

# Planning report

1155078701 Njo Ling Fung 1155078895 Chan Paul

Topic: Machine Learning for Storage system Reliability Prediction (I)

Code: PCL1803

## **Project goal:**

Explore the use of machine learning in improving system reliability by predicting the relationship hard disk drive sector error and improving storage system reliability through the use of sector error predictors. Meanwhile, Develop practice tools for improving system reliability or improving the predicting accuracy beside the experimented algorithm.

## **Significance of the project:**

Storing data reliably is one of the key requirement of a storage system. However, storage system is not always reliable due to partial drive failures, where individual sectors on a drive cannot be read. There is the chance of data cannot be recovered when the affected sector is too corrupted to be corrected by drive-internal error correcting codes (ECC). Many previous studies showed that both solid state drives and hard disk drives experience sector errors at a significant rate.

Currently, countless of data in the whole world is under the threat of sector error. And when there are more sectors on a drive, with the larger amount of storage, a higher possibility to occur sector error. With the growing consumption demand in data storage, storage systems' capacity will increase in the future, the concern of sector errors will be even more serious.

It is not easy to protect data against sector error as a drive will not ware of the error until the affected sector is being accessed. Storage systems need to proactively periodically read and verify data (a process called disk scrubbing), to discover the error and recover it through redundancy in the system. However, this process is time consuming and affect the performance of the storage system. The efficiency of the process can be improved if errors were predictable.

In this paper, the accuracy of predicting sector error in machine learning will be examined. It could increase the performance of disk scrubbing and the result will be applied to improve the system reliability.

## **Problem statement:**

- 1) Predicting the sector error accurately.
- 2) Decreasing unrecoverable error with lower cost in system performance

**Proposed solution:**

- 1) Predict sector error through various machine learning algorithm(classification and regression trees (CART), random forests, support vector machines, neural networks and logistic regression) against the SMART parameters of storage device.
- 2) Make use of the result in disk recovery (e.g. scrubbing)
- 3) Create a practical tools to reduct the effect of sector error

**Proposed timeline:**

Week	Target	Date
1	Download and analyze the data from Blackblaze	24/9/2018
2	Setting up environment for machine learning e.g. tensorflow	1/10/2018
3	Passing the data to simple machine learning model to test	8/10/2018
4	Evaluate and make the model more sophisticated	15/10/2018
5	Testing more models on different data set, may make use of more powerful server	22/10/2018
6	Mid term progress report	29/10/2018
7	To be decided	5/11/2018
8	To be decided	12/11/2018
9	To be decided	19/11/2018
10	Term End Report	26/11/2018