E-Team: Practical Energy Accounting for Multi-Core Systems

Till Smejkal[1], Marcus Hähnel[1], Thomas Ilsche[2], Michael Roitzsch[1], Wolfgang E. Nagel[2], and Hermann Härtig[1]

[1] Operating Systems Group, TU Dresden

[2] Center for Information Services and High Performance Computing (ZIH), TU Dresden

HAEC

July 13th 2017

TECHNISCHE UNIVERSITÄT DRESDEN

Motivation

Existing energy measurement methods are not suited for energy-based billing or accurate energy profiling of applications.

Motivation

Existing energy measurement methods are not suited for energy-based billing or accurate energy profiling of applications.

External Measurement

- ☺ Accurate energy/power values
- ☹ Limited to machine-level granularity
- ☹ Expensive deployment for data centers

Motivation

Existing energy measurement methods are not suited for energy-based billing or accurate energy profiling of applications.

External Measurement

- ☺ Accurate energy/power values
- ☹ Limited to machine-level granularity
- ☹ Expensive deployment for data centers

Inference-Based Estimation

- ☺ Adjustable granularity and resolution
- ☹ Calibration required for every system
- ☹ 8 % to 10 % inaccuracy

Running Average Power Limit

Running Average Power Limit

In 2011 Intel® introduced *Running Average Power Limit* (RAPL) into their processors, a feature to limit the power consumption of the processor.

Running Average Power Limit

In 2011 Intel® introduced *Running Average Power Limit* (RAPL) into their processors, a feature to limit the power consumption of the processor.

- Processor internally estimates the consumed energy
- Four energy counters are exported by the processor
- Each CPU socket has separate energy counters

Running Average Power Limit

In 2011 Intel® introduced *Running Average Power Limit* (RAPL) into their processors, a feature to limit the power consumption of the processor.

- Processor internally estimates the consumed energy
- Four energy counters are exported by the processor
- Each CPU socket has separate energy counters

RAPL is very promising for on-line energy measurements.

Running Average Power Limit

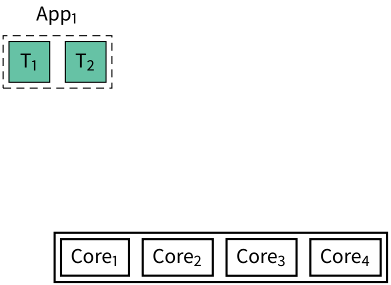
Spatial Granularity

RAPL only reports the energy consumption of the whole CPU socket.

Running Average Power Limit

Spatial Granularity

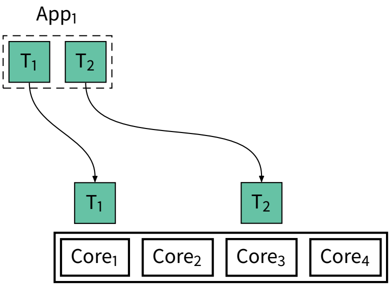
RAPL only reports the energy consumption of the whole CPU socket.



Running Average Power Limit

Spatial Granularity

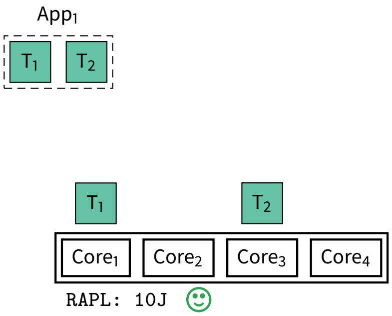
RAPL only reports the energy consumption of the whole CPU socket.



Running Average Power Limit

Spatial Granularity

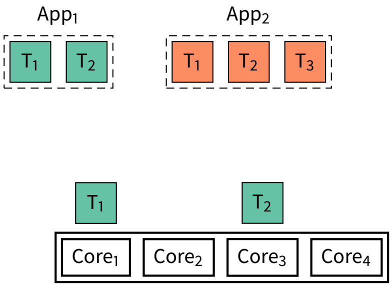
RAPL only reports the energy consumption of the whole CPU socket.



Running Average Power Limit

Spatial Granularity

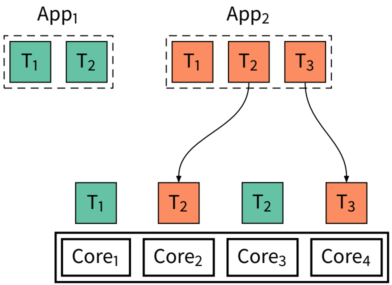
RAPL only reports the energy consumption of the whole CPU socket.



Running Average Power Limit

Spatial Granularity

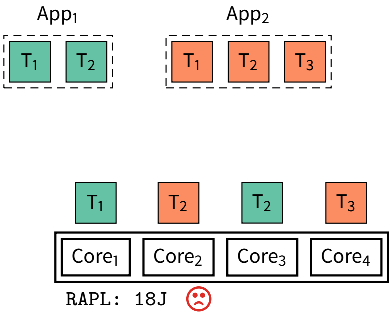
RAPL only reports the energy consumption of the whole CPU socket.



Running Average Power Limit

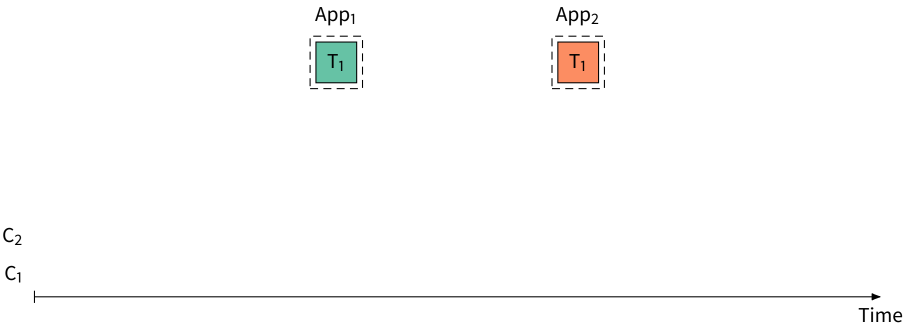
Spatial Granularity

RAPL only reports the energy consumption of the whole CPU socket.



