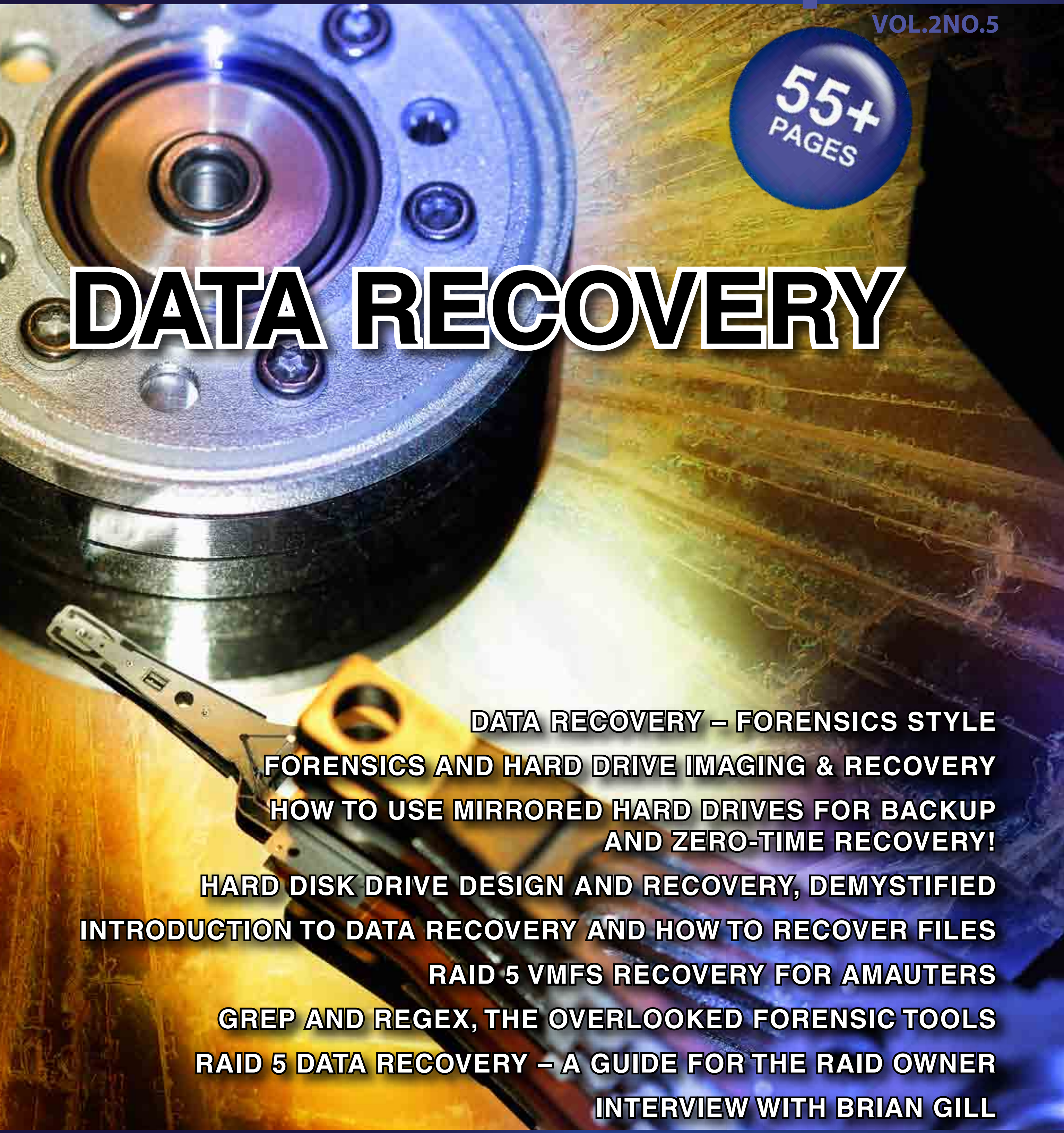
eForensics Magazine

Computer

VOL.2NO.5



55+ PAGES

DATA RECOVERY

DATA RECOVERY – FORENSICS STYLE

FORENSICS AND HARD DRIVE IMAGING & RECOVERY

HOW TO USE MIRRORED HARD DRIVES FOR BACKUP AND ZERO-TIME RECOVERY!

HARD DISK DRIVE DESIGN AND RECOVERY, DEMYSTIFIED

INTRODUCTION TO DATA RECOVERY AND HOW TO RECOVER FILES

RAID 5 VMFS RECOVERY FOR AMAUTERS

GREP AND REGEX, THE OVERLOOKED FORENSIC TOOLS

RAID 5 DATA RECOVERY – A GUIDE FOR THE RAID OWNER

INTERVIEW WITH BRIAN GILL

Issue 05/2013 (9) May

What do all these have in common?

