A test of the interoperability of grid middleware for the Korean High Energy Physics Data Grid system

Kihyeon Cho† ,

Korea Institute of Science and Technology Information, Daejeon, 305-806, Korea

Summary
In this paper, the current architecture of the High Energy Physics (HEP) data grid system in Korea is described. An overview is given for the current and planned uses of grid technology of the Korean Grid Project and especially for computing on HEP. The HEP data grid in Korea has been developed for CDF (Collider Detector at Fermilab), CMS (Compact Muon Solenoid), Belle, and AMS (Alpha Magnetic Spectrometer) experiments. We have first suggested and designed the DCAF(Decentralized CDF Analysis Farm) outside Fermilab for CDF. We have also first constructed interoperability between DCAF and other farms. Therefore, this paper addresses the grid farms for the experiments and the interoperability issues between the farms. The main grid farm locates at the Center for High Energy Physics (CHEP) at Kyungpook National University (KNU) in Korea.
Key words:
grid, computing, network.

1. Introduction

The High Energy Physics (HEP) is to understand the basic properties of elementary particles and their interactions. HEP is usually conducted at the major accelerator sites, in which detector design, construction, signal processing, data acquisition, and data analysis are performed on a large scale. The size of collaboration is 100~ 2000 physicists. To perform computing at the required HEP scale, the data grid is a strong requirement. Grid is the coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations[1]. In this work, we have assembled grid farms for the HEP experiments, CMS (Compact Muon Solenoid) [2], CDF (Collider Detector at Fermilab) [3], Belle [4], and AMS (Alpha Magnetic Spectrometer) [5]. This paper describes computing and interoperability issues for the HEP grid farms in Korea.

2. Objectives

The objective of HEP data grid is to construct a system to manage and process HEP data and to support user group(i.e., high energy physicists). For the current and future HEP activities that require large scale data, the HEP data grid is indispensable and mass storage system of hard disks and tapes in a stable state is necessary. To make the HEP data transparent, CPU power should be extendable and accessible. The transparent HEP data means that the data should be analyzed even if high energy physicists as users do not know the actual place of data. The HEP data grid project in Korea is related with the CDF experiment at Fermilab, the AMS experiment at the International Space Station, the Belle experiment at KEK, and the CMS experiment at CERN. This work focuses on the CMS experiment, in which HEP data grid is an absolute requirement. Finally, for networking, we use multi-leveled hierarchy of distributed servers to provide transparent access to data. Table 1 shows the current network bandwidth between CHEP(Center for High Energy Physics) in Korea and laboratories abroad. Recently connection has established to a 10 Gbps network around the world called GLORIAD (Global Ring Network for Advanced Applications Development) [6].

Table 1: The current network between CHEP in Korea and Laboratories abroad.

Laboratories	***Network***	*Current bandwidth*
KEK (Japan)	Hyeonhae-Genkai [7]	2.5 Gbps
Fermilab (USA)	Hyeonhae-Genkai, TransPAC2 [8]	2 X 2.5 Gbps
CERN (EU)	TEIN [9]	155 Mbps
Fermilab, CERN	GLORIAD	10 Gbps

Manuscript received March 5, 2007
Manuscript revised March 25, 2007

3. Overview

Since March 2002, the Center for High Energy Physics (CHEP) at Kyungpook National University (KNU) has been working with the HEP data grid project collaborating with Seoul National University (SNU), Yonsei University (YU), Sungkyunkwan University (SKKU), Konkuk University, Ewha Womans University (EWU), Chonnam National University (CNU), Dongshin University (DU), Gyeongsang National University (GSNU). Fig. 1 shows the computing infrastructure for HEP data grid in Korea. In order to process the HEP data grid project in Korea efficiently, we study it for both future experiments (CMS, AMS) and current experiments (CDF, Belle). Therefore, the first study is for the construction of the regional data center and the second is to apply the grid technology to the current experiments.

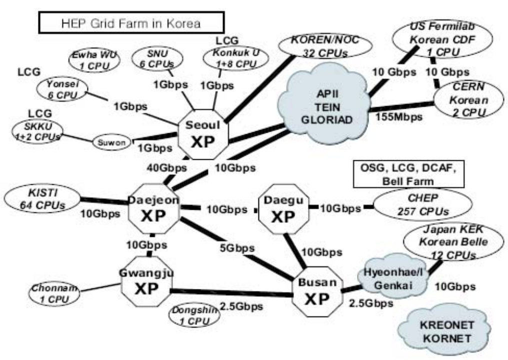

Fig. 1. The computing infrastructure for HEP data grid in Korea. The regional data center will be located at the Center for High Energy Physics(CHEP) at KNU in Korea.

