

Selective scrubbing based on algorithmic randomness



Rahul Vishwakarma[○], Bing Liu⁺, Peter Gatsby[○], Jinha Hwang[○]

⁺Dell Technologies, [○]California State University Long Beach

Introduction

- Frequently scrubbing the entire storage array during high incoming IO on a system can degrade the performance of the individual disks
- Decide "which disk to scrub" and "when to scrub"

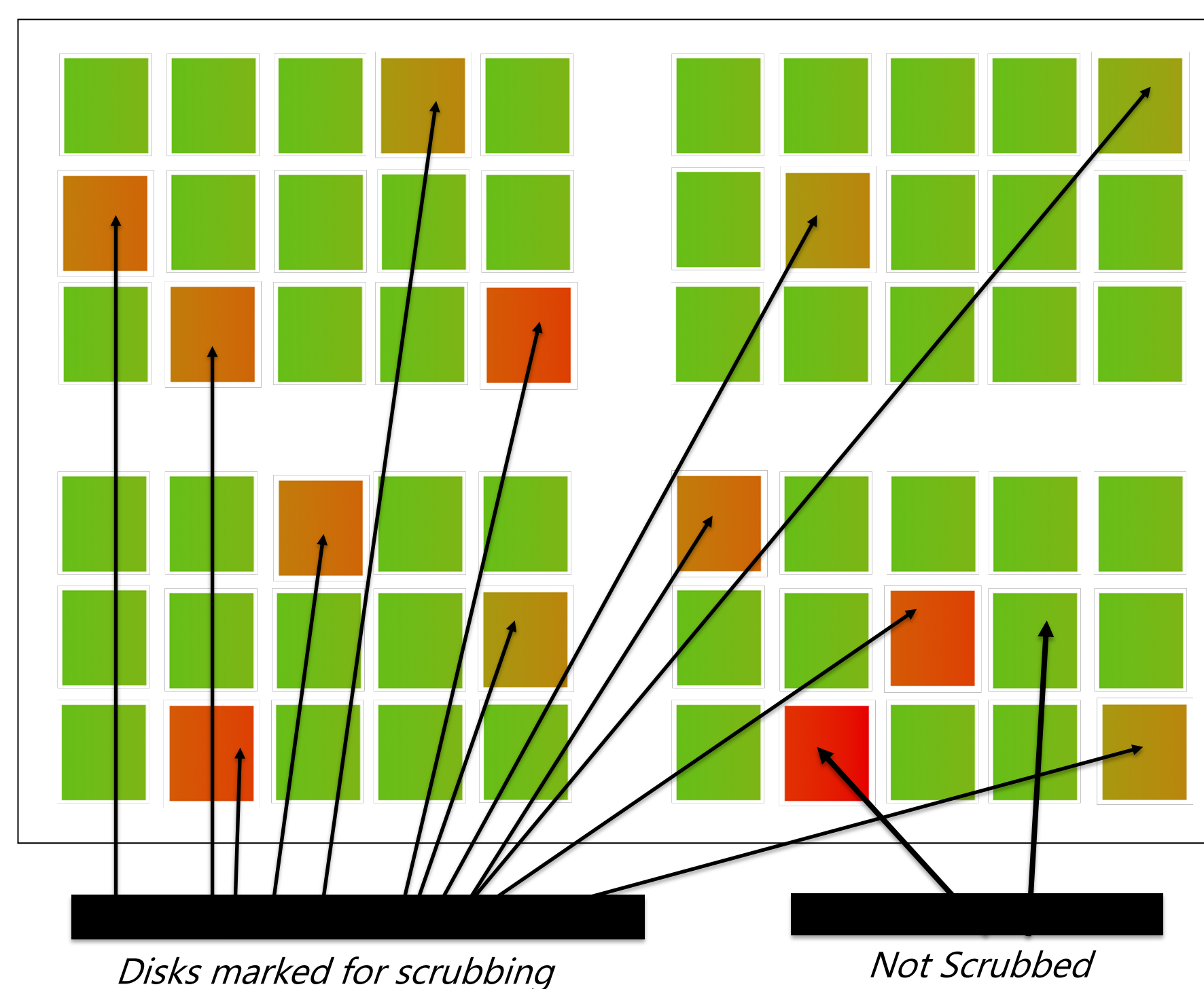


Figure 1. Selective scrubbing of least healthy disks

Method

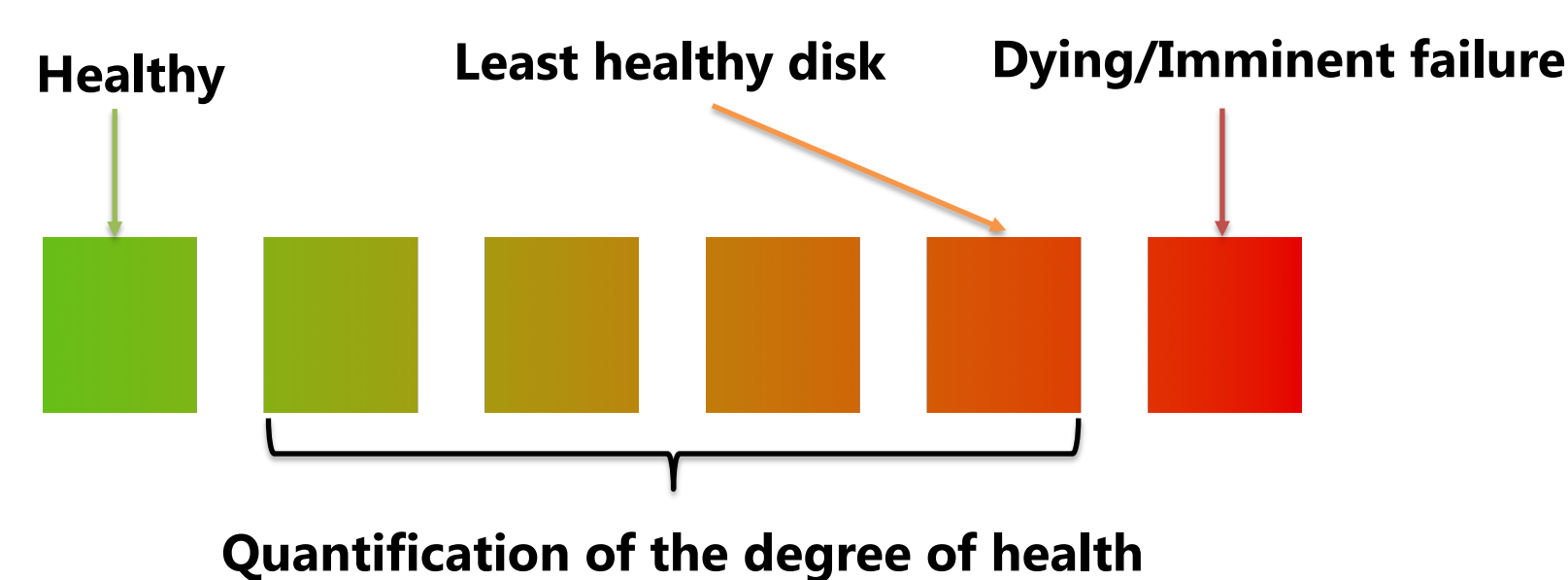


Figure 2. Drive health quantification based on conformal prediction

- Rank the health scores based on confidence
- Create a set of relatively least healthy disks for scrubbing
- Map the selected disks with scrub frequency cycle

