



Implementation of Diskless Cluster System

Paralel Processing Lab.

Kwangho Cha, Sangog Na, Dongsoo Han, and Chansu Yu

ICU Information and Communications
University

Overview

◆ Cluster of 50 Pentium III processors

- 24 SMP client nodes (48 CPUs)
- SMP server node (2 CPUs)
- Fast Ethernet , Serial Hub

◆ Diskless nodes

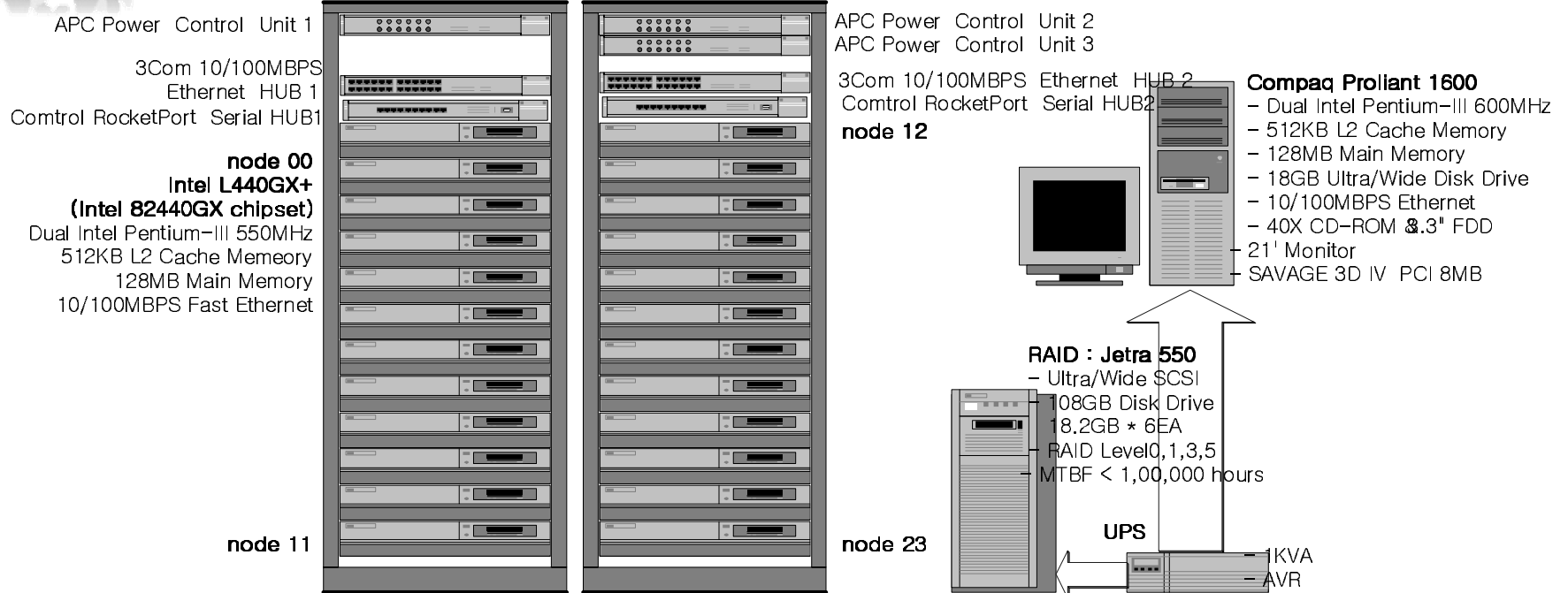
- Hard disk : causes of error and heat
- No HDD/FDD/CDROM/Network Boot ROM
- Single point of OS upgrade and maintenance
- Each client node downloads kernel image from server's hard disk

