Grid Checkpointing Service

Heinrich-Heine University Duesseldorf

XtreemOS IP project is funded by the European Commission under contract IST-FP6-033576
What is XtreemGCP?

- A grid service integrated within AEM (WP3.3) providing job migration and job fault tolerance for grids.
- Service aims at integrating existing kernel checkpointers.
- By defining a common kernel checkpointer API implemented by translation libraries.
- **Job** = collection of job units.
- **Job unit** = collection of processes running on a grid node
- **Process** = managed by the LinuxXOS and LinuxSSI kernels of a grid node
- **Grid node** = PC or LinuxSSI cluster
XtreemGCP Architecture

Job Checkpointer

Job Unit Checkpointer

Common Kernel Checkpointer API
( implementing translation library per kernel checkpointer )

Kernel Checkpointer LinuxSSI-XOS

Linux SSI Cluster

Kernel Checkpointer Linux-XOS

Linux PC
- Checkpoint files stored in XtreemFS

- Checkpointing strategies controlled by job checkpointer
  - Coordinated checkpointing:
    - for job migration & fault tolerance
  - Uncoordinated checkpointing:
    - Avoiding coordination overhead
    - For fault tolerance, only

- Adaptive checkpointing:
  - Based on monitoring parameters (e.g. failure frequency)
  - Used to adapt checkpointing parameters and/or strategies
Common Kernel Checkpointer API

- For accessing different kernel checkpoints in an uniform way.
- Implemented by a translation libraries.

Job submission: jsdl file + **checkpoint properties**

Allocate grid node with appropriate kernel cp(s)

- PC kernel cp
- SSI kernel cp
- Condor libckpt

**Common Kernel Checkpointer API**
- CRExecMng.so
- Condor.so
- libckpt.so

**Kernel Checkpointers**
- Condor
- libckpt

---

XtreemOS IP project
is funded by the European Commission under contract IST-FP6-033576
- Book-keeping of process dependencies.
- Handle different process grouping techniques.
- Callbacks:
  - Processed during checkpoint and restart operation
  - Allows applications to optimize checkpointing
  - Used to drain communication channels
- Jobs share grid nodes → resource conflicts may occur

No Resource Isolation:

- jobA running ...
  - process_99200

- checkpoint jobA
  - process_99200

- migrate jobA
  - image and ...

- jobB running ...
  - process_99200

- jobB running ...
  - process_99200

- ... restart jobA
  - process_99200

PID 99200

RESOURCES CONFLICT
**Resource isolation:**

- **Cgroups tied to hierarchical name space:**

  NO RESOURCE CONFLICT

---

XtremOS IP project
is funded by the European Commission under contract IST-FP6-033576
XtreemGCP is an open service architecture integrating existing kernel checkpointing solutions.

Used for job migration and fault tolerance.

Status:
- Translibs for PCs using BLCR and LinuxSSI clusters available
- Basic Checkpoint and restart prototype working

Future work:
- Integration of cgroups
- Communication channel draining
- Integration of OpenVZ and Linux-native checkpointer