DEPARTMENT OF DEFENSE
UNITED STATES OF AMERICA

DEPARTMENT OF JUSTICE
FEDERAL BUREAU OF INVESTIGATION
FIDELITY BRAVERY INTEGRITY

DEPARTMENT OF ENERGY
UNITED STATES OF AMERICA

DEPARTMENT OF THE AIR FORCE
UNITED STATES OF AMERICA
MCMXLVII

THE DEPARTMENT OF THE TREASURY
1789

They all use Nipper Studio

to audit their firewalls, switches & routers

Nipper Studio is an award winning configuration auditing tool which analyses vulnerabilities and security weaknesses. You can use our point and click interface or automate using scripts. Reports show:

1) Severity of the Threat & Ease of Resolution

2) Configuration Change Tracking & Analysis

3) Potential Solutions including Command Line Fixes to resolve the Issue

Nipper Studio doesn't produce any network traffic, doesn't need to interact directly with devices and can be used in secure environments.

SME
pricing from
£650
scaling to
enterprise level

evaluate for free at
www.titania.com

2012 Computing Security Awards
WINNER
Enterprise Security
Solution of the Year

2012 Computing Security Awards
WINNER
Network Security
Solution of the Year

2012 Computing Security Awards
Runner-up
SME Security
Solution of the Year

titania
nipper®
STUDIO

CLOUD

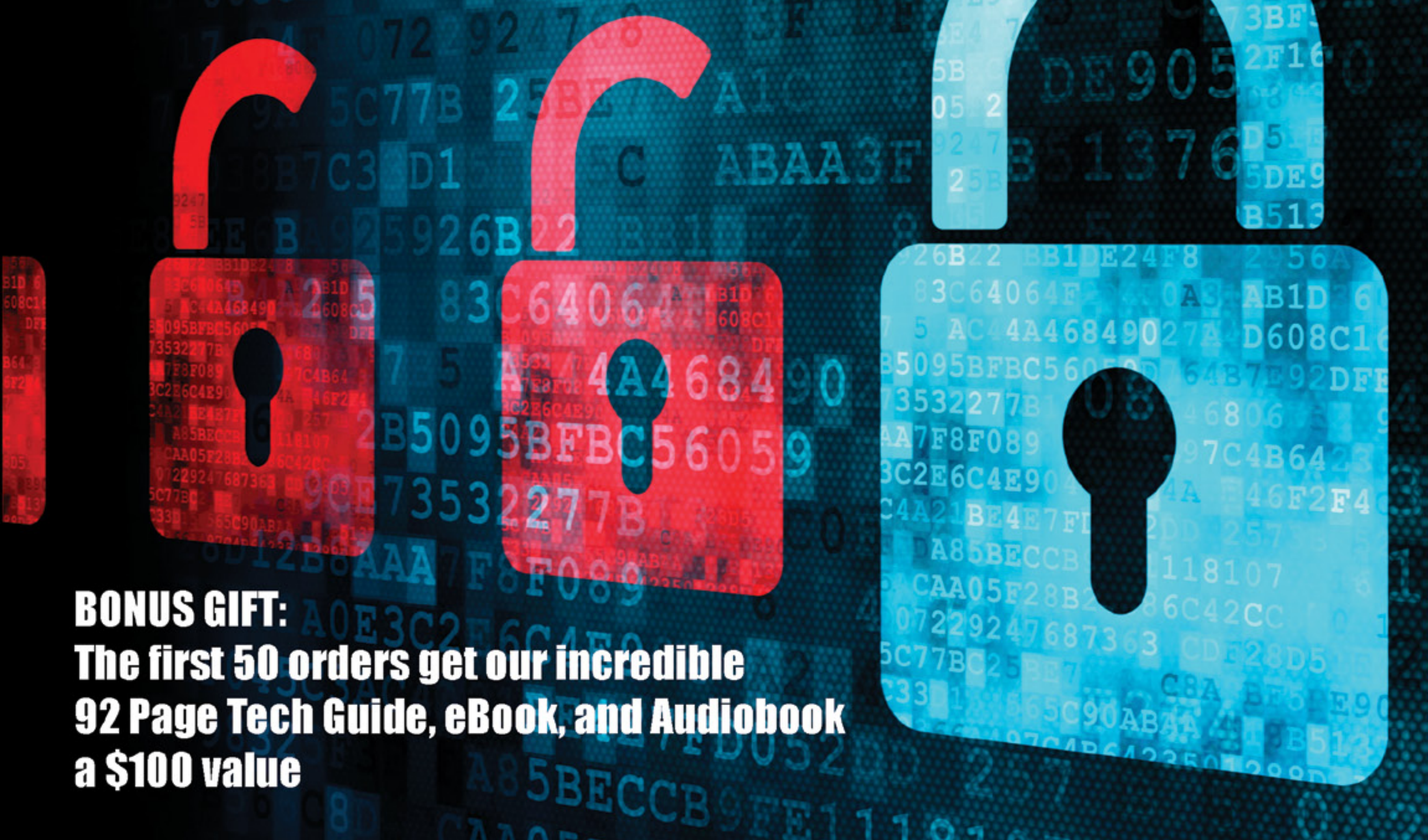
www.titania.com
T: +44 (0) 1905 888785

Make them hang on your every word...
Put them on the edge of their seats
when you speak...

