Porting Linux to the Power Macintosh

Paul Mackerras
Australian National University

Introduction to Power Macintosh family

- First generation: NuBus based, similar to 68k mac architecture
- Second generation (1995): PCI based
- PCI-based Power Macintosh features:
  - PowerPC processor
  - PCI bus
  - Hardware cache coherence
  - Descriptor-based DMA controller
  - Open Firmware

Open Firmware Device Tree

```
FF828FD8: /cpus@0
FF8290F0:   /PowerPC,603ev@0
FF829528:     /l2-cache@0,0
FF829D98: /memory@0
FF829EE0: /openprom@0
FF829FA0: /AAPL,ROM@FFC00000
FF82A1B8: /options@0
FF82AB40: /aliases@0
FF82AFE0: /packages@0
FF82B068:   /deblocker@0,0
FF82B790:   /disk-label@0,0
FF82BD08:   /obp-tftp@0,0
FF82DF78:   /mac-files@0,0
FF82E698:   /mac-parts@0,0
FF830860:   /iso-9660-files@0,0
FF8311C8:   /xcoff-loader@0,0
FF831B18: /bandit@F2000000
FF840510:   /pci106b,1@B
FF8409A8:     /chips65550@11
FF8458C8:     /ti1130@13
FF84A4C0:       /pccardffff,101@0
FF84AD20:     /ti1130@13,1
FF833108:     /ohare@10
FF833998:       /mesh@10000
FF8354C0:         /sd@0,0
FF836D20:       /escc@13000
FF836E78:         /ch-a@13020
FF8374C0:         /ch-b@13000
FF83A758:       /ata@20000
FF83BD88:         /ata-disk@0,0
FF83CEF8:       /via-pmu@16000
FF83EB80:         /adb@0,0
FF83EC70:           /keyboard@0,0
FF83F498:           /mouse@1,0
FF83FD68:         /power-mgt@0,0
FF83FE28:         /battery@0,0
```

```
0 > dev /bandit/ohare/mesh  ok
0 > .properties
name                    mesh
device_type             scsi
model                   AAPL,343S1146
reg                     00010000  00000100
                        00008000  00000200
clock-frequency         02FAF080
AAPL,interrupts         0000000C 00000000
AAPL,connector          scsi

 ok
0 > words
show-children   set-tid-lun     decode-unit     close
open            dma-sync        dma-map-out     dma-map-in
dma-free        dma-alloc       retry-cmd       error?
db-rem          do-cmd          reset-scsi      set-timeout
mesh-iooffset   mesh-dmaoffset

 ok
0 >
```

Open Firmware

1. Build device tree
2. Boot operating system
   - Load from device into memory and start execution
   - Operating system can call back into Open Firmware (client interface)
3. User interface
   - Forth interpreter
   - Somewhat unfriendly for new users

Bugs in Apple's implementations have been a problem.

PowerPC processor architecture

- Modern RISC instruction set architecture
  - Load/store architecture
  - Many registers (32 GPR, 32 FPR)
  - Fixed-length instructions (32 bits)
- Allows misaligned accesses, traps if cross page boundary
- MMU uses a hash table rather than a page table tree
- Caches are physically addressed and snoopy (although I-cache doesn't snoop D-cache)
- I/O is memory mapped
- Memory access may be reordered unless prevented by barrier instructions
- CPU is natively big-endian, can run in little-endian mode.

PowerPC-based machines

- PCI-based Power Macintosh
  - Apple's 2nd generation powermac, PCI bus
- NuBus-based Power Macintosh
  - Apple's 1st generation powermac, similar to 680x0 mac
- PReP (PowerPC Reference Platform) machines
  - PCI bus, PC-style peripherals
- CHRP (Common Hardware Reference Platform) machines
  - Combined PReP and PCI Power Mac technology
- APUS (Amiga Power-Up Systems)
  - PowerPC processor upgrade for Amiga systems
- VME single-board computers for embedded applications

History

1995–1996 : Gary Thomas ported 1.3.x and 2.0.x kernels to PReP.

early 1996 : Apple MkLinux group released MkLinux DR1 for NuBus powermacs.

