Distributed Operating Systems

Overview
Ye Olde Operating Systems
OpenMOSIX
OpenSSI
Kerrighed
Distributed Operating Systems vs Grid Computing

User Space

Operating System

Nodes

Amoeba, Plan9, OpenMosix, OpenSSI, Kerrighed.

Grid System

User Space

Operating System

Nodes


OS OS OS OS OS OS

Xgrid, SGE, Condor, Distcc, Boinc, GpuGrid.
Distributed Operating Systems vs Grid Computing

Problems with the grid.

Programs must utilize that library system. Usually requiring separate programming.

OS updates take place N times.

Problems with dist OS

Security issues – no SSL.

Considered more complicated to setup.
Each node, even with distributed operating systems, boots a kernel.

This kernel can vary depending on the role of the node and overall architecture of the system.
Amoeba

Andrew S. Tanenbaum
Earliest documentation: 1986
What modern language was originally developed for use in Amoeba?
Anyone heard of Orca?
Sun4c, Sun4m, 386/486, 68030, Sun 3/50, Sun 3/60.
Amoeba

bootuser: using vdisk:02
bootuser: starting user process ...
bootuser: process started: wait ...
bootuser: ReInit
bootuser: Using disk vdisk:02 - size 2010
bootuser: NeuConf: 5 confs
BOOTSERVER INITIALIZED
bootuser: Soap_0: cappoll failed: server not found
bootuser: Soap_0 considered down
bootuser: Boot(Soap_0)
SOAP 0: initializing
SOAP 0: cache_init: estimated bytes per row = 100
SOAP 0: cache_init: rowlen = 100K, Max_rows = 568
SOAP 0: cache_Init: Max_dirs = 94
SOAP 0: caching 47 of 130 super blocks
SOAP 0: Bullet 0 is up
SOAP 0: super_init: global seqno in superblock is 10, 81
SOAP 0: starting in 1 copy mode
SOAP 0: coming up in 1-copy mode
SOAP 0: 4 threads started
bootuser: auto-switch off verbose

Welcome to standalone Amoeba
Plan9

Started development in the 1980's
Released in 1992 (universities) and 1995 (general public).
All devices are part of the filesystem.
X86, MIPS, DEC Alpha, SPARC, PowerPC, ARM.
Union Directories, basis of UnionFS.
/proc first implemented here.
Plan9

Rio, the Plan9 window manager showing "faces(1), stats(8), acme(1)" and many more things.
Plan9

Split nodes into 3 distinct groupings.
  Terminals
  File servers
  Computational servers
Uses the "9P" protocol.
  Low level, byte protocol, not block.
  Used from filesystems, to printer communication.
Author: Ken Thompson
Both Plan9 and Amoeba make groupings of nodes, into specific categories. This can mainly be attributed to the time period.

Starting with OpenMOSIX, there was a push to make the nodes identical, or at least breakout from the "grouping" model.
OpenMOSIX

SSI System. (Single System Image).
Automatic load leveling. (Procs not threads).
Patch for Linux 2.4.x
EOL March 1 2008.
Linux PMI - Linux 2.6.x branch.
LiveCD autoconfiguration available.
An OpenMOSIX cluster, running John The Ripper.
OpenMOSIX

Unique /mfs filesystem.

/mfs/here → / filesystem, current node.
/mfs/home → / filesystem, home node.
/mfs/selected → / filesystem on selected node, done by "echo # > /proc/self/selected"

Added /proc support.

/proc/hpc/nodes/[mosix ID]/(load|mem|speed), specific node statistics from remote /proc.
/proc/hpc/nodes/[mosix ID]/ is not a remote /proc (only pieces).
OpenMOSIX

Enable migration of sub processes:

"echo 0 > /proc/self/lock"

Useful for a shell.

Perl and Python modules available to ease programming specific applications.

Libmosix for C

Commonly used for large scale LTSP/POVRay.
OpenSSI

Last updated a year ago.
Kernel 2.6.12
http://openssi.org/cgi-bin/view?page=docs2/1.9/Introduction-to-SSI

Single process space.
  Global PID's, local information.
Single root.
No specific programming required.
  libcluster.so and cluster.h available.
  (rexec(), rfork(), etc.)

x86_64, x86 architectures
Lenny, Etch, Sarge, FC3, FC2, RH9
Access to remote /dev
OpenSSI

Stable release is FC2

Good example on how much activity.

Outdated timeline

- 2008 August - 2.0.0pre release number reserved for base kernel 2.6.16 or higher for OpenSSI stable.
- 2008 October - Preview of OpenSSI-1.9.6 (aka. 2.0.0pre3) - kernel bug fixes and performance improvements to VPROC, CFS, and PROCFS. (In CVS)
- 2009 Q1 - OpenSSI-1.9.6 for CentOS 4 - more kernel bug fixes, performance; re-enable CFS buffered I/O (**in testing**)
- 2009 Q1 - OpenSSI-1.9.6 for Debian Etch(?) (**in testing**)
- 2009 - OpenSSI-1.9 x86_64 64-bit port.
- 2009 - OpenSSI-1.9 port to kernel 2.6.18 or higher.
- TBD - OpenSSI-1.9 port to CentOS 5.
- TBD - OpenSSI-1.9 socket migration bug fixes.
OpenSSI

'localview' command
Prefix like nice.
Restricts that process to local devices, processes, and scope of ipcs.

'loadlevel' command
Algorithms borrowed from openMOSIX.
Turned off by default.
Can be turned on globally, or on individual nodes.
OpenSSSI

Extensive guides and tutorials

Out of date OS support.

Contrib contains xen kernel's. Possible to run massive paravirtualized guests.

Cluster virtual IP support

Similar to LVS (Linux virtual Server)
Kerrighed

Modification to the linux kernel.

2.6 branch
Current release of 2.6.30
x86, x86_64.

Single System Image

Single process space.

Checkpoint / Restart

Distributed memory.
Kerrighed

Single Process Space.
Global PID's
Mashup of statistics.
Kerrighed

Small Kerrighed cluster running on commodity hardware.
Kerrighed

Ubuntu 8.04
Mandrivia 2008.0-
Debian Lenny-

Support for x86 in $\leq 2.3.0$
Support for x86_64 $\geq 2.4.0$
Kernel arguments

session_id
The cluster identifier. Currently 256 clusters can be on the same network.

node_id
Individual node id. Used in internal workings.

Autonodeid
If set makes node_id=x in 192.168.0.x