Evaluating 100Gbps Flash Disaggregation on ARM SoC

Minghao Xie, Heiner Litz, Chen Qian
Flash Memory Summit, August 6 – 8, 2019

Motivation
Can we achieve a competitive performance on ARM64 to Intel x86?

What are the challenges of porting an application from x86 to aarch64?

What is the TCO optimal platform for Flash storage disaggregation?

Overview
Flash Disaggregation enables sharing flash storage across the data center, improving resource utilization and reducing the total cost of ownership (TCO). The previous work ReFlex[2] tightly integrates the networking layer and storage layer by employing IX data plane architecture with the high-performance storage access framework SPDK.

With an energy-efficient and powerful smart NIC[3], we can think more... like offloading the processing tasks to free the host entirely. We want to build an integrated storage, computation and network platform over ReFlex and find a new tradeoff among these three. This NIC has 8 ARM A72 cores and it is so small and low-powered that using it as a PCI-e root complex with a fanout of several SSDs would largely lower down the previous OPEX in data centers. Without degrading much of the performance, we can come up with a flash disaggregation solution with a lower TCO.

Evaluation

Portability is one of the development goal throughout this project. Working on different architectures, it’s hard to reuse all the codes because we need assembly code in some critical areas. And available CPU features may vary. We’ve been modified the legacy codes and upgrade the APIs to the newest and make it portable across hardware/DPDK versions.

Performance is the thing we concerned most. We’ve achieved a decent performance with offloading features enabled. While good IOPS and throughput does not mean a good latency. In our 4k random read test, the system can enforce a 95th tail latency to less than 400 μs in a 400K IOPS test, which is still better than the conventional SSD read. And it can be scaled up to 2100k IOPS with 4 cores.

TCO is an important factor to build a datacenter. With current price, using this new smart NIC is not able to bring the benefits of CAPEX. But it does deliver better OPEX with lower energy consumptions and smaller space occupation. We will see better TCO if we can further optimize our system. With a more scalable design upcoming, the CAPEX would be very promising soon.

Reference: