

SSD Data Recovery: Benefits of Industrial Cooperation

Executive Summary

Recovering data from failed storage devices is a complicated process. Over the last four decades, the data recovery industry has conducted extensive research to develop techniques for recovering data from failed hard disk drives (HDDs). Solid-state drive (SSD) technology introduces a new set of challenges for data recovery engineers to solve. Although some SSD recovery techniques have been developed, the ever-changing landscape of SSD technology and the proliferation of self-encrypting drives is making data recovery from today's SSDs difficult and in some situations impossible.

In order to facilitate efficient and cost-effective data recovery from SSDs, Gillware is requesting cooperation from SSD manufacturers. Assistance can be provided in many forms, but ranges from simple technical specifications to specialized manufacturer commands or data recovery specific software toolkits.

Although assistance will go a long way to improve the techniques employed to recover data from failed SSDs, Gillware is aware that SSD manufacturers are wary of disclosing sensitive information and valuable intellectual property. Gillware wishes to work with SSD manufacturers to 1) clearly define the information and tools needed and 2) figure out ways to share this information in a confidential and secure manner.

This white paper is intended to serve as a starting point for future conversations between Gillware and SSD manufacturers. It outlines some of the challenges the industry faces when performing data recovery from SSDs and some areas where SSD manufacturers can assist in the effort.

Introduction

Hours of painstaking work crafting the perfect PowerPoint presentation, thousands of irreplaceable memories stored as impeccably organized digital images or years of research locked in hundreds of Word documents—regardless of what kind of data it is, losing it is an extremely emotional experience. Solid-state drive technologies offer a reliable way to store electronic data. This fact, however, does not mean the SSDs are impervious to failure. Sophisticated drive firmware and the unpredictable behavior of the average computer user means that SSD failures can and do occur. When this results in data loss, there is an expectation by the end user that there “must be a way to recover the data”. For data recovery labs like Gillware, the challenge is providing customers who have lost data as the result of an SSD failure a fast, reliable data recovery option at an affordable price.

Data recovery is rarely an inexpensive endeavor. Regardless of the storage technology, SSD or HDD, the techniques required to recover data from a failed device require a significant amount of research to develop and many hours of a recovery technician's time to perform. That being said, the fact that HDD technology is mature and has been around for many decades means that on average the data recovery success rates are much higher, and so the data recovery costs are much lower for HDDs compared to SSDs. Although the data recovery industry as a whole has invested significant resources into developing SSD data recovery techniques, certain demands from the market have resulted in design decisions that make data recovery from SSDs extremely expensive and, in some cases, impossible without assistance from the device manufacturer.

Gillware's goal is to foster collaboration between the SSD and data recovery industries to ensure that the data recovery needs of SSD customers can be met in a reliable, affordable and timely manner. While at the same time being cognizant that the engineering resources being committed to the data recovery effort by the SSD manufacturer needs to be kept to a minimum and valuable intellectual property must be protected.

HDD vs. SSD Data Recovery: A Comparison of Price, Turnaround Time and Success Rates

Many factors impact the cost of data recovery from failed storage devices, including equipment, facilities and human resource expenditures. However, research and development is the biggest contributor to the relatively high price of data recovery.

HDDs and SSDs are incredibly sophisticated devices with multiple potential failure points. Each failure mode requires different techniques in order to recover data stored on the device. The research and development time required to establish reliable and cost-effective recovery procedures for each specific drive and failure mode is substantial. This work is generally performed by experienced teams of electrical and mechanical engineers and computer scientists. Hundreds of new drive models are released every year and drive manufacturers are continuously pushing the envelope in terms of performance and capacity. As a result, successful data recovery organizations must invest enormous amounts of resources into research and development, with sometimes hundreds of hours spent on the development of a single new technique. Taking the time in the R&D phase to develop efficient data recovery tools and techniques usually results in lower average data recovery costs to the consumer. More specifically, reducing the amount of time spent by an engineer or technician to perform the recovery reduces the overall cost.

