Faster Storage Devices Profiling with Parallel SeRRa

Jean Bez Francieli Boito Rodrigo Kassick
Vinícius Machado Philippe Navaux

Federal University of Rio Grande do Sul (UFRGS), Brazil

XIII WSPPD - July 21st, 2015
Porto Alegre, Brazil

Summary

Introduction

SeRRa

Parallel SeRRa

Performance Evaluation

Conclusion

Introduction

Hard Disk Drives

- Main non-volatile storage option
- Mechanical parts limit their performance
- Most Systems were developed or adapted in order to maximize their performance

RAID Arrays

- Another popular solution
- Combine multiple hard disks onto a virtual unit
- Data is striped among the disks and can be retrieved in parallel

Solid State Drives

- Recent Alternative
- No mechanical parts
- Lower power consumption

Performance differences

- SSDs and RAID arrays are inherently different from HDDs
- They cannot be treated as ”faster disks”

Spatial Locality

- HDDs perform better sequentially
- RAID arrays' performance are usually better with sequential accesses

Spatial Locality

- Works that aim to characterize SSDs reach different conclusions
- In some disks, there is no difference between sequential and random accesses
- Others have differences in orders of magnitude

Classifying Optimizations

- Optimizations cannot be classified between suitable for HDDs or SSDs
  - Sequential approaches can benefit both HDDs and certain SSDs
  - The optimization might not compensate its overhead

A New Tool

- We could classify these according to the sequential to random throughput ratio
- Obtaining this metric can be time-consuming
- We developed SeRRa to provide this metric as quickly as possible with small errors

Parallelizing

- Faster profiling would facilitate the tool's use for dynamic decision making
- SeRRa's accuracy can be increased by allowing more repetitions of its benchmark
- For these reasons, we developed a parallel version of SeRRa

SeRRa

SeRRa

- A storage device profiling tool
- Written in Python
- Provides the sequential to random throughput ratio of storage devices

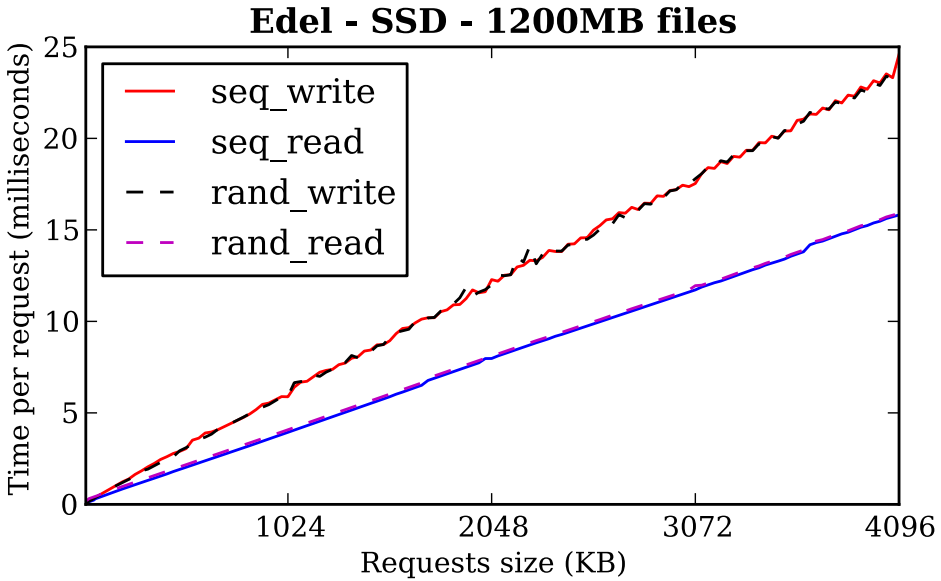
Goals

- Performance: Provide results as fast as possible
- Accuracy: Results must reflect the real behavior of the profiled devices
- Generality: The tool must be easy to use and do not require user-provided information about the device

Difficulties

- Keeping both performance and accuracy goals is challenging
- Profiling a storage device properly can take a long time
- Results depend on a system configuration
- Changes could require a new profiling

Solution



- Most of our tests' graphs present a linear function appearance
- Possible to estimate accurate results with linear regressions