Leave them wanting more...
It can happen, but ONLY IF YOU GET THIS:

THE ELECTRONIC ADVANTAGE: 101 the Basics

This fast-paced, 4-hour, online tutorial is for any
Skill level and even includes 10 case examples!
All this for just $360



BONUS GIFT:
The first 50 orders get our incredible
92 Page Tech Guide, eBook, and Audiobook
a $100 value

Go here now and order:
www.technologicalevidence.com

eForensics Magazine
TEAM

Editors: Artur Inderike
artur.inderike@eforensicsmag.com

Betatesters/Proofreaders:
Richard C. Leitz Jr., Olivier Caleff, Kishore P.V., Gabriele Biondo, Salvatore Fiorillo, Sundaraparipurman Narayanan

Senior Consultant/Publisher:
Paweł Marciniak

CEO: Ewa Dudzic
ewa.dudzic@software.com.pl

Art Director: Ireneusz Pogroszewski
ireneusz.pogroszewski@software.com.pl

DTP: Ireneusz Pogroszewski

Production Director: Andrzej Kuca
andrzej.kuca@software.com.pl

Marketing Director: Joanna Kretowicz
joanna.kretowicz@eforensicsmag.com

Publisher: Hakin9 Media Sp. z o.o. SK
02-682 Warszawa, ul. Bokserska 1
Phone: 1 917 338 3631
www.eforensicsmag.com

DISCLAIMER!

The techniques described in our articles may only be used in private, local networks. The editors hold no responsibility for misuse of the presented techniques or consequent data loss.

Dear Readers,

We would like to introduce our newest issue of eForensics Computer. Now you hold in your hands special edition about Data Recovery. As you probably know, the topic is very complicated, various and ... secret! Yes, yes! Companies and individual experts don't want to disclosure their recovery tools and techniques or moreover trade secrets of their clients. I tried to gather in this issue basics of data recovery for beginners who just start to delve into the area with advanced analysis of recovering processes and tools for professionals.

In the first article Donna Eno expains the importance of data recovery awareness in the modern world for everyone, espesially owners of businesses. This article is a kind of introduction to the topic and explains some basics of recovering data process.

Than we will focus on more practical approach. The second topic is rather for more advanced user who knows hard drive basics. The author Jonathan Yaeger from DataSavers LLC focused on real examples from his practice. Of course there will be some tips and advices given by author.

The forth article is a kind of demystification prepared by John from Epic Data Recovery Labs. He gives us based on his own practise advices and... warnings. The article will be extremely helpful and interesting for either beginners or professionals.

Secrets of successful data recovery you'll find in the article prepared by iCube Development. They will explain us how to prevent serious damage and failures which sometimes appeared during the data recovering process.

Dr Craig Wright our expert and regular author this time prepared overlook of GREP and Regular Expressions tools. He explains why using GREP and RegEx together make sence for forensics envestigators.

At the very end of this issue you'll find extremely interesting interview with CEO of Gillware, Inc. He gave us some details from his biography and spoke about urgent problems in data recovery sector. For example he presents his thoughts about comon mistakes in data recovery and his forensic recovery practice.

I hope you'll find this information interesting and useful. Thanks for your support and opinions. We're always trying to fit our magazine to all your requirements and needs.
It's my debut, so I'm open to all your suggestions and feedbacks. You can reach me at my email.

Enjoy your reading!
Artur Inderike
eForensics Team

DATA RECOVERY – FORENSICS STYLE 06
by Donna Eno, CCNA
So there you are...running your business or maybe you're at home catching up on emails etc. Minding your own business. Then you notice files are missing, your screen flashes, your network connection slows to a crawl.

10 **FORENSICS AND HARD DRIVE DATA IMAGING & RECOVERY: THE PERILS AND PITFALLS OF WORKING WITH DEFECTIVE HARD DRIVES**
by Jonathan R. Yaeger
This article will discuss some of the details of hard drive operation and failure, as related to digital data recovery or forensics. This will help the investigator to minimize compromises in evidence integrity.

HOW TO USE MIRROR DRIVE FOR BACKUP WITH ZERO-TIME RECOVERY! 14
by Dr. Wei Deng
With Mirror Drive technology, you can recover and replace a failed device with close to zero down time.

20 **HARD DISK DRIVE DESIGN AND RECOVERY, DEMYSTIFIED**
by John EDRL
Many in the Data Recovery industry know that HDD design is obviously not perfect, and can definitely use some improvement. Some Hard Drive brands and models are better than others. One flaw is the fact that most of the System Area for the Hard Drive reside on the Platters, where they are referenced.

