Syracuse University

A Study of SSD Reliability in Large Scale Enterprise Storage Deployments

Omkar Desai

Advisor: Prof. Bryan Kim

Why this paper in particular

- Learn how to measure systems like a pro.
- Measuring and studying data can help find problems with existing systems
- “Measure, then build” – Prof. Remzi Arpaci-Dusseau.

Previous large-scale studies

- 4 large scale field studies from Google, Facebook, Alibaba, and Microsoft data centers.
- These pervious studies are performed on distributed data center storage systems.
- This study is performed on enterprise storage systems from NetApp.

Enterprise storage sys. vs distributed DC storage sys.

Enterprise Storage Systems

- High end, more reliable drives
- Reliability is achieved through RAID

Distributed Storage Systems

- Commodity hardware, often consumer class, over the shelf
- Reliability may be achieved through distributed storage, replication etc.

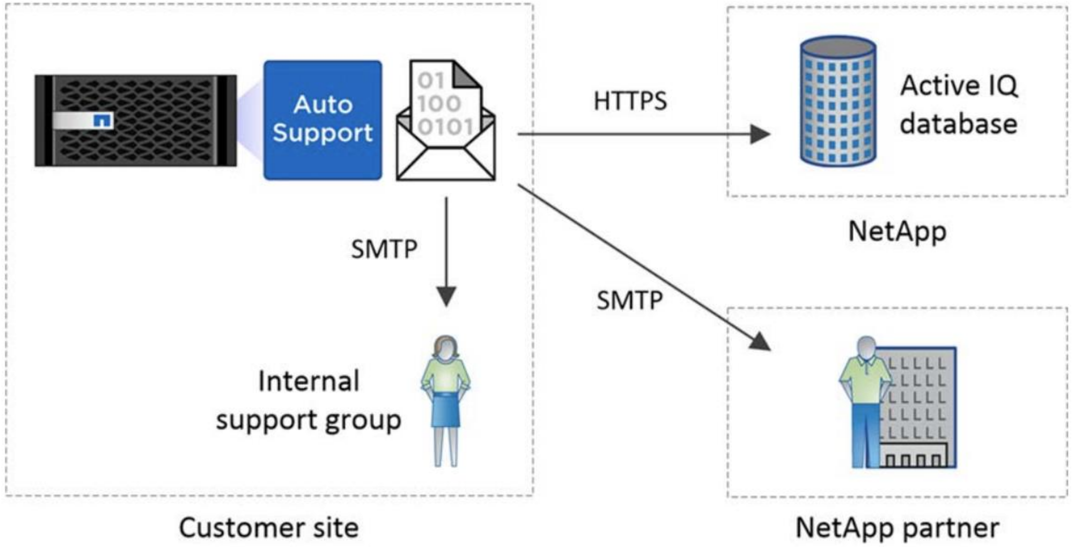
Abbreviations used:
SLC=Single level cell (more reliable and expensive)
MLC=Multi level cell (less reliable and inexpensive)
TLC=Triple level cell (less reliable and inexpensive)
e=enterprise class, c=consumer class

The dataset by the numbers

- 1.4 Million SSD’s
- 2.5 years of data
- SLC, cMLC, eMLC and 3D-TLC drives
- 3 Manufacturers
- 18 drive models
- 12 drive capacities
- A variety of wear, ageing, and system configurations

