



Project no. IST-033576

XtreemOS

Integrated Project

BUILDING AND PROMOTING A LINUX-BASED OPERATING SYSTEM TO SUPPORT VIRTUAL ORGANIZATIONS FOR NEXT GENERATION GRIDS

Market Analysis and Technology Transfer plan

D5.1.11

Due date of deliverable: August 31, 2010
Actual submission date: November 15, 2010

Start date of project: June 1st 2006

*Type: Deliverable
WP number: WP5.1
Task number:*

*Responsible institution: Edge-IT
Editor & and editor's address: Stephane Lauriere
Edge-IT
55, bd Saint-Martin
F-75003 PARIS, FRANCE*

Version 1.0 / Last edited by Sandrine L'Hermitte / November 23, 2010

Project co-funded by the European Commission within the Sixth Framework Programme		
Dissemination Level		
PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Revision history:

Version	Date	Authors	Institution	Section affected, comments
0.1	10/09/23	Stephane Lauriere	Edge-IT	Initial version
0.2	10/09/27	Stephane Lauriere	Edge-IT	Executive summary added
0.3	10/09/29	Stephane Lauriere	Edge-IT	CNR input added
0.4	10/10/11	Stephane Lauriere	Edge-IT	INRIA review taken into account
0.5	10/10/17	Stephane Lauriere	Edge-IT	Kerlabs, CNR, XLAB reviews taken into account
0.6	10/10/20	Stephane Lauriere	Edge-IT	Added NuGO external use case
0.7	10/10/29	Stephane Lauriere	Edge-IT	Added other external use cases
1.0	10/11/15	Sandrine L'Hermitte	INRIA	Added other external use cases

Reviewers:**Tasks related to this deliverable:**

Task No.	Task description	Partners involved[°]
T5.1.3	Exploitation plans	All - Edge-IT*

[°]This task list may not be equivalent to the list of partners contributing as authors to the deliverable

*Task leader

Executive Summary

This document summarises the plans of the XtreamOS partners for sustaining, evolving, disseminating and commercializing the technologies developed within the project. It emphasizes the uses of the technology internally by each individual partner, the way the XtreamOS components will be maintained and enhanced through upcoming initiatives, and the experiments/initiatives that were conducted externally.

The building blocks of the platform will be supported with the goal to widen their audience of developers and users. The LinuxSSI component will be maintained mostly by Kerlabs, with a strong focus on dependability improvement. XtreamFS, acknowledged as one of the most promising outcome of the project, will be supported and enhanced by at least BSC, Mandriva, NEC, SAP, ZIB, with a focus on I/O performance improvements. The XOSAGA implementation, the Virtual Node Framework and the checkpointing service will be made more Cloud-oriented: VUA will adapt XOSAGA to various existing Cloud APIs such as Amazon EC2; ULM will make the Virtual Node Framework adaptive to different environments and will optimize it; INRIA and UDUS will extend the XtreamOS checkpointing service by supporting adaptive replication for OSS and virtual machines. CNR will conduct further research on the vertical development of P2P/Cloud functionalities building on OverlayWeaver, the Service/Resource Directory Service, and on XtreamOS as a whole. The CDA service will be updated by STFC to deal with more flexible ways of authenticating. ZIB will support the Scalaris community, EADS the Amibe and the seccomp-nurse ones. BSC will build on the Application Execution Management. INRIA will continue to improve the Virtual Organization Service and will collaborate with ULM, XLAB, Mandriva and others for maintaining the XtreamOS software repository up to date and for bringing quality documentation both to developers and to users.

In order to extend the user community, the consortium has opened access to a testbed: an XtreamOS ready to use configuration. The partners are altogether ready to commit approximately 20 machines, and a total of 2 persons for contributing to the administration of the testbed.

Several partners have drawn a commercial analysis of new products that would build on the advantages brought by XtreamOS: Kerlabs studies the commercialization of a Kerrighed version harnessing the advanced checkpointing of complex applications; ZIB, Mandriva and XLAB are planning to draw revenues from XtreamFS; Mandriva, Redflag and XLAB are willing to derive benefit from XtreamOS in the field of Cloud Computing. However, the lack of wide support for the industry at this stage is identified as a weakness.

Beside commercial exploitation, several components of XtreamOS are of strategic value to several partners. SAP, EADS and EDF will pursue the monitoring of XtreamFS, XOSAGA and SSI. STFC intends to recommend XtreamOS for the UK National Grid Service and to set up a virtual market place for computing resources around it. Telefonica will keep assessing the integration of multimedia Grid powered applications into devices with limited computing power using XtreamOS MD flavour. DIXI, AEM, VOPS and RCA will be harnessed by XLAB for improving their internal infrastructure. Mandriva will prototype an engineering cluster based on XtreamOS.

Several follow-up projects have been set up: COOP on the adaptation of XtreamOS to the needs of HPC, ECO-GRAPPE, where the cluster flavour of XtreamOS will be used for a study on energy management, CONTRAIL for leveraging XtreamOS

for Cloud Computing, HEMERA focusing on AEM improvements and on several XtreamOS use cases, a joint EDF/INRIA PhD in the field of XtreamOS fault tolerance improvements and execution of XtreamOS in virtual machines, Compatible One where XtreamOS will be used for creating an innovative infrastructure-as-a-service layer, MosGrid, which will put XtreamFS in production in the field of astrophysics, and XtreamOS Easy, a support action for the maintenance of the XtreamOS testbed and of the XtreamOS packages.

XtreamOS has also been used and assessed externally: the Nutrigenomics Organisation Network of Excellence has evaluated the suitability of XtreamOS to upgrade its existing distributed IT infrastructure for data sharing and bio-informatics.

The partners have laid down strong foundations for continuing the scientific, technological and commercialization efforts around XtreamOS components. With its ability to provide a wide range of Cloud services, XtreamOS has the potential to become a major open-source asset of the Cloud Computing industry of the 2010 decade.

Contents

1	Introduction	3
2	Summary and analysis of the questionnaires	5
2.1	Motivation	5
2.2	Maintenance and evolution of XtreamOS	6
2.3	Open testbed	9
2.4	Exploitation paths	11
2.5	Business opportunities	14
2.6	Follow-up projects	17
2.7	Dissemination	19
2.8	Analysis	21
3	External use of XtreamOS	25
3.1	Assessment of XtreamOS by the Nutrigenomics Organisation Network of Excellence	25
3.2	Use cases finalists of the XtreamOS Computing Challenge 2010 . . .	26
3.3	Other external participants	26
4	Conclusion	29
A	Questionnaire template	31
B	Application of XtreamOS for development of NBX Grid	35

Chapter 1

Introduction

This report summarises the plans of the XtreamOS partners for sustaining, evolving, disseminating and commercializing the technologies developed within the project, and presents the use of XtreamOS by external organizations. XtreamOS technology transfer plan hinges both on the maintenance and on the evolution of individual components of the system, on the business exploitation of these components and on future projects sustaining the work required for facing emerging technological and scientific challenges in the field of distributed computing. The plans of each partner were gathered through a questionnaire (collected by e-mail) using the template that is available in Appendix ¹. The answers brought are summarized and analyzed throughout the first chapter of this document, which consists of the following sections:

- Motivation, a reminder of the initial interest of each partner for grid technologies and reasons why they got involved in XtreamOS,
- Maintenance and evolution of XtreamOS, a description of a strategy for keeping XtreamOS' main pieces active beyond the EC funding period,
- Exploitation paths, a presentation of the way each partner considers to draw scientific and technological benefits from the project,
- External use of XtreamOS, a summary of use cases that were experimented externally,
- Business opportunities, an overview of the commercial perspectives that are studied by some partners for building upon the technologies developed within the project,
- Dissemination, a description of the plans for disseminating the technology and for reaching out new users and developers.

The second chapter presents external uses of XtreamOS.

¹CDC partner did not provide any answer to the questionnaire since CDC is not involved in research activities – only in administrative and financial tasks

Chapter 2

Summary and analysis of the questionnaires

2.1 Motivation

In order to make sure the technology transfer plan is inline with the partners' objectives, the partners were asked to remind their motivation in participating in the XtremOS project. With the advent of Cloud Computing as a major shift in IT, all industry partners emphasized how much the collaboration with researchers was important for improving their know-how in the field. They were also very interested in taking part in a large international consortium for carrying out a continuous technology watch and critical observation of the industry trends. Their ultimate goal was to take advantage of the acquired know-how and integrated technologies for developing new business lines, for increasing their market shares or for improving their internal infrastructure. Researchers had the objective to improve the state of the art in large scale distributed systems. The table below gives the details of each partner's motivation.

Partner	Motivation for XtremOS
INRIA	Research on large scale distributed systems and Grid systems.
STFC	Interested in Grid computing, both at research and operational levels.
CNR	IT Research related to HPC, Distributed computing, P2P and Distributed Data Management techniques, Grids, Service and Cloud Computing architectures.
EADS	Evaluation of SSI solutions for large resource enablement at low cost. Demonstration of Grid technologies for future products.
EDF	Collaboration with researcher in Grid computing. Technology watch and industry trends analysis.
EDGE	Opportunity to reinforce expertise in the field of Linux SSI. Development of professional storage solution based on XtremFS. Collaborations with researchers in various fields of Cloud Computing.
NEC	Research and development on distributed file systems for clusters and grids. Research on SSI OS such as LinuxSSI. Collaboration with academic institutions.

SAP	Assessment of XtremOS concepts for operating enterprise system landscape. Identification of scalable, reliable and cost-efficient technologies which could improve the manageability of data centers.
BSC	BSC joined to perform research and potentially offer a useful product to the community.
ULM	Challenging research in the area of fault tolerance; ideal application of fault-tolerance mechanisms in Grid infrastructure services.
VUA	Research and development around execution environments for grid applications, in the context of an integrated grid operating system.
XLAB	Gain a competitive edge by improving our existing know-how. Improvement of our own infrastructure on which our software-as-a-service runs, as well as selling the expertise to major infrastructure-providing customers.
ZIB	Opportunities for research in grid computing and distributed (storage) system.
ICT	The XtremOS project has unique approach to design and implement the Grid and Cloud Computing technologies, which is based on the real applications and comprises of many potential research topics in distributed computing.
RED	Promotion of a flavour of XtremOS based on Redflag Software, initially within the China Academy of Science (CAS) and other Chinese universities. Business development through local ISVs around XtremOS based applications.
TID	Knowledge acquisition of Grid computing technology through partnerships with experts and potential application of Grid-computing paradigm to mobile devices world.
UDUS	Joint research on fault tolerance and data management in grids. Industrial partners providing real applications for evaluation of research.
KER	Improvement of the Kerrighed technology.

2.2 Maintenance and evolution of XtremOS

The success of the XtremOS technology transfer plan requires that all the building blocks of the platform are well maintained beyond the project, and keep evolving. This means that each component needs to be under the responsibility of at least one partner. The paragraphs below summarize the will of the partners to commit resources to specific XtremOS components both for developer community support and for the development of new key features.

LinuxSSI Kerlabs, NEC, Mandriva

Kerlabs is leading the development of Kerrighed and also animates the community through a Web site, mailing-lists and an IRC channel. The development and the code are public. In particular code repositories are setup at gforge.inria.fr and mirrors.git.kernel.org. Free community support is provided as far as time permits it. Other community members outside Kerlabs already participate in this support. Kerlabs also provides commercial support. Kerlabs will continue developing and supporting Kerrighed. The major feature that will be developed in the future is high availability capabilities, with the objective to improve the dependability of Kerrighed so that it could

be used in critical applications. Kerlabs is currently the only actor in this new development. NEC will continue to follow and evaluate LinuxSSI and might use it or pieces of it in future for offering solutions based on it or on components from LinuxSSI (like KDDM). Work by NEC related to LinuxSSI is: following the development, evaluating periodically. NEC might work on a simpler standalone distributed memory component, and will continue the work on a network abstraction which is more appropriate for RDMA capable interconnects than the current network stack of LinuxSSI.

Scalaris ZIB

ZIB will monitor the Scalaris community mailing-list and will keep providing support and will evolve the component in the frame of future research projects.

XtreemFS BSC, Mandriva, NEC, SAP, ZIB

BSC and ZIB will monitor the XtreemFS community mailing-list and will keep providing support in form of bug solving, helping new developers and develop new features. BSC will focus on making the replica management more proactive and thus achieve better I/O performance. One BSC researcher and 1 or 2 PhD students will work on this, which will result into new research articles covering XtreemFS advancement. As far as NEC is concerned, XtreemFS will not be used or deployed soon, as it still is in an academic software state which is useful for research but might be problematic for deployments in industrial environments. Nevertheless NEC plans to continue to follow the XtreemFS development and, in case of increased stability, to evaluate it and to consider its usage. NEC might contribute occasionally to the XtreemFS open source project. An XtreemFS related paper will be published soon by NEC. Mandriva considers developing commercial activities with hardware manufacturers willing to bind their solutions with an open-source distributed storage solution. Depending on the outcome of the evaluation, SAP might be interested in joining forces regarding further developments of XtreemFS. For SAP, XtreemFS offers an interesting alternative to expensive filer technologies currently used in hosting scenarios.

XOSAGA VUA

The XOSAGA implementation will be maintained. The work will be available via the common SAGA web site: <http://saga.cct.lsu.edu/>. Support will be provided via the mailing lists available through <http://saga.cct.lsu.edu/>. VUA plans to add support for Cloud platforms (Amazon EC2 interfaces) to the XOSAGA implementation. The goal is to integrate XtreemOS installations with clouds. VUA is currently seeking local funding to perform this work.

Virtual Node Framework ULM

ULM will maintain the Virtual Node Framework Software (VN). There will be further publications, a website, a tutorial on how to use the software. ULM will provide support in form of bug fixes. There will be no commercial support but, in case of large interest, an open source community will provide the support. ULM would like to extend VN to be adaptive to different environments, to improve the efficiency by optimisations, and to add Byzantine failure models to the system. ULM is open for future collaborative projects on these topics.

XtreemOS checkpointing service INRIA, UDUS

UDUS and INRIA plan to maintain all the components that are part of the XtreemOS checkpointing service, i.e. OSS, the container-based checkpointer, the xos-autoconfig tool. Improvement of the xos-autoconfig tool. UDUS will provide a separate Web site for OSS and grid checkpointing. Furthermore, UDUS will use/extend the code for future projects in the field of Cloud Computing. UDUS will re-use the code for future projects. The future developments will focus on the following improvements: support for memory-mapped files and adaptive replication for the OSS successor, support of virtual machines for Grid checkpointing.

Amibe EADS

The amibe mesher was existing before the project and will continue to exist after including the version developed for the XtreemOS project. Amibe is available as an open-source component and will continue to exist under the same license (see <http://jcae.sourceforge.net>). EADS might focus to interfacing to other open-source mesher to enlarge the scope of applications. Currently no volume capabilities are existing. Nevertheless, EADS scope will always be very large meshes – hence their participation to grid qnd/or high performance computing projects. The context will depend upon the opportunities varying for BU funded work to co-financed national or European projects.

seccomp-nurse EADS

seccomp-nurse will similarly be maintained. For seccomp-nurse, no new development will be pursued.