Faster turnaround times also mean that the value of the data to the consumer is preserved. In most data recovery scenarios, there is an inverse relationship between the value of the data and the time it takes to recover the data. In other words, the data is never more valuable than at the instant it is lost. As potential sales are missed, payrolls come and go, and projected deadlines are delayed, the once critical data becomes less important as it is naturally recreated. Therefore, for data recovery to make economic sense, the recovery process must be accomplished both quickly and cost-effectively. Most data recovery professionals agree that with the exception of cases in which data cannot be recreated, there is a precipitous drop-off in the number of customers willing to pay for their lost data when recovery times exceed three weeks. Figure 1 depicts the delicate balance that exists in the data recovery industry between the value of the lost data to the consumer and the cost and turnaround time of performing the recovery.

Data Recovery Value/Cost

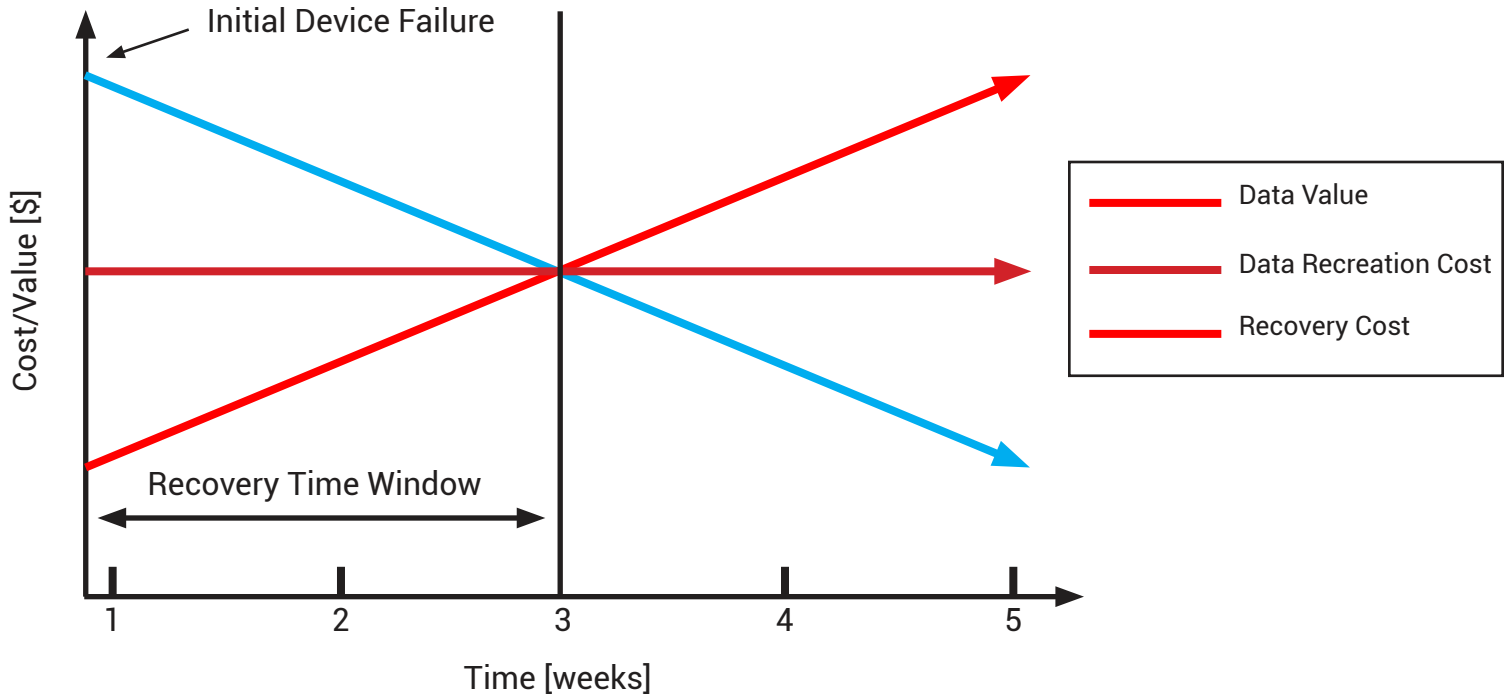


Figure 1: A graphical representation of the relationship between the value of lost data to the consumer and the cost of the recovery compared to the cost of manual data recreation.

