Computer Science and Information Systems 11(3):1113–1126 DOI: 10.2298/CSIS130916067P

Study on Network Architecture of Big Data Center for the Efficient Control of Huge Data Traffic

Hyoung Woo Park[1,2], Il Yeon Yeo[1], Jongsuk Ruth Lee[1,2] and
Haengjin Jang[1]

[1] KISTI Supercomputing Center,245 Daehak-ro,
Yuseong-gu, Daejeon, South Korea
{hwpark, ilyeon9, jsruthlee, hjjang}@kisti.re.kr
[2] Faculty of Korea University of Science and Technology (UST), 217 Gajong-ro,
Yuseong-gu, Daejeon, South Korea

Abstract. The network architecture of typical data centersis characterized bytiered networks andaggregation-based traffic controls.The emergence of big data makes it difficult for these data centers to incorporate big data service. The tier and aggregation based traffic management systems can magnify the seriousness of the traffic congestion and extend the congested region when big data moves around in the data center. As a consequence, big data has been forcing data centers to change their architecture dramatically. In this paper, we first address the important paradigm shifts of network architecture caused by big data traffic. We then show the new network architecture which resulted from our experience of the CERN LHC data service. Finally, we illustrate the effect of the throughput improvements of the proposed network architecture using a NS2simulation.

Keywords: big data traffic QoS, big data network architecture, big data-front networking, edge traffic separation, big data paradigm shifts.

1. Introduction

If we look into the network architecture of data centers with respect to traffic control, the network architecture features tiered networks and aggregation-based traffic control. Tiered networks mean that networks of data centers consist of backbone networks, sub-networks, sub of sub-networks, and so on. Fig.1. shows the typical tiered network and illustrates that the traffic of the lower tier network is to be automatically aggregated at the upper tier network. Therefore, the tiered network traffic [2][3][4] has a tendency to rapidly flood over all of the networks of the data center. Some studies [5][6] were conducted in order to avoid such situations by making tools to provide multi paths under the tiered network. Tiered networks and traffic aggregation have been useful for data centers to economically construct the network and to efficiently control traffic until now. As the era of big data has arrived, the tiers and the aggregation systems are not functioning well any more. In the big data environment, the tier and aggregation based network architecture magnifies the seriousness of the traffic congestion and may extend the congested region because of the way big data moves around in the data center.

Especially, in case of science big data, scientists tend to move big data from the origin site of the big data to the nearest data center because it is hard for scientists to analyze big data remotely due to the long delay time during the read/write process of big data. After moving big data to the nearest data center, scientists analyze big data with thousands of CPUs that are connected by a very high speed local network in the data center. One well-known big science data is the data from the Large Hadron Collider such as LHC [7] in Swiss CERN. CERN LHC generates multi-peta(10^{15}) bytes of data per year. It is said that a peta-byte of LHC data analysis needs approximately 3000 CPUs. Therefore, science big data centers utilize Grid computing technology to collect thousands of CPUs scattered in the data center and to orchestrate all the gathered-CPUs working together as if they are single supercomputer. As a result, Grid computing can increase traffic in many parts of the data center network. Big data traffic that is caused by big data transmission and big data processing disturbs data center networks more frequently than we imagine. The impact of both traffics is so strong that it can suffocate other services of the data center for quite a long time.

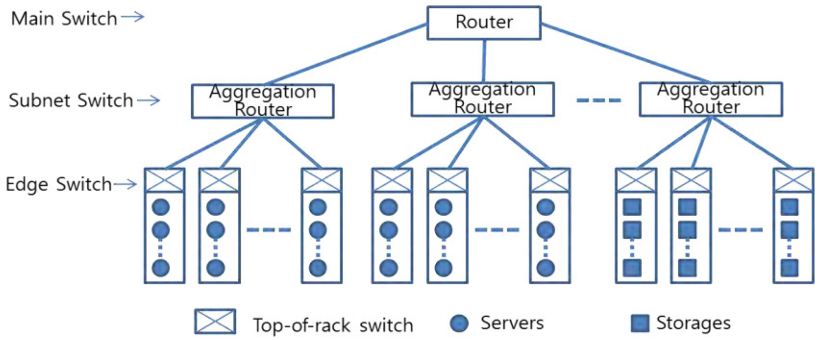

Fig. 1. Typical network architecture of the legacy data center

This paper is the extended version of a previous paper [1]. That paper mainly illustrated the phenomena and the impact of big data traffic from our experience only and coarsely proposed the necessity of a new network architecture for big data. But this paper focuses on showing the details of paradigm shifts due to big data traffic and an analysis of the impact in detail by simulation. For example, this paper shows which part of data centers is destined to change and shows what the architecture of big data centers is like after the changes. In this paper, we try to show the reason why those paradigm shifts happen through the simulations. We will continue this study further for the enhancement of R&E infrastructure [13].

