SATA Port Multipliers Considered Harmful

Peng Li† (Student), James Hughes‡, John Plocher‡, David J. Lilja†
†*University of Minnesota - Twin Cities,* ‡*FutureWei Technologies*

Abstract

By using SATA port multipliers (PMs), the storage capacity of a system can be increased significantly. However, there has been little work studying the reliability of the SATA PMs. To study the potential reliability issues, this work proposes a recreatable process to create actual failures of hard disk drives (HDDs). Experimental results show that at least one combination of SATA controllers, the SATA PMs, and the HDDs does not provide resiliency when a single HDD fails.

1 Introduction

SATA connectivity normally consists of a single drive connected to a single SATA controller port via a single cable. As a result, the maximum number of drives in a storage system is limited by the number of available SATA controller ports. The SATA port multiplier (PM) can be used to expand the drive scalability of storage systems cost-effectively. However, there has been little work studying the reliability of the SATA PMs. As storage systems reach the petabytes and exabytes era, system reliability becomes more challenging. The consequence of ignoring the reliability issue of the SATA PMs and using them for a large scale storage system could be catastrophic.

The goal of this paper is to study the reliability of the SATA PMs by conducting several experiments. Our experimental results show that, if one of the HDDs connected to the SATA PM fails, the other HDDs will not be accessible. Architecture designers should consider this effect when designing a system using the SATA PMs. The contributions of this paper are: (1) a demonstration of a reproducible process for creating actual failures in the HDDs for reliability testing; (2) a demonstration that at least one combination of the SATA controllers, the SATA PMs, and the HDDs does not provide resiliency when a single HDD fails. This paper does not explain why this error occurs. Rather, its contribution is showing that this type of error can occur in real systems. It also demonstrates a successful testing methodology. Section 2 presents the detailed test approaches and the results. Conclusions are drawn in Section 3.

2 Experimental Methodology and Results

We first do experiments on a system which does not contain the SATA PM as shown in Fig. 1. Then we do the same experiments on the system which uses the SATA PM as shown in Fig. 2. The Marvell board has two SATA controllers, which are SATA0 and SATA1. In Fig. 1, the HDD connected to SATA0 is the OS HDD, on which we install the OS and the **fio** program. The test results also are recorded on this HDD. We use the Seagate enterprise HDD as the OS HDD. The HDD which is connected to SATA1 is used for tests. During the experiments, we remove the cover of the test HDD to emulate fatal errors. Both Western Digital consumer HDDs and Seagate enterprise HDDs are used as the test HDD.

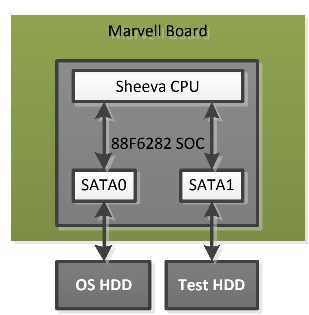

Figure 1: The system without the SATA PM.

The software flow of the first test system is shown in Fig. 3. Tasks 1 and 2 are generated by the **fio** program. They issue random I/O operations to the two HDDs si-

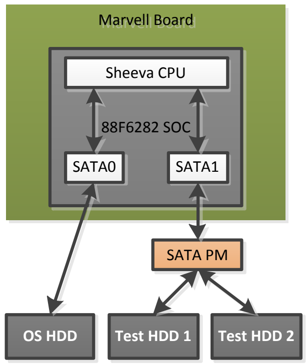

Figure 2: The system using the SATA PM.

multaneously. While the **fio** program is running, we remove the cover of the test HDD to emulate fatal errors. The test HDD stops working quickly after its cover is removed due to environmental contamination. In the entire process, we use the Linux command 'iostat' to record the I/O status of the two HDDs on the OS HDD.

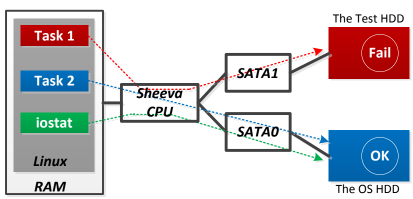

Figure 3: Software flow of the system in Fig. 1.

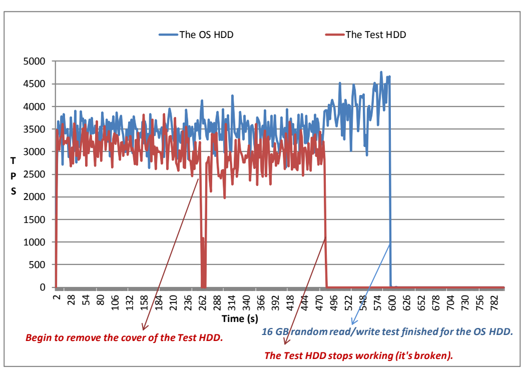

Figure 4: Test results of the system in Fig. 1 when using the Seagate enterprise HDD as the test HDD.

The Seagate enterprise HDD is first used as the test HDD. Fig. 4 shows the number of transfers per second (TPS) for the two HDDs as the test progresses. We started to remove the cover of the test HDD at about 260s. The HDD stopped working after the entire cover was removed (at 480s). It can be seen that the TPS of the OS HDD was not affected at all during the entire test process. It finishes all the required I/O operations issued by the **fio** program. We then repeat the same experiment using the Western Digital consumer HDD as the test HDD, and get the same results.

The software flow of the second test system is shown in Fig. 5. Similar to the first experiment, Tasks 1 and 2 are generated by the **fio** program. While the program is running, we remove the cover of the test HDD 2 to emulate fatal errors.

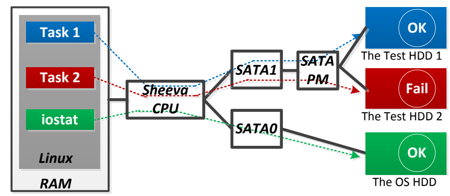

Figure 5: Software flow of the system in Fig. 2.

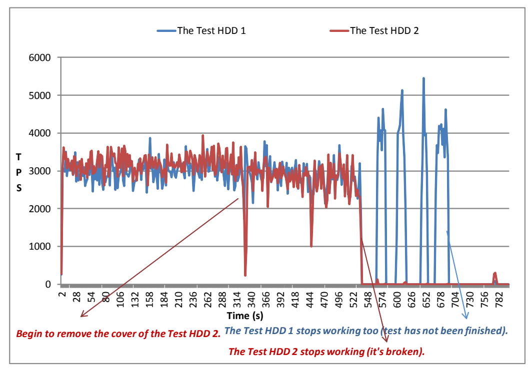

Figure 6: Test results of the system in Fig. 2 when using the Seagate enterprise HDD as the test HDD.

We first use the Seagate enterprise HDD as the test HDDs, with results shown in Fig. 6. We started to remove the cover of the test HDD 2 at about 320s. The test HDD 2 stopped working at about 520s after its cover had been removed due to environmental contamination. It can be seen that the test HDD 1 also stopped working at about 680s without finishing the required I/O operations issued by the **fio** program, i.e., Task 1 shown in Fig. 5. We next use the Western Digital consumer HDD as the test HDD and get the same experiment results.

3 Discussion and Conclusion

This paper presents a reproducible process for creating actual failures in the HDDs for reliability testing. When there was no SATA PM in the system, a single HDD failure had no impact on the rest of the system. However, when the SATA PM was used, a forced failure of one HDD caused the other HDD connected to the PM to also fail. This means that at least one combination of SATA controllers, the SATA PMs, and the HDDs does not provide resiliency when a single HDD fails. We must be very careful when using the SATA PM to build a large scale storage system, or the consequence could be catastrophic.