July 1996 : I started porting Linux to my PowerMac 7500, based on Gary's 2.0.0 code.

August 1996 : I released a 2.0.0 kernel for the PowerMac 7500.

November 1996 : I released a 2.0.0 kernel for PowerMac 7500 and 7200, with MILO (Quik) hard disk boot system and Mac OS bootstrap installer.

December 1996 : Apple MkLinux group released support for PCI PowerMacs.

Initial PowerMac port

- Startup code
- Interrupt controller
- Open Firmware interface
- MMU initialization (use information from OF)
- Periodic interrupts (using PowerPC decrementer interrupts)
- Ethernet (MACE + DBDMA) driver
- SCSI (MESH + DBDMA) driver
- Macintosh partition table support
- CUDA interface
- ADB keyboard driver
- Powermac console driver (for “control” display initially)
- ADB mouse driver
- Serial port (Z8530 SCC) driver

Sources of information

- Allen Briggs' web page at `http://www.macbsd.com/openmac/refs.html`
- Open Firmware: device addresses and interrupt numbers
- "Macintosh Technology in CHRP": DBDMA, MESH SCSI
- MkLinux source code: MACE ethernet chip, CUDA interface, ADB keyboard and mouse
- Manufacturer's data sheets: MACE ethernet, SCC serial ports
- Open Firmware and related IEEE standards documents
- PowerPC 601 User's Manual

Progress

- Merged PowerMac and PReP ports
  - unified kernel binary for PowerMac, PReP and CHRP
- Optimized exception and MMU handling
- New models of PowerMac supported
  - including MP PowerMacs
- PPC port extended to various kinds of PowerPC-based machines
- PCI device drivers for Linux adapted to PowerPC
- Growth of user community (est. 100,000–500,000 systems)
- Growth of developer community

PowerPC Memory Management

The PowerPC MMU uses a hash table for storing PTEs rather than a tree.

The PowerPC MMU meets Linux

Linux/PPC uses a two-level tree to store PTEs.

- Can't guarantee that all PTEs will fit into the hash table.
- Linux memory management code assumes a tree structure.
- Hash table acts as a level-2 cache of PTEs.

MMU contexts implemented using consecutive VSID (virtual segment ID) values in the segment registers.

- Segments 0–11 set on MMU context switch (3GB user address space)
- Segments 12-15 map the kernel, not switched (1GB kernel address space)

Hash table must be flushed in TLB flush operations (or the task can be given a new MMU context).

Device drivers on the PowerMac

The PCI bus is little-endian by definition, while the PowerPC CPU is big-endian.

- `in[bwl]`, `out[bwl]` etc. macros include memory barrier (`eieio`) and byte-swapping.
- Data structures in memory which are accessed by DMA should be in little-endian format unless the device can be programmed to do the byte-swapping.
  - Use `cpu_to_le32` etc. macros.
- Byte-swapping is *not* necessary when transferring a buffer of bytes.
  - `insl_ns`, `outsl_ns` variants don't do byte-swapping

On PowerMacs, CPU will take a machine check exception if the kernel tries to access an I/O port where there is no device.

- Machine check is an asynchronous, imprecise, nonrecoverable exception.
- Device probing code must be disabled on PowerMacs
- PCI configuration space and OF device tree provide enough information that probing should be unnecessary.