Through a commitment to research and development, Gillware Inc. has been able to significantly reduce the turnaround time and total cost for a single HDD data recovery. The industry average cost of a single HDD data recovery is around \$1500 and average turnaround time is close to three weeks. For the fiscal year 2013, the average HDD data recovery at Gillware Inc. cost \$694 and took only six business days to complete, staying well within the recovery time window shown in Figure 1.

Years of experience and well-defined techniques have stabilized the average cost and turnaround time for data recovery from HDDs. SSD data recovery, on the other hand, is a discipline that is being developed as the SSD technology evolves. As a result, the cost, recovery time and success rates from SSDs can vary dramatically depending on the specific SSD technologies employed (e.g., ECC, encryption, FTL, etc.) and whether or not the controller or SSD manufacturer is supporting the data recovery effort by collaborating with data recovery providers.

For example, Gillware's average service fee is \$700 for data recovery from full-disk encrypted SSDs whose manufacturers have assisted Gillware engineers with technical specifications and tools. The average turnaround time for such recoveries is five business days. More importantly, Gillware is currently reporting over a 90% success rate for these cases. This effectively brings the price and turnaround time for SSD data recovery in line with that of HDD data recovery while at the same time improves upon the success rates seen with HDD. Conversely, the average service fee for data recovery from an SSD whose manufacturer is not assisting in the data recovery effort is more than \$3,500. In many cases data recovery from modern SSDs employing state-of-the-art security measures is simply not possible without assistance from the device manufacturer. Figure 2 clearly shows the impact that support from SSD manufacturers can have on data recovery service fees.