Finally, this paper is consisted of 4 parts. We first described problems caused by big data traffic in local network of big data center. Second, we showed some research activities related with the problem. Third, we suggested some paradigm shifts as a new approach to solve the problem. Finally, we showed the result of simulations for proves of our approach.

2. Problems Caused by Big Data Traffic

First, we introduce some problems that we experienced during big data service. It took 6 months for us to get just 200 TByte data from KEK institute in Japan via Internet even though we have a 10 Gbps international link. During the transmission, the LANs of our data center as well as WAN suffered from the traffic. The Grid computing for data analysis also caused local traffic bursts for long periods. The load for QoS processing of network devices and IP packet filtering of the firewall became gigantic because of the huge volume of big data. Therefore, we had to quickly buy more expensive network devices such as high performance routers as soon as we launched big data service.

The typical features from the perspective of the network administrator for big data traffic are long burst traffic, jumbo IP packet frame and low priority. The LHC data that we serviced as a big data is one of the most well-known big data. The size of the data is almost peta(10^{15})-byte scale data. Therefore, it always took a long time to move it. So, long burst traffic on the network of big data centers is the most typical feature of big data traffic. The second feature, Jumbo frame means a 9K byte packet. It is recommended for the high performance transmission of big data. The size of packets in ordinary Internet usage is generally less than 1.5K byte. Therefore, the effect of packet loss with big data is more serious than that of ordinary Internet packet loss. The third feature, low priority in QoS control of big data traffic, means that the big data traffic should be dropped first, not ordinary traffic, when a congestion of public networks happens. That is because Internet Service providers don't want big data to disturb ordinary Internet traffic.

The scope of the problem mentioned in this paper is limited to the local area traffic of the legacy data center. Generally, storages and file servers are located at the lowest subnet in the tiered network architecture of the legacy data center. Therefore, most parts of the local network of the legacy data center are suffering whenever big data moved from the local storage to the computing servers for the analysis of the data. The aim of the new network architecture is to reduce both of the congested area and the congestion time caused by big data traffic. Therefore, Problems caused by big data traffic can be enumerated by long time of the network congestion, the vast range of the congested region and the strong aggressiveness of big data for occupying the network bandwidth.

Considering big data traffic in the LAN of the legacy data centers once more, their tiered network architecture and the aggregation based traffic management are not the best strategy for traffic management any longer due to the increase of big data traffic. Aggregation based traffic management demands data centers raise the network bandwidth of the data center or enhance QoS function in networking devices. This expenditure continuously increases according to the increase of the volume of big data as we mentioned above. The summary of the problem is that big data analysis as well as big data transfer drops the quality of the data center service because data centers use grid computing for the collection of thousands of CPUs spread in the data center.

3. The Related Researches for the Separation of Big Data Traffic

CERN LHC data is one of big data and CEN LHC produces multi peta(10^{15})-byte data per year. Therefore, it is difficult for a single data center to analyze peta-scale of scientific big data such as CERN LHC data [15][16] within a single data center. Therefore, The CERN LHC data should be moved to multiple data centers over the world. There are 10 data centers [17][18] for the analysis of CERN LHC data. They are called CERN LHC Tier1 centers. Our center (GSDC in KISTI) is one of them. 10 Tier1 centers consist of global infrastructure for LHC data share and analysis computing. Therefore, it is essential for 10 Tier1 centers to work together as if they are single system [19][20][21]. This single system is called WLCG (World-wide LHC Computing Grid). For the analysis of peta-scale data, 10 Tier1 centers have been researching on separating LHC data traffic from the legacy Internet data traffic because peta-scale of LHC data that produced annually severely suffers other traffic. Therefore, related researches that described in this paper are focused on the research activities for building additional infrastructure for the separation of LHC data traffic. We are going to insist that the structure for the separation of LHC data traffic should be extended into the local network architecture of the data center if a data center wants to service CERN LHC data.

To survey related research about the problems mentioned above, we first studied research for the construction of the dedicated network for big data transfer. Among well-known dedicated networks for big data, there is LHCOPN [9], LHCONE [10] and ScienceDMZ [8]. LHCOPN is operated by the research community of CERN LHC data. LHCOPN is a kind of the dedicated optical network and it is built globally by ten data centers and services at 10Gbps for peta-byte data transfer. LHCONE is a kind of dedicated Internet for CERN LHC data. It provides services between Tier 1 centers and Tier2/Tier3 centers. ScienceDMZ has been implemented for local networks of data centers, which use vLAN for building virtually dedicated networks. The difference between LHCONE and ScienceDMZ is that the aim of LHCONE is to support the data transfer between LHCOPN and research organizations. But, ScienceDMZ is mainly used for local networks within a data center. We also surveyed Grid computing technology as related research because we found Grid computing made big data move for big data processing. Grid computing builds a dedicated cluster based computing farm for big data analysis. We surveyed Grid computing, virtual computing [14] and cloud computing [11][12].All of them are used for gathering thousands of CPUs. Therefore, it is inevitable for the legacy traffic and big data traffic to brim over in legacy data centers. If other data centers' resources are collected to use, the range of the network congestion is further extended.

