Project no. IST-033576

XtreemOS

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

LinuxSSI Integration and Packaging in OSCAR

D4.1.4

Due date of deliverable: November 30th, 2007
Actual submission date: January 14th, 2008

Start date of project: June 1st 2006

Type: Deliverable
WP number: WP4.1
Task number: T4.1.4

Responsible institution: INRIA
Editor & and editor's address: Christine Morin
INRIA
Campus de Beaulieu
35042 Rennes
France

Version 1.5 / Last edited by Christine Morin / January 15, 2008

Revision history:

Version	Date	Authors	Institution	Section affected, comments
1.0	07/11/07	Jean Parpaillon	INRIA	first draft
1.1	21/11/07	Jean Parpaillon	INRIA	Include RPM updates from Mandriva (Nicolas Vigier)
1.2	04/12/07	Christine Morin	INRIA	Correction of typos and minor things after proof reading
1.3	13/12/07	Christine Morin	INRIA	Corrections taking into account Michael Schoettner's review
1.4	19/12/07	Matthieu Fertré	INRIA	Corrections taking into account Samuel Kortas's review
1.5	14/01/08	Christine Morin	INRIA	Final corrections

Reviewers:

Samuel Kortas (EDF),Michael Schoettner (UDUS)

Tasks related to this deliverable:

Task No.	Task description	Partners involved°
T4.1.2	Packaging and integrating LinuxSSI in OSCAR	INRIA*,EDGE,NEC

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

*Task leader

Executive Summary

LinuxSSI is the core XtreemOS foundation layer for clusters. It is based on the Kerrighed technology, originally developed by INRIA in collaboration with EDF R&D and now with the Kerrighed open source project (`www.kerrighed.org`). The OSCAR open source project eases the installation, configuration and administration of Linux based HPC clusters (`http://oscar.openclustergroup.org/`). Kerrighed has been packaged for OSCAR since November 2004 through the SSI-OSCAR project. Kerrighed is now distributed as a standard OSCAR package. This document describes the work that has been done since April 2007 to integrate recent changes from the OSCAR and Kerrighed projects into Kerrighed OSCAR package. We worked on packaging Kerrighed for OSCAR rather than LinuxSSI for OSCAR. In fact, the deployment method of LinuxSSI is the same one as for Kerrighed. The most recent version of Kerrighed (release 2.2.0) on which LinuxSSI is based cannot be released in the most recent version of OSCAR (release 5.0) due to the fact that OSCAR does not support the new diskless installation mode of Kerrighed. However, we worked on preparing the integration of the 2.1.1 version of Kerrighed and thus LinuxSSI in the next OSCAR release (the upcoming 6.0 OSCAR release).

Contents

1 Introduction

LinuxSSI [1] is the core of XtreemOS for clusters. As an SSI (Single System Image) based on the Kerrighed project[2], it gives the illusion that a cluster is a powerful grid node appearing as an SMP machine. OSCAR is a set of best practices tools to ease the deployment and management of clusters [5].

Packaging LinuxSSI for OSCAR offers a complete solution for deploying, configuring, managing and using a cluster.

The initial integration of LinuxSSI into OSCAR is described in Deliverable D4.1.2 [6]. The present document describes the changes in OSCAR and Kerrighed and their implication for the packaging of LinuxSSI for OSCAR.

1.1 Presentation of Kerrighed and Oscar Software

1.1.1 Kerrighed

LinuxSSI is based on the Kerrighed [2, 4] cluster operating system. As a Single System Image, it gives the view of a single machine over a fast network interconnected set of machines. Kerrighed implements a virtual multiprocessor SMP like machine.

Kerrighed system offers load-balancing through cluster nodes. The load balancing policy is configurable. Distributed services implementing global resource management in Kerrighed rely on the Transparent Inter Process Communication (TIPC) system available in Linux kernel.

Kerrighed can be customized as a user can selectively enable or disable functionalities of the SSI through an API or with command-line tools. These functionalities are:

migration: allow a process to migrate,

distant fork: allow a process to fork on a distant node,

see local `/proc`: allow a process to see the local `/proc` filesystem instead of the cluster-wide one,

change capabilities: allow a process to change its capabilities.

The Kerrighed system is implemented with a Linux kernel patch, a kernel module and some optional tools.