For the area of regional data center, we study the high performance network [10], SRB (Storage Resource Broker) [11], Data Challenge [12], Grid3 [13], and OSG (Open Science Grid) project [14].

For the high performance network, we test networks among Korea, USA, and Europe, and then find the optimized network transfer system [10].

For the SRB, we share the storages at KISTI (Korea Institute of Science and Technology Information) and find the effective data handling system. For the Grid3/OSG, we support the infrastructure of the regional data center along with USA as a member of iVDGL (International Virtual Data Grid Laboratory) [15].

For the application of the HEP data grid, we have first suggested and provided the DCAF (Decentralized CDF Analysis Farm) as a remote farm for the CDF collaboration [16] so that the CDF users around the world may use it like Central Analysis Farm (CAF) [17] at Fermilab. We also study the Belle Data Grid and the AMS Data Grid. We also conduct batch jobs based on Globus for the Belle experiments among SKKU, KISTI, and KNU farms.

4. Achievements

4.1 Introduction

We have constructed and provided the DCAF with the CDF collaborations. Not only users in Asia, such as Taiwan, Japan, and Korea, but also users around the world use the DCAF at the CHEP, which contributes to the CDF experiment. This know-how also contributes to the R&D for the CMS regional data center. We have also constructed both the OSG and the LCG (LHC Computing Grid) farms for the CMS experiment. Therefore, we have studied the interoperability between farms. The main computing facility of Data Grid is located at KNU and other facilities are located at KISTI and Konkuk University.

4.2 Grid Activities with CDF

The CDF is an experiment on the Tevatron at Fermilab, USA. After being thoroughly upgraded, the CDF experiment began its Run II phase in 2001. Up to now, Run II has gathered more than a fb^{-1} of data, equivalent to 1.5×10^9 events or a Pbyte of data, and that is expected to double by the end of the year [18]. More than 700 physicists are involved in this experiment. The CDF computing model is based on the concept of CAF [17]. The CAF is a large farm of computers running Linux with access to the CDF data handling system and databases to allow the CDF collaborators to run batch analysis jobs [17]. The submission uses a CAF portal which has two special things. The first one is to submit jobs from anywhere. The second thing is that job output can be sent directly to a desktop or stored on CAF FTP server for later retrieval.

However, due to the limited resources of CAF, the CDF experiment produced the concept of the CDF Grid. As a first step of grid, we have first suggested and designed the DCAF based on CAF system [16]. A user can submit a job from anywhere to the cluster either at CAF or at the DCAF.

We have first embedded the DCAF outside of Fermilab, in Korea [16]. The Fig. 2 shows a scheme of the DCAF in Korea which we have developed and installed. However, DCAF in Korea is much more difficult than CAF due to data files, which are physically apart by more than 10,000 km. In order to run the remote data that is currently stored in Fermilab, USA, we use SAM (Sequential data Access via Meta-data) [19] which consists of Kerberos rcp, bbftp, gridftp and rcp as transfer methods. We use the same GUI (Graphic User Interface) of CAF. The difference is only to select the analysis farm for the DCAF in Korea. First, we make the CafGui at the submission site. The submit machines can be in any place around the world. The CafGui has several sets of input information. A user can submit a job using CafGui. The DCAF is used as a Monte Carlo simulation and real data analysis by the CDF users around the world. The system contains head node, worker node, SAM station, and job submission site. All the users around the world as well as users in Korea use this farm directly. Even if the data is 10,000 km apart, we can handle it since the bandwidth between CHEP and Fermilab is 10 Gbps. We have handed this know-how to other CDF sites around world. Now, DCAF around world processes 40% of total CPU of the CDF experiment.

