This white paper explores the area of a hard disk drive not usually examined for data, what it is and how to access it

Forensic Imaging of Hard Disk Drives

-What we thought we knew

By Todd G. Shipley and Bryan Door

January 2012

All rights reserved

Copyright 2012, Todd G. Shipley and Bryan Door

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.

Mailing:
4790 Caughlin Pkwy #323
Reno, Nevada 89519

Some of the research in this document was done under a project supported by Award No. 2010-MU-MU-K021 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

WHAT WE HAVE BEEN TAUGHT

Imaging of hard drives has been the main stay of the "Science" part of digital forensics for many years. It has been articulated by many, including us, that we "forensically" image a hard drive to get that "Bit for bit" image of the ***ENTIRE*** contents of a hard drive. That concept is drilled into digital examiners in every class, vendors have built tools around that concept and the National Institute of Technology and Standards (NIST) has built a whole department (website included) around proving the science of hard drive imaging.

But what if all of this was not correct? What if you testified in court, "I imaged the entire drive and verified the process by producing a hash value that matched the original evidence.", and this statement is not entirely accurate.

Now, we are not intending to imply that every case that was ever testified about by a digital forensic examiner should be reopened and the testimony thrown out. On the contrary, what we want to do is tell a tale of ***data not known is data not seen*** or in our case imaged. Since the beginning of computer forensics, which has morphed into the broader field of "digital" forensics, the industry has espoused that imaging is the foundation of everything we do. With no forensic image we have no verifiable evidence. Certainly with live system analysis and cell phones the concept of repeatable imaging has been changed, but with a hard drive the basic tenant of "make no changes" to the hard drive is still the watch word for dead system imaging and an analysis.

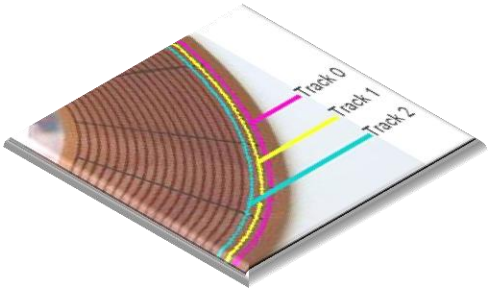

Figure 1 Typical misrepresentation of HDD track layout

We all use the standard industry forensic imaging tools available to us. We test the veracity of each of the tools we use. We read and reference the NIST projects that review the imaging tools and substantiate their ability to acquire a "forensically" sound image of the hard drive. All of this as we know it is accurate and true. The tools work and do acquire that repeatable image of the user accessible data on the hard drive. The issue for this paper is what is "user accessible" and what additional data on the hard disk drive is there and what doesn't get imaged.

The description of the fundamentals of hard drive operation and hard drive imaging is taught in every basic forensic course. Often the same or similar slides are used throughout these courses to describe the physical and mechanical parts of a hard disk drive. We all have seen the slide with the open hard disk drive case showing the platter, heads and spindle. Some slides will even describe in detail how the head assembly writes the ones and zeros to the disk surface. With that physical description we then use the image as the foundation upon how we determine the location of data on a hard drive (by LBA or Logical Block Address, a scheme used for specifying the location of blocks of data stored on computer storage devices) and then use that data to interpret a user's behavior. Using Google's advanced search feature simply search for "logical block addressing computer forensics" to find any number of PowerPoint's describing the LBA process.

Copyright 2012 – Todd G. Shipley and Bryan Door

So those of you who do true data recovery on hard drives are going "Ya so, I knew that", however, many in the digital forensics field have not been taught these physical structure details, nor do most digital forensics examiners have access to or even know about a PC-3000 (the PC-3000 is a specialized data recovery tool designed for the recovery of data from drives) or similar tools and their capabilities.

Don't confuse the hidden area on a hard disk drive that we are describing with an HPA (Host Protected Area), a DCO (Drive Configuration Overlay) or separate partition. The HPA and DCO can be used to set the user accessible size of a hard drive, effectively hiding areas of the disk from untrained investigators. This issue has been addressed and discussed in multiple forums and won't be discussed here. These are both logical constructs of the user data area and not the physical geometry of the hard disk drive.

THIS PROBLEM IS BROUGHT TO YOU BY THE LETTERS "P" AND "G"

So let's look at some new concepts and throw a wrench in our thought process. Did you know that potentially 30% of the hard drive you are examining is not in the image you are making? What? "I imaged the entire drive from beginning to end". Let's explore that fantasy and why what we have been saying is not totally correct, but it's not totally wrong either.

