Towards A Software Defined Secure Data Staging Mechanism

Susumu Date, Takashi Yoshikawa, Kazunori Nozaki, Yasuhiro Watashiba, Yoshiyuki Kido, Masahiko Takahashi, Masaya Muraki, and Shinji Shimojo

Abstract Recently, the necessity and importance of supercomputing has been rapidly increasing in all scientific fields. Supercomputing centers in universities are assumed to satisfy scientists' diverse demands and needs for supercomputing. In reality, however, medical and dental scientists who treat security-sensitive data have difficulties using any supercomputing system at a supercomputing center due to data security. In this paper, we report on our on-going research work towards the realization of a supercomputing environment where separation and isolation of our supercomputing environment is flexibly accomplished with Express Ethernet technology. Specifically, in this paper, we focus on an on-demand secure data

S. Date (✉) • Y. Kido • S. Shimojo
Cybermedia Center, Osaka University, 5-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan
e-mail: date@cmc.osaka-u.ac.jp; kido@cmc.osaka-u.ac.jp; shimojo@cmc.osaka-u.ac.jp

T. Yoshikawa
System Platform Research Laboratories, NEC Corporation, 1753 Shimonumabe, Nakahara-ku, Kawasaki, Kanagawa 211-8666, Japan

Cybermedia Center, Osaka University, 5-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan
e-mail: yoshikawa@cd.jp.nec.com; tyoshikawa@cmc.osaka-u.ac.jp

K. Nozaki
Division of Medical Informatics, Osaka University Dental Hospital, 1-8 Yamadaoka, Suita, Osaka 565-0871, Japan
e-mail: knozaki@dent.osaka-u.ac.jp

Y. Watashiba
Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan
e-mail: watashiba@is.naist.jp

M. Takahashi
System Platform Research Laboratories, NEC Corporation, 1753 Shimonumabe, Nakahara-ku, Kawasaki, Kanagawa 211-8666, Japan
e-mail: m-takahashi@ex.jp.nec.com

M. Muraki
Strategic Technology Center, TIS Inc., 17-1, Nishishinjuku 8-chome, Shinjuku-ku, Tokyo 160-0023, Japan
e-mail: muraki.masaya@tis.co.jp

M.M. Resch et al. (eds.), *Sustained Simulation Performance 2017*,
DOI 10.1007/978-3-319-66896-3_2

staging mechanism that interacts with a job management system and a software defined networking technology, thus enabling minimum data exposure to third parties.

1 Introduction

The necessity and importance of supercomputing has been rapidly rising in all scientific fields. This rising necessity and importance can be explained from the following four factors. First, the recent advance in scientific data measurement devices has allowed scientists to obtain high resolution scientific data in both temporally and spatially, and this has resulted in an increasing amount of data obtained with such devices. For example, according to [16], the Large Synoptic Survey Telescope (LSST) is expected to generate 15 TB a day. Also, approximately 15 PB of experimental data is annually generated and processed at the Large Hadron Collider (LHC), an experimental facility for high energy physics [2]. Second, the development of current processors and accelerators has advanced dramatically. Although discussions about 'Post Moore's era' are more and more frequent these days, new-generation processors and accelerators have been continuously researched and released. The Intel Xeon Phi (Knight Landing) processor and the NVIDIA GPU accelerator (Pascal and Volta) are representative examples of such processors and accelerators. NEC's future vector processor is another example of a cutting-edge processor [11]. These cutting-edge processors and accelerators have allowed researchers to perform an in-depth and careful analysis of scientific data observed and measured through the use of scientific measurement devices on supercomputing systems with such processors and accelerators. Third, networking technologies have advanced greatly. Today, 10 Gbps, and even 100 Gbps-class networks are available as world-scale testbeds for scientific research [4, 14, 15]. This advancement means that scientists and researchers can move scientific data more quickly in a research collaboration environment where research institutions, universities, and industries are connected. Furthermore, the potential and feasibility of data movement using advanced networking technology benefits the aggregation and sharing of scientific knowledge and expertise for problem solving. Finally, the needs and demands of high-performance data analysis (HPDA) have led to increased demands on supercomputing. Scientists from every field are enthusiastic about applying computationally intensive artificial intelligence (AI) technologies that exemplify deep learning in their own scientific domains.

