Паралелно и дистрибуирано процесирање

Домашна 4

Стефан Милев | 206055

1.

The Globus toolkit contains a set of libraries and programs that provides the developers of specific tools or apps with solutions for common problems that are encountered when creating a distributed system services and applications.

Globus is a software with components and capabilities that includes:

- A set of service Implementations that Indicate resource management, data alterations service finding and relevant issues
- Tools for building web services
- A powerful standards-based security prerequisites for authentication and authorisation.
- Various services in java c and python for clients of API and command line programs
- Detailed documentation on these various components

2.

ChinaGrid aims at building a public service system for Chinese education and research. ChinaGrid Support Platform (CGSP) is a grid middleware developed for the construction of the ChinaGrid. Function modules of CGSP for system running are Domain Manager, Information Center, Job Manager, Data Manager, Service Container and Security Manager. Developing tools for gird constructor and application developers consist of Service Packaging Tool, Job Defining Tool, Portal Constructor and Programming API. CGSP architecture is introduced first. Then, CGSP function modules and developing tools are described.
The modules of uniform authorization management include user management module, authorization management module and file-level access control.
User management module offers the identity certification, proxy credentials management and Identity mapping.

- Identity certification. Every entity in ChinaGrid gets X.509 certificate and uses their certificate to verify each other.

- Proxy credentials management. We introduce the proxy mechanism through saving a limited proxy certificate signed by user in proxy certificate management server. In this way, user need not use his own certificate every time and implement the single-sign-on.

- Identity mapping. When user logins, domain manager get user information from user's register information. Domain server uses rule set to transfer user id to a domain id.

Authorization management module includes two parts:

- Multi-level access control. CGSP inherits WSRF (Web Services Resource Framework) security. Through implementing GSI Secure Conversation and GSI Secure Message, CGSP ensures the job submitted is legal and provide the security of job during execution.

- Dynamic authorization mechanism. It predicts the probability of the success of task execution, and makes a decision to limit the user's right based on the prediction result. File-level access control mechanism is designed based on the files metadata information. The files metadata information related to security includes owner, access privilege, authorization object, and so on.

3.

SETI@home searches for possible evidence of radio transmissions from extraterrestrial intelligence using observational data from the Arecibo radio telescope and the Green Bank Telescope. The data is taken "piggyback" or "passively" while the telescope is used for other scientific programs. The data is digitized, stored, and sent to the SETI@home facility. The data are then parsed into small chunks in frequency and time, and analyzed, using software, to search for any signals—that is, variations which cannot be ascribed to noise, and hence contain information. Using distributed computing, SETI@home sends the millions of chunks of data to be analyzed off-site by home computers, and then have those computers report the results. Thus what appears a difficult problem in data analysis is reduced to a reasonable one by aid from a large, Internet-based community of borrowed computer resources.

The software searches for five types of signals that distinguish them from noise:

- Spikes in power spectra
- Gaussian rises and falls in transmission power, possibly representing the telescope beam's main lobe passing over a radio source
- Triplets – three power spikes in a row
- Pulsing signals that possibly represent a narrowband digital-style transmission
- Autocorrelation detects signal waveforms.

There are many variations on how an ETI signal may be affected by the interstellar medium, and by the relative motion of its origin compared to Earth. The potential "signal" is thus processed in many ways (although not testing all detection methods nor scenarios) to ensure the highest likelihood of distinguishing it from the scintillating noise already present in all directions of outer space. For instance, another planet is very likely to be moving at a speed and

acceleration with respect to Earth, and that will shift the frequency, over time, of the potential "signal." Checking for this through processing is done, to an extent, in the SETI@home Software.

The process is somewhat like tuning a radio to various channels, and looking at the signal strength meter. If the strength of the signal goes up, that gets attention. More technically, it involves a lot of digital signal processing, mostly discrete Fourier transforms at various chirp rates and durations.