Logical Block Addressing is fundamentally where we say that we are getting the "entire hard drive from beginning to end", and we are in that sense acquiring all the data addressed with a logical block (LBA 0 through Mas LBA). This is the area that you speak of when you make the statement, "I made a bit for bit image of the entire hard drive".

Figure 2 Sample Representation of a Hard Disk Drives Track layout

The drive mechanically has space (tracks and sectors) in what is called the "Service Area" by most manufacturers. This area is the true beginning of the drive space and can occupy a significant amount of physical space on the hard drive. This area essentially contains the hard disk drives operating system.

When a hard drive is manufactured the hard disk drive surface is examined for defective areas on the platter's surface. Hard disk drives do not come out of the manufacturing process without defects on the drive platters. It would not be cost effective for the hard disk drive manufacturers to just toss out hard drives with defects. Instead, they simply mark the defective areas of the hard drive as bad, make a list of them in the Service Area, and ship the drive.

Hard drive manufacturers began compensating for these defects on the drive surface by building into the hard drive firmware (code used by the hard disk drive to operate) the ability to record these defects in lists. These lists are referenced by the hard disk drives operating system in order to prevent the hard disk drive from attempting to read or write data to the damaged sections of the hard disk drive.

Copyright 2012 – Todd G. Shipley and Bryan Door

Depending upon the manufacturer, there can be several lists. However, the common lists are referred to as the "P" and "G" lists. The "P" list is a list of defects, or what we refer to as "Bad" sectors, on the hard drive surface that are identified prior to the hard drive leaving the factory. The "G", or "grown", defect list is a list of defects that are "grown" on the hard drive during its use after leaving the factory. Okay so you are saying up to this point that this is not a bad idea, the hard drive is looking out for me and my data. Bad areas are marked bad and not used to store data. That is a good thing, and yes it is. It's how the hard drive then deals with the space it has marked as bad.

How the lists help to find your hard drive's data occurs in very different ways between the "P" and "G" lists. The "P" list has a real severe effect on how the hard drive later stores data. The factory review for defects is unique to each hard disk drive. The physical geometry layout of a hard disk drive is still defined by cylinders, heads and sectors. Add into that mix the number of platters and the functioning heads (a drive may have two platters and four heads but the manufacturer may only address three heads for that model) and you get a clearer picture of the complexity of the drive layout.

Once the drive layout is defined in the manufacturing process the search for defects can occur. Keep in mind that this is all occurring before the definition of the Logical Block Addressing scheme is applied. The "P" list consists of identified physical locations of "Bad" sectors on the hard disk drive. This list is then used to assist in the Logical Block Addressing of drive space for the writing of data. Here is the trick…During manufacturing, the layout of the LBA blocks occurs by shifting the LBA block sequence past the "Bad" sector. The LBA layer simply skips over the identified "Bad" location in the "P" list and continues sequentially in LBA order. The "P" list is a list of "skipped" physical locations on the hard disk drive platters.

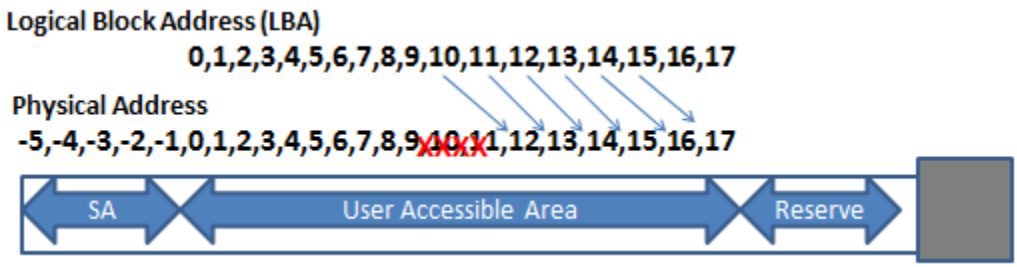

Figure 3 Simplified diagram of "P" list shifting

The "P" list creates the first layer of translation from the logical addressing to the physical address on the hard disk platter. The next level of translation exists with the "G" list.

Reserved Space and the "G" list

When addressing user accessible areas on a drive we generally make reference to these areas using Logical Block Addressing (LBA). User accessible area is almost always sector 0 through (max LBA) the last user accessible sector on a drive. Sectors on a disk are referenced by LBA number (i.e. sector

432,287,315) refers to a specific physical location of a drive. There is, however, other space on the drive called “Reserved Space”. This space is not in the Logical Block Addressing scheme (well not yet anyway).