4. The Paradigm Shifts of Network Architecture for Big Data Center

Thus, this paper suggests the new network architecture of data centers for big data. New requirements for the network architecture of big data centers can summarized as follows. First, it must avoid collision between big data traffic and the other traffic when big data move around. Second, it must minimize the region of the big data traffic residence in the data center for the reduction of the influence of big data traffic. Finally, it must reduce the QoS cost that is exponentially rising in proportion to the increase of the volume of big data. We suggest three paradigm shifts of the network architecture to meet the requirements above. These are a paradigm shift on resource provisioning, a paradigm shift on service provisioning, and a paradigm shift on QoS provisioning.

4.1. The 1st Paradigm Shift on Resource Provisioning

The 1st paradigm shift is related to the point of separating big data traffic from the ordinary traffic. The volume of big data is more than a million times of the size of the ordinary data. Therefore, it is impossible to separate big data traffic by simply allocating a virtual circuit because the size of big data traffic is beyond full utilization of the physical network device. Big data traffic always demands full allocation of the capability of the network device for an extended period. For convenience, this paper uses the term, BDC, for Big Data Center. BDC is also used for future data centers. IDC or Internet Data Center is used for legacy data centers.
Fig.2 and Fig.3 show the difference before and after the 1stparadigm shift on resource provisioning. Fig.1 illustrates the shared use of virtual resources by the separation of logical networks in the legacy data center. Fig.2 indicates the proprietary use of physical resources by dynamic allocation.

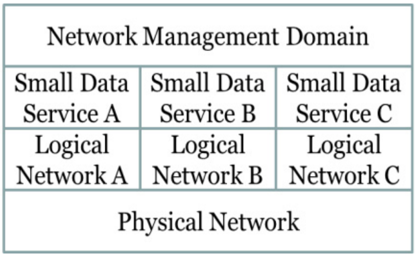

Fig. 2. The share use of virtual resource by the separation of logical network in legacy data center

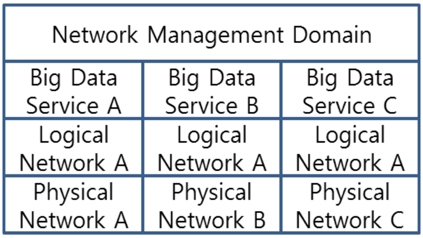

Fig. 3. The proprietary use of physical resource by dynamic allocation for big data center

4.2. The 2nd Paradigm Shift on Service Provisioning

The 2nd paradigm shift on service provisioning addresses the change of the sequence of the service processing. In the legacy data center, legacy data is usually attached the computing server. It is hidden to users. But, in the era of big data, big data should be first, the server for big data processing will be invisible. Users don't to need to know which computers service their job. The typical differences caused by the 2nd paradigm shift on service provisioning are a change from client/server computing to data-driven computing in computing architecture and a change from menu-driven service to user-defined service in service architecture. Fig.4 and Fig.5 show the impact of the 2nd paradigm shift on service provisioning.

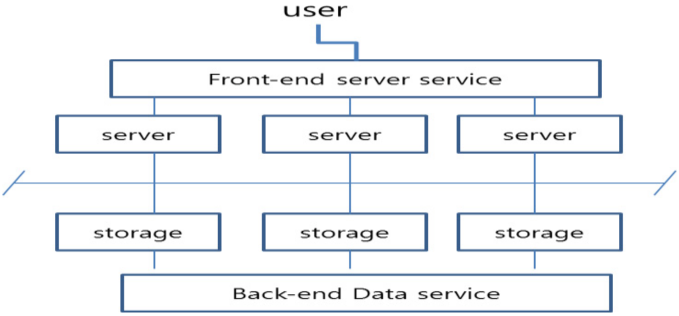

Fig. 4. Traditional client-server service architecture for menu-based service in legacy data center

