

BEAM LINE SOFTWARE UPGRADE FOR HIGH THROUGHPUT PROTEIN CRYSTALLOGRAPHY EXPERIMENTS AT THE PHOTON FACTORY*

Yu.A.Gaponov[#], N.Honda, N.Igarashi, M.Hiraki, K.Sasajima, N.Matsugaki, T.Kosuge, S.Wakatsuki,
Structural Biology Research Center, Photon Factory, High Energy Accelerator Research
Organization, Tsukuba, Japan
M.Suzuki, Osaka University, Osaka, Japan

Abstract

An integrated controlling system based on the unified database newly designed, is developed and realized for high-throughput protein crystallography experiments on synchrotron beam lines. Main features of protein crystallography experiments (purification, crystallization, loop preparation, data collecting, data processing) are dealt with the software. All information necessary to perform protein crystallography experiments is stored (except raw data, that are stored in a Network file servers) in a relational database (MySQL). The control system consists of several local servers and clients connected through network [1,2]. All communications are performed through TCP/IP sockets. Secure network connections such as OpenSSL, an open-source implementation of the SSL (Secure Sockets Layer) and TLS (Transport Layer Security) protocols are used wherever secure server-client communication is needed. Secure remote access to the system is possible from any operating system with X-terminal and SSH/X11 (Secure Shell with graphical user interface) support.

INTRODUCTION

There are several different tasks one should solve during programming of control system for operating scientific experiments. In general, control software may be considered as an interface between experimenter and equipment: experiment related commands are transferred onto equipment related commands and operations. That is why the important demand of the control system is the ability to support the friendly and reliable interface, which converts experiment related commands (top level) to equipment related commands (low level). Especially it is important for high throughput protein crystallography experiments, which are expected to be performed in schedule mode (without participation of experimenter during different stages of the experiment): it is necessary to keep all parts of the system in reliable conditions and protected from any mistakes or errors due to software/hardware features and human error. Another demand for high throughput protein crystallography experiment control system is the possibility to control and assist during an experiment remotely allowing different members of a research team to join the experiment.

To satisfy the demands described above with one

standalone application on one computer is quite problematic: to control several tens of different equipment units makes the application large, heavy, and difficult to modify. It is more advantageous to split the large task to smaller ones, which is a preferred choice in the field of controlling design and large-scale programming [3]. Typically, it is based on a data-bus with integrated CPU board computers. These systems tend to be expensive and quite complicated for extensive programming and modification. An alternative approach is Internet based distributed systems. In a distributed system, many different computers are integrated through the network, which makes it feasible to develop control software simultaneously by a group of system engineers and computer scientists [4].

High throughput protein crystallography experiment includes several stages. To perform tasks during all stages in a high-throughput mode, it is necessary to integrate all schedules and experimental data and results in the database to keep the controlling software updated according to current experimental status and to allow one to make scientific analysis of all related with experiment data.

Several parts of the current project (database management, experiment scheduling, data collection of snapshot, oscillation and MAD experiments, crystal mounting and centering) were successfully checked and evaluated at protein crystallography beamline BL5, Photon Factory, KEK, Tsukuba.

UNIX TCP/IP SECURE SOCKET CLIENT/SERVER MODEL

The UNIX TCP/IP socket client/server was chosen as the general communication model. To make communication secure OpenSSL, an open-source implementation of the SSL (Secure Sockets Layer) and TLS (Transport Layer Security) protocols is used. This model is reliable and well designed under all operating systems for inter-process/program synchronous and asynchronous communications [4, 5].

There are four types of communication interfaces between client application and the server: a) a Sending/Receiving type (a sending command is followed by waiting a reply); b) a Receiving/Sending type (a receiving command is followed by a sending reply); c) a Sending type (a sending command or message isn't followed by waiting a reply); d) a Receiving type (a

*Work supported in part by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan

[#]yugru@yahoo.com

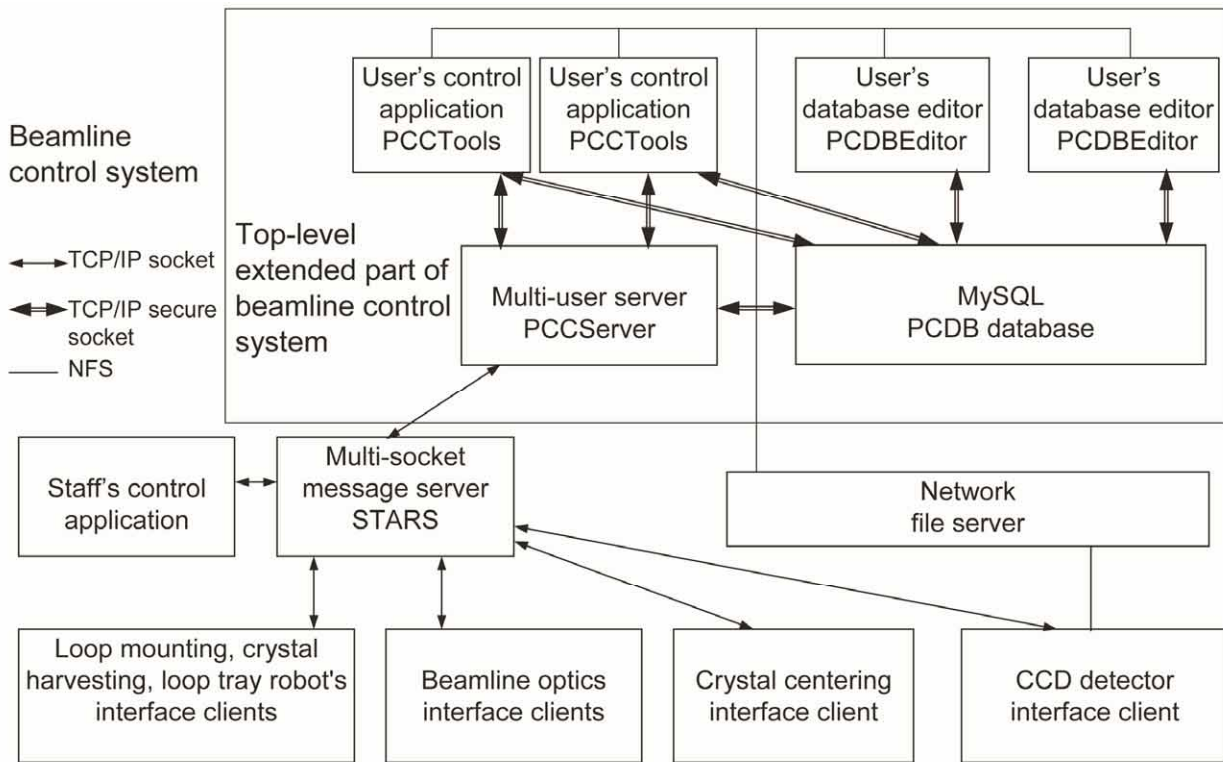


Figure 1: The overall scheme of the high throughput protein crystallography experiments control system.

receiving a message or a command isn't followed by sending a reply.

A combination of the first two types of interfaces allows one to build synchronous network communication with a high reliability. A combination of the last two types allows for asynchronous network communication with a higher communication speed, mainly for exchanging of status information for monitoring purposes. Our control system consists of a mixture of the two above combinations.

Fig.1 shows the overall scheme of the control system. There are two levels of control. On the top level, the main server module, PCCServer, allows user's client application PCCTools to be connected in a multi-user mode. This server has a socket connection with the database (MySQL), which stores all the top-level information. The server is able to execute both X-ray data collection experiments (it has its own "run-engine" not to stop the experiment in the case of breaking the socket connection between user's client application and server) and data analysis of the experimental data by executing the commercial applications for indexing, integration and scaling. On the top-level, control commands reflect the modes of operation of the experimental equipment, for example, data collection, direct beam measurement, beam stopper alignment, etc. A multi-socket server, STARS, (Simple Transmission and Retrieval System, which was developed at the Photon Factory, KEK, Tsukuba, Japan) is used for communication with different equipment interface modules [6]. One client can be connected to

STARS through several sockets to realize any type of communication interface mentioned above.

UNIFIED DATABASE

Fig.2 shows the database layout. The relational database for high throughput protein crystallography experiments was developed using MySQL software to keep all the experiment details, experiment schedules and data, data analysis results. The database allows one to make systematic analyses of the information related with protein crystallography experiment, which is essential to solve scientific tasks and avoid human errors and mistakes [1,2].

There are several database hierarchy trees. One of it, Crystal-tree, is related to protein crystal describing the information about protein overexpression and purification, crystallization, crystal harvesting and crystal mounting in the crystal loop (cryo-loop). DataCollection-tree is related to X-ray experiment describing the details of the experiment schedule steps (snapshot, oscillation, MAD, XAFS), their executing progress and results. DataAnalysis-tree is related to X-ray data analysis describing the details of the data indexing and integrating, and data reduction procedures. There are additional sets (ExperimentSettings and Administration) of tables to keep information about experiment and equipment settings, user accounts and groups.

