

Sine Nomine Associates

OpenSolaris for IBM System z October 2008 Tech Update

David Boyes
Neale Ferguson

Agenda

- Why do this?
- Timeline
- Design Decisions
- Porting Process
- Progress Made
- Planned Future Work
- Q & A

OpenSolaris vs Solaris

- What's the Difference?
 - “You can think of OpenSolaris as Solaris.NEXT.” – Sun Marketing
 - OpenSolaris provides the building block technology for what will become the next release of commercial Solaris (in fact, ‘uname –a’ IDs as Solaris 11)
 - Solaris is the core technology available in OpenSolaris PLUS a bunch of add-ons from other parts of Sun and third-parties.
 - OpenSolaris is not (yet) part of the commercial support regimen from Sun (or IBM) support. Plenty of other 3rd party options, though...

Why?

- Apart from the “cool hack” factor...
 - What's in it for IBM?
 - What's in it for Sun?
 - What's in it for users?

What's in it for IBM

- New workload for IBM System z
 - With the success of the Linux initiative, “mainframe” is less of a dirty word
 - System z capacity increasing to level some previous argument about CPU-intensive workloads
 - Opens up new avenues for “Solaris shops” to push effective virtualization

What's in it for IBM?

- Demonstration of z/VM being “best of breed”
 - Sun domains para-virtualization strategy not working out to be particularly scalable
 - Recent cost structure changes to z/VM pricing leverage better per-virtual machine ROI
 - “Just one more” comment

“Why not? It’s just another virtual machine. We welcome any workload into the System Z family, we’re not picky.”

What's in it for IBM?

- Continue the Server Consolidation push
 - Makes Solaris workloads accessible for consolidation
 - Targets human workload as well as computational workload for better ROI

What's in it for Sun

- Re-enfranchise StorageTek customers
 - Counteract mixed message to STK customers wrt continued zSeries I/O and device development
- “Stop loss” for existing enterprise customers in non-NIC fields
 - Sun has steadily lost ground in large enterprise deployments over the last 4 years
- Foot in the door to new customers
 - Allows choice by superior manageability and operations concerns, not hardware platform

What's in it for Users

- Competition
 - Forces both IBM and Sun to concentrate on technical merit, not just price

What's in it for Users?

- Integrated consolidation strategy
 - Permits concentration to fewer platforms and management tooling
 - Simplified D/R
 - Reuse of:
 - Skill set
 - Procedures

What's in it for Users?

- Elimination of “religious” arguments:
 - Anti-Linux
 - Anti-Sun
 - Anti-Open Source

What's in it for Users?

- New tools for improved productivity
 - Availability of new application suites
 - Availability of desirable technology advances
 - Dtrace
 - System management enhancements
 - Printing system enhancements

Timeline

- 2006
 - Download OpenSolaris code
 - Spare time review of code
 - Build tools: gcc/binutils
 - Sun donates Sunblade
 - Get kernel build happening

Timeline

- 2007
 - Present progress at System z conference in Munich
 - Call with IBM execs
 - Meeting with interested parties in Somers
 - Meeting with Sun CTO and developers
 - Joint Sun/IBM announcement
 - Analyst conference call
 - Demo at Gartner Data Center Conference
 - Formal project begins Oct 2007

Timeline

- 2008
 - January delivery of working kernel, disk driver, libraries and userland commands
 - March delivery of network driver
 - April delivery of “fully functioning” system
 - SMF
 - gdb
 - gcc testsuite
 - perl
 - Extensive testing by dedicated IBM resource

YouTube Demo Video (Late November 2007)

- <http://www.youtube.com/watch?v=cH71qP-yDDI>
- 5 parts – shows the state of the world in November 2007
- Much, MUCH further along now, but it's handy to show to people who can't come here...

Development Team

- Neale Ferguson
 - Kernel and Integration
- Leland Lucius
 - I/O Subsystem and CCW Layer
 - Disk Driver
 - Network Driver
- Max Cohen
 - GCC and C/C++ Libraries
 - Dynamic loader
 - Libraries
- Adam Thornton
 - Device Drivers and Release Mgmt
- David Boyes
 - Documentation and Vendor Pacification
- Mary K. Holicky
 - Project Management

Code Base

- Current drop based on August 2008 (build 95) release
- Using the “mercurial” tool to keep current
- Native 390 target – no SPARC or Intel binaries (yet...)