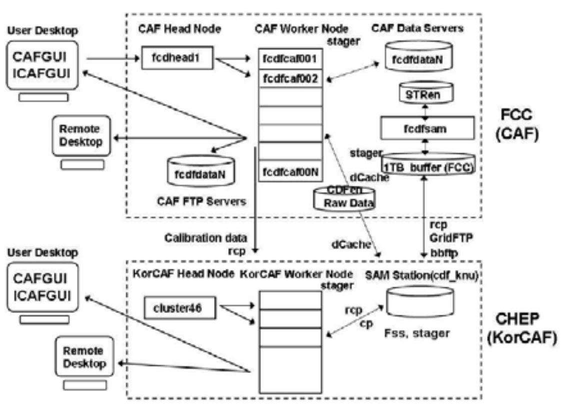

Fig. 2 The design for the DCAF system.

For monitoring system of the DCAF, we use both Ganglia [20] and CAF web monitor system. Fig. 3 shows one month's usage of the DCAF in 2006 by the Ganglia monitoring system.

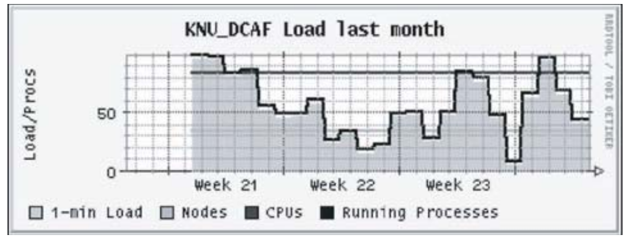

Fig. 3 . The Ganglia monitoring system shows one month's usage of the DCAF in 2006. The total number of CPUs is 84 and the total number of node is 34.

4.2 Grid Activities with CMS

We participate in an international research environment by sending experimental data and participating in the research analysis. The participation in Grid3/OSG project as a member of iVDGL contributes to the CMS experiment for the development of remote site regional data center. We have also constructed the LCG farm for the CMS experiment.

The Grid3 [13] is a multi-virtual organization, application driven Grid Laboratory. In 2003, Grid3 project has deployed for the first time a persistent, shared, multi-virtual organization (VO), multi-application Grid laboratory capable of providing production level services [21]. We have joined Grid3 and installed the test bed. KNU test bed has started with only three nodes for Grid3. One is for the head node and the other two are work nodes. However, we added 81 CPUs for the Grid3 work node using grid interoperability. Recently Grid3 farm is moved to OSG farm.

As shown in Fig. 4, the LCG monitoring system shows one
month's usage of LCG farm in 2006. The number of maximum total CPUs is 28 and the number of maximum free CPUs is 27.

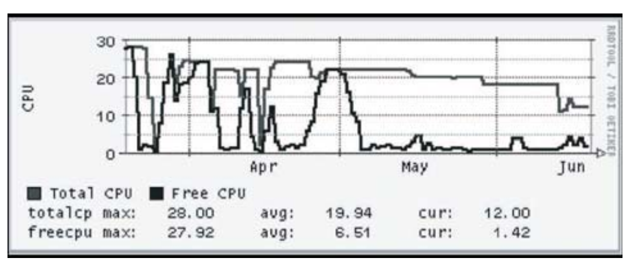

Fig. 4. The LCG monitoring system shows one month's usage of LCG farm in 2006. The number of maximum total CPUs is 28 and the number of maximum free CPUs is 27.

As shown in Fig. 5, the Ganglia monitoring system shows one month's usage of OSG in 2006. The number of total

CPU is three and the number of total node is three.

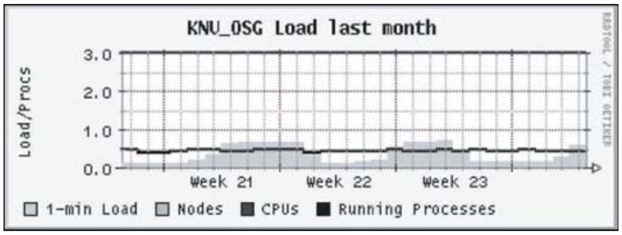

Fig. 5. The Ganglia monitoring system illustrates one month's usage of OSG in 2006. The number of total CPU is three and the number of total node is three. The number of average running processes is about 0.5.

4.3 Interoperability

The CMS experiment is deeply involved in several grid test beds, both in the USA and in Europe [22]; the Korea CMS group is also engaged in both test beds. Therefore, we test program and integration of grid tools in its computing environment.