◆ Cluster management software

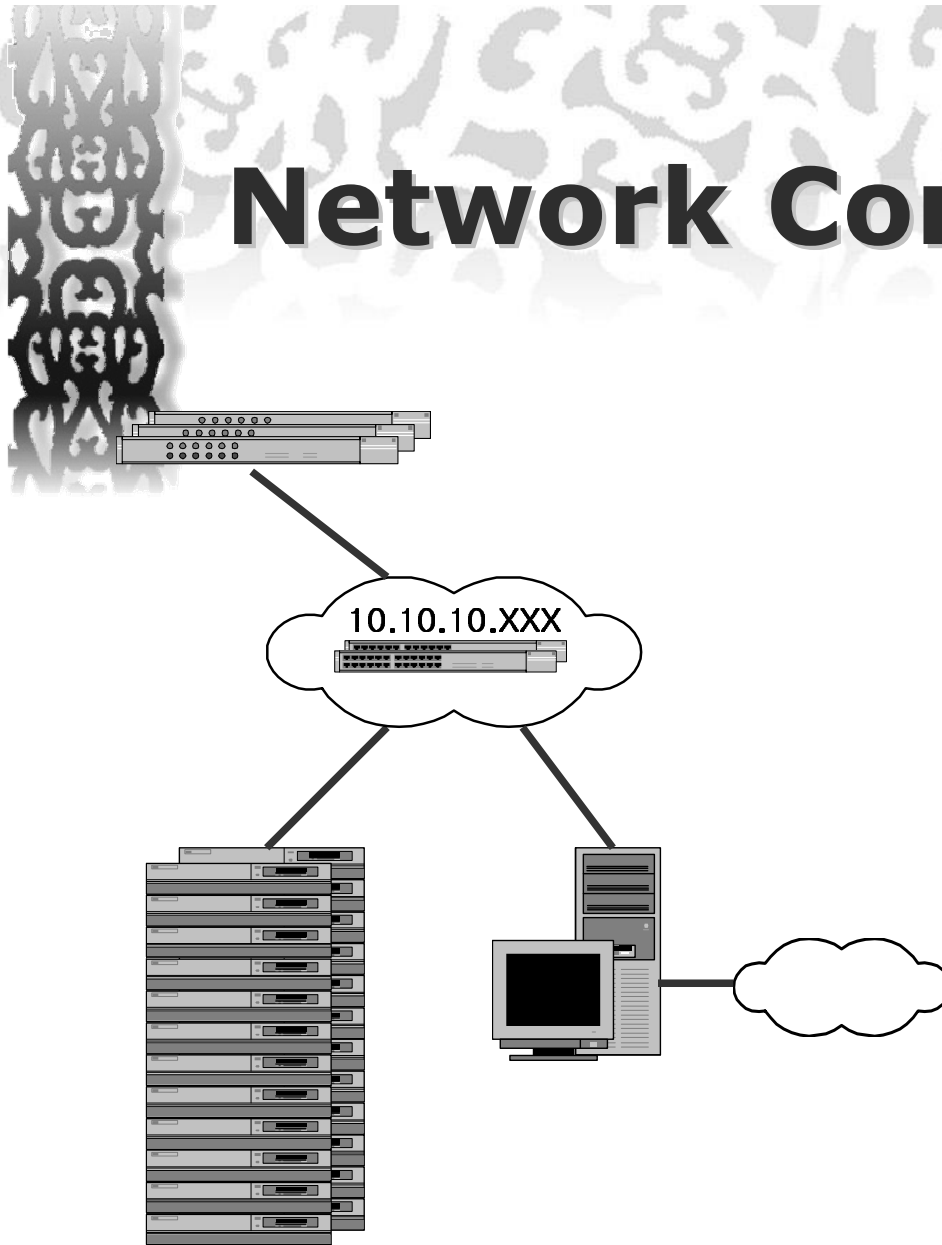
- Web-based management tool
- Power on/off, reset
- Program execution and Monitoring