Despite the increasing demand on supercomputing, however, the Cybermedia Center (CMC) [9], a supercomputing center at Osaka University in Japan, is facing two serious problems because of the recent diversification of users' requests and requirements from their high-performance computing environments, in addition to the strong user requests and demands for larger computational power. The first problem is the low utilization of supercomputing systems due to an inflexibility in resource configuration, and the second problem is the loss of supercomputing

opportunities due to data security. These two problems hinder the efficient and effective use of supercomputing systems for scientific research.

The first problem can be explained from the many choices scientists have today in terms of the hardware and software for acceleration of their programs and thus, each scientist tends to have his/her own favorite supercomputing environment. For example, some scientists may want to use OpenMP to achieve a high degree of thread-level parallelism on a single computing node equipped with two 18-core processors, while others may want to use MPI to achieve inter-node parallelism on 16 computing nodes. Also, a scientist may want to perform GROMACS [1] on a single computing node with four GPUs, while others may want to perform LAMMPS [8] on a single computing node with two GPUs. Taking these users' diverse requests and requirements into consideration, supercomputing centers like the CMC should have a flexibility in resources to accommodate the diversity and heterogeneity of user requests pertaining to supercomputing systems. Based on this observation and insight described above, our research team has been working on the research and development of a more flexible supercomputing system. We have published the results and achievements obtained so far in [3, 10] and therefore do not present how we have approached this problem in this paper.

The second problem for scientists is how to treat a large amount of privacy-rich and confidential data, especially in the medical and dental sciences, such as is the case at Osaka University. Medical and dental scientists acquire a lot of such privacy-rich and confidential data which they want to analyze with a supercomputer system for the prediction of patients' future disease risks, for assessment of the severity of the patient's case, and for the understanding of brain function, etc. All of these situation require deep learning techniques on a high-performance simulation. Unfortunately, in reality, medical and dental scientists have great difficulties using the supercomputing systems at the CMC because of privacy issues. To assist scientific researches treating this type of security-sensitive scientific data, a technical solution that enables security-sensitive data to be treated is essential from the point of a supercomputing service provider.

In this paper, we present and report the research work in progress for the second problem. More specifically, an on-demand secure data staging mechanism that enables minimum exposure of security-sensitive scientific data, which we have envisaged and been prototyping, is overviewed. Technically, the mechanism leverages Software Defined Networking in cooperation with a job management system for the mechanism. The organization of this paper is as follows. Section 2 briefly overviews the challenge and issues in this early stage of our research. Next, we describe our basic idea to approach this challenge and then we explain the key technologies composing the mechanism in Sect. 3. Subsequently, in Sect. 4, the overview of the mechanism is shown. Section 5 concludes this paper.

2 Challenges and Issues

To tackle data security, we have started a discussion about data security issues with the Division of Medical Informatics at the Osaka University Dental Hospital so that dental scientists can utilize supercomputing systems for their own research. The biggest hurdle and problems to overcome so far are the regulations and guidelines set forth by the Ministry of Health, Labour and Welfare, Japan, which has strictly required organizations and/or scientists that treat privacy-rich data in adherence with the government's regulations and Ministry guidelines. The Division of Medical Informatics at the Osaka University Dental Hospital has set up their own security policy and rules based on their regulations and guidelines for managing controlling their security-sensitive data.

The second hurdle and problem is how we can make a supercomputing environment to include the network between at the Division of Medical Informatics at the Osaka University Dental Hospital and the CMC dedicated only to dental scientists who are willing to use supercomputing systems at the CMC. As described above, the CMC is in charge of delivering a supercomputing environment to researchers in universities and research institutions. Thus, the supercomputing systems are inherently expected to be shared by many scientists and researchers at the same time. Dental and medical scientists, however, do not want to share data and need to securely move data from the hospital to the CMC's supercomputing environment, perform their computation on a supercomputing environment dedicated to them, and move their data and computational results back to their departments. In other words, we, at the CMC, need to find a way to service the privacy of data from the dental hospital.

3 Key Technologies

Our approach to the second problem described in the previous section synergically makes use of three key technologies: Software Defined Networking (SDN), Express Ethernet (ExpEther) technology, and Job Management System to realize a secure and isolated supercomputing environment where dental scientists can perform their own computations. The following subsections explain these three key technologies.

3.1 Software Defined Networking

Software Defined Networking [12, 17] is a new concept of the construction and management of computer networks. Traditional computer networking facilities have a built-in implementation of network protocols such as Spanning Tree Protocol (802.1D) and Tagged VLAN (802.1Q). SDN provides a systematic separation of

two essential functionalities of such networking facilities, namely, data forwarding and decision making. In SDN, the data forwarding part is called the *Data Plane* and the other part, which decides how each data should be forwarded in a network and conveys the decision to appropriate networking facilities, is called the *Control Plane*. Control Plane is usually implemented as a software program, hence, the name Software-Defined.