SeRRa's execution steps

1. Monte Carlo
    - Request sizes inside a given interval are randomly picked

SeRRa's execution steps

2. Benchmark
    - Tests are executed for the picked request sizes

SeRRa's execution steps

3. Linear regression
   - Estimate the complete set of access times

SeRRa's execution steps

4. Report
    - Ratios for write and read tests are reported

Parallel SeRRa

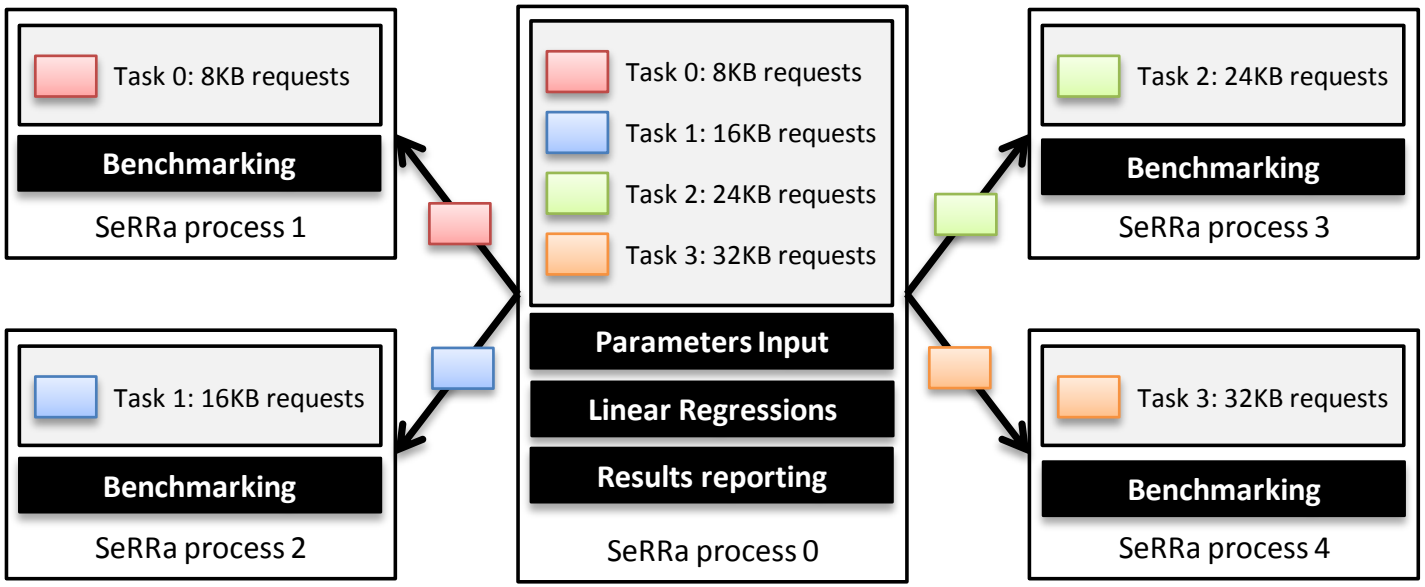
Motivation

- It is usual for HPC architectures to dedicate a set of nodes for the parallel file system deployment
- This shared storage infrastructure often uses identical storage devices on all involved machines
- These devices are expected to present the same performance behavior

Implementation

- A parallel implementation which benefits from this characteristic
- Parallelization with MPI4PY
- Master-Slave paradigm for communication between processors
- Parallelism is limited by:
  - Number of intervals
  - Measuring points per interval
  - Benchmark repetitions

Organization



- Only the benchmark step was parallelized
- Master is responsible for all other steps

Performance Evaluation

Tests' Environment

- Four different clusters from Grid'5000
  - Graphene and Pastel use HDDs
  - Suno uses an RAID-0 array
  - Edel uses SSDs