Hardware Configuration



Network Configuration



◆ Server

- **pplab-5 : 210.107.131.172**
- **abc : 10.10.10.1**
- The server has two NIC, one for external area, and the other for client nodes

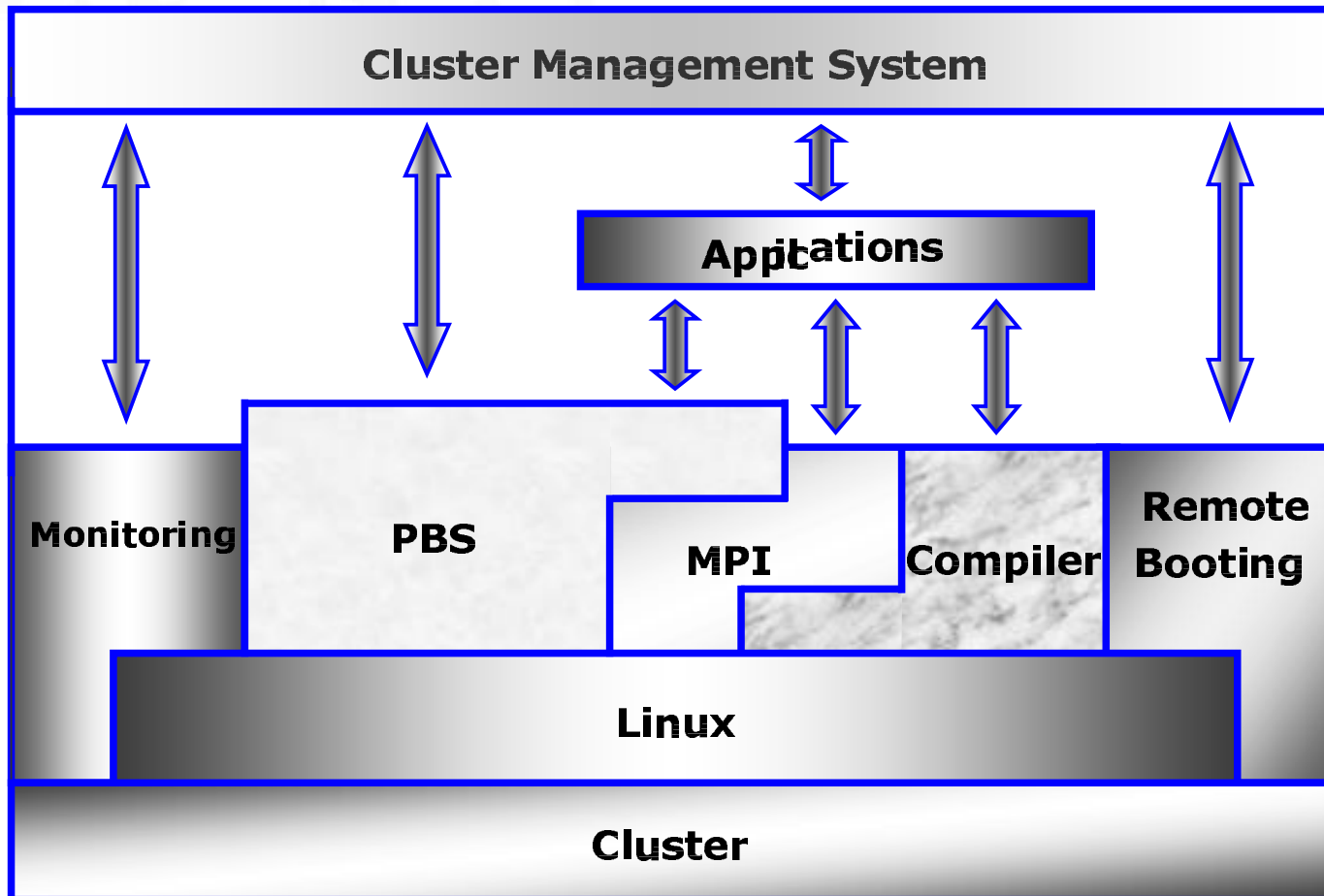
◆ Clients

- **node00 : 10.10.10.100**
- ~
- **node23 : 10.10.10.123**

◆ Master Switch

- **pms0 : 10.10.10.50**
- **pms1 : 10.10.10.51**
- **pms2 : 10.10.10.52**
- They have IP address and can be accessed by telnet or web-applications