In the Logical Block Addressing scheme the hard drive manufacturers do not want you to miss out on that data storage capability. So, it marks the sectors “Bad” and it allocates the same amount of sectors from a “Reserved” set of sectors in the “Service Area” to be the new address in the “Bad” logical block.

The “G” list or Grown Defects list contains a table of physical sectors on the drive that have been marked as “bad” and re-allocated since the drive was manufactured or re-conditioned. Once a sector has been marked “bad” and re-allocated by the hard drive it is no longer referenced by its original LBA number but rather a physical address which is stored in the “G” list module found in the “Service Area” of the hard disk drive.

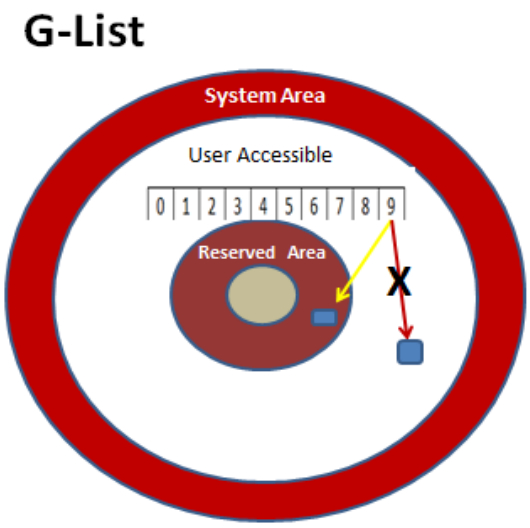

Figure 4 Representation of "G" lists effect on LBA

The original LBA number, for example LBA block # 355,422,123, needs to maintain it’s availability to the hard drive. It is simply re-mapped to a previously unused area of the disk. This action is referred to as “remapping”. The unusable drive space is remapped to usable space from the “Reserved Area” of the hard drive. This then maintains the logical sequence of Logical Block Addressing that the computer uses to find data. This “Reserved Area” is space that is inaccessible to the user unless an LBA block has been remapped to the space. Even then the only blocks available to the user are the remapped LBA blocks and nothing else. The remainder of the space is still addressed by the hard drive by the cylinder head and sector addressing scheme. The “Reserved Area” location depends on the manufacturer’s design and can be in different locations depending on the drive manufacturer’s specifications.

What the hard drive has done without the user knowing is swap one good sector for one bad sector. This maintains the balance of Logical Block Addresses and keeps the hard drive functioning normally. No data is missing and the drive has the same number of logical blocks. The imaging process we know follows this same path. Bad sectors are skipped and if they are remapped in the LBA map the hard drive provides the imaging tool with the data from the remapped sector and adds it as if it were part of the normal block. This is done regularly throughout the drive. How does this happen? The hard drive says it is only so large? While underlying the Logical Block Addressing is a separate physical addressing scheme.

There is also a translation table to ensure that all of this redirection occurs correctly and is done without any user knowledge, and up to this point any understanding of this process. So no harm no foul, one block is traded for another so you have an equal amount of sectors always in the LBA. The hard disk drive is very good at ensuring that we only have the requisite number of logical block addresses as listed by the manufacturer.