Dynamic Virtual Organization management service INRIA will maintain and evolve the VO management service.

Application Execution Management BSC

The Application Execution Management (AEM) service manages the distribution, execution, choice of resources, and eventually the migration of the different job execution entity. BSC will provide support in form of bug solving and helping new developers.

OverlayWeaver P2P framework (OW) CNR

CNR will further develop and maintain the framework. This matches the current strategic interests of ISTI toward P2P technologies, and allows to improve XtreemOS sustainability in the open-source community. In particular, the module for range queries and a gossip-based module developed for OW, but not used within XtreemOS, are going to be released to the OW community for integration within the OW platform. The range query and multidimensional query capabilities of OverlayWeaver are being further developed.

Service/Resource Directory Service (SRDS) CNR

CNR will keep maintaining and developing the SRDS code, and will provide the maintenance of the integration with the other services (RSS and Scalaris). Dynamic namespace allocation for SRDS is foreseen. Extension of the SRDS interfaces to upcoming standards for Clouds (e.g. new releases of JSDL and GLUE) are planned, as well as a stronger integration of SRDS with the security mechanisms of XtreemOS. CNR's

objectives are summarized as vertical development of P2P functionalities: (1) of OW as a framework for building overlays, (2) of SRDS as an overlay integration device and XtreamOS adapter, and (3) of XtreamOS as a scalable ubiquitous software solution to support scalable, dynamically evolving and secure computational platforms. CNR plans to reuse all the three levels for further research related to P2P and Cloud Computing. At the moment, CNR foresees the involvement of 1 junior developer for the next months, and 2 senior developers/researchers are committed in continuing the effort.

CDA STFC

STFC will revise and update the CDA component. The CDA service will be updated to deal with more flexible ways of authenticating. STFC is including authentication via Shibboleth identity management. It is planned to include other authentication models such as Kerberos. Another new feature planned is an autonomic policy management component.

XtreamOS software repository INRIA, Mandriva

Mandriva will keep updating the XtreamOS packages for the Mandriva distribution. As far as possible, the consortium will provide a selection of open-source developers to whom assign maintenance and guidance of the different XtreamOS packages (at least those that will not have a clear official developing institution), in order to ensure a sustainable transition to the open-source community.

XtreamOS documentation and tutorials INRIA, CNR, ULM, XLAB

INRIA and XLAB will provide tutorials and documentation as well as directions towards bug fixes. XLAB plans to provide help to the users in terms of documentation and on-line support. INRIA will provide tutorials and lectures on XtreamOS system at Master level. ULM plans to create a small tutorial video clip. XtreamOS will be taught as a topic by CNR in two different University courses in years 2010-2011 in the framework of the international, inter-University Ms.Sc. on Computer Science and networking, by University of Pisa and SSUP S.Anna.

XtreamOS Web site INRIA

INRIA will keep maintaining the XtreamOS Web site by publishing new releases, research articles and documentation.

General contributions Beside the contributions to individual components, several partners will keep developing the XtreamOS platform as a whole, in particular within the CONTRAIL follow-up project, aiming at leveraging XtreamOS for Cloud Computing. In addition, ICT will work on making XtreamOS more user-friendly. SAP will evaluate the XtreamOS components as technologies to improve the operation of SAP system landscapes.

2.3 Open testbed

In order to extend the user community, the Consortium has opened access to a testbed to external users in a first step. Doing so, external users will be able to experiment

XtreemOS on a ready to use configuration without the need to go through the complex process of installing XtreemOS on their own machines. In a second step, external users will be allowed to integrate their own nodes (core nodes as well as resource nodes) in the XtreemOS testbed. This would allow the testbed to grow as the user community grows. The structure aiming at managing this testbed will be established after the end of the project. This section describes the way the management structure will be organized, and the resources individual partners are willing to allocate to the testbed.

Management of the permanent XtreemOS open testbed The partners propose the set up of a consortium involving all the partners providing machines to the XtreemOS permanent testbed. A board committee will be set up and several engineers will participate in the maintenance of the testbed. A global administrator will be in charge of the testbed management. He will be assisted by local site administrators in charge of local nodes. For instance, Kerlabs machines will be administered by Kerlabs employees, with some administrative rights that may be delegated to an external, top-level administrator of the whole testbed. A representative of each partner would assemble (virtual meeting/telepresence) twice a year (possibly more frequent interaction of subcommittees) in order to manage the platform. CNR remarks that the management structure should allow any institution, including non-profit ones, to enter the platform with minimal costs beside the machine and man- power provided, and also act as a gatherer for institutions too small / single individuals which wish to join but cannot directly impact on testbed governance (and should probably be monitored/shepherded).

Resources The table below presents the human and computing resources each partner is ready to commit to the operational activities required by the testbed.

Partner	Resources
INRIA	INRIA will allocate resources to the management of testbed, and to the development of system management tool to maintain the testbed. Several computers will be provided and 1 engineer will be involved.
CNR	CNR is committed to sustaining the XtreemOS open testbed. The allocated resources will be kept, in number of machines (12 nodes) and manpower (about 3 man-months per year) in order to keep a constant core presence within the testbed. CNR plans to contribute more resources into the open testbed on a part-time basis, in the order of 100 nodes, by merging HPC resources from a collaboration with another institution. Leveraging the self-management properties of XtreemOS, resources will dynamically become available to authorized VOs. Additional manpower (6 to 12 man-months per year) will be employed to develop and maintain the system.
EADS	EADS is ready to allocate 1 man-month per year and approximately 5 desktops to the testbed.
EDF	The involvement of EDF R&D in the testbed will depend on the creation of another European project or ANR. If that happens, EDF R&D may be interested to participate indeed.

BSC	BSC is still evaluating how to solve the internal problems to share machines. BSC is interested but is not yet sure whether this will be possible.
VUA	Two recent machines of VUA will be dedicated to being integrated in the XtremOS open testbed.
XLAB	XLAB is ready to provide 2 nodes and one supporting person.
ZIB	2 virtual machines for one year will be allocated.
ICT	ICT will participate. Some students will be involved, and ICT will provide several machines.
RED	Redflag would like to provide Asianux with XtremOS integrated as the base of testbed.
UDUS	UDUS will provide one machine for the testbed, beyond the project end. UDUS can provide a couple of hours per month for maintenance of this machine.
KER	Kerlabs will continue participating in the testbed as far as unused resources are available. Up to 8 bare metal servers will be made available when unused for other purposes. Reservations of these machines will be made possible too. Up to a quarter-time administrator will be allocated.

2.4 Exploitation paths

Both industry and academic partners have presented the scientific and technological exploitation paths they are considering for the XtremOS components. This section gives an overview of these individual exploitation paths. Some of them may lead to business opportunities, described in Section 2.5.

STFC XtremFS and XtremOS as a whole will be recommended for the UK National Grid Service. A virtual market place for computational resources will be set up. This is a demo application, acting as a meta-broker that in addition that in addition to the normal requirements of a job, takes into consideration financial (cost) information. It would be exploited in two ways: to show the potential use of XtremOS in business, and as an interoperability case between XtremOS and other Grid middleware such as gLite. STFC has developed this component. This is a result which can stand on its own.

CNR CNR will in particular exploit the following components:

- Service/Resource Directory Service (SRDS), software module which was developed by CNR and can be used stand-alone (with other modules as subsystems). SRDS will be reused and extended within other research projects, whose funding paths are under investigation by CNR.
- DDT set of P2P algorithms, which was developed by CNR and contributes to reducing update network traffic in DHTs holding dynamically variable content. It is applied within the SRDS.
- XCONE, a P2P algorithm to implement scalable multiattribute search in DHTs; developed by CNR, but not yet applied within XtremOS.

- OverlayWeaver software modules implementing DDT and XCONE, developed by CNR, will contribute to the open source P2P community. The modules for range query and multi-attribute selection will be developed by internal funding and will help the sustainability of the system.
- XtreamFS, software module for which CNR took part in design and first implementation, and in testing of the second implementation. Can be used stand alone.
- XOS-GATE, application and software architecture for large scale SPECT simulations using the GATE simulator and heterogeneous XtreamOS resources. It was developed by CNR and University of Pisa. It can improve the resource usage for GATE users and it can be a prototype for further scientific applications to be ported on top of XtreamOS. A collaboration with the GATE user community and Applied Physics is being studied by CNR.

EADS Three main components are of interest to EADS for scientific computing in a distributed infrastructure:

- The Single System Image of XtreamOS that enable large jobs for legacy software that usually run on a single node;
- GridFS that enable collaboration between sites
- SAGA for enacting jobs using standards

EADS will pursue the monitoring of those components – an industrial usage will occur under the condition inclusion that the software is included in the standard Linux distribution. For all the previously mentioned components, EADS has contributed only indirectly. Those components provide greater agility to the processes in which there are embedded in. The seccomp-nurse has been fully developed by EADS Innovation Works. Exploitation will only be possible through industrial products (Linux distribution) including those components. Further experiments will be pursued internally to assess the maturity of the previously described components. Those testing might be done through Grid5000 or any other virtual platform that will propose XtreamOS images.

EDF R&D EDF R&D plans to continue to follow the further development of XtreamOS and if it is widely accepted in the Linux community. EDF will test it again and examine the opportunity to include it to their information systems. The becoming of XtreamOS as a whole particularly interests EDF. Its active technological survey will keep an eye on the usability of the distributed system and concepts. SAGA specifications are particularly promising and how large SAGA is adopted by the users will be carefully observed. EDF contributed directly to none of them but were involved in the definition of the requirement as well as the tests of the installation of the distribution and the porting of some of their applications to evaluate the maturity of the solution.

Mandriva Mandriva will use the SSI technology developed by Kerlabs for its cluster professional offering, and duplication tools which are developed and maintained by Mandriva. The node duplication capabilities of XtreamOS will be exploited for deploying a Linux image on large number of machines in a raw. The capabilities of

XtreemFS will be exploited internally at Mandriva for the set up of a build and analysis cluster, and possibly externally for creating storage clouds for customers. The integration will take place in particular in the national Compatible One project, which is further described in the future projects of Section 2.6.

NEC NEC has participated in the software architectures of XtreemFS client and server, in KDDM, running standalone, and in the OSS components from WP3.4. Direct benefits to NEC are unclear. Indirect benefits are linked to NEC increased knowledge and might result into products inspired by the know-how or reusing the components.

SAP SAP plan to exploit XtreemFS and checkpointing/restart. XtreemFS can stand on its own. Scalability of XtreemFS may offer an interesting alternative to avoid expensive filer up-scaling. XtreemFS may allow for exploiting scale-out effects using commodity storage hardware. Checkpointer: depends on the level of the checkpointer hierarchy. XtreemFS may offer a highly scalable file system for applications accessing data distributed over wide area networks. CR may offer a possibility for coordinated checkpoints of distributed applications. CR might be useful to facilitate and improve system management operations.

ULM The Virtual Node software architecture will definitely be reused in the future. ULM implemented a bunch of scheduling algorithms and is planning to invent some others. At the end of the project, ULM will have a tool for converting ordinary applications into fault-tolerant ones. ULM is investigating whether the Distributed Server software (DS) can be used in conjunction with Virtual Node (VN). This was already tried during the project, but the latest release of DS was not integrated into VN. ULM directly contributed to VN and partially to DS. Both components can stand on their own. There is no direct benefit for the Grid community. It can however be used for the implementation of infrastructure services as shown within the XtreemOS project.

VUA VUA will exploit in particular the three following pieces:

- **XOSAGA API specification:** VUA has directly contributed to the XOSAGA API. It can stand on its own if needed. Benefit for the grid community and the OGF standardization body: extension of the existing SAGA API according to grid forum document GFD.90. XOSAGA interfaces will be exported in the framework of the Ibis platform (<http://www.cs.vu.nl/ibis>) objectives: provided standardized API's to Ibis applications. Funding type: various Dutch national funding agencies (NWO, Ministry of Economic Affairs, etc.)
- **RSS:** VUA has directly contributed to it. The RSS itself is tightly coupled to XtreemOS but specific internal algorithms can stand on their own. VUA plans to develop the use of similar algorithms for other large-scale decentralized system designs.
- **Distributed servers:** VUA has directly contributed to it. It can stand on its own. VUA plans to use this technology as a support for service mobility in future research.

XLAB XLAB intends to harness DIXI, AEM, VOPS, RCA, XtreamFS, Scalaris, monitoring and auditing. XLAB directly contributed to security, DIXI, AEM, monitoring and auditing. DIXI is a stand-alone component, monitoring and auditing are currently coupled with AEM and VOPS, but can be easily decoupled, the rest are stand-alone. The components give XLAB the tools needed to both exploit their infrastructure and software better, as well as provide the foundation for our service providing towards customers. XLAB will also exploit XtreamOS as a whole system, where they will provide services on top of the operating system (i.e. site deployment services). Furthermore XLAB will use developed components as a building blocks for other products and services, which will yield new revenue and markets for the company.

ICT The expected benefit for the Grid community or standard bodies for ICT is a Grid platform with better performance and better integration with the Operating system. It will help ICT design the future distributed system research, and the projects that can be exploited further for real useful applications.

Redflag XtreamFS might be useful for some customers of Redflag to set up distributed computing environment if it is stable enough.

Telefonica I+D The main result that could be exploited by Telefonica is the integration of multimedia applications with Grid-assisted content transcoding, applied to devices with limited computing power. The potential benefit for end-users is that they will be able to run Grid/Cloud-aware multimedia applications. The GridPlayer application for mobile devices would be directly exploited, with the objectives to apply Cloud Computing technologies to multimedia mobile device applications.

UDUS Currently, UDUS is trying to use OSS for bio-informatics applications and in the field of software engineering (model checking) UDUS believes it can help to simplify the development of distributed and parallel applications.

2.5 Business opportunities

INRIA INRIA will establish new research collaboration with XtreamOS former partners as well as nationally and internationally. An XtreamOS spin-off will be investigated based on the results obtained in the next two years.

Kerlabs Kerlabs will exploit commercially the advanced checkpointing of complex applications, full-capacity communications after process migration. Kerlabs contributed most of advanced checkpointing features in Kerrighed, and all of full capacity communications features. All above-mentioned results can be exploited separately, although they are integrated in the same software and can benefit from each others. These results will allow Kerlabs to enlarge the range of applications and customers that can benefit from fault-tolerance and advanced resource management through checkpoint/restart; improve the performance of communicating applications, making Kerrighed suitable for parallel message-passing applications and for Internet servers. It is planned to further improve the range of features as well as the quality of Kerrighed's enhancements, but mostly on customers' demand. Kerrighed software suite for Internet servers will be the main product, and will be commercialized with integration and support services.

