Ecole polytechnique fédérale de Lausanne

EDIC Seminar

Prof. E. Bugnion
About me

18y  14y

Academia

Industry
The big open problems according to Google

The Case for Energy-Plus Computing

Software techniques that tolerate latency variability are vital to building responsive large-scale Web services.

BY JEFFREY DEAN AND LUÍZ ANDRÉ BARROSO

Microsecond-scale I/O means tension between performance and productivity that will need new latency-mitigating ideas, including in hardware.

BY LUÍZ BARROSO, MIKE MARTY, DAVID PATTERSON, AND PARTHASARATHY RANGANATHAN

The Tail at Scale

At least according to L.Barroso

Attack of the Killer Microseconds
Thoughts from the IX Project
[OSDI 14, SoCC 15, TOCS 17]

George Prekas, Mia Primorac, Marios Kogias,
Sam Grossman, Ana Klimovic,
Adam Belay, Christos Kozyrakis,
Edouard Bugnion

Stanford
IX main contributions [OSDI 14]

#1: Protection and direct HW access through virtualization

#2: Execution model for low latency and high throughput
Separation of Control and Data Plane

Virtualization HW provides protection
Design (1): Run to Completion

Improves D$ locality; eliminates schedule
Design (2): Adaptive Batching

Improves I$ locality; prefetching; bounds latency
Design (3): Flow-consistent hashing

- Uses RSS to separate flow groups onto queues
- Coherence-free and synchronization free execution
- API follows scalable commutativity rule [Clements ‘13]

Multicore scalability by design
Where’s the catch?

In systems, sweeping simplifications:
• Enable massive gains
• Nearly always introduce hidden tradeoffs

**Tradeoff #1:** dataplane polling model limits system efficiency
Latency-critical applications

- Memcached, Facebook USR workload, 2752 connections
- Server: Sandy Bridge @ 2.4 Ghz, 8 cores / 16 HTs, Intel x520
Power vs. Throughput

Throughput (MRPS)

0 1 2 3 4 5 6 7

Power (W)

0 20 40 60 80 100 120

Linux max
Power vs. Throughput

Linux max

IX max
Power vs. Throughput – across static configurations

- **Linux max**: AVG -35% on Linux
- **IX max**: AVG -49% on IX
Dynamic resource control in the control plane

Increase the frequency

Host
Kernel

Userspace

Dataplane

Latency critical workload

IX

Host

CP
Policy

DVFS
Key challenges

1. Which resources to add/remove?
   • inferred by Pareto analysis

2. When to add/remove resources?
   • Need to design a stable control loop
   • Different triggers to add/remove resources

3. How to add/remove cores
   • Fast, TCP-friendly rebalancing mechanism
#1: Resource Adjustment Policies

![Diagram showing the relationship between Power (W) and different policies: +cores, +HT, +DVFS, +Turbo.](image)
Evaluation – step pattern

Achieved MRPS

Adequate compliance: violations ~ 1 second
Lots of collaborators

Adam Belay
Ana Klimovic
Christos Kozyrakis
George Prekas
Jim Larus
Jonas Fietz
Katerina Argyraki
Maggy Church
Marios Kogias
Mia Primorac
Sam Grossman
Sam Whitlock
Stanko Novakovic
Stuart Byma