Copyright 2012 – Todd G. Shipley and Bryan Door

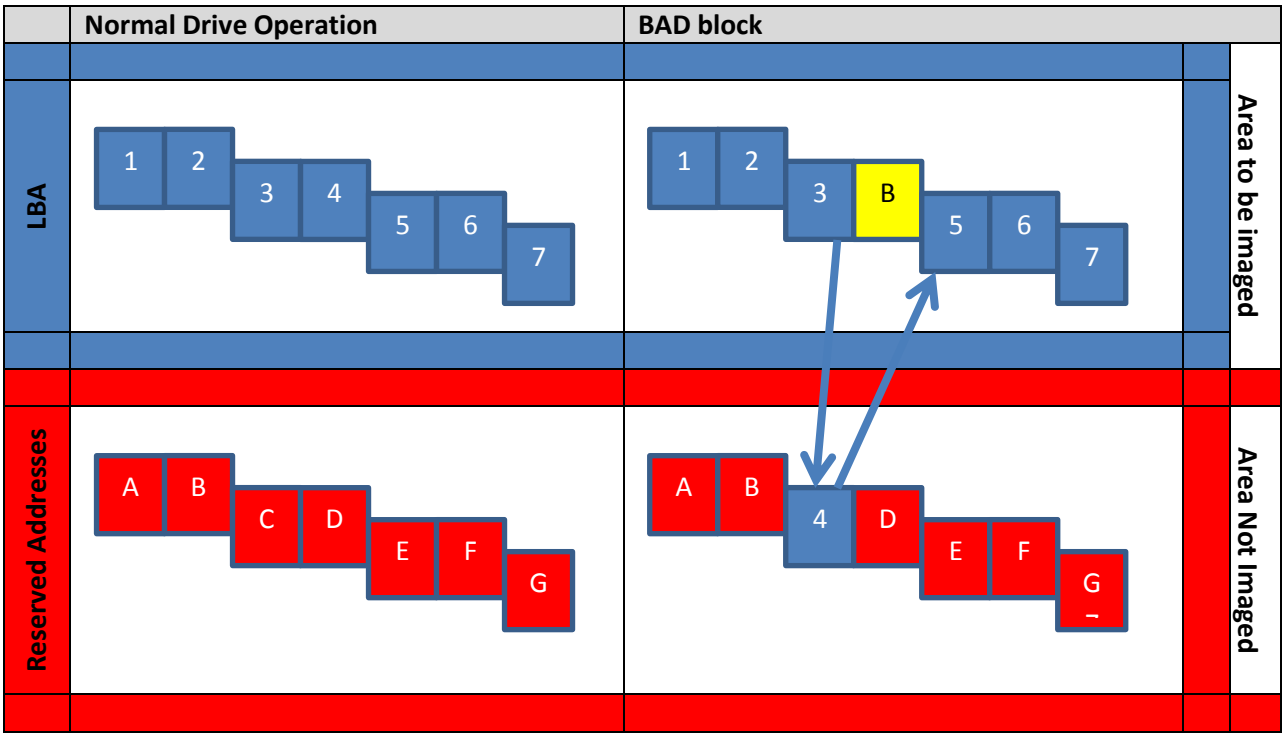

Figure 5 Remapping of a BAD sector to a sector in the reserved area

SERVO INFORMATION

Another area of space on the hard disk drive that exists but is not accessible is the Servo data. The Servo (data) Information is written to the disk by the hard drive manufacturer during the initial structural layout of the platters geometry. The Servo Data is basically markers, or road signs, that contain information used by the hard drive to control spindle rotation and head and actuator arm movement. The information guides the heads and helps to keep them properly on the track. It also tells the heads what track they are on. If this Servo Information is damaged it cannot be repaired.

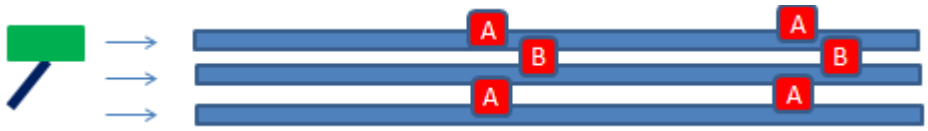

Figure 6 Simplified Description of Servo Data.

Older drives had several methods of laying out the Servo Data for the heads to follow. In some older manufacturing schemes, complete sides of a platter were used, or entire tracks became the Servo guides. These methods used a significant amount of drive platter physical space to properly direct the

Copyright 2012 – Todd G. Shipley and Bryan Door

heads around the platters. In modern drives the Servo Data is embedded in between the tracks and sectors allowing for more efficient direction to the heads and utilizing less physical drive space.

SERVICE AREA

The area most often misunderstood, and most of us are least aware of is the "Service Area" of hard disk drives. This is also commonly referred to as the "System Area", the terms are synonymous. The Service Area of a hard disk drive is used to store manufacturer data such as Servo information, firmware, and the drive defects tables to include the "P" and "G" Lists and translation table. The hard disk drive "SMART" data we are familiar with is also stored here.