DATA RECOVERY INTRODUCE AND HOW TO RESCUE DATA 26
by Tommy Redden
A helpful and fast method for recovering data in a file system. This method includes the steps of performing a given way that scans the deleted and lost data files on a computer, and preview the recovered data before you decide to recover it back with professional data recovery software.

32 **IF YOU ARE STUPID AND YOU KNOW IT...: RAID 5 VMFS RECOVERY**
by iCube Development
Basic hard drive data recovery is difficult, but a damaged RAID 5 Array containing VMWare ESX Virtual images, require an expertly trained technician and lab to properly perform a recovery without causing further damage to a client's data.

GREP AND REGEX, THE OVERLOOKED FORENSIC TOOLS 38
by Dr Craig S Wright GSE GSM LLM MStat
This article takes the reader through the process of learning to use GREP and Regular Expressions (RegEx). GREP May not seem to be a tool that relates to the process of data recovery, but we will show that this is an essential tool in recovering data. If you cannot find data, how can you recover it?

46 **RAID 5 DATA RECOVERY – A GUIDE FOR THE RAID OWNER**
by Wayne Horner
Your business stores data in a RAID 5 storage array and you just found out that the RAID has failed. Many RAIDs get damaged by the repair attempts. Your job is to keep an eye on well meaning IT consultants and in-a-hurry tech support centers.

INTERVIEW WITH BRIAN GILL 52
by Kishore P.V. and Richard C. Leitz Jr.

DATA RECOVERY – FORENSICS STYLE

by Donna Eno, CCNA

So there you are...running your business or maybe you're at home catching up on emails etc. Minding your own business. Then you notice files are missing, your screen flashes, your network connection slows to a crawl. You begin to wonder. You remember the conversation, just the other day with an IT friend over lunch about how more and more networks and the computers that sustain those networks are being compromised. You remember that conversation and some of the symptoms that a network and/or computer would display at the onset of a network attack. It bears a real resemblance to what you are experiencing sitting at your computer desktop. You get that sinking feeling in your gut as you wonder if it has happened to you and what it will take to minimize the damage....

What you will learn:

- Basic understanding of what digital forensics is
- Basic Forensic Recovery Methods
- Basic Forensic Rules of Evidence
- Basic Hash Encryption

What you should know:

- Basics of hard drive recovery
- Basic computer maintenance

Data Recovery speaks to recovering the data. Period. Regardless of how much is recovered and/or what the state of the files, whether or not they are readable, and what is/was the last disposition of those files. It is that... recovered data. However, if litigation of your losses is in the near future, then... Data Recovery will not be 'good enough.' Why? US Courts have determined that in order to substantiate whether or not electronic data is authentic and/or original... said data must adhere to some pretty strict guidelines. And so... this discussion will be regarding, Forensic Data Recovery. Data Recovery that will, at the end, hold up in court allowing you, the victim of network fraud, hacking etc., win in court and watch the bad guy go to jail.

First, let us revisit our victim, who has now realized that the computer and possibly the network has been infiltrated and files are missing, mayhem has ensued. He calls his IT friend who, very graciously has offered his assistance as well as his staff.

But this is a good place to stop. Before anything progresses from this point forward a few facts. FRE (Federal Rules of Evidence) state the following in order: "To satisfy the requirement of authenticating or identifying an item of evidence, the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is. " and part of that process or supportive of that process: "*Evidence About a Process or System*. Evidence describing a process or system and *showing that it produces an accurate result*. " (FRE Rule 801(a)(9)). Courts have ruled that hash values assigned to the electronic evidence satisfy this requirement. The term Hash Value is the application of an algorithmic equation being applied to the data file in such a way as to make it unique to all other files contained within a piece of evidence. There are varying types of hash values according to the algorithm equation such as, MD5, SHA1 and SHA2. In a Forensic Data Recovery ALL evidence artifacts (which are ALL files on an image, regardless of type) are assigned MD5 hash values. There have been studies shown where there is a duplicate MD5 hash value applied. These are called MD5 collisions and there have been many papers written regarding this anomaly. The short version of MD5 or HASH collisions is this: When the algorithm that is applied to a binary string produces the same number to a wholly different binary string, then that is what is defined as a MD5 collision because the algorithmic numbers match exactly when they should be different.

MD5 hash values are as 128-bit / 16byte hash values therefore MD5 collisions are rare but they do occur. So.. SHA1 and SHA2 hash algorithms were created and applied. SHA1 hash values are 160bit and SHA2 are 256bit hash algorithms using an entirely different methodology to arrive at their respective numbers. SHA1 and SHA2 hash algorithms are required by government entities due to the cryptographic security these hash algorithms provide. Problem solved.

But how does the hash value get applied so that the end result is: data recovered and bad guy goes to jail or pays a hefty fine? The answer is: Forensic Imaging.