Average Data Recovery Service Fee

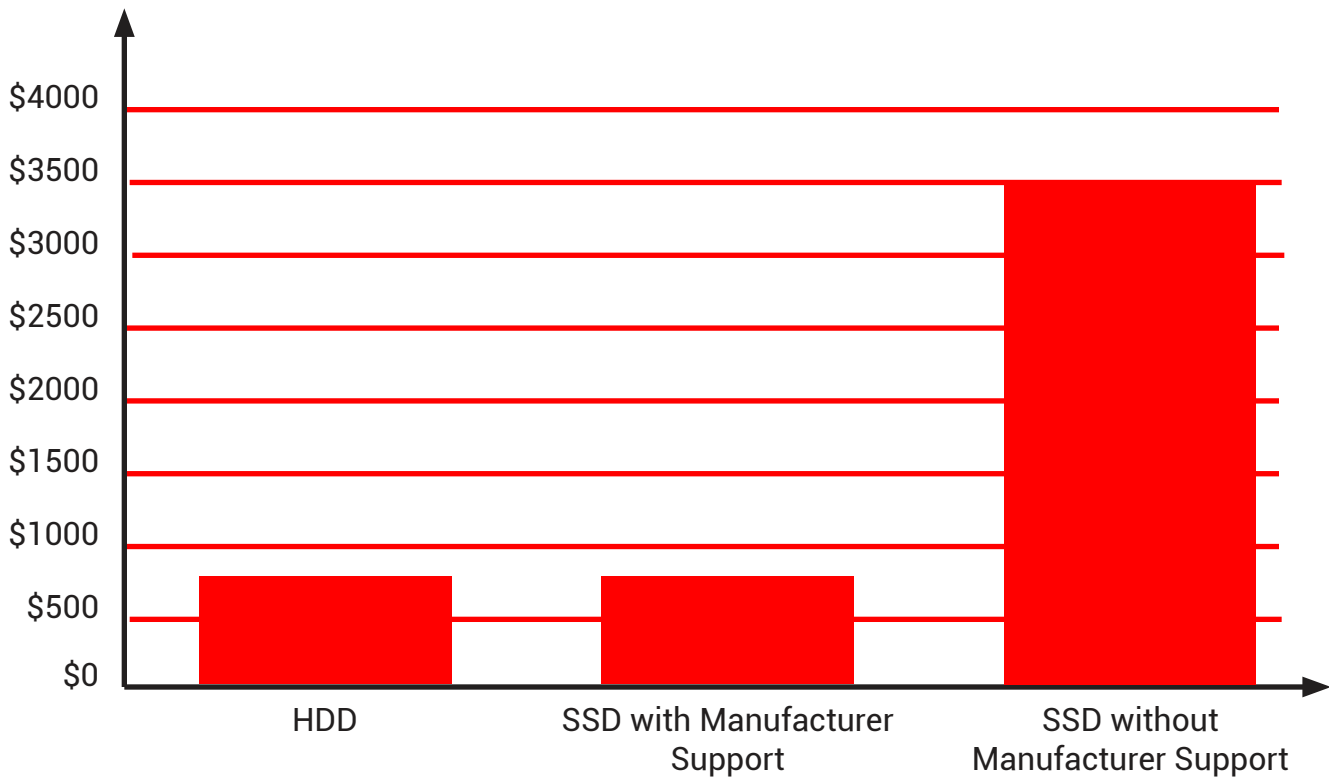


Figure 2: Comparing the average data recovery service fee for HDD recovery vs. SSD recovery where the manufacturer is supporting the data recovery effort vs. SSD recovery where the manufacturer is not supporting the data recovery effort.

Gillware's goal is to collaborate with more SSD manufacturers to ensure that we can offer data recovery services equal to or better than those currently offered to Gillware's HDD recovery customers.

SSD Data Recovery Overview

The majority of SSDs that come in for recovery are physically and electrically healthy, but suffer from corruption to key areas of the device's firmware. A device in this state will generally detect with a generic model name, an incorrect capacity and will not allow access to user data. These symptoms are not entirely dissimilar to those exhibited by many failed hard disk drives. Indeed, firmware in both types of devices share many common tasks. For example, both have to perform a translation of a logical block address to a physical data location and both have to adapt to media defects that arise throughout the life of the device. However, the key distinction between HDD and SSD technology from a recovery standpoint is that data stored on the underlying storage media can be easily accessed in the case of an SSD.

The ideal technique for recovering data from a failed HDD would be a device that can read HDD platters independent of the hard drive. Although accomplished in laboratory environments with varying degrees of success, this technique has not shown promise as a cost-effective solution for commercial data recovery. HDD data recovery continues to hinge on restoration of the failed device, completely at the mercy of ever-shrinking mechanical tolerances, sophisticated control electronics and complex firmware all being made to work together in harmony.

In the case of SSDs, however, not only does a method for directly reading the data from NAND flash memory chips exist but one is readily available from a host of electronics suppliers across the globe. The underlying storage medium, an array of industry-standard NAND flash memory chips, can be accessed without restoring the SSD through the use of any off-the-shelf device programmer. This powerful distinction between HDD and SSD technology should relieve the SSD recovery engineer from the burden of device repair and allow for recovery in all but the most catastrophic of circumstances. However, due to some peculiarities of NAND flash memory and certain characteristics of SSD technology, data recovery from SSDs cannot be performed by simply reading the raw NAND flash and concatenating the images.

SSD Recovery Challenges

Challenge #1: Determining the NAND Page Layout

The data coming off the NAND directly hardly resembles what engineers are used to seeing through a sector editor and is far from being usable by a client. Interspersed with user data are bits and pieces of information used internally by the SSD and never seen during normal operation. There is no industry-standard way of organizing information in each NAND page, and determining the exact page layout is a crucial first step in the recovery process.

A good portion of this extra information is used for error correction code (ECC). Bit errors are seamlessly detected and corrected in hardware by the SSD controller during normal operation and the same procedure must be applied by the engineer during recovery. The exact ECC implementation varies from drive to drive and determining it is often a time-consuming process of trial and error.

Challenge #2: Deciphering the Flash Translation Layer

SSD recoveries can be explained using analogies to recovery from a failed RAID array: Both storage technologies combine multiple physical components into one large pool of storage, and any individual file is often striped across many of these components. But unlike a RAID, a logical block's location on an individual NAND chip does not directly correlate with its location in the overall volume.