Separating the Data Plane and the Control Plane can deliver many benefits to those who construct and manage computer networks. Most significantly, the separation of the Control Plane from physical networking facilities such as Ethernet switches makes replacing protocol handling modules possible. This can be done quickly by replacing the software program installed in the Control Plane, without updating any firmware or configurations on the actual networking facilities. This feature is beneficial for operators who want to realize their own automated network management suited to their particular businesses, or researchers who want to try their new networking protocols or traffic management scheme. Other benefits may include high-level interoperability between the Data Plane and the Control Plane, and applicability of software engineering techniques when developing new networking protocols.

3.2 ExpEther Technology

Express Ethernet (ExpEther) technology basically virtualizes PCI Express over the Ethernet [5] and creates a single hop PCI Express switch even if the Ethernet network is composed of multiple switches. A promising feature of this architecture is that we can put as many computers and devices as necessary in a single Ethernet network without limits to the connection distance.

Another feature of ExpEther technology is that it allows us to attach and detach such devices to and from computers in a software-defined manner. This feature utilizes the characteristics of PCI in that its configuration is automatically executed among ExpEther chips that have the same group ID. In other words, by controlling the group ID, the computer hardware can be reconfigured.

3.3 Job Management System

The Job Management System is usually deployed to a high-performance cluster system for the purpose of load balancing and throughput improvement. This system is in charge of receiving resource requests from users, scheduling assignments of processor resources and then assigning an appropriate set of resources to each resource request based on its scheduling plan. Numerous job management systems have been proposed and implemented. Examples of such job management system include PBS [6], NQS [7] and Open Grid Scheduler/Grid Engine (OGS/GE) [13].

In general, the job management systems are also known as queuing systems, and these job management systems use multiple queues categorized by resource limitations and requests' priorities for resource allocation of the system. Thus, the job management systems completely understand when and which job should start. In this research, our job management system interacts with SDN and ExpEther technology to realize a computing environment dedicated to the job owner.

4 Proposal

Figure 1 shows an overview of our envisioned supercomputing environment. As illustrated in the figure, the job management system receives user's requests in the form of a batch script. The received job requests are stored in a queue in the job management system and then wait to be dispatched. At this time the contents of the job requests are parsed and understood by the job management system in terms of what kind of devices, for ex., GPU and SSD (Solid State Drive) in a resource pool connected to a ExpEther network are required. Based on the information on resources requested by jobs in the queue and the usage information of computing nodes, the job management system attempts to coordinate an optimal allocation plan of resources to each job request, taking into account the availability of processors and devices in the resource pool. When a certain job request's turn comes, the job management system interacts with the ExpEther manager to prepare for a set of computing nodes reconfigured with user-requested devices in a resource pool through the use of OpenStack Ironic, which is the module for bare metal machine management functionality. Figure 2 diagrams the inside interaction of the resource management software. The above-mentioned interaction mechanism between the job management system and ExpEther technology allows users to specifically request their own desired computing environment. For example, two computing nodes, each of which has four GPU nodes, can be specified through the batch script file to the job management system. The details of this mechanism have already been reported in [10].

For the security issue described in Sect. 2, our envisioned supercomputing environment plans to take advantage of network programmability brought by SDN in cooperation with the job management system. As described in the above paragraph, the job management system can learn when a certain job is run on which computing nodes. In our envisioned computing environment, we design the data stage-in and stage-out functionality so that the connectivity of a network between data storage where security-sensitive data is located and SSD to be connected to computing nodes via the ExpEther network is guaranteed only before and after the job treating the security-sensitive data is executed. As explained in Sect. 3.1, SDN enables the control of packet flows in a software programming manner. This means that the on-demand control of network connectivity can be achieved in response to the necessity of data movement. This prominent feature of SDN is considered promising in lowering the risk of data exposure to third parties. In our envisioned