Most Data Recovery companies will wish to take an image of the hard drive or electronic device in question, just to be able to then process that data, and extract out the recovered data. The Forensic image is a bit different and almost always, a Forensic Image is more in-depth. A Forensic Image is a bit stream image from the very first zero or bit at the very first position of the platter on the hard drive or electronic device to the very end. Assigning hash values along the way. It is very important to note that a Forensic Image captures absolutely everything including UnAllocated Space, UnPartitioned Space, File-slack, Deleted data... everything. Including the Service Area of a Hard drive or electronic device. A Service Area is the area of the electronic device that is for the indexing of the drive itself. With the advent of Smart drives, this area is reserved for drive index data that is 'outside' of the operating system and all other data included in the drive Smart software. This comes in handy when decrypting whole encrypted drives as the encryption data itself (the key) is stored outside of the operating system. So the process is as follows:

- The hard drive and/or electronic device is connected to a write blocker device. The write blocker device is a device that only allows data from the hard drive to be read and no electronic signal is able to go 'back' to the hard drive itself disallowing any ability for the data on the hard drive to be manipulated in any way.
- A clean (wiped so that all bits are zero) target drive is connected in such a way as to be able to 'capture' the data from the source drive as it is being read from the source drive. Thus a complete exact replica of the source drive is created along with the necessary hash values being assigned to each file, folder, directory and finally the whole drive.
- Verification of the written image is performed. This verification process is a process whereby all the hash values are computed again and verified with the first set of hash values computed to verify that they match. If no data has changed on the source drive since the imaging process began, then the hash values will match. IF so much as ONE BIT of data changed on the source drive since the beginning of the imaging process, the hash values will not match and questions will need to be answered as to why.

So what would cause the hash values to NOT match? Several things and they could be quite innocuous.

- Bad spot on the source drive itself. Since this is a bit-stream image, any bit not read correctly and/or properly would cause a hash value miss-match.
- Electrical surges. All of this equipment, computers, hard drives, electronic devices...all are subject to electrical power surges. To remedy this, verify that all pertinent equipment is plugged into a surge protector that is good up to 700 joules. Most lightening strikes are under that... so if you happen to be imaging in an area with lightening strikes... work with a surge protector, always.
- Lastly, operator interference. No one likes to get caught. So if someone realizes the game is up, they may be attempting to cover their tracks by removing data, the source drive etc., while the imaging process is occurring. This actually happened once when I was imaging a drive. One of the custodians (bad guys) logged onto the machine that was being imaged and began to erase data. Pretty impressive. Did that end the process in whole or in part? No. Why? Because all of the allegations still had to be proven in court. It did not matter at all that I had witnessed someone erasing data even as the image was being created and verified.

Keep that in mind. Just because an image may have been compromised does not stop the forensic process whatsoever. So...what did we do? Re-Start the image again with a new target drive (the drive being written to) so that a comparative analysis could be made between the first image created and the second one to determine what files and/or data was destroyed.

Remember the End Game. The bad guy going to jail and/or paying a hefty fine for their illegal activities. And this AFTER the court case is won. So... what do we know about the court system? Paperwork. Forensic Data Recovery is NOT exempt. In fact, it is *key* to a successful litigation process. So what kind of paperwork would be involved in a Forensic Data Recovery?

It is always good verify the requirements of the law to determine what specifically needs to be accomplished. FRCP Rule 26 speaks specifically to the evidence type of electronic discovery. It is rather lengthy and may be viewed here: FRCP Rule 26. The short version, with the addition of FRE 801, can be stated thusly: Copious documentation of said evidence will keep discussions regarding the Forensic Data Image to a minimum and on point. The following are steps towards good documentation:

- Photograph the evidence. Remember, once it has been determined that the computer and/or electronic device may have been compromised or has been compromised, and a Forensic Image required, the hard drive and/or electronic device is now considered evidence and must be treated as such. This is important because there is prosecution for tampering with evidence. So... begin by photographing the evidence. Labels that are sold with the hard drive can come off or become damaged. To successfully verify that the electronic data gathered via hard drive imaging matches the hard drive and/or electronic device that was imaged, take a photograph of the hard drive and/or electronic device. There are serial numbers that are on the label that will match the data kept in the Service Area of the hard drive which will be brought forward upon successful completion of the forensic imaging process. These numbers should match. If they do not, and there is a photograph of the hard drive / electronic device, there will be a good starting point to work back and discover how the label was changed etc. Photograph all sides of the evidence. This quells any argument that the evidence in question was/was not tampered with prior to / or after being forensically imaged.
- Chain of Custody. There needs to be paperwork showing the responsible party releasing the hard drive or evidence (it should now be called) into the hands of the individual performing the forensic image. This Chain of Custody paperwork should include the following at the minimum:
  - Make / Model and Serial Number of the evidence, any markings and /or labeling.
  - The name of the individual responsible for turning over the evidence for forensic imaging along with their signature
  - Date / Time-stamp of the transaction
  - The name of the individual responsible for performing the forensic image and signature.
  - The Make / Model and Serial Number of the drive to be used to capture the image (sometimes referred to as the Target Drive)
  - Date / Time-stamp of the Beginning of the Forensic Image process
  - Date / Time-stamp of the Ending of the Forensic Image process
  - Date / Time-stamp of the End of the Verification of the Forensic Image and whether or not the hash values assigned matched.
  - Any and all pertinent notes.
  - Case Number. Remember, if the decision has been made to go to court with the findings, a case number will be required.