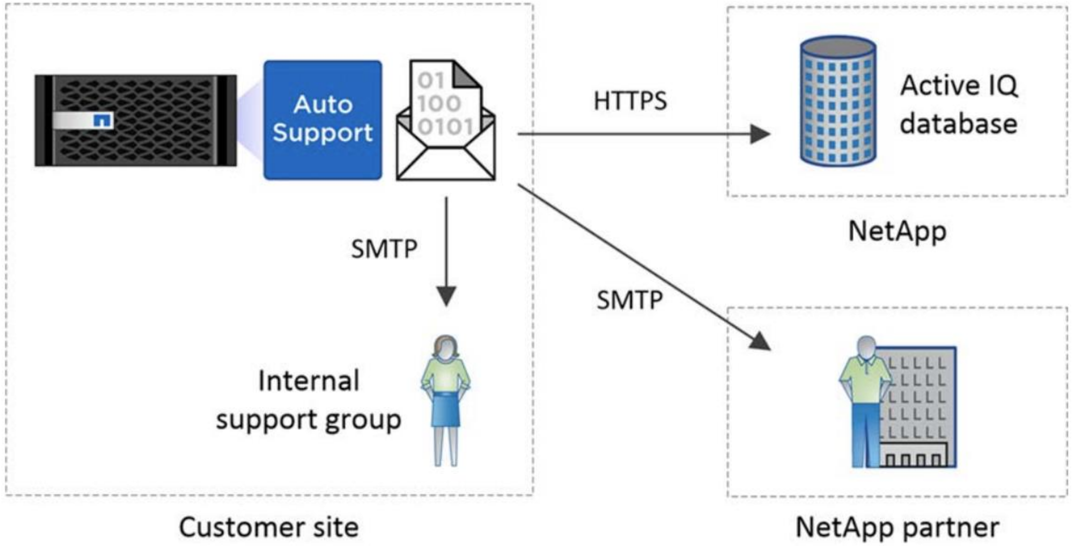
Data collection mechanism

- The data collected is telemetry data of a large scale of NetApp's systems deployed in field (on prem)
- Only metadata is collected
- The system sends weekly bundles which track a very large set of system and device parameters



Data collection mechanism

- Another dataset contains failure logs along with the diagnosed reason for replacements
- Metrics collected are:
  - Capacity, interface, flash tech, litho, rated PE life
  - Overprovisioning, first deployment, Drive power on years, %of life used
  - Spare blocks consumed, Bad sector count, ARR
  - More details on the next slide



Metrics collected

Drive Characteristics

Manufacturer & Model	Capacity (GB)	Interface	Flash Technology	Lithography	Rated P.E. Cycles

Usage Characteristics

Overprovisioning Factor	First Deployment	Drive Power Years	Rated live used (%)