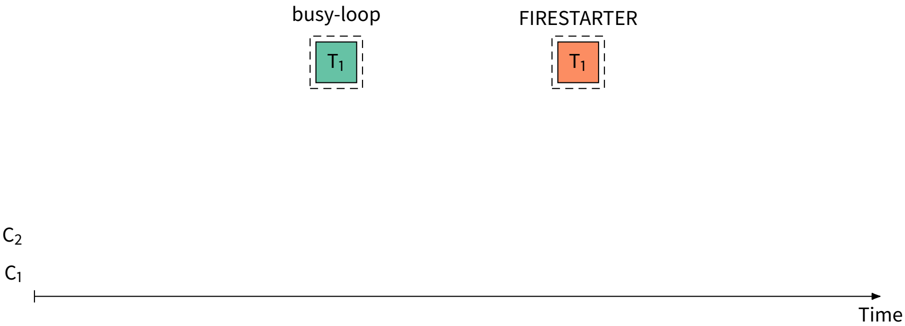
Running Average Power Limit

Spatial Granularity



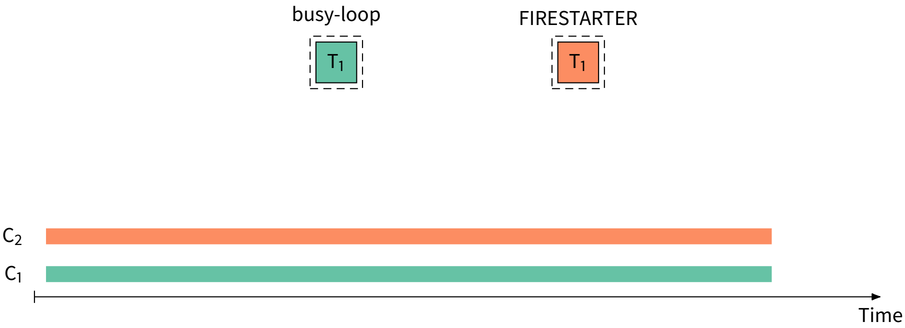
Running Average Power Limit

Spatial Granularity



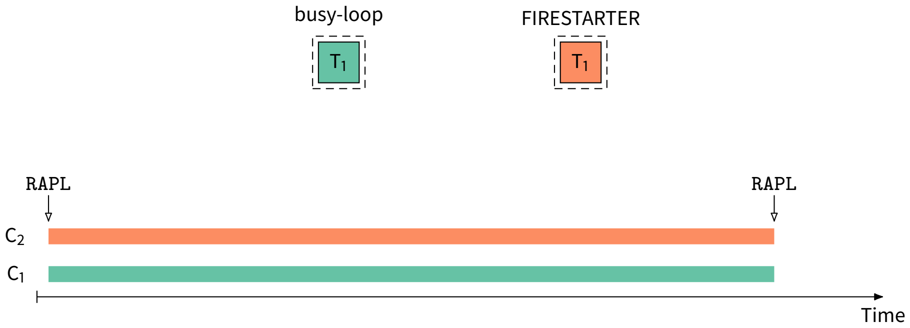
Running Average Power Limit

Spatial Granularity



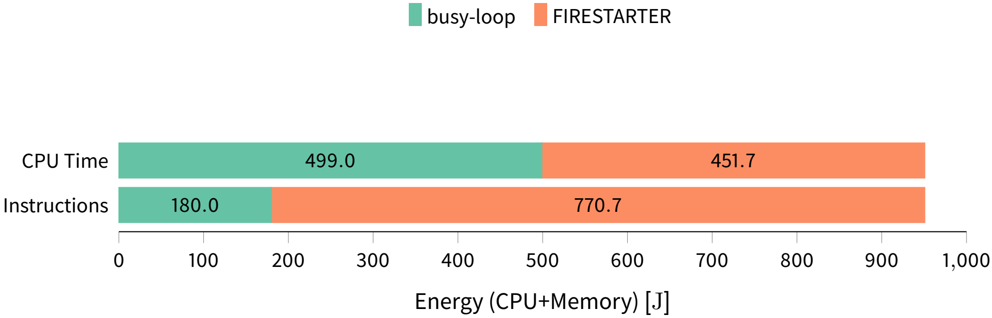
Running Average Power Limit

Spatial Granularity



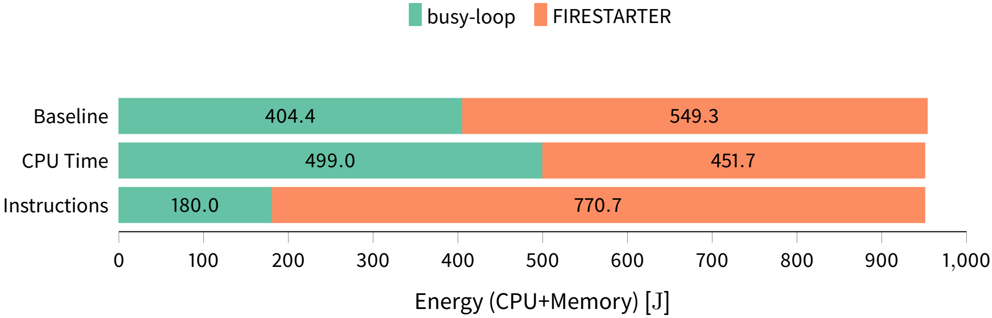
Running Average Power Limit

Spatial Granularity



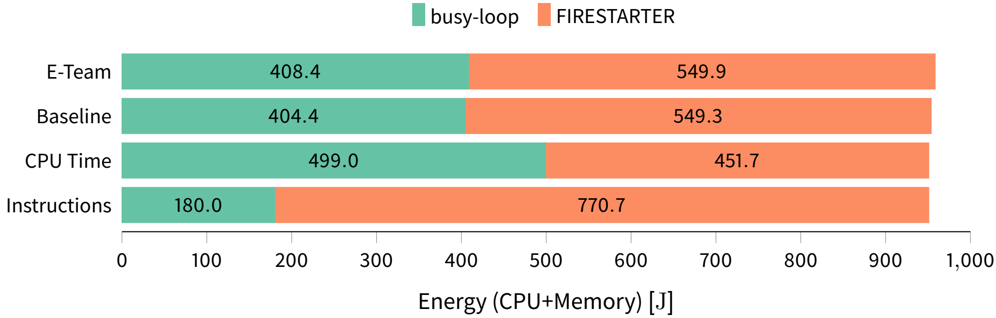
Running Average Power Limit

Spatial Granularity



Running Average Power Limit

Spatial Granularity



Running Average Power Limit

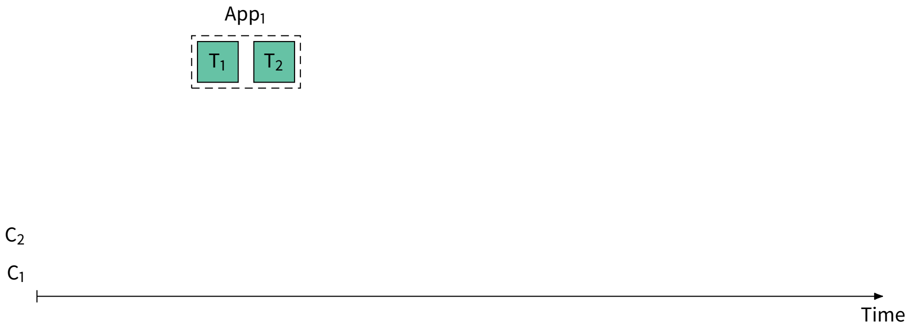
Temporal Granularity

RAPL energy counters are updated only at discrete intervals.

Running Average Power Limit

Temporal Granularity

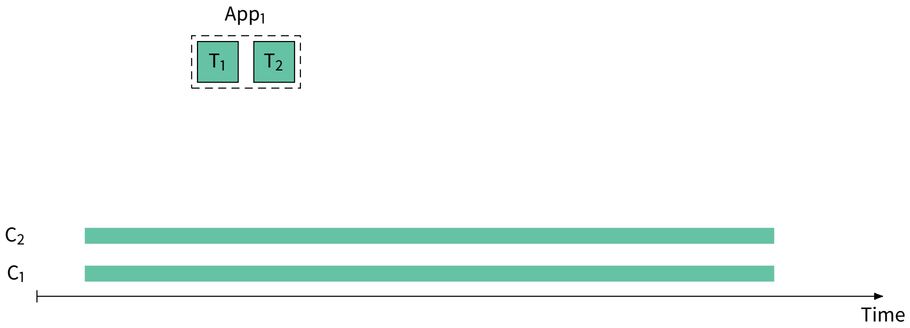
RAPL energy counters are updated only at discrete intervals.



Running Average Power Limit

Temporal Granularity

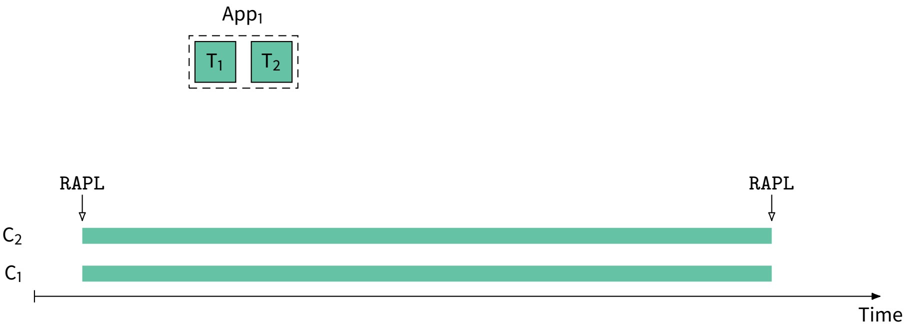
RAPL energy counters are updated only at discrete intervals.



