Scheduler Load tracking update and improvement

Vincent Guittot

Linaro
connect
Budapest 2017

ENGINEERS
AND DEVICES
WORKING
TOGETHER

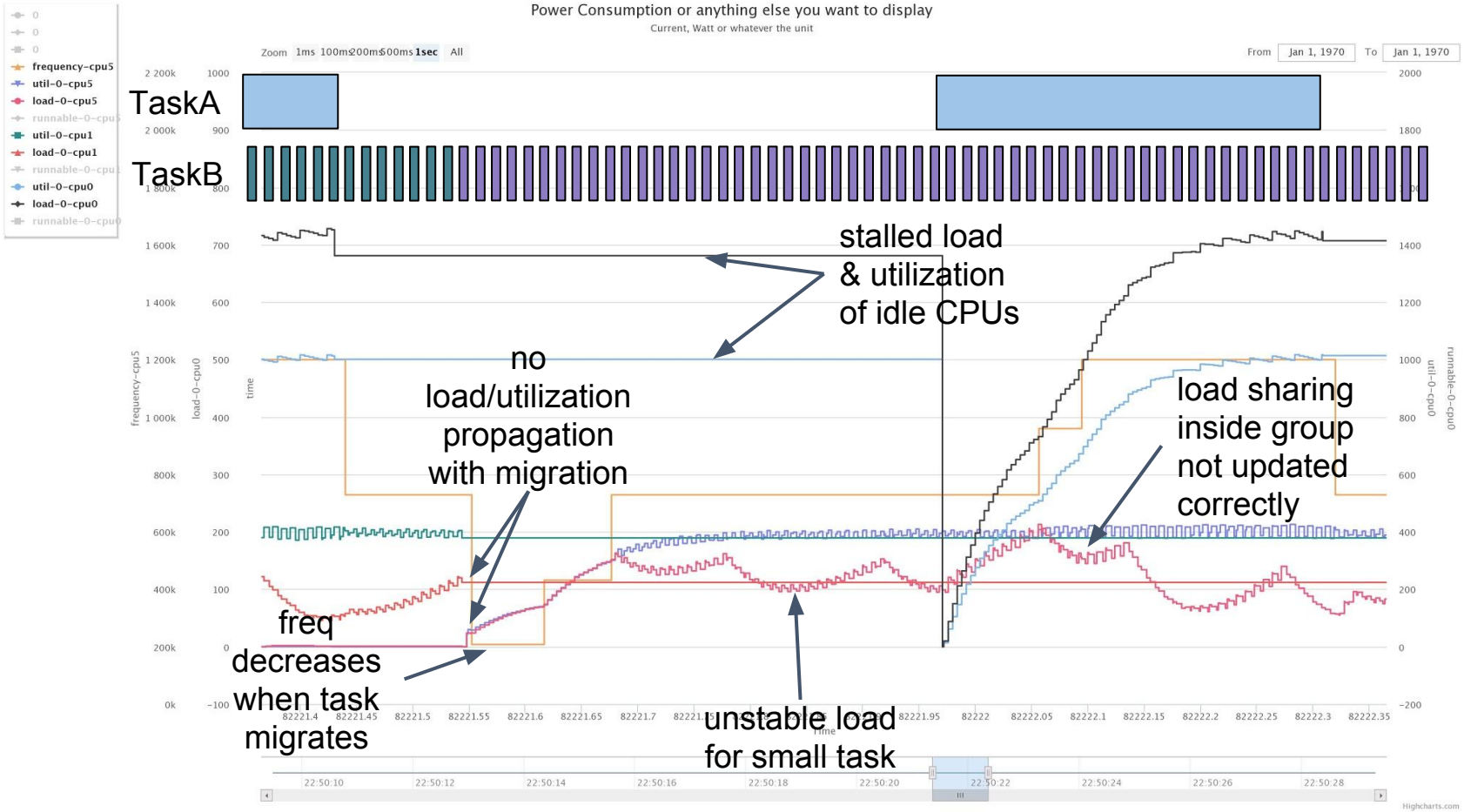
Agenda

- The problem
- The enhancements
- Some figures
- What’s next ?

The problem

- Can’t get best performance when using PELT
  - Compared to WALT as an example
- There are/were several issues with PELT
  - not enough responsive
  - Task migrations in groups not reflected
  - load/utilization metrics not always stable
  - ...