Kerlabs' competitive advantage is its ability to allow users to use aggregated resources in a simplistic manner, without requiring dedicated hardware. The core Kerrighed software is publicly available through a community website (<http://www.kerrighed.org>), and its development can be followed/contributed to through public code repositories at gforge.inria.fr and mirrors.git.kernel.org. Note that kernel.org is the reference site hosting Linux kernel developments and Linux kernel releases. A SWOT analysis of the path to commercialization of an enhanced version of Kerrighed taking advantage of the XtremOS developments is summarized in the table below.

Competitors	Existing competitors all base their solutions on dedicated hardware: <ul style="list-style-type: none"> • ScaleMP http://www.scalemp.com • Numascale http://www.numascale.com/ • 3leafsystems http://www.3leafsystems.com/
Threats	Major threats come from hardware competitors providing low-cost hardware.
Weaknesses	The major weakness of Kerlabs' solution is the lack of high-availability in Kerrighed. It should be noted that competitors also have this weakness.
Intellectual Property	Intellectual property of Kerlabs' software is protected by the General Public Licence and its derivatives, which are well established world-wide.
Distribution channels	Direct distribution is currently the only channel considered.
Sale target	SMEs as well as big companies are targeted.

ZIB Exploitation opportunities will focus in particular on XtremFS, Scalaris. Theoretical algorithms will be studied, in the field of scalable resource detection. As an alternative to manual file management (gridFTP) or non-POSIX grid storage systems, XtremFS will provide a benefit to the Grid community.

Redflag As an OS vendor, Redflag thinks some XtremOS components integrated in their products may be attractive for customers. Cloud Computing is a possible market. Competitors: Hadoop. Weakness: Lack of wide supporting from industry.

Telefonica I+D The main business opportunity is on the side of applying cloud/grid computing to multimedia applications on mobile devices. However, no specific product/service has been defined yet with business units.

XLAB XLAB will improve the ISL Light product (ISL Grid technology) and the Gaea+ 3-D reconstruction service. ISL is in production since 2007, and being further developed; currently marketed as Cloud-based communication¹. The targeted markets and customers are all existing and new ISL Light customers. The SWOT analysis of

¹<http://www.islonline.com/technology/>

the path to commercialization of enhanced versions of ISL and Gaea+ 3-D is presented in the tables below.

Product that will be commercialized	ISL
Unique competitive advantage	ISL Grid technology, integrated into the ISL Light product with the knowledge gained in the XtremOS project, allows ISL Light modules to run geographically distributed, improving the responsiveness, reliability, and enhancing the capacity of our underlying infrastructure. .
Competitors	Teamviewer Host, Citrix GoToMyPC, Laplink Everywhere
Weaknesses	Small and specialized product line, i.e. lack of capability to sell all-encompassing solutions to large customers
Intellectual Property	No specific measures for IP related to this feature (ISL Light is commercial, proprietary software)
Distribution channels	Integration into existing product
Sale target	No separate target set for this functionality - target for the whole product line apply.

Product	Distributed computing backbone for Gaea+ 3-D reconstruction service (http://www.gaeaplus.si/sl/3d-rekonstrukcija/predstavitev)
Maturity	In development
Targeted markets and customers	Municipalities, urbanism, architects; simplified version for physical entities
Unique competitive advantage	As part of Gaea+, XLAB is developing service for reconstruction of 3-D objects from sets of photographs. This involves computationally and data-intensive processing, thus XLAB will evaluate whether XtremOS as a whole or parts of it (e.g., AEM, XtremFS) can be exploited for this task.
Competitors	Google as well as Microsoft are supporting similar development.
Threats	Above-mentioned companies or other competitors launch similar services sooner than XLAB; competing technologies, e.g. laser scanning, prove to be more successful than photo-based reconstruction.
Weaknesses	Some further development is required to reach the goal of reliable reconstruction without user intervention; high computational complexity
Intellectual Property	The product will only be offered as software-as-a-service.
Distribution channels	The offer will be integrated into an existing service.
Sale target	The targeted turnover in three years amounts to 150000 euros.

Mandriva Mandriva will consider creating a professional solution around XtremFS, in partnership with hardware manufacturers. In addition, Mandriva may set up a wider professional offer around XtremOS, in collaboration with other XtremOS industry partners.

2.6 Follow-up projects

This section gives an overview of several industry or research initiatives that have been undertaken for adding new features to one or several XtremOS components, for improving key non-functional characteristics of the system, or for putting into practice the platform for industrial use cases.

COOP INRIA

Project funded by ANR (started in March 2010, 3 years). Exploitation of XtremOS system as a whole to study the interactions between HPC runtimes and resource management systems. INRIA resource: 1 research engineer, working on the adaptation of XtremOS to the needs of HPC runtimes for better interactions, experimentation of SALOME applications on top of XtremOS, experimentation of TLSE on top of XtremOS.

ECO-GRAPPE INRIA, Kerlabs, EDF

Project funded by ANR (started in December 2009, 3 years) focusing on the exploitation of the XtremOS cluster flavour for a study on energy management in clusters. 1 INRIA PhD student is working on the improvement of the cluster flavour towards energy conservation, with the technical support of Kerlabs and EDF. Kerlabs develops the platform, and both EDF and Kerlabs implement use cases.

CONTRAIL INRIA, CNR, STFC, VUA, XLAB, ZIB

European project (starting Oct 2010, 3 years) coordinated by INRIA. The objective is to leverage XtremOS for Cloud Computing, evolving the components to a flexible open-source system for Cloud federation supported by powerful tools and advanced APIs. The focus will also be brought on the ease to deploy and administer the system, and on the system robustness.

HEMERA INRIA research initiative

Execution of various distributed applications in the framework of XtremOS. Research work on job and resource management is conducted, that may lead to improvements in AEM. Experimentations of various Grid applications are carried out on top of XtremOS.

Collaboration between INRIA and EDF INRIA, EDF R&D

Collaboration between INRIA and EDF RD on reliable and timely execution of HPC applications in virtualized distributed architectures. 1 PhD student. Exploitation and improvement of some XtremOS fault tolerance features, extension of XtremOS for executing applications in virtual machines and/or containers.

EDF HPC Portal HPC Portal is an internal EDF R&D project running in 2010-2011 and aiming at providing an easy access to the internal HPC resources.

PhD thesis PhD thesis to start in October 2010 on dynamic adaptation of XtreamOS services (partially funded by the Brittany regional council)

XtreamOS Easy INRIA

Development action funded by INRIA, focusing on the open testbed maintenance and management, the maintenance of XtreamOS packaging for popular Linux distributions, the user and developer community support, the development of system management tools and improvement of the GUI. 2 engineers are devoted to this initiative. The work will result in an improvement of XtreamOS robustness and user interface, extension of the user community, sustainability of the developer community, further dissemination of XtreamOS system as a whole.

Collaboration with the University of Chicago Informal collaboration with Kate Keahey's research group (Argonne, University of Chicago) on the topic of studying XtreamOS in the context of Sky computing (federation of clouds, Nimbus IaaS software). This will result into improvements of XtreamOS automatic deployment, dissemination of XtreamOS system as a whole.

MosGrid ZIB

National German R&D project with film business, joint astrophysics project with Astrophysical Institute Potsdam. The objective is to use XtreamFS in production for distributed data repositories, highly scalable management of metadata.

Compatible One INRIA, Mandriva

Compatible One is a collaborative research project starting November 2010 until October 2012. Its goal is to create an open-source stack of APIs and services featuring infrastructure as a service and platform as a service, and compatible with a wide range of existing open-source cloudware. In this frame, Mandriva plans in particular to reuse, benchmark and evolve XtreamFS. The other partners are Bull (leader), eNovance, Institut Telecom, INRIA, Nexedi, Nuxeo and others. The Compatible One cloudware stack will be validated against a set of use cases including the set up of a cloud software engineering for building, testing, analysing very large software systems (typically a Linux distribution such as Mandriva).

Other initiatives Beside the projects listed above, the following initiatives have been undertaken:

- CNR is formalizing a collaboration with the new ITC center of the University (CNR-Unipi agreement ongoing) on the topic of XtreamOS as a dynamic service platform / XtreamOS on virtualized platform for HPC. The objective is to explore the use of XtreamOS and of Kerrighed-SSI as tools for HPC in a virtualized HPC environment. The expected impact on XtreamOS is the following: enhance visibility and availability of the platform, increase project momentum. In addition, CNR is setting up a local collaboration aiming at increasing the availability of SPECT simulation software by gathering distributed resources into an XtreamOS platform. It will consist in developing and deploying large scale

SPECT simulations with adaptive load balancing over heterogeneous resources. CNR will also investigate the testing of XtremOS in collaboration with other IT centres (CASPUR) which may lead to national project proposals.

- Mandriva has submitted a project to the French Equipex call, aiming at large computing facilities. The goal is to create an operating system for a super-computer supporting virtual organizations, advanced task scheduling and large storage capabilities. The work will directly harness several components of the XtremOS stack.
- UDUS plans to use grid checkpointing in a future project (no proposal submitted so far) in the context of reliable Cloud Computing, maybe at a national or international (EU) level. Furthermore, UDUS plans to use OSS in a future national project (project proposal under preparation, national level, German Science Foundation). UDUS is in contact with other people from UDUS not participating in XtremOS. Prof. Martin Lercher (Bio-Informatics), Prof. Bill Martin (Biology) and Prof. Michael Leuschel (Software Engineering).
- ULM will try to get funding in national projects to extend VN. The expected impact on XtremOS is an enhancement of VN software that could be re-integrated into XtremOS, e.g. AEM and CDA integration.
- Xtrem BioLinux: STFC has submitted a research proposal as a national project to the UK Natural and Environmental Research Council aiming at an integration of XtremOS with the BioLinux system. The expected impact on XtremOS is an adoption of XtremOS by an established community; in this case, the bioinformatics community using BioLinux.

2.7 Dissemination

Publications

- INRIA encourages researchers in the system/cluster/Grid community to use XtremOS as a system research vehicle
- UDUS plans to submit a journal paper in fall 2010, to Wiley “Software Practice and Experience” about OSS.
- ULM will submit further publications about VN work.
- EDF R&D plans to publish its currently ongoing work on the porting of OpenTurns on XtremOS.
- CNR plans to submit new papers and demos related to XtremOS.

Software distribution Mandriva will keep mirroring all the packages of the Mandriva XtremOS distribution consisting of release, updates and backport packages. Distributed computing and Cloud initiatives will also be contacted by the Consortium partners for promoting the adoption of XtremOS aside existing Linux OS distributions.

Community uptake CNR proposes the set up of a real foundation or formal international committee, with research institutions/companies represented and taking charge of expenses in order to maintain source code development available to the open community (maintaining lists, blogs, web site etc). CNR is ready to participate in this kind of activity. A full specification of the XtreamOS addenda as a generic complement to a Linux distribution and a formal definition of their configuration would be the first target. This broader strategy would also include the open testbed action, but could more directly address industrial use-cases and fulfil the needs of potential key users (e.g. specific scientific or industrial communities, like the bioinformatic one that was proposed by other partners).

Events Redflag plans to promote XtreamOS during the future marketing events it will attend, e.g. the North-East Asia OSS forum.

Experimentations INRIA will encourage researchers needing huge computing power to use XtreamOS system to experiment their applications.

Demonstrations

- ULM will prepare a demonstration for the final review that will be also available as a video clip. This will be shown at various occasions.
- XLAB is developing a workflow-based demonstration of XtreamOS used in a business environment, as well as multiple technical demos showing particular features of XtreamOS; some of these demos are already published on YouTube and on the project web site.
- CNR will continue showcasing the XtreamOS technology in the future as long as it is maintained. CNR will promote XtreamOS among local start-up companies.

Internal dissemination

- NEC will organize internal seminars on Kerrighed, KDDM, network abstractions and technologies used in XtreamFS.
- Redflag's product marketing department will communicate with other departments, such as sales and technical support, about XtreamOS solutions and look for possible customers.
- SAP plans to present XtreamOS to internal product groups.
- Telefonica's internal dissemination is being carried out by promoting XtreamOS as a whole by using as main promotion tools the XtreamOS portal and selected demos to internal experts.
- At various internal occasions, ULM is providing insights to the XtreamOS project. Those occasions are Open Day of the department, career events for students, internal information days etc.
- XLAB widely uses an internal XtreamOS mailing list inside XLAB. Furthermore, XLAB regularly organizes internal tutorials on the XtreamOS components most interesting to XLAB, e.g. monitoring, AEM etc.

2.8 Analysis

This sections analyses the situation of each major component of the XtreamOS stack regarding its chances to keep alive beyond the administrative end of the project. For each component, the table below lists the partners who are committed to support users and developers, the ones who are interested in exploiting technologically or commercially the component, the research projects that will sustain the effort, and the projects that are still under investigation at the time of the report.

Component	Support	Exploitation	Commercialization	Future projects	Potential projects
LinuxSSI	KER, NEC, MDV	KER, NEC, MDV, EADS	KER, MDV		
XtreamFS	BSC, MDV, NEC, SAP, ZIB	BSC, MDV, NEC, STFC, CNR, EADS, SAP, XLAB, ZIB, RED	MDV, XLAB, RED, ZIB	Compatible One, Contrail, MosGrid	MDV national project under evaluation
XOSAGA	VUA	EADS, EDF, VUA			
Scalaris	ZIB	XLAB, ZIB			
Virtual Node Framework	ULM	ULM			ULM project under preparation
Checkpointing service	INRIA, UDUS		INRIA, SAP, UDUS		UDUS EU project under preparation
OSS	UDUS	UDUS			UDUS project under preparation
Dynamic VO management	INRIA	INRIA			
AEM	BSC	BSC, XLAB	XLAB	Hemera	
Amibe	EADS	EADS			

seccomp-nurse	EADS	EADS			
Overlay-Weaver P2P framework	CNR	CNR		Internal CNR funding	
Service and Resource Directory Service	CNR	CNR			CNR project planned
CDA	STFC	STFC			
XtreemOS GATE	CNR	CNR			CNR project under investigation
DS	ULM	ULM, VUA			
CR		SAP			
RSS	VUA	VUA			
DIXI	XLAB	XLAB			
VOPS	XLAB	XLAB			
RCA	XLAB	XLAB			
GridPlayer for mobile devices	TID	TID			
XtreemOS as a whole	INRIA, EDF	INRIA, EDF, MDV, RED	INRIA, MDV, RED	Contrail, COOP, ECO-GRAPPE, INRIA / EDF collaboration, XtreemOS Easy, collaboration with the University of Chicago	CNR project, MDV national project, STFC Xtreem BioLinux

The components can be distributed in three main categories summarized in the table below. The consortium managed to make sure that all components will be supported by at least one partner. This reduces the risk that any of them becomes unusable and obsolete. Besides, XtreemOS as whole will be maintained and evolved within several future projects, in the frame of which all components will keep evolving as an integrated system.