This is just the very minimum. If the image has to be shipped to a Forensic Lab, then tracking numbers, signatures of people handling the evidence along the way also needs to be recorded.

So at the end of all this imaging, paperwork etc. .... there should be a bit-stream image exactly the same as the original down to the very last bit. This image can then be processed to determine what, if anything was taken as to documents etc., when, where it went to and if any other network devices that were connected to this device may have been compromised along with ability to take any findings in to a court of law to be litigated upon.

Let us revisit our victim. He has contacted his IT friend and based upon his IT friend's expertise what has been gleaned from the conversation has been determined that indeed there is a good chance something is up, and action needs to be taken. Question: Should the IT person (as good as they may be) take over from here and move to secure the network and gather evidence to be used in court if necessary? To secure the network? Absolutely! To produce the Forensic Images? Probably not. Why? This question is best answered with this question: What does the victim really want? Just to get the data back? Fine... let the IT friend have at it. Go to Court? In the over-litigious society in which we live, it

ON THE WEB

- Federal Rules of Civil Procedure Rule 26
- Federal Rules of Evidence Rule 801
- Federal Rules of Evidence Rule 702

would be best to have someone that can meet / exceed FRE (Federal Rules of Evidence) Rule 702 and let that standard be the guide as to defining a Forensic Expert. Why? Because all it takes is one good attorney and all bets are off and the victim could lose. The paperwork must be completed correctly, the investigation (which is the next step after completion of the Forensic Image) must be completed in a timely manner... and the list goes on.

If the victim chooses to go to court with the evidence that the network was compromised, the data is missing etc., the expertise in evidence acquisition will pay big dividends towards winning the case.

Author's Bio

IT person with 20+ years' education and experience in the field. Owner of D.Eno Forensics, a full service Digital Forensics company, and recently working with AccessData, LLC a forensic software company. Leveraging the previous 20 years of IT education and experience to provide expert IT analysis to Computer Forensic investigative work. Experience includes over 400 cases, written opinions, depositions and court testimony. www.denoforensics.com

adverisement

FORENSICS AND HARD DRIVE DATA IMAGING & RECOVERY

THE PERILS AND PITFALLS OF WORKING WITH DEFECTIVE HARD DRIVES

by Jonathan R. Yaeger

This article will discuss some of the details of hard drive operation and failure, as related to digital data recovery or forensics. This will help the investigator to minimize compromises in evidence integrity. The article will also serve as an introduction to best practices when data recovery is required.

What you will learn:

- Best practices for data acquisition, including turning off SMART and automatic sector reallocation
- The difference between data acquisition and data recovery
- When to stop imaging a drive
- Best practices for data recovery of forensic drives

What you should know:

- Hard drive basics
- How to image or copy a hard drive for forensic purposes
- The importance of using a write-blocker

A requisite of any forensic digital investigation is to preserve the evidence "as received." Alterations compromise or ruin the value of the evidence by introducing uncertainty, which a skilled, knowledgeable attorney may use to try to challenge or manipulate a legal proceeding or trial.

Indeed, the reputation and integrity of a digital investigator largely depends upon the ability to deliver "bulletproof" reports and analyses that can withstand courtroom challenges. Anything less might expose the investigator to professional liability and the possible stress and discomfort of having to defend one's work under cross-examination.

Making a forensic copy of a hard drive usually is done using a hardware imager, which incorporates a write-blocker function to thwart modifications to the source. Standards for imagers and other forensic procedures are described on the National Institute of Standards and Technology (NIST) website under "Information Technology Laboratory: Computer Forensics Tool Testing Program" (*http://www.cftt.nist.gov/disk_imaging.htm*).

As noted in the Digital Data Acquisition Tool Specification, "The ideal goal of the imaging process is to perform a complete and accurate acquisition of the digital source." [1] Typically, a source drive is copied bit-for-bit, creating a *bit* or *clone copy*.

The cloning process is fairly straightforward if the drive is in perfect operating condition, but that often is not the case. "Real-world" drives might develop bad sectors and other issues after prolonged use, and sometimes investigators have to work with damaged, defective or compromised drives.

