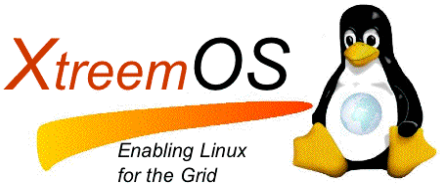

Building and Promoting a Linux-based Operating System to Support Virtual Organizations for Next Generation Grids

Executive Summary, May 31 2007

1- Project Objectives

The emergence of Grids enables the sharing of a wide range of resources for solving large-scale computational and data intensive problems in science, engineering and business. While much has been done to build Grid middleware on top of existing operating systems, little has been done to extend the underlying operating systems for enabling and facilitating Grid computing, for example by embedding important functionalities directly into the operating system kernel. XtreemOS project aims at investigating and proposing new services that should be added to current operating systems to build a Grid infrastructure in a simple way. This approach can be seen to have some advantages over conventional Grid middleware. Application programmers spend a good deal of time managing the services provided by the middleware toolkit, which may have different programming interfaces and lack of a unifying model. A common interface can be provided to simplify the task of the application developer on the Grid by making the Grid support native to the operating system, and also by removing layers of abstraction, leading to higher dependability of services.

Therefore, the goals of XtreemOS project are to design, implement, evaluate and distribute an open source Grid operating system which supports Grid applications, and capable of running on a wide range of underlying platforms, from clusters to mobiles. Installed on each participating machine (personal computer, cluster of workstations, mobile device), the XtreemOS system will provide for the Grid what a traditional operating system offers for a single computer: abstraction from the hardware and secure resource sharing between different users. It will thus considerably ease the work of users belonging to virtual organisations by giving them the illusion of using a traditional computer, and releasing them from dealing with the complex resource management issues of a typical Grid environment. The approach being investigated is to base XtreemOS on the existing well-accepted open source Linux OS. The underlying Linux OS will be extended as needed to support virtual organisations spanning across many machines and to provide appropriate interfaces to Grid OS services. A set of system services will provide users with all the Grid capabilities associated with current Grid middleware, but fully integrated into the OS. By integrating Grid capabilities into the Linux operating system, XtreemOS will also provide a more robust, secure and easier way to manage a Grid infrastructure for system administrators. This will be experimentally demonstrated with a set of real applications, provided by well-known industrial partners that cover a large spectrum of application fields.

2- Work Done and Main Achievements

During the first year of the project, we first worked on the capture of requirements from a set of reference applications and use cases. The 14 reference applications cover different sectors: aerospace, energy, bio-informatics, business, finance, virtual reality and telecommunications. We also worked on the specification of the XtreemOS Grid operating system. The results of this work are described in a set of deliverables available in XtreemOS public web site. We then started the design and implementation of the XtreemOS components for the first basic version of the system for individual PCs and clusters. We regard it as essential for a practical system that there should be some assurance

provided that the system does meet recognised security criteria. Work is ongoing to derive a systematic analysis of threats to the XtreemOS system, with a view to validating the integrity of XtreemOS.

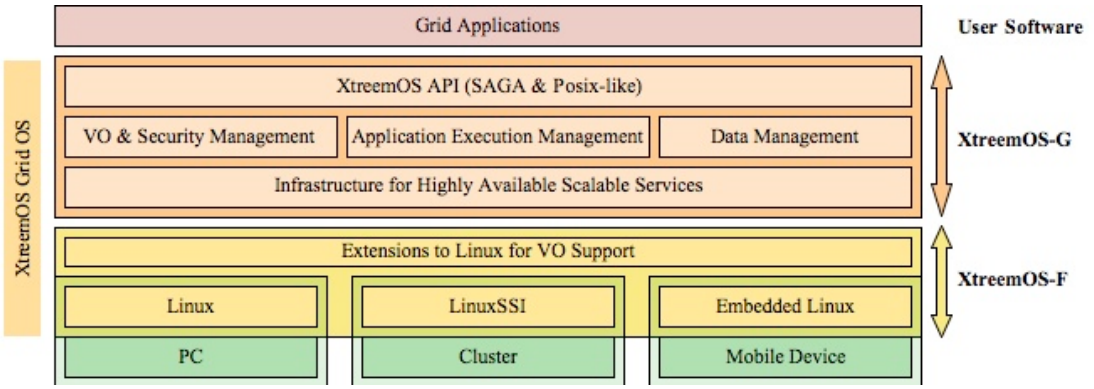

Figure 1: XtreemOS Architecture

The overall XtreemOS architecture is depicted in Figure 1. The XtreemOS architecture is divided logically into two layers.

The XtreemOS Foundation layer, XtreemOS-F, provides a modified Linux Kernel in the kernel embedding native Virtual Organisation support. This will be provided in three major variants: a version aimed at PCs and workstations; a version aimed at providing a Single-System Image for cluster computing, based on the existing Kerrighed system; and a version which can be deployed on small mobile devices such as PDA.

The XtreemOS Grid support layer, XtreemOS-G provides Grid OS distributed services to securely manage computation and data resources. The main services provided in the G-Layer are Application Execution Management (AEM), a Grid file system (XtreemFS), and application and VO-level security. In order to deploy the XtreemOS-G layer over a number of participating nodes, an infrastructure of highly available services will manage these nodes. It provides for example support for publish/subscribe services, node virtualization, and node directory services.

XtreemOS will provide a common API to Grid applications; this will be based on the emerging Simple API for Grid Applications (SAGA) standard currently under development within the OGF.

