SSI-OSCAR: a Distribution For High Performance Computing Using a Single System Image

Geoffroy Vallée (INRIA / ORNL / EDF), Christine Morin (INRIA), Stephen L. Scott (ORNL), Jean-Yves Berthou (EDF), Hugues Prisker (EDF)

OSCAR Symposium, May 2005

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

OAK RIDGE NATIONAL LABORATORY

Context

- Clusters: distributed architecture
  - difficult to use
  - difficult to manage
- Different approaches
  - do everything manually
  - use software suite to simplify management and use (*e.g.* OSCAR)
- This solution does not completely hide the resources distribution
  - use a Single System Image (SSI)
  - all resources are managed at the cluster scale
  - transparent for users and administrators
  - gives the illusion that a cluster is an SMP machine

What is a Single System Image?

- SSI features
  - Transparent resource management at the cluster level
  - High Availability: tolerate all undesirable events that can occurs (node failure or eviction)
  - Support of programming standards (*e.g.* MPI, OpenMP)
  - High performance
- A Solution: merge OSCAR and an SSI
  - Simple to install
  - Simple to use
  - Collaboration INRIA / EDF / ORNL

SSI - Implementation

- Key point: global resource management

User level: middle-ware	Limitations for functionalities and efficiency (*e.g.* CONDOR)
Kernel level: OS	Complex to develop and maintain (*e.g.* MOSIX, OpenSSI, Kerrighed)
Hardware level	More expensive (*e.g.* SGI)

Kerrighed – Overview

- SSI developed in France (Rennes), INRIA/IRISA, in collaboration with EDF
- Management at the cluster scale of
  - memories (through a DSM)
  - processes (through mechanisms for global process management and a global scheduler)
  - data streams (mechanism for global management of sockets, pipes, ...etc.)
  - disks (ongoing work)
- Extension of the Linux kernel 2.4.29 for x86
- About 100,000 lines of kernel modules
- GPL project - http://www.kerrighed.org/

Kerrighed - Issue

- Why OSCAR is interesting for Kerrighed?
- Kerrighed installation
  - Users: “Kerrighed seems to be interesting, I want to test it”
  - Kerrighed team:

    “Well, download Kerrighed sources, the linux kernel blabla, patch the kernel, compile the kernel, compile kerrighed, install the kernel, configure the kernel, configure kerrighed, reboot the machine, launch kerrighed, ... and pray!”
- Simple for... kernel hackers only
- If I do a mistake?
  - restart installation steps from scratch

SSI-OSCAR - Introduction

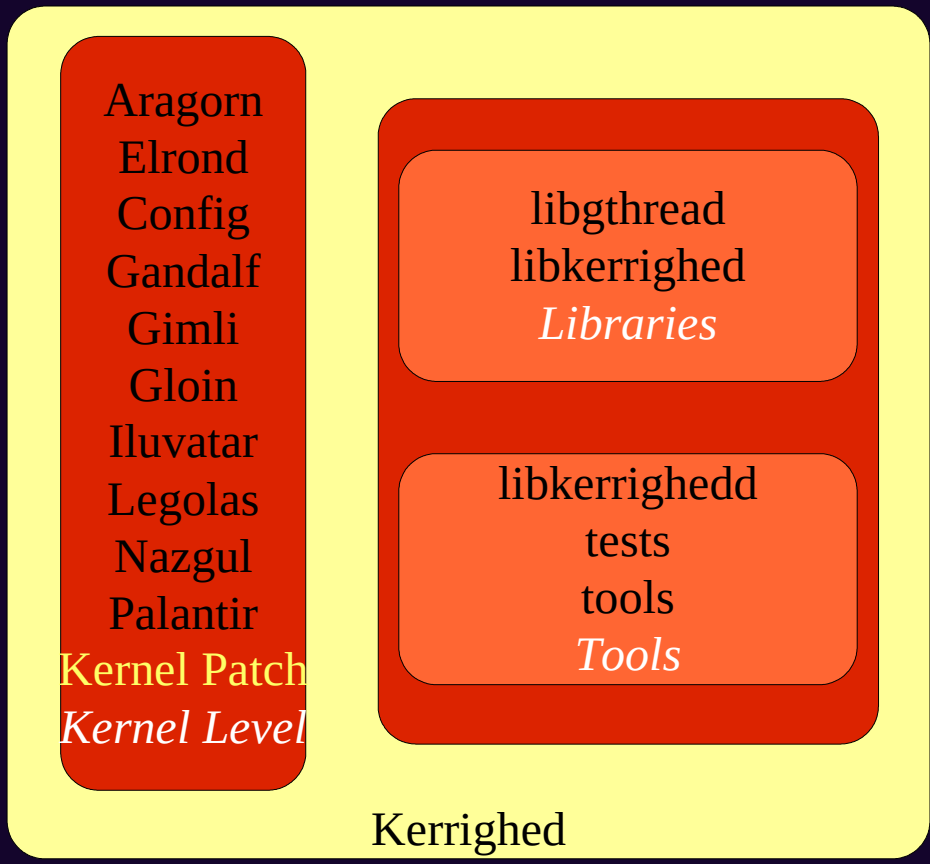
- Main problem: SSI (Kerrighed) difficult to install / setup
- A solution
  - Use OSCAR to install / setup the SSI cluster
    - create an image including Kerrighed
    - create configuration scripts to automatically configure the Kerrighed cluster
  - Allows to transparently and automatically configure an SSI cluster
    - No longer need any expertize in cluster management even to install / manage SSI clusters