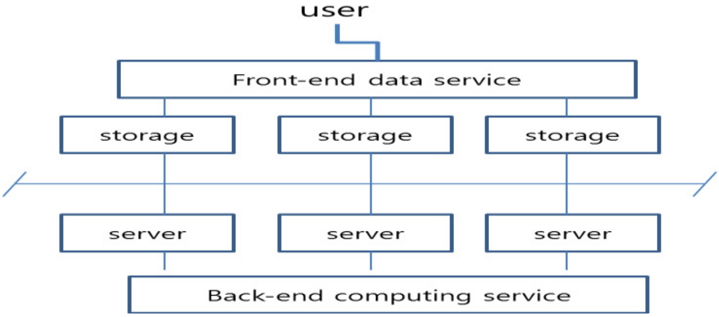

Fig. 5. User-defined service by data-driven computing architecture for big data center

4.3. The 3rd Paradigm Shift on QoS Provisioning

Finally, the 3rd paradigm shift on QoS provisioning is related to QoS initiative. In a legacy data center, QoS initiative belongs to the Internet service provider or network administrator. These kinds of QoS management costs are high in dealing with big data because they have always tried to solve QoS problems by purchasing more expensive network devices which have more performance than that of the existing devices. This paradigm shift suggests moving QoS initiative from ISP to users or end systems. Fig. 6 shows the overload of QoS at each tier in the tiered network architecture. The increase of load of the lower tier increases the load of top tier dramatically. Fig. 7 illustrates the possibility of reduction of QoS load in each tier if part of the load for QoS control is moved from the network devices to end systems. Fig.8 shows traffic flow when traffic separation occurs at end systems.

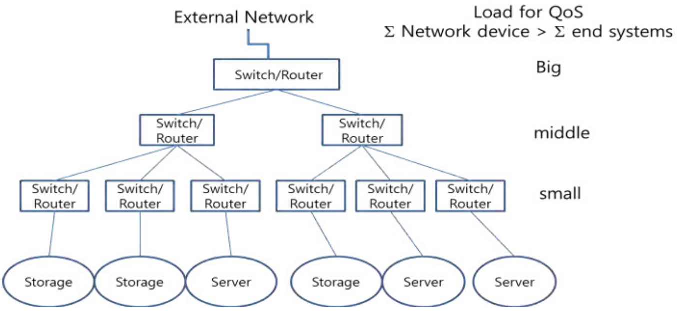

Fig. 6. Tree-like centralized QoS provisioning in traditional data center

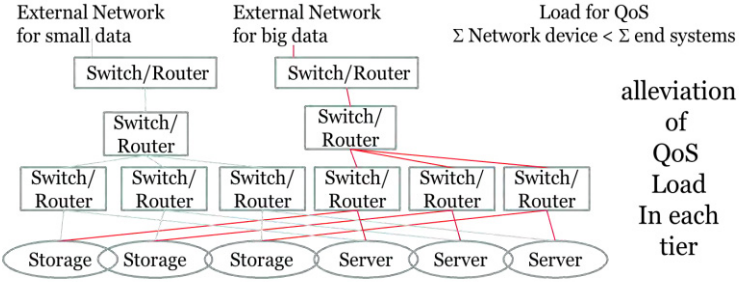

Fig. 7. Traffic separation based QoS provisioning by end system

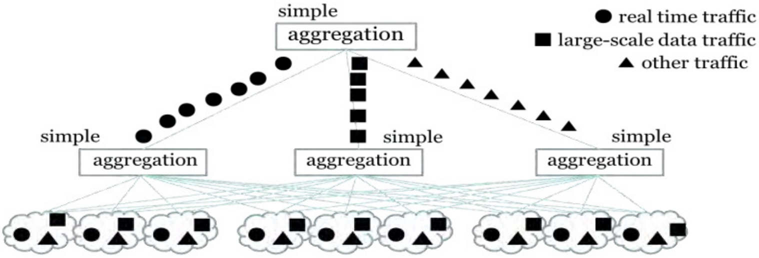

Fig. 8. Traffic separation based QoS provisioning by end system

4.4. The Candidate Network Architecture for Big Data Center

To accomplish these paradigm shifts, we designed anew network architecture for big data centers. We first divided the data center network into 3 parts for the separation of traffic. Part 1 is for big data transfer and sharing, part 2 is for big data analysis computing, and part 3 is for user access and job control. Part1 and part 3 are configured with a public IP address and a private IP address. Part 2 is configured with only a private IP address for security. We also suggest dual interconnection between each part using front end networking and back end networking. These dual interconnections can eliminate the traffic collision between big data traffic and small data traffic. User access is allowed only by front networking. Big data is serviced only by back end networking Fig. 9 shows the implementation example of the new approach.

The key benefit of our approach is a dramatic reduction of the cost for the big data network operation and management. Traffic separating at the end systems reduces the requirement of high performance routers and extends the life of the legacy network devices.

