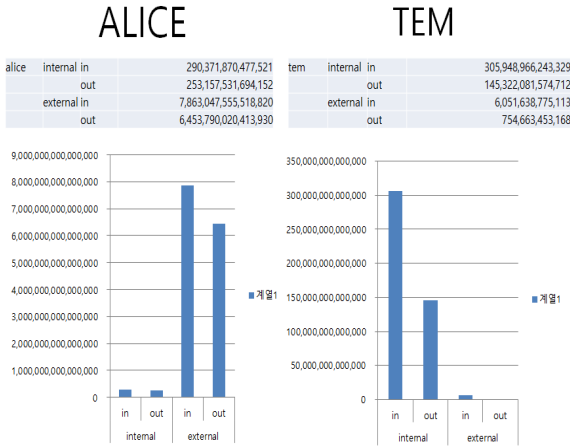


Figure 1 traffic comparison with ALICE and TEM

The name of ALICE is the one of experiments what is the acronym for “A Large Ion Collider Experiment”, one of the largest experiments in the world devoted to research in the physics of matter at an infinitely small scale, and the name of TEM is the acronym for “Transmission Electron Microscope”. GSDC supports TEM experiment which is one of the largest experimental research device using TEM. In the figure1, there are two different shape graph, in the case of ALICE experiment, the system is included in WLCG using Xrootd and computing grid. Because grid system is used for ALICE system, amount of external traffic – out bound – shapes more bigger than amount of internal traffic. In the case of grid job, if there is not required data, then grid searches the data set in other external grid site and download them, and because GSDC is one of ALICE tier 1 center, raw data which is generated I CERN accelerator should be stored in Tier 1 center through dedicated network connection. ALICE system is HTCondor as batch system. On the other hands, in the case of TEM, the amount of internal traffic is more than the amount of external traffic. That means mostly TEM job has to be connect to storage and transfer TEM data. The “internal in” means TEM job requires lots of data in storage and

“internal out” means the result of TEM job generates data which is stored in storage.

In the figure 2, the entire traffic pattern which GSDC use is described. The width means amount of traffic in byte and the length means days. In GSDC, there are two main back bone routers which are E01 and E02. The highest traffic value is E02 device. E02 is charge on ALICE experiment means grid system consume a generous portion of network capability. In bound traffic is more than out bound traffic in E02. It means, after RUN2, the amount of raw data which is coming from CERN is larger than other traffic. The cut part is missing of data, and abnormal value in the fore part is because of data reset.

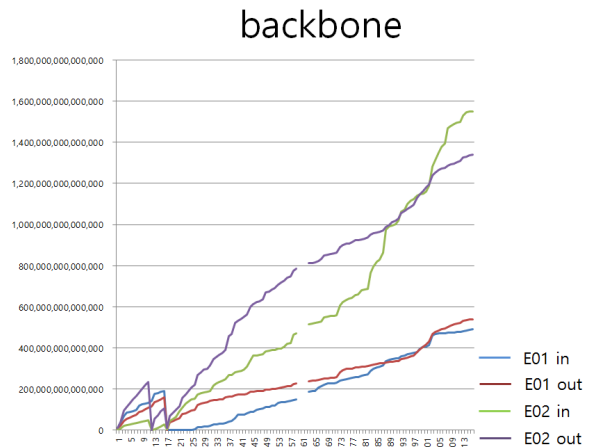


Figure 2 Traffic pattern in back bone network

### 3 CONCLUSION

In this research, we try to find data center network traffic character and analysis them. There are two main characters: first character is, in a data center, if several services are served and the systems are similar, the traffic pattern might be same each other. Second character is, if the system uses grid infrastructure, then the value of east-west traffic is more smaller than the value of north-west traffic. It is useful information if network admin constructs own network design.

The future work of this research is that: first is, the used data set is not used periodic analysis because there are several movements of servers between racks. Second is, there is no L3 analysis. For that, we have a plan to install another network measurement device in GSDC network.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) through contract N-16-NM-CR01-S01 and the Program of Construction and Operation for Large-scale Science Data Center (K-17-L01-C05-S01).

## REFERENCES

- [1] Theophilus Benson, Aditya Akella, David A. Maltz, “network traffic characteristics of data centers in the wild”, the 10th ACM SIGCOMM conference on internet measurement, New York, Nov. 1~30, 2010, pp267-280

**Jin Kim** received the B.S. degree from the college of engineering at KOREA University, in 2008 and the M.S. degrees from graduate school of information technology at POSTECH in 2010. He has been serving as a Senior Researcher of Global Science experimental Data hub Center at Korea Institute of Science and Technology Information (KISTI) since 2010. His research interests include network management, autonomic computing, traffic measurement and analysis, datacenter network, fast data transfer, SDN.

**Ilyeon Yeo** received the B.S. and M.S. degrees from the School of Electronic Engineering at Kyungpook National University, in 2000 and 2002, respectively. He has been serving as a Senior Researcher of Global Science experimental Data hub Center at Korea Institute of Science and Technology Information (KISTI) since 2012. He served as a Senior Researcher of Knowledge Information Center and National Science and Technology Information Service at KISTI since 2002. His research interests include Parallel Computing, Information Retrieval, Database, Grid Computing, and Security

**Kujin Cho** is a senior researcher at Global Science experimental Data hub Center, Korea Institute of Science and Technology Information (KISTI), where he does research on open research data strategy and management & policy of data center.