Introduction au Power Management dans Linux

Implémentation, Utilisation et Benchmark

maxime.chevallier@smile.fr

21 mars 2017

Smile

OPEN SOURCE SOLUTIONS

Plan

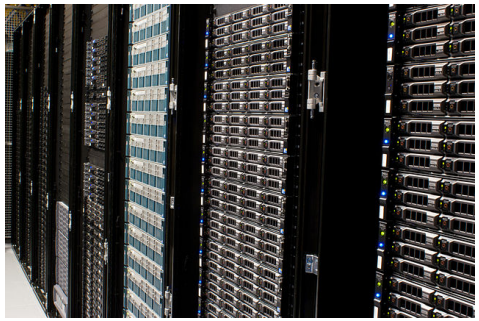
Smile
OPEN SOURCE SOLUTIONS

Enjeux

- N'utiliser que les ressources nécessaires
- Etre générique (ACPI, APM, SCPI)
- Respecter les contraintes utilisateur
- Rester transparent

Enjeux

- N'utiliser que les ressources nécessaires
- Etre générique (ACPI, APM, SCPI)
- Respecter les contraintes utilisateur
- Rester transparent

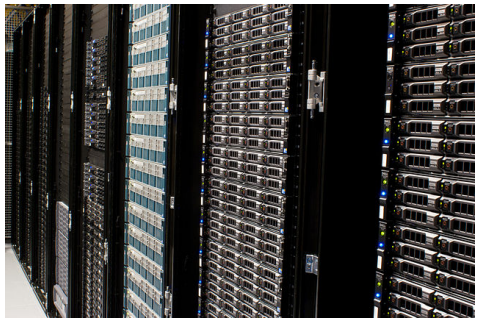


x86, milliers d'unités

Power Management sous linux

Enjeux

- N'utiliser que les ressources nécessaires
- Etre générique (ACPI, APM, SCPI)
- Respecter les contraintes utilisateur
- Rester transparent



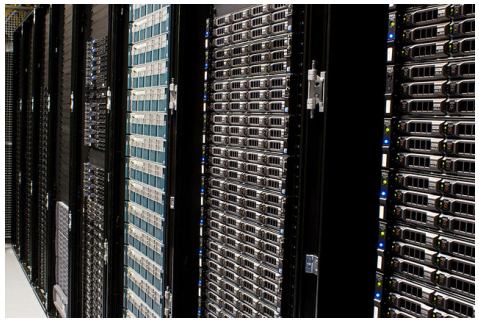
x86, milliers d'unités

x86, fléxibilité

Smile
OPEN SOURCE SOLUTIONS

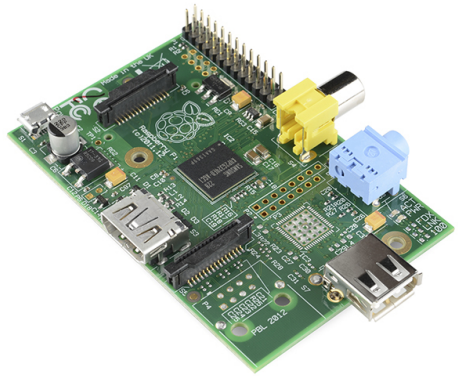
Enjeux

- N'utiliser que les ressources nécessaires
- Etre générique (ACPI, APM, SCPI)
- Respecter les contraintes utilisateur
- Rester transparent



x86, milliers d'unités

x86, fléxibilité



ARM, autonomie

Plan

Smile
OPEN SOURCE SOLUTIONS

Power Management Dynamique

Minimiser la consommation d'un système actif

Compromis

- Ressources utilisées
- Ressources nécessaires
- Latences acceptables
- Consommation actuelle
- Température actuelle

PM pour les périphériques

PM core

- API pour drivers
- Interface sysfs
- PM dynamique
- Modes d'endormissement

Documentation : `Documentation/power`
Headers : `include/linux/pm.h`
Implémentation : `kernel/power/`

Smile
OPEN SOURCE SOLUTIONS

runtime_pm

Etats

- `active` : Est capable d'I/O
- `suspended` : Pas d'I/O

Callbacks

```
runtime_suspend(dev)
runtime_resume(dev)
runtime_idle(dev)
```

Helpers

```
pm_runtime_*
pm_request_*
pm_schedule_*
pm_runtime_{get,put}*
pm_*_autosuspend
```