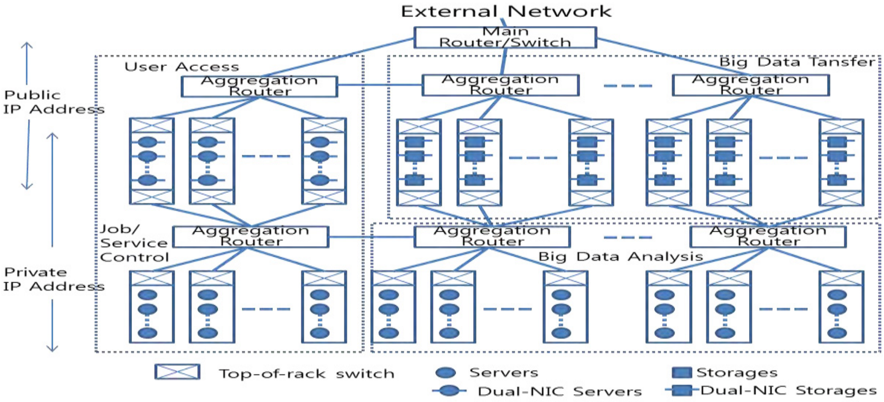

Fig. 9. The candidate architecture for big data center, empowered by organic interrelationship among functional areas

5. Simulations for the Analysis of the Impact of Big Data Traffic

We did the simulation in order to prove the advantage of the paradigm shifts on network architecture mentioned above. It is a well-known fact that traffic separation improves the throughput of the congested network. Therefore, we simulated to show how big data and small data interact when they are co-existing. For this purpose, we set some conditions for the simulations. We set a 9k-byte packet for big data traffic because the 9k-byte frame is strongly recommended by CERN LHC data center. It is also called jumbo frame. For small data, we use the 1.5 k-byte frame. Most ordinary packets are less than 1.5k byte. Big data transfer usually uses TCP protocol. Fig. 10 showed the configuration for the simulation. N0 is a congested node. N1 is a destination node for all of the source nodes. Each link delay is set at 2ms for the simulation of local data center traffic. Except packet size and link delay, we use the default parameter of NS2. We set the volume of traffic to be the same for a reasonable comparison. Table 1 shows the combination of 9k-byte frame and 1.5k byte frame. Simulation is to run for 90 seconds.

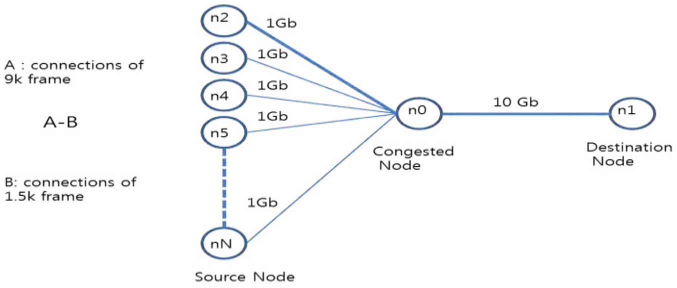

Fig. 10. Simulation for the study of the interrelation between big data traffic and small data traffic

Table 1. The list of the various combinations between big data traffic and small data traffic

type of frame	Combinations							
- no. of sessions with 9k frame	21	18	15	12	9	6	3	0
- no. of sessions with 1.5k frame	0	18	36	54	72	90	108	126
- total No. of sessions	21	36	51	60	81	96	111	126

Fig. 11 shows the amount of the data received at node 1 during the simulation. X-axis stands for the combination for 9k-byte frame and 1.5k-byte frame. The former number is for the number of 9k-byte frames, the latter number for the number of 1.5k-byte frames. We find that the amount of transferred data is decreased according to the increase of the number of total sessions. Therefore, total packet loss in Fig. 12 increased according to the increase of the number of total sessions. One of the interesting results of simulation is shown by Fig.13. The number of packets dropped in a single session can be reduced even though the number of sessions increases when sessions have the same kind (size) of packets. In other words, it is proven that it is better traffic management to classify and to group the traffic into the similar traffic packets.

