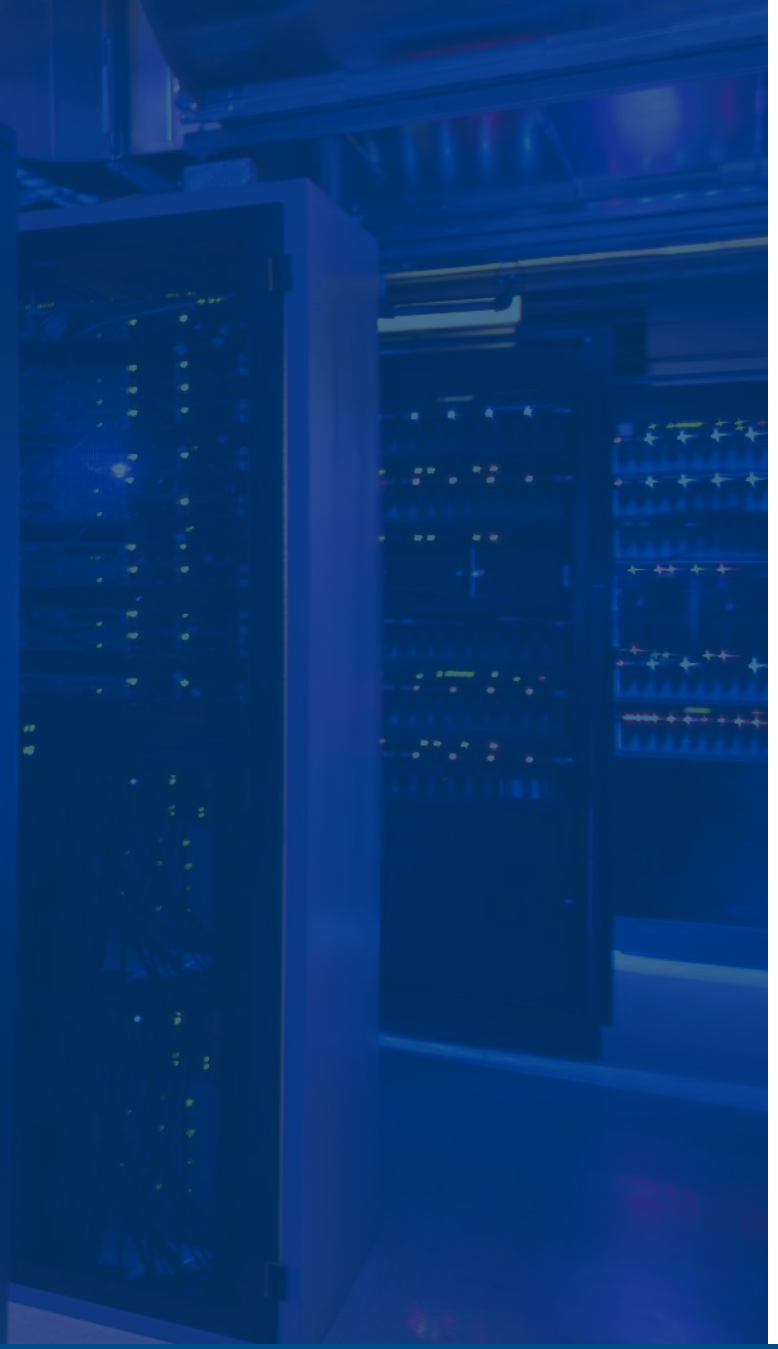
SPDK Energy Efficiency Improvement on Scheduler and Interrupt Mode

Liu Xiaodong

Senior Software Engineer
Intel

2021 Dec. 15th



Agenda

1. Why Energy (Power) Efficiency
2. Scheduler Module Overview
3. Schedulers
4. Interrupt Mode

SPDK achieves BEST IO performance

System Performance
System Efficiency
Ease of Development
For
Fabrics
Storage Media
& Everything In Between

