The Xen Credit CPU Scheduler

Emmanuel Ackaouy
ack@xensource.com
Goals

• Maintainability
  – Simple & Fast

• Zero-admin **good default behavior**
  – **Fairness**: Weight based
  – **I/O versus CPU** intensive VCPUs
  – Tuning via per-VM *weight & cap*

• Better **SMP support**
  – **Work conserving**
  – Transparent cross **CPU migration**
  – **System wide** accounting & fairness
Round Robin

CPU Usage

VCPU1  VCPU2  VCPU3  VCPU4

VCPU 1.0
VCPU 2.0
VCPU 3.0
VCPU 4.0
Priorities!

Run Queue

CPU0

VCPU I/O

VCPU SPIN

Active Domains

Host

DOM SPIN

DOM I/O

VCPU I/O

Priority UNDER

DOM I/O

Credit 30

Cap 0
CPU Usage

CPU0
- VCPU 1.0
- VCPU 2.0

CPU1
- VCPU 3.0

VCPU1
VCPU2
VCPU3
Run Queues

CPU0
  VCPU 1.0
  VCPU 2.0

CPU1
  VCPU 3.0

Active Domains

Host
  DOM1
  DOM2
  DOM3

DOM4
  Credit: -10
  Cap: 0
  Priority: OVER

VCPU 3.0
  Credit: -10
  Cap: 0
  Priority: OVER
Self Balancing!

CPU Usage

VCPU1 VCPU2 VCPU3

CPU0

VCPU 1.0

VCPU 2.0

VCPU 3.0

CPU1

VCPU 1.0

VCPU 2.0

VCPU 3.0
Global Accounting

• Periodically recompute priorities for all active VCPUs
  – **UNDER** (below fair share usage)
  – **OVER** (above fair share)
• VCPU fair shares derived from VM weight and cap in proportion to all active VMs
• Accounting cost
  – A function of # active VCPUs in system
  – **NOT** a function of # scheduling decisions
Local CPU Scheduling

- Per-CPU sorted runq
  - 2 VCPU priorities: OVER and UNDER fair share
- Clock tick
  - Charge running VCPU
- Scheduling decision
  - End of time slice or VCPU blocks
  - Common case: run next local VCPU under fair share
  - Otherwise: pick queued remote under fair share VCPU, local over VCPU, or remote over VCPU (in that order)
- Scheduling cost
  - Common case: Pick next VCPU on local runq
Credit Scheduler Internals

Run Queues

CPU0
  VCPU 1.0
  VCPU 2.0
  VCPU 3.0
  VCPU 0.0

CPU1
  VCPU 1.1
  VCPU 2.1
  VCPU 3.1

CPU2
  VCPU 1.2
  VCPU 2.2
  VCPU 3.2

CPU3
  VCPU 1.3
  VCPU 2.3
  VCPU 3.3
  VCPU 4.1
  VCPU 4.0

VCPU 4.1
Priority OVER

DOM4
Credit -10
Cap 0

Active Domains

Host
DOM0 DOM1 DOM2 DOM3 DOM4