Running Average Power Limit

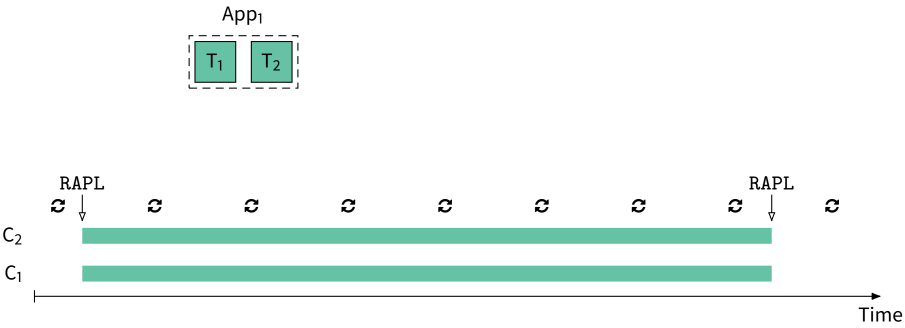
Temporal Granularity

RAPL energy counters are updated only at discrete intervals.



Running Average Power Limit
Temporal Granularity

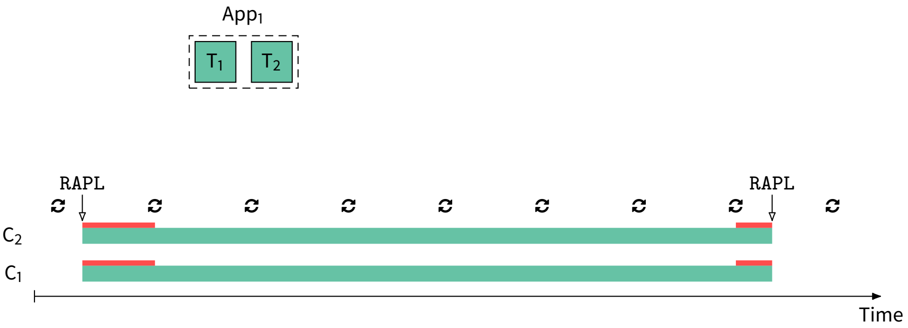
RAPL energy counters are updated only at discrete intervals.



Running Average Power Limit

Temporal Granularity

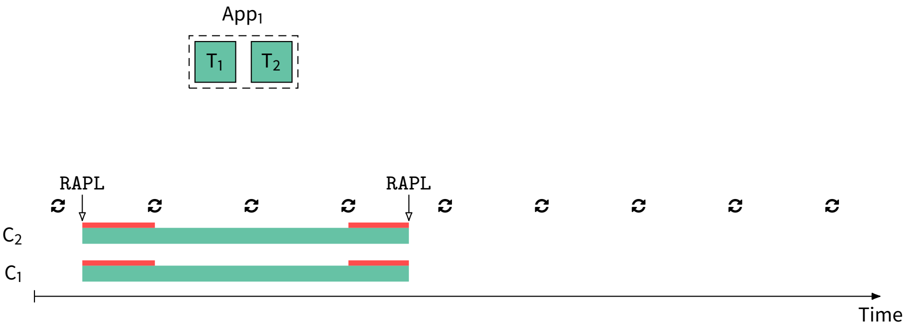
RAPL energy counters are updated only at discrete intervals.



Running Average Power Limit

Temporal Granularity

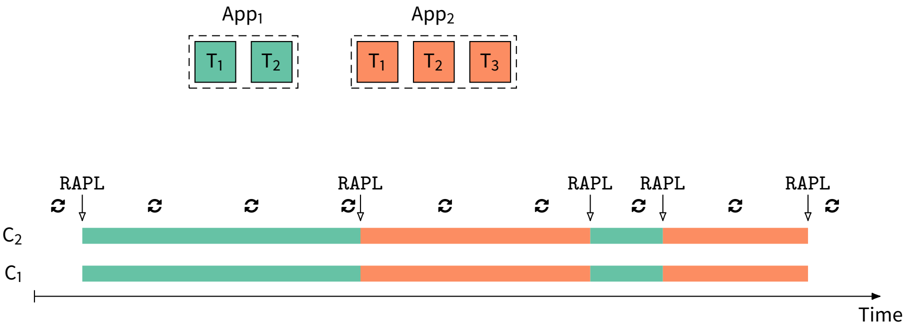
RAPL energy counters are updated only at discrete intervals.



Running Average Power Limit

Temporal Granularity

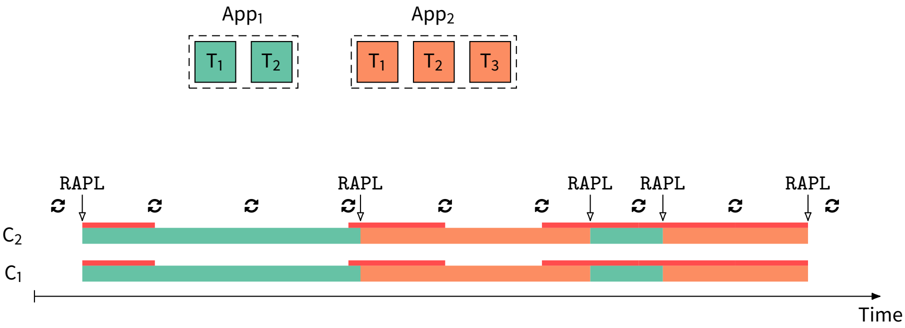
RAPL energy counters are updated only at discrete intervals.



Running Average Power Limit

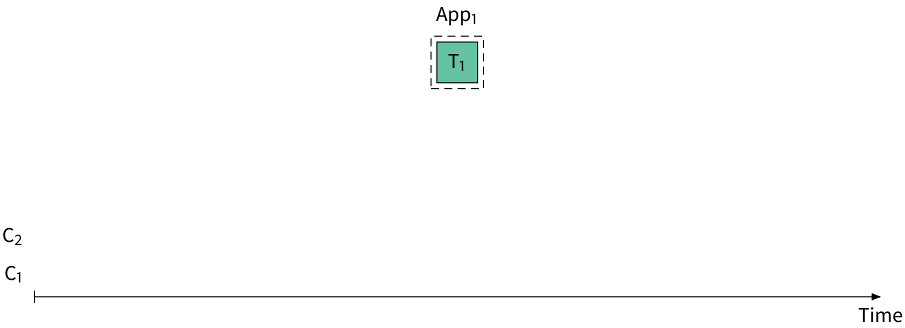
Temporal Granularity

RAPL energy counters are updated only at discrete intervals.