Up to	7x	Better IOPS/core	NVMe-oF (vs Linux Kernel)
Up to	50%	Better Latency	NVMe-oF w/ Optane (vs Linux Kernel)
Up to	10x	Better IOPS/core	NVMe (vs Linux IO_URING)
Up to	14M	IOPS/core	NVMe
Over	11GB/s	Read bandwidth/core	NVMe-oF TCP target
Up to	3x	IOPS/core & Tail Latency	Virtualized Storage (vhost with standard Qemu)

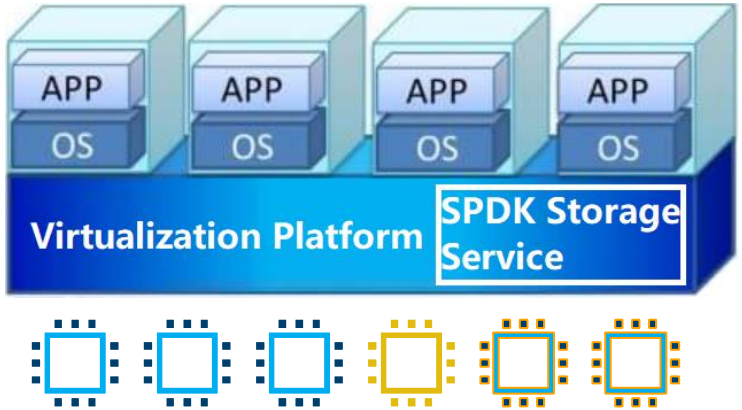
But,

See https://spdk.io for performance reports describing system configurations for these comparisons.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to http://www.intel.com/performance

Problem Statement

- In virtualization/container, SPDK vhost target will exclusively pre-occupy host CPU cores to serve VM and achieve expected storage performance



How to get this CPU occupancy elastic ?

- In storage offloading, SPDK target inside DPU will run out of permitted CPU resource to reach device-defined best IO performance.

How to meet device idle power consumption?

Why Energy (Power Consumption) Efficiency

Ever-changing IO Workload is the most common situation

Power consumption is one of the key matrix in system & device specification

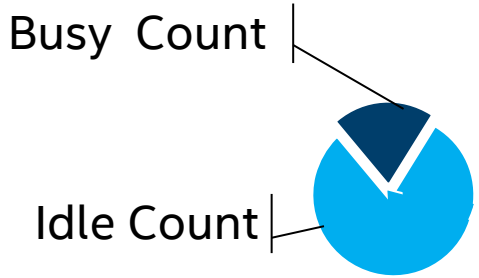
Meet the workload variation **sufficiently & efficiently**

SPDK Scheduler

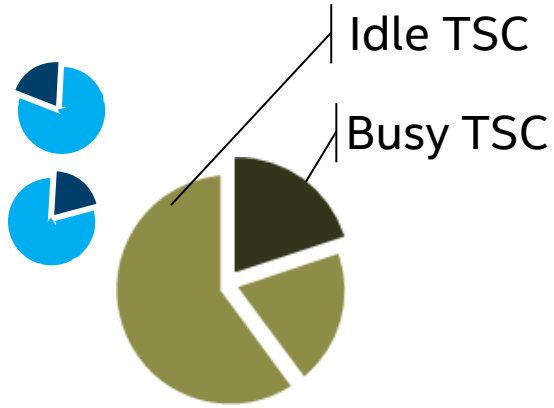
- Aiming to improve SPDK efficiency
- Abstracted as plugin modules from SPDK event/application framework
- Schedulers:
  static; gscheduler; dynamic scheduler
- Static Scheduler:
  Default setting
  None scheduling operations