1.1.2 OSCAR Suite of Tools

OSCAR [5, 3] is a set of best practices tools for clusters. It contains software for deploying, configuring, monitoring and managing clusters. Most of these tools are distribution-independent and OSCAR officially supports 7 distributions: Fedora Core, CentOS, RHEL, Mandriva Linux, SLES, Scientific Linux and Debian. It is implemented in an architecture scheme where a master node controls the cluster nodes, as shown in Figure 1.
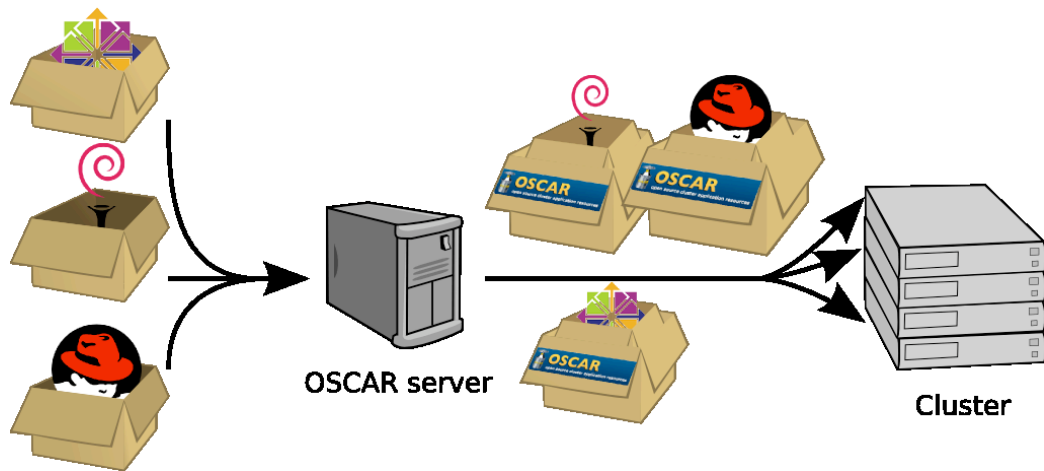

Figure 1: OSCAR architecture

To deploy software on a cluster, OSCAR uses a specific package format. This format includes cluster-level configuration information as well as a multi-distribution description of the software.

The latest version of OSCAR is OSCAR 5.0 released in November 2006.

1.1.3 SSI-OSCAR

Since November 2004, Kerrighed has been packaged in the SSI-OSCAR package [7, 8]. The first versions of the SSI-OSCAR package were developed for the OSCAR versions 4.0 and 4.1. As they included some specific scripts, they were not an official part of OSCAR, but third party packages. Furthermore, the Kerrighed version packaged in OSCAR at that time was the 1.0.2 one, based on Linux 2.4.29. This first version of SSI-OSCAR could only be deployed on distributions supported by OSCAR 4.x series.

Since 2007, the SSI-OSCAR package can be deployed with OSCAR 5.0 version as a standard add-on package.

In [6], we described the initial integration and packaging of LinuxSSI in OSCAR.

4

1.2 Projects Changes

Since the initial integration of LinuxSSI into OSCAR, Kerrighed, OSCAR and SSI-OSCAR have changed. The standard way of deploying Kerrighed uses diskless nodes. OSCAR packaging architecture has been modified. SSI-OSCAR project is now part of OSCAR.

In the following sections, we describe these changes and how they impact the packaging of LinuxSSI.

2 Kerrighed and OSCAR Changes

2.1 Kerrighed Deployment

The current version of Kerrighed is 2.2.0[1] released in November 2007. It includes some changes concerning its deployment.

2.1.1 Compilation

The use of *Autotools*, as described in Section 2.1 of [6], to configure sources compile and install them provides a set of documented options to enable or disable some features and setup installation paths.

First, some features were automatically enabled or disabled by the `configure` script, regarding presence or absence of required tools. It appeared that this auto-detection confused users. Now, if required tools are not present, the `configure` script fails and the user must explicitly disable the feature or install the tool. These features are (i) compilation of *man* pages (requires `xmlto`) and (ii) installation of *init* scripts (requires LSB tools and a supported distribution).

Meanwhile, all `configure` option names remain the same.