Profiling Time

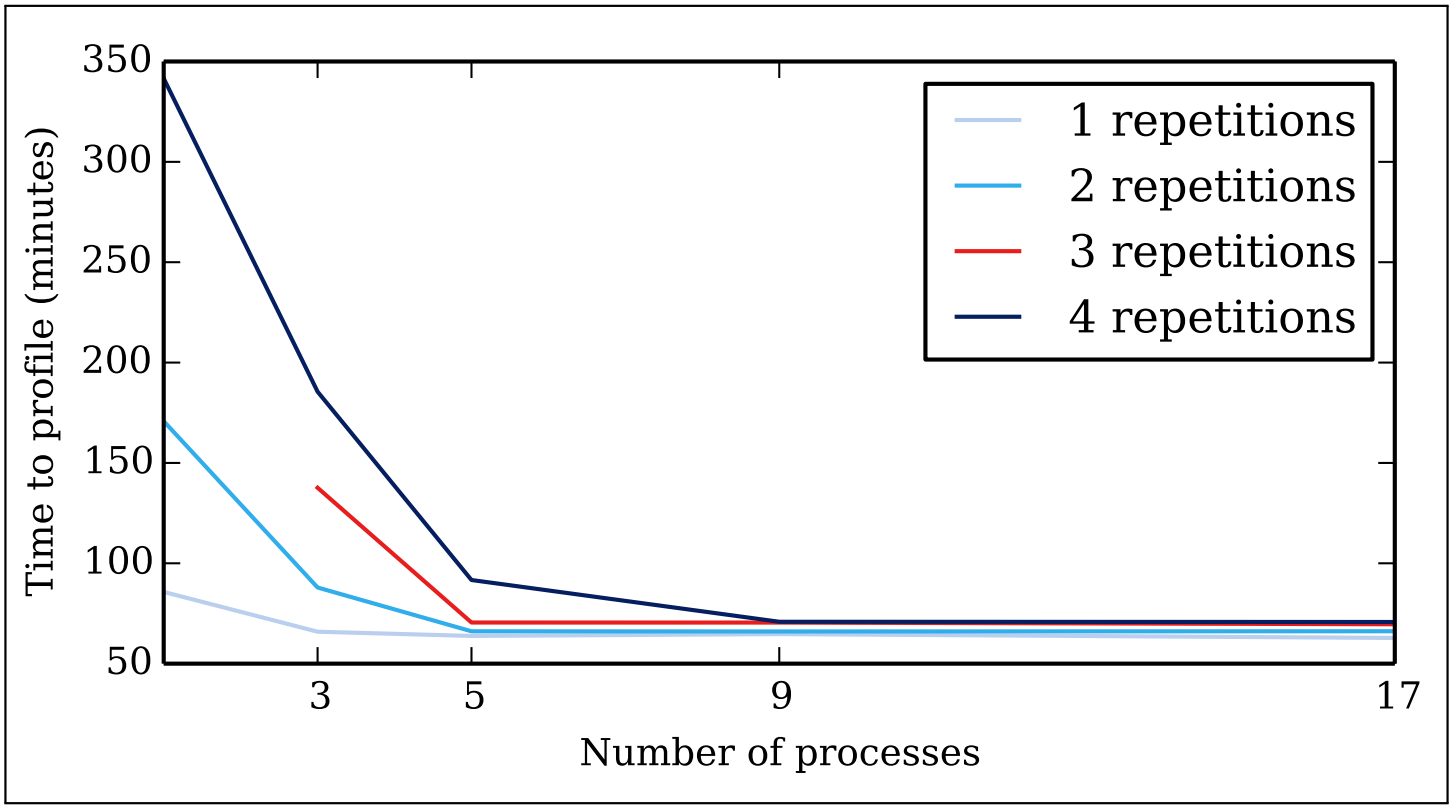
Time to profile (in minutes)

	No SeRRa	Sequential SeRRa		Parallel SeRRa
		1 repetition	4 repetitions	4 repetitions
Pastel (HDD)	19769.4	93.21 (1/212)	376.99 (1/52)	71.57 (1/276)
Graphene (HDD)	7222.8	85.69 (1/84)	341.50 (1/21)	70.75 (1/102)
Suno (RAID-0)	3679.2	28.12 (1/130)	109.42 (1/33)	21.56 (1/171)
Edel (SSD)	2426.4	5.92 (1/409)	23.58 (1/102)	2.72 (1/892)
Sum	33097.8	212.94 (1/155)	851.49 (1/39)	166.6 (1/199)

- Parallel SeRRa's times for 4 benchmark repetitions were faster than sequential SeRRa with 1 repetition

Profiling Time

Time to profile disks with Parallel SeRRa - Graphene cluster (HDDs)



Other curves omitted because of similarities

Analysis

- Performance increases with the number of processes
- This increase is limited by the number of benchmark repetitions

Analysis

- In most cases with 4 benchmark repetitions, there's no difference between using 9 or 17 processes
- Same for 2 benchmark repetitions and 5 or 9 processes
- Maximum speedup can be reached using a number of tasks which is twice the number of slave processes

Speedup

Speedup provided by SeRRa's parallel implementation (with the best number of processes to each case).

Benchmark repetitions	Pastel (HDD)	Graphene (HDD)	Suno (RAID-0)	Edel (SSD)
1	1.47	1.37	1.31	2.29
2	2.76	2.59	2.6	4.51
3	2.15	1.98	2.18	3.31
4	5.34	4.83	5.08	8.71

- Speedup increases with the number of benchmark executions
- Edel cluster has the highest speedups

Limitation

- Storage device presents a high sequential to random throughput ratio
- A Task consisting of many small, random accesses will take much longer than other possible tasks

Limitation

- This creates a situation of Load Imbalance
- The imbalance can impair speedup due to longer execution times

Analysis

Sequential to random ratio with 1200MB files for 8KB requests - measured vs. estimated with SeRRa tool (4 repetitions).

		Pastel (HDD)	Graphene (HDD)	Suno (RAID-0)	Edel (SSD)
Write	Measured	21.29	15.12	8.17	0.66
	SeRRa	22.62	15.21	8.42	0.67
Read	Measured	38.91	40.68	25.46	2.37
	SeRRa	39.08	40.65	25.51	2.38

- Edel cluster has the lowest ratios
- Load unbalance is also smaller
- Presents the highest speedups

Conclusion

Final Remarks

- We presented a parallel implementation of a storage device profiling tool named SeRRa
- SeRRa is available at http://serratool.bitbucket.org

Final Remarks

- We have evaluated our approach with four different clusters, using HDDs, RAID arrays and SSDs
- Our results show performance improvements
  - Up to 8.71 times over sequential SeRRa
  - Up to 895 times over not using SeRRa
- It is possible to achieve more accurate results with same profiling time