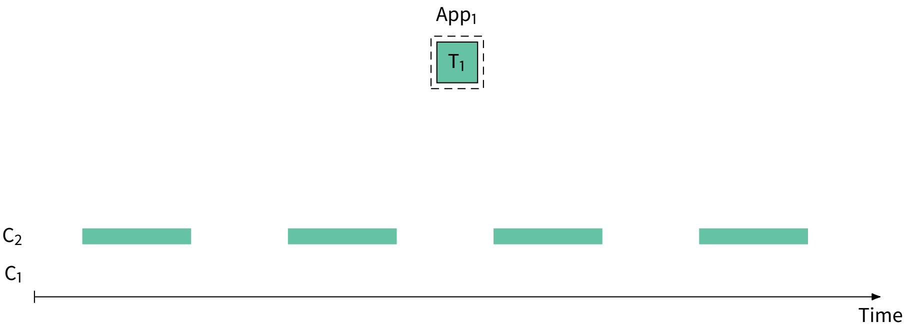
Running Average Power Limit

Temporal Granularity



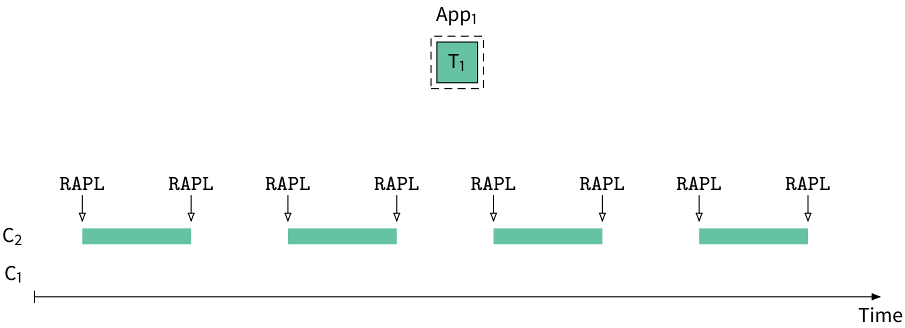
Running Average Power Limit

Temporal Granularity



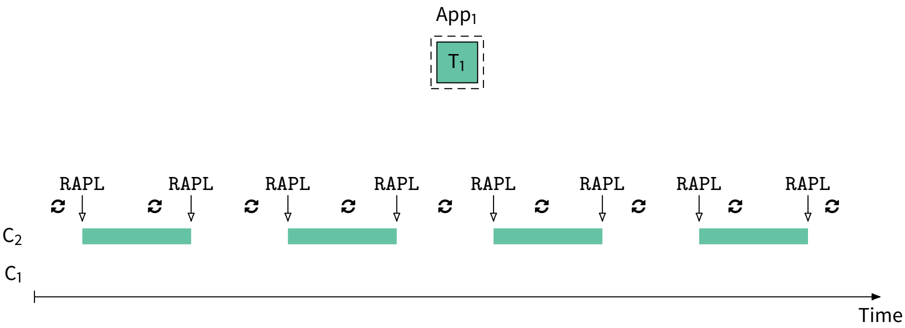
Running Average Power Limit

Temporal Granularity



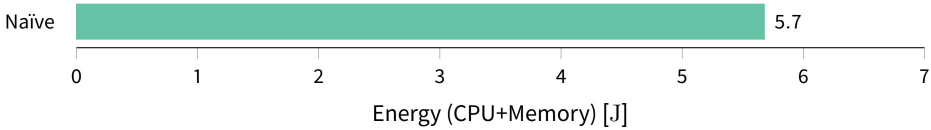
Running Average Power Limit

Temporal Granularity



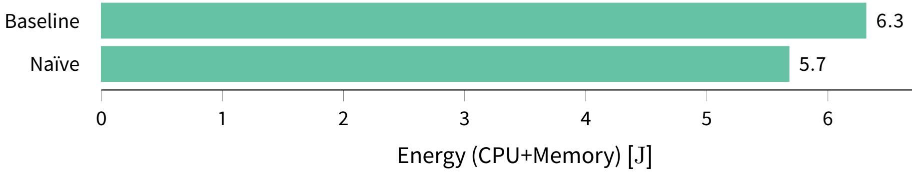
Running Average Power Limit

Temporal Granularity



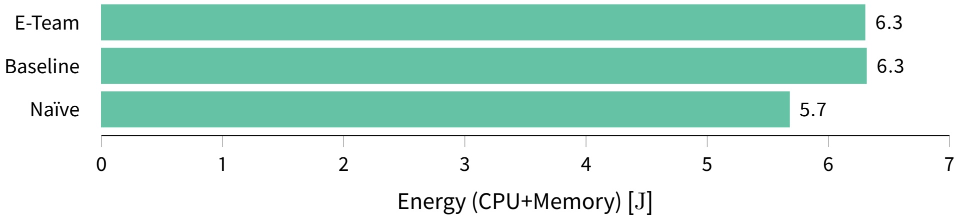
Running Average Power Limit

Temporal Granularity



Running Average Power Limit

Temporal Granularity



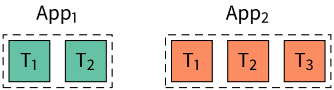
E-Team

E-Team

E-Team allows per-application energy accounting based on socket-local measurements by leveraging a specialized scheduling scheme.

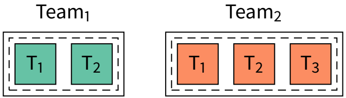
E-Team

E-Team allows per-application energy accounting based on socket-local measurements by leveraging a specialized scheduling scheme.



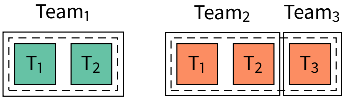
E-Team

E-Team allows per-application energy accounting based on socket-local measurements by leveraging a specialized scheduling scheme.



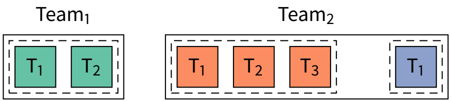
E-Team

E-Team allows per-application energy accounting based on socket-local measurements by leveraging a specialized scheduling scheme.



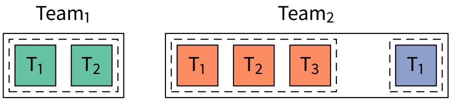
E-Team

E-Team allows per-application energy accounting based on socket-local measurements by leveraging a specialized scheduling scheme.



