Energy Management in Virtualized Environments

Gaurav Dhiman, Giacomo Marchetti, Raid Ayoub, Tajana Simunic Rosing (CSE-UCSD)

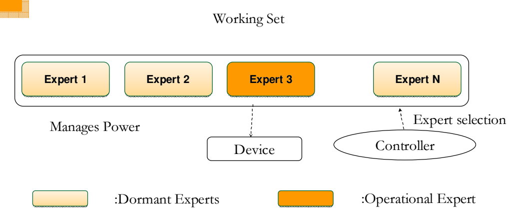
Motivations and Goals

- ✓ Lower datacenter energy consumption
- ✓ Handle non-stationary workloads
- ✓ Service - VM - Customization
- ✓ Virtual Machine Power Oriented Scheduling
- ✓ Workload migration across physical machine
- ✓ Minimize impact on performance

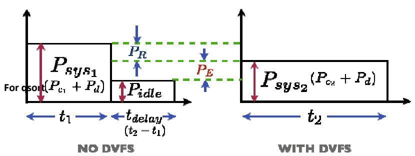
OS implementation and Results

Online Learning Algorithm

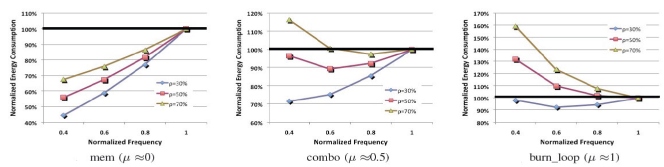
- ✓ Performs dynamic evaluation of a set of DPM and DVFS policies at run time and selects the best suited for the current workload
- ✓ Guarantees convergence and performance close to that of the best available policy in the set



DVFS

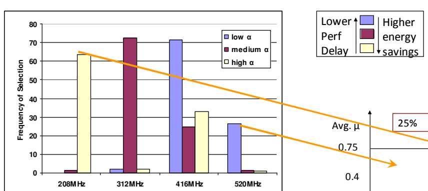


- ✓ CPU intensive (μ ->1) vs Memory intensive (μ -> 0)
  - ✓ μ = measure of CPU intensiveness
- ✓ Leakage impact (ρ)

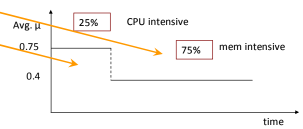


Experimental Setup

- ✓ Workloads: qsort, djpeg, blowfish, dgzip
- ✓ CPU Xscale



Identifies both CPU-intensive and memory intensive phases correctly



Benchmarks	Low α		Med α		High α	
	%delay	%energy	%delay	%energy	%delay	%energy
qsort+djpeg	6	17	15	33	25	41
dgzip+djpeg	13	24	19	40	27	49
qsort+dgzip	7	20	18	35	26	42
dgzip+bf	11	17	20	31	26	45

Multi-Tasking

Energy Saving/Performance Delay Results for CPU

DPM

Device	Trace Name	t_{RI}	σ t_{RI}
HDD	HP-1Trace	20.5	29
	HP-2 Trace	5.9	8.4
	HP-3 Trace	17.2	2

t_{RI} : Average Request Inter-arrival Time (in sec)

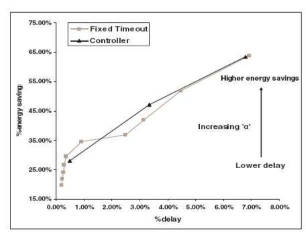
Expert	Characteristics
Fixed Timeout	Timeout = $7*T_{be}$
Adaptive Timeout (Douglis, USENIX'95)	Initial timeout = $7*T_{be}$; Adjustment = $+1T_{be}/-1T_{be}$
Exponential Predictive (Hwang, ICCAD'97)	$I_{n+1} = a\, i_n + (1-a).I_n$ with a = 0.5
TISMDP (Simunic, TCAD'01)	Optimized for delay constraint of 3.5% on HP-1 trace