The USA CMS grid environment is the part of the OSG infrastructure. The national and international grid infrastructures (e.g., OSG and the LCG Projects) provide the basis of the worldwide CMS computing system. The USA CMS grid software infrastructure is collaboration across CMS, with the Fermilab Computing Division, with other physics groups in the USA, and with computer scientists at various universities [23]. In the USA, work is ongoing to deploy the MOP (MOnte carlo distributed Production) tool for the distributed production of CMS Monte Carlo [24]. This system uses Globus [25], Condor-G [26], and DAGMan (Directed Acyclic Graph Manager) submission client [24]. In Europe, the CMS experiment is actively evaluating the middleware developed by the EGEE project [27].

Since the CMS's aim is to build a unique grid and the Korea CMS group has both test beds, we study interoperability for both test beds. The CMS software and computing activities are currently largely focused on allowing the massive Monte Carlo productions. The computing side of the CMS experiment is strongly involved in the integration of grid tools in its computing environment. We intend to use some of these tools in the next Monte Carlo production and also to investigate a remote data analysis.

Among several methods available for the regional data center facility to interoperate with multiple grid installations, Korea CMS group is working on deploying multiple sets of interfaces to the same physical resources and ultimately on defining and deploying compatible interfaces between multiple grid projects.

As shown in Fig. 6, we first succeeded in constructing the dual batch operating systems between Grid3 system and the DCAF system. In 2004, we added 81 CPUs for the Grid3 work node. These nodes are not exclusive for the Grid3 machine. This is a test bed for the DCAF system. Therefore, the DCAF system has higher priority than that of Grid3. The Grid3 job will run when the DCAF systems are free. When DCAF jobs are submitted, Grid3 job is suspended and after CDF job is finished, Grid3 job will run again. This is dual batch system that improves the efficiency for the clusters. Two information providers run at Grid3 and CDF job. As the Grid3 view of its jobs, there are three kind of jobs. The first is “Claimed Busy” which means that Grid3 job is executing. The second is “Owner Idle” which means that a local CMS or CDF job is running. The third is “Suspended”, which means that a Grid3 job has been suspended due to priority of local CMS or CDF jobs.

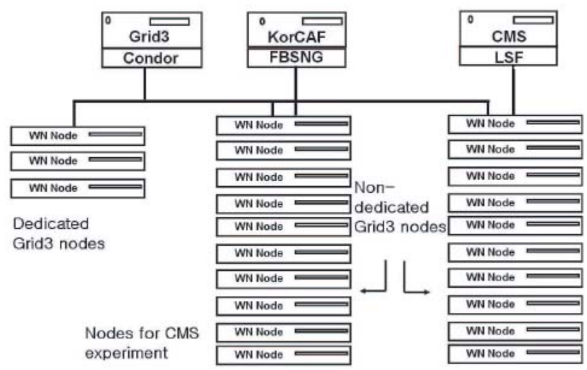

Fig. 6. The dual batch operations systems at CHEP of KNU in Korea.

Table 2 shows an interoperability of CPUs among Grid3, CMS Batch system, and the DCAF.

Table 2: An interoperability of CPUS in 2004.

Grid Farms	***Exp.***	***OS***	***Batch System***	*Current bandwidth*
Grid3	CMS	RedHat7.3	Condor	Interoperability
CMS Batch	CMS	RedHat7.3	LSF	Interoperability
DCAF	CDF	RedHat7.3	FBSNG	Interoperability

Belle Farm	Belle	RedHat9.0	Interactiv e	
AMS Farm	AMS	RedHat9.0	OpenPBS	

Interoperable storage element has been developed in CHEP at KNU using SRB. We succeeded in the SRB federation among CHEP, KISTI, and KBSI (Korea Basic Science Institute) domestically as well as among Korea, Japan, Australia, Taiwan, China, and Poland internationally as shown in Fig. 7. The Fig. 8 shows the SRB performance tests between KEK and other sites - ANU (Australia National University), Krakow (Poland), KNU (Kyungpook National University) and IHEP (China) [28]. The performance between KEK and KNU shows the best performance among them.

