



Toward Seamless Integration of RAID and Flash SSD

Sang-Won Lee

Sungkyunkwan Univ., Korea

(Joint-Work with Sungup Moon, Bongki Moon, Narinet, and Indilinx)



Table of Contents

- Introduction
- Bandwidth Requirements
- Lame RAID Phenomenon
- Striped Sequential Write
- Summary

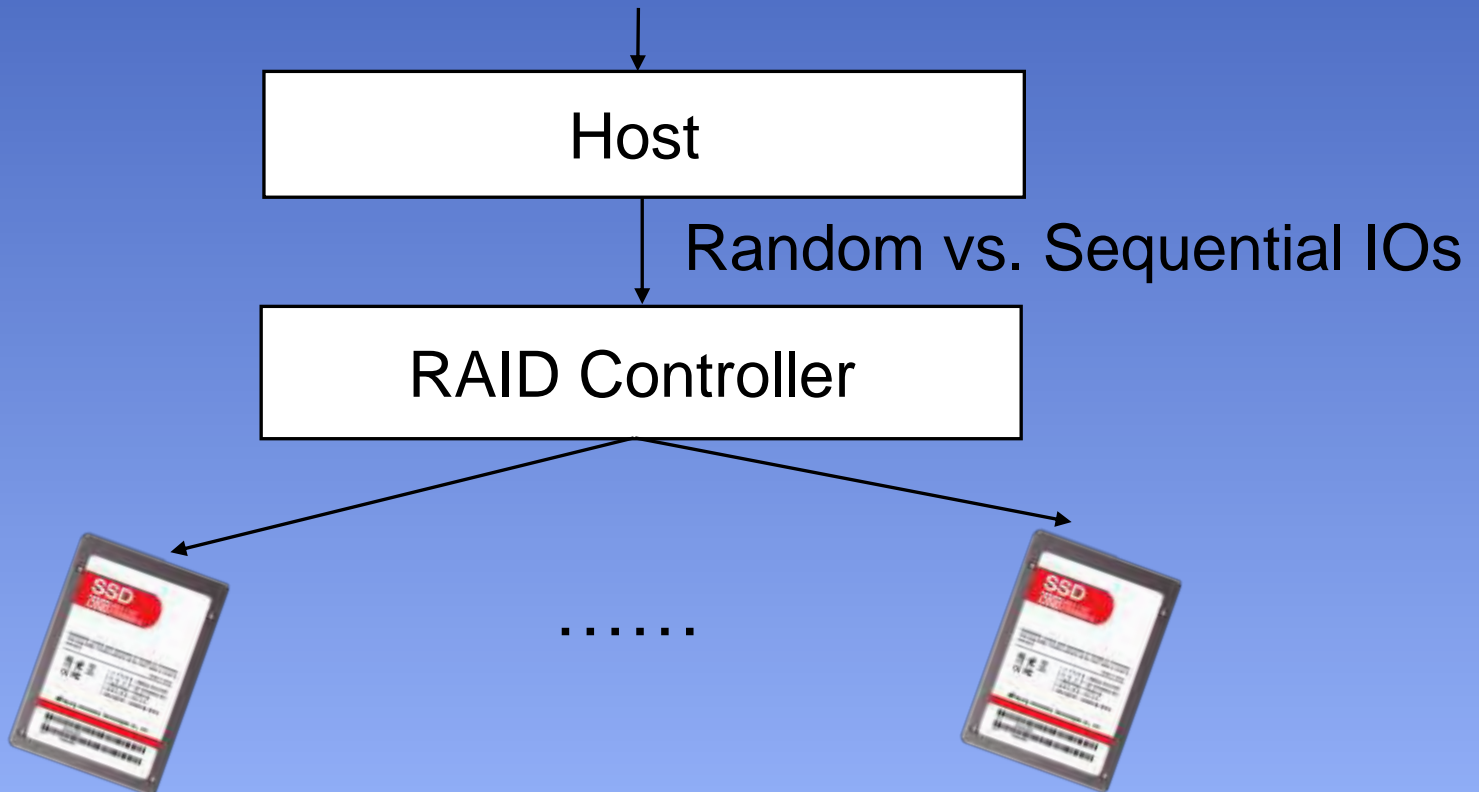
- Flash SSDs
 - Faster, bigger, cheaper SSDs are available
 - MLC, Multi channels/ways, FTL & controller
 - 1TB size SSD will soon appear
 - Will their capacity ever increase?
 - And, is it the right direction?
 - What if any flash chip inside SSD fail?

