The University of Southern Mississippi

The Aquila Digital Community

Honors Theses

Honors College

Spring 5-2018

An Empirical Study On the Recovery Speed of USB Flash Drives Utilizing Raid-5 Compared to HDDs and SSDs

Joshua Manuel Martins
University of Southern Mississippi

Follow this and additional works at: https://aquila.usm.edu/honors_theses

Part of the Data Storage Systems Commons

The University of Southern Mississippi

An empirical study on the recovery speed of USB flash drives utilizing RAID-5 compared to

HDDs and SSDs.

by

Joshua Martins

A Thesis
Submitted to the Honors College of
The University of Southern Mississippi
in Partial Fulfillment
of the Requirements for the Degree of
Bachelor of Science
in the Department of Computing

December 2017

Approved by:

_____
Jeremy Graves, M.S., Thesis Advisor
Instructor, Information Technology

_____
Andrew H. Sung, Ph.D., Director
School of Computing

_____
Ellen Weinauer, Ph.D., Dean
Honors College

Abstract

Since their creation and implementation, storage drives have undergone and continue to undergo drastic changes in speed, size, and reliability. The original storage drives, known as hard disk drives (HDDs), are constructed using moving parts. The second modern type of storage drives, known as solid state drives (SSDs), are constructed using a series of silicon chips that utilize no moving parts. The third and most recent innovation in storage drives, known as USB flash drives (USBs), use only a single silicon chip to provide storage which grants them the smallest form factor of the three drive types.

This study compared these three types of storage devices with a RAID-5 drive redundancy configuration. A RAID-5 configuration logically joins three or more drives of a single type using either software or hardware, a situation in which both HDDs and SSDs are frequently placed. This research placed USBs in a similar configuration to compare their functional speeds within this arrangement with two similar configurations of HDDs and SSDs. These recorded speeds were then mathematically compared with the price of the drives to determine if USBs are a cost-effective alternative to HDDs and SSDs in the current marketplace. While the testing did not demonstrate consistent results with the selected batch of USB drives, the evolutionary trajectory of storage technology promises that such devices will eventually match their peers in processing capabilities.

Key Words: empirical comparison, hard disk drives, RAID-5, recovery speed, solid state drives, USB flash drives

Dedications

I would like to dedicate this thesis to both of my parents and my sister. Thank you all for always pushing me to do my best.

Acknowledgements

Table of Contents

List of Figures

List of Tables

List of Abbreviations

EPROM	Erasable Programmable Read Only Memory
FIO	Flexible Input/Output
GB	Gigabyte
HDD	Hard Disk Drive
HP	Hewlett-Packard
IBM	International Business Machines
IOPS	Input/Output Operations per Second
MDADM	Multiple Device Admin
MTTF	Mean Time to Fail
PCMCIA	Personal Computer Memory Card International Association
RAID	Redundant Arrays of Inexpensive Disks
SATA	Serial AT Attachment
SSD	Solid State Drive
USB	Universal Serial Bus
XOR	Exclusive "or"

Chapter 1: Introduction

Hard disk drives (HDDs, a form of computer storage that utilizes moving parts) have been a common storage item throughout much of computing history (Ekker, Coughlin, & Handy, 2009). Unfortunately, the moving parts of these drives wear out over time and as a result can render these drives unreliable (Patterson, Gibson, & Katz, 1988, p. 110). The drives have expanded in size since their inception, but their speed has fallen behind that of computer processors (Ekker et al., 2009; Patterson et al., 1988, p. 109). To combat the storage loss caused by drive failure and to glean faster speeds from HDDs, computing techniques known collectively as RAIDs (Redundant Arrays of Inexpensive Disks) were introduced (Patterson et al., 1988, p. 110). These RAID configurations come in many forms, but this study will focus on RAID-5. RAID-5 is a rotating parity configuration in which a parity block is stored adjacent to the multiple data blocks to reconstruct lost data when a drive fails (Patterson et al., p. 114).

While RAIDs offered necessary reliability improvements to early HDDs, the creation of solid state drives (SSDs, drives that utilize silicon chips) increased the speeds at which storage can be accessed (Ekker et al., 2009). However, the original SSDs were marketed at prices rivaling small fortunes (Ekker et al., 2009), and their small counterparts, USB flash drives, offered only a fraction of useful space at a reasonable price (Mohr, 2008, p. 94). As time went on, memory densities improved and as a result SSDs have become cheaper (Ekker et al., 2009) and USB flash drives have grown to hold comparable amounts of space (Mohr, 2008, p. 94).