When drives develop bad sectors but are otherwise operational, forensic imagers typically insert zeros onto the image, in place of the unread data, and note it in a log. Although questions remain about the unread or missing data, this practice is widely accepted as the standard for handling bad sectors.

BEST PRACTICES FOR DATA ACQUISITION

Hard drives incorporate *housekeeping routines* into their firmware. These run in the background and are invisible to the end user. One standard routine is defect management, in which data in weak or failing sectors are moved to another region of the drive. The bad sectors are "marked out" and included in the *grown defect* list, also called the *G-List*.

Although the G-List sectors are marked out, they still contain data (often corrupt), even after a low-level format of the drive. These sectors are released only if they are deliberately cleared. Complete forensic analysis entails examining the data within the G-Listed sectors, but that topic is beyond the scope of this article.

Another housekeeping function is regularly updating the drive's S.M.A.R.T. (Self-Monitoring, Analysis and Reporting Technology) logs.

These logs keep track of a myriad of indices, called *attributes*, of a drive's performance. Examples of attributes include the number of times the drive is started up, the inability to read sectors, and updates to error logs. These attributes are expressed in terms of "threshold exceeded" and "threshold not exceeded" values, which can provide an early warning of a drive's impeding failure (Figure 1).

Using a write blocker when imaging a digital source is a standard requirement to preserve the fidelity of the drive. However, there are other precautions that should be taken, which might or might not be among the standard features of a digital imager:

- Disable the drive's defect auto-reallocation, i.e. updates to G-List.
- Disable the drive's S.M.A.R.T. logging capability.
- In coordination with the file identification and recovery phase, correlate and log the locations of the specific bad (or missing) sectors with their respective file names.

Andrei Shirobokov – chief technology officer of DeepSpar Data Recovery Systems, a manufacturer of data imagers with forensic features – said: "Unfortunately, the ATA specification does not have a command to turn off auto-relocation. Therefore imaging software should use vendor-specific ATA commands to do this." [4]

"A similar problem exists with Self-Monitoring Analysis and Reporting Technology (S.M.A.R.T.) attributes," he continued. "The drive firmware constantly recalculates S.M.A.R.T. attributes, and this process creates a large amount of overhead that increases imaging time and the possibility of further drive degradation. Imaging software should be able to disable S.M.A.R.T. attribute processing." [5]

In other words, a failing drive will constantly try to update the logs, and because it is failing, it might not be able to update them successfully. Thus, the drive gets caught in a vicious cycle that interferes with imaging and can lead to complete drive failure.

Note that commands to turn off defect auto-reallocation are vendor-specific and often unpublished.

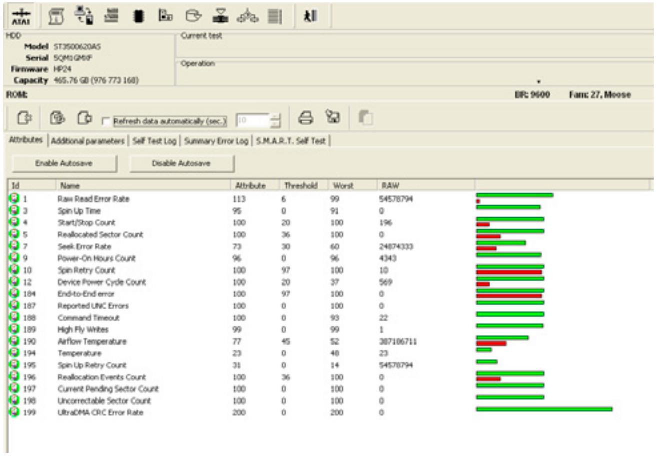
The BIOS of most PCs offers a menu choice to turn off S.M.A.R.T. notifications of impending drive failure, but that feature does not affect the internal S.M.A.R.T. operations of the hard drive.

A standard ATA command to disable a drive's S.M.A.R.T. operations is described in section 7.52 of the T13 ATA standard, which states: "This command disables all S.M.A.R.T. capabilities within the device including any and all timer and event count functions related exclusively to this feature." [6]

The command is *B0h with a Feature register value of D9h.* Correlation of bad sectors to specific files, or the lack of that capability, is typically a part of the software feature set of some, but not all, forensic imagers. From a forensic standpoint, obviously it is important to know which files might be affected by bad or missing sectors.

