LARGE-BANDWIDTH DATA ACQUISITION NETWORK FOR XFEL FACILITY, SACLA

T. Sugimoto, Y. Joti, T. Ohata, R. Tanaka, M. Yamaga, JASRI/SPring-8, Hyogo, Japan
T. Hatsui, RIKEN/SPring-8, Hyogo, Japan

Abstract

X-ray Free-Electron Laser (XFEL) experiments demands large bandwidth network for data acquisition (DAQ). At the SPring-8 Angstrom Compact Free Electron Laser (SACLA) facility, the experimental data rate is to be up to 5.8 Giga bit per second (Gbps). Some of the experiments demands preprocessing and on-line analysis by high-performance computers. In order to fulfill these requirements, a dedicated network system for DAQ and data analysis has been developed. A DAQ network consists of a dedicated 10 Gbps Ethernet (10GbE) physical layer to secure the data bandwidth and a 1GbE layer for instrument controls. The DAQ network is connected to a primary storage and indirectly to a PC cluster for data preprocessing. A firewall system with virtual private network (VPN) features is also implemented in order to secure remote access from off-site institutes. The preprocessed data plan to be transferred to the off-site supercomputer "K computer" with high efficiency.

OVERVIEW OF SACLA

SACLA is a X-ray free electron laser facility (XFEL) in Japan. The SACLA is located at SPring-8 site, in the west region of Japan. Characteristic specification of the SACLA is electron energy at 8 GeV, peak bunch length of 30 fsec, 18-unit undulator with 18-mm magnetic period. The SACLA is designed to produce X ray with a wavelength as short as 0.06 nm at the repetition rate of 60 Hz. The SACLA consists of accelerator building, undulator building, and experimental facility. To utilize both the SACLA and existence facility "SPring-8", SACLA-SPring-8 experimental facility will be in service in 2012. In the SACLA-SPring-8 experimental facility, X rays from SACLA and SPring-8 can be used simultaneously.

The SACLA facility has successfully established. First X-ray lasing with a wavelength of 0.12 nm was observed on June 7, 2011.[1] By continuing beam commissioning, the wavelength 0.08 nm was achieved on July 13. First test experiment is planned from the October, 2011, and public experimental use will be open on March, 2012.

Experimental Requirement and Data Acquisition Network Design

In March 2012, the SACLA facility will deliver beam time to public users. Multi-Port CCD (MPCCD) detector [2] and its data acquisition (DAQ) system will be used in most of the anticipated experiments. Characteristic requirements are flexibility to change experimental setup within an approximately one-week cycle, robustness to minimize down time, and guaranteed network bandwidth to store all the data. Experimental users also need to use a PC-Cluster system for both on-line and off-line analysis. In order to meet these requirements, we have developed a DAQ system for the SACLA experiments. One MPCCD sensor has 512×1024 pixels with 16-bit data depth. With additional meta data, single-frame data size is about 1 Mega Bytes (MB). At the maximum acquisition rate of 60 Hz, a single sensor produces 480 Mega bit per second (Mbps). Using 12 MPCCD sensors, the data rate reaches 5.8 Gbps. To satisfy the required data rate, we divided the DAQ network (DAQ-LAN) into two physical layers, one is a dedicated 10GbE for large-bandwidth data transfer, and another is a low-latency 1GbE for instrument controls. We applied high-availability techniques to the network trunk lines to reduce down time. The target value of the down time is less than one day per year.

Some of the important SACLA applications, such as three-dimensional (3-D) coherent diffraction imaging at near-atomic resolution and protein nano-crystallography, demand high performance computing (HPC) infrastructures. At the SACLA facility, we plan to connect the DAQ system to the off-site 10-Peta-flops (Pflops) supercomputer "K computer"[3] to make HPC infrastructure available to users. In the foreseen analysis procedure, experimental data are first transferred to an on-site 10-Tera-flops PC cluster with high I/O capability for preprocessing, and then transferred to the K computer via a dedicated network or academic network SINET4.[4] The DAQ-LAN should be connected to the PC cluster and internet, which demands careful network design. This was achieved by the implementation of firewall systems with virtual private network (VPN) features.

THE DAQ NETWORK SYSTEM AT SACLA

The SACLA accelerator facility is controlled by many control-system components. All control-system components are connected to machine control network (CNTL-LAN). Considering how to use the DAQ system and how to ensure the security and the performance, we have developed additional three networks; 1. large-bandwidth data transfer network (DAQ-LAN), 2. experimental user network (DAQ-USER-LAN), and 3. network for data analysis (PC-Cluster-LAN). Policies of these three networks are described in this section.