Future work

- Better support for recent machines (especially PowerBooks)
- Support new machines as they come out (G4)
- Participate in USB driver development
- Resolve keycode issue: ADB vs. USB vs. PC keycodes

Performance

L M B E N C H 1 . 9 S U M M A R Y

(Alpha software, do not distribute)

Processor, Processes - times in microseconds - smaller is better

Host	OS	Mhz	null call	null I/O	stat	open clos	selct	sig inst	sig hndl	fork proc	exec proc	sh proc
PMac G3	Linux 2.1.129	267	1.0	1.4	12	15	0.08K	1.7	6	0.9K	9K	27K
Pentium II	Linux 2.0.34	401	1.3	1.7	14	15	0.05K	2.3	4	0.4K	2K	9K
PMac 7600	Linux 2.1.129	200	2.0	2.6	17	21	0.10K	3.3	11	2.3K	19K	61K
RS/6000	AIX 2	233	1.1	6.4	32	38	0.66K	3.2	21	3.4K	6K	18K

Context switching - times in microseconds - smaller is better

Host	OS	2p/0K ctxsw	2p/16K ctxsw	2p/64K ctxsw	8p/16K ctxsw	8p/64K ctxsw	16p/16K ctxsw	16p/64K ctxsw
PMac G3	Linux 2.1.129	1	8	134	31	322	61	413
Pentium II	Linux 2.0.34	2	18	53	19	132	28	166
PMac 7600	Linux 2.1.129	1	138	808	220	992	251	1007
RS/6000	AIX 2	11	81	490	148	693	179	845

Local Communication latencies in microseconds - smaller is better

Host	OS	2p/0K ctxsw	Pipe	AF UNIX	UDP	RPC/ UDP	TCP	RPC/ TCP	TCP conn
PMac G3	Linux 2.1.129	1	11	27	33		57		
Pentium II	Linux 2.0.34	2	12	19	47		74		228
PMac 7600	Linux 2.1.129	1	23	51	76		195		732
RS/6000	AIX 2	11	61	81	172		219		1141

File & VM system latencies in microseconds - smaller is better

Host	OS	0K File Create	0K File Delete	10K File Create	10K File Delete	Mmap Latency	Prot Fault	Page Fault
PMac G3	Linux 2.1.129	20	3	45	5	6796	2	0.1K
Pentium II	Linux 2.0.34	34	2	47	3	23623	1	0.1K
PMac 7600	Linux 2.1.129	30	6	209	320	17393	3	0.3K
RS/6000	AIX 2	1000	961	961	970	2188	9	6.8K

Local Communication bandwidths in MB/s - bigger is better

Host	OS	Pipe	AF UNIX	TCP	File reread	Mmap reread	Bcopy (libc)	Bcopy (hand)	Mem read	Mem write
PMac G3	Linux 2.1.129	146	74		77	136	57	58	136	100
Pentium II	Linux 2.0.34	254	108	72	148	324	136	142	324	171

```
PMac 7600 Linux 2.1.129   49   18   13      24      60      25      27   60     36
RS/6000           AIX 2   33   38   24      37      65      30      32   65     50

Memory latencies in nanoseconds - smaller is better
    (WARNING - may not be correct, check graphs)
---------------------------------------------------
Host                 OS   Mhz  L1 $   L2 $    Main mem    Guesses
--------- -------------   ---  ----   ----    --------    -------
PMac G3   Linux 2.1.129   267     7     79         236
Pentium II Linux 2.0.34   401     7     49         152
PMac 7600 Linux 2.1.129   200    10    293         386
RS/6000           AIX 2   233     8    150         372
```

Conclusion

- Linux runs efficiently and well on Power Macintoshes.
- Active and growing user and developer communities.
- Device tree helps kernel adapt to variations between models without resorting to probing.
- Information from Open Firmware made the porting effort much easier than for the 680x0-based Macintoshes.

Acknowledgements

- Linus Torvalds and the many many other Linux developers, for Linux.
- Gary Thomas, who did the initial port of Linux to the PowerPC processor.
- Cort Dougan, who has put a lot of effort into developing and supporting Linux/PPC, particular for PReP, CHRP and embedded systems.
- Many people who have worked on Linux for PPC, including Benjamin Herrenschmidt, Geert Uytterhoeven, Randy Gobbel, Anthony Tong, Anthony Sun, Dan Jacobowitz, Harry Eaton, Tom Rini, Mark Hatle, Franz Sirl, Matt McLean, Troy Benjegerdes, Jack Howarth, Kevin Hendricks, and others (apologies to those I've missed).