The Service Area will contain many files referred to as "modules" or "adaptives". These modules are used by the hard disk drive to properly function. Hard disk drives boot up at power-on similar to a computer. The firmware in Read Only Memory (ROM) on the drive's Printed Circuit Board (PCB), starts on power-on and identifies the type of drive and certain parameters to startup. It makes a call to the Service Area of the hard disk drive, and various modules, containing information specific to that drive which are required to properly start the drive. These modules include the "P" and "G" list and translation table that tell the computer where the data is on the hard disk drive. This Service Area of the drive is sometimes referred to as being in "Negative Tracks" or "Negative Cylinders" of the disk. Hard disk drives can have hundreds of these negative tracks. They are referred to as negative because the user data area tracks start at number 0 and the data recovery tools place a minus sign in front of the track number. The Service Area tracks, depending on the manufacturer, may not physically appear before the 0 track of the User area. The Service Area location is based on the manufacturer's design of the drive and may be in the middle or at the end of the drive.

Although there are standards as to how to communicate with the Service Area of the drive (the Advanced Technology Attachment or ATA standards, specify the implementation of the controller on the disk drive itself-Technical Committee T13 at www.t13.org) each drive manufacturer implements the commands they desire for their specific make and model to communicate with this space. Although standard in their implementation, these commands are not found as an option in any of the digital forensic imaging tools to allow access to the data in the Service Area.

FIRMWARE

All of the information we have discussed about "P"and"G" lists and translation tables are the code found in the modules of the Service Area and which correspond with the data found in the PCB's ROM. The Read Only Memory on a hard disk drive PCB may be found in a single chip on the PCB or incorporated into a larger chip on the board. The firmware code found in the ROM of the PCB allows the hard disk drive to boot up. It tells the computer certain information about the hard disk drive's physical and logical locations of space on the drive and where to find the translation data on the disk. The ROM will

Copyright 2012 – Todd G. Shipley and Bryan Door

contain various information about the drive, sometimes including the firmware version, its serial number, a mapping of the heads and the location of modules in the Service Area. Post manufacture of the hard disk drive means that the firmware becomes unique to the particular hard disk it is "married to". With this uniqueness come problems with swapping PCB's from one drive to another. Improper swapping of PCB's can cause Service Area module corruption and the unrecoverability of the data from that hard disk drive.

RESOURCES AND TOOLS TO GET TO THIS DATA

The tools required to get to this data are unique to the data recovery field. Most examiners are not aware of the tools or have access to them. Additionally, there are very limited training opportunities on how to use these tools. Some of the common tools for accessing the Service Area of a hard disk drive include:

- PC-3000 by Ace Laboratory
- Atola by Atola Technology
- HD Doctor by Salvation Data

For an examiner in a government agency in the United States, these solutions may cause a problem. None of these solutions are manufactured in the U.S. Additionally, each of these tools approaches working in the Service Area in a different fashion. Also, the cost to use these tools can be prohibitive for a digital forensic examiner not regularly dealing with non-functioning hard disk drives.

CONCLUSION

The firmware and modules of a hard disk drive contain programs and configuration settings needed by the drive to operate. There are a very small number of tools that can read or modify the Service Area of a hard disk drive. So, for the time being this is not an everyday threat to digital forensic examiners cases. It is something however that we need to all be aware of now and in the future. Testifying about the imaging of hard disk drives needs to carry the caveat that we are imaging the "User accessible area" of the hard disk drive. We should not be saying we do a "bit for bit image of the entire hard drive". Data can be written to the Service Area, but those areas of the hard disk drive aren't accessible when forensically imaged utilizing the current industry tools. The current forensic imaging tools don't allow for imaging outside of the user accessible area and will not in the foreseeable future.

Digital forensic examiners need to be aware of current hard disk drive geometry as it truly exists. Additionally, digital forensics basic instructional courses should be updated to include a more thorough description of hard disk drive geometry and its physical layout. This will all provide the digital forensic examiner with a better foundation for statements and testimony made about their imaging process of hard disk drives. It will also provide the industry with a more clear understanding of the data accessible on a hard disk drive and the tools available to access that data.

Copyright 2012 – Todd G. Shipley and Bryan Door

References:

Hard Disk Drive Mechatronics and Control, Abdullah Al Mamun, GuoXiao Guo, Chao Bi, CRC Press, 2007

Hard Disk Drive Servo Systems, Ben M. Lee, Tong H. Lee, Kemao Peng, Venkatakrishnan Venkataramanan, Springer, 2nd Edition, 2006

http://en.wikipedia.org/wiki/Logical_block_addressing

http://www.wwpi.com/pdf/White%20Paper%20hard%20drive%20duplication.pdf

http://hddscan.com/doc/HDD_Tracks_and_Zones.html

Copyright 2012 – Todd G. Shipley and Bryan Door