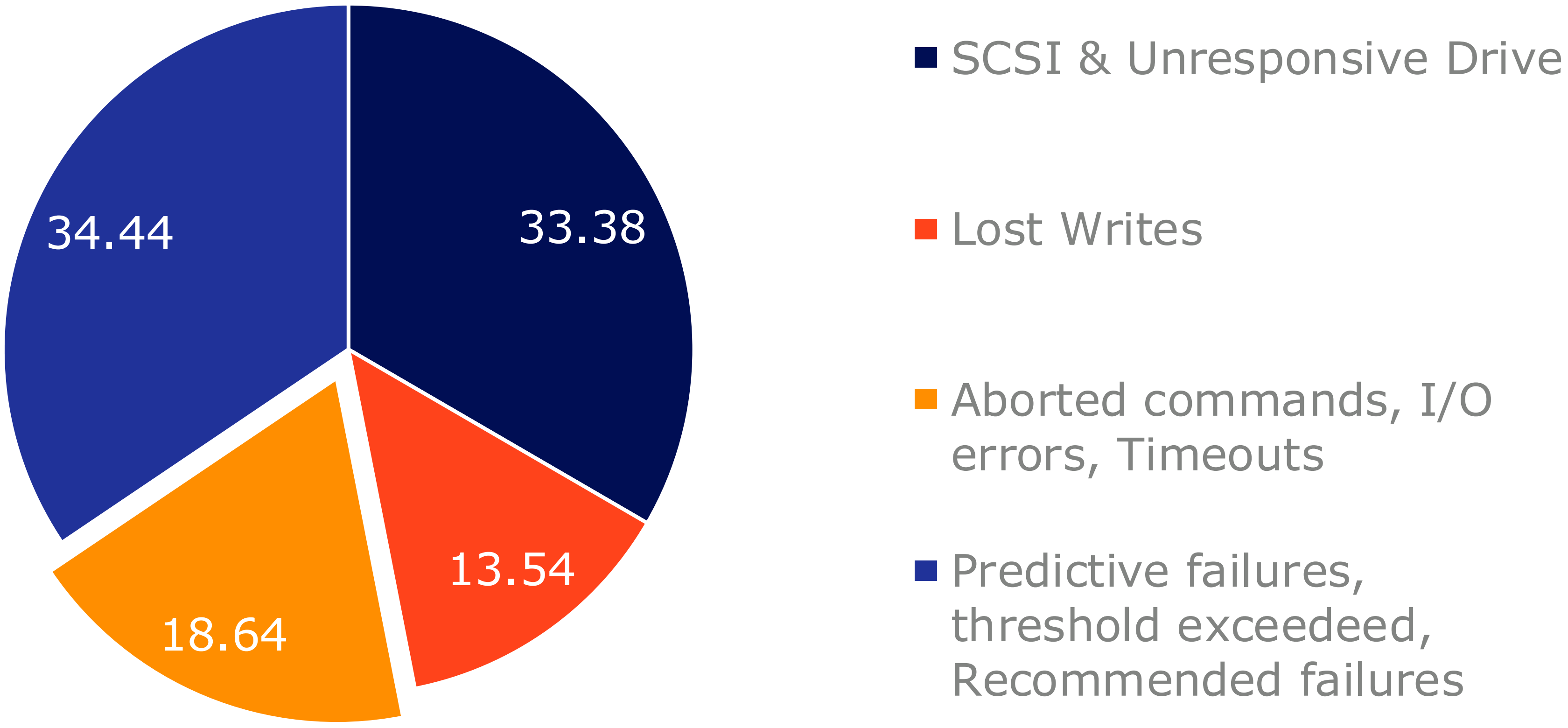
Reliability metrics

% of spare blocks used	Number of bad sectors	Annual replacement rate (ARR)

$$ARR = \frac{\#\ failed\ Devices}{\#\ Device\ Years}$$

ARR=Annual replacement rate

Reasons for replacement of drives



Reasons for replacement of drives

Severity

Category	Type	Percentage
A	SCSI error	**37**
	Unresponsive drive	0.6
B	Lost Writes	13.5
C	Aborted Commands	13.5
	Disk Ownership I/O Errors	3.2
	Command Timeouts	1.8
D	Predictive Failures	12.7
	Threshold Exceeded	12.7
	Recommended Failures	8.9

Factors impacting replacements

1. Usage and age
2. Drive type (e.g.: SLC, MLC, 3D-TLC etc.)
3. Capacity
4. Lithography
5. Firmware version
6. Number of bad blocks
7. RAID

Abbreviations used:
RAID=Redundant Array of Inexpensive disks

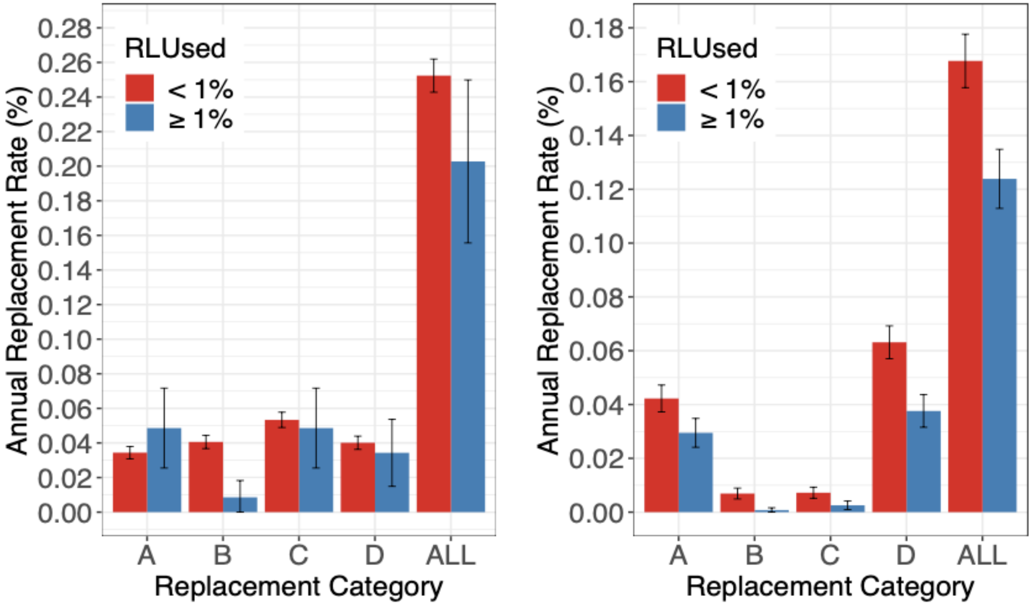
Factors impacting replacements

1. Usage and age
2. Drive type (e.g.: SLC, MLC, 3D-TLC etc.)
3. Capacity
4. Lithography
5. Firmware version
6. Number of bad blocks
7. RAID