Software Configuration





Software Components

Operating Systems : Alzza[‡] Linux 6.2 (kernel revision 2.2.13)

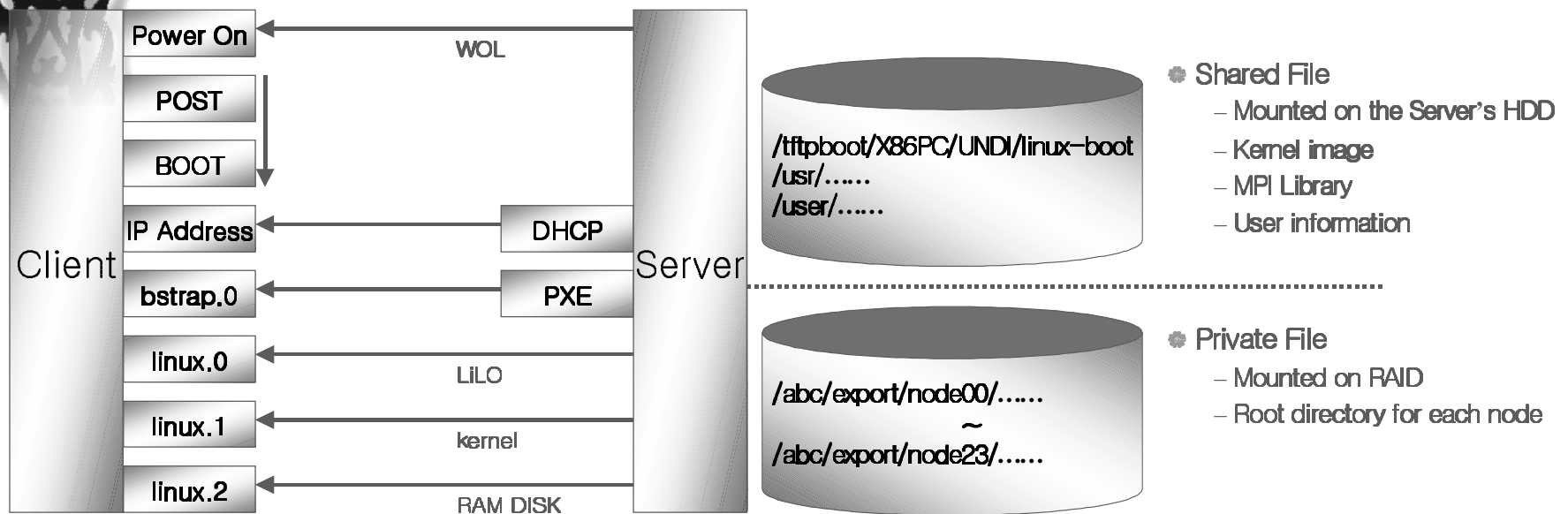
([‡] Linux's Korean version)

- ◆ **System Start-up : PXE-based remote booting**

- ◆ **Programming Environment**
 - **message passing library : LAM 6.3.2, MPICH1.1.2**
 - **compilers : GNU gcc, g77**

- ◆ **Cluster Management System**
 - **Web based GUI**
 - **Job Management System : OpenPBS 2.2 (Open Portable Batch System)**
 - **Monitoring System : Status of CPU, Memory, and Node**

Remote Booting Process & File Configuration



After the server turns on client nodes by sending 'WOL' packet, it provides 'IP address' and 'bootstrap(bstrap.0)' via DHCP daemon and PXE daemon, respectively

※DHCP : Dynamic Host Configuration Protocol
 ※PXE : Pre-boot Execution Environment

The server has two kinds of files, shared and private files. All client nodes and server can access the shared files.