Smile
OPEN SOURCE SOLUTIONS

Quality of Service

Indiquer au noyau les latence et débits à respecter

Paramètres globaux

- `cpu_dma_latency` (µs)
- `memory_bandwidth` (mbps)
- `network_latency` (µs)
- `network_throughput` (kbps)

Interface Userspace : `/dev/*` + sysfs

Smile
OPEN SOURCE SOLUTIONS

CPU Idle

Que faire quand le CPU n'a rien à faire ?

- Choix du mode (governor)
  - `select()`
  - `reflect()`
- Implémentation (driver)

Smile
OPEN SOURCE SOLUTIONS

CPU Idle

Que faire quand le CPU n'a rien à faire ?

- Choix du mode (governor)
  - `select()`
  - `reflect()`
- Implémentation (driver)

struct cpuidle_state

- `exit_latency`
- `power_usage`
- `target_residency`
- `int enter([...], int index)`

Smile
OPEN SOURCE SOLUTIONS

CPU Idle

Que faire quand le CPU n'a rien à faire ?

- Choix du mode (governor)
  - `select()`
  - `reflect()`
- Implémentation (driver)

struct cpuidle_state

- `exit_latency`
- `power_usage`
- `target_residency`
- `int enter([...], int index)`

`powertop`, `/sys/devices/system/cpu/cpuX/cpuidle`

Smile
OPEN SOURCE SOLUTIONS

Driver intel_idle

i7 4702MQ

name	latency	residency	utilisation
C0	-	-	1.5%
POLL	0	0	0.2%
C1-HSW	2	2	0.6%
C1E-HSW	10	20	0.2%
C3-HSW	33	100	0.0%
C6-HSW	133	400	0.0%
C7s-HSW	166	500	97.4%

Driver ACPI `processor_idle`

i5 6500

```
name      latency  residency
C0        -        -
POLL      0        0
C1        1        2
C2        151      302
C3        256      512
```

Smile
OPEN SOURCE SOLUTIONS

DVFS

Dynamic **V**oltage and **F**requency **S**caling

Smile
OPEN SOURCE SOLUTIONS

DVFS

Dynamic **V**oltage and **F**requency **S**caling

cpufreq

`/sys/devices/system/cpu/cpuX/cpufreq/`

Smile
OPEN SOURCE SOLUTIONS

DVFS

Dynamic **V**oltage and **F**requency **S**caling

cpufreq

- Implémentation hardware (policy)

`/sys/devices/system/cpu/cpuX/cpufreq/`

Smile
OPEN SOURCE SOLUTIONS

DVFS

Dynamic **V**oltage and **F**requency **S**caling

cpufreq

- Implémentation hardware (policy)
- Implémentation software (governor) :

`/sys/devices/system/cpu/cpuX/cpufreq/`

Smile
OPEN SOURCE SOLUTIONS

DVFS

Dynamic **V**oltage and **F**requency **S**caling

cpufreq

- Implémentation hardware (policy)
- Implémentation software (governor) :
  - `performance`

`/sys/devices/system/cpu/cpuX/cpufreq/`

Smile
OPEN SOURCE SOLUTIONS

DVFS

Dynamic **V**oltage and **F**requency **S**caling

cpufreq

- Implémentation hardware (policy)
- Implémentation software (governor) :
  - `performance`
  - `powersave`

`/sys/devices/system/cpu/cpuX/cpufreq/`

Smile
OPEN SOURCE SOLUTIONS

DVFS

Dynamic **V**oltage and **F**requency **S**caling

cpufreq

- Implémentation hardware (policy)
- Implémentation software (governor) :
  - `performance`
  - `powersave`
  - `userspace`

`/sys/devices/system/cpu/cpuX/cpufreq/`

Smile
OPEN SOURCE SOLUTIONS

DVFS

Dynamic **V**oltage and **F**requency **S**caling

cpufreq

- Implémentation hardware (policy)
- Implémentation software (governor) :
  - `performance`
  - `powersave`
  - `userspace`
  - `ondemand`

