DON’T LET RAID RAID THE LIFETIME OF YOUR SSD ARRAY

Sangwhan Moon and A. L. Narasimha Reddy
Texas A&M University

sangwhan@tamu.edu  reddy@ece.tamu.edu
SOLID STATE DRIVE (SSD)

High performance
Low power consumption

MLC SSD ??

Cost per bit

Write Endurance
Reliability of MLC Flash Memory

* The measurement data from Hairong Sun et al., “Qualifying Reliability of Solid-State Storage from Multiple Aspects,” MSST’11.
Device Level Protection Scheme
  – Error Correcting Code (ECC)
  – Flash Translation Layer
    • Log-like Writing and Garbage Collection
    • Wear Leveling

System Level Protection Scheme
  – Parity Protection (RAID5)

These protection schemes require additional writes internally which in turn reduce the lifetime of SSD.
PARITY PROTECTION (RAID5)

• Protect a device array from a device failure
  – Protect each page group from a page error
• PARITY = XOR of ALL data
Parity Protection (RAID5)

1. Parity update results in additional writes
   – Write amplification: \([*N/(N-1), 2]\)

* N is the number of SSDs in an SSD array
2. Parity consumes more space
   – Higher space utilization reduces the lifetime

Parity Protection (RAID5)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Striping: 50%
- Parity Protection: 75%
Parity protection is supposed to improve the lifetime of SSD array

1) Parity update amplifies the number of writes by up to 2

2) Space overhead for parity protection initiates frequent garbage collection

Is parity protection beneficial or not in terms of reliability?
For the same number of SSDs given, to store the same amount of data, which is better in lifetime, striping (RAID0) or parity protection (RAID5)?
CONTRIBUTIONS

Markov models

We estimated the lifetime of SSD arrays.

Parity protection vs. Striping

Systems with different parameters are explored.

Our preliminary results

Parity protection *conditionally* provides benefit in lifetime over striping.
Page Error Rate Model at $x$ write count

- Bit errors accumulate until access time
- ECC detects/corrects the bit errors

$\begin{align*}
S \lambda(x) & \quad (S - 1) \lambda(x) \\
(S - K) \lambda(x) & \\
\mu + g(i, x)
\end{align*}$
• Source of failures
  – Page error
  – Device failure
• Any failure results in data loss in striping
• Parity protection loses data when
  – Two page errors *in the same page group*
  – Two device failures
  – Page error + device failure
  – Device failure + page error
Parity protected SSD array

Uncorrectable page error rate

Data loss

Device recovery rate

Page group recovery rate

Device failure rate

$Nu_sTv(x)$

$(N - 1)(v(x) + d)$

$(N - 1)Tu_sTv(x)$

$(N - 2)v(x) + (N - 1)d$

$v$: page error

$d$: device failure

$F_R$: data loss

The number of device replacement for device failure recovery
Mean Time to Data Loss (MTTDL)

- The expected time to encounter the first data loss in an SSD array

\[
MTTDL = \sum_{j=1}^{\infty} \left( j p(j) \prod_{i=1}^{j-1} (1 - p(i)) \right) t_w
\]

The probability of data loss at \( j \) write count

Time to write an SSD array
• SSD Parameters
  – 3x nm MLC flash memory
  – Capacity = 80GB
  – Page size = 4KB
    • ECC: 61-bit errors correctable BCH code
  – Annual device failure rate = 3%
  – TRIM command is exploited
Simulation Environment

• Simulation Parameters
  – The amount of data = 30GB/SSD
  – Workload
    • Read + Write = 125 MB/s/SSD
    • Read : Write = 3:1
  – 8 SSDs in an SSD array

• Relative MTTDL
  – The ratio of the lifetime of the target SSD array to that of single SSD with default parameters
Analysis of Single SSD

Lifetime decreases when space utilization increases.
Evaluation: Different Number of Devices

* The amount of data = 30GB/SSD

- Mirroring is a way worse than striping due to its penalties.
- RAID5 is potentially better than striping when ≥ 8 SSDs.
EVALUATION: DIFFERENT AMOUNT OF DATA

Considerable amount of overprovisioning is required for RAID5 to win over striping.
EVALUATION: DIFFERENT PARAMETERS

TECC, hotness, read intensive workload, or less intensive workload may help RAID5 win over striping.

![Bar chart showing lifetime of 8 SSDs with different parameters]

Reference | TECC | Hotness* | R:W=9:1 | 62.5MB/s/SSD

Relative MTTDL

0.0
0.5
1.0
1.5

Lifetime of 8 SSDs with different parameters


* 80% of workload accumulates on 20% of space
RAID5 is more effective on protecting SSDs with higher device failure rate.

Annual Device Failure Rate = 5%
Evaluation: Large Scale Storage System

RAID5 wins over striping in large scale storage systems

- RAID5 (4 SSDs)
- Striping

Relative MTTDL

1 RAID (4 SSDs)
5 RAIDs (20 SSDs)
10 RAIDs (40 SSDs)
15 RAIDs (60 SSDs)
20 RAIDs (80 SSDs)
25 RAIDs (100 SSDs)
50 RAIDs (200 SSDs)
• Parity protection is potentially worse than striping with small number of SSDs
• Parity protection wins against striping when
  1) considerably lower space utilization is guaranteed.
  2) TECC, hotness, read intensive workload, or less intensive workload is provided.
  3) SSDs have higher device failure rate.
• Parity protection wins against striping in large scale storage systems
• Other lifetime evaluations
  – Different write sizes
  – Other storage systems (e.g. RAID6)
• Monetary cost of ownership
• Validation of our analytic models
• Advanced techniques to reduce write amplification from parity protection
CONCLUSION

• Markov models to estimate the lifetime of an SSD array with protection schemes

• Lifetime comparison of striping (RAID0) and parity protection (RAID5) with different parameters

• Parity protection is conditionally superior to striping.
THANK YOU FOR LISTENING!

QUESTIONS?