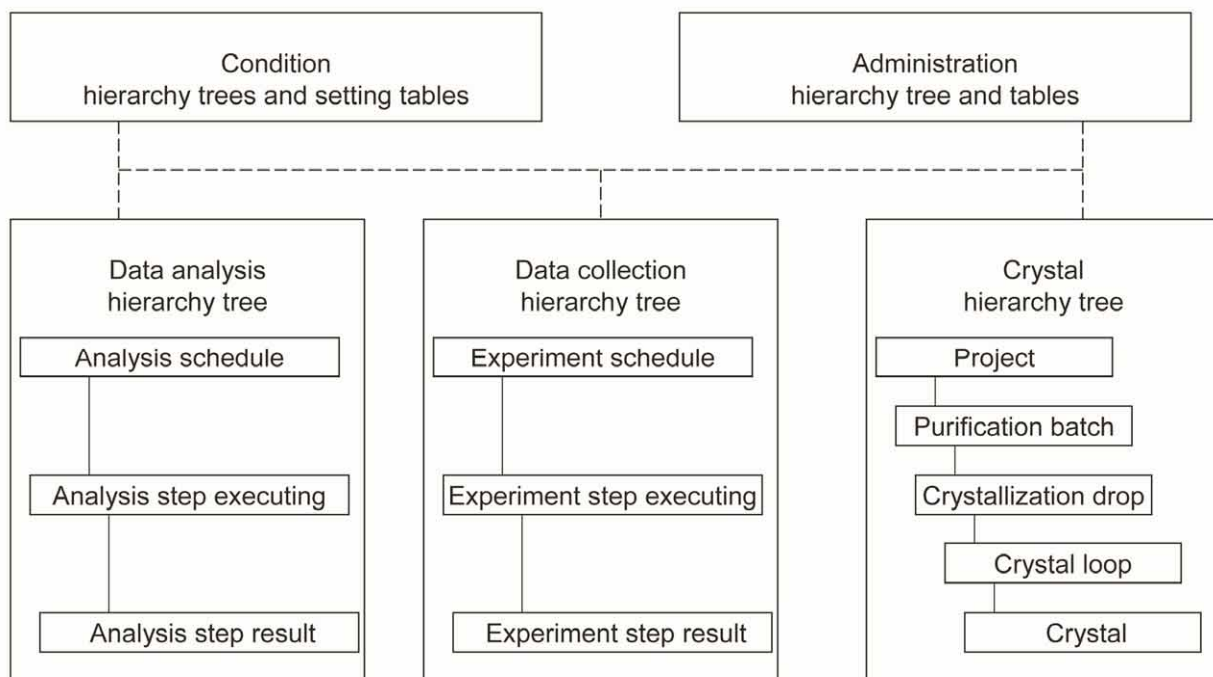


Figure 2: The overall scheme of the relational PCDB database.

BEAMLIN SOFTWARE UPGRADE

The main operating system is Linux Red Hat 9. The STARS server, several client modules and applications for controlling the equipment, MySQL database software, PCCServer, PCCTools and PCDBEditor all operate under the Linux. Another operating system, which is used for controlling the equipment, is MS Windows.

Majority of the source code is written in C/C++ (GNU C/C++ compiler v.3.2.2). GUI modules of the system were built mainly using Glade user interface builder for GTK+ and Gnome under Red Hat Linux 9 operating system. STARS server, several client modules and applications for controlling the equipment are written in Perl. Some existent client modules, connected to STARS server, were modified to accept top-level commands. Top-level part of the control system was just connected to STARS server. Upgraded control system was evaluated at the protein crystallography beamline BL5, Photon Factory, KEK, Tsukuba. Main functions of the system, such as X-ray experiment scheduling, X-ray data collection (snapshot, oscillation and MAD experiments), crystal loop mounting and crystal centering (manual and automatic) procedures, were checked.

Current scheduling system is designed to create experiment schedule in advance (before experiment). In the case of unknown crystal under investigation it is necessary to define optimized parameters for the diffraction experiment (time exposure, rotation angle range, number of frames, etc.). So, for high-throughput

mode (schedule mode) of experiment the procedure of crystal evaluation (so-called expert-procedure) is necessary. For manual mode (step-by-step) it is necessary to develop schedule step creation procedure (so-called wizard procedure), which allows user quickly make, store in the database the experiment schedule step without opening the database editor. In both cases it is necessary to develop the procedure of optimized search and generation of reference records in the relational database.

Fig.3, 4, and 5 represent the control clients PCCTools (data collection and crystal loop centering windows) and PCDBEditor.