Anybody with an at least intermittently Internet-connected computer can participate in SETI@home by running a free program that downloads and analyzes radio telescope data. Observational data were recorded on 2-terabyte SATA hard disk drives fed from the Arecibo Telescope in Puerto Rico, each holding about 2.5 days of observations, which were then sent to Berkeley. Arecibo does not have a broadband Internet connection, so data must go by postal mail to Berkeley. Once there, it is divided in both time and frequency domains work units of 107 seconds of data, or approximately 0.35 megabytes (350 kilobytes or 350,000 bytes), which overlap in time but not in frequency. These work units are then sent from the SETI@home server over the Internet to personal computers around the world to analyze.

The analysis software can search for signals with about one-tenth the strength of those sought in previous surveys, because it makes use of a computationally-intensive algorithm called coherent integration that no one else has had the computing power to implement.

Data is merged into a database using SETI@home computers in Berkeley. Interference is rejected, and various pattern-detection algorithms are applied to search for the most interesting signals.

The project uses CUDA for GPU processing since 2015.

Since 2016, the project has also been helping to process data from the Breakthrough Listen project which has been recorded at Green Bank Telescope.

4.

FutureGrid

5.

In essence, BOINC is software that can use the unused CPU and GPU cycles on a computer to do scientific computing—what one individual does not use of his/her computer, BOINC uses. In late 2008, BOINC's official website announced that Nvidia had developed a system called CUDA that uses GPUs for scientific computing. With NVIDIA's assistance, some BOINC-based projects (e.g., SETI@home, MilkyWay@home) now have applications that run on NVIDIA GPUs using CUDA. Beginning in October 2009, BOINC added support for the ATI/AMD family of

GPUs also. These applications run from 2 to 10 times faster than the former CPU-only versions. In 7.x preview versions, GPU support (via OpenCL) was added for computers using Mac OS X with AMD Radeon graphic cards.

BOINC consists of a server system and client software that communicate with each other to distribute and process work units and return the results.

There are 31 active projects listed in the BOINC official website:

- Amicable Numbers – Mathematics
- Asteroids@home – Astrophysics
- BOINC@TACC – Multiple scientific areas
- Citizen Science Grid – Molecular biology, computer science
- Climateprediction.net – Climate study
- Collatz Conjecture – Mathematics
- Cosmology@Home – Astronomy
- Einstein@home – Astrophysics
- Gerasim@Home – Computer engineering
- GPUGrid.net – Molecular simulations of proteins
- Ibercivis BOINC – Biomedicine
- LHC@home – Physics
- Milkyway@home – Astronomy
- MindModeling@Home – Cognitive science
- Minecraft@Home – Games
- MLC@Home – Artificial Intelligence
- Moo! Wrapper – Cryptography and combinatorics
- nanoHUB@Home – Nanoscience
- NFS@home – Factorization of large integers
- NumberFields@home – Mathematics
- ODLK1 – Mathematics
- PrimeGrid – Mathematics
- QuChemPedIA@home – Molecular Chemistry
- Radioactive@Home – Environmental research
- RakeSearch – Mathematics
- RNA World – Molecular biology
- Rosetta@home – Biology
- SRBase – Mathematics
- Universe@Home – Astronomy
- World Community Grid – Medical, environmental and other humanitarian research
- Yoyo@home – Mathematics

6.

- ChinaGrid
- NSF TeraGrid.

7.

- GLUE for resource representation
- SAGA (Simple API for Grid Applications)
- GSI (Grid Security Infrastructure)
- OGSI (Open Grid Service Infrastructure)
- WSRE (Web Service Resource Framework)

8.

- Alluxio (Virtual Distributed File System)
- Ceph
- Coda
- GlusterFS
- MooseFS
- Quantcast File System
- IPFS
- Kertish-DFS
- LizardFS
- Lustre
- MinIO
- OpenAFS
- OpenIO
- RozoFS
- SeaweedFS
- Tahoe-LAFS
- HDFS
- XtreemFS
- Ori
- BeeGFS
- ObjectiveFS
- Spectrum Scale (GPFS)
- MapR-FS
- PanFS
- Infinit
- Isilon OneFS
- Scality
- Quobyte
- Amazon S3
- Google Cloud Storage
- SWIFT (part of OpenStack)
- Microsoft Azure
- IBM Cloud Object Storage