

SSI-OSCAR

A Single System Image for OSCAR Clusters

Geoffroy Vallée

INRIA – PARIS project team

COSET-1

June 26th, 2004

Clustering: Issues

- Clusters = distributed architecture
- 2 main problems:
 - Cluster management
 - system installation : FAI/System imager, KaTools
 - system update : apt
 - user management
 - Cluster programming / use
 - support of MPI, OpenMP, ...etc.
 - SSI Operating System

SSI-OSCAR

- Goals: to provide an open source SSI distribution for clusters which guaranties both
 - ease of use, of programming
 - easy administration
- SSI-OSCAR merges
 - advantages of a cluster distribution
 - advantages of a SSI approach

Global solution for clusters

Cluster Distribution: OSCAR

- OSCAR: Open Source Cluster Application Resources
- A snapshot of best known methods for building, programming, and using clusters
- Consortium of academic/research & industry members

Oscar's Functional Areas

- Cluster installation
- Programming environment
- Workload management
- Security
- Administration
- Maintenance
- Documentation
- Packaging



Oscar Packages

- Administration/Configuration
 - SIS, C3, OPIUM, Kernel-Picker, NTPconfig cluster services (dhcp, nfs, ...)
 - Security: Pfilter, OpenSSH
- HPC Services/Tools
 - Parallel Libs: MPICH, LAM/MPI, PVM
 - OpenPBS/MAUI
 - HDF5
 - Ganglia, Clumon, ... [monitoring systems]
 - *Other 3rd party OSCAR Packages*
- Core Infrastructure/Management
 - System Installation Suite (SIS), Cluster Command & Control (C3), Env-Switcher,
 - OSCAR DAtabase (ODA), OSCAR Package Downloader (OPD)

SSI Systems

- Goals
 - Global resource management
 - Resource access transparency
 - High availability
 - Multi-programming
- Examples of SSI systems: OpenSSI, OpenMosix, Genesis, Kerrighed

SSI-OSCAR Overview

Distributed Services
(e.g. OpenPBS/MAI)

