Research for Extreme-Scale Computing

Zhiling Lan

“Who We Are and What We Do” Seminar, 2010
Illinois Institute of Technology
Self Introduction

- Zhiling Lan
  - Ph.D., Northwestern University, 2002
  - Associate Professor
  - Email: lan@iit.edu
  - Office: SB #226D
  - Office hour: 3:00-4:00pm on Tu. & Th.

- Research group:
  - Scalable Computing Software Laboratory (SCS)

- Research website: http://www.cs.iit.edu/~zlan
Research Overview

- Research in the area of high performance computing
  - System software & applications
- Address **real problems** in **real systems**, with the goal to improve **performance** and **reliability**
  - Close collaboration with research staff at ANL, ORNL, SDSC, ...
  - Close collaboration with application scientists
  - Involving lots of system designs & experiments
- Increasing research opportunities, given the emerging of multi-core processors
High Performance Computing

- Use of supercomputers for running advanced applications efficiently, reliably and quickly
  - A supercomputer is a high-end computer with extremely fast processing capabilities, usually contains a large number of processors
  - Typical applications are programs from the fields of science and engineering, such as cosmological modeling, molecular dynamics, etc.
  - Parallel processing is often used to divide large problem into smaller ones so that they can be solved concurrently ("in parallel")
Extreme-Scale Computing

- The leading edge of high performance computing
  - Petascale \((10^{15})\), exascale\((10^{18})\), ....
  - One exaflop equals:
    - The combined performance of 50 million laptops - enough to reach 1000 miles from the ground when stacked, weighing over 100,000 tons.
    - 1000 times the power of today's most powerful supercomputer.

- Key challenges:
  - The energy and power challenge
  - The memory and storage challenge
  - The concurrency and locality challenge
  - The resilience challenge

Reliability Concerns

- Systems are getting bigger
  - 2k-16k processors is today’s “medium” size (92% of TOP500)
  - O(100,000) processor systems are being designed/deployed

- Even highly reliable HW can become an issue at scale
  - 1 node fails every 10,000 hours
  - 6,000 nodes fail every 1.6 hours
  - 64,000 nodes fail every 5 minutes

Fault tolerance/resilience
Losing the entire job due to one node’s failure is costly in time and CPU cycles!

From “Simplicity and Complexity in Data Systems at Scale”, Garth Gibson, Hadoop Summit, 2008
FENCE & RAPS

**FENCE**
Pre-failure prediction & tolerance

Take precaution action based on failure forecasting

**RAPS**
Post-failure diagnosis & recovery

Quickly resume computing after failure occurrence

In collaboration with
- Xian-He Sun (IIT)
- N. Desai, D. Buettner, R. Thakur, S. Coghlan, R. Gupta and P. Beckman (ANL)
- B.H. Park and A. Geist (ORNL)
- J. White and E. Hocks (SDSC)
FENCE Overview

- **emph{FENCE: Fault awareness ENabled Computing Environment}**
  - *Proactive actions* prevent applications from anticipated failures
  - *Reactive actions* minimize the impact of unforeseeable failures
### RAPS Overview

**RAPS for Post-Failure Diagnosis and Recovery**

- **Failure Diagnosis**
  - Multi-Layer Data Collection
  - 2D Diagnosis Time vs. Space

- **System Orchestration**
  - Resource Planner
  - Recovery Coordinator
  - Recovery-Aware Job Scheduler

- **Analysis & Assessment**
  - Failure Analysis
  - Recovery Assessment

- **Application Recovery**
  - Process Recovery
  - Data Recovery
  - System Support

---

- **RAPS: Recovery Aware Parallel computing Systems**
  - Algorithm design
  - System development
  - Empirical study

---

ILLINOIS INSTITUTE OF TECHNOLOGY
Failure Diagnosis - SYSDP

- **Challenges:**
  - Overwhelming amount of data
  - Redundant & unformatted data
  - Faults are many and complex
  - System changes are common

- **Our prototype tool**
  - Data mining, pattern recognition, statistical learning, ...
  - Preprocessing, ensemble learning techniques, dynamic relearning...
Job Scheduling

- Adaptive fault management
- Fault-aware scheduling
- Recovery-aware scheduling
- Joint work with the Cobalt group at Argonne
  - Cobalt is a production job scheduler used on HPC systems
Cosmology Simulations

- **PetaART Objective:**
  - To bring cosmological simulations (e.g., galaxy formation and evolution, formation of cosmic structures,...) into the petascale era

- **Research issues:**
  - Dynamic load balancing, fault tolerance, parallel I/O, ...

- **In collaboration with**
  - Univ. of Chicago, FermiLab, and Yale Univ.
Contact: Zhiling Lan
Email: lan@iit.edu