`/sys/devices/system/cpu/cpuX/cpufreq/`

Smile
OPEN SOURCE SOLUTIONS

DVFS

Dynamic **V**oltage and **F**requency **S**caling

cpufreq

- Implémentation hardware (policy)
- Implémentation software (governor) :
  - `performance`
  - `powersave`
  - `userspace`
  - `ondemand`
  - `conservative`

`/sys/devices/system/cpu/cpuX/cpufreq/`

Smile
OPEN SOURCE SOLUTIONS

DVFS

Dynamic **V**oltage and **F**requency **S**caling

cpufreq

- Implémentation hardware (policy)
- Implémentation software (governor) :
  - `performance`
  - `powersave`
  - `userspace`
  - `ondemand`
  - `conservative`
  - `schedutil` (linux 4.6)

`/sys/devices/system/cpu/cpuX/cpufreq/`

Smile
OPEN SOURCE SOLUTIONS

DVFS

Dynamic **V**oltage and **F**requency **S**caling

cpufreq

- Implémentation hardware (policy)
- Implémentation software (governor) :
  - `performance`
  - `powersave`
  - `userspace`
  - `ondemand`
  - `conservative`
  - `schedutil` (linux 4.6)

`/sys/devices/system/cpu/cpuX/cpufreq/`

devfreq

Similaire pour les devices non-CPU

ile
OPEN SOURCE SOLUTIONS

Operating Performance Points

Tuples (Fréquence, Tension) pour un périphérique

```
operating-points = <
 /* kHz uV */
 792000 1100000
 396000 950000
 198000 850000
 >;
```

Smile
OPEN SOURCE SOLUTIONS

Actions en fonction de la température

Thermal zone

Actions en fonction de la température

Thermal zone

- Température (`trip_point`)

Actions en fonction de la température

Thermal zone

- Température (`trip_point`)
- Politique :

Actions en fonction de la température

Thermal zone

- Température (`trip_point`)
- Politique :
  - `step_wise`

Smile
OPEN SOURCE SOLUTIONS

Actions en fonction de la température

Thermal zone

- Température (`trip_point`)
- Politique :
  - `step_wise`
  - `fair_share`

Smile
OPEN SOURCE SOLUTIONS

thermal

Actions en fonction de la température

Thermal zone

- Température (`trip_point`)
- Politique :
  - `step_wise`
  - `fair_share`
  - `userspace`

Smile
OPEN SOURCE SOLUTIONS

thermal

Actions en fonction de la température

Thermal zone

- Température (`trip_point`)
- Politique :
  - `step_wise`
  - `fair_share`
  - `userspace`
- Cooling device

Smile
OPEN SOURCE SOLUTIONS

thermal

Actions en fonction de la température

Thermal zone

- Température (`trip_point`)
- Politique :
  - `step_wise`
  - `fair_share`
  - `userspace`
- Cooling device

Cooling device

- Hardware : Ventilateur
- Software : cpufreq

Smile
OPEN SOURCE SOLUTIONS

thermal

Actions en fonction de la température

Thermal zone

- Température (`trip_point`)
- Politique :
  - `step_wise`
  - `fair_share`
  - `userspace`
- Cooling device