Abbreviations used:
RAID=Redundant Array of Inexpensive disks

Usage as % of rated life

- SSD manufacturers rate an SSD for about 10-50K PE cycles
- Drives used <1% have a higher failure rate
- This shows the presence of **infant mortality**.
- Drives used >50% experienced 0 failed writes. Likely cause is firmware upgrades or failed writes being an infant mortality characteristic

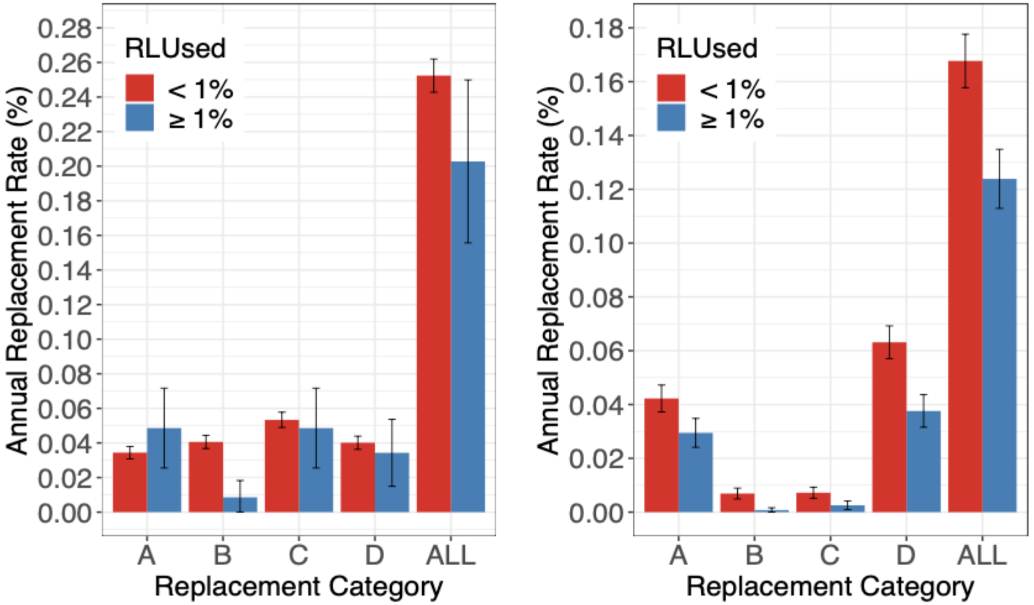


(a) *3D-TLC Drives.*

(b) *eMLC Drives.*

Usage as % of rated life

- Drives used >50% experienced 0 failed writes.
- Likely cause is firmware upgrades
- Another reason could be failed writes being an infant mortality characteristic
- Heavily used drives were also found to be replaced due to predictive failures (D).