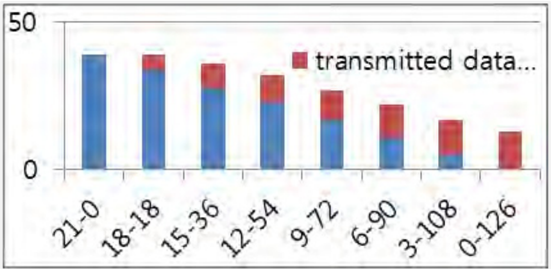

Fig. 11. The total amount of the data received at node 1 during the simulation. Y-axis unit is Mega Byte

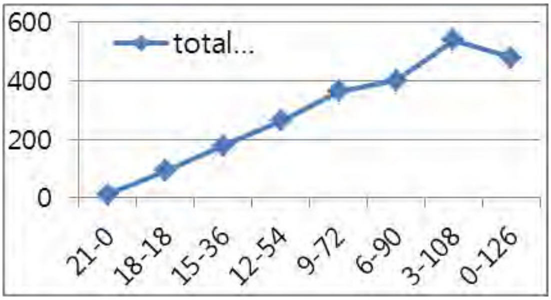

Fig. 12. Total packet drops during simulation. Y-axis unit is number of drops

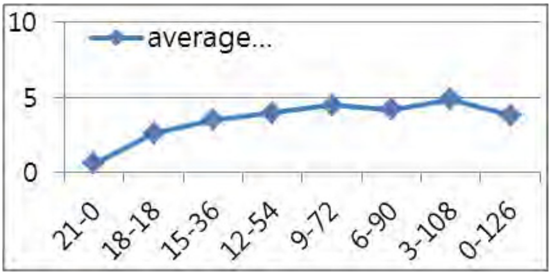

Fig. 13. Average packet drops per sessions. Y-axis unit is number of drops

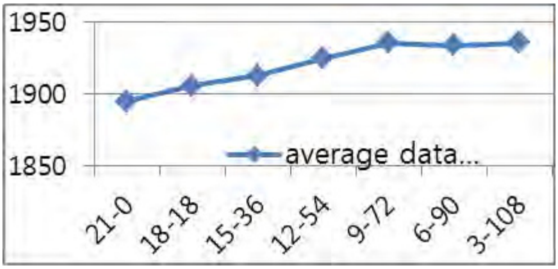

Fig. 14. Average data transmitted by sessions with 9k frames. Y-axis unit is Kilo Byte

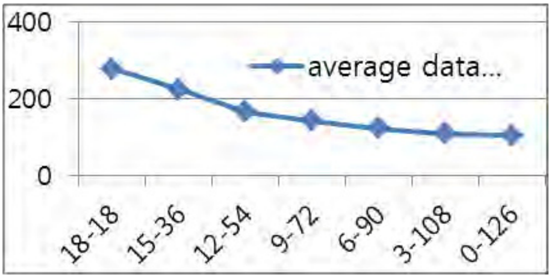

Fig. 15. Average data transmitted by sessions with 1.5k frames. Y-axis unit is Kilo Byte

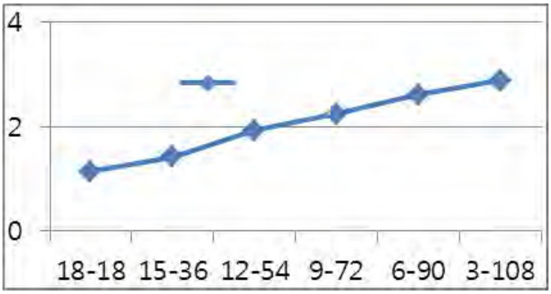

Fig. 16. Relative ratio between transmitted data by the unit size of 9k frame and transmitted data by the unit size of 1.5k frame

Fig. 14 also shows an interesting result. It shows that the amount of transmitted data by the 9k-byte frame is continuously increasing compared with the amount of transmitted data by the 1.5k-byte frame which is decreasing. The decrease is shown in Fig. 15. These phenomena illustrate that big data is stronger than small data in comparison of aggressiveness to occupy network bandwidth. This result can be a proof that the separation of big data traffic from ordinary data traffic is necessary. Fig.16 shows the performance ratio between the performance obtained by small frame and performance obtained by big frame on the same simulation. Small frame means 1.5K byte Packet Data Unit and big frame means 9K byte PDU. Therefore, Fig. 16 also indicates that the aggressiveness of big data is higher than small data. The ratio of the aggressiveness varies from 114% to 288%.

For more detail explain of results, we describe meaning of parameters among result Figures showed in this paper. "A-B" marked in the left of Fig. 10, A-B means the number of a pair (or a set of A and B). One(for A) is the number of sessions for 9k-byte frame transmitted by source nodes and the other (for B) is the number of sessions for 1.5K-byte frame transmitted by the rest of source nodes that are not joined 9k-byte frame transmission. In other words, A is denoted for the number of TCP sessions that transmits 9k-byte frame transmission and B is denoted for the number of TCP sessions that sends data with 1.5k-byte frame. This "A-B" form is used for the value of X-axis in the Fig.12, Fig.13, Fig.14, Fig.15 and Fig. 16. For example, "21" stands for A and "0" stands for B at the first value (21-0) of X-axis in Fig. 11. And, "0" means for A and "126" means for B at the last value (0-126) of X-axis in Fig. 11. This denotation style of the value of X-axis is applied to the all of result Figures from Fig.11 to Fig. 16. The meaning for the value of Y-axis is denoted at the bottom of each Figure. In Fig. 12, the value of Y-axis stands for the number of total packet drops whenever we simulated under the condition of each A-B combination that is denoted at the value of X-axis. The meaning of the result showed in Fig. 12 indicated that packet drops is proportional to the number of sessions that were joined in the simulation and at the same time Fig. 12 indicated that it is relatively less related with the amount of the total data that are transmitted during the simulation. Because we configure NS2 simulation program to send same amount of data for all of simulation even though the pair value of A-B changed. This result is very important results for the management of big data traffic. In Fig. 13, we can find more confidence on that reducing the number of sessions is better for the management of big data traffic. In Fig. 13, the value of Y-axis stands for the average value of packet drops at each combination of A (the number of 9k-byte-frame used sessions) and B (the number of 1.5K-byte-frame used sessions). Therefore, we also reach same result that showed in Fig. 12. The most important result is also showed in Fig. 13. That is, the average packet loss at the both ends of X-axis is lower than average packet loss at the middle of X-axis. It means that average packet loss is decreased abs the ratio of homogeneity of packet size is increased. This result can be also the proof of our proposed architecture. This result is also appeared in Fi. 14. But, the meaning of the value for Y-axis is changed from packet drops to throughput.

6. Conclusion

The goal of this study is to develop new network architecture for big data centers. As we mentioned above, big data traffic will have a major influence in creating paradigm shifts of the system and network architecture of big data centers. These paradigm shifts are related to resource provisioning, service provisioning and QoS provisioning. Therefore, big data will change the architecture of data centers fundamentally. Due to these paradigm shifts, the tiered architecture of IDC will be changed into full-matrix architecture for BDC, and we can also expect that dynamic physical resource allocation will be preferred to the allocation of virtual systems, the decision power of the network path will belong to users not network providers, and the demand for expensive backbone routers can be reduced by edge traffic separation. An interesting feature of our new approach for network architecture is a kind of recycling-friendly architecture because the proposed architecture requires a plentiful number of legacy network cables and legacy low-end network devices instead of buying expensive and cutting-edge network devices.

According to our investigation, the future network architecture of the big data center will be a dual matrix architecture in which the big data part will be located at the front and the center of the architecture in order to reduce the number of interactions between the big data traffic and the legacy traffic. Therefore, the thing that we are going to study further is how we can carry out the transforming of the current network architecture of our data center (GSDC: Global Science Data Center) from tier architecture to data-centered architecture. It is difficult for data centers to change their network architecture without interruption of service. We are building the proposed architecture in parallel. We will first construct the big data share part, and then move the big data analysis part. Finally we will upgrade the user service part. Dual interconnection among the 3 parts will be parallel implemented. Therefore, we will spend much time in rearranging the network devices in order to group similar kinds of traffic onto same network. Finally, we hope this architecture and our experience will help legacy data centers introduce big data service.

Acknowledgements. This work was supported by the program of the Construction and Operation for Large-scale Science Data Center, 2014funded by the KISTI and by the program of the Global hub for Experiment Data of Basic Science, 2014 funded by the NRF.