Category	Description	Components	Total
A	More than three partners plan to support the component, and the commercialization of a product is planned	LinuxSSI, XtremFS	2
B	One or two partners plan to support the component, and at least one future project related to the component is planned	AEM, Checkpointing service, GATE, OSS, OverlayWeaver P2P framework, Service and Resource Discovery Service, Virtual Node Framework	7
C	One or two partners plan to support the component, but no future project related to the component is planned	Amibe, CDA, DIXI, DS, Dynamic VO Management, RCA, RSS, VOPS, Scalaris, seccomp-nurse, XOSAGA	7

Chapter 3

External use of XtremOS

In addition to the commitment of the consortium partners to support individual XtremOS components, the XtremOS community has been broadened through the dissemination actions that have been conducted during the project's life. These dissemination actions resulted in the assessment of the XtremOS technology by several organizations external to the project's consortium.

3.1 Assessment of XtremOS by the Nutrigenomics Organisation Network of Excellence

The Nutrigenomics Organisation Network of Excellence (NuGO) evaluated the suitability of XtremOS to upgrade its existing distributed IT infrastructure for data sharing and bioinformatics. The work was conducted at the University of Aberdeen Rowett Institute of Nutrition and Health and the Mario Negri Institute in Milan, who accepted that their report were appended to this document, in appendix. The summary of the report is the following:

NuGO used an NBX (NuGO Black Box) Beowulf SSI (Single System Image) cluster constructed using COTS (Commodity Off The Shelf) hardware and FLOSS (Free/Libre Open Source Software). The stability of the Kerrighed SSI cluster was severely compromised when later 2.4 versions of the Kerrighed software were tested on the NBX cluster under 32-bit NuGO-Linux. However, the older 2.3 version of Kerrighed was stable enough to be considered for use as a production platform. Stability problems prevented a detailed evaluation of the full 'grid' functionality of the XtremOS-G layer, but experiments concerning cooperation between NBX systems connected via the WAN (Wide Area Network) demonstrated that Kerrighed might provide a load-balancing option for 'embarrassingly' parallel applications where little or no communication occurs between processes. The object-oriented XtremFS filesystem was also evaluated and performed very well. However, concerns about the requirement for multiple network ports to be open in the perimeter firewall at NuGO partners limit the practicality of this to be considered as a replacement for the sshfs user-space filesystem that is currently in use for data sharing between NBX's over the WAN. It seems unlikely that Kerrighed will be sup-

ported beyond version 2.3 on the 32-bit Linux platform, which restricts its usefulness under the 32-bit NuGO-Linux currently installed on the NBX network. However, the NBX network will be upgraded to 64-bit NuGO-Linux and more recent versions of Kerrighed will then offer much greater stability on the 64-bit platform. Nevertheless, Kerrighed 2.3 provides a 32-bit SSI platform compatible with the existing NBX infrastructure that may be useful to aggregate the CPU and memory resources of multiple NBX's connected locally via a high-speed LAN. This was demonstrated as a practical way of improving the performance of the GenePattern bioinformatics software running on the prototype NBX Beowulf SSI cluster constructed for this evaluation.

3.2 Use cases finalists of the XtremOS Computing Challenge 2010

An XtremOS computing challenge took place during the Euro-Par 2010 conference. It increased the pool of demonstrable use-cases and enlarged the user and developer community. This computing challenge provided students with the opportunity to test their applications on a Linux-based Grid operating system and perform large-scale experiments on the Grid. Researchers, PhD and M.Sc. Students were invited to compete to run their applications on the XtremOS Grid operating system (the announcement was widely advertised world-wide). The following list is the ranking of the finalists:

1. "Monte Carlo Simulation for Single-Photon Emission Computed Tomography", Emanuele Carlini, Sebnem Erturk, Giacomo Righetti, University of Pisa, Italy. A demo video is available at the following address: <http://www.youtube.com/watch?v=o2183WfVBuk>
2. "Grid Security Operation Center", Syed Raheel Hasan, Jasmina Pazardzievska, Maxime Syrame (supervised by Julien Bourgeois), Laboratoire Informatique, Universite de Franche-Comte, France. A demo video is available at the following address: <http://www.youtube.com/watch?v=VUUaRpzusdo>
3. "Parallel Kriging", Alvaro Parra, Exequiel Sepulveda, and Felipe Lema, ALGES lab at Universidad de Chile, Chile. A demo video is available at the following address: http://hpc.isti.cnr.it/xos/XtremOS_Challenge_Video.avi
4. "XtremOS@home", Zubair Nabi from Lahore University of Management Sciences, Pakistan. Due to other commitments, the participants from Lahore University of Management Sciences, Pakistan had to bow out of the challenge and did not submit any video demo.

3.3 Other external participants

Beside the external uses mentioned above, XtremOS has been used in the following contexts:

- Prof. Lonnie Cumberland - Physicist at US NIST Lab, New Mexico - tested and deployed a 25 node XtremOS setup with 4 servers running Kerrighed and the

rest of the machines as clients for accessing the Kerrighed cluster. The purpose was to allow graduate students to run larger scale experiments and data sets in scientific applications.

- A PhD student at INRIA uses XtremOS for developing his PhD thesis on autonomous services in Grids.
- A PhD student at INRIA uses XtremOS for developing SAGA applications.
- 2 master students at VUA used specific components of XtremOS for their respective master projects.
- 2 ZIB graduate students at ZIB installed and tested XtremFS within the MOS-Grid project.
- 2 XLAB PhD students experimented XtremOS. The members of XLAB established 2 spin-off businesses to which XtremOS is beneficial: Dimenzija, with Gregor Berginc as the CEO, is developing a distributed version of the 3D object reconstruction from multiple views, where the parallelisation through jobs submitted to XtremOS will be beneficial. Second, Alanta, with Ales Stimec as CEO, which focuses primarily on Cloud Computing, is considering building solutions on top on XtremOS.
- Several persons at the CNR Physics Department and at the Computer Science Department tested and used XtremOS. A PhD student ported a GATE-based SPECT simulation to XtremOS, used to gather data which are the subject of her Ph.D. Thesis. A Ms.Sc student designed XtremOS images to be deployed with the Octopus software. Another Ms. Sc. student conducted a preliminary evaluation of the PBS integration into XtremOS on a cluster. INRA
- Pauline Ezanno is a researcher at INRA (the French National Institute for Agricultural Research) and her research relates to "mathematical modelling and epidemiology" (for example, Modelling of the spatio-temporal spread and persistence of virus in a bovine metapopulation accounting for heterogeneity in immune statuses). She is often limited by the power of our computational servers when experimenting even at a regional level and will try her simulations for stochastic dynamic models (developed on Scilab) on XtremOS.

Chapter 4

Conclusion

This document completes a series of reports focusing on the market analysis of distributed computing and on the XtreamOS technology transfer plan. It emphasizes the uses of the technology internally by each individual partner, and the way the XtreamOS components will be maintained and enhanced through upcoming initiatives. It shows that the partners are aware of the remaining challenges the XtreamOS stack still needs to face for being completely ready for commercial take-off. The partners have laid down strong foundations toward this goal: a plan for maintaining a dedicated open testbed has been defined, more than 10 new collaborative research projects revolving around XtreamOS have been set up for the 2010-2012 period, and a large number of activities have been undertaken for disseminating the work both within each participating organization and across the software industry. With its ability to provide a wide range of Cloud services including resource mutualisation, a distributed file system and transparent external resource access, XtreamOS has the potential to become a major open-source asset in the Cloud Computing industry of the 2010 decade.

Appendix A

Questionnaire template

The questionnaire used within the Consortium for gathering the exploitation plans of each partner is appended hereafter.



Partners' exploitation plans - Questionnaire

Objective: Collect your individual exploitation plans.

Please answer the below questions (at least a few lines per question).

Name of your company or institution:

Industrial partner (only): brief description of your company business model (if available, please include a picture) + currently offered products and services, as well as services customers ask for.

What is the main reason you joined the XtreemOS project?

1. Involvement in XtreemOS after the end of the EC funding period

- Will you maintain the components you contributed to? (Please indicate any specific action, e.g. individual website, blog, tutorial, notes, publications etc)
- Will you provide support for the component(s)? (open source community...)
- Will you work on further development(s)?

- Which new features do you plan to develop?

- Objectives?

- Context and people involved?

- Will you continue participating in the XtreemOS open testbed?

In order to extend the user community, we plan to open access to this testbed to external users in a first step. Doing so, external users will be able to experiment XtreemOS on a ready to use configuration without the need to go through the complex process of installing XtreemOS on their own machines. In a second step, we plan to allow external users to integrate their own nodes (core nodes as well as resource nodes) in the XtreemOS testbed. This would allow the testbed to grow as the user community grows.

- Resources to be allocated (Effort and machines)?

- What would be the most appropriate structure to manage this permanent XtreemOS open testbed?

2. Identification of exploitation opportunities by your company/institution

- Name the results that could possibly be exploited by your company/institution, classifying them according to the following categories.

- Conceptual or software architectures

- Software modules, components software

- Specifications

- Theoretical algorithms

- Tools

- Other (specify)

- For each exploitable result, indicate:

- if you have directly contributed,

- if the result can stand on its own or on the contrary is tightly coupled into the whole system,

- the expected benefit for the Grid community or standard bodies (academic partners only),

- the benefit it creates for your company, your customers and / or partners (industrial partners only).

3. Actions taken towards exploitation of results

How will you exploit XtreamOS results?

- Integration of XtreamOS functionalities/results into another system/project?

XtreamOS result exploited:

Framework (national/international project, collaborations...duration):

Objectives:

Funding type:

Expected impact on XtreamOS:

- Adaptation of XtreamOS functionalities/results?

XtreamOS result exploited:

Framework (national/international project, collaborations...duration):

Objectives:

Funding type:

Expected impact on XtreamOS:

- Experimentations/Testing/Validation on XtreamOS system?

XtreamOS result exploited:

Framework (national/international project, collaborations...duration):

Objectives:

Funding type:

Expected impact on XtreamOS:

- Collaboration with external people on XtreamOS?

Please describe ongoing work with people outside the consortium

- *Who's involved?*

- *Framework:*

- *Topics:*

- *Objectives:*

- *Expected impact on XtreamOS:*

Please describe future work with people outside the consortium:

- Other follow-up activities or projects? *(Please describe)*

Expected impact on XtreamOS:

- Business opportunities/ideas?
- Commercialization?
 - Product/Service that will be commercialized?
 - Maturity of product/service?
 - Targeted markets and customers?
 - Unique competitive advantage & differentiating characteristics?
 - Competitors?
 - Partners?
 - Threats?
 - Weaknesses?
 - Specific measures for protection of Intellectual Property?
 - Distribution channels? (*Creation of start-up? Integration to existing products...*)
 - Sale target?
- Dissemination activities
 - Publications?
 - Demonstrations?
 - Participation in conferences, keynotes and invited talks?
 - Courseware?
 - Training activities?
 - Other? (Please give details)

What kind of **internal** dissemination activities will you carry out to communicate the XtremOS results and exploitable solutions?

Other comments:

Appendix B

Application of XtreamOS for development of NBX Grid

The outcome of the XtreamOS feasibility assessment carried out by participants in the Nutrigenomics Organisation Network of Excellence (NuGO) is appended hereafter.

Application of XtreamOS for development of NBX Grid

Feasibility assessment report

Alicia Mason^{*}, Luca Clivio^{**} and Anthony Travis^{***}

^{*}University of Aberdeen Computing Science Department; ^{**}Mario Negri Institute, Milan and
^{***}University of Aberdeen Rowett Institute of Nutrition and Health

SUMMARY

The suitability of the EU FP6-funded Kerrighed and XtreamOS projects to upgrade the existing NuGO distributed IT infrastructure for data sharing and bioinformatics was evaluated using an NBX (NuGO Black Box) Beowulf SSI (Single System Image) cluster constructed using COTS (Commodity Off The Shelf) hardware and FLOSS (Free/Libre Open Source Software). The stability of the Kerrighed SSI cluster was severely compromised when later 2.4 versions of the Kerrighed software were tested on the NBX cluster under 32-bit NuGO-Linux. However, the older 2.3 version of Kerrighed was stable enough to be considered for use as a production platform. Stability problems prevented a detailed evaluation of the full 'grid' functionality of the XtreamOS-G layer, but experiments concerning cooperation between NBX systems connected via the WAN (Wide Area Network) demonstrated that Kerrighed might provide a load-balancing option for 'embarrassingly' parallel applications where little or no communication occurs between processes. The object-oriented XtreamFS filesystem was also evaluated and performed very well. However, concerns about the requirement for multiple network ports to be open in the perimeter firewall at NuGO partners limit the practicality of this to be considered as a replacement for the "sshfs" user-space filesystem that is currently in use for data sharing between NBX's over the WAN. It seems unlikely that Kerrighed will be supported beyond version 2.3 on the 32-bit Linux platform, which restricts its usefulness under the 32-bit NuGO-Linux currently installed on the NBX network. However, the NBX network will be upgraded to 64-bit NuGO-Linux and more recent versions of Kerrighed will then offer much greater stability on the 64-bit platform. Nevertheless, Kerrighed 2.3 provides a 32-bit SSI platform compatible with the existing NBX infrastructure that may be useful to aggregate the CPU and memory resources of multiple NBX's connected locally via a high-speed LAN. This was demonstrated as a practical way of improving the performance of the GenePattern bioinformatics software running on the prototype NBX Beowulf SSI cluster constructed for this evaluation.

INTRODUCTION

Overview of the data sharing problem and data sharing technology

Data sharing is vital to any organisation, but a 'virtual' organisation such as NuGO (Nutrigenomics Organisation) an EU-FP6 funded NoE (Network of Excellence) with geographically distributed membership, requires an appropriate type of distributed data sharing infrastructure in order to function successfully. This report describes an investigation into the possibility of creating a distributed resource and data sharing 'grid' for NuGO using the Kerrighed SSI (Single System Image) Linux kernel, and the XtreamOS Grid-enabled Linux operating system, installed on a small group of NBX's (NuGO Black Boxes) configured as an NBX Beowulf parallel computing cluster.

The main type of data sharing infrastructure used by larger virtual organisations at present is an information or data 'grid', with many independent grid nodes connected together over a WAN (Wide-Area Network) and all working collectively to perform distributed data processing tasks that are passed to the nodes via the data 'grid' using 'middleware such as the open source "Globus" grid software for distributed resource sharing (<http://www.globus.org>). Nodes in an organisation's data and resource sharing grid can range from an individual user's workstation to large mainframe computers, or local area networks of computers, or specialised parallel computing clusters. This allows organisations that already have a data grid to scale up and add new technology as needed.