DPM: With Individual Experts

Policy	HP1 Trace		HP2 Trace		HP3 Trace	
	%delay	%energy	%delay	%energy	%delay	%energy
Oracle	0	68.17	0	65.9	0	71.2
Timeout	4.2	49.9	4.4	46.9	3.3	55
Ad Timeout	7.7	66.3	8.7	64.7	6	67.7
TISMDP	3.4	44.8	2.26	36.7	1.8	42.3
Predictive	8	66.6	9.2	65.2	6.5	68

DPM: With Online Learning

Preference	HP-1 Trace		HP-2 Trace		HP-3 Trace	
	%delay	%energy	%delay	%energy	%delay	%energy
Low delay	3.5	45	2.61	37.41	2.55	49.5
↓	6.13	60.64	5.86	54.2	4.36	61.02
High energy savings	7.68	65.5	8.59	64.1	5.69	66.28

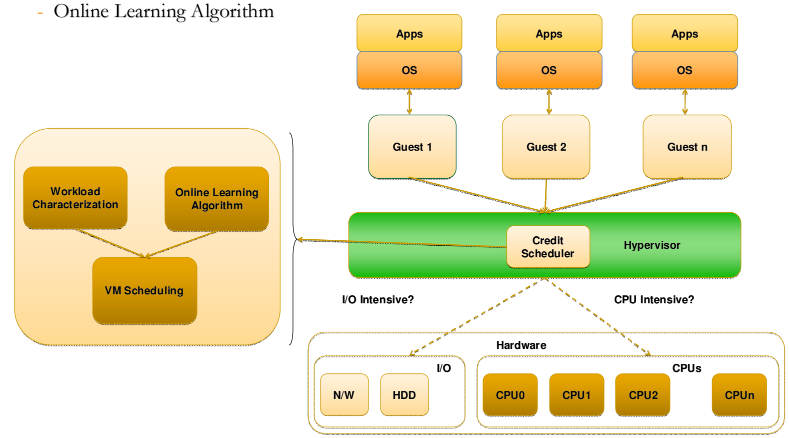


Power/Performance Results for HDD HP-1 trace
Comparison with fixed timeout experts

Inside Xen Hypervisor

Virtualization

- ✓ **Energy Oriented Scheduler**
  - Implements a scheduler capable of adapting to workload (guest) characteristics
  - Migration: Guest balancing and clustering
  - Co-locate guests to free up resources
  - Online Learning Algorithm
- ✓ **Workload characterization**
  - I/O Intensiveness: Maintain metrics for I/O accesses per guest
  - CPU Intensiveness: Use CPU performance counters



DPM & DVFS

Policy	Description
PM-1	switch CPU to ACPI state C1 (remove clock supply) and move to lowest voltage setting
PM-2	switch CPU to ACPI state C6 (remove power)
PM-3	switch CPU to ACPI state C6 and switch the memory to self- refresh mode

Experimental Setup

- ✓ AMD quad core CPU
- ✓ SPEC benchmarks

Benchmark	Freq	%delay	$\%Energy_{savingsPM-i}$		
			PM-1	PM-2	PM-3
mcf	1.9	29	5.2	0.7	-0.5
	1.4	63	8.1	0.1	-2.1
	0.8	163	8.1	-6.3	-10.7
bzip2	1.9	37	4.7	-0.6	-2.1
	1.4	86	7.4	-2.4	-5
	0.8	223	7.8	-9.0	-14
art	1.9	32	6	1	-0.1
	1.4	76	7.3	-1.7	-4
	0.8	202	8	-8	-13
sixtrack	1.9	37	5	-0.5	-2
	1.4	86	6	-4.3	-7.2
	0.8	227	7	-11	-16.1

Recent CPUs might perform better with a "run to sleep" policy due to:

- ✓ Improved CPU efficiency
- ✓ Idle power management support

Summary

Hypervisor VM scheduler implementation

- ➢ Power Management: DPM/DVFS
- ➢ Workload characterization aware
- ➢ Adaptive Behavior

Supported by NSF-GreenLight project, CNS, Sun Microsystems, UC Micro, Cisco, GSRC/DARPA