SSI-OSCAR - Implementation

- Step 1: Kerrighed packaging
  - modify Kerrighed scripts to match to OSCAR tools
  - port on the compiler used by target Linux distributions
  - create binary packages
- Step 2: Integration in OSCAR
  - creation of an OSCAR package for the SSI
  - integration in OSCAR
- Step 3: Test & Validation
  - test and validation of the complete “distribution”

SSI-OSCAR - OSCAR Packages

- Organization: binary packages
  - kernel
  - kernel modules
  - tools
  - headers
  - libs
- Kernel issue: use of the OSCAR tool *kernel_picker*
- Creation of scripts to automatically create config files (*/etc/kerrighed/<config_file>*)



SSI-OSCAR Releases

- SSI-OSCAR 1.0
  - Based on Kerrighed 1.0rc8 (kernel 2.4.24), OSCAR 3.0
  - Support of Red Hat 9
- SSI-OSCAR 2.0
  - Based on Kerrighed 1.0.0 (kernel 2.4.24), OSCAR 4.0
  - Add the support of Fedora Core 2
- SSI-OSCAR 3.0
  - Based on Kerrighed 1.0.0, OSCAR 4.0
  - “Spin-off” OSCAR package
  - Improve the integration of the Kerrighed kernel
- Available at http://ssi-oscar.irisa.fr/