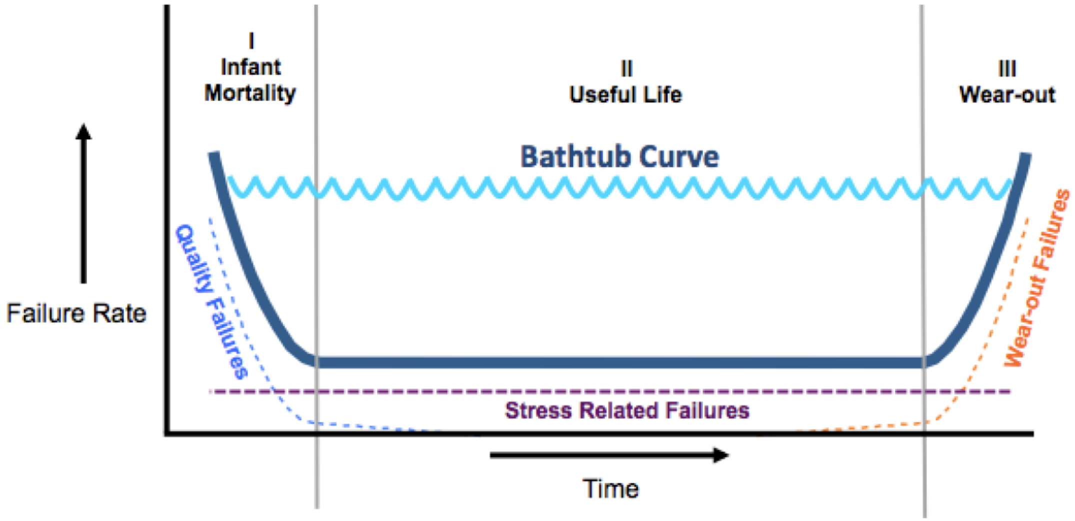


(a) *3D-TLC Drives.*

(b) *eMLC Drives.*

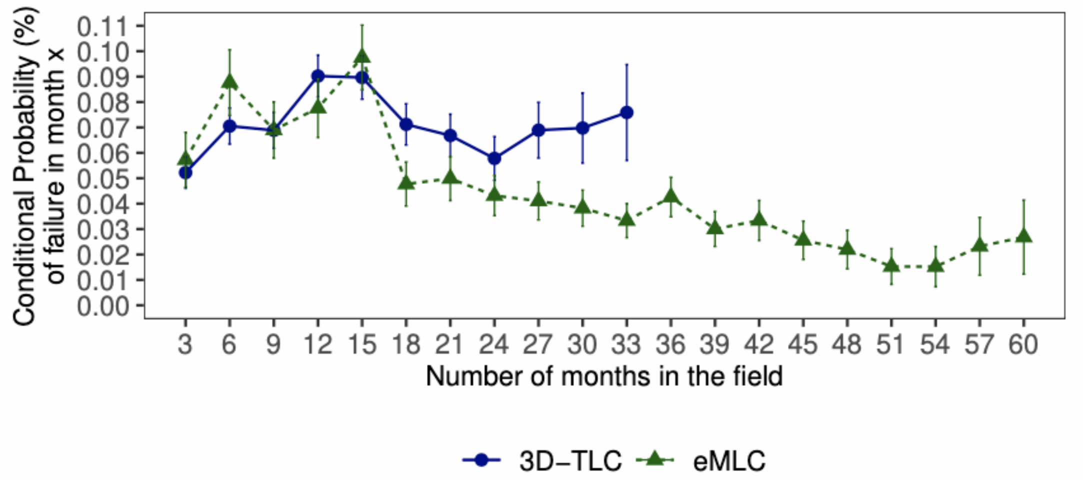
Drive age (in months)

- The commonly assumed model for drive ageing is the bathtub curve
- In the bathtub curve, there is a high failure rate initially, followed by a drop during midlife. Failure rate picks up again at the end