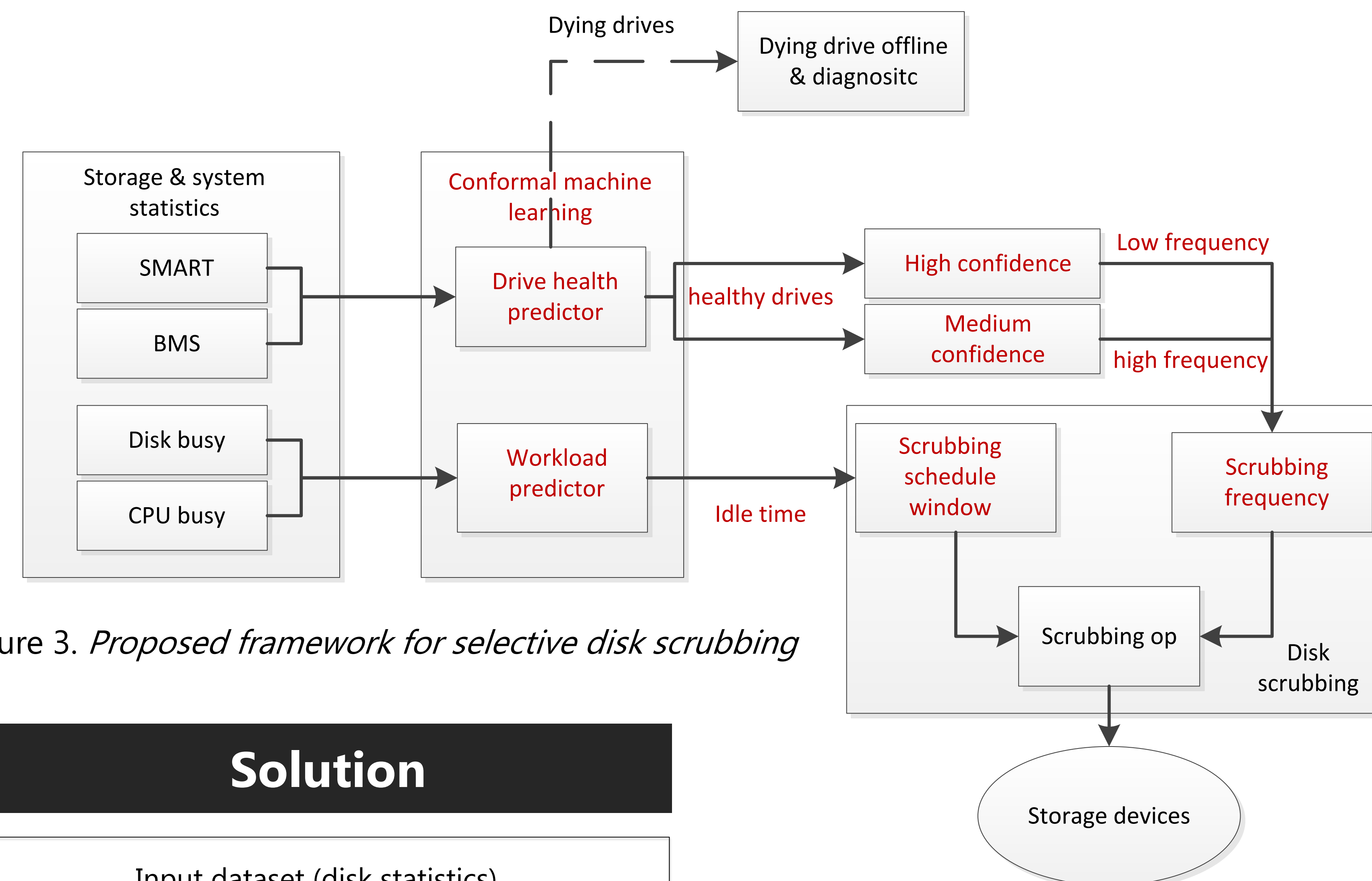


Figure 3. Proposed framework for selective disk scrubbing

Solution

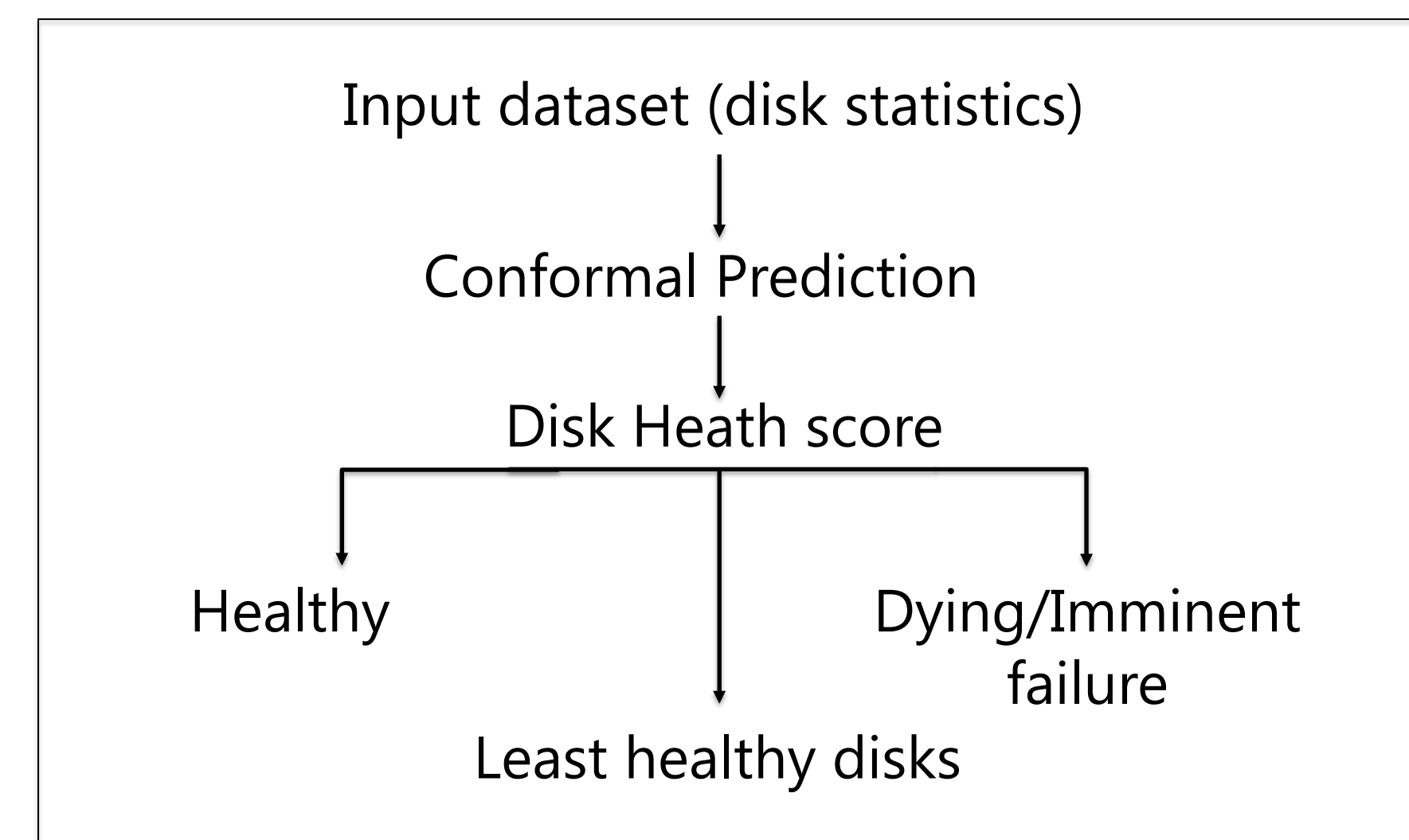


Figure 4. Disk health classification

Disk Serial	Health Rank	Status
Z296Z3KN	0.934426	HEALTHY
Z296Z3A6	0.819672	HEALTHY
Z296Z5MQ	0.754098	HEALTHY
Z29987ZE	0.474212	HEALTHY
Z29609UI	0.321424	HEALTHY
Z29IUW4	0.287129	HEALTHY

Figure 5. We quantify the degree of health of the disks and rank them as per the confidence of the prediction.

Disk Health	Scrub Frequency	Heath
Best	LOW (CYCLE - A)	1 - 0.7
Medium	MEDIUM (CYCLE - B)	0.7 - 0.3
Poor	HIGH (CYCLE - C)	0.3 - 0

Figure 6. Disks are segregated based on their health and assigned scrub frequency.

Disk Serial	Health Score	Status	Scrub Frequency
9330KBD2	0.934426	BEST	CYCLE - A
9330EHZU	0.819672	BEST	CYCLE - A
9330KB5X	0.606557	MEDIUM	CYCLE - B
9330K3D5	0.491803	MEDIUM	CYCLE - B
09330K6PW	0.360656	POOR	CYCLE - C
...
ZSE98JUJS	0.023149	POOR	CYCLE - C

Figure 7. Implementation of selective scrubbing based on algorithmic randomness.

Contribution

Method

- Conformal Prediction for quantifying the disk health

Optimal Scrubbing

- Select only specific disks which really need scrubbing based on their health

Performance

- Automated scheduling of disk scrubbing based on n-step ahead system load prediction using probabilistic weighted fuzzy time series

Advantages

Enhanced storage reliability

- More frequent scrubbing on poor healthy drives
- It helps detect and correct sector errors in advance

Improved performance

- Healthy drives are scrubbed less often freeing up resources for user workloads

Acknowledgement

The project has been supported by Dell Technologies, Israel (**Yosef Shatsky**) and Dell EMC China (**Bing Liu**).