SSI-OSCAR - Experimentations

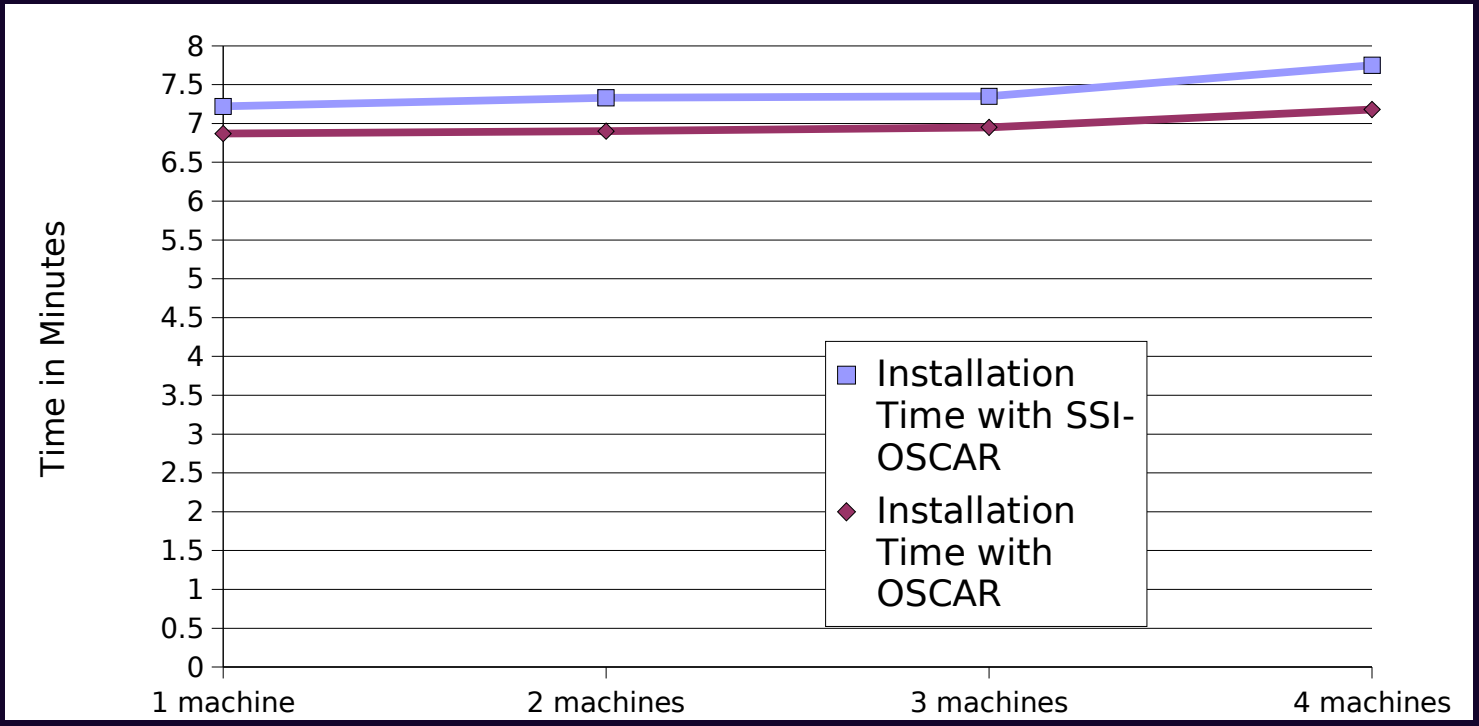
- Experimentations using SSI-OSCAR 2.0 RC 1
  - OSCAR 4.0
  - Kerrighed 1.0rc8 (kernel 2.4.24)
- Platform
  - Head node: P4, 1.7GHz, 1GB of memory, 40 GB HD
  - Compute nodes: 4 dual PIII, 450MHz, 512 MB of memory, 20 GB HD
  - 100MB Ethernet
- Try the evaluate the overhead created by the integration of the SSI

SSI-OSCAR – Image Size

- Evaluation of the size of a typical image for compute nodes
- Default OSCAR setup (default set of packages) + Kerrighed
- Image size = 678 MB vs. OSCAR image size = 615 MB
  - All sources for Kerrighed are included in the image
  - Impossible to remove distributed services already provided by Kerrighed => redundancy of mechanisms
- SSI-OSCAR 3.0 is smaller
  - “Spin-off” OSCAR package
  - Sources not included

SSI-OSCAR – Installation Time

- Time to fully install all compute nodes
  - Start time = machines boot
  - Stop time = reboot of the last machine after a successful installation



SSI-OSCAR – Unresolved Issues

- OSCAR: install a Beowulf cluster with a complete software suite to use the cluster
  - MPI, PVM
  - Distributed services of resource management (NFS, Torque)
- An SSI cluster: illusion of an SMP machine
  - No head node
  - Distributed service of resource management (*e.g.* global scheduler)

How to handle duplication of distributed services?

Package Set

- We need a mechanism to define dependences, conflicts between distributed services
  - Each distributed service is available as a OSCAR package
  - How to define these dependences/conflicts between OSCAR packages?
  - How to define the complete set of OSCAR packages for a typical configuration of the cluster?

  Current OSCAR features do not allow to create a *Package Set*

  => OSCAR Meta-Package
- Can be useful to integrate other “spin-off” projects based on OSCAR

Future Work

- Integrate a newest version of Kerrighed (1.0.2)
  - more stable
  - kernel 2.4.29
  - some new features
  - new architectures
- Integrate Kerrighed test in the testing OSCAR mechanism
  - allows users to test themselves their Kerrighed system
- “Maintenance” tasks
  - Integrate future Kerrighed releases
    - add/remove nodes
    - 2.6.x Linux kernel, x86_64
  - Integrate future OSCAR releases

http://ssi-oscar.irisa.fr/

http://www.kerrighed.org/

http://oscar.openclustergroup.org/