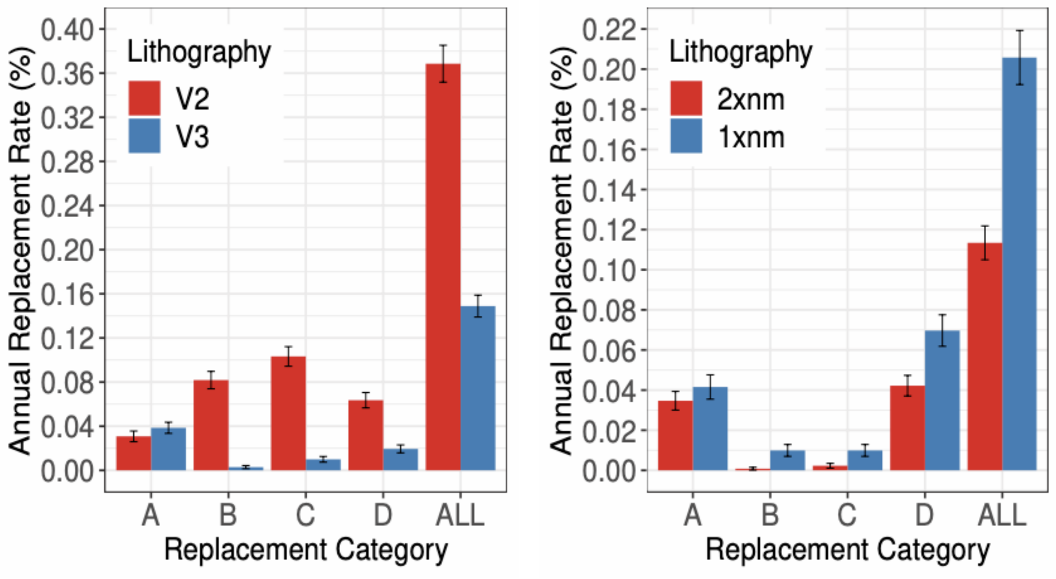
Drive age (in months)

- For 3D-TLC and MLC drives, the study experienced a long 12-15-month infant mortality period
- The next period was of slowly decreasing failure rates (6-12 months)
- Infant mortality is long! 20-40% of their lifetime
- An increased failure at the end of the graph is not seen as the drives are not very old



Lithography

- Here the common expectation is that higher density leads to more failures
- For eMLC drives, the results are as expected.
- Drives with a smaller lithography have a higher failure
- These drives have also consumed a larger number of spare blocks which is a sign of developing bad blocks

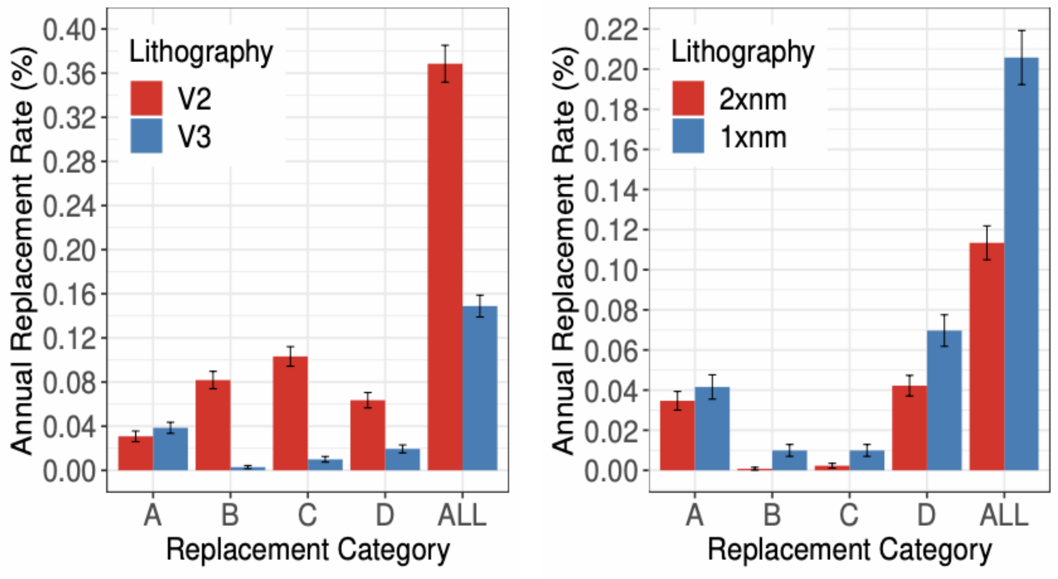


(a) *3D-TLC Drives.*

(b) *eMLC Drives.*

Lithography

- In terms of 3D-TLC drives, the higher density drives(v3) show lower replacement rates.
- This could be because 3D-TLC is new and vertical stacking in v3 is not yet reached saturation
- Only replacement category A (unresponsive drive) is unaffected by lithography

