Architecture of Scalable Operating Systems: Multikernel

Rasmus Pfeiffer

Friedrich-Alexander-Universität Erlangen-Nürnberg

6. Dezember 2016

Table of Contents

Shared Memory vs Message Passing

Inter-Process Communication

Multikernel Model

Barrelfish

Description

Shared Memory uses data structures at well known places in memory to communicate between CPU cores.

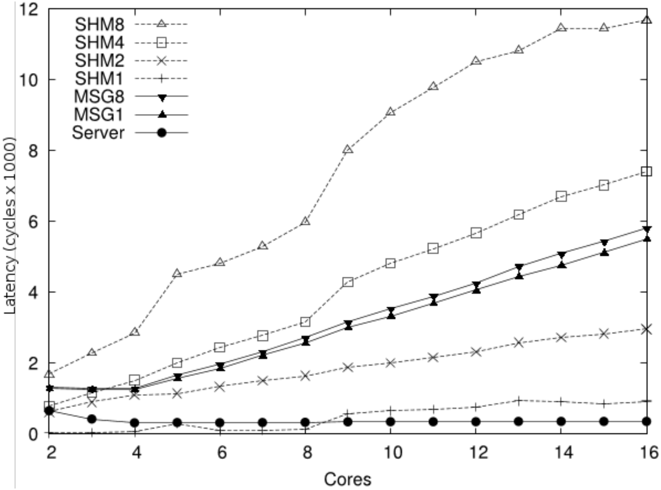
Message Passing uses explicit messages to communicate between CPU cores.

History

Shared Memory and Message Passing are duals [3]

In 1978 Lauer and Needhalm argued, that it depends on the hardware, if shared memory or message passing is faster.

Current Situation



Comparison of the cost of updating shared state using shared memory and message passing [1]

Kernel-Based IPC

Kernel-based inter-process communication (IPC) is limited by the cost of invoking the kernel and reallocating a processor from one address space to another [2].

User-Space Remote Procedure Call (URPC)

- Messages are sent directly between address spaces.
- Unnecessary processor reallocation between address spaces is eliminated.
- When processor reallocation is needed, the overhead is reduced.

URPC - Assumptions

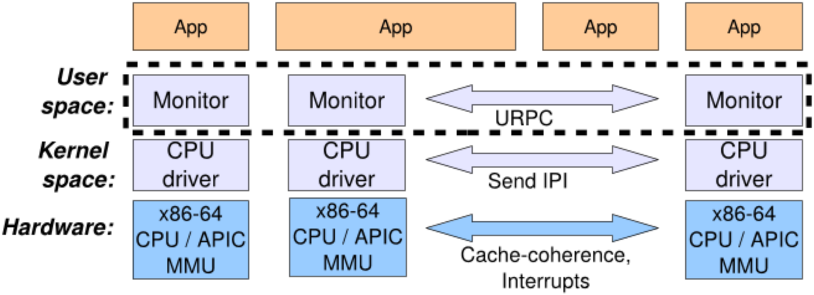
- Client has other work to do

- The server has, or will have, a CPU core available.

Multikernel Model

1. Make all inter-core communication explicit.
2. Make OS structure Hardware-neutral.
3. View state as replicated instead of shared.

Barrelfish structure



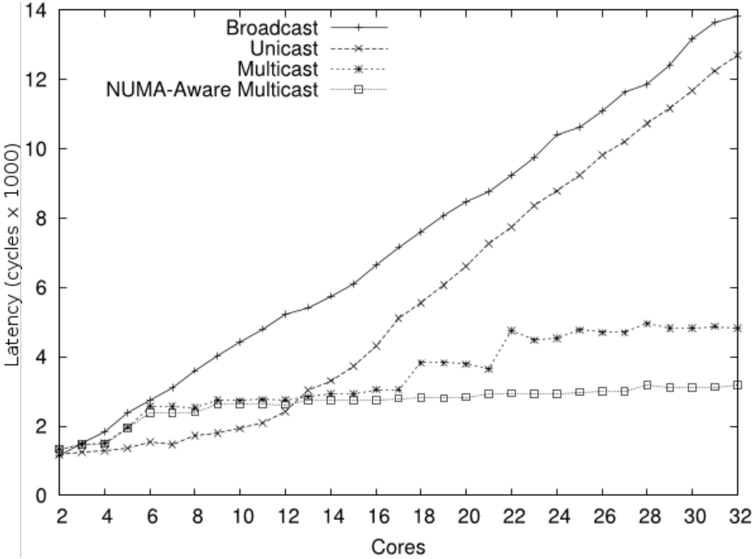
CPU driver

- single threaded
- controls: APIC, MMU, etc
- shares no state with other cores
- specialized for CPU architecture

Monitors

- processor-agnostic
- manages system-wide state

Inter-Core Communication



Process structure

- processes consist of dispatcher objects

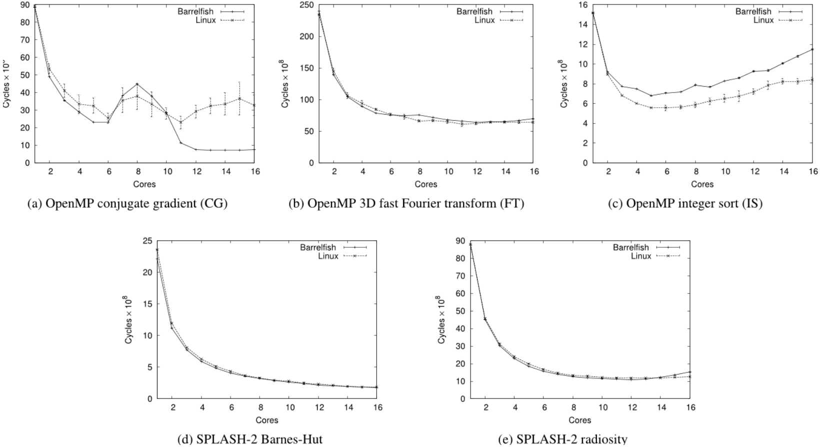
- dispatcher objects are scheduled by CPU driver

Memory Management

Memory management is performed explicit in user level:

1. acquire memory for page table
2. insert page table in root page table
3. acquire more memory and insert in page table

Performance I - compute-bound workloads



(a) OpenMP conjugate gradient (CG)

(b) OpenMP 3D fast Fourier transform (FT)

(c) OpenMP integer sort (IS)

(d) SPLASH-2 Barnes-Hut

(e) SPLASH-2 radiosity

Performance II - IO workloads

Webserver:

- static content
  - Linux: 8924 requests per second
  - Barrelfish: 18697 requests per second
- dynamic content
  - 3417 requests per second

Questions

ANY QUESTIONS?

References

- A. Baumann, P. Barham, P.-E. Dagand, T. Harris, R. Isaacs, S. Peter, T. Roscoe, A. Schüpbach, and A. Singhania. The multikernel: a new os architecture for scalable multicore systems. In *Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles*, pages 29–44. ACM, 2009.

- B. N. Bershad, T. E. Anderson, E. D. Lazowska, and H. M. Levy. User-level interprocess communication for shared memory multiprocessors. *ACM Transactions on Computer Systems (TOCS)*, 9(2):175–198, 1991.

- H. C. Lauer and R. M. Needham. On the duality of operating system structures. *ACM SIGOPS Operating Systems Review*, 13(2):3–19, 1979.