Since the accelerator is a radiation generator, we must apply strict network security to the CNTL-LAN. Control-system components are fixed during the SACLA operating cycle (typically a few month). On the other hand, experimental machine time is one-week cycle, which is shorter than the operating cycle. The DAQ-LAN should be segregated from the CNTL-LAN for the DAQ-system flexibility. The DAQ system needs accelerator status data, such as wave form, beam intensity, and beam position acquired by non-destructive detectors. To satisfy the requirement and our network security, the CNTL-LAN and the DAQ-LAN are connected each other by a firewall system.

The DAQ-USER-LAN is another network for user experiments. Experimental users can use their own PCs and experimental instruments connected to the DAQ-USER-LAN. The differences between the DAQ-LAN and the DAQ-USER-LAN are 1. all instruments at the DAQ-LAN are managed by the SACLA staff to ensure the DAQ performance (large-bandwidth data transfer and low-latency DAQ control), 2. users' instruments can be connected to the DAQ-USER-LAN by themselves, and the network performance is best-effort 1GbE. The DAQ-USER-LAN is also connected to the CNTL-LAN and the DAQ-LAN by the firewall to acquire accelerator status and to control beamline components.

The PC-Cluster-LAN is another network for the PC-Cluster system. Experimental users can login from the DAQ-USER-LAN, and the users can perform on-line data evaluation using the PC-Cluster system. Off-line users can login from General-LAN in the SPring-8 site. The off-line users can analyze experimental data in the high-speed storage system. To read out the data from storage system, large-bandwidth network is required for the PC-Cluster-LAN. The PC-Cluster-LAN is also used to transfer experimental data from the SACLA to off-site supercomputer "K computer". To meet these requirements, we made policies on the PC-Cluster-LAN; 1. people who have permission can login the PC-Cluster system, 2. login path from the Internet must be protected using VPN system, 3. data-transfer path between the high-speed storage system and the PC-Cluster system should be more than 5.8 Gbps bandwidth, and the bandwidth is upgradable, 4. data-transfer path between the SACLA and the K computer should have more than the DAQ data rate (5.8 Gbps). We have two plans for the data-transfer path to the K computer, one is a best-effort 10 Gbps wide area network (WAN) SINET4, to be connected with in the autumn of 2012. Another plan is data-transfer path using a dedicated line between the SACLA and the K computer to guarantee the bandwidth.

To develop the real network systems, we isolated logical-network layer and physical-network layer. By isolating network layers, the network systems can have upgradability and redundancy. In the next subsections, we show network planning and implementations from each logical and physical view.

Logical Network Planning and Implementations

We made logical network plan of SACLA. We decided IP-address assignment based on same idea of the existence SPring-8 network system.[5] The reasons to adopt the same idea are 1. interconnection method between LANs match each other, 2. to reduce problem such as broadcast traffic[6], 3. to reduce administration cost using same assignment rules. From the IP-address-assignment rule, we applied class-B private addresses to the CNTL-LAN, the DAQ-LAN, and the DAQ-USER-LAN. We also applied class-A private addresses to the PC-Cluster-LAN, because class-A private address indicate open network in the SPring-8.

Each LAN has one (or two with high-availability configuration) layer-three network switch (L3SW). To isolate logical-network layer from physical-network layer, we applied virtual LAN (VLAN) technology. We assigned different VLAN IDs to the large-bandwidth network (10GbE) and the low-latency instrumental-control networks (1GbE). Using VLAN technology, we have flexibility to fold some VLANs on one physical layer or only one VLAN occupies one physical layer such as 10GbE line.

To satisfy requirements on data acquisition and beamline instrumental control on the DAQ-LAN, we assigned twelve VLANs. A segment named "DAQ Data" is a exclusive 10GbE network for data transfer. Other eleven segments are 1GbE networks for experimental instrument control. "DAQ Control" is a 1GbE network to control DAQ components. "Live View" is a best-effort 1GbE network for on-line monitoring. "BL1" (beamline 1), "BL3", and "LH1" (laser hutch 1) are networks to control beamline components such as monochromators and optical mirrors. For future beamline expansion, "BL2", "BL4", "BL5", and "LH2" segments are reserved. Other two VLANs are used for network managements. Near future, network segments for the SACLA-SPring-8 experimental facility will be added to the DAQ-LAN.

Physical Network Implementations for the DAQ System

From the experimental requirements, the physical network should have the following functions; 1. the DAQ-LAN have a bandwidth of 10 Gbps, 2. the DAQ-LAN should be upgraded to 40GbE/100GbE (IEEE802.3ba), 3. the DAQ-LAN can be used several spot in the experimental hall, and 4. the DAQ-LAN and the DAQ-USER-LAN should have robustness to minimize down time less than one day per year. From these requirements, we have developed the physical network in the SACLA experimental facility.