The SSD firmware maintains a fluctuating logical-to-physical location mapping, commonly referred to as a Flash Translation Layer (FTL). The necessity for an FTL rather than a conventional RAID level stems from the peculiarities of NAND flash memory. The memory is divided into a number of equally-sized units known as blocks, which themselves are divided into a number of equally-sized pages. Data access is performed at the page level and, like HDD sectors, pages are random-access and can be read from or written to in any order. To rewrite a page, however, the entire block must first be erased. Furthermore, the number of write/erase cycles tolerated over the lifetime of the memory is limited.

It would be terribly inefficient and would dramatically reduce the lifespan of the SSD if an entire block was erased and rewritten to accommodate a change in a single page. A better approach is to store the new data in an available page and update the FTL with the new location. When a sufficient number of pages from a given block have been remapped, the block can be erased and be made available for use again.

Challenge #3: Stitching the Raw NAND Page Dumps Back Together

Transforming the raw pages read from the NAND back into a linearly-addressed disk image is the most difficult part of an SSD recovery. This is generally accomplished by identifying key filesystem structures in the page data area that must exist at a specific LBA. For example, finding a Master Boot Record at the start of a particular page is strong evidence that the page stores the data for LBA 0. This information often allows us to identify an LBA number or other logical-to-physical mapping information in the page spare area.

A side effect of the page-update scheme described in the previous section is that old versions of an updated page persist in the storage array for an indeterminate amount of time. From a recovery perspective, this often results in multiple pages claiming to belong at the same physical location. Similar techniques already discussed can sometimes be used to remedy this. A page containing a filesystem inode, for example, will have a modification timestamp that can be used to distinguish conflicting pages and isolate a revision number in the spare area.

Challenge #4: Dealing with Encryption

Since Gillware started performing SSD recoveries in 2008, another obstacle has emerged that has the potential to make recovery impossible without manufacturer assistance: self-encrypting drives. From an IT perspective, it's a major improvement. There's no software to install, no key packages to manage and from day one everything that reaches the underlying medium is fully encrypted. Although encryption is great from a security and IT process efficiency point of view, the same cannot be said for encryption's impact on data recovery. The tools and techniques developed for recovery from raw NAND dumps discussed in the previous section are no longer applicable and the only option for recovery is to restore the device to operation. Unless, of course, the SSD manufacturers provide a means to perform raw NAND dumps with the data in a decrypted format.

The following section outlines the capabilities that would help the data recovery industry address some of the SSD recovery challenges that exist today, and achieve acceptable levels of data recovery services in terms of reliability and cost. The list should not be seen as an all-or-nothing situation. Each individual capability represents incredible value to the data recovery industry.

Capabilities the Data Recovery Industry Needs

- ***A means to access the raw, unencrypted NAND data when the device can be properly authenticated.*** A command such as this would need to be reviewed by the Trusted Computing Group to ensure that it does not violate standards and principles governing self-encrypting drives. It is important to note that the vast majority of the failed SSDs arriving in Gillware's lab are still seen by the host controller. These devices have encountered an unanticipated or unexpected condition, and as a result contingencies to handle the event were not implemented by the firmware engineers. We refer to drives in this condition as being in a "panicked" state. Although the device will not allow access to user data in this state, virtually ALL can be properly authenticated and unlocked. It is only after being unlocked that they "panic".

Gillware believes that because the command would only be available in instances where the user has not set any pre-boot security controls, or when the device allows set pre-boot security controls to be satisfied, that such a command would not violate the guidelines put forth in the Opal storage specification.

With this proposed raw dump command implemented, the same methodology used today for recovering unencrypted drives could then be applied to recover data from self-encrypting SSDs. In order to avoid the need to remove each NAND chip from the SSD and read it individually, the unencrypted raw dump capability would ideally be implemented as a vendor-specific ATA command, given that the decryption credentials are not likely to be externally accessible.

- **A description of the NAND page layout** including a breakdown of the fields present in the spare area and the exact ECC implementation. Ideally, error correction would be handled by the device as part of the raw dump command.
- **A means to employ as much of the most recent runtime translator as possible.** Relying solely on an LBA marker in the spare area has shown to produce hundreds, if not thousands, of conflicts for a given sector, which negatively impacts the turnaround time and the quality of the recovery.