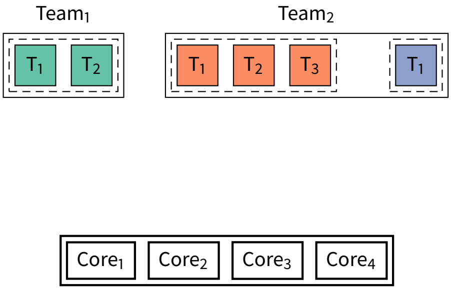
E-Team

Solving Spatial Granularity



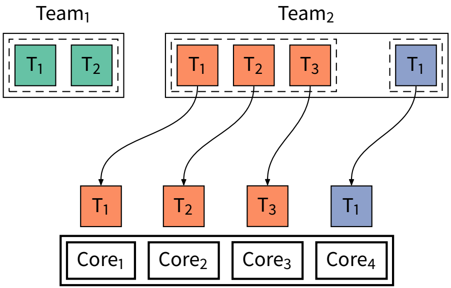
E-Team

Solving Spatial Granularity



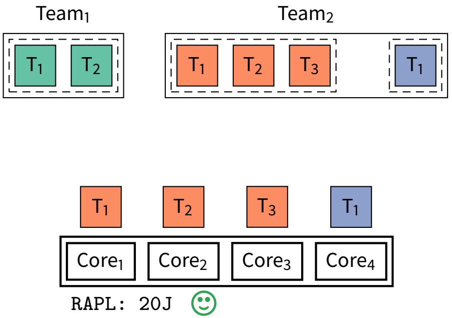
E-Team

Solving Spatial Granularity



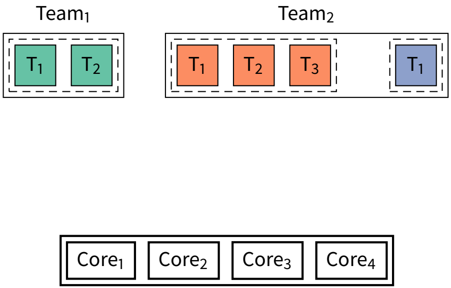
E-Team

Solving Spatial Granularity



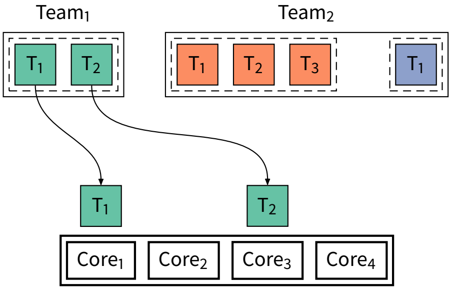
E-Team

Solving Spatial Granularity



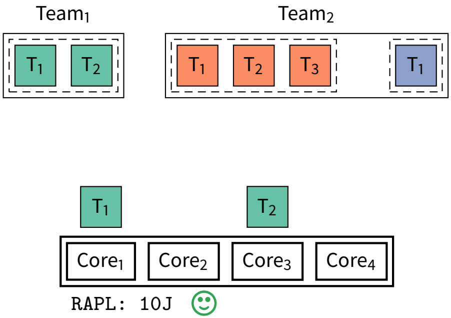
E-Team

Solving Spatial Granularity



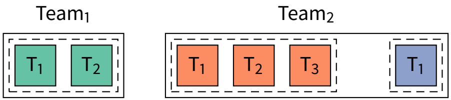
E-Team

Solving Spatial Granularity



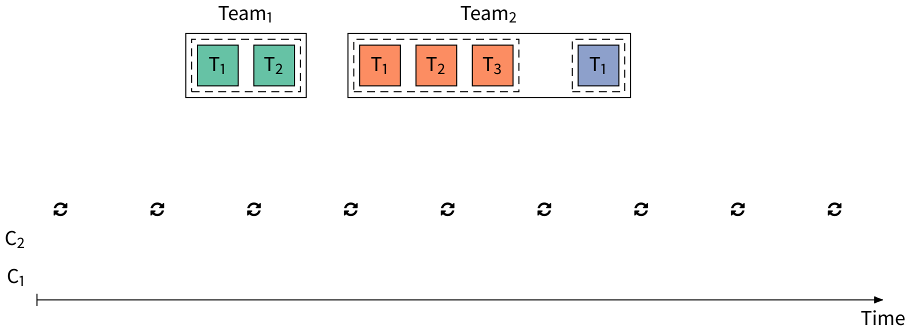
E-Team

Solving Temporal Granularity



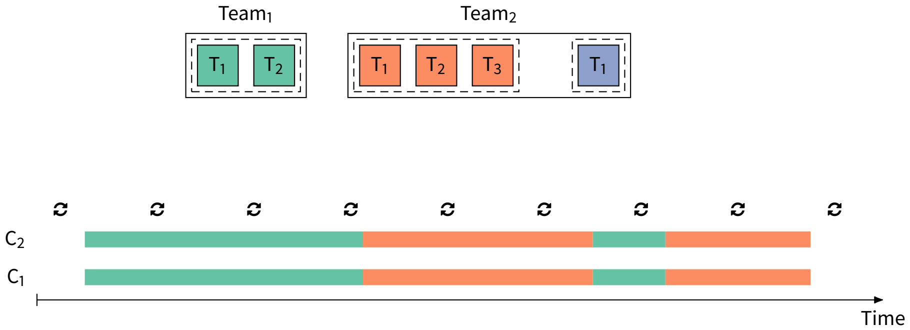
E-Team

Solving Temporal Granularity



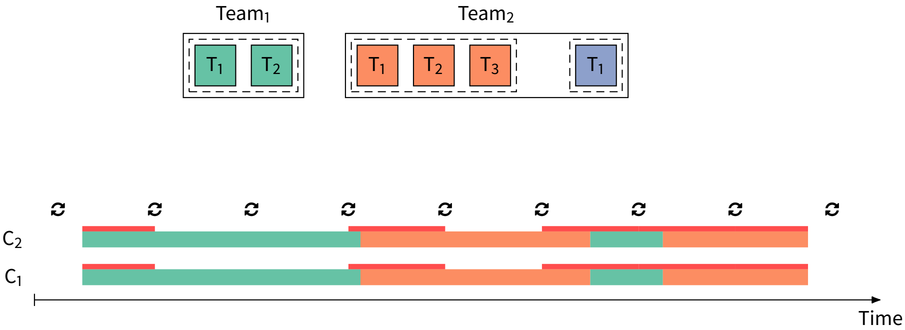
E-Team

Solving Temporal Granularity



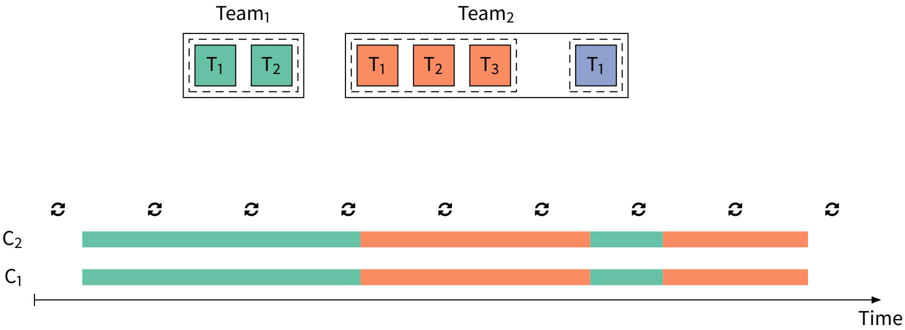
E-Team

Solving Temporal Granularity



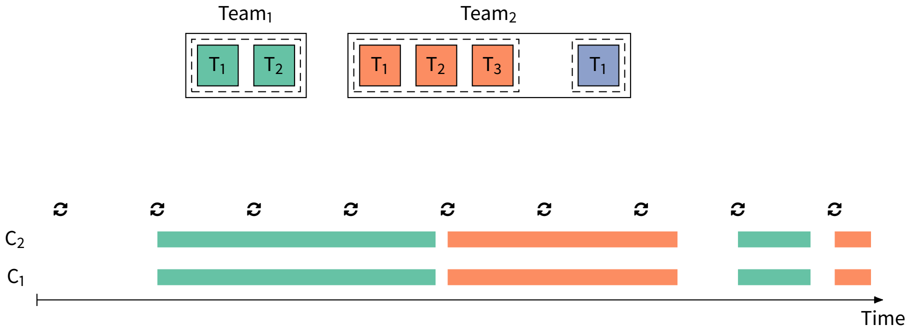
E-Team

Solving Temporal Granularity



E-Team

Solving Temporal Granularity



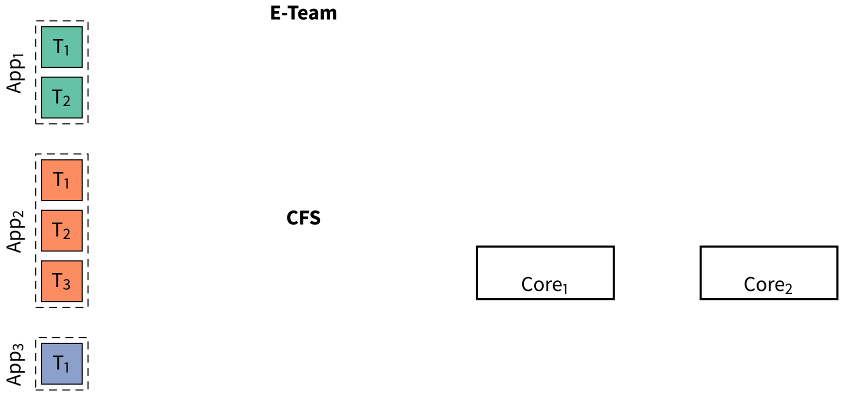
E-Team

Implementation

E-Team is implemented as a new scheduler in the Linux kernel.

E-Team
Implementation

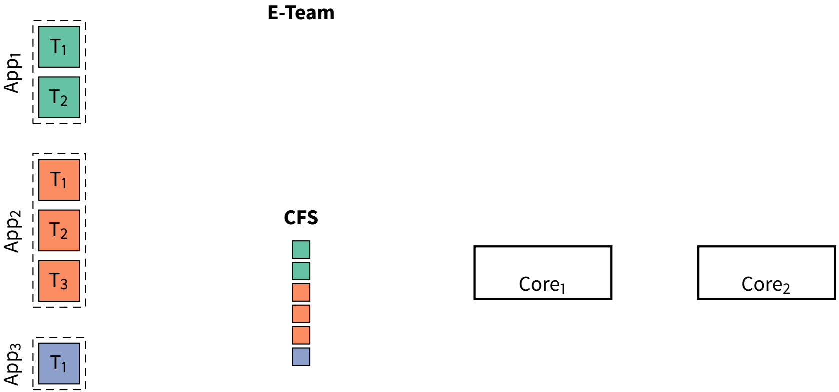
E-Team is implemented as a new scheduler in the Linux kernel.