From the requirement of bandwidth and upgradeability, we applied $9/125$ μm single-mode fiber (SMF) as physical trunk lines. For example, the 10GBASE-LR needs two core of SMF, and we can upgrade to 40GBASE-LR4 or 100GBASE-LR4 with existent two-core SMF in the future. If we apply multi-mode fiber (MMF), it is difficult to upgrade the network, because 40GBASE-SR4 and 100GBASE-SR10 need 8 and 20 core of MMF, respectively.

Figure 1 shows network-wiring diagram at the SACLA experimental facility. The DAQ-LAN consists of two kind of trunk line: one is a exclusive 10GbE for large-bandwidth data transfer (red) and another is versatile 1GbE for instrumental control (blue). We also use same physical trunk line for the DAQ-USER-LAN (purple) and the PC-Cluster-LAN (green). All of wiring lines are distributed from the computer room. Each trunk lines have redundant configuration using link aggregation (LAG, IEEE802.3ad/802.1AX) technology to meet the requirements on robustness of the DAQ-LAN.

The DAQ-LAN with 10GbE is available at several DAQ stations; in the experimental hutch 1–4 (EH1–EH4), 19-inch racks beside EH2 and EH3, and 19-inch rack at experiment preparation room (EP) 1. The DAQ-LAN with 1GbE is available at the LH1, EP5, and the optical hutch 1 (OH1) in addition to the DAQ stations.

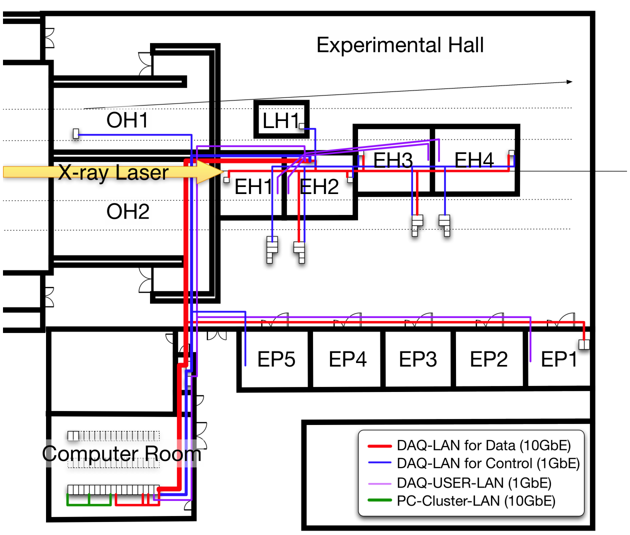

Figure 1: Overview of the SACLA experimental facility. X-ray laser comes from left side of this figure. The X-ray laser is focused in the optical hutches (OHs), and the X-ray laser is used for experiments at the experimental hutches (EHs). Experimental preparation rooms (EPs) are used for instrumental test. Red line is the DAQ-LAN data-transport line (10GbE). Blue line is the DAQ-LAN control line (1GbE). Purple line is the DAQ-USER-LAN (best-effort 1GbE). Green line is the PC-Cluster-LAN (10GbE).

Using the DAQ-LAN, the experimental data flow is described as below. Each detector outputs scattering image at 500 Mbps data rate via the Camera Link interface. MPCCD consists of 12 sensors, total data throughput is to be 5.8 Gbps. VME Camera-Link systems[7] receive data via the Camera Link interfaces. The VME Camera-Link systems send data to data file servers via the DAQ-LAN. Since the trunk line of data-transfer Ethernet is based on 10GbE, the network has enough capacity to transfer experimental data with 5.8 Gbps rate. Digital data from two VME systems are received by one data file server. The data file servers write data on the high-speed storage system via the FibreChannel. Each data file server has three GbE interfaces. Two GbE interfaces are used to receive experimental data at the rate of 500 Mbps from 10GbE line. Another GbE interface is used to control the DAQ system. Thus, we achieved 5.8 Gbps data rate from the MPCCD detectors to the high-speed storage system. Details of the DAQ framework is shown by M. Yamaga et al.[8]

Physical Network Implementation for the PC-Cluster System

We installed a PC-Cluster system at the SACLA experimental facility. The PC-Cluster system is aimed at 1. on-line data preprocessing, and 2. preprocessed data transfer to the K computer. Off-line data analysis will be carried out using off-site 10 Pflops K computer. Generally, network bandwidth over hundreds-kilometer distance is narrow. To make effective use of the K computer, we perform data reduction using the SACLA PC-Cluster system.