Existing software for creating both resource sharing grids and specialised computing clusters is, in general, too complex for all but dedicated research use, and has been slow to evolve. In many cases, development of promising new systems has halted entirely¹. Grid computing software has been introduced into current systems as middleware² that functions as a 'mediator' layer between the grid, as a single conceptual entity, and the services provided by individual grid nodes to complete a given task. This type of software is often FLOSS (Free/Libre Open Source Software)^{3,4}, which can be an advantage, but its isolated nature - external to both the grid-enabled programs and the operating system installed on the grid nodes - makes it difficult to integrate the middleware into a production system. A unified layer of middleware could also, in theory, become a single point of failure for the entire grid, which is a large enough problem to require considerable expenditure to an organisation should a significant downtime occur. Accordingly, most data grids built to date are national efforts.⁵

COTS (Commodity Off-The-Shelf) Hardware

Beowulf clustering, is the practice of using low-cost, COTS (Commodity Off-The-Shelf) hardware components to build computers and interconnect them on a fast local area network to create a high-performance parallel computer using FLOSS. This is usually achieved by modifying the underlying operating system⁶, most often Linux, rather than by using an intermediate layer of 'middleware'.

In addition to the Kerrighed SSI Linux kernel patches discussed later in this report, there are two other well established FLOSS Beowulf cluster operating systems currently in use: openMosix⁷, development of which closed on March 1st 2008, and the openSSI project⁸, which appears to have ceased its updates and active development in 2006. Both openMosix and openSSI are sufficiently stable to run on Beowulf clusters intended for research work. For example, openMosix was used until May 2010 on the RINH/BioSS openMosix Beowulf⁹. However, these two FLOSS projects suffer from serious bugs and design flaws that are unlikely to ever be resolved¹⁰ - incomplete or

1 See Open Middleware Institution of Europe 2006, "D:JRA2.0 Report on Grid Activities relevant to the identification of new services", <http://www.omii-europe.org/docs/DJRA20.pdf>

2 For a definition of the term, see the FOLDOC: <http://foldoc.org/middleware>

3 The EU Gridipedia project hosts a list of leading grid middleware projects, all of which are FLOSS of some variant: <http://www.gridipedia.eu/middleware.html>

4 FLOSS is *Free and Libre Open Source Software*; that is, it is software which costs nothing and is freely available as source code to the public. It is usually licensed under the GNU GPL (<http://www.gnu.org/licenses/gpl.html>)

5 A listing of the extant national research grids can be found at the EU Gridipedia project: <http://www.gridipedia.eu/grid-infrastructures.html>

6 Note: this report only deals with Single-System Image (SSI) clustering. Other forms of computer clustering are not readily available on COTS hardware and are not user-friendly enough for analysis here.

7 See the OpenMosix archive, which also includes an explanation from its creator Moshe Bar about what happened to the project's development: <http://openmosix.sourceforge.net/>

8 The OpenSSI project's information page: <http://openssi.org/cgi-bin/view?page=openssi.html>

9 For a definition of Beowulf clustering, see <http://www.beowulf.org/overview/index.html>; the RINH/BioSS Beowulf is available online at <http://bioinformatics.rri.sari.ac.uk/>

10 The INRIA comparison of openMosix, Kerrighed and OpenSSI is a useful overview of the problems with these

missing features, security vulnerabilities, and kernel instability - which will not be patched and corrected unless active project development resumes, and this now seems unlikely to happen.

New processor architectures and hardware types cannot be supported by such abandoned operating systems unless users themselves rewrite the code, but this is not a viable task for NuGO to attempt. Although openMosix is still a very useful and workable system, it will not remain so for long. In particular, openMosix is confined to the 32-bit Linux 2.4 kernel and does not support modern 64-bit multi-core processors. An updated commercial version of MOSIX2 that does support the 64-bit architecture is now available, but this is proprietary closed source software and is only available at significant cost. Similarly, Penguin Computing's Scyld Clusterware, based on the pioneering work of Donald Becker who invented the Beowulf architecture, is also available as an alternative, but this too is commercial proprietary software¹¹. It is now clear that other potential FLOSS solutions to the data sharing problem need to be considered. The XtremOS grid operating system and associated Kerrighed SSI Linux kernel are contemporary EU-funded FLOSS projects that may provide an appropriate solution for NuGO.

REQUIREMENTS

A distributed IT infrastructure for NuGO

NuGO has distinct requirements for a distributed IT infrastructure, arising from its overarching goal of working as a virtual organisation in a transparent and open technical environment: Any potential distributed system for a virtual organisation of any sort needs to be abstracted from the hardware, to such an extent that users of the system are not required to understand all of the underlying technical details in order to use it¹². The majority of NuGO members are biologists, or bioinformaticians, not computer scientists. The proposed system, therefore, needs to support users with a reasonable range of computing skills from novice to expert. Any candidate system used to provide the distributed IT infrastructure must also be able to run the scientific software currently used by NuGO members for bioinformatics and, because the tasks it will carry out may often be time-consuming or constrained by limited access rights, the system needs to be stable enough that it can be relied upon to store and work with valuable and irreplaceable or critical data. A further requirement is that the system must continue to be supported in the long-term by its team of developers and maintainers, so that NuGO itself is not burdened with the task of performing routine maintenance work, or upgrading the basic functionality of the software chosen to create a virtual infrastructure for the organisation.

Access rights and security considerations

Access to the NBX grid must be restricted only to appropriate NuGO members and trusted NuGO partners, both to protect the organisation's intellectual property and to prevent inappropriate use of its resources¹³. However, the virtual IT infrastructure used must also allow its users autonomy and control over access to their own data. All of the experimental data stored on the system should be modifiable and, when necessary, removable by its owners/creators. The system's overall strategy of protecting NuGO's intellectual property as a whole should not override the rights and freedoms of users to access their *own* experimental data, nor should it restrict or inhibit the trust relationships developed within NuGO by undermining a user's confidence in their access rights. Members should

systems: 2004, "OpenMosix, OpenSSI and Kerrighed: A Comparative Study ", Lottiaux, Boissinot, Gallard, Vallée and Morin <ftp://ftp.inria.fr/INRIA/publication/publi-pdf/RR/RR-5399.pdf>

11 http://www.penguincomputing.com/software/scyld_clusterware

12 Abstraction in this case refers to SSI clustering. This idea was taken from the work of the XtremOS project: <http://www.xtreemos.org/index.html>.

13 NuGOweek documents {citation needed}

be confident that they actually do have control of their own intellectual property, and NuGO as an organisation should also have confidence in the system's protection of its intellectual property.

Economic constraints

Although the NBX project is an important part of NuGO's existing data sharing strategy, the project was conceived as a low-cost alternative to investing in an expensive, centralised NuGO data centre. The grid system proposed as a solution to NuGO's data-sharing requirements must also be cost-effective, and inexpensive in comparison to an 'industrial' scale mainframe computer, because there will be more than one grid node in production and most grid nodes will be a 'lab' scale resource consisting of one or more NBX's connected to the Internet. To achieve this objective requires both the NBX and grid hardware to be constructed using inexpensive COTS components, and for the software used to implement the infrastructure to be FLOSS. Finally, the proposed NuGO-Grid solution must make use of NuGO's existing NBX infrastructure rather than simply replace it, by adding to and building upon the previous work of the existing NIN (NuGO Information Network) and NBX (NuGO Black Box) projects (Harttig *et al.*, 2009; Ommen *et al* 2010).

PROPOSED SOLUTION

Kerrighed and XtreamOS

NuGO's infrastructure requirements were initially met by the existing NBX's and the NIN creating a NuGO file-sharing data grid. However, two recent computing science research projects that might provide a more suitable virtual IT infrastructure for NuGO were examined during this investigation: XtreamOS, a cluster- and grid-enabled GNU/Linux operating system¹⁴ developed at INRIA¹⁵, and the Kerrighed Linux SSI kernel that powers it¹⁶. The Kerrighed project was founded by the same team of developers who later developed it commercially via the open source 'Kerlabs' initiative¹⁷.

Kerrighed is currently the most active of the Linux SSI (Single System Image) clustering projects. SSI architecture gives users (and administrators) a conceptual view of a Beowulf cluster that is used and administered as a single machine, but with all the resources of the entire cluster available on it. A particular advantage of SSI is that users do not need to learn many new commands, or how to write parallel programs, in order to be able to make use of an SSI cluster. Kerrighed SSI operates as a modification to the standard OS kernel of an existing Linux distribution and is similar, in many respects, to SMP (Symmetric Multi-Processing) on a single computer with multiple processors or multiple processor cores, but SSI shares work between all of the computers in a Beowulf cluster using an Ethernet interconnection instead of the local inter-processor bus on a single motherboard. This inevitably creates an order of magnitude greater latency between processors, but has major economic benefits for the cost-effectiveness of the system as a whole. Kerrighed operates in kernel 'space' via two loadable Linux kernel modules, but there are also several 'userland' tools available for administration of a Kerrighed cluster. Kerrighed is FLOSS, and is currently at release version 3.

14 See the XtreamOS project site at <http://www.xtreemos.org/>

15 INRIA is the *Institut National de Recherche en Informatique et en Automatique* of France, the organisation which controls the development of XtreamOS and was the original source of development for Kerrighed. It can be found at <http://www.inria.fr/>.

16 It is not always publicly acknowledged that the SSI cluster-enabled kernel of XtreamOS, called *LinuxSSI*, is Kerrighed; however, the two projects are identical, in code and version, as verified by Dr. Clivio. Any reference to LinuxSSI should be assumed to mean Kerrighed.

17 The Kerlabs public site can be found at <http://www.kerlabs.com/-Home-.html>.

XtreemOS is also FLOSS; and is currently available as a full Linux distribution derived from the Mandriva 2008.0 Linux distribution. XtreemOS is intended to be a fully-fledged “grid OS”¹⁸ which will run on Beowulf clusters as well as on mobile devices and single PC desktops. The “cluster” version of XtreemOS is actually powered by 'LinuxSSI', which is simply a re-branded version of Kerrighed 2.4 but with otherwise identical functionality. XtreemOS can be downloaded from the project site either as a pre-mastered '.iso' image for full installation on a dedicated PC, or as a complete set of “rpm” (Red-Hat Package Manager) format packages from the publically accessible XtreemOS "svn" (subversion) development repository, and is currently at version 1, release candidate 1¹⁹.

The existing 32-bit NuGO-Linux operating system used on the current NBX platform is derived from the UK NEBC (Natural Environment Research Council Environmental Bioinformatics Centre) Bio-Linux 5²⁰ operating system, which is a remastered version of the commercial FLOSS 32-bit Ubuntu 8.04 LTS (Long-Term Support) Desktop CD. Bio-Linux is well documented in its own right by NEBC who NuGO have assisted with development of Bio-Linux. Support of Bio-Linux 5 is linked to upstream support of Ubuntu 8.04 LTS, which is guaranteed on the server until 2011. Although a new 64-bit version of Bio-Linux 6 was under development during our Kerrighed and XtreemOS evaluation project, a production version of Bio-Linux 6 was not then available.

Kerrighed and XtreemOS could, in principle, provide an abstract cluster interface to an NBX grid that is largely independent of the underlying computer hardware. If Kerrighed and XtreemOS can be successfully integrated into NuGO-Linux then the scientific packages required by NuGO should run exactly the same way that they do on a single NBX, and the NBX Beowulf cluster nodes would be entirely compatible with the existing NBX data sharing platform. An important feature of SSI clusters is that many applications will run on them unmodified, and applications do not need to be programmed specifically to work in an SSI environment. This is useful for 'embarrassingly' parallel applications where little or no communication takes place between processes running on different nodes. However, SSI is also compatible with MPI (Message Passing Interface) parallel applications that can exploit the SMP-like environment of an SSI cluster and are simply unaware that MPI processes spawned by a single job are actually running on different nodes in the Beowulf cluster.

It is clear that, together, Kerrighed and XtreemOS can fulfil some of NuGO's basic requirements immediately. Kerrighed itself is being promoted and developed as “Linux clusters made easy”²¹, and XtreemOS is simply another Linux distribution but with grid functionality provided at Linux kernel level. In our prototype Kerrighed/XtreemOS implementation on the NBX, we have simply added extra packages to the existing NuGO-Linux operating system installed on the NBX.

The evaluation reported here mainly concerns the reliability and stability of our proposed virtual infrastructure - not only in relation to its speed of operation and functionality when performing bioinformatics tasks, but also if the proposed infrastructure would comply with NuGO's existing security policy. If the proposed system is appropriate for the NuGO virtual organisation, it may also be a useful tool for controlling access rights to scientific data in other similar virtual organisations.

18 The XtreemOS project overview, <http://www.xtreemos.eu/overview/plonearticlemultipage.2006-06-08.9297943452/project-summary>

19 The latest version of XtreemOS can be obtained from the project's repository, for which a primary mirror is at <http://www.mirrorservice.org/sites/carroll.cac.psu.edu/MandrivaLinux/devel/xtreemos/>

20 This is the official NEBC Linux distribution. It can be obtained directly from them at <http://nebc.nox.ac.uk/>.

21 The project slogan. See the main header of the project wiki: <http://www.kerrighed.org/>

IMPLEMENTATION

Research questions

"Testing stability" is a rather vague concept: However, this objective was approached by building a prototype Kerrighed SSI NBX Beowulf cluster using the software systems proposed for the virtual infrastructure. The prototype cluster was then used to address two fundamental questions: first, if Kerrighed and/or XtremOS could actually be implemented on a Beowulf cluster consisting of a number of standard NBX's and also how reliably such an NBX cluster would run; second, if the NBX cluster would be able to communicate either locally or remotely with another NBX system, in a similar way to communication over the proposed grid infrastructure it might eventually run on.

Overview of the prototype cluster

An important objective of the proposed distributed IT infrastructure is that the user interface should resemble, as closely as possible, the existing NBX web GUI. The screen-shot in Figure 1., shows the NBX web-GUI of "kitcat". In particular, the familiar NuGO-Linux web-GUI layout based on the Drupal CMS (Content Management System) is used and the GenePattern bioinformatics suite, for example, can be used without the user needing to know how to migrate parallel processes using SSI on the an NBX Beowulf cluster. This is an important design objective of the proposed NuGO distributed IT infrastructure and an evaluation of running the GenePattern bioinformatics software on an SSI Beowulf cluster will be described in more detail later.