Each client node has its own root directory for private files.

Cluster Management System

Management

User

Process

Power

Jobs

Edit

Submit

State

Monitoring

Node State

CPU

Memory

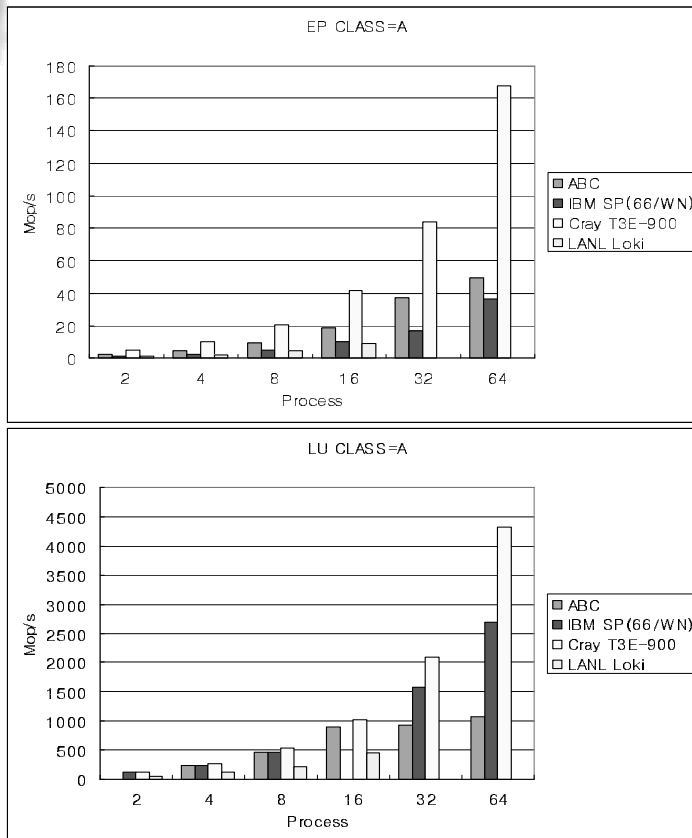
Manage user-id, process and client system power.

Edit and create jobs, and display Job states

Display current information of Node, CPU, and Memory

Performance & Booting Time

NPB (NAS Parallel Benchmark)



◆ Booting Time

Group Size	Gap		Total
	Gap	Total	
1	10 sec	386 sec	
	8 sec	338 sec	
	0 sec	Fail	
2	30 sec	421 sec	
	25 sec	419 sec	
3	40 sec	434 sec	
	30 sec	350 sec	
4	50 sec	341 sec	
	45 sec	315 sec	
	40 sec	291 sec	
5	45 sec	335 sec	
	40 sec	316 sec	
6	70 sec	366 sec	
	60 sec	Fail	
7	70 sec	300 sec	
	60 sec	Fail	

◆ Remote boot does not work if all nodes try to obtain the kernel image simultaneously.

◆ **Group size** : the number of nodes to obtain the kernel image simultaneously.

◆ **Gap** : a group's waiting time, after a previous group gets the kernel image.

◆ **Total** : complete time to boot

We select
Group size = 4,
Gap = 45 sec.



Problem History

2000. 4.20. : Purchased all components.

5.22. : Ethernet Hub's port error ➡ changed 1 unit.

6.20. : RAID failure ➡ changed 2 disks and RAID controller.

7.10. : UPS failure ➡ repaired UPS.

**11. 8. : Serial Hub's port error and Intel board error
➡ changed a serial Hub and one intel board.**

2001. 2.20. : Changed another intel board.

Current Problem

"There is a CPU problem in some nodes."