While SSDs have fallen in price, they continue to cost a considerable amount more than HDDs and USB flash drives. And solid state storage is not completely free of reliability issues; both SSDs and USB flash drives can wear down their silicon chips through "excessive read/write cycles" (Ekker et al., 2009). Originally intended for use in HDDs, RAID has found a place in modern solid state storage systems. While SSDs still command a higher price per gigabyte, it is reasonable to consider logically binding USB flash drives to create dependable and fast storage in the form of an array. "So what's to stop us [from] treating these portable storage units as small hard drives?" (Mohr, 2008, p. 94) Thanks to the speed of USB 3.0 and relatively cheap prices per gigabyte, USB flash drives can and have been joined using RAID level 0 techniques. However, there have yet to be studies undertaken to track the operating speeds of drives in a RAID 5 array made up of USB flash drives.

This research created three RAID 5 arrays using three HDDs, three SSDs, and three USB flash drives. These logical drives were tested as storage for general files. This study simulated drive failure and recovery to document the speed of the operations of the surviving drives.

Each drive was connected to the computer using a USB 3.0 hub. The drives were bonded and formatted using a Linux-based operating system due to the fact that Windows operating systems do not offer an easy method of creating RAIDs on USB flash drives (Mohr, 2008, p. 94). The time required for the system to configure and format the array was documented. Data on the efficiency of these logical drives was collected to set a baseline. Once the baseline was set, a single drive from each configuration was set faulty to simulate the failure of the drive. If the RAID worked properly, the remaining drives were able to shift into a RAID 0 array and continued to function. Once this test had completed, the faulty drive was re-added to the array and the time required to do so was documented. A final reconstructed test was then run on the array. The recorded efficiencies of the tests were compared to determine if USB flash drives are a viable form of dependable, fast, and cheap storage.

Chapter 2: Literature Review

2.1 Storage Devices

This study is incorporating three storage device types for analysis. A storage device is a computer component that stores information when the machine is powered down. The three devices to be tested are hard disk drives, solid state drives, and USB flash drives.

2.1.1 Hard Disk Drives (HDDs)

The original magnetic hard disk drive was first created by IBM scientists in 1956 (Noyes & Dickinson, 1956, p. 42). The original setup featured 50 disks spaced 0.3 inches apart to allow space for reading and writing via magnetic heads (Fig. 2.1, Noyes & Dickinson, 1956, p. 42). Each disk had a magnetic coating, and the device utilized a motor to rotate the disks for access (Noyes & Dickinson, 1956, p. 43). Modern hard disk drives are composed of less disks (Fig. 2.2) but are capable of reading and writing far more information at much higher speeds (Anderson, Dykes, & Riedel, 2003, p. 247). However, the response times of these drives has fallen behind that of processors due to rotational latency (Ekker, Coughlin, & Handy, 2009, p. 2). The drives also suffer from natural "wear and tear associated with mechanical devices" (U.S. Patent No. 5,459,850, 1995) which culminated in the creation of a storage reliability method called Redundant Arrays of Inexpensive Disks (RAID) (discussed later in this chapter) to combat the unreliability of hard disk drives (Patterson, Gibson, & Katz, 1988, p. 110).
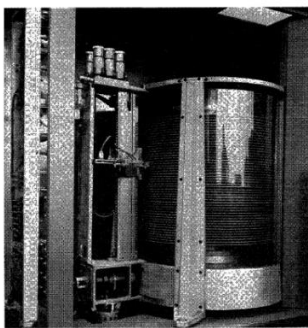

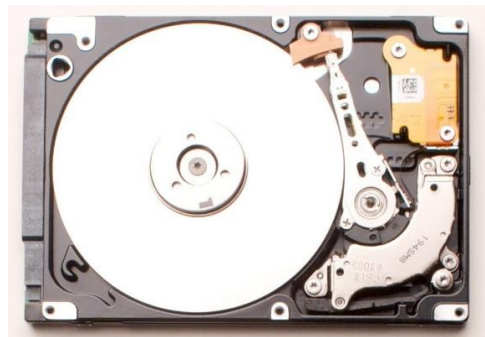**Fig. 2.1** *(Noyes & Dickinson, 1956, p. 42)*

Fig. 2.2 *(Intel News Release, 2008)*

2.1.2 Solid State Drives (SSDs)

Derived from the nonvolatile memory of erasable programmable read-only memory (EPROM) (U.S. Patent No. 3,660,819, 1972), Flash EPROM was created by Microsoft scientists