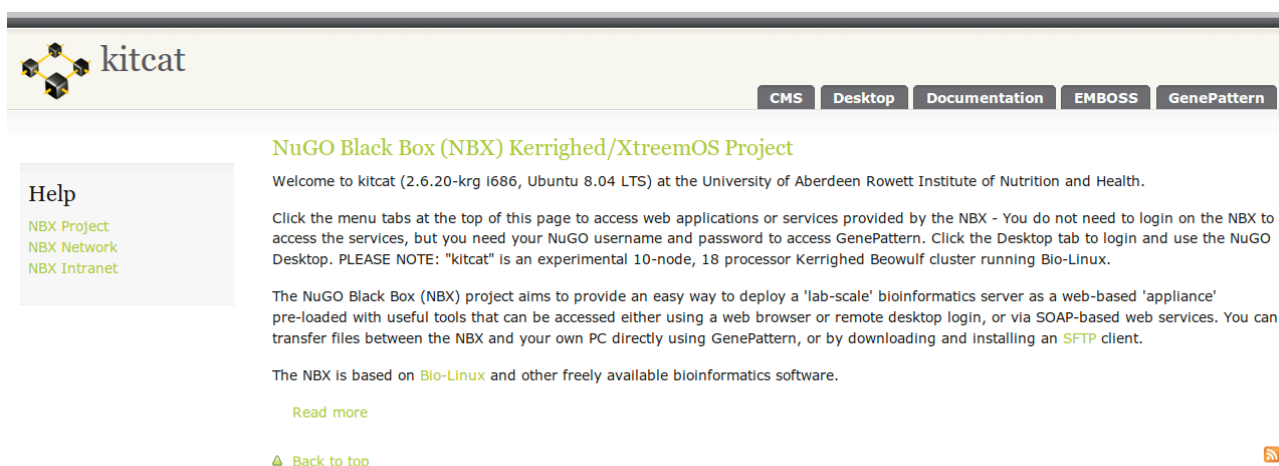


Figure 1. Standard NBX web GUI (Graphical User Interface) on the "kitcat" NBX Beowulf cluster

The prototype cluster is a ten-node NBX Beowulf computer, created from two standard diskfull NBX servers and eight diskless compute nodes. The head node of the prototype NBX cluster is a standard NBX running a web server available on the Internet at "http://kitcat.rri.sari.ac.uk", but "kitcat" can also participate in the Beowulf cluster as a compute node under certain circumstances (Kerrighed 'server as node' configuration). All the machines in the NBX Beowulf cluster run Bio-Linux 5. The diskless cluster nodes run a standard (NFSROOT) Kerrighed kernel, but the "kitcat" NBX cluster server itself runs a stand-alone Kerrighed Linux kernel. The diskless cluster nodes boot over the interconnection LAN (Local Area Network) network via PXE (Preboot eXecution Environment) and the cluster nodes and servers communicate over standard Gigabit Ethernet.

Inventory of prototype hardware and software

The NBX cluster nodes used in this evaluation are server-grade PC's similar in to a standard NBX, but the two NBX cluster servers used are standard NBX's exactly as have been provided to research partners by NuGO²² during the NoE. The standard NBX has a Tyan S3970 server-grade motherboard with two Opteron 2212 dual-core AMD processors, eight GiB RAM and two TB of hard disk space configured as RAID1 and RAID5 and are connected to the cluster LAN by two independent Gigabit (1,000 Mbit) Ethernet network interfaces. The "kitcat" server also has a third Ethernet card for connection to the Internet. The NBX hardware inventory for "kitcat" is shown below in Table 1.

Table 1. The standard NBX hardware inventory for "kitcat"

H/W path	Device	Class	Description
=====			
		system	empty
/2		bus	S3970-E
/2/0		memory	64KiB BIOS
/2/4		processor	Dual-Core AMD Opteron(tm) Processor 2212
/2/4/5		memory	128KiB L1 cache
/2/4/6		memory	1MiB L2 cache
/2/4/7		memory	L3 cache
/2/8		processor	Dual-Core AMD Opteron(tm) Processor 2212
/2/8/9		memory	128KiB L1 cache
/2/8/a		memory	1MiB L2 cache
/2/8/b		memory	L3 cache
/2/28		memory	8GiB System Memory
/2/28/0		memory	DIMM [empty]
/2/28/1		memory	DIMM [empty]
/2/28/2		memory	2GiB DIMM Synchronous 333 MHz (3.0 ns)
/2/28/3		memory	2GiB DIMM Synchronous 333 MHz (3.0 ns)
/2/28/4		memory	1GiB DIMM Synchronous 333 MHz (3.0 ns)
/2/28/5		memory	1GiB DIMM Synchronous 333 MHz (3.0 ns)
/2/28/6		memory	1GiB DIMM Synchronous 333 MHz (3.0 ns)
/2/28/7		memory	1GiB DIMM Synchronous 333 MHz (3.0 ns)
/2/2		processor	
/2/2/0		memory	128KiB L1 cache
/2/2/1		memory	1MiB L2 cache
/2/3		processor	
/2/3/0		memory	128KiB L1 cache
/2/3/1		memory	1MiB L2 cache
/2/1		bridge	BCM5785 [HT1000] PCI/PCI-X Bridge
/2/1/d		bridge	BCM5785 [HT1000] PCI/PCI-X Bridge
/2/1/d/3	scsi6	storage	7xxx/8xxx-series PATA/SATA-RAID
/2/1/d/3/0.0.0	/dev/sde	disk	500GB Logical Disk 0
/2/1/d/3/0.0.0/1	/dev/sde1	volume	38GiB EXT3 volume
/2/1/d/3/0.0.0/2	/dev/sde2	volume	7812MiB Linux swap volume
/2/1/d/3/0.0.0/3	/dev/sde3	volume	419GiB EXT3 volume
/2/1/e	scsi2	storage	BCM5785 [HT1000] SATA (Native SATA Mode)
/2/1/e/0	/dev/sda	disk	250GB ST3250820AS
/2/1/e/0/1	/dev/sda1	volume	698GiB EXT3 volume
/2/1/e/1	/dev/sdb	disk	250GB ST3250820AS
/2/1/e/1/1	/dev/sdb1	volume	232GiB Linux raid autodetect partition
/2/1/e/2	/dev/sdc	disk	250GB ST3250820AS
/2/1/e/2/1	/dev/sdc1	volume	232GiB Linux raid autodetect partition
/2/1/e/3	/dev/sdd	disk	250GB ST3250820AS
/2/1/e/3/1	/dev/sdd1	volume	954GiB EXT3 volume
/2/100		bridge	BCM5785 [HT1000] Legacy South Bridge
/2/100/2.1	scsi0	storage	BCM5785 [HT1000] IDE
/2/100/2.1/0.0.0	/dev/cdrom	disk	CD/DVDW SH-S182M
/2/100/2.2		bridge	BCM5785 [HT1000] LPC
/2/100/3		bus	BCM5785 [HT1000] USB
/2/100/3.1		bus	BCM5785 [HT1000] USB
/2/100/3.2		bus	BCM5785 [HT1000] USB
/2/100/4	eth0	network	82541GI Gigabit Ethernet Controller
/2/100/5	eth1	network	82541GI Gigabit Ethernet Controller
/2/100/6		display	Volari Z7
/2/100/7	eth2	network	3c940 10/100/1000Base-T [Marvell]
/2/101		bridge	K8 [Athlon64/Opteron] HyperTransport Technology
Configuration			

22 For a simple description of the NBX and its standard hardware, see the NBX project presentation on NuGONet (<http://www.nugo.org/folders/34113>).

/2/102		bridge	K8 [Athlon64/Opteron]	Address Map
/2/103		bridge	K8 [Athlon64/Opteron]	DRAM Controller
/2/104		bridge	K8 [Athlon64/Opteron]	Miscellaneous Control
/2/105		bridge	K8 [Athlon64/Opteron]	HyperTransport Technology
Configuration				
/2/106		bridge	K8 [Athlon64/Opteron]	Address Map
/2/107		bridge	K8 [Athlon64/Opteron]	DRAM Controller
/2/108		bridge	K8 [Athlon64/Opteron]	Miscellaneous Control
/0	c0/p0	disk	137GB	ST3500320AS
/1	c0/p1	disk	137GB	ST3500320AS

The software installed on the cluster nodes and on the “kitcat” server has been kept as close to the NBX standard as possible: They all run NuGO-Linux (NEBC Bio-Linux 5 with NuGO's standard modifications). This includes all the bioinformatics packages that users of the NBX have access to on a standard NBX installation. In addition to this, “kitcat” required several extra packages to be installed in order to function as a PXE and NFS server for the cluster nodes²³ as shown in Table 2.

Table 2. Extra cluster server packages installed on “kitcat”

- A standard Linux server kernel
- A Kerrighed-patched kernel and its userland tools, both to be exported to the nodes
- DHCP server software to allow *kitcat* to assign IP addresses to the nodes
- TFTP server software to allow the nodes to PXE boot from “kitcat”
- NFSv3 server software to enable a single filesystem to be shared over the cluster

Method of implementation

Step-by-step documentation of the process followed in order to set up the Kerrighed cluster for the evaluation has been added to the Ubuntu Community Wiki²⁴ (NuGO-Linux/Bio-Linux is based on Ubuntu Linux). An important part of the evaluation was to involve both the Ubuntu and Kerrighed communities in our work. This aspect of FLOSS can be an advantage over the use of proprietary software and significant interest was generated in our 'easy' Ubuntu clustering discussion thread on the Ubuntu “Science and Education” forum (<http://ubuntuforums.org/showthread.php?t=1030849>).

Table 3. Summary of prototype NBX cluster installation

1. The two NBX cluster servers were built to the standard NBX specification.
2. Bio-Linux 5 was installed to each server according to the standard Bio-Linux installation procedure²⁵.
3. The modifications specified in the NuGO NBX development wiki²⁶ were made in order to properly configure the NBX OS, making the servers identical to production NBX's.
4. A server version of the Kerrighed SSI kernel and the associated software to enable SSI clustering were installed on “kitcat”.
5. A server version of the Kerrighed SSI kernel was also installed on the second NBX cluster server and this server was configured as an NFS file server for the NBX cluster.
6. A separate boot environment for the nodes was created on the “kitcat” server's disks using the Debian/Ubuntu *debootstrap* tool²⁷, and a minimal Ubuntu 8.04 operating system was installed into this.
7. Following the Kerrighed project's own installation instructions, the Kerrighed kernel and userland tools were installed into the node boot environment.

²³ See the documentation created by this investigation for the full list.

²⁴ <https://wiki.ubuntu.com/EasyUbuntuClustering/UbuntuKerrighedClusterGuide>

²⁵ The standard procedure is described in the Bio-Linux 5 install guide: http://nebc.nox.ac.uk/downloads/bio-linux/Bio-Linux5_InstallationGuide.pdf

²⁶ {citation needed}

²⁷ <http://packages.debian.org/lenny/debootstrap>

Three versions of Kerrighed were tested: the original development version 2.3, a stable release 2.4 and a minor update 2.4.1. Additionally, testing Kerrighed raised questions about whether an NBX server could be run as a 'compute node' participating in the SSI cluster, while still functioning as the cluster 'head node', rather than simply as the cluster PXE and NFS server providing network-mounted root filesystems to the 'diskless' compute nodes. Testing this Kerrighed 'server as node' configuration was only attempted after Kerrighed 2.3 had already been successfully tested with the prototype NBX cluster configured according to the recommendations in the Kerrighed project documentation that a Kerrighed SSI cluster should only include diskless nodes. In order to test the Kerrighed 'server-as node' functionality, the following procedure outlined in Table 4 was used.

Table 4. NBX 'server as node' configuration of "kitcat"

1. The normal PXE boot environment for the nodes was disabled. Instead, they were configured to use the "kitcat" server's own filesystem as their NFS root filesystem.
2. The root filesystem of the "kitcat" server was reconfigured to be mounted read-only by the nodes in order to protect system files that only the "kitcat" server should modify from the filesystem used by the cluster nodes should, and a separate writeable filesystem partition was added where the cluster nodes could read and write the volatile parts of their individual NFS-mounted filesystems.
3. Further, to prevent the nodes from reading system files which they do not need access to, UNFS3 with ClusterNFS²⁸ extensions was installed on "kitcat" and the ClusterNFS 'context-dependent' symbolic link file tagging system was used to redirect such files to a non-existent file.

Testing configurations of the prototype NBX Kerrighed cluster, either with the "kitcat" server not participating in the Kerrighed SSI cluster or with "kitcat" participating as a cluster node, involved following the procedure outlined in Table 4. to establish if the cluster would run reliably. Then performing 'stress' tests on the running Kerrighed SSI cluster. This procedure was repeated for all three versions of Kerrighed that were available at the time of our evaluation. Upon a cluster crash, notes were made about the circumstances of the crash, then attempts were made to work around the problems encountered so that subsequent NBX cluster testing could continue. This was an iterative process of progressive establishment testing that we used to create a reliable cluster for evaluation.

When the prototype NBX cluster was reliable enough for extended use, it was tested using a small CPU 'hog' stress-test program²⁹. The 'hog' program, written in C, is a 'wrapper' that calls multiple instances of the well-known computationally intensive 'Dhrystone' integer CPU benchmark³⁰. On a fully functional Kerrighed cluster, CPU intensive processes such as the CPU 'hog' running on one cluster node should be automatically migrated onto other, less busy nodes. Although fundamentally different to openMosix at the Linux kernel level, by default Kerrighed uses the same load-balancing algorithm that is used in openMosix to distribute the computational burden across the SSI cluster.

Multiple instances of the CPU 'hog' program were run after permitting process migration to occur under an interactive Linux command shell running on the Kerrighed cluster, and automatic process migration by Kerrighed was monitored using the Linux "top"³¹ program to confirm that automatic process migration was actually taking place. This process was to repeated for all three versions of Kerrighed tested, and for both "kitcat" server configurations, resulting in a total of six evaluations of Kerrighed SSI stability overall on the NBX platform.

28 ClusterNFS project site, <http://sourceforge.net/projects/clusternfs/>, contains full instructions for setup.

29 The *hog* source code is included as Appendix A.

30 Weicker, Reinhold. "Dhrystone: A Synthetic Systems Programming Benchmark" *Communications of the ACM (CACM)*, Volume 27, Number 10, October 1984, p.1013-1030

31 See the man page for the *top* command, at <http://linux.die.net/man/1/top>

XtreemOS was evaluated separately from Kerrighed, because it contains the re-branded Kerrighed Linux-SSI kernel. Our evaluation of XtreemOS was done by downloading, burning and then installing the entire XtreemOS operating system onto the “kitcat” server, and also by attempting to install XtreemOS “rpm” packages under NuGO-Linux on “kitcat” from the XtreemOS project's developer repository. XtreemOS sources were obtained from both the XtreemOS project "svn" repository, and from the downloaded XtreemOS installation DVD.

EVALUATION

Kerrighed SSI kernel

The results from our evaluation of the Kerrighed SSI kernel were mixed, and differed depending on whether the conventional configuration of the “kitcat” server not participating in the Kerrighed SSI cluster as recommended by the Kerrighed development team or the server-as-node system was used.

When set up with “kitcat” controlling but not participating in the cluster, the 32-bit Kerrighed 2.3 kernel compiled and installed on a standard NBX with only a few errors in configuration that were corrected as installation proceeded. The Kerrighed SSI cluster could be started, and appeared to run reliably until a heavy CPU load (one or more CPU ‘hog’ instances) was placed on the cluster. At this point the Kerrighed cluster would crash predictably, rendering the “kitcat” server inaccessible to the NBX cluster nodes and a manual reboot was necessary. However, migration of the “hog” processes across the Kerrighed cluster was successful until a crash occurred. Some of the Kernel instability may have been caused by the UNFS3 user-mode NFS filesystem used to support the diskless compute nodes, rather than the Kerrighed kernel itself which ran reliably on “kitcat”.

Tests with Kerrighed 2.4 were less successful. The kernel could be installed with little effort, but running a Kerrighed cluster proved to be all but impossible. A Kerrighed 2.4 cluster could not be started, regardless of the cluster configuration. The problem was, eventually, traced to a missing, and undocumented, “configs” cluster configuration filesystem that is used by the Kerrighed 2.4 kernel. The requirement for this filesystem was not documented anywhere in the Kerrighed and XtreemOS project manuals, and it took us some considerable time to discover this fundamental requirement. Even when the “configs” filesystem was eventually created, and the Kerrighed 2.4 NBX SSI cluster was properly configured, the Kerrighed cluster failed to run for more than a few seconds before the system crashed irretrievably, making it impossible to perform any tests on its process migration functionality.

The cause of this particular crash was actually fixed by the developers in Kerrighed version 2.4.1 during our evaluation. However, our undocumented workaround for the missing “configs” was still needed. Under this later version of Kerrighed, the NBX cluster was considered to be nearly stable enough for production use: However, the average cluster uptime was still only about two days, and it was only capable of successfully migrating two instances of the CPU “hog” stress-test program - three CPU “hogs” appeared to be too many and the NBX Kerrighed cluster crashed.

When testing the server-as-node functionality, Kerrighed 2.3 behaved in a similar manner to its performance under the standard configuration recommended by the Kerrighed developers: The kernel could be installed with only minor modifications, but the cluster crashed again under heavy ICPU oad, although process migration was functional. Critically, however, the standard Kerrighed SSI kernel configuration does not support the “iptables” firewall³² that is used on the NBX. As a consequence, the “fail2ban” intrusion detection script³³, which depends on “iptables”, failed to

32 See the *iptables* project page: <http://www.netfilter.org/projects/iptables/>

33 See the *fail2ban* wiki: http://www.fail2ban.org/wiki/index.php/Main_Page

work. This is not acceptable on an Internet-facing NBX server such as “kitcat”, because “fail2ban” is a critical defence used against brute-force and denial of service attacks from hostile ‘botnets’. No documented Kerrighed configuration was found to address the “iptables” problem during the initial stages of our evaluation and “kitcat” was not connected to Internet until much later when we created a new Kerrighed kernel configuration that does support “iptables” ourselves. The NBX cluster was, however, stable running for periods of ten or more days under this version of Kerrighed.

Kerrighed version 2.4.1 was even less reliable when the “kitcat” server was participating as an SSI cluster node. The 2.4.1 cluster crashed slightly more frequently during our stress testing, although compared with the other versions of Kerrighed its uptime was good. It could migrate two instances of the CPU “hog” process reliably. However, as with the Kerrighed 2.3 version, it was incompatible with “iptables”, and not suitable for use on unprotected Internet-facing NBX servers. After due consideration of stability issues that we encountered evaluating newer versions of Kerrighed, we decided to adopt the most stable Kerrighed 2.3 version as the basis of our evaluation. According to the Kerrighed developers, many of the problems that we encountered are a consequence of their development work targeting the more modern 64-bit platform and newer versions of Kerrighed beyond version 2.3 are no longer tested on a 32-bit platform such as NuGO-Linux/Bio-Linux 5.

Running GenePattern on the “kitcat” NBX Beowulf cluster

Having established a reasonably stable NBX Beowulf cluster running on “kitcat”, we proceeded to evaluate the potential of load-balancing processes spawned by the GenePattern bioinformatics software used by NuGO for analysis of transcriptomics data. An SSI view of “kitcat” is shown in Figure 2., which shows the output of the “htop” process monitoring program with 18 processors apparently present on “kitcat”. Most of these processors are, in reality, present on other nodes in the SSI cluster, but they all appear to be on “kitcat” because of the distributed process-space created by the Kerrighed SSI kernel.

GenePattern is a Java servlet running under Apache Tomcat , and typically spawns computationally intensive R programs encapsulated as GenePattern 'modules'. The GenePattern developers already do support the use of SGE (Sun Grid Engine) as a DRM (Distributed Resource manager) to run these GenePattern modules. However, SGE does not provide an SSI environment and is not able to provide the transparent aggregation of both CPU and memory resources that can be achieved using the Kerrighed SSI Linux kernel. These two alternative solutions to improving the performance of GenePattern on the NBX operate at different levels.

We were able to verify that R programs running as Genepattern modules were automatically migrated by the Kerrighed SSI kernel running on the prototype NBX SSI Beowulf cluster using the Unix/Linux “top” process monitor. This aspect of our Kerrighed/XtreemOS evaluation work was demonstrated at a NuGO WTP meeting in Newcastle, part of which concerned how to plan the future development and sustainability of the existing NBX data sharing infrastructure.

Migration of Java processes is possible under Kerrighed but, at present, Kerrighed cannot migrate light-weight Java threads. This can be seen in Figure 2., where all the Java processes are running on processor 12 in the “htop” monitor, creating 100% processor occupancy. This is in contrast to R programs in modules spawned by the GenePattern Java servlet. These computationally intensive R programs migrate automatically to unoccupied processors according to process metrics used by the default openMosix load-balancing algorithm adopted by Kerrighed SSI.

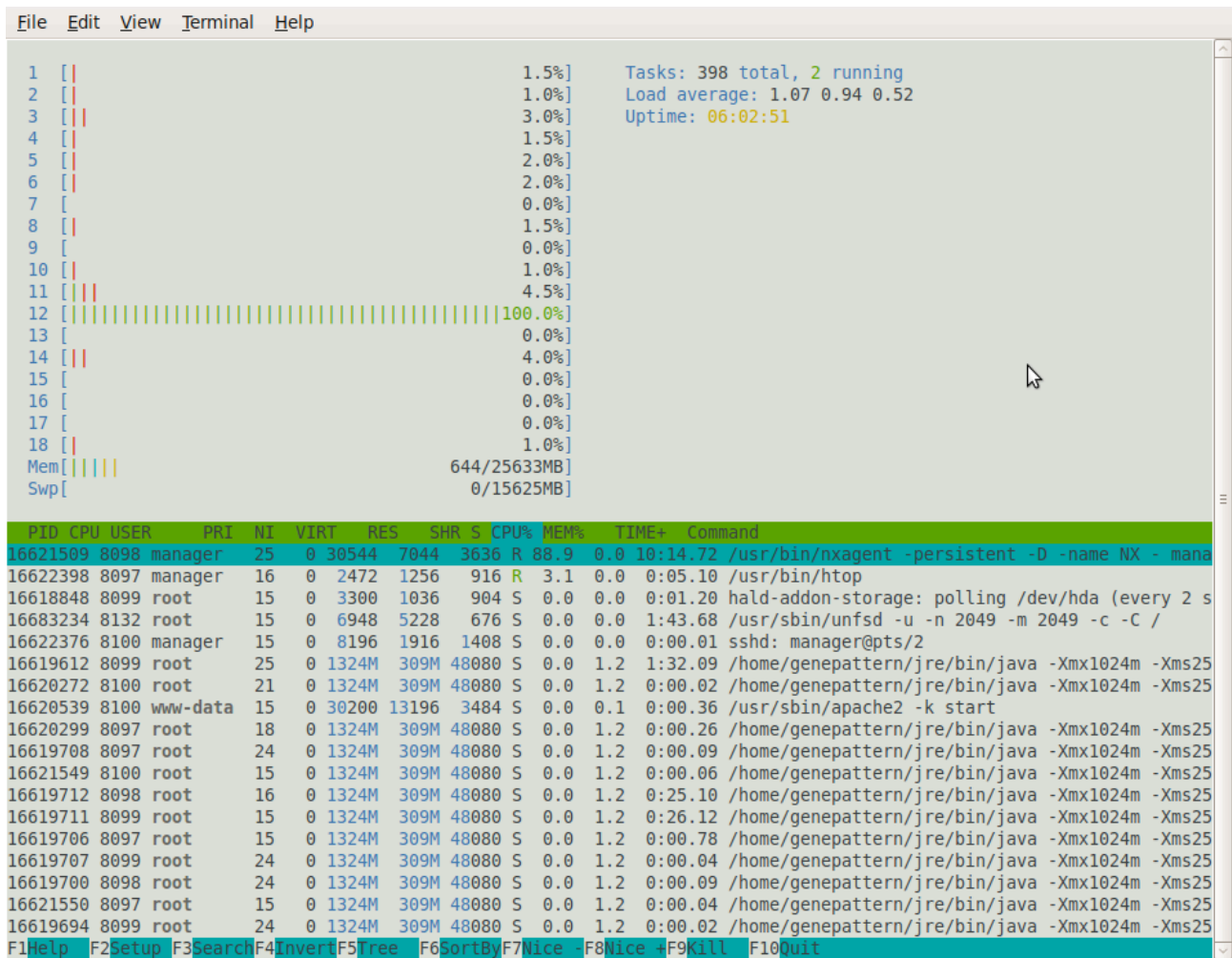


Figure 2. GenePattern running on the “kitcat” NBX kerrighed Beowulf cluster.

XtreemOS

The XtreemOS operating system provides VOM (Virtual Organization Management) There are a number of related components present in XtreemOS. An overview of these components is shown in Figure 4. We did not have the opportunity to fully investigate the "Grid Layer" of XtreemOS-G, because of time spent dealing with fundamental problems that we encountered in the stability of the Kerrighed/Linux-SSI kernel used in the XtreemOS-F layer. Our strategy was to conduct a 'bottom-up' evaluation of XtreemOS. It still remains to be seen if the XtreemOS-G layer alone would be useful to NuGO without running a Kerrighed or Linux-SSI kernel in the XtreemOS-F layer. This might be useful if the XtreemOS-G 'grid' layer could be used to connect NBXs on the NIN together. However, we were not able evaluate this aspect of using XtreemOS without SSI during the project.

The standalone 'desktop' edition of XtreemOS was installed and functioned correctly on “kitcat” without any modifications or re-configuration. However, the re-branded “Linux-SSI” kernel used by XtreemOS-F is Kerrighed 2.4 kernel, and the 'cluster' edition of XtreemOS did not run reliably for the same reasons as Kerrighed, even on a two-node stand-alone NBX XtreemOS cluster, but XtreemOS 'cluster' edition ran reliably on a single NBX without any problem. The first part of our evaluation concerned the feasibility of 'porting' XtreemOS to a Debian/Ubuntu-based Linux. This would be a fundamental requirement of using XtreemOS to provide a distributed IT infrastructure for NuGO based on the existing NBX data sharing network running NuGO-Linux/Bio-Linux 5.

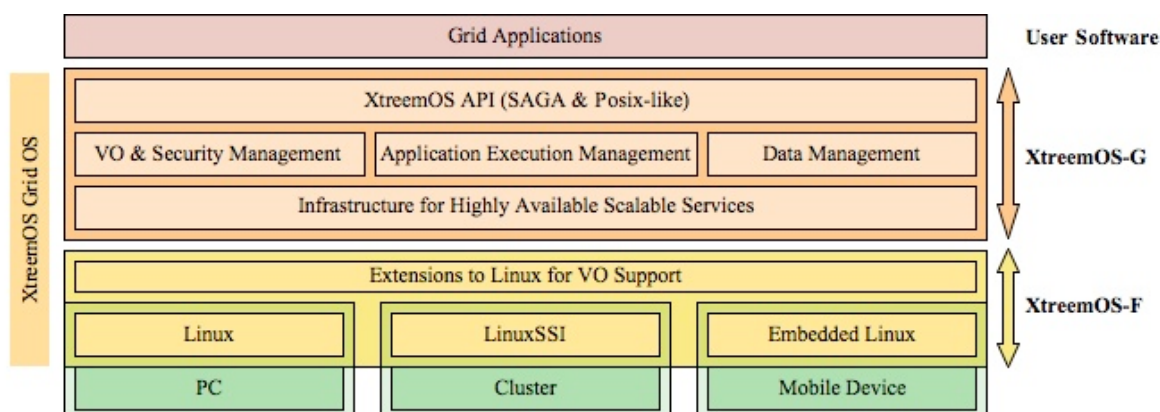


Figure 4. overview of XtreamOS

XtreamOS repository packages³⁴, containing applications unique to the XtreamOS distribution, were installed on the “kitcat” NBX cluster server under Bio-Linux 5. It was necessary to unpack the source RPM packages that are contained in the Mandriva Linux XtreamOS distribution and then compile XtreamOS from source manually, because the Debian/Ubuntu Linux distribution that NuGO-Linux is based on uses a different “apt” package management system. However, after unpacking the sources, XtreamOS installation could proceed normally. There did not appear to be any major compatibility problems with these manually compiled source packages and the binary distribution of XtreamOS, but further testing would be required in order to confirm compatibility. Other problems, such as bugs in the distribution, and security risks caused by installation under a Debian/Ubuntu-based Linux instead of the Mandriva-based Linux distribution were not tested.

XtreamFS

XtreamFS is an object-oriented distributed file system designed for a distributed IT infrastructure connected by the WAN, and forms part of the XtreamOS-G layer. This functionality is currently provided on the NBX data sharing network by the 'user-space' “sshfs” filesystem and Unix/Linux 'automounter'. Like “sshfs”, XtreamFS can be mounted remotely from anywhere in the Internet. However, XtreamFS is considerably more robust than “sshfs” and supports much more advanced capabilities for data management. The Kerrighed project initially supported a KFS (Kerrighed File System) directly in the modified Linux kernel. However, KFS caused Kerrighed kernel instability and this feature was dropped from later versions of the Kerrighed SSI kernel. The KFS feature of Kerrighed is now replaced by XtreamFS and is developed in the same family of related projects.

The XtreamFS-0.99.0 client and server programs were installed on “kitcat” and also on the second NBX cluster server “badcat”. The XtreamFS server is implemented as a Java servlet in user-space, in contrast to the original kernel-space KFS module, and operates as a user-space filesystem. Our tests of XtreamFS revealed that it worked very well between NBX's connected locally via a high-speed LAN and the demonstration tests and tutorials worked without any problems on our NBX servers. However, we were not able to connect to NBX's running XtreamOS over the WAN.

XtreamFS is a sophisticated filesystem supporting distributed filesystem meta-data and uses replication to provide fault tolerance or increased performance by 'striping' filesystems across different servers in a similar way that filesystems can be striped across local disks. However, many different ports need to be open in the firewall to allow the full functionality of XtreamFS to be used,

34 Can be accessed directly at <http://gforge.inria.fr/plugins/scmsvn/viewcvs.php/?root=xtreemos>

in contrast to the single SSH port 22 used by “sshfs”. One problem facing distributed infrastructure projects is security and it is unusual for NuGO partners to have many ports open to NBX servers.