Although support for recovery is limited across the SSD industry, Gillware has partnered with some leading SSD manufacturers who have provided many of the capabilities listed above. The positive impact this support has had on the recovery success rates, costs and turnaround times for the drives produced by these manufacturers is easy to measure. Success rates are better than 90%, turnaround times are less than a week and the average cost is in line with the recovery costs of HDD recoveries.

Protecting Intellectual Property

The data recovery industry understands that SSD manufacturers have spent an incredible amount of engineering and financial resources researching and developing their technology. The business case for assisting data recovery labs is unlikely to gain acceptance if it comes at the cost of putting extremely valuable intellectual property at risk. Although opinions on what qualifies as protected intellectual property vary from manufacturer to manufacturer, much of the information that would be useful to data recovery providers is likely to be common knowledge to most in the SSD community, yet not readily available to those in the data recovery industry. For example, having easy access to a simple breakdown of NAND flash layouts could prove extremely beneficial to the recovery effort and is unlikely to involve the disclosure of protected intellectual property. This kind of information can also be provided to the data recovery labs in the form of basic engineering documentation, meaning the engineering resources required on the part of the SSD manufacturer are minimal.

In situations where protected intellectual property is involved, alternative solutions may be necessary. Data recovery professionals and SSD manufacturers need to find solutions that make data recovery possible, but at the same time do not require the SSD manufacturer to disclose protected intellectual property. One potential solution is to have the SSD manufacturer develop software toolsets that provide the assistance the data recovery labs need without disclosing protected intellectual property. For example, the self-encrypting nature of modern SSDs means that performing raw dumps of the NAND memory and the reassembling the NAND images is not possible. It is necessary to perform the raw dump through the controller to ensure that the data is read unencrypted. If the unencrypted raw dump was performed by a built-in manufacturer command or by an external software utility developed by the SSD manufacturer, there would be no need disclose sensitive intellectual property.

The one obvious drawback to this solution is that it requires the commitment of valuable engineering resources on the part of the SSD manufacturer to implement the build-in command or to develop the external software utility. However, this is an investment that some SSD manufacturers might be willing to make in order to ensure the protection of their intellectual property.

Conclusion

The data doesn't lie. Solid state drives are reliable storage devices. That being said, failures can still occur, and when they do, there must be options for users to recover their lost files. Today's SSD data recovery techniques are expensive, slow and, in many cases, ineffective. This can all be changed through a collaborative effort between data recovery labs and SSD manufacturers. The collaboration starts with an understanding of what information the data recovery industry needs to do their work. This can be followed by a discussion of what information the SSD manufacturers are willing to provide, keeping in mind the sensitive nature of some topics. Gillware is confident that an open dialogue between SSD manufacturers and the data recovery industry will lead to data recovery solutions that not only match, but in many situations exceed what is possible with current HDD recovery techniques.

About the Authors

Scott Holewinski, President, Gillware Inc.

scott@gillware.com, (608) 237-8784

As President of Gillware Inc., Scott leads a team of sales and marketing professionals whose goal is to expand Gillware's data recovery and online backup businesses by establishing relationships with key strategic partners and building an active and thriving affiliate network. His team's efforts have resulted in contracts to support the data recovery needs for Dell, Western Digital and Intel customers. In addition, Gillware has built partnerships with over 1,500 affiliates nationwide who use their data recovery and backup services to support their clients. Beyond his work at Gillware, he has also helped to form three additional Madison-based start-ups. In 2006 he co-founded Phoenix Nuclear Labs (PNL) with Dr Greg Piefer, and Gillware Data Services, LLC with his other Gillware partners. In 2011 he and his partners spun off Shine Medical Technologies from PNL to pursue medical isotope production.

Greg Andrzejewski, Director of Research and Development, Gillware Inc.

As Director of Research and Development at Gillware Data Recovery, Greg has helped develop industry-leading hardware and software platforms to rescue data files from otherwise inaccessible storage devices. He pioneered solid-state recovery at Gillware in 2008 and has remained on the front lines of storage technology ever since. He is also well-versed in filesystems and their implementations. Greg holds a BSE degree in Computer Engineering from the University of Michigan.