VO and Security Management

XtreemOS aims to provide native support for the management of VOs in a secure and scalable way, without compromising on flexibility and performance. VO Management (VOM) covers all the infrastructural services that are needed to manage the entities involved in a VO and ensure a consistent and coherent exploitation of the resources, capabilities, and information inside the VO under the governance of the VO policies. In XtreemOS we are aiming to integrate VOM as part of the OS. More specifically, VOM can be implemented as a service that can be integrated directly with existing authentication infrastructure. First, this approach reduces the management and performance overheads introduced by the layers of controls. Second, the hassle of accessing VO resources can be reduced. The policies specified by a VO, such as security, resource limitations, scheduling priorities and rules on how shared resources could be used by VO members, will be finally checked and ensured at resource nodes. In order to adapt to different VO models and reduce kernel code changes, XtreemOS uses the PAM (Pluggable Authentication Modules) system, which allows a system administrator to add (possibly VO-specific) authentication methods by installing new PAM modules. Local user accounts in XtreemOS are allocated dynamically on each resource to match the actual global users exploiting that resource. The XtreemOS PAM plugins are in charge of implementing (or interfacing to) a local service allocating fresh local UID/GID couples upon request. The dynamic allocation of user accounts ensures XtreemOS scalability and reduces the complexity of VO management: no need

to configure resources when users are added or removed from VOs. Dynamic management of local UID/GID also provides some level of isolation between Grid users: they do not share access to local files, and it is possible to hide the real identity of a user in the local name space.

Application Execution Management
XtreemOS aims at offering easy and efficient job execution management. The AEM service provides a simple interface for executing, controlling and monitoring jobs, on top of which the SAGA interface defined by OGF will be implemented. The AEM service also supports remote execution of jobs submitted in a standard Unix way without modifications in the binary code and with minimal user intervention. For the sake of scalability, most of the AEM services are not long-term running and do not have a global view of the system. In particular, there is no global Grid scheduler in XtreemOS; the scheduler being job-oriented and only active from the submission to the queuing of the job on some Grid resources. For resource discovery in a VO, AEM exploits overlay networks provided by the infrastructure for scalable and highly available services. Another important feature of AEM is its ability to accurately monitor the execution of jobs, which is not normally the case in Grid systems. AEM will also be able to migrate jobs automatically in case of changes in resource availability, for instance after a resource failure.

Data Management
The XtreemOS data management service, XtreemFS, is a distributed file system structured according to the *object-based file system* approach. The object-based file system architecture splits files into their pure content, the objects, which are stored on so called object storage devices (OSDs), and the file metadata, which is put on dedicated metadata servers. XtreemFS extends the architectural concept of object-based storage to Grid environments by replacing the centralised metadata servers with a federation of metadata servers in order to ensure independence of participating organisations while maintaining a global view of the system. In order to achieve scalability and fault tolerance, XtreemFS also features replication and partitioning/striping for both file metadata and the file content. Dynamically created communication overlays coordinate concurrent accesses and ensure the data's consistency in a scalable way. Data can be replicated across organisation boundaries, and therefore special attention is paid to the latencies that the connecting wide area networks introduce and to failure cases like those of possible network partitioning. An automatic system will monitor file access and resource conditions to automatically optimise data layout in the Grid. In addition, semantic naming and advanced query functions allow users to find data in huge archives, with the aim of overcoming limitations for the organization of data of traditional hierarchical file systems.

3- Software Dissemination

XtreemOS Grid operating system will be released as open source software. We plan to deliver a public release of a first basic version of XtreemOS for individual PC and clusters in mid 2008. A public release of XtreemOS flavour for mobile devices (PDA) is planned in 2009.
We aim at creating of community of XtreemOS system users and developers beyond XtreemOS consortium during the second half of the project.

The foundation layer for the cluster flavour of XtreemOS leverages Kerrighed single system image operating system for clusters. Kerrighed is open source software under the GPL licence, originally developed by INRIA and EDF and now developed by a community (http://www.kerrighed.org). Since the beginning of the project, XtreemOS consortium contributes to the development of Kerrighed, being in particular strongly involved in the design and implementation of the high performance distributed file system, process checkpoint/restart functionalities and customizable scheduling framework. We also built Kerrighed packages for Mandriva, RedFlag and Debian Linux distributions. These packages are distributed in the SSI-OSCAR package of the OSCAR software suite that makes it easy to build Linux clusters for high performance computing. And last but not least, XtreemOS consortium works actively to integrate Kerrighed kernel patches into the mainstream Linux development.

4- XtreemOS Fact-sheet

XtreemOS is a four-year European Integrated Project funded by the European Commission that started in June 2006.

Contractors

Organisation name	Country
Caisse des dépôts et consignations	FR
Institut National de Recherche en Informatique et Automatique	FR
Council for the Central Laboratory of the Research Councils	UK
Consiglio Nazionale delle Ricerche	IT
European Aeronautic Defence and Space Company	FR
Electricité de France	FR
Edge-IT	FR
NEC High Performance Computing Europe	DE
SAP	DE
Barcelona Supercomputing Center - Centro Nacional de Supercomputación	ES
Universitaet Ulm	DE
Vrije Universiteit Amsterdam	NL
Xlab	SI
Konrad-Zuse-Zentrum für Informationstechnik Berlin	DE
T6	IT
Institute of Computing Technology of Chinese Academy of Sciences	CN
Red Flag Software	CN
Telefónica I+D	ES
Universitaet Düseeldorf	DE

Administrative and Financial Coordinator
Jean-François Forté, Caisse des dépôts et consignations
15 Quai Anatole France
75007 Paris
France

Scientific Coordinator
Christine Morin, INRIA
Campus universitaire de Beaulieu
35042 Rennes cedex
France

Contact: xtreemos-info@irisa.fr

Public web site: http://www.xtreemos.eu