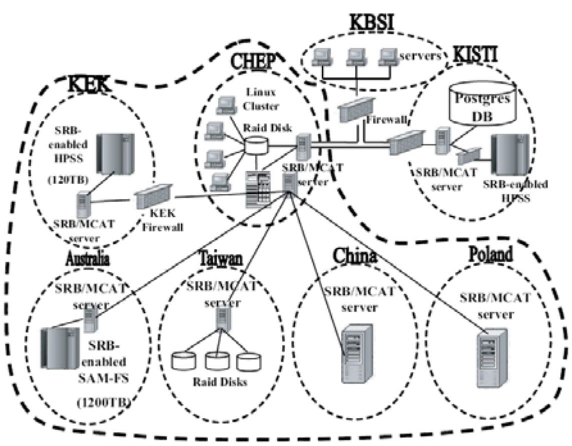

Fig. 7. The SRB federation, which is connected among CHEP, KISTI, and KBSI domestically as well as Korea, Japan, Australia, Taiwan, China, and Poland internationally.

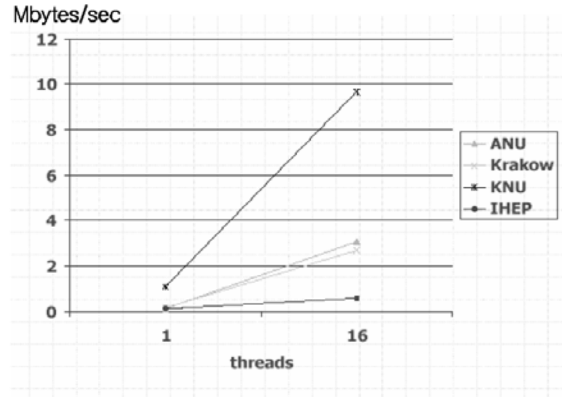

Fig. 8. The SRB performance tests between KEK and other sites - ANU (Australia National University), Krakow (Poland), KNU (Kyungpook National University) and IHEP (China).

Table 3 illustrates a plan of interoperability of CPUs. OSG farm will be interoperability with LCG farm. Other farms, such as the DCAF, the Belle farm, and the AMS farm, will be work nodes as LCG computing elements. Then, each experiment will use each VO based on LCG farm respectively. The LCG storage element information provider will be deployed on the top of the SRB. The tape systems do not have to report how much space they have left.

Table 3: A plan of interoperability of CPUs. We are going to use all work nodes as LCG computing element

Grid Farms	*Exp.*	*OS*	*Batch System*	*Comment*
OSG	CMS	RedHat7.3	Condor	Interoperability
LCG	CMS	SL3.0.4	Torque	
DCAF	CDF	LTS3.0.5	Condor CAF	To be LCG farm
Belle Farm	Belle	RedHat9.0	Interactive	To be LCG farm
AMS Farm	AMS	Debian	Condor	To be LCG farm

5. Conclusions

We succeeded in developing and installing HEP data grid farm in Korea using 257 CPUs for four different experiments (e.g., CMS, CDF, Belle, and AMS) at KNU, Korea. Konkuk University has six CPUs for LCG farm and KISTI has 64 CPUs for the Belle farm, which is currently to be changed to LCG farms.

With different types of grid being developed around the world, the interoperability issue is a very important issue. KNU deployed different grids and cluster systems, and interoperability among them is achieved through deploying multiple batch system to the same physical resources. This work regarding HEP data grid in Korea contributes to the global HEP data grid. In the future, different types of grids will be incorporated into the CMS Grid. This effort will effectively harness all the computing resources through interoperability for the CMS experiment.

Acknowledgments

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD).(R08-2003-000-10258-0). I am grateful to numerous colleagues at HEP data grid working group in Korea for discussing with the HEP data grid project, especially, Daehee Han, Kihwan Kwon, Yuchul Yang, Daejung Kong, DongHee Kim and Dongchul Son.

I also would like to thank to Mark Neubauer (UCSD) and Frank Wuerthwein (UCSD) for CAF (Central Analysis Farm), Harvey Newman (Caltech) for network test between Korea and USA, Paul Avery(Florida) and Bockjoo Kim(Florida) for Grid3 and OSG farms. I also would like to thank to Michael Wan (SDSC), Stephen J. McMahon (ANU), Glenn Moloney(Melbourne), Gang Chen (IHEP), Henryk Palka (Krakow, Poland), Eric H. Yen (ASCC, Taiwan), Nobu Katayama(KEK), Takashi Sasaki(KEK), Yoshimi Iida (KEK) and Yoshiyuki Watase(KEK) for the federations of SRB.