- RAIDed SSDs > One monster SSD
- Benefits of RAID
 - Bandwidth for sequential IOs
 - IOPS for random IOs
 - Reliability by redundancy
- Benefits of RAID would still hold in SSDs

- But, the chasm between RAID and SSD exist
 - RAID have been developed mainly for harddisks
 - SSDs have not been seriously tested under RAID
 - So, the performance of RAIDed SSDs might still be **suboptimal**

RAIDed SSDs: Workload Categorization

Single vs. Multiple Threads





RAIDed SSD: Workload Categorization

	Random IOs (e.g. 4KB unit)	Sequential IOs (e.g. 1MB unit)
Single thread	No chasm	No chasm
Multiple concurrent threads	No chasm	Chasm

Ultra Sequential Write Bandwidth Requirement

- Narinet MSTM-1000 Board



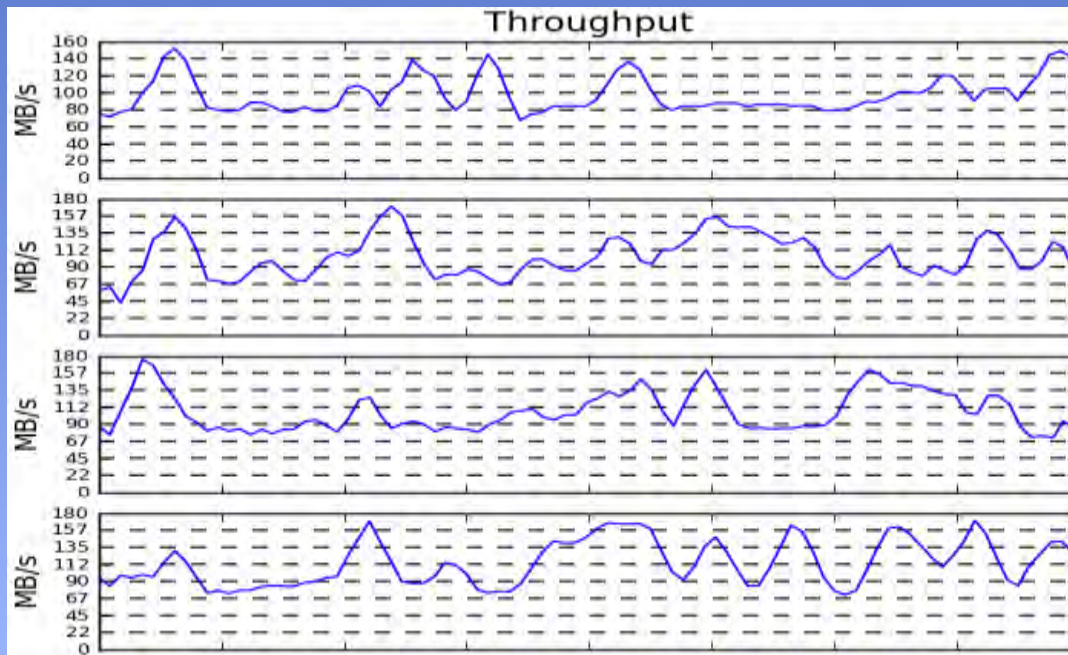
- Multicore CPUs captures network traffics in parallel
- Write the captured packet data in storage (1GB /sec)
- >> 6 Disks, 1.4GB/sec throughput SAS interface
- Multiple concurrent sequential write threads
- The captured data should be stored in one big file

Naïve Attempt

- Linux internal software RAID: mdadm
- MLC SSD:
 - 690\$, 256GB, 150MB/sec seq. writes
 - With 6 RAIDed SSDs, we expected 800MB/sec
- Real performance: **150 MB/sec**

Lame RAID Phenomenon

- Much slower than expectation: why?
- Although each SSD shows similar average throughput, each SSD experiences the worst performance at different points of time



Lame RAID Phenomenon

- The lame RAID phenomenon would be **nature** in low-end and medium class SSDs because of the limitations in SSD controller (e.g. SRAM resource, FTL, garbage collection)
- Related website:
<http://www.overclock.net/hard-drives-storage/620320-my-raid0-ssds-slower-than-individual.html>

