

Norbert Podhorszki and Gábor Vida
MTA SZTAKI, Laboratory of Parallel and Distributed Systems

Application programming Interface for SETI-like distributed programs and an execution system based on BOINC

For over the last couple of years the largest computing systems in the world have been based on the assembly of large numbers of PCs over the Internet. Such systems (Desktop Grid) can provide multiple teraflops computing power for applications that are feasible for such an infrastructure. PCs in a Desktop Grid do not communicate with each other. Thus, the executing application processes should be sequential (except for multi-threaded ones optimised for dual-processor machines). So the enormous computing power can be used only for problems that can be partitioned into independent subtasks.

The best example of a Desktop Grid is the SETI@home project, where the application is running for years while the incoming daily 35 GB input data is processed in chunks of 350 KB on the PCs. The SETI@home is a scientific experiment using several millions of computers connected over the Internet for the Search for Extraterrestrial Intelligence. SETI is the name of the whole scientific area, whereas SETI@home is a technical approach to it by analysing radio frequencies, a search for intelligent signals in a very narrow frequency interval. The processing of data is independent in practice: signals coming from two different parts of the sky are independent as well as signals with different frequency from the same place (at least from the search criteria of the project).

The open source BOINC is being designed and implemented by the SETI@home's earlier developers in order to create a unified infrastructure for similar distributed programs of attractive scientific projects that need enormous computing power. This would enable millions of users to maintain the same software on their PCs while they can choose among different projects for support without downloading and reinstalling new software. The network of several millions of PCs running the BOINC client could become the most powerful supercomputer on Earth, which is able to execute many different applications (not just a dedicated one). However, the installation and usage of BOINC is inconvenient. There is a need for sophisticated system administrators to maintain and configure BOINC servers. Moreover, there is no aid for creating appropriate applications as the BOINC developers wanted to satisfy the needs of SETI@home. Other developers need to have persistence to tailor their applications to the needs and peculiarities of BOINC.

The DC-API (Distributed Computing API) developed in MTA SZTAKI provides a simple programming interface for SETI-like, so called Master-Worker, distributed applications. This API hides the properties of the actual execution system thus, the programmer can concentrate the efforts to create subtasks and to process sub-results only. The execution of such a program is the task of the implementation of the API. The life of the programmers who are used to sequential programming becomes much easier. Moreover, the application is not specifically bound to a specific infrastructure, because it can be executed on any distributed platform (Desktop Grids, another Grids, clusters and supercomputers, etc) for which the API is implemented.