References

[1] I.Foster *el al.*, International J. Supercomputer Applications, 15(3), (2001).
[2] M.Della Negra *el al.*, CMS Physics Technical Design Report, CERN/LHCC 2006-001 (2006).
[3] R.Blair *el al.*, FERMILAB-PUB-96/390-E (1996).
[4] A.Abashian *et al.* Nucl. Instr. and Meth. A 479, 117-232 (2002).
[5] M.Aguilar *et al.* Phys. Reports 366/6 331-404 (2002).
[6] G.Cole, GLORIAD(Global Ring Network for Advanced Applications). International ICFA workshop on HEP Networking, Grid and Digital Divide Issues for Global e-Science, (Daegu, 2005).
[7] Y.J.Park, Korea Activities in Hyeonhae-Genkai Project. APII Workshop, (Tokyo, 2004).
[8] J.Williams, CIREN/TransPAC2 Annual Report (SCI-0441096) (2006).
[9] D.West, Technical Achievements of TEIN2(Trans-Eurasia Information Network), ASEM6 SUMMIT Brie_ng, (Brussels, 2006).
[10] K.Kwon *et al.*, HEP Data Grid and Network Performance Evaluations on Research Networks. In Proc. 18th APAN International Conference Network Research Workshop, (Cairns, 2004).
[11] R.Moore, Integrating Data and Information Management, International Supercomputer Conference, (Heidelberg, 2004).
[12] H.Newman, *et al.*, The Motivation, Architecture and Demonstration of the UltraLight Network Testbed. In Proc. of CESNET (Prague, 2006).
[13] R.Gardner *el al.*, Grid3: An Application Grid Laboratory for Science. In Proc. of Conference on Computing on High Energy Physics, (Interlaken, 2004).
[14] P. Avery, iVDGL / Open Science Grid. International ICFA workshop on HEP Networking, Grid and Digital Divide Issues for Global e-Science, (Daegu, 2005).
[15] R.P.Mount, US Grid Projects: PPDG and iVDGL. In Proc. of Computing for High Energy Physics, (Beijing, 2001).
[16] K.Cho *et al.*, Construction of DCAF in Korea. In Proc. 18th APAN International Conference Network Research Workshop, (Cairns, 2004); A.Sill *et al.*, Globally Distributed User Analysis Computing at CDF. In Proc. of Conference on Computing on High Energy Physics, (Interlaken, 2004).
[17] M. Neubauer *et al.*, Nucl. Instr. and Meth. A 502, 386-390 (2003).
[18] I.S_ligoi, CDF Computing. Conference on Computational Physics, (Gyeongju, 2006).
[19] G.Garzoglio, SAMGrid project. International HEP Data Grid Workshop, (Daegu, 2002).
[20] M.L.Massie, B.N.Chun and D.E.Culler, Parallel Computing, Vol. 30, 817-840 (2004).
[21] I.Foster *et al.*, Production Grid: Principles and Practice. In Proc.13th IEEE Intl. Symposium on High Performance Distributed Computing, (Honolulu, 2004).
[22] C.Charlot, Nucl. Instr. and Meth. A 502, 353-357 (2003).
[23] L.Bauerdick, CMS Computing. International ICFA workshop on HEP Networking, Grid and Digital Divide Issues for Global e-Science, (Daegu, 2005).
[24] G.Graham, MCPS(Monte Carlo Processing Service). International ICFA workshop on HEP Networking, Grid and Digital Divide Issues for Global e-Science, (Daegu, 2005).
[25] I.Foster *et al.*, The Globus Project: A Status Report. In Proc. IPPS/SPDP '98 Heterogeneous Computing Workshop, pp.4-18, (Orlando, 1998).
[26] J.Fery *et al.* Cluster Computing, 5(3), 237-246 (2002).
[27] D.Foster, EGEE(The Enabling Grids for E-sciencE). International ICFA workshop on HEP Networking, Grid and Digital Divide Issues for Global e-Science, (Daegu, 2005).
[28] S.McMahon, Demonstration on SRB. KEK Grid workshop, (KEK, 2004).

Kihyeon Cho received the B.S. and M.S. degrees in Physics from Yonsei University, Korea in 1985 and 1987, respectively. He received the Ph.D in Physics from University of Colorado, Boulder, USA in 1996. During 1996-2001, he stayed in University of Tennessee, Knoxville, USA as a Post-doc. During 2001-2006, he stayed at the Center for High Energy Physics (CHEP), Kyungpook National University (KNU), Korea as a professor. He is now with Korea Institute of Science and Technology Information (KISTI) as a Senior Researcher.