- ➔ **We used OEM processor (not genuine Intel box CPU).**
- ➔ **The same kind CPU is out of stock.**
- ➔ **Faster CPU box requires BIOS upgrade.**

"2 Intel boards also have problem."

These boards have problems with the power switch.

Monitoring DATA

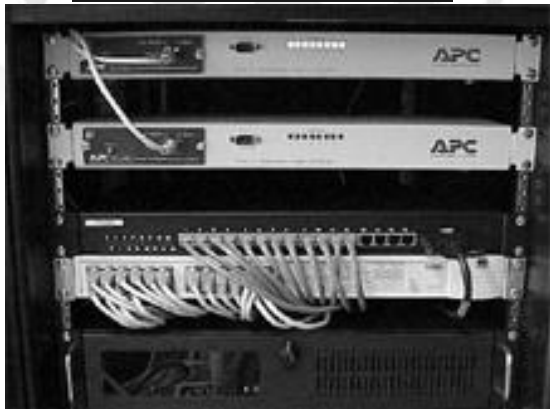
(using ASUS P3V4X ACPI BIOS
Revision 1003)

	Normal CPU (Genuine Intel BOX)	Abnormal CPU(OEM)
MB Temperature	26°C / 78°F	26°C / 78°F
CPU Temperature	58.5°C / 136°F	7.5°C / 44.5°F
CPU Fan Speed	4066 RPM	4753 RPM
VCORE Voltage	2.05 V	2.05 V
.....

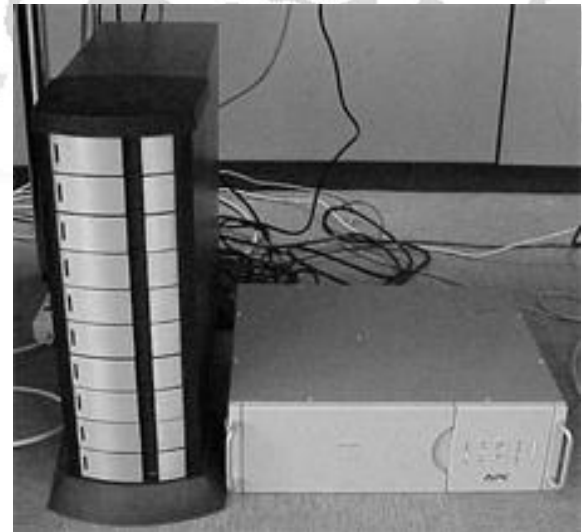
24 Client nodes



**Power switch(APC),
Rocket Port and
3Com Ethernet hub**



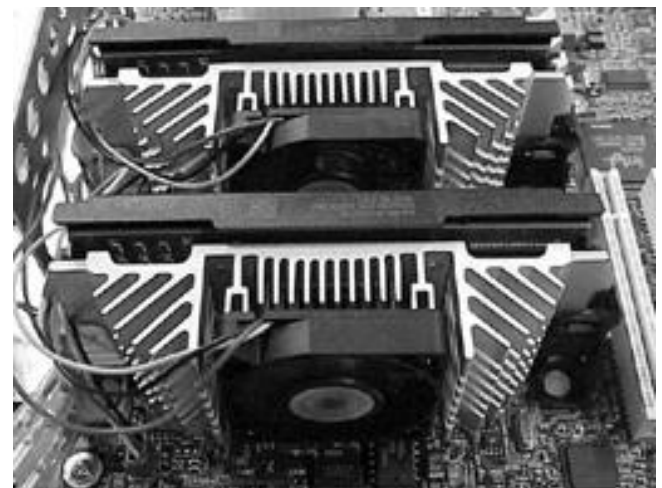
RAID(Jetra550) and UPS(APC)



**Inside of
client board(Intel L440GX+)**



**CPUs of our client board
(Intel Pentium III slot type)**



**Server system
Compaq Proliant 1600**