(Canada Patent No. 2,088,442, 1993). This new form of storage was designed with a casing shape similar to hard disk drives (U.S. Patent No. 5,459,850, 1995) and has resulted in modern solid state drives (Fig. 2.3). Solid state storage writes to silicon chips, and as a result has a higher input and output rate than hard disk drives (Ekker et al., 2009, p.1). These drives cost far more than typical mechanical drives (Ekker et al., 2009, p. 7), though their prices continue to decrease as time goes on (p. 10). For example, a single 32GB Intel SSD in 2008 was marketed for $695 (Intel News Release, 2008) while the 128GB SanDisk SSDs purchased in 2016 for this experiment cost $40 each. The drives can endure far more read/write activity than mechanical drives, but they are not infallible and do eventually fail (Gasior, 2015, para. 5). One study found that one brand of drive died after constantly reading/writing 800 terabytes of information (Gasior, 2015, para. 12) while two other brands died after constantly reading/writing over 2 petabytes (2000 terabytes) of information (Gasior, 2015, para. 16). However, these extreme amounts are far more than the typical consumer should ever require (Gasior, 2015, para. 10).

Fig. 2.3 (Intel News Release, 2008)

2.1.3 USB Flash Drives

An offshoot of the flash memory developments resulted in the creation of the modern USB Flash Drive, originally called a tandem circuit card (U.S. Patent No. 5,513,074, 1996). The drive was created by IBM scientists according to standards set by the Personal Computer Memory Card International Association (PCMCIA) (U.S. Patent No. 5,513,074, 1996). In a few short years, the device became available from several manufacturers (Stone, 2003, p. 145) and has seen a decrease in price similar to the decrease recognized in solid state drives (Ekker et al., 2009, p. 10). The

drives can be built as solid sticks of silicon chip flash (USB2U, 2013, para. 4) or the flash memory chips can be mounted to printed circuit boards (Fig. 2.4) (USB2U, 2013, para. 6).
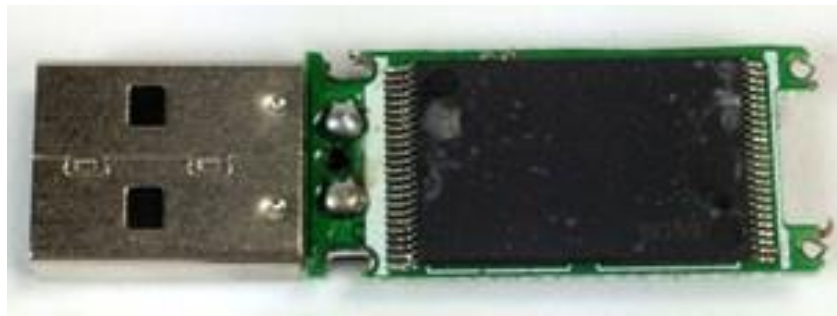
Fig. 2.4 *(USB2U, 2013)*

2.1.4 Mean Time to Fail (MTTF)

Drive manufacturers calculate the assumed life and write expectancies of their drives and provide this information with the drive specifications (Patterson et al., 1988, p. 110). These "mean time to fail" calculations are typically a conservative estimate that define a good checkpoint in drive usage; users should begin to monitor the drive at this time for any signs of failure, even though the drives may survive far beyond this estimate (Gasior, 2015, para. 10).

2.2 Hardware Media

This study used two cabling standards and an external hardware device to connect the storage devices to the computer conducting the examination.

2.2.1 USB 3.0

Intel formed the USB 3.0 Promoters Group in the mid-2000s with the intent of increasing the speeds offered by USB 2.0 (480 Mbps (megabits per second)) tenfold to 4.8 Gbps (gigabits per second) while maintaining backwards compatibility (Garber, 2007, p. 21). The work was successful (U.S. Patent No. 20090088024), and today both USB 3.0 and USB 2.0 ports are in common use on computer systems.

2.2.2 SATA

Serial AT Attachment (SATA) is a cabling protocol (Serial ATA Workgroup, 2003, p. 13) that defines the transmission of information in a "serial data stream" (p. 14) and can be connected to any storage device that conforms to this standard (p.15).

2.2.3 USB Hubs

Universal Serial Bus (USB) is another cabling protocol that applies to numerous devices, both storage and peripheral (U.S. Patent No. 6,141,221, 2000). The high prevalence of USB devices in modern computing combined with limited space for numerous USB ports on computers has resulted in the use of hubs, devices that are "wiring concentrators" that offer multiple access ports to hosts (U.S. Patent No. 6,141,221, 2000).

2.3 Software

This study used the predefined RAID-5 configuration that logically combines storage devices with rotating parity, the Linux software command FIO that tracks the input/output capabilities of the storage devices, and the Ubuntu operating system that enables the use of both the aforementioned logical combination and drive monitoring software.