The Problem


Power Consumption or anything else you want to display
Current, Watt or whatever the unit
Zoom 1ms 100ms 200ms 500ms 1sec All
From Jan 1, 1970 To Jan 1, 1970
frequency-cpu5
util-0-cpu5
load-0-cpu5
runnable-0-cpu5
util-0-cpu1
load-0-cpu1
runnable-0-cpu1
util-0-cpu0
load-0-cpu0
runnable-0-cpu0
TaskA
TaskB
stalled load & utilization of idle CPUs
no load/utilization propagation with migration
load sharing inside group not updated correctly
freq decreases when task migrates
unstable load for small task
time
Time
Highcharts.com

Linaro
connect
Budapest 2017

ENGINEERS
AND DEVICES
WORKING
TOGETHER

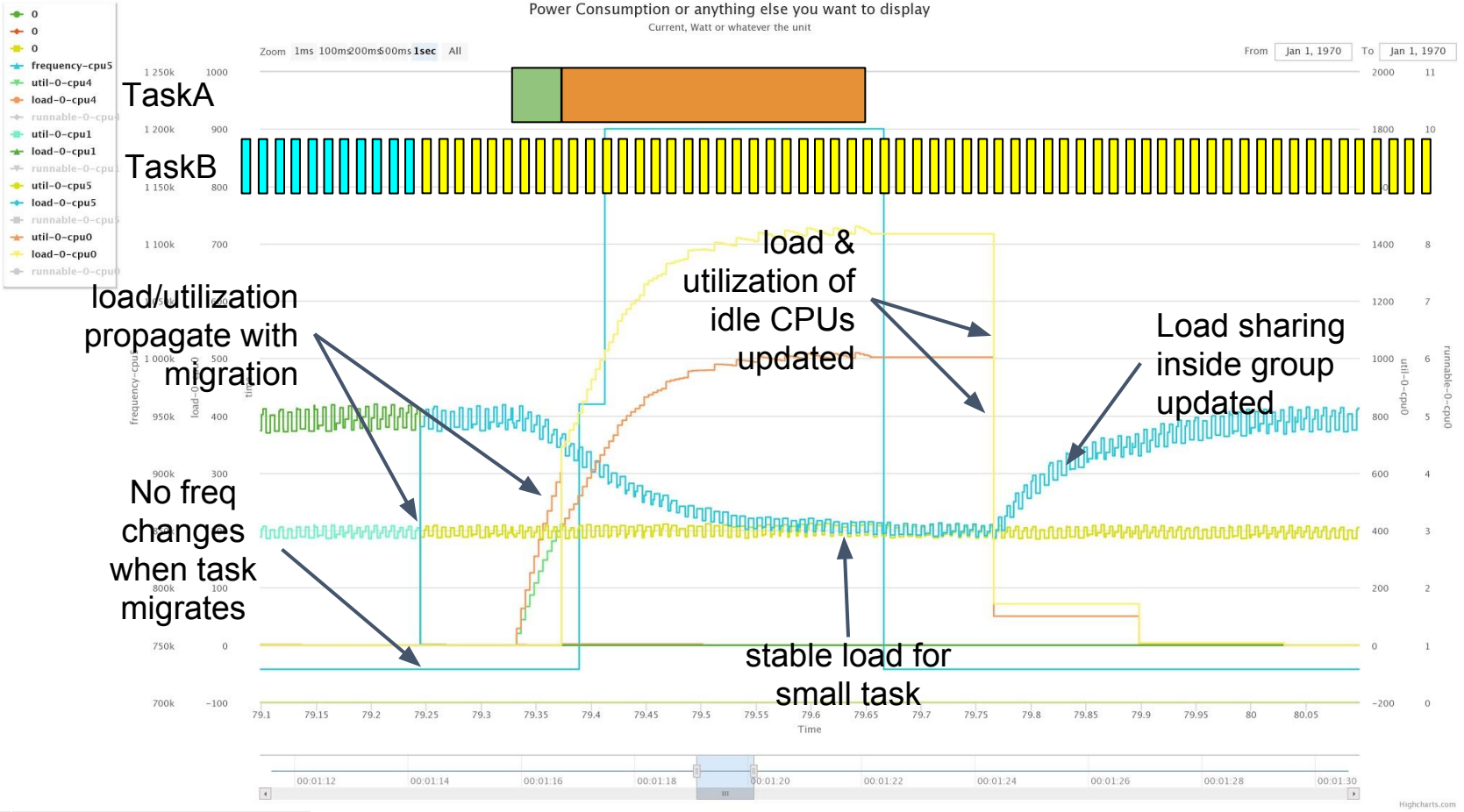
The enhancements

- Almost 20 patches that impacts PELT merged since v4.7

- Some noticeable changes
  - Propagate utilization/load in cgroup (merged)
  - Fix small task tracking (merged)

- More to come
  - Update blocked load
  - New invariance scheme

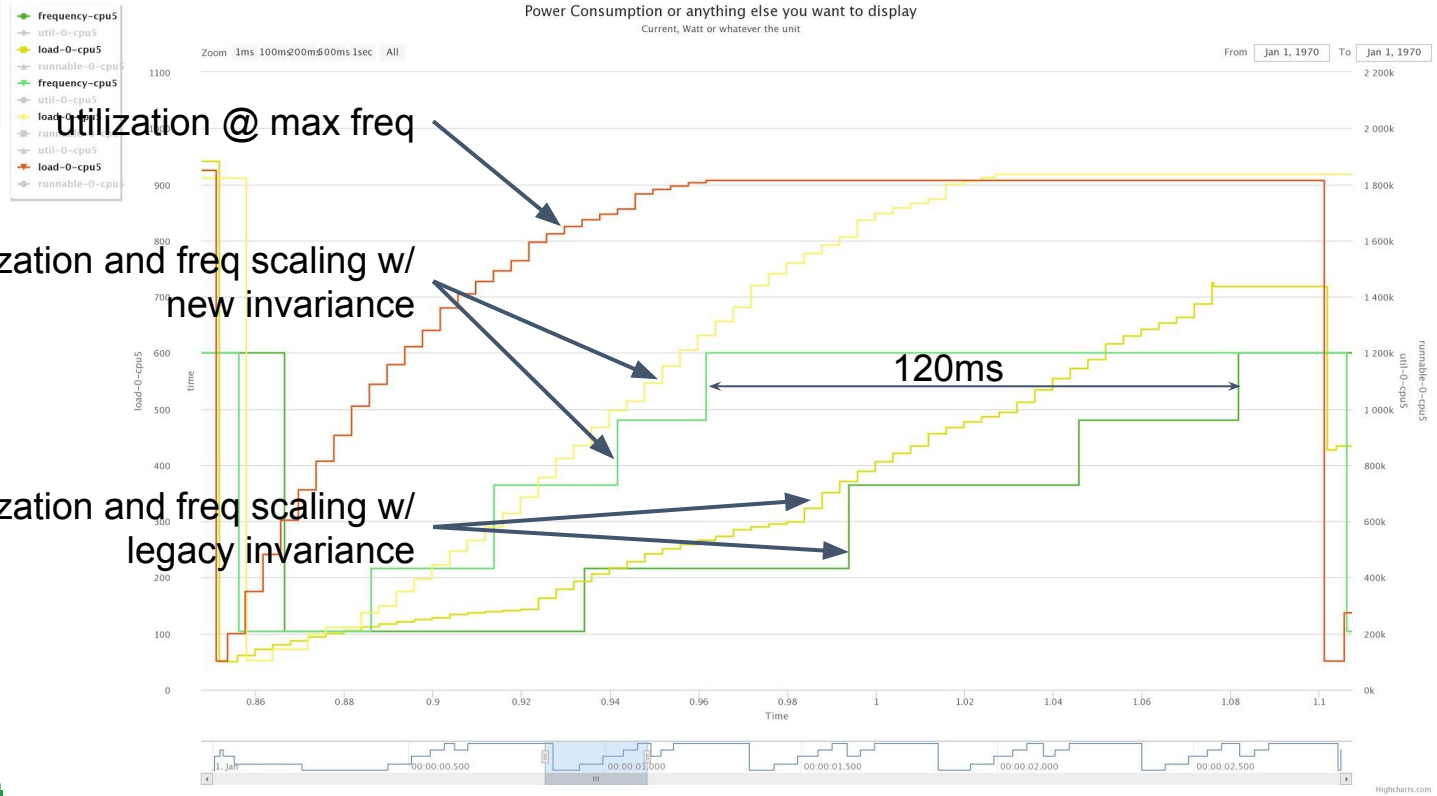
The target


Power Consumption or anything else you want to display
Current, Watt or whatever the unit
TaskA
TaskB
load/utilization propagate with migration
load & utilization of idle CPUs updated
Load sharing inside group updated
No freq changes when task migrates
stable load for small task
Time