The PC-Cluster-LAN requires; 1. large-bandwidth data transfer between the high-speed storage system and the PC-Cluster system, 2. large-bandwidth data transfer from the SACLA to the K computer, 3. robustness to perform on-line analysis, down time should be equal to or less than that of the DAQ-LAN, and 4. versatile use of the PC-Cluster system, and users can login from the DAQ-USER-LAN, the General-LAN, and the Internet. It should be noted that users cannot login the DAQ-LAN from the General-LAN nor the PC-Cluster LAN.

From the requirement of large-bandwidth transfer (1), we have developed the PC-Cluster-LAN between the PC-Cluster system and the high-speed storage as 10GbE network. We also developed the PC-Cluster-LAN between the SINET4 and the PC-Cluster system for the requirement (2). The SINET4 is used for data transfer from the SACLA to users' institutes and the K computer. We also plan to connect the PC-Cluster-LAN with a dedicated network between the SACLA and the K computer to guarantee the data rate to the K computer. From the requirement of robustness (3), we choose ring-topology protocol as a high-availability mechanism for the PC-Cluster-LAN. The ring-topology protocol have much redundancy than that of LAG. From the requirement of versatile use of the PC-Cluster system (4), we installed firewalls between the PC-Cluster-LAN and other LANs to satisfy security policy of the SPring-8 site.

DATA TRANSFER PLAN FROM SACLA

To make effective use of data-transfer network, we have studied large-bandwidth data transfer technique from the SACLA to the off-site supercomputers. We have planed on-line analysis using the K computer. From the result of on-line analysis at the K computer, we can judge the quality of experimental conditions. We set a present goal to finish the quality-judgement during the beam time. Thus, we can optimize experimental conditions during the experiment.

In order to determine data-transfer bandwidth requirement, we plan to use other two supercomputers. One is "FOCUS" supercomputer located at Kobe, just beside the K computer.[9] The Internet access line of the FOCUS is commercial WAN at 100 Mbps. Another is "e-Science" supercomputer located in the K computer site.[10] The e-Science is connected to the SINET4 at 4 Gbps. The SINET4 between the SACLA and the e-Science can be configured as a 4-Gbps dedicated line. In early of 2012, we plan to evaluate data-transfer throughput of the WAN and a dedicated academic network using the FOCUS and the e-Science.

SUMMARY

We have developed large-bandwidth networks for high-speed DAQ and the PC-Cluster systems at the SACLA. Experimental requirements are 5.8 Gbps bandwidth, flexibility, low down time, accessibility to CNTL-LAN, accessibility to the PC-Cluster system, and network security. We satisfied these requirements, and the DAQ system has no bottle neck from the detectors to the high-speed storage system. In March 2012, the SACLA facility will deliver beam time to public users. MPCCD detector and the DAQ system will be used. In the autumn of 2012, the PC-Cluster system will be connected with the K computer. Using the K computer, we plan to perform XFEL data analysis of 3-D coherent diffraction imaging at near-atomic resolution and protein nano-crystallography.

REFERENCES

[1] H. Tanaka, "Status Report on the Commissioning of the Japanese XFEL at SPring-8", IPAC 2011, San Sebastian, Spain, Septempber 2011.

[2] T. Kameshima et al., "Development status of X-ray 2D Detectors for SPring-8 XFEL", IEEE Nuclear Science Symposium, Knoxville, TN, USA, Octber 2010.

[3] The supercomputer "K computer", Advanced Institute for Computational Science, RIKEN, Kobe, Japan, http://www.aics.riken.jp/.

[4] Science Information NETwork 4 (SINET4), National Institute of Informatics, Japan, http://www.sinet.ad.jp/.

[5] T. Sugimoto et al., "Upgrade of the SPring-8 Control Network for Integration of XFEL", ICALEPCS 2009, Kobe, Japan, October 2009, WED006.

[6] T. Sugimoto et al., "A study of network vulnerability in embedded devices", (CS)2/HEP Workshop, Kobe, Japan, October 2009, Track 08.

[7] A. Kiyomichi et al., "Development of Image Data Acquisition System for Large Scale X-ray 2D Detector Experiments at SACLA (SPring-8 XFEL)", ICALEPCS 2011, WTC Grenoble, France, October 2011, WEPMN028.

[8] M. Yamaga et al., "Event-Synchronized Data Acquisition System of 5 Giga-bps Data Rate for User Experiment at the XFEL Facility, SACLA", ICALEPCS 2011, WTC Grenoble, France, October 2011, TUCAUST06.

[9] FOCUS super Computer, Foundation for Computational Science, Kobe, Japan, http://www.j-focus.or.jp/.

[10] Supercomputer of the e-Science project, Information Technology Center, The University Tokyo, Japan, http://pslogin.cspp.cc.u-tokyo.ac.jp/escience/.