E-Team

Implementation

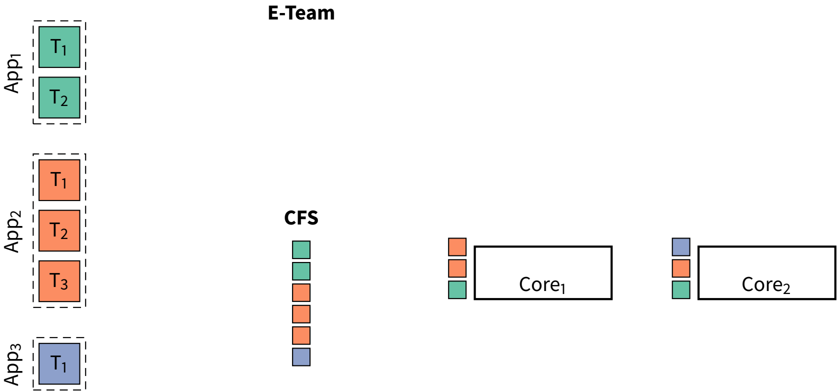
E-Team is implemented as a new scheduler in the Linux kernel.



E-Team

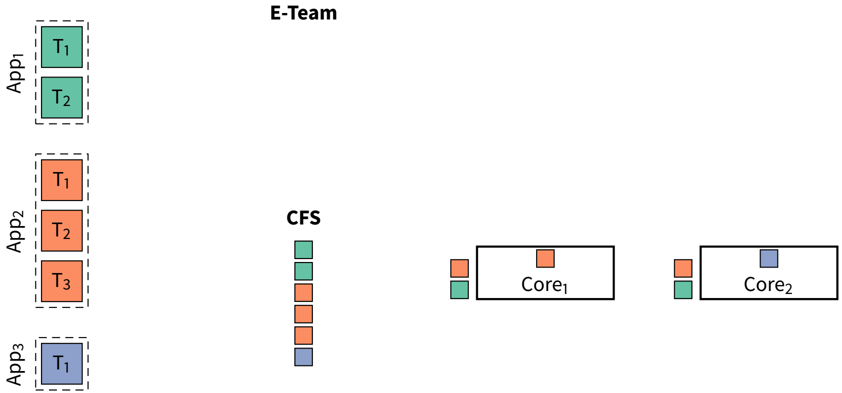
Implementation

E-Team is implemented as a new scheduler in the Linux kernel.



E-Team
Implementation

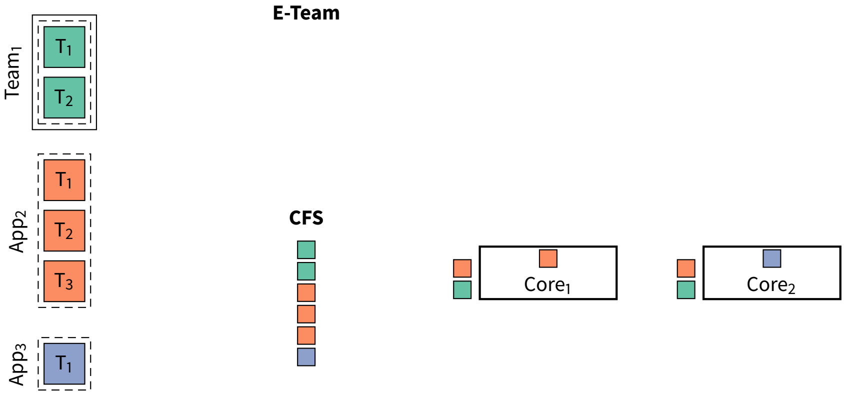
E-Team is implemented as a new scheduler in the Linux kernel.



E-Team

Implementation

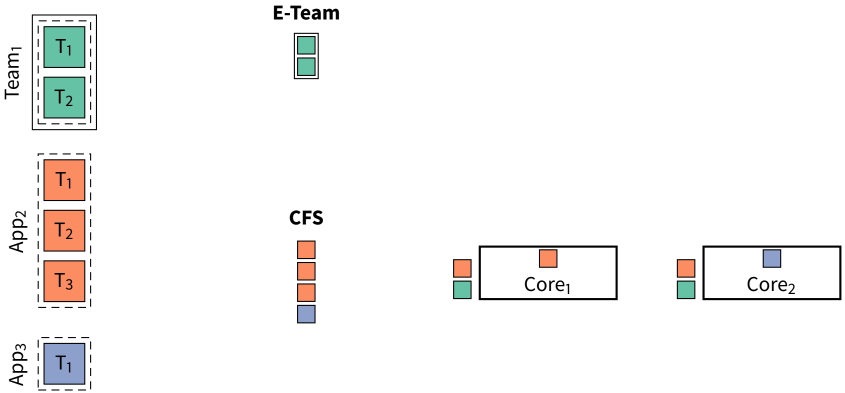
E-Team is implemented as a new scheduler in the Linux kernel.



E-Team

Implementation

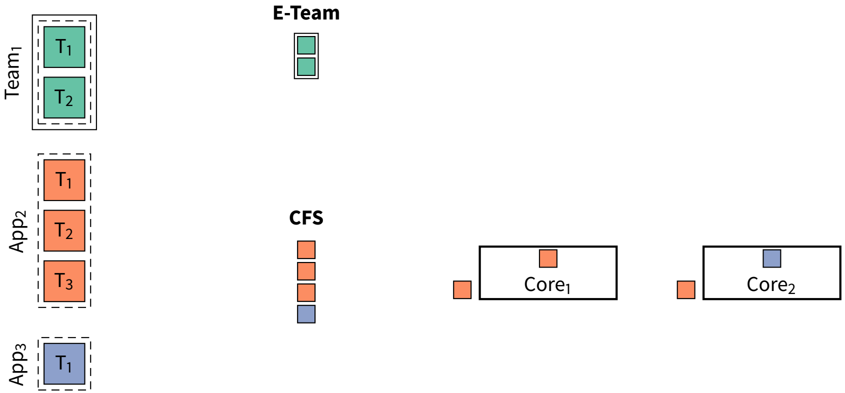
E-Team is implemented as a new scheduler in the Linux kernel.



E-Team

Implementation

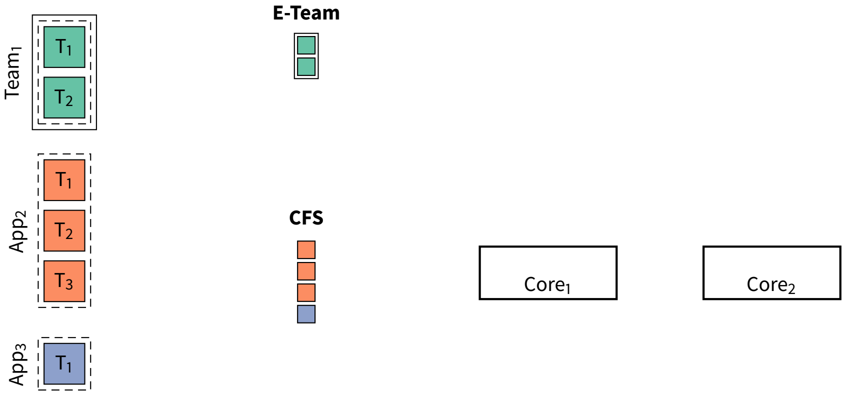
E-Team is implemented as a new scheduler in the Linux kernel.



E-Team

Implementation

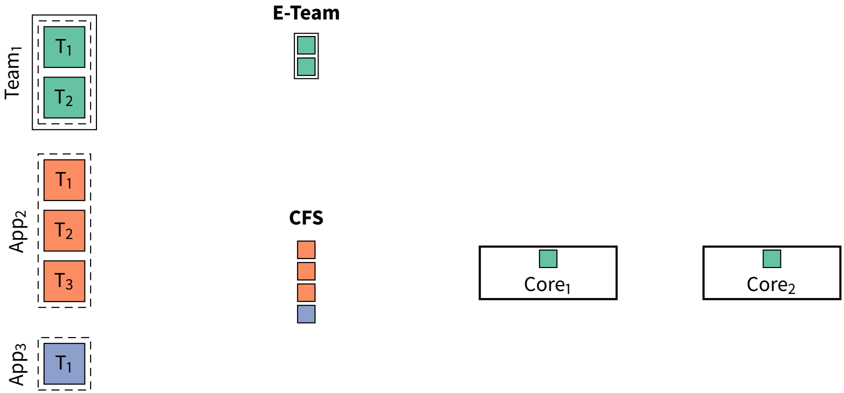
E-Team is implemented as a new scheduler in the Linux kernel.