Polling Profiling & Statistics

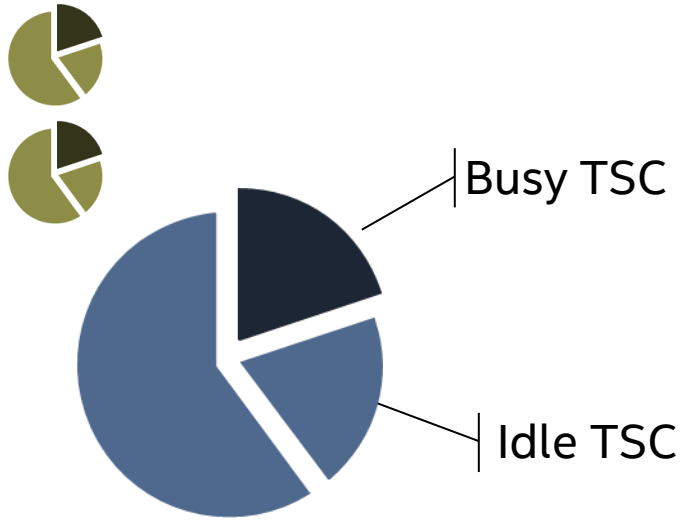
- CPU usage *V.S.* Effective CPU usage
- idle_tsc & busy_tsc



Spdk_poller



Spdk_thread



Spdk_reactor

gscheduler

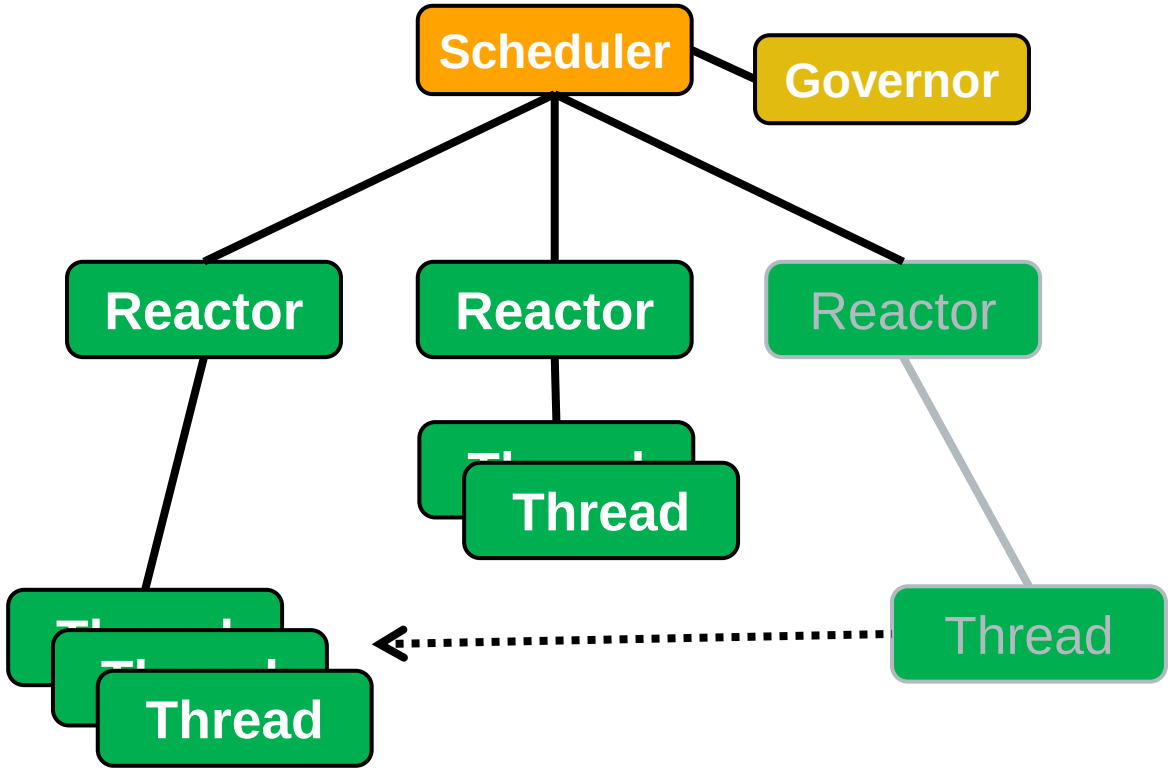
- Based on Intel Speed Select Technology (Intel SST)
- Set the frequency of specific core
  - Set MIN frequency if IDLE all the time
  - Set MAX frequency if BUSY all the time
  - Increase frequency if busy
  - Decrease frequency if idle
- Saving power directly and simply