Lame RAID Phenomenon

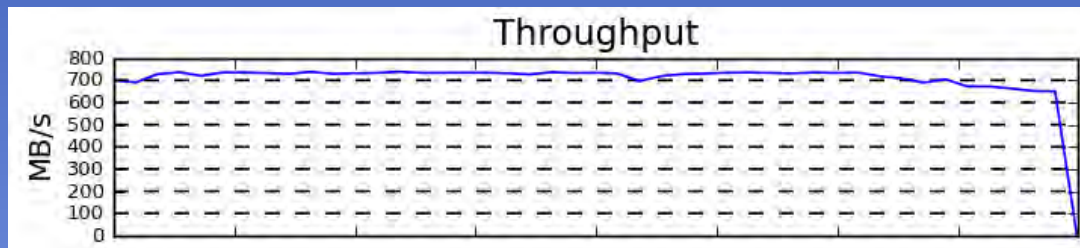
- Users of RAIDed HDDs also report a related issue
 - <http://www.tomshardware.com/reviews/HETEROGENEOUS-RAID-ARRAYS-WORK,1789.html>
 - Each component HDD has different average performance: e.g. 7.2K rpm, 10K rpm, 15K rpm

- But, this is technically different from our Lame RAID phenomenon

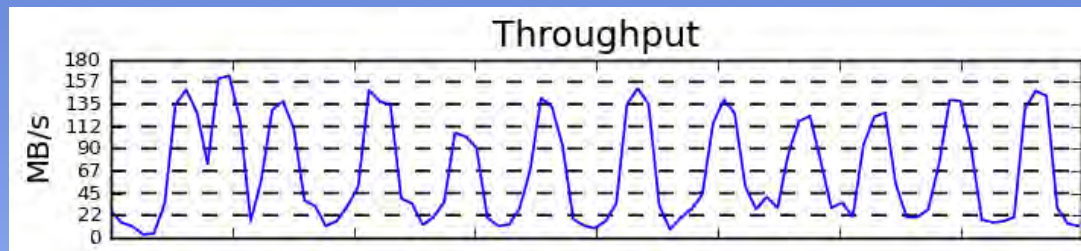
- Some solutions
 - Application-side
 - RAID controller-side
 - SSD side
 - Minimize the performance variance even by sacrificing the average performance
 - Uniform performance is a key virtue of storage devices

Lame RAID Phenomenon

- Performance of RAIDed SSDs with mild variance



High-end MLC SSD (4Disks)



Low-end MLC SSD (4Disks)

Next Attempt with High-End SSDs

- High-end commercial SSD
 - SLC based SSD (10 channel)
 - 170 MB/s
 - No lame RAID
 - 6 SSDs
 - Expectation $6 \times 170 \text{ MB/s} = 1020 \text{ MB/s}$

- Real performance: 670 MB

Striped Sequential Write

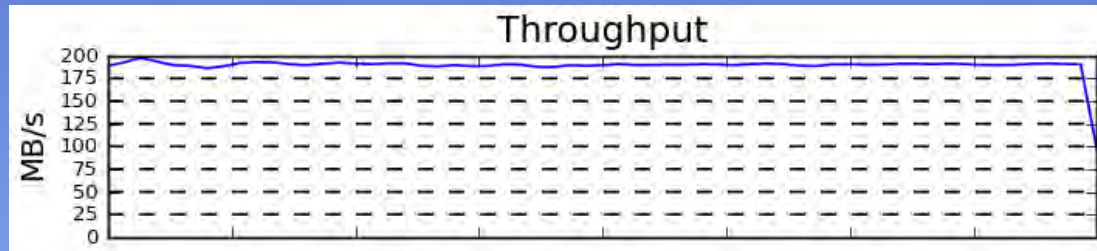
- Slower than expectation: why?
- Striped sequential writes from RAID controller
- Pure vs. sequential write bandwidth
 - HDD: almost same
 - SSD: quite different

Striped Sequential Write

- Some solutions
 - Application-side
 - RAID controller-side
 - SSD side

Striped Sequential Write

- INDILINX New Firmware
 - 180 MB/s
 - No lame RAID
 - 6 SSDs
 - Expectation $6 \times 180 \text{ MB/s} = 1080 \text{ MB/s}$



- Real performance: **950 MB**

- RAIDed SSDs would be popular; but RAID and SSD are not aware of each other
- For multiple concurrent sequential writes, RAIDed SSD reveals two issues
 - Lame RAID
 - Striped sequential write
- We presented some basic solutions, but a lot of things to do is ahead

Some Notes about this template

- The first action you should take is to save this presentation
 - You have opened a design template (.pot)
 - Need to save as .ppt
- A master exists for:
 - Slides
 - Handouts - default is 3 to a page
 - You can print a different number, but no guarantees about appearance
 - Notes



Free-format slide title