E-Team

Implementation

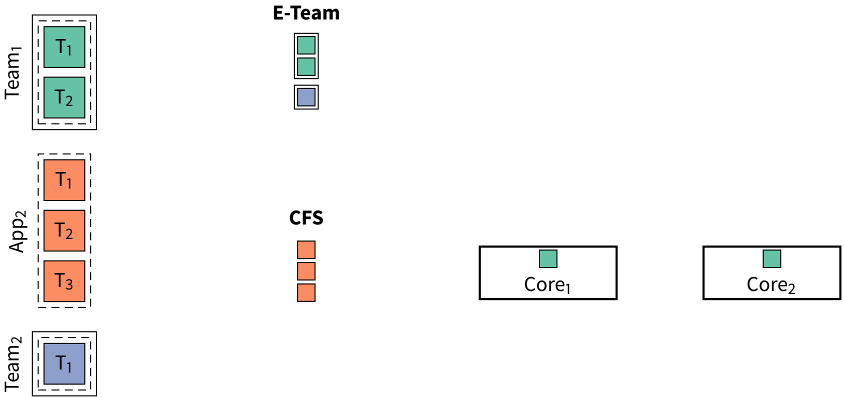
E-Team is implemented as a new scheduler in the Linux kernel.



E-Team

Implementation

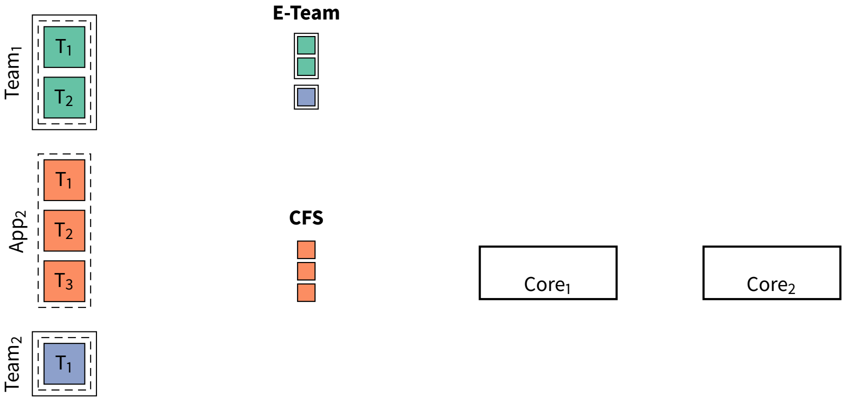
E-Team is implemented as a new scheduler in the Linux kernel.



E-Team

Implementation

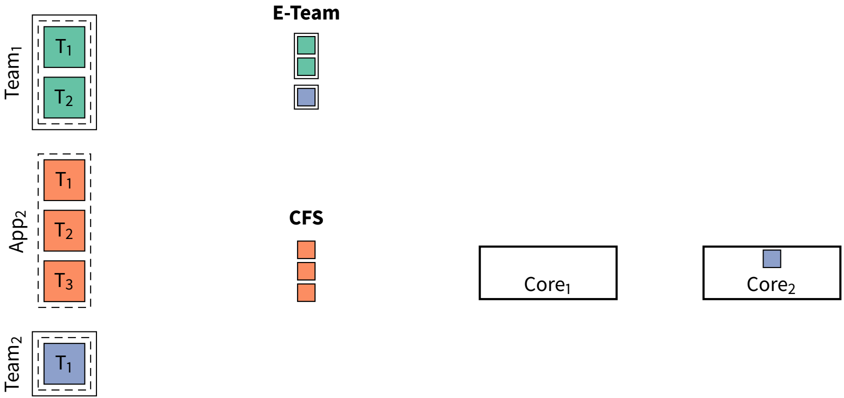
E-Team is implemented as a new scheduler in the Linux kernel.



E-Team

Implementation

E-Team is implemented as a new scheduler in the Linux kernel.



Evaluation

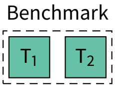
Evaluation

- Intel® Haswell Core™ i7-4770 with 3.4 GHz nominal frequency
- Disabled Hyper-Threading and Turbo Boost
- NAS Parallel Benchmarks[1] (OpenMP version) with 1 to 4 threads
- Results are relative differences to execution in stripped-down environment

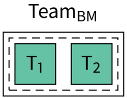
[1] Bailey et al., “The NAS parallel benchmarks”.