2.1.2 Installation

Kerrighed requires that all nodes run with identical filesystem. Read-only files can be deployed on each node and read-write files must be shared. Former versions of Kerrighed included an experimental distributed filesystem, KerFS. For reliability reasons, this filesystem is not used anymore hence, a shared filesystem like NFS must be used instead.

---

[1]See http://kerrighed.org/wiki/index.php/Download

2.1.3 Configuration

In Table 1, you can read which parameters are needed by Kerrighed to run and the way they can be specified.

Name	Boot parameter	/etc/kerrighed_nodes	Auto-detection
session id	X	X	
node id	X	X	X
network interface			X
nbmin	X	X	

Table 1: Kerrighed parameters

Major improvements relate to the auto-detection of some parameters. If `nbmin` is given, the cluster automatically starts when this number of nodes has started. If not given, the cluster is started manually with `krgadm cluster start`.

Auto-detection of node id simplifies the boot process of Kerrighed cluster as `/etc/kerrighed_nodes` is not required anymore and boot parameters are the same for all nodes of the cluster.

2.1.4 Impact on LinuxSSI for OSCAR

Post-installation script of LinuxSSI for OSCAR consisted in setting boot parameters for each node. This required deep modifications in OSCAR tools but they are not required anymore as boot parameters are the same for all nodes in the cluster.

The major issue is that deploying diskless nodes is not yet included in OSCAR, whereas it is required by Kerrighed. This support in OSCAR is planned for the next release (OSCAR 6.0) currently under development.

Thus, no official version of the SSI-OSCAR package based on the latest version of Kerrighed (2.2.0) has been released. In April 2007 (when Deliverable D4.1.2 [6] was edited), the SSI-OSCAR package was based on Kerrighed 2.0.0 extending Linux 2.6.11 kernel. The current version of the SSI-OSCAR package is based on the Kerrighed 2.0.0 version extending Linux 2.6.11 kernel and released in August 2007. This is the last version of Kerrighed still allowing a non diskless deployment.

2.2 Distribution Packages

Autotools options for Kerrighed have been stable since the initial integration of LinuxSSI into OSCAR. Then, RPM and Debian packages build process have not required specific development.

Nevertheless, as Kerrighed is now based on Linux 2.6.20, kernel configuration must be adapted to this kernel.

2.2.1 Debian Packages

We try to provide with Kerrighed Debian packages the same kernel drivers as with the Linux Debian packages. So, updating Kerrighed kernel packages consists in applying the same patches as to the Debian kernel, plus the Kerrighed patches.

The latest Debian packages for Kerrighed are based on Kerrighed 2.1.1 (relying on Linux2.6.20) and are available at the following url `http://kerrighed.gforge.inria.fr/debian/)`.

2.2.2 RPM Packages

RPM packages are now based on the Mandriva kernel. The package name has changed and the package version includes kernel version plus Kerrighed version. Thus, several versions of the kernel can be installed.

The Kerrighed kernel has been patched with *unionfs* to allow a Live CD.

The latest RPM packages for Kerrighed are based on Kerrighed 2.1.1 (relying on Linux2.6.20) and are available at the following url `http://people.mandriva.com/~nvigier/kerrighed-2.1.1/2007.1/`.

2.3 OSCAR Packages Architecture

2.3.1 Former OSCAR Package Format

OSCAR packages were made of distribution native packages (RPM or Deb), some configuration scripts and a description of the package in an XML file, all enclosed in a *tarball* file, as shown in Figure 2. This is briefly described in [6]. This format, that is used up to and including OSCAR 5.0 release had two limitations:

1. the tarball provides all architecture and distribution native packages, but only one is used on a cluster. This is a waste of space and bandwidth;

2. with its own packaging system, OSCAR must provide its own mechanism of dependencies, installation, uninstallation and repository management. These mechanisms are already implemented in the distribution native packaging systems.
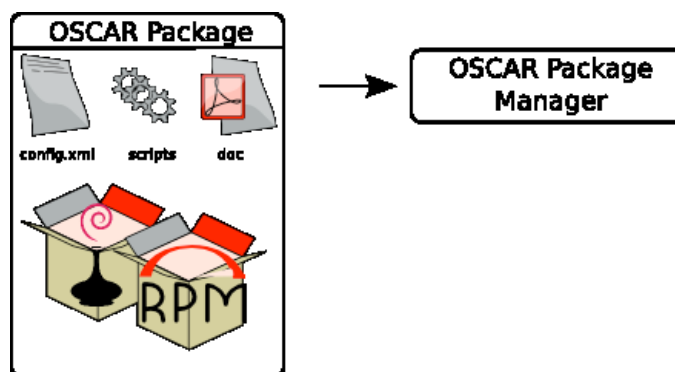
Figure 2: Old OSCAR Package Format