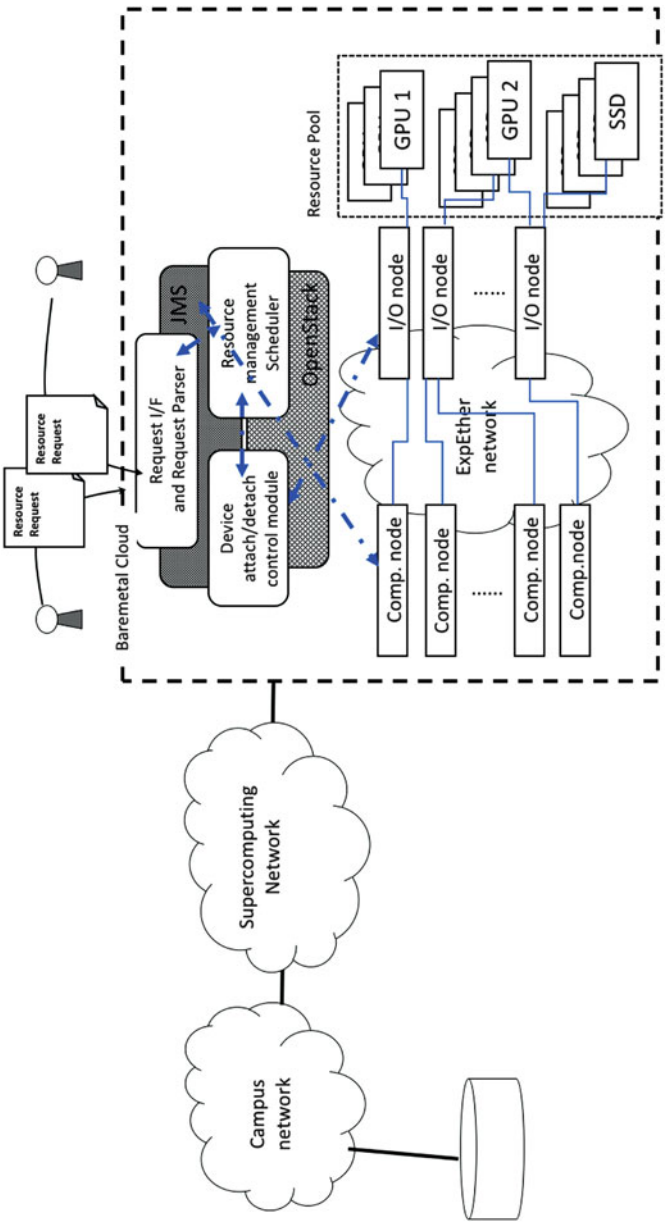

Fig. 1 Proposal overview

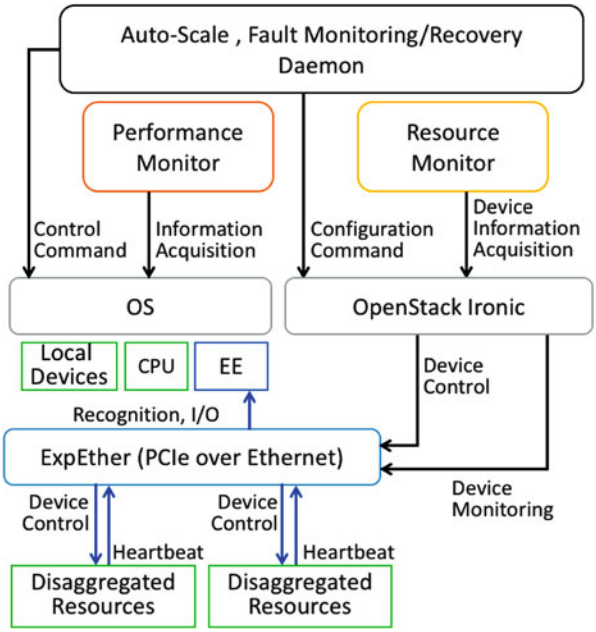

Fig. 2 Architecture of the resource management scheduler

computing environment, we synergistically have the job management system and SDN interlocked so that exclusive and secure data movement from and to computing nodes can be achieved.

In our plan, the envisioned computing environment is expected to work as follows. A dental scientist submits a job request to our supercomputing systems at the CMC. At this time security-sensitive data are still located on a secure storage in the dental hospital. Before the job request is dispatched to a set of computing nodes, the envisioned supercomputing environment isolates the target computing nodes by preparing a bare metal environment on OpenStack Ironic and then preventing other users from logging into it. Next, the environment establishes the connectivity and reachability of the network between the storage and SSD in the resource pool to be connected to a computing node used for secure computation. Simultaneously, the environment also reconfigures computing nodes with the user-specified devices including the SSD in the resource pool of our computer system using the ExpEther technology. Currently, security-sensitive data can be moved from the storage to SSD on a software-defined network established only for data movement. Immediately after data stage-in is completed, our envisioned environment disconnects the network so that no one can access both the storage and SSD.