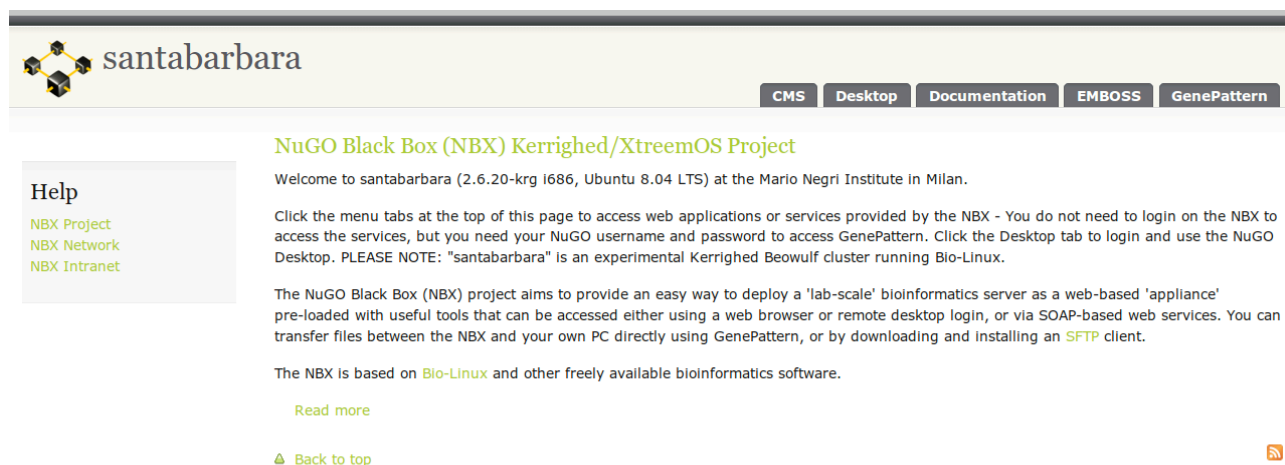
We were unable to open all the ports required to run XtremFS properly between NBX's that are behind firewalls outside our administrative control. No such problems occurred when using XtremFS inside a firewall, because all necessary ports were accessible (or we were able to open the required ports in the NBX “iptables” firewall). This is not an insurmountable problem, but it is a factor to be considered when planning a distributed IT infrastructure. The single most difficult problem facing the NBX data sharing project to date is obtaining permission for NBX's connected to the Internet at NuGP partner sites to accept incoming connections on TCP ports 22 (SSH) and 80 (HTTP). The number of ports required to be open for XtremFS is considerable more than this.

Kerrighed SSI cluster between two NBX's over the WAN (Wide Area Network)

Because of our limited opportunity to evaluate the full grid functionality of XtremOS directly, we decided to evaluate the potential of Kerrighed SSI load-balancing and resource sharing over the WAN (Wide Area Network). Although this is not directly comparable to using the XtremOS-G layer, it provided some insight into how NBX's could cooperate and resource share via the WAN.

A second NBX cluster server “sanatabarbara” was built at the Mario Negri Institute in Milan, and connected to the Internet. As with the “kitcat” NBX cluster server, “sanatabarbara” has the same Drupal-base Web-GUI as a standard NBX. This is shown in Figure 3. Our objective was to evaluate how Kerrighed, and potentially XtremOS, might be used for cooperation between geographically distributed NBX's connected to the Internet, with a relatively low bandwidth available via the WAN, compared to NBX's connected together locally on a high-speed LAN.

The first problem to resolve in creating an NBX Kerrighed SSI cluster over the WAN is that of traversing perimeter firewalls with most ports closed according to security policies for Internet facing machines. Kerrighed process migration occurs via TIPC (Transparent Inter Process Communication) using random ports. A simple solution to this problem would be to use openVPN, but VPN is not an efficient transport because of significant overhead involved in VPN encryption. The solution we chose is to run an *unencrypted* VPN tunnelled through SSH on a TCP port connected between two NBX's on two separate private networks that are not directly accessible from Internet. Although this seems a complicated solution SSH encryption is very efficient.



The screenshot shows the web-GUI interface for the NBX cluster server "sanatabarbara". The page has a header with the server name and a navigation menu with tabs for CMS, Desktop, Documentation, EMBOSS, and GenePattern. A "Help" sidebar is visible on the left. The main content area features a title "NuGO Black Box (NBX) Kerrighed/XtremOS Project" and a welcome message. Below this, there are two paragraphs of text: the first explains how to access services and use GenePattern, while the second describes the project's goal of providing a web-based bioinformatics server. A "Read more" link and a "Back to top" link are at the bottom of the main content area. A small RSS icon is in the bottom right corner.

Figure 3. Web-GUI interface of the NBX cluster server “sanatabarbara”

This is our simplified use-case example: Two institutes each have one NBX server, but one is not powerful enough for all of the tasks that are required for a particular bioinformatics analysis. The two Institutes want to collaborate and share their data, as well as some of their NBX computational resources, but for security reasons they don't have permission to expose ports on their server to the internet in the DMZ (Demilitarized Zone) outside the perimeter firewall because of internal security policies. The two servers concerned are Ubuntu based NBX's, with a Kerrighed-enabled kernel.

The aim of the experiment was to obtain automatic process migration between two NBX's over a WAN connection as if they were connected together locally on a private LAN. Achieving this would mean it might be possible to create a virtual infrastructure between geographically distant parts of an interconnected network of NBX's, working together to share data and computational resources. In particular, we want to demonstrate that two distant NBX's, each on private LAN's without direct external access, and without routers using NAT (Network Address Translation) or PAT (Port Address Translation), can be connected together directly, efficiently and securely.

The NBX located at each partner (P1 and P2) creates an outgoing SSH connection, that is not blocked by local LAN security policy, to a trusted machine at a third partner (P3) in a secure location, but accessible from the Internet via SSH on port 22. The trusted machine (P3) acts as a proxy for communication between the two NBX's (P1 and P2), but P3 is only needed if P1 or P2 cannot be accessed externally as in our use-case example.

A secure tunnel is first established between SSH connections "P1<=> P3" and "P2<=> P3", so that P1 requests a socket on an agreed port on P3 and when a connection is made to that port the connection is forwarded over the SSH tunnel to a socket on P1. This connection will be used to create a server for VPN between the two NBX's. P2 also allocates a socket to listen to on an agreed port and if a connection is made to this port, the connection is forwarded over the secure tunnel, and a connection is made to P3 on the same port forwarded by the other connection. P2 will be used as a server for VPN between the two NBX's. Once this bi-directional SSH tunnel is established, P2 can connect to any local port and the connection will be securely forwarded to P1, as if P1 was directly accessible from the Internet, but instead using the SSH server listening on port 22.

The second step is establishing VPN between the two partners over the established SSH tunnel. An SSH tunnel can only forward TCP connections, but OpenVPN can also be used to forward UDP connections. If OpenVPN is configured for connection over the established SSH tunnel, the two NBX's can exchange both TCP and UDP packets. The final step is allowing TIPC to flow over the SSH encrypted OpenVPN tunnel. TIPC is an efficient protocol designed for use in Beowulf cluster environment, and this is the protocol used by Kerrighed to allow communication between cluster nodes. TIPC is not normally used inside an SSH tunnelled VPN, but TIPCutils software can be used to make this possible. The commands required to set up the tunnels are shown in Figure 4.

Despite the multi-layered and high latency tunnel, Kerrighed performed quite well between the two NBX's over the WAN. Responses to simple commands were acceptable even when communication between the two NBX's was requested, which happens quite often because the RAM as well as the CPU's of the clustered NBX's are managed by Kerrighed as a single system). The experiment was successful and the WAN cluster allowed process migration to occur with an overhead (or latency) not much greater than that needed to migrate a process locally if it was already involved in another CPU-bound task. We compared two similar processes running on one NBX without any cluster connection, and then with the two NBX's connected over the WAN. The time necessary to start the processes on one partner, let one of the processes migrate to the second partner, and migrate back depended on the quantity of data migrated, but we demonstrated that it is possible to connect remote

NBX's as if they were connected locally without placing them in the DMZ or breaching security.

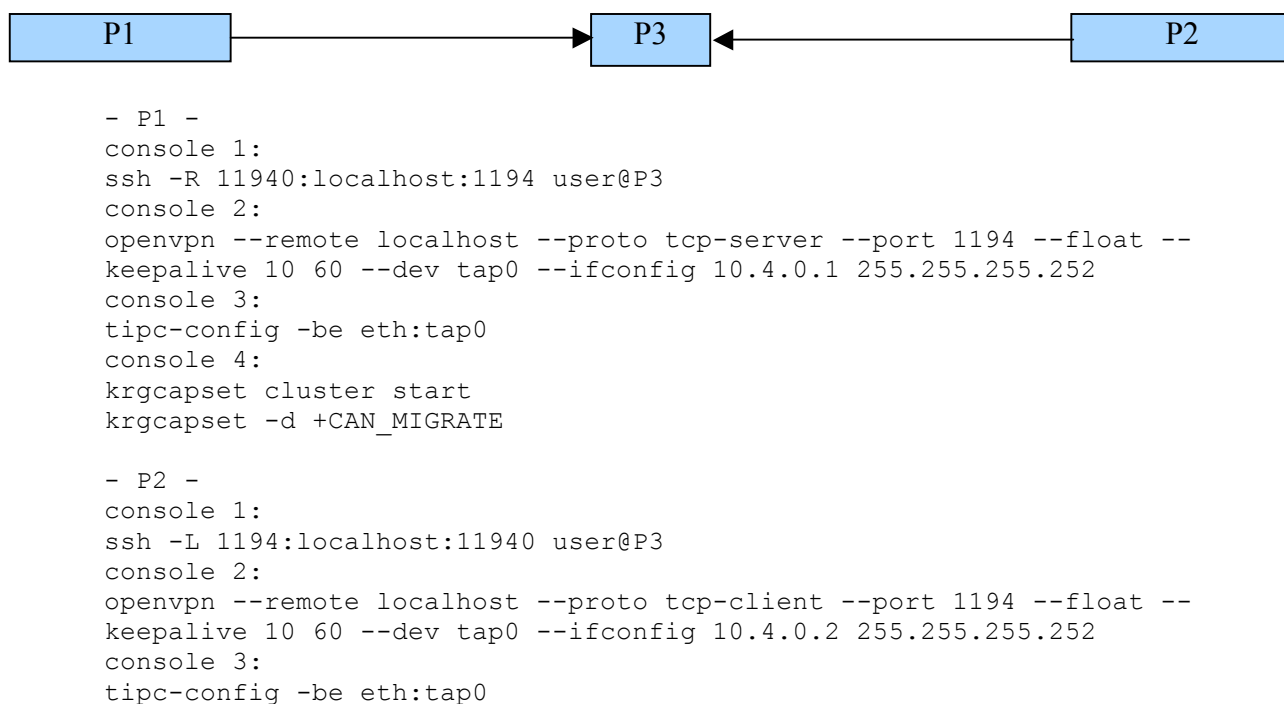


Figure 4. Connecting two NBX's over the WAN using a multi-layered secure SSH and VPN tunnel

The stability of the SSH connection can be improved using “autossh” to monitor and if necessary restart the SSH connection. OpenVPN is needed because although SSH has VPN capabilities they do not offer sufficient connection stability when an SSH tunnel bandwidth is saturated. OpenVPN has excellent stability, even on unstable network connections. TIPC-utils is needed to redirect the TIPC traffic over an SSH+VPN tunnelled connection. All of these packages are FLOSS.

Kerrighed is unstable if an NBX's crashes or if the network connectivity is compromised, and use of Kerrighed to create SSI clusters over the WAN is not recommended. However, virtualization of the SSI kernel and sharing only part of the computational capabilities of a machine using lightweight 'containers', that may to crash or even restart their communications without compromising the rest of the machine are being considered for future versions of Kerrighed. With these developments, it may be worth considering the use of Kerrighed to create WAN Beowulf SSI clusters.

CONCLUSIONS

Kerrighed stability

The stability of the Kerrighed SSI cluster was severely compromised when later 2.4 versions of the Kerrighed software were tested on the NBX cluster under 32-bit NuGO-Linux. However, the older 2.3 version of Kerrighed was stable enough to be seriously considered for use as a production platform. Stability problems prevented a detailed evaluation of the full 'grid' functionality of the XtreamOS-G layer, but experiments concerning cooperation between NBX systems connected via the WAN (Wide Area Network) demonstrated that Kerrighed might provide a useful load-balancing option for running 'embarrassingly' parallel bioinformatics applications on multiple NBX's when little or no communication occurs between the processes.

XtreemFS

The object-oriented XtreemFS filesystem was also evaluated and performed very well. However, concerns about the requirement for multiple network ports to be open in the perimeter firewall at NuGO partners limit the practicality of this to be considered as a replacement for the “sshfs” user-space filesystem that is currently in use for data sharing between NBX's over the WAN.

Application performance

Kerrighed 2.3 provides a stable 32-bit SSI platform compatible with the existing NBX infrastructure and might be useful to aggregate the CPU and memory resources of existing 32-bit NBX's connected locally via a high-speed LAN at particular NuGO partners. This was demonstrated as a practical way of improving the performance of the GenePattern bioinformatics software running on the prototype NBX Beowulf SSI cluster constructed for this evaluation.

Processor architecture

It seems unlikely that Kerrighed will be supported beyond version 2.3 on the 32-bit Linux platform, which restricts its usefulness under the 32-bit NuGO-Linux currently installed on the NBX network. However, the NBX network will be upgraded to 64-bit NuGO-Linux and more recent versions of Kerrighed will then offer much greater stability on the 64-bit platform.

ACKNOWLEDGEMENTS

This work was conducted at the University of Aberdeen Rowett Institute of Nutrition and Health and the Mario Negri Institute in Milan. We wish to thank NuGO for their financial support for the Kerrighed/XtreemOS project, and RINH for providing the necessary facilities for us to build and operate the prototype NBX cluster. We also thank our colleagues at the University of Aberdeen and in NuGO for their many helpful suggestions during the course of the project.

REFERENCES

- Hartig, U., Travis, A., Rocca-Serra, P., Renkema, M., van Ommen, B., Boeing, H. (2009). Owner controlled data exchange in nutrigenomic collaborations: the NuGO information network. *Genes & nutrition*, **4(2)**, 113-22. doi: 10.1007/s12263-009-0123-8.
- Ommen, B., Bouwman, J., Dragsted, L. O., Drevon, C. A., Elliott, R., Groot, P., Jim Kaput, K., Mathers, J.C., Muller, M., Pepping, F., Saito, J. Scalbert, A., Radonjic, M., Rocca-Serra, P., Travis, A.J., Wopereis, S., Evelo, C.T. (2010). Challenges of molecular nutrition research 6: the nutritional phenotype database to store, share and evaluate nutritional systems biology studies. *Genes & Nutrition*. doi: 10.1007/s12263-010-0167-9.