2.3.2 New OSCAR Package Format - opkg

A new OSCAR package format has been designed to solve the above-mentioned issues. We call a new OSCAR package, that will be used in the next OSCAR version (OSCAR 6.0 version that has not yet been released) "opkg". An "opkg" consists of a directory containing:

- configuration scripts,

- an XML description of the package,

- optionally, documentation,

- optionally, test scripts.

This "opkg" is then compiled to produce a set of three native packages:

API package: installed on OSCAR master node and containing scripts and documentation,

server package: installed on the cluster server,

client package: installed on the cluster nodes.

All packages, which were provided in the *tarball* file in the former format, are now set as requirements of the generated packages. This way, all package management is done by the distribution native package system, as shown in Figure 3.
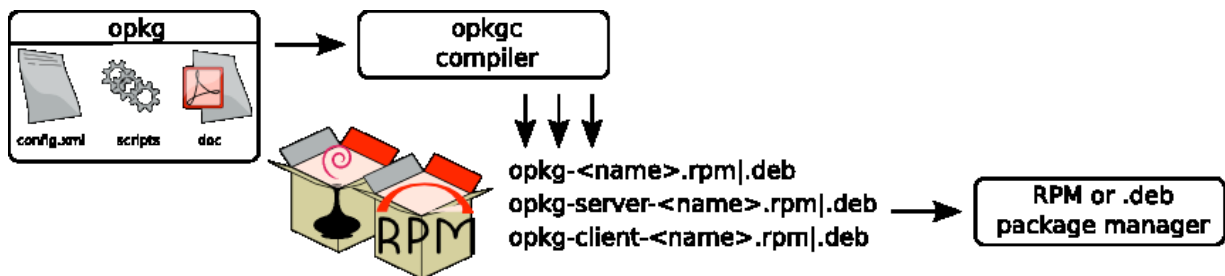
Figure 3: OPKG

2.3.3   Impacts on LinuxSSI for OSCAR

From a developer's point of view, few modifications are needed to convert a package from the old to the new format, the great part of the new architecture being implemented by the *opkg* compiler. Furthermore, a script provided with the compiler converts old format to new format, whenever possible.

We prepared a new version of the SSI-OSCAR package conforming to the new OSCAR format based on Kerrighed 2.1.1 version. This package is stored in the OSCAR software repository but has not been released officially as the diskless deployment mode used in Kerrighed 2.1.1 is incompatible with the current official version of OSCAR (OSCAR 5.0).

2.4   SSI-OSCAR Project

As SSI-OSCAR now requires only standard tools of OSCAR, it has been integrated in the OSCAR project. The code is managed on the OSCAR subversion server (see `http://svn.oscar.openclustergroup.org/trac/oscar/browser/pkgsrc/kerrighed` for online browsing or `http://svn.oscar.openclustergroup.org/svn/oscsar/pkgsrc/kerrighed` for checkout) and support is now provided on OSCAR mailing-lists.

SSI-OSCAR main developer is also member of the OSCAR core team. This allows a higher integration of LinuxSSI into OSCAR.

Since October 2007, the SSI-OSCAR package is maintained by the OSCAR open source community. The SSI-OSCAR project does not exist anymore as a standalone project. It has been integrated in the OSCAR project.

3   Current Status

- Debian packages are available for Kerrighed 2.1.1 `http://kerrighed.gforge.inria.fr/debian/`.

9

- RPM packages are available for Kerrighed 2.1.1 (`http://people.mandriva.com/~nvigier/kerrighed-2.1.1/2007.1/`).

- Kerrighed opkg (for Kerrighed 2.1.1) is currently in the OSCAR svn repository but has not been released because:

  - NFSROOT support (that would make OSCAR compatible with the Kerrighed diskless deployment mode) in OSCAR is still under development (planned to be released in OSCAR 6.0 version),

  - OSCAR 6.0 with the new package architecture has not yet been released. The planned release date is unknown.