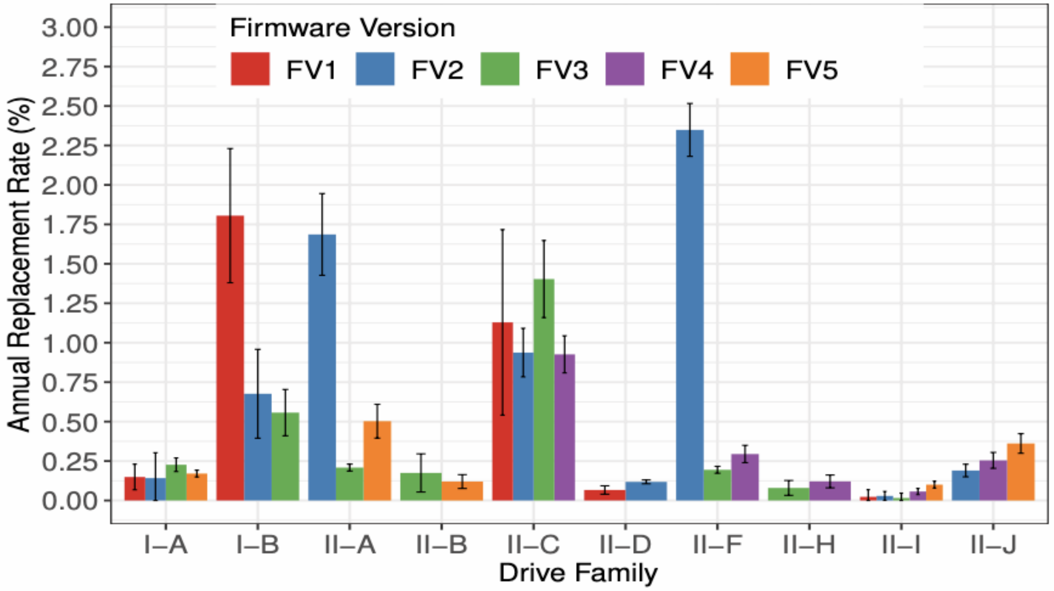


(a) *3D-TLC Drives.*

(b) *eMLC Drives.*

Firmware version

- The firmware version can have a huge impact on performance
- The reasons behind this are bug fixes, performance improvements, etc.
- In some rare cases, updating firmware decreases the ARR. Likely reason being new bugs

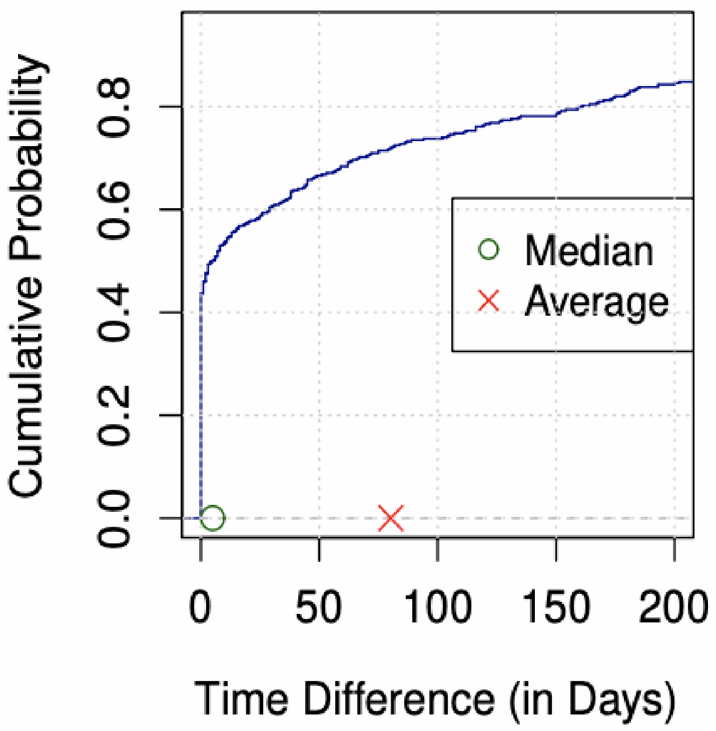


Correlations between failures in a RAID setup

Motivation:

RAID is in most places to improve redundancy. A failure of 1 drive should not be a cause for data loss.