Evaluation

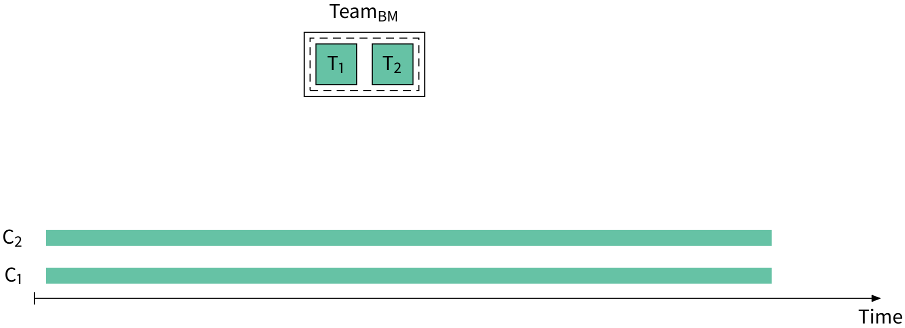
Solo



Evaluation
Solo

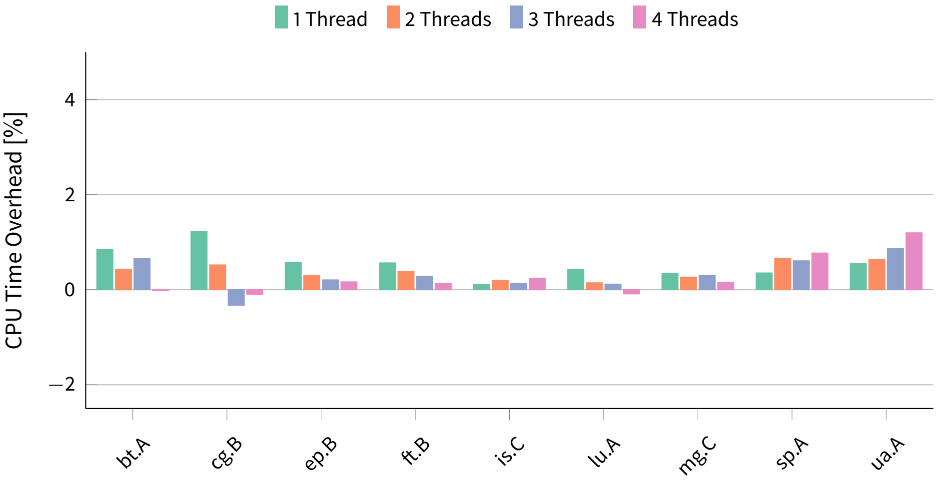


Evaluation
Solo



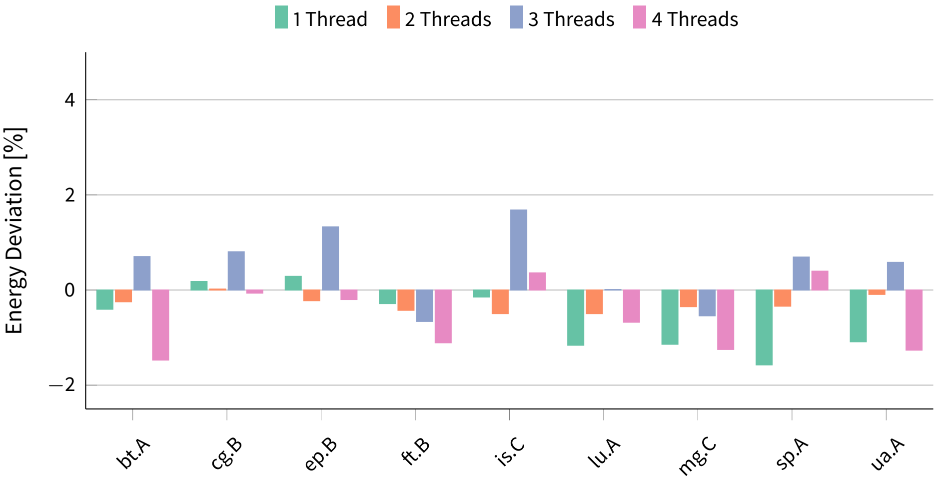
Evaluation

Solo

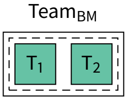


Evaluation

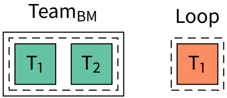
Solo



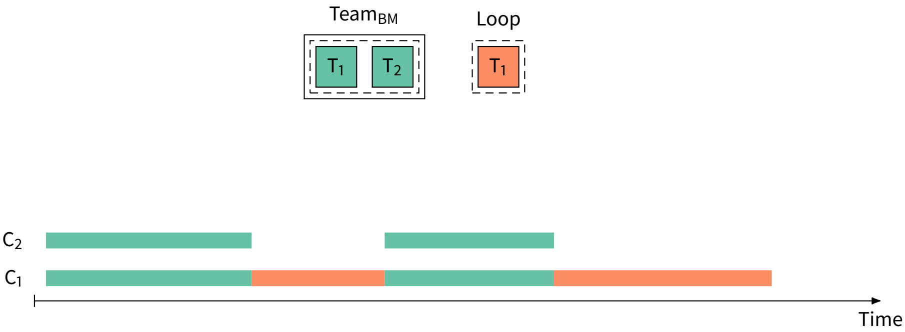
Evaluation
Background



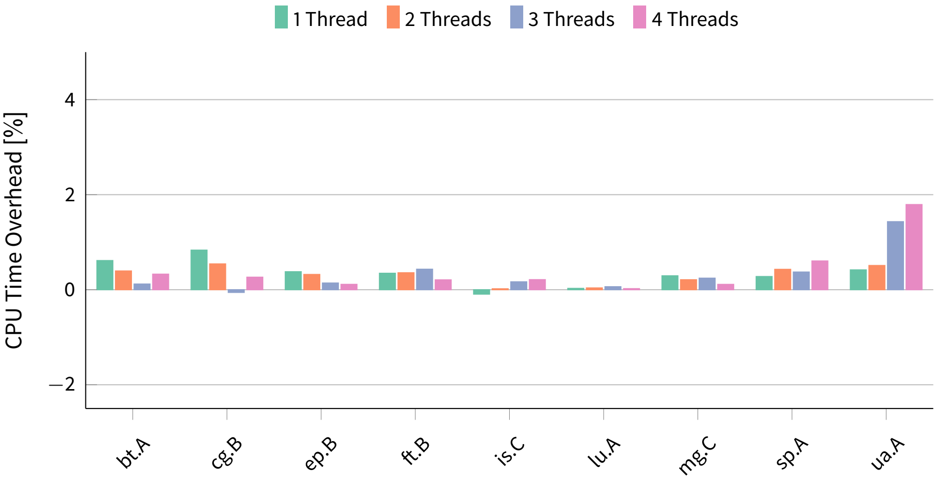
Evaluation
Background



Evaluation
Background

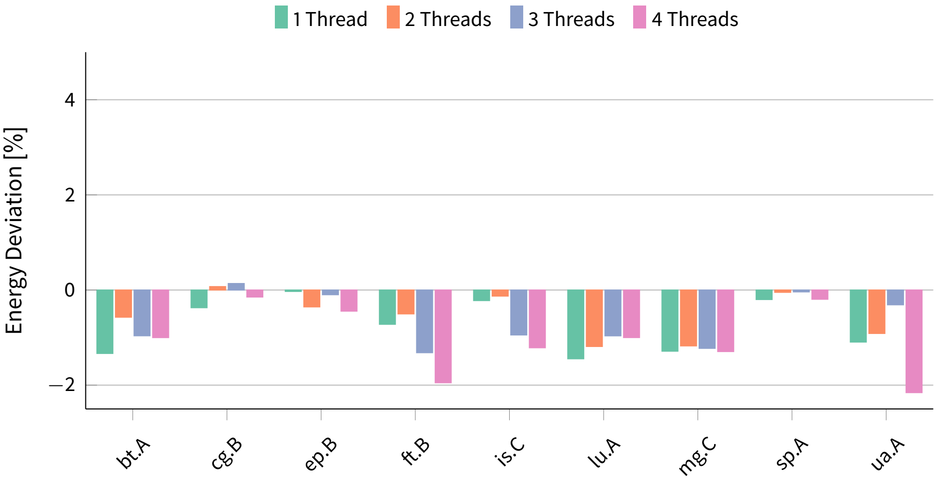


Evaluation
Background



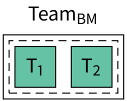
Evaluation

Background



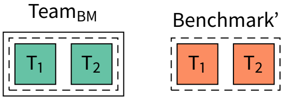
Evaluation

Comparison with External

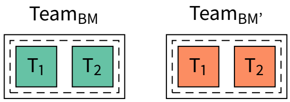


Evaluation

Comparison with External

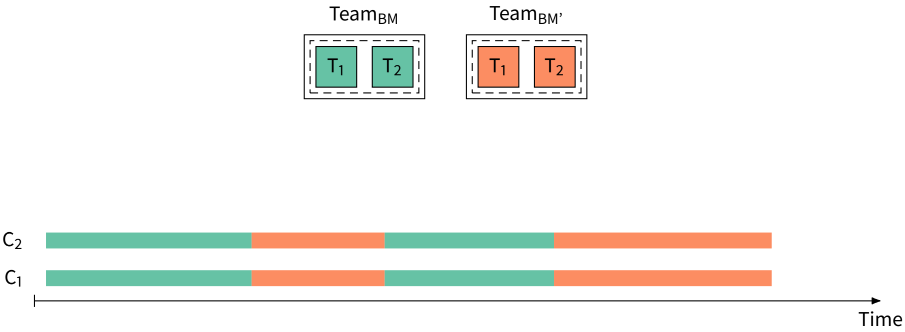


Evaluation
Comparison with External



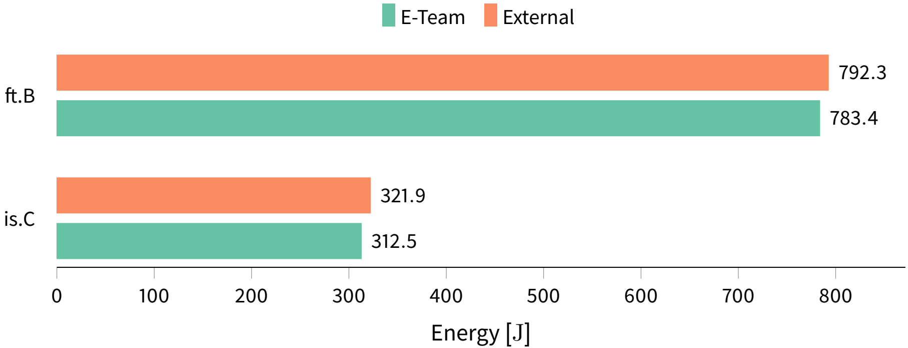
Evaluation

Comparison with External



Evaluation

Comparison with External



Limitations & Conclusion

Limitations

- Exclusive access to whole CPU socket potentially wastes resources
- I/O-intensive applications require frequent use of *Short-Time RAPL*

Limitations

- Exclusive access to whole CPU socket potentially wastes resources
- I/O-intensive applications require frequent use of *Short-Time RAPL*

Both limitations can be mitigated using *random sampling*.

Conclusion

- Accurate energy accounting for arbitrary groups of threads and applications
- No additional equipment, calibration, adaptation required
- Easy-to-use library and command line tool
- Extendable to other energy measurement methods

https://github.com/TUD-OS/{eteam,eteam-rt}