Kerrighed Debian packages are maintained by the Kerrighed open source community (currently by Kerlabs). Kerrighed RPM packages are maintained by the XtreemOS consortium (currently EDGE-IT and RedFlag Software). The SSI-OSCAR package based on Kerrighed Debian and RPM packages are maintained by the OSCAR open source community (currently by Kerlabs).

The latest version of SSI-OSCAR package is based on OSCAR 5.0 and Kerrighed 2.0.0 extending Linux 2.6.11. This version has been released in January 2007.

4 Conclusion

LinuxSSI is based on Kerrighed. Kerrighed has been packaged for OSCAR in the SSI-OSCAR package. There has been no new version of the SSI-OSCAR package since January 2007. Since the writing of [6], there has been new releases of Kerrighed with important changes in the deployment process: compilation, configuration and last but not least installation. In Kerrighed 2.1.1 based on Linux 2.6.20, Kerrighed's installation mode is diskless. RPM and Debian packages have been built and released based on Kerrighed 2.1.1. These packages can be downloaded from Kerrighed website. There has been no new release of OSCAR since November 2006. As the latest release of OSCAR, OSCAR 5.0, is incompatible with the diskless installation mode of Kerrighed, it was not possible to release the new version of Kerrighed Debian and RPM packages in OSCAR. A new version of OSCAR, OSCAR 6.0, is under open source development. This version already offers a new format for OSCAR packages. It will also support the diskless installation mode in the future (the release date is unknown). We have built Kerrighed 2.1.1 packages based on the new format for OSCAR 6.0 (*opkg*). As soon as OSCAR 6.0 will be released with a support of the diskless mode, these new packages will be tested and then released.

The most recent version of Kerrighed is Kerrighed 2.2.1 released in January 2008. Debian packages for this version of Kerrighed will be built by the Kerrighed community. XtreemOS consortium plans to build the corresponding RPM packages. From these Kerrighed Debian and RPM packages, the OSCAR community will build the OSCAR packages for OSCAR 6.0 (they will only be released officially when OSCAR 6.0 will be officially released). Building OSCAR packages is an automated process taking as input Debian and RPM packages. Thus, most of the work to be done consists in testing the OSCAR package.

Concerning LinuxSSI, most of its functionalities are in Kerrighed. In addition, it includes essentially a distributed file system, KDFS, and an infrastructure for customizing the scheduling policy. Today, these functionalities do not have any important impact on the packaging. The XtreemOS consortium plans to build RPM packages for LinuxSSI. It will take advantage of the work done and experience gained for building Kerrighed RPM packages to build the LinuxSSI ones. Building OSCAR packages for LinuxSSI is not difficult as it will be done using the same automatic process as for building Kerrighed OSCAR packages. So most of the work to be done by XtreemOS consortium consists in testing LinuxSSI RPM and OSCAR packages.

References

[1] X. consortium. D2.2.7: Prototype of the basic version of LinuxSSI. Deliverable, December 2007.

[2] Kerrighed. Kerrighed website (http://www.kerrighed.org/), 2007.

[3] T. Mattson. Oscar: A packaged cluster software stack for high performance computing. Technical report, Open Cluster Group, June 2001.

[4] C. Morin, P. Gallard, R. Lottiaux, and G. Vallée. Towards an efficient single system image cluster operating system. *Future Generation Computer Systems*, 20(2), January 2004.

[5] OSCAR. Oscar website (http://oscar.openclustergroup.org/), 2007.

[6] J. Parpaillon. D4.1.2: Initial LinuxSSI integration and packaging in OSCAR. Deliverable, Feb. 2007.

[7] J. Parpaillon. Ssi-oscar website (http://ssi-oscar.gforge.inria.fr/), 2007.

[8] G. Vallée, S. L. Scott, C. Morin, J.-Y. Berthou, and H. Prisker. Ssi-oscar: a cluster distribution for high performance computing using a single system image. In *The 3rd Annual OSCAR Symposium*, University of Guelph,

Guelph, Ontario, Canada, May 2005. Held in conjunction with the 19th International Symposium on High Performance Computing Systems and Applications (HPCS 2005).