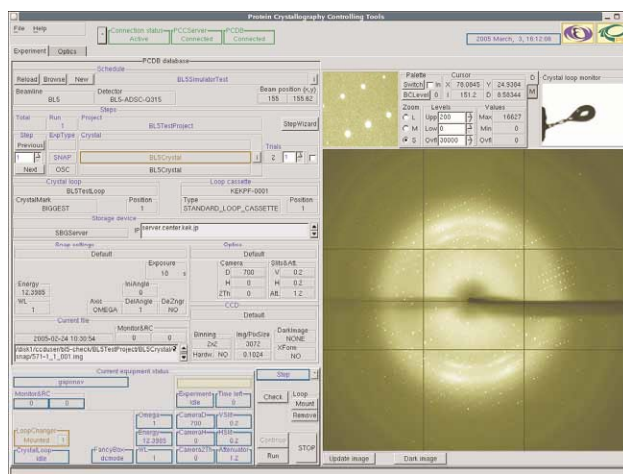


Figure 3: Screenshot of data collection window of user's application PCCTools.

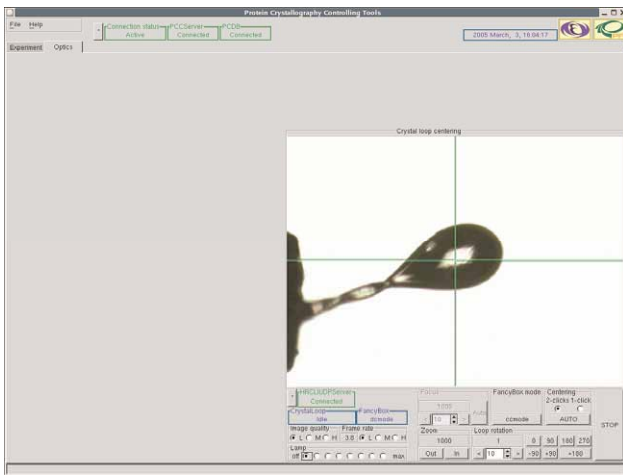


Figure 4: Screenshot of crystal centering window of user's application PCCTools.

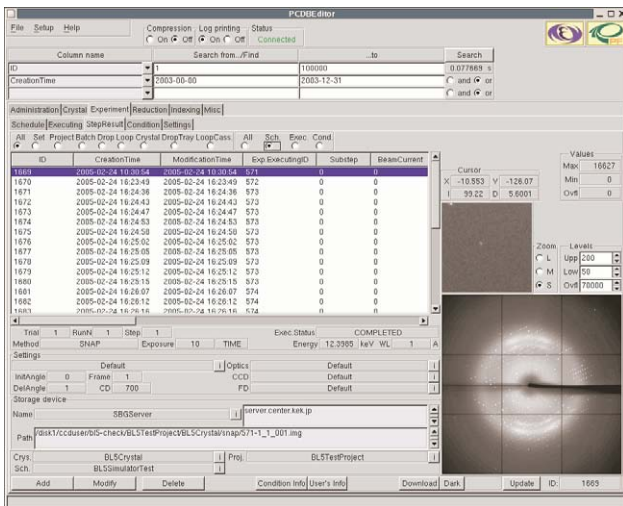


Figure 5: Screenshot of user's application PCDBEditor.

ACKNOWLEDGEMENTS

This work was supported in part by Grants-in Aid for Scientific research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, Special Coordination Funds for Promoting Science and Technology, and Protein 3000 Project of the MEXT.

REFERENCES

- [1] Gaponov, Yu.A., Igarashi, N., Hiraki, M., Sasajima, K., Matsugaki, N., Suzuki, M., Kosuge, T. and Wakatsuki, S., *J. Synchrotron Rad.*, 11,17-20, (2004)
- [2] Gaponov, Yu.A., Igarashi, N., Hiraki, M., Sasajima, K., Matsugaki, N., Suzuki, M., Kosuge, T. and Wakatsuki, S., *AIP Conference Proceedings*, 705, San Francisco, California, 25-29, August, USA, 1213-1216, (2003).
- [3] R.Pugliese, L.Gregoratti, R.Krempaska, F.Bille, J.Krempasky, M.Marsi, A.Abrami, A novel approach to the control of experimental environments: the ESCA microscopy data-acquisition system at ELETTRA. *J. Synchrotron Rad*, 5 (1998) 587-589
- [4] E.Abola, P.Kuhn, T.Earnest, R.C.Stevens, Automation of X-ray crystallography. *Nature. Struct. Biol.*, 7 (2000) 973-977.
- [5] R.M.Sweet, J.M.Skinner, M.Cowan, The trust of automation in synchrotron-based macromolecular crystallography. *Synchrotron Radiation News*, V.14, 3 (2001) 5-11.
- [6] Kosuge, T., Saito, Y., Nigorikawa, K., Katagiri, H., Shirakawa, A., Nakajima, H., Ito, K., Kishiro, J. & Kurokawa, S. (2002). Oral talk on 4th International Workshop on Personal Computers and Particle Accelerator Controls, 14-17 October, 2002, Frascati (RM), Italy, http://www.lnf.infn.it/conference/pcapac2002/TALK/WE-03/WE-03_talk.pdf.