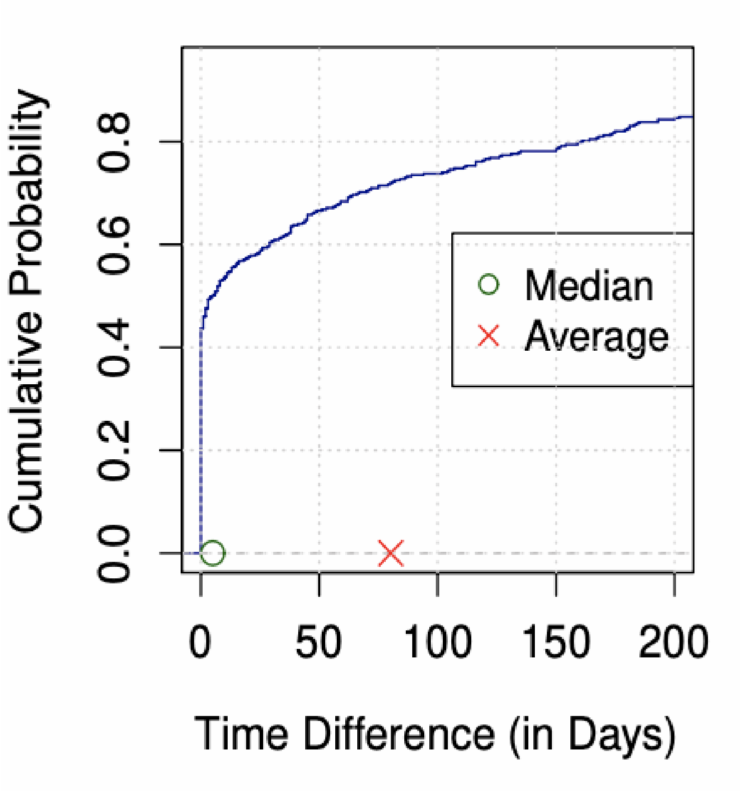
Understanding correlated failures becomes important as RAID has a maximum limit to the number of drives that can fail in an array before data loss happens



Time difference between successive replacements within RAID groups.

Correlations between failures in a RAID setup

- The probability that a RAID group will experience drive failure in a random week is 0.0504%. This depends on the number of drives in a group.
- The probability that a RAID group will experience drive failure within a week following a replacement is 9.39% (180x)
- 46% of replacements take place only a day after a replacement
- 52% consecutive replacements take place within a week



Time difference between successive replacements within RAID groups.

Concerns

- Studies based on the manufacturer rated life of SSD's may not be very accurate as manufacturer ratings tend to be inaccurate.
- Looking at performance degradation along with failures could reveal some interesting facts about SSD's in enterprise storage. (Probably out of scope for this study)
- The study measures the effect of age in months on failures. However, using enterprise storage which is always on and in use to study this would not give the right distinction between the effects of use and age of the SSD.
- Information on how much time the drives are ON and how much time they spend being used would be helpful.

Takeaways

1. Optimizing for easy firmware upgrades is important
2. Most critical failures (category A) would be a result of infant mortality as failure predictions get better over time as the drives age.
3. RAID mechanisms are vulnerable to consecutive failures and this study observes that the possibility of a 2nd drive failure after the 1st one is 180x more
4. As enterprise storage products tend to use higher quality products, the overall ARR is lower (0.22%) than google’s DC drives (1-2.5%)
5. Despite of all the failures, SSDs are still much more reliable than HDDs (ARR of ~20%)

References

- https://www.usenix.org/sites/default/files/conference/protected-files/fast20_slides_maneas.pdf
- Stathis Maneas, Kaveh Mahdaviani, Tim Emami, Bianca Schroeder. A Study of SSD Reliability in Large Scale Enterprise Storage Deployment. 18th USENIX Conference on File and Storage Technologies (FAST 20). https://www.usenix.org/system/files/fast20-maneas.pdf
- https://www.youtube.com/watch?v=agZLbknJ-kM&feature=emb_title&ab_channel=USENIX
- https://www.netapp.com/pdf.html?item=/media/17080-tr4699pdf.pdf

Syracuse University

Q & A

Omkar Desai
odesai@syr.edu