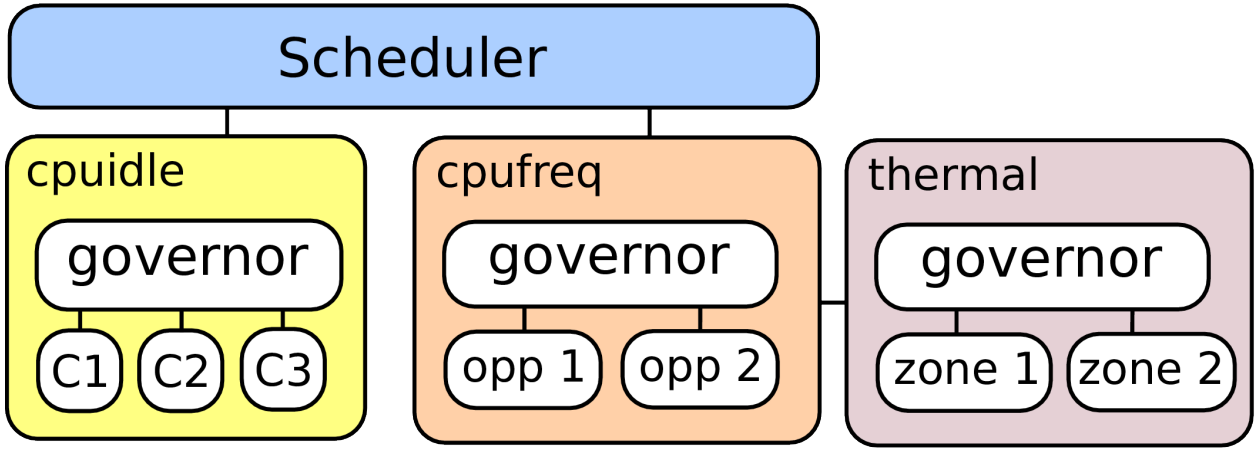
Cooling device

- Hardware : Ventilateur
- Software : cpufreq

`/sys/class/thermal/`

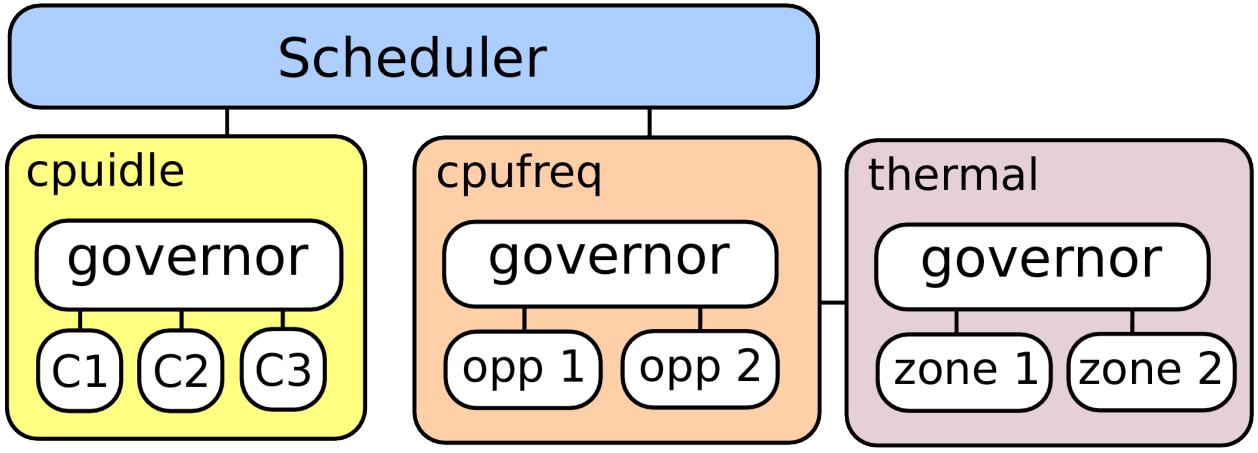
Smile
OPEN SOURCE SOLUTIONS

Vue d'ensemble



Smile
OPEN SOURCE SOLUTIONS

Vue d'ensemble



A venir

- Unifier cpuidle et cpufreq
- Energy Aware Scheduler

Smile
OPEN SOURCE SOLUTIONS

Plan

Smile
OPEN SOURCE SOLUTIONS

PM core

struct dev_pm_ops

Ensemble de callbacks :

```
prepare()     freeze()
complete()    thaw()
suspend()     poweroff()
resume()      restore()
```

Smile
OPEN SOURCE SOLUTIONS

PM core

struct dev_pm_ops

Ensemble de callbacks :

```
prepare()     freeze()
complete()    thaw()
suspend()     poweroff()
resume()      restore()
```

Wakeup

- `enable_irq_wake()`
- `disable_irq_wake()`

Smile
OPEN SOURCE SOLUTIONS

Modes d'endormissement

ACPI State : S1

`freeze > /sys/power/state`

Suspend to Idle

- Entièrement software
- Freeze userspace
- Périphériques lowpower
- Toujours supporté

Réveil en quelques millisecondes

Smile
OPEN SOURCE SOLUTIONS

Modes d'endormissement

ACPI State : S2

`standby > /sys/power/state`

Standby

- Suspend to Idle +
- Coupure des coeurs non-boot
- Coupure de certains composants bas niveau
- Support dépendant de la plateforme