Linaro connect
Budapest 2017

ENGINEERS AND DEVICES WORKING TOGETHER

New invariance scheme



More realistic UC figures

- Previous results and tests were unitary ones
- Move to more realistic bench
  - Android Hikey : Kernel v4.4 (w/ EAS)
  - Vellamo bench
  - UI jank test

Thermal mitigation impact

- Thermal mitigation impacts on results
  - No simple rule
  - Done tests with external active cooling
    - Prevent SoC temp to go above 65°C and trig thermal mitigation
  - Performance impact of thermal mitigation is not obvious
    - Sometime results at the opposite.
    - Vellamo Threadbench seems to take advantage of thermal mitigation ...

Vellamo Multicore bench

	Android kernel v4.4 WALT		Android kernel v4.4 legacy PELT		Android kernel v4.4 new PELT	
	score	stdev	**score**	stdev	**score**	stdev
Multicore	**1,407**	4%	**1,295**	13%	**1,548**	3%
multi.Linpack	**130**	2%	**125**	2%	**132**	0%
multi.LinpackJava	**176**	2%	**166**	3%	**176**	2%
multi.Stream	**165**	3%	**155**	0%	**165**	4%
multi.Membench	**208**	7%	**266**	62%	**206**	7%
multi.Sysbench	**182**	6%	**166**	6%	**232**	12%
multi.Threadbench	**201**	31%	**105**	19%	**279**	8%
multi.Parsec	**219**	3%	**176**	13%	**223**	9%
multi.ProcessCommunication	**125**	1%	**136**	1%	**134**	1%

UI interaction bench

	Android kernel v4.4 WALT		Android kernel v4.4 legacy PELT		Android kernel v4.4 new PELT	
	ms	Stdev (ms)	**ms**	Stdev (ms)	**ms**	Stdev (ms)
hwuitest						
50th percentile	**22.00**	0.00	**22.21**	0.41	**21.63**	0.49
90th percentile	**24.00**	1.32	**25.04**	1.33	**22.67**	0.56
95th percentile	**24.79**	1.50	**27.29**	1.97	**23.04**	0.81
99th percentile	**29.75**	4.06	**34.63**	4.44	**26.29**	2.33
recentfling						
90th percentile	**33.50**	0.88	**33.75**	1.07	**32.21**	2.83
95th percentile	**36.17**	0.82	**36.67**	0.96	**35.42**	2.59
99th percentile	**39.58**	1.44	**40.75**	1.75	**39.17**	2.12

Linaro
connect
Budapest 2017

ENGINEERS
AND DEVICES
WORKING
TOGETHER

What next ?

- RT task utilization tracking
- Improve responsiveness furthermore

Linaro
connect
Budapest 2017

ENGINEERS
AND DEVICES
WORKING
TOGETHER

RT task utilization tracking

- RT tasks run at max freq
  - At least, it tries
  - Frequency change on hikey needs around 1.5ms
  - slow path of schedutil uses RT thread
- Current load tracking: rt_avg
  - Fine for slow variation around hundreds of ms
  - Fine when averaging period is larger than updating period
- Use a similar load tracking than CFS
  - Only at rq level
  - Same responsiveness than CFS

Linaro
connect
Budapest 2017

ENGINEERS
AND DEVICES
WORKING
TOGETHER

Improve PELT responsiveness

- Decay windowing
- Change half decay period
  - Current period is 32ms
- Use other scheduling class
-

Linaro
connect
Budapest 2017

ENGINEERS
AND DEVICES
WORKING
TOGETHER

Next

- Continue the discussion during hacking session

Thank You

#BUD17

For further information: www.linaro.org

BUD17 keynotes and videos on: connect.linaro.org