2.3.1 RAID

Redundant arrays of inexpensive disks (RAIDs) were created to combat the poor reliability of mechanical disk drives (Patterson et al., 1988, p. 110). Divided into levels, these arrays can be created by combining the disks through hardware and software configurations (Patterson et al., 1988, p. 110). There are many RAID level configurations, but this paper focuses only on RAID level 5 (RAID-5) and its degradation to RAID-0 (Fowler, 2015a, para.3). In RAID-5, there is rotating parity. In the array, there is a single sector on one of the disks that calculates the parity of the other sectors in that same row of disks (Fig. 2.5) (Patterson et al. 1988, p. 114).

D= Disk, *A* = Data, **P** = Parity, S = Sector

S#	D1	D2	D3	D4
S1	*A*	*A*	*A*	**P**
S2	*A*	*A*	**P**	*A*
S3	*A*	**P**	*A*	*A*
S4	**P**	*A*	*A*	*A*
S5	*A*	*A*	*A*	**P**

Fig. 2.5: *The Layout of a RAID-5 Array (Patterson et al., 1988, p. 114)*

For the next row, another disk houses the parity sector and so on (Fig. 5) (Patterson et al., 1988, p. 114). This allows for the parity sectors to be spread across the array instead of sacrificing a whole disk to be a parity calculator as in RAID-3 (Patterson et al., 1988, p. 112). When a disk fails in RAID-5, the parity sectors of each disk can be compared to the surviving sectors and the information can be recreated on a new disk (Patterson et al., 1988, p. 114). When there is not a replacement disk immediately ready, the RAID-5 array tolerates the failure by shifting into a RAID-0 configuration (Fowler, 2015a, para. 3). RAID-0 uses a configuration known as striping in which the drives are logically bound to each other but offer no fault tolerance. (Fowler, 2015b, para. 5). If a drive in a RAID-0 array fails, the data it contains will be lost (Fowler, 2015b, para. 5).

Parity calculations take the binary values of each bit in a sector and compare them in the fashion depicted in Fig. 2.6 (Germundsson & Weisstein, 2016). The XOR value of each comparison is stored in a parity bit that corresponds to other sector bits in that that group (Patterson et al., 1988, pp. 112-113).

Disk 1 Sector Bit	Disk 2 Sector Bit	Disk 3 Parity Bit
1	1	0
1	0	1
0	1	1
0	0	0

Fig. 2.6 *Parity Calculations*

2.3.2 Benchmarking

Benchmarking is a term for the storage analyses conducted by software such as FIO; programs of this type calculate the input/output operations per second (IOPS) of storage devices by reading and writing data to the drives and tracking the speed of these functions (K, 2016).

2.3.3 Linux

Created by Linus Torvalds in 1991, the onetime college project Linux has grown into a massive operating system kernel that functions on numerous platforms (Calore, 2009, para. 16).

Countless other operating systems have been constructed using this base kernel, including the operating system known as Ubuntu (Canonical Ltd., 2016, para. 2). Designed to be "an easy-to-use Linux desktop" (Canonical Ltd, 2016, para. 2), Ubuntu has grown to have a worldwide user base (para. 7) and provides the functionality necessary to this study. According to Mohr (2008), Microsoft operating systems do not allow for USB flash drives to be logically combined using RAID software configurations (para. 4) while Linux operating systems provide the necessary support (para. 7).

2.4 Investigative Reasons

This study compared the IOPS of RAID-5 arrays on hard disk drives, solid state drives, and USB flash drives when used as simple consolidated storage units to determine the potential to use USB flash drives as cheap forms of external RAID-5 flash storage. At the time of this study, there have been no fully documented studies conducted on the IOPS for USB flash drives in RAID-5 configurations. This study included three of each type of drive. Due to one out of every three sectors on the drives being used for parity, the aggregated drives were expected to offer only 66% total storage capacity.

2.5 Other Studies

A study by Boboila & Desnoyers (2010) explored the durability of flash memory in USB flash drives and established that the drives are capable of enduring high levels of reading/writing to their silicon chips (Boboila & Desnoyers, 2010, p. 13). However, a study by Kim (2015) noted that the local garbage collection processes of solid state drives can have an impact on the effectiveness of a RAID array (Kim, 2015, p. 975). Garbage collection processes on solid state drives erase outdated memory blocks to prepare blank space for future writes, whereas hard disk drives simply overwrite the old data (Rouse, 2012, para. 2).

Chapter 3: Methodology

3.1 Setting up the Environment

The testing computer (an HP Envy 17.3" i7) was configured to be capable of a dual boot. The second operating system to be installed was Ubuntu (16.04). This was required so that the three forms of storage (hard disk drives, solid state drives, and USB flash drives) could be logically joined using the RAID functionality offered by the Linux kernel.