But,

- Without consideration on CPU cost

Dynamic Scheduler

- More actions:
  - Spdk_thread moving
  - Reactor mode switching
  - Core Frequency setting
- Schedule Behavior
- Expectation
  - Automatic
  - Universal
  - Efficient



User Scheduler

- Customized with specific workload characteristic
- Traffic Prediction
  - Morning and Evening Rush Hours
  - Day or Night
- Business Value
- Workload Type

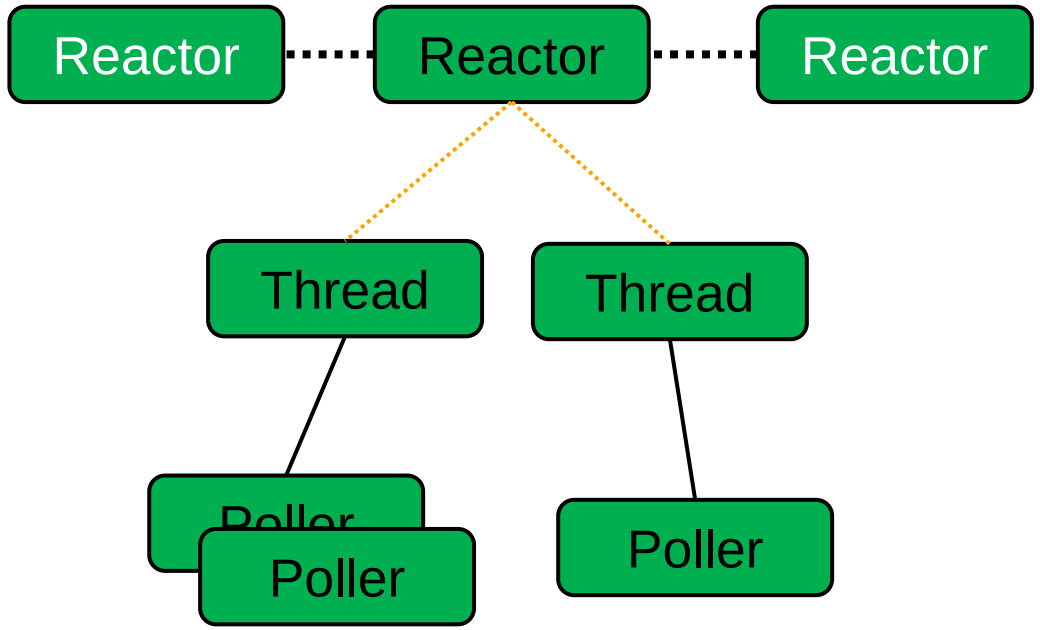
Interrupt Mode

- Targeted Situations:
  - High density of devices
  - **Lightweight workload**

- Running SPDK App/thread in interrupt mode (Event-driven)
  - Preempt CPU resource for upcoming IO workload
  - Yield to avoid CPU blank polling