Libs

Tools

Libs

Tools

Libs

Tools

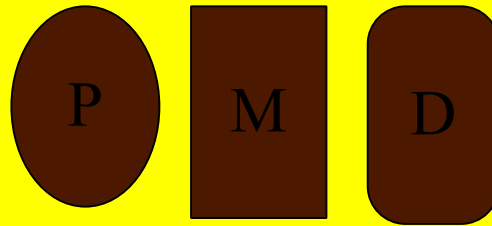
Kernel

P

M

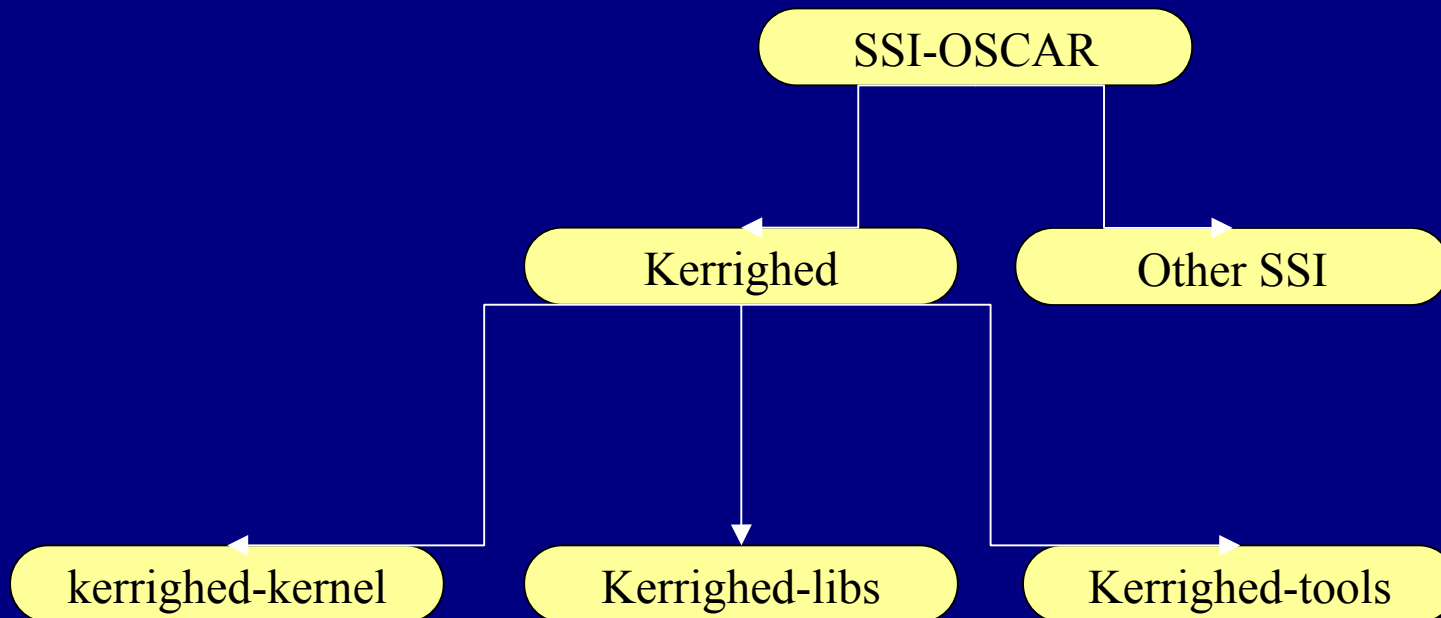
D

SSI



Architecture

- First step: OSCAR + Kerrighed



SSI: Kerrighed



- Operating system for clusters providing an SSI system
- Native support of message passing and shared memory programming paradigms
- Extension of the Linux kernel
 - Kernel patch
 - Kernel modules (independent from the kernel)
 - User API (libraries)
 - Kerrighed's Tools

Kerrighed Overview

- Resource global management
 - memories: remote paging, SDSM through *containers*
 - disks: distributed file system, exploiting *containers*
 - processors:
 - configurable global scheduler,
 - based on efficient process management mechanisms (migration, remote creation, checkpointing)
 - data streams: socket migration, global pipe management
- Provides the full Posix Thread interface

Kerrighed Architecture

OpenMP

MPI

POSIX Thread

libKerrighed

Global
Process
Management

Global
Memory
Management

Global Data
Stream
Management

Global Disk
Management

Kerrighed

Linux Kernel

User
Space

Kernel
Space

Kerrighed Tools

- Simple tools for monitoring
 - global use of memories
 - global use of processors
 - compute the global load of the cluster
- No tools for node installation
 - node installation
 - kernel copy on each node
- ↻ take advantage of OSCAR tools

SSI-OSCAR

- Creation of SSI-OSCAR package
 - Package Kerrighed (Linux kernel and Kerrighed system)
 - Create the OSCAR package for the SSI system
 - Adapt OSCAR to the SSI approach (not completely done)
 - modifications of the installation procedure
 - Integration of OpenPBS with the Kerrighed's cluster scheduler

Kerrighed Package

- Porting Kerrighed on the platform used by OSCAR (currently only RedHat 9.0)
 - Compiling issues
 - Initially, Kerrighed is developed for Debian with the gcc-2.95 compiler
 - Kernel patch : to compile the Linux kernel with Kerrighed patch on RedHat 9.0
 - Packaging (RPM)
 - Kerrighed modules and scripts
 - Kerrighed sources
 - Linux kernel with Kerrighed patch
 - Sources of Linux kernel with Kerrighed patch

SSI-OSCAR Distribution

- Current state
 - Kerrighed packages ready
 - Currently modification of the install procedure
 - No other modifications of OSCAR at this time
 - thanks to the transparency property of Kerrighed
 - Integration of OpenPBS
 - Validation of the Kerrighed OS with OSCAR packages

Partners

- Électricité de France (EDF R&D)



- Institut National de Recherche en Informatique et en Automatique (INRIA)



- Oak Ridge National Laboratory (ORNL)

OAK RIDGE NATIONAL LABORATORY

Conclusion

- SSI-OSCAR is a complete distribution for clustering
 - easy to use
 - easy to manage
- Website <http://ssi-oscar.irisa.fr/>
- First release candidate coming soon
- Future works
 - include another SSI system?
 - collaboration with HA-OSCAR?
 - federation of clusters?

Outlines

- Introduction
 - OSCAR presentation
 - SSI presentation
- SSI-OSCAR
 - Presentation of OSCAR
 - Presentation of an SSI system (Kerrighed)
 - Current state

Cluster Management

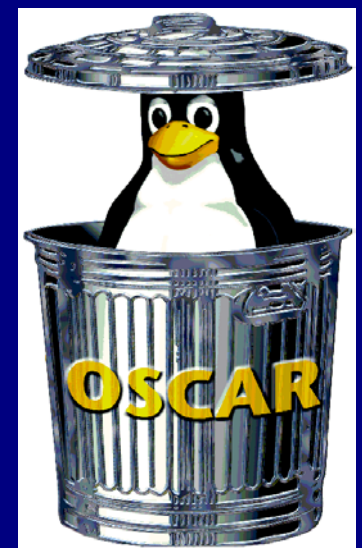
- System installation/update
 - node installation (initial installation, failures)
 - node update (software update)
- Network configuration
- User management
- Load balancing management

Cluster Use

- Distributed resources => hard to efficiently use, hard to program
- Standard hardware => regular failures to manage
- An approach : hide the resource distribution
 - Single System Image approach: to give the illusion that a cluster is an SMP machine
 - Global management of cluster's resources
 - High availability

OSCAR

- Wizard based cluster software installation (OS & environment)
- Automatically configures cluster components
- Increases consistency among cluster builds
- Reduces time to build/install a cluster
- Reduces need for expertise



OSCAR Package

- Creation of a new OSCAR package (like C3 tools)
- Currently in development, should be available in few weeks