3.1.1 Installing the Necessary Software

The benchmarking software FIO was downloaded from the Ubuntu repository in the command terminal and the Linux operating system Ubuntu was downloaded from ubuntu.com. FIO and Ubuntu are both free software.

3.1.2 Cabling the Drives Together

The three hard disk drives and three solid state drives were placed in their respective docking stations (Sabrent USB 3.0 to SATA External Docking Station) complete with cooling fans. The six USB cables (AmazonBasics USB 3.0 Cable) were added next to attach the drives to the USB 3.0 hub (Sabrent High Speed 10 Port USB 3.0 HUB). The hard disk drives (Western Digital AV-GP WD2500AVVS 250 GB Internal Hard Drive), the solid state drives (SanDisk Internal SATA 120GB), and the USB flash drives (SanDisk Cruzer Glide 128GB) were grouped according to their type. The USB hub was then connected to the computer through a USB 3.0 port. The hub was plugged into the surge protector (Belkin 8-Outlet Commercial Power Strip Surge Protector) and the docking stations of the SSD and HDD groups were plugged in only when being tested. The groups were tested one at a time to accurately measure the IOPS of the logical drives over the USB 3.0 cable that connected the hub to the computer.

Once attached to the computer, it was necessary to erase the file systems provided by the manufacturer so that the drives could be recognized by the computer as capable of being joined in an array.

3.2 Logically Connecting the Drives Using RAID-5

The logical drives to be formed by the arrays were designated as drives **/dev/md(0-1-2)** for the USB flash drives, drives **/dev/md(3-4-5)** for the solid state drives, and drives **/dev/md(6-7-8)**

for the hard disk drives. The command line editor was used to create the logical drives. The commands **sudo apt-get install mdadm** and **sudo apt-get install gparted** were run to install the programs Multiple Device Admin and Gnome Partition Editor (Mohr, 2008, p. 94). Running the **lsblk –o NAME,SIZE,FSTYPE,TYPE** command outputs a list of the drives mounted to the system (Ellingwood, 2016).

To create the logical drive, the command **sudo mdadm –create –verbose / dev/md# - level=5 –raiddevices=3 /dev/sd# /dev/sd# /dev/sd#** was run (Mohr, 2008, p. 94; Ellingwood, 2016). The symbol (#) denotes the different drive letters that were present. The new logical drive was then formatted and mounted with the commands **sudo mkfs.ext4 /dev/md#** and **sudo mount /dev/md# /mnt/md#**, respectively (Ellingwood, 2016). References to the logical drives were appended to the files **/etc/fstab** and **/etc/mdadm/mdadm.conf** and the filesystem was updated from within the terminal with the **sudo update-initramfs –u** command (Ellingwood, 2016).

When the drives were to be recreated at the end of each simulated failure and recovery, the unmount, stop, and remove commands were required to reset the logical drives: **sudo umount /dev/md#** (Mohr, 2008, p. 95; Ellingwood, 2016), **sudo mdadm --stop /dev/md#** (Mohr, 2008, p. 95; Ellingwood, 2016), and **sudo mdadm --remove /dev/md#** (Ellingwood, 2016), respectively. The separated drives were then individually erased using the **sudo mdadm --zero-superblock /dev/sd#** command and persistent references to the removed logical drives were commented out in both the **/etc/fstab** and **/etc/mdadm/mdadm.conf** files (Ellingwood, 2016).

3.3 General Storage Benchmarking Test

The time necessary to prepare the drives and perform the tests was recorded as well as the IOPS of the logical drive in three scenarios. The test sections were: creation time, formatting time, functional state (baseline) benchmarking, failed state benchmarking, recovery time, and reconstructed state benchmarking. The speed and drive usage of each state was recorded. Each physical drive experienced a single simulated failure for a total of three per logical drive. The benchmarking fio command used was **fio --randrepeat=1 --ioengine=libaio --direct=1 -- gtod_reduce=1 --name=test1 --filename=test1 --bs=4k --iodepth=64 --size=250M -- readwrite=randrw --rwmixread=50**, which is based upon the command used in Christopher Sam K's 2016 study comparing disk I/O between AWS and DigitalOcean drives. This command has been slightly modified for usage within this study.

3.3.1 Creation Time

The Linux command terminal was then opened on the computer. The constituent drives were cleared of any file systems before being assembled within the logical drive. The process was monitored in person to document the time required for drive creation. Upon completion of drive creation, a command to record the time of completion was manually sent to the file documenting the testing times.