On the other hand, when the computation is completed, data staging-out is performed. When the computation is finished, the job management system establishes the network between the storage and the SSD again and then move the

computational results back to the storage. After finishing the data movement, the staged-in data and computational results are completely removed and then SSD is detached from the computing node so that no one can access SSD.

In this research we aim to reduce the risk of data exposure to third parties by realizing the above-mentioned computing environment. At the time of writing this paper, we have been working on the prototype of the data stage-in and stage-out mechanism interlocked with the job management system and SDN. Furthermore, FlowSieve [18], a network access control mechanism leveraging SDN, which we have prototyped, can be applied in the future for enhancement of data security.

5 Conclusion

This paper has reported the research in progress towards the realization of an on-demand secure data staging mechanism that enables minimum exposure of security-sensitive data to third parties. Currently, we have been working on the prototype of the mechanism and the integration of it with SDN into JMS, in the hope that the achievement of this research will enable scientists to analyze such data on supercomputing systems at the CMC. At the same time, through continuing collaboration with scientists who need to treat security-sensitive data, we have recognized that there are still many unsolved security issues related to the regulations and guidelines mentioned in Sect. 2.

Acknowledgements This work was supported by JSPS KAKENHI Grant Number JP16H02802 and JP26330145. This research achievement is partly brought through the use of the supercomputer PC cluster for large-scale visualization (VCC).

References

1. Abraham, M.J., Murtola, T., Schulz, R., Pall, S., Smith, J.C., Hess, B., Lindahl, E.: GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX **1–2**, 19–25 (2015)
2. Bird, I.: Computing for the Large Hadron Collider. Annu. Rev. Nucl. Part. Sci. **61**(1), 99–118 (2011). doi:10.1146/annurevnucl-102010-130059
3. Date, S., Kido, Y., Khureltulga, D., Takahashi, K., Shimojo, S.: Toward flexible supercomputing and visualization system. Sustained Simulation Performance 2015, pp. 77–93. Springer, Cham (2015). doi:10.1007/978-3-319-20340-9_7
4. ESnet. https://www.es.net/
5. ExpEther (Express Ethernet) Consortium. http://www.expether.org/
6. Henderson, R.L.: Job scheduling under the portable batch system. In: Job Scheduling Strategies for Parallel Processing, vol. 949, pp. 279–294. Springer, Cham (1995)
7. Kingsbury, B.A.: The network queuing system. Technical Report, Sterling Software (1992)
8. LAMMPS Molecular Dynamics Simulator. http://lammps.sandia.gov/
9. Large-Scale Computer System, The Cybermedia Center at Osaka University. http://www.hpc.cmc.osaka-u.ac.jp/en/

10. Misawa, A., Date, S., Takahashi, K., Yoshikawa, T., Takahashi, M., Kan, M., Watashiba, Y., Kido, Y., Lee, C., Shimojo, S.: Highly reconfigurable computing platform for high performance computing infrastructure as a service: Hi-IaaS. In: The 7th International Conference on Cloud Computing and Services Science (CLOSER 2017), April 2017, pp. 135–146. doi:10.5220/0006302501630174
11. Momose, S.: NEC supercomputer: its present and future. In: Sustained Simulation Performance 2015, pp. 95–105. Springer, Cham (2015). doi:10.1007/978-3-319-20340-9_8
12. Nunes, B.A., Mendonca, M., Nguyen, X.N., Obraczka, K., Turletti, T.: A survey of software-defined networking: past, present, and future of programmable networks. IEEE Commun. Surv. Tutorials **16**(3), 1617–1634 (2014)
13. Open Grid Scheduler: The Official Open Source Grid Engine. http://gridscheduler.sourceforge.net/.
14. Science Information Network 5 (SINET5). http://www.sinet.ad.jp/en/top-en/
15. Singapore Advanced Research and Education Network (SingAREN). https://www.singaren.net.sg/
16. The Large Synoptic Survey Telescope. http://www.lsst.org/
17. Xia, W., Wen, Y., Foh, C.H., Niyato, D., Xie, H.: A survey on software-defined networking. IEEE Commun. Surv. Tutorials **17**(1), 27–51 (2015)
18. Yamada, T., Takahashi, K., Muraki, M., Date, S., Shimojo, S.: Network access control towards fully-controlled cloud infrastructure. Ph.D. Consortium. In: 8th IEEE International Conference on Cloud Computing Technology and Science (CloudCom2016), December 2016. doi:10.1109/CloudCom.2016.0076