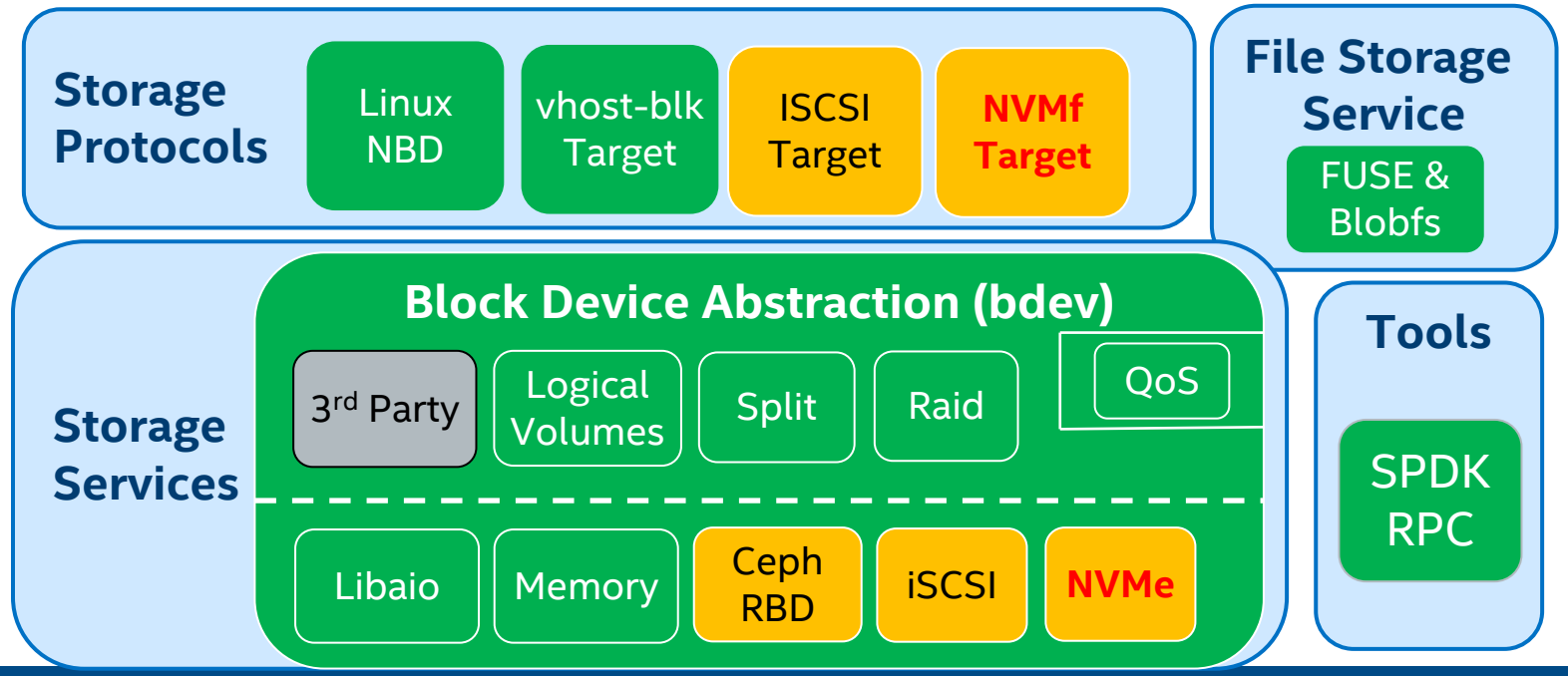
Interrupt Mode

- SPDK polling entities:
  - spdk_reactor
  - spdk_thread
  - spdk_poller
- spdk_thread can move dynamically between reactors
- spdk_poller belongs to specific spdk_thread inherently
- **Interruptibility**



Interrupt Mode

- Switch between polling and interrupt mode
  - Performance for heavy workload
  - Efficiency for lightweight workload
- Limited support (vhost-blk, nbd, aio)
- Continuing for a long time to get all SPDK libs & modules interruptable.



Summary

- SPDK Scheduler is supported to improve the energy efficiency of SPDK application
- Interrupt Mode can be used to balance performance and efficiency needs
- With Scheduler and Interrupt Mode, SPDK is solving the challenges, not only on performance

THANKS

Q & A


intel
SPDK, PMDK, Intel® Performance Analyzers
Virtual Forum