3.3.2 Formatting Time

The logical drive was then formatted with the Ext4 file system and the time it required was documented. The logical drive was mounted to its corresponding mount point and had references to its existence added to the necessary files. The filesystem was then updated to reflect these changes.

3.3.3 Functional State (Baseline) Benchmark

A folder named "fio" was then created on the logical drive using the **mkdir fio** command. The benchmarking test was run from within this folder to capture the IOPS of the logical drive. The results of this test were sent to the baseline record file (**md#-baselinegood**) within the user's home folder and the time of its duration was recorded.

3.3.4 Failed State Benchmark

The designated failure drive of the round of testing was set to faulty. The same benchmarking test was then run from the "fio" folder, and the results of this test were sent to the failed baseline record file (**md#-failedstate**) within the user's home folder. The duration of the test was once again recorded.

3.3.5 Recovery Time

The designated failure drive was then hot removed from the logical drive, and then promptly reattached. This reattachment process was manually monitored to document the time required to reconstruct the array.

3.3.6 Reconstructed State Benchmark

The same benchmarking test was run from the "fio" folder and sent to the reconstructed baseline record file (**md#-reconstruction**) in the same location as the previous two benchmarking log files. The duration of this test was once again recorded.

3.4 Determining Value of Devices

Following testing completion, the time-to-complete of the drives was averaged within their groups. These averages were inserted into the two following equations, where *A* is equal to gigabyte per hour, *h* is equal to average time in hours, *d* is equal to the storage size of a single drive, *V* is equal to price per gigabyte per hour, and *m* is equal to the price of a single drive.

$$A = d/h$$

$$V = m/A$$

The individual prices of the SSDs and HDDs include the price of a single docking station.

Chapter 4: Results

4.1 Time to Completion

TIMES TO COMPLETE TESTING								
TEST	DRIVE	DATE	Creation	Formatting	Baseline	Failed State	Recovery	Reconstructed	TOTAL
MD0	USB1	07/03/17	03:11:25	00:05:08	00:21:18	00:18:19	00:01:38	01:41:46	05:39:34
MD1	USB2	06/15/17	05:12:56	00:17:06	00:25:00	00:02:13	00:46:53	00:09:11	06:53:19
MD2	USB3	06/20/17	03:04:24	00:02:48	00:03:20	07:27:50	00:02:24	00:21:45	11:02:31
MD3	SSD1	06/22/17	00:28:30	00:00:16	00:00:10	00:00:07	00:00:14	00:00:09	00:29:26
MD4	SSD2	06/22/17	00:28:19	00:00:15	00:00:09	00:00:08	00:00:23	00:00:11	00:29:25
MD5	SSD3	06/22/17	00:28:19	00:00:03	00:00:11	00:00:06	00:00:25	00:00:10	00:29:14
MD6	HDD1	06/22/17	01:30:36	00:00:18	00:02:45	00:03:28	00:00:31	00:02:41	01:40:19
MD7	HDD2	06/27/17	01:31:05	00:00:09	00:02:48	00:03:25	00:00:27	00:02:44	01:40:38
MD8	HDD3	06/27/17	01:31:00	00:00:09	00:03:16	00:02:48	00:00:58	00:02:45	01:40:56

Times in red are potentially +2 seconds over actual time

Table 4.1: *Table of Process Completion Times*

The tests were all completed in single sessions (the USB1 drive test was repeated due to having been conducted over two days with the system being powered down in between tests). The USB tests demonstrated extremely variable ranges for time of process completion, while both the SSD and HDD tests shared fairly consistent completion times within their groups.

VALUE PER DRIVE					
TEST	AVG TIME	TIME IN HRS	FILE SIZE IN GB	PRICE	GB PER HR	PRICE/GB/HR
USB	1:12:18	1.2050	0.2500	19.9900	0.2075	96.3518
SSD	0:00:09	0.0025	0.2500	66.9800	100.0000	0.6698
HDD	0:02:58	0.0494	0.2500	43.9400	5.0562	8.6904

Prices in orange include the price of the external docking station

Table 4.2: *Table of Price Comparisons*

The average operating times in Table 4.2 include only the read/write speeds of the nine baseline tests for each group. The average operating results for the USB drive tests are skewed due to the unexpectedly long durations of the MD0 Reconstruction test and the MD2 Failed State test, and as a result the average GB/hr of the logical drives is limited to 0.2075 GB. This results in a price per GB/hr of $96.35 for USB drives, which is obscene in comparison to the $0.66 for the SSDs and $8.69 for the HDDs. The price of the docking stations is included in both the SSD and HDD prices, raising their prices by $26.99 (initially $39.99 for each SSD and $16.95 for each HDD).