References

1. Park, H, Yeo I, Lee, J, Jang, H: Study on big data center traffic management based on the separation of large-scale data stream. In the proceedings of the seventh International Conference on Innovative Mobile and Internet Service in Ubiquitous Computing. pp. 591-594. Asia University, Taichung, Taiwan (2013)
2. Kandula, S, Sengupta, S, Greenberg, A, Patel, P, Chaiken, R: The nature of Data Center Traffic Measurement & Analysis. In the proceedings of IMC'09. pp. 202-208. Chicago, Illinois, USA (2009)
3. Benson, T, Anand, A, Akella, A, Zhang, M: Understanding Data Center Traffic Characteristics. ACM SIGCOMM Computer Communication Review, volume 40 issue 1, 92-99 (2010)
4. Benson, T, Akella, A, Maltz, A: Network Traffic Characteristics of Data Centers in the Wild. In the proceedings of IMC'10. pp. 267-280. Melbourne, Australia (2010)
5. Al-Fares, M, Radhakrishnan, S, Raghavan, B, Huang, N, Vahdat, A: Hedra: Dynamic Flow Scheduling for Data Center Networks. In the NSDI'10 Proceedings of the 7th USENIX conference on Networked systems design and implementation.(2010)
6. Benson, T, Anand, A, Akella, A, Zhang, M: MicroTE Fine grained Engineering for Data Centers. In the proceedings of ACM CoNEXT 2011 on emerging Networking experiments and Technologies, ACM. Tokyo, Japan, Article No. 8. (2011)
7. CERN LHC: http://public.web.cern.ch/public/en/lhc/lhc-en.html
8. ScienceDMZ: http://fasterdata.es.net/science-dmz/
9. LHCOPN: http://lhcopn.web.cern.ch/lhcopn/
10. LHCONE: http://lhcone.net/
11. Agrawal, D, Das, S, Abbadi, A: Big Data and Cloud Computing: Current State and Future Opportunities. In the proceedings of EDBT 2011. pp. 530-533. Uppsala, Sweden (2011)
12. Cisco Cloud Computing - Data Center Strategy, Architecture, and Solutions, Point of View White Paper for U.S. Public Sector, http://www.cisco.com/web/strategy/docs/gov/CiscoCloudComputing_WP.pdf
13. Park, H, Kim, S, Lee, J, Jang, H, Cho, K: u-TransitNET: Researches on the user-controlled path-shortening for the reduction of the path-delay in detoured R&E networks. INFORMATION, An International Interdisciplinary Journal, volume 4, 933-944 (2009)
14. http://en.wikipedia.org/wiki/Virtuality_(computing)
15. Juve, G, Rynge, M, Deelman, E, Vockler, J, Berriman, G: Comparing FutureGrid, Amazon EC2, and Open Science Grid for Scientific Workflows. Computing in Science & Engineering, Vol. 15, issue 4, 20-29, (2013)
16. Adamova, D, Horkey, J: An optimization of the ALICE Xrootd storage cluster at the Tier-2 site in Czech Republic. In Journal Conference on Computing in High Energy and Nuclear Physics 2012, NewYork, USA, 1-14, (2012)
17. WLCG(Worldwide LHC Computing Grid) Tier 1 sites, http://lcg-archive.web.cern.ch/lcg-archive/public/tiers.htm
18. Aiftimiei, C, Ceccanti, A, Dongiovanni, D, Meglio, A, Giacomini, F: Improving the quality of EMI Releases by leveraging the EMI Testing Infrastructure. Journal of Physics: Conference Series, Vol. 396, No. 5, 1-12, (2012)
19. EMI (European Middleware Initiative): http://www.eu-emi.eu
20. OSG(Open Science Grid): http://www.opensciencegrid.org/
21. Rodriguez, A, Gouveia, V, Meneses, D, Capannini, Aimar, A, Meglio, A: Article 1. Multi-platform Automated Software Building and Packaging. Journal of Physics: Conference Seriese, Vol. 396, No. 5,1-7, (2012)

Hyoung Woo Park is a Principal Researcher in KISTI Supercomuting center in South Korea. He obtained his Ph. D. in Computer Networks at SungKyunKwan University in South Korea. He has participated lots of R&D projects including the construction of national R&D network (KREONET), the implementation of National Grid Computing Infrastructure, and so on. Currently, He joins the project for construction of peta-scale science data Grid center for the global collaboration researches on CERN LHC data, KEK Belle data etc.

Il Yeon Yeo is a senior researcher at Global Science Data hub Center, KISTI. He obtained his Master degree in Electronic Engineering at Kyungpook National University. He is working for ALICE which is one of the WLCG experiment. He is interested in Grid Computing and Information Retrieval.

Jongsuk Ruth Lee is the corresponding author of this paper. She is a principal researcher and a head of the Department of Advanced Application Environment Development, National Institute of Supercomputing and Networking, Korea Institute of Science and Technology Information (KISTI), Daejeon, South Korea. She is also an adjunct faculty member at University of Science & Technology of Korea, South Korea. She received her Ph.D. in Computer Science and Software Engineering from the University of Canterbury, New Zealand. She, as a researcher, worked for University of Canterbury, NZ from 1998 to 2002, and Korea Electronics and Telecommunications Research Institute (ETRI) from 1992 to 1993. She is the author or co-author of more than 160 research publications, including more than 32 domestic and international patents. Dr. Lee is on the editorial board and a reviewer for various domestic and international journals. Her research interests include parallel computing, distributed simulation, simulation based cyber learning, and grid computing. Board Member of KIPS and KSII.

Haengjin Jang is a director of WLCG Tier1-Korea which is the peta-scale science data Grid Center for the global collaboration researches on CERN LHC data, KEK Belle data etc. He obtained Ph.D. in Computer Science at CHONBUK national university in South Korea. He joined the construction of national supercomputing center since 1988. He was a member of IT expert committee that is managed by Korea Telecommunication Technology Association and worked for Korea business Grid association as the vice chair.

Received: September 16, 2013; Accepted: February 25, 2014.