Réveil en quelques millisecondes

Smile
OPEN SOURCE SOLUTIONS

Modes d'endormissement

ACPI State : S3

`mem > /sys/power/state`

Suspend to RAM

- Standby +
- Périphériques en lowpower
- CPU en lowpower
- RAM en auto-rafraichissement
- Support dépendant de la plateforme

Réveil en quelques centaines de millisecondes

Smile
OPEN SOURCE SOLUTIONS

Modes d'endormissement

ACPI State : S4

`disk > /sys/power/state`

Suspend to disk

- Image mémoire persistée
- Système en lowpower, voire éteint

Smile
OPEN SOURCE SOLUTIONS

Modes d'endormissement

ACPI State : S4

`disk > /sys/power/state`

Suspend to disk

- Image mémoire persistée
- Système en lowpower, voire éteint

parametres

`/sys/power/disk`

- `platform`
- `shutdown`
- `reboot`
- `suspend`

Smile
OPEN SOURCE SOLUTIONS

Modes d'endormissement

ACPI State : S4

`disk > /sys/power/state`

Suspend to disk

- Image mémoire persistée
- Système en lowpower, voire éteint

parametres

`/sys/power/disk`

- `platform`
- `shutdown`
- `reboot`
- `suspend`

Processus

Smile
OPEN SOURCE SOLUTIONS

Modes d'endormissement

ACPI State : S4

`disk > /sys/power/state`

Suspend to disk

- Image mémoire persistée
- Système en lowpower, voire éteint

parametres

`/sys/power/disk`

- `platform`
- `shutdown`
- `reboot`
- `suspend`

Processus

1. Suspend

Smile
OPEN SOURCE SOLUTIONS

Modes d'endormissement

ACPI State : S4

`disk > /sys/power/state`

Suspend to disk

- Image mémoire persistée
- Système en lowpower, voire éteint

parametres

`/sys/power/disk`

- `platform`
- `shutdown`
- `reboot`
- `suspend`

Processus

1. Suspend
2. Snapshot

Smile
OPEN SOURCE SOLUTIONS

Modes d'endormissement

ACPI State : S4

`disk > /sys/power/state`

Suspend to disk

- Image mémoire persistée
- Système en lowpower, voire éteint

parametres

`/sys/power/disk`

- `platform`
- `shutdown`
- `reboot`
- `suspend`

Processus

1. Suspend
2. Snapshot
3. Resume

Smile
OPEN SOURCE SOLUTIONS

Modes d'endormissement

ACPI State : S4

`disk > /sys/power/state`

Suspend to disk

- Image mémoire persistée
- Système en lowpower, voire éteint

parametres

`/sys/power/disk`

- `platform`
- `shutdown`
- `reboot`
- `suspend`

Processus

1. Suspend
2. Snapshot
3. Resume
4. Write

Smile
OPEN SOURCE SOLUTIONS

Modes d'endormissement

ACPI State : S4

`disk > /sys/power/state`

Suspend to disk

- Image mémoire persistée
- Système en lowpower, voire éteint

parametres

`/sys/power/disk`

- `platform`
- `shutdown`
- `reboot`
- `suspend`

Processus

1. Suspend
2. Snapshot
3. Resume
4. Write
5. Poweroff

Smile
OPEN SOURCE SOLUTIONS

Wakeup

Device Tree

- `wakeup-source` (Générique)
- `gpio-key,wakeup`
- `enable-sdio-wakeup`
- `linux,wakeup`
- etc. (Anciens bindings)

Smile
OPEN SOURCE SOLUTIONS

Wakeup

Device Tree

- `wakeup-source` (Générique)
- `gpio-key,wakeup`
- `enable-sdio-wakeup`
- `linux,wakeup`
- etc. (Anciens bindings)

sysfs

- `enabled > /sys/devices/.../power/wakeup`
- `disabled > /sys/devices/.../power/wakeup`

Consultation : `/sys/kernel/debug/wakeup_sources`

Smile
OPEN SOURCE SOLUTIONS

C'est fini

Merci

Smile
OPEN SOURCE SOLUTIONS