13

4.2 Drive Usage

DRIVE USAGE DURING TESTING											
BASELINE				FAILED STATE				RECONSTRUCTED			
TEST	MD#	D1	D2	D3	MD#	D1	D2	D3	MD#	D1	D2	D3
MD0	94.81%	94.21%	94.81%	93.19%	99.75%	0.00%	99.75%	23.69%	100.00%	1.79%	100.00%	1.75%
MD1	97.14%	93.20%	97.14%	94.67%	98.21%	97.94%	0.00%	98.21%	95.97%	95.95%	82.63%	81.86%
MD2	89.56%	89.56%	60.12%	51.74%	100.00%	0.71%	100.00%	0.00%	70.19%	47.32%	70.19%	55.47%
MD3	97.68	73.60%	73.54%	97.68%	98.06%	0.00%	74.58%	98.06%	97.59%	69.51%	69.70%	97.59%
MD4	97.27%	73.03%	73.03%	97.27%	97.59%	66.60%	0.00%	97.59%	77.67%	64.80%	66.53%	77.67%
MD5	76.31%	64.93%	64.50%	76.31%	96.05%	95.75%	96.05%	0.00%	97.54%	72.98%	74.85%	97.54%
MD6	96.27%	92.27%	95.04%	96.27%	99.14%	0.00%	99.14%	94.34%	97.17%	97.17%	94.36%	94.32%
MD7	96.95%	93.18%	93.89%	96.95%	98.38%	98.38%	0.00%	97.10%	98.21%	98.21%	95.73%	93.18%
MD8	95.01%	94.97%	95.01%	93.88%	97.92%	97.92%	97.05%	0.00%	95.90%	94.29%	95.90%	94.63%

*Percentages in blue are aggregate usage of the drives

Table 4.3: *Table of Drive Usage Percentages*

The usage of the drives proved an interesting point of difference between the USB drives and both the HDDs and SSDs. In the USB drives, the arrays seemed to have favored the usage of USB drive 2 over the other two USB drives. This resulted in USB drive 2 operating at 100.00% usage during two separate testing periods while the other drive(s) recorded usage of less than 2.00%. This is believed to have negatively impacted the performance speed of the arrays, and the strange behavior of USB drive 2 is potentially indicative of it being a faulty drive. Neither the SSDs nor HDDs experienced this favoritism.

Chapter 5: Discussion

5.1 Discussion of Notable Events that Impacted Testing

The HDDs (250GB) used in this study have nearly double the storage size of the SSDs (120GB) and USB flash drives (128GB); the production of HDDs with storage less than 250GB appears to have ended.

Both the testing location and choice of testing operating system were changed many times prior to the first successful test; the USB1 and the SSD1 tests were both started on a university computer but cancelled due to improper setup, and the USB1 test was started a separate time on the personal laptop. The USB drives were wiped and rewritten on two separate occasions in preparation for the USB1 test. This information is being shared to notify readers that the SSD and USB drives experienced minor use prior to the start of the official tests, while the HDDs experienced no activity prior to their testing.

The seeming election of USB drive 2 as the dominant drive in the RAID5 array was a surprise that appears to have impacted the potential of the USB drives. This "election" was not witnessed in the drive usage statistics of the SSD and HDD testing groups, and such an election process was not described in the preceding literature.

5.2 Discussion of Testing Parameters

The size of the test file generated by the fio command was reduced to 250MB after test sizes of 1GB and 4GB were attempted and found to take unnecessarily large amounts of time. These prior tests were not completed and have not been included in this study.

Chapter 6: Conclusion

6.1 Conclusion

The use of USB drives in RAID-5 general storage arrays has been demonstrated to be a time-consuming idea that outweighs the benefits of reduced cost. Even if the USB logical drives operated at the fastest speeds recorded, the vast difference in time-to-completion and consistency in operation is a strong deterrent.

6.2 Future Work

Due to the ever-evolving nature of drive storage, there is a foreseeable future of USB drives improving in speed and storage functionality to the point that future testing will show them to be a fantastic alternative. As the price of storage per gigabyte continues to fall and the choice of purchasing SSD, HDD, and USB drives for use in an array is reduced to the preference of the purchaser, I look forward to the day where a RAID configuration of USB drives can perform at a level equal to that of its predecessors.

References

Ainsbury, A.W., & Kerklaan, A.J. (1996). *U.S. Patent No. 5,513,074.* Washington, DC: U.S. Patent & Trademark Office.