Design Decisions...

- SNA Codename “Sirius”
- `_LP64` datamodel
 - 32-bit compatibility layer for kernel and some Sun utilities
- Architecture Level Set - - IBM System z9 Required
 - Fullword immediate instructions
 - Compare-swap-and-purge (CSP/CSPG) instruction
 - Long displacement (RY) instructions
 - Long relative displacement instructions
 - Load Page Table Entry instruction (LPTEA)
 - Purge DAT instruction (IDTE)
 - Cryptographic instructions

...Design Decisions...

- ABI is identical to Linux for IBM System z

- Assumes presence of z/VM
 - 5.3 base
 - DIAG interfaces:
 - Block I/O
 - Network I/O (VM64466)
 - PFAULT (soon)
 - I/O discovery (DIAG 210)
 - Memory discovery (DIAG 260)
 - VMDUMP
 - SALIPL
 - Co-operative Memory Management (later)

...Design Decisions

- I/O Layer similar to Linux CCW layer
- Separate address spaces for kernel and user processes
 - Allows for split code and data in separate address spaces to prevent buffer overwrite attacks
- Full 64-bit (16EB) address space
 - 3 levels of region table
 - Linux is 53 bit with most recent patch levels

Overall Porting Process

- Build cross-compilation environment
- Write some glue code to allow proper module construction.
- Build library support and process entry/exit code
- Build I/O model and device drivers
- dmake;dmake install...

Current Build Environment

- Initially done in cross-build environment on Sparc64
 - Sparc is “big endian”
 - “ON Build” tools: part of OpenSolaris
 - Ported a couple of tools for s390x support
- Switched to native build of tools/apps in 3rd drop
 - OpenSolaris build requires dmake so can’t yet self-host
 - GNU tools with new target of “ibm-s390x-solaris2”
 - GCC 4.2.3 with patches (important!)
 - Binutils very current (2.18.50 or later)
 - gdb-6.1.7

Major Development Areas

- PROM emulation routines
- Virtual memory support: “HAT layer”
- I/O support
 - Device detection and initialization
 - Adapter layer similar to Linux CCW device interface
 - DIAG 250 disk driver
 - Network driver
- Machine check handling/error management
- External interrupt handling
- Thread switching
- Syscall handling
 - Including 32-bit compatibility layer

Major Development Areas

- Libraries and loader
 - libc etc. part of OpenSolaris source tree
- gcc
 - New target `s390x-ibm-solaris2`
 - 4.3.2
 - Added `#pragma _init, _fini, and ident` support
- binutils 2.17.50 +
 - Added a couple of Sun extensions
- gdb 6.1.7

Patches to GNU Tools

```
519 binutils_2.18_s390_20080725.diff
2860 gcc_4.2.3_s390_20080725.diff
19289 gdb_6.7.1_s390_20080725.diff
158 gdb_bfd_only_s390_20080725.diff
```

Some Statistics

- 52909 source files in OpenSolaris tree
- 2240 files added
 - 1091 makefiles
 - 202 assembler (mostly syscall invocations)
 - 282 C
 - 418 headers
- 192 common files modified

Progress Made So Far

- Completed clean build of kernel, usr/lib, and user commands supplied with OpenSolaris source tree
- Server-oriented device drivers
 - Disk
 - Console
 - Network
- GNU compiler/debugger suite and libraries for C/C++ and other gcc-based languages
- Important open-source utilities (gmake, emacs, perl, python etc)
- Countless open-source servers, libraries and tools (Apache, more secure FTP server , ssh, etc)

Remaining Development Areas

- dtrace and mdb
- Port of Solaris linker to s390x - underway
 - Link process uses several options not supported with GCC
 - Build process adapted to GNU ld
- Additional applications and device drivers
 - Tape
 - Crypto acceleration hardware
- Java
- Bug fixes as people report them

Where to now?

- Improve support from Sun and IBM
- Get infrastructure in place
- Create “Sirius” community on OpenSolaris.org
- Open for public participation in port

Summary

- Putting OpenSolaris on IBM System z opens up a lot of interesting options for exploiting virtualization and existing Solaris knowledge in a really reliable environment.
- The porting process has been long, but is proving to be of great interest to IBM, Sun and others
- We hope to be able to make it available very soon.
- Questions?

Contact Info

David Boyes
Sine Nomine Associates
info@sinenomine.net
www.sinenomine.net