Anderson, D., Dykes, J., & Riedel, E. (Eds.). (2003). Proceedings from File and Storage Technologies '03: *More than an interface – SCSI vs. ATA.* San Francisco, CA: USENIX.

Boboila, S., & Desnoyers, P. (Eds.). (2010). Proceedings from 8th USENIX Conference on File and Storage Technologies: *Write endurance in flash drives: measurement and analysis.* Berkeley, CA: USENIX.

Calore, M. (2009, August 25). *Aug. 25, 1991: Kid from Helsinki foments Linux revolution.* Retrieved from https://www.wired.com/2009/08/0825-torvalds-starts-linux/

Canonical Ltd. (2016). *The Ubuntu story*. Retrieved from http://www.ubuntu.com/about/about-ubuntu

Clay, D.W., & Anderson, S.A. (1995). *U.S. Patent No. 5,459,850.* Washington, DC: U.S. Patent & Trademark Office.

Ellingwood, J. (2016, August 16). *How to create RAID arrays with mdadm on Ubuntu 16.04.* Retrieved from https://www.digitalocean.com/community/tutorials/how-to-create-raid-arrays-with-mdadm-on-ubuntu-16-04

Ekker, N., Coughlin, T., & Handy, J. (2009, January) *Solid state storage 101: An introduction to solid state storage.* Retrieved from http://www.snia.org/sites/default/files/SSS%20101%20Wht%20Paper%20Final.pdf

Fowler, S.E. (2015a). *Do's and don'ts for RAID-5.* Retrieved from http://www.recover-raid.com/failed-RAID-help.html

Fowler, S.E. (2015b). *Understanding basic RAID definitions for commonly used types/levels of RAID.* Retrieved from http://www.recover-raid.com/RAID_understand.html#R0

Frohman, B.D. (1972). *U.S. Patent No. 3,660,819.* Washington, DC: U.S. Patent & Trademark Office.

Garber, L. (2007, December). New approach will speed up USB. *Computer, 40*, 21-22.

Gasior, G. (2015, March 12). *The SSD endurance experiment: They're all dead.* Retrieved from
http://techreport.com/review/27909/the-ssd-endurance-experiment-theyre-all-dead

Germundsson, R., & Weisstein, E.W. (2016, September 27). *XOR.* Retrieved from
http://mathworld.wolfram.com/XOR.html

Intel News Release. (2008, October 15). *Intel ships enterprise-class solid-state drives.* Retrieved
from http://www.intel.com/pressroom/archive/releases/2008/20081015comp.htm

K, C.S. (2016, September 16). *IOPS: Benchmarking disk I/O – AWS vs DigitalOcean.* Retrieved
from https://dzone.com/articles/iops-benchmarking-disk-io-aws-vs-digitalocean

Kim, Y. (2015, January 31). An empirical study of redundant array of independent solid-state
drives (RAIS). *Cluster Computing, 18*, 963-977.

Krueger, W.J., & Rajagopalan, S. (1993). *Canada Patent No. 2,088,442.* Gatineau, QC: Canadian
Intellectual Property Office.

Ling, Y., Tong, D., & Lynch, J. (2009). *U.S. Patent No. 20090088024.* Washington, DC: U.S.
Patent & Trademark Office.

Mohr, N. (2008) Turbo-boosting USB flash drives. *APC (Bauer Media Group), 28*(12), 94-95.

Noyes, T., & Dickinson, W.E. (Eds.). (1956). Proceedings from AIEE-IRE '56 (Western):
Engineering a design of a magnetic-disk random-access memory. New York, NY: ACM.

Patterson, D.A., Gibson, G., & Katz, R.H. (Eds.). (1988). Proceedings from 1988 ACM SIGMOD
International Conference on Management of Data: *A case for redundant arrays of
inexpensive disks (RAID).* New York, NY: ACM.

Rouse, M. (2012, January). *Solid state storage (SSS) garbage collection.* Retrieved from
http://searchsolidstatestorage.techtarget.com/definition/solid-state-storage-SSS-garbage-
collection

Serial usb2uATA Workgroup. (2003, January 1). *Serial ATA: High speed serialized AT
attachment.* Retrieved from
http://www.ece.umd.edu/courses/enee759h.S2003/references/serialata10a.pdf

Stone, W.R. (2003). From the screen of Stone. *IEEE Antennas and Propagation Magazine, 45*(3), 142-149.

Tong, E., & Van Harlingen, B.L. (2000). *U.S. Patent No. 6,141,221.* Washington, DC: U.S. Patent & Trademark Office.

USB2U. (2013, February 26). *PCB, COB & UDP.* Retrieved from http://www.usb2u.co.uk/articles/tag/chip-on-board/