DATA ACQUISITION VS. DATA RECOVERY

At some point, the drive might not be able to be completely imaged, and data acquisition might become or require data recovery. To differentiate, the optimum is a complete acquisition; for example, "If for every bit of the digital source there is a corresponding bit in the destination object, and for every bit representing acquired data in the destination object there is a corresponding bit in the digital source." [2] Alternately, "Data recovery is the process of salvaging data from damaged, failed, corrupted or inaccessible secondary storage media when it cannot be accessed normally." [3]



HDD
Model ST3500620AS
Serial 5QM1GMXF
Firmware HP24
Capacity 465.76 GB (976 773 168)

ROM: BR: 9600 Fam: 27, Moose

Id	Name	Attribute	Threshold	Worst	RAW
1	Raw Read Error Rate	113	6	99	54578794
3	Spin Up Time	95	0	91	0
4	Start/Stop Count	100	20	100	196
5	Reallocated Sector Count	100	36	100	0
7	Seek Error Rate	73	30	60	24874333
9	Power-On Hours Count	96	0	96	4343
10	Spin Retry Count	100	97	100	10
12	Device Power Cycle Count	100	20	37	569
184	End-to-End error	100	97	100	0
187	Reported UNC Errors	100	0	100	0
188	Command Timeout	100	0	93	22
189	High Fly Writes	99	0	99	1
190	Airflow Temperature	77	45	52	387186711
194	Temperature	23	0	48	23
195	Spin Up Retry Count	31	0	14	54578794
196	Reallocation Events Count	100	36	100	0
197	Current Pending Sector Count	100	0	100	0
198	Uncorrectable Sector Count	100	0	100	0
199	UltraDMA CRC Error Rate	200	0	200	0

Figure 1. *S.M.A.R.T. attribute display*

The need for data recovery might be immediately obvious: A hard drive might be physically damaged, clicking or failing to spin up. At other times, the drive might stop during the imaging process. In any case, a drive failure places the investigator in a bit of a dilemma: Data recovery, by definition, changes the collection of data from a hard drive from “as received” to “as recovered.” The recovery process might entail making changes to the drive’s data in order to get it to function again.

Data recovery is a salvage process, and there might be no other practical alternative for retrieving the data. The data recovery specialty is distinct from forensic data acquisition in that it requires additional equipment and expertise. Although many investigators are skilled in data recovery, some might subcontract the data recovery phase to a third party.

At this point, some investigators begin or subcontract the data recovery process without regard to the possible forensic consequences of the various recovery steps employed. What an investigator chooses to do in these situations can affect the integrity of the results, and therefore the ability of competent counsel or an expert witness to poke holes in the validity of forensic reports and conclusions generated from the data.

THE IMPORTANCE OF IDENTIFYING HARD DRIVE FAILURE – AND WHEN TO STOP IMAGING

Hard drives often fail in a cascading mode. For example, a drive’s head might become dirty and fail to read or write data. Good sectors will be reported as bad, and the drive will try to move the data to another sector (i.e., through reallocation). If it cannot successfully write to the new location and verify the written data, then it will report the new sector as bad, too. Each time it moves an allegedly bad sector, it will update both the G-List and S.M.A.R.T. logs.

The G-List log firmware module has a limited capacity, so continuous updating can fill up the space and overflow to corrupt other essential firmware modules. Sometimes that is the reason why a drive that takes an inordinately long time to image will die before the process has been completed.

Turning off S.M.A.R.T. and bad sector reallocation can help prevent this failure. However, a drive could also have a bad or contaminated head that is slowly ruining the surfaces of the platters, leading to contamination and complete drive failure, in which further recovery efforts are fruitless.

The problem is that the drive acts much the same in both situations; the defect logs will be updated, and there is no easy way to discern whether this is a problem that will eventually ruin the drive.

Therefore, if a drive continuously clicks, is slow to image or appears to be slowing down, it is important for the client to *stop the imaging process* and proceed to data recovery. Failure to recognize when a drive is failing can ruin the chances of a successful investigation.

BEST PRACTICES FOR DATA RECOVERY

Best practices for the data recovery phase might include:

- Making as few changes as possible in the source drive.
- Trying to conduct the recovery in a manner so you can always revisit or undo prior recovery steps.
- As with the data acquisition process, logging any data recovery steps that can change the data written to as well as reported from the source drive.

Often, the data recovery process is not logged or annotated, especially if the recovery work is outsourced to a third party, and particularly if that third party is not well-versed in forensics procedures. The failure to notate the relevant actions performed might be a result of ignorance, lack of practice or resistance due to the secretive nature of the data recovery industry. Most data recovery firms are unwilling to expose or share proprietary recovery methods.

Forensic evidence, such as a hard drive, requires a documented chain of custody. Sending a drive to a third party that does not document relevant changes creates a gap in the record, which can reduce the reliability of the evidence and related reports.

The following are examples of hard drive recovery steps that would not normally change or impact the integrity of the source:

- Changes or repairs to the printed circuit board (PCB) when there is no significant adaptive information in the ROM, or when the ROM from the source drive is physically transplanted to a replacement PCB;
- Head swaps;
- Motor or platter swaps;
- Replacement of firmware modules that are not specific to a particular drive and may be copied from one hard drive to another;
- ROM regeneration using service area (SA) data from the subject drive.

Alternately, these steps are likely to impact or change the data and are to be performed only as required. Logging is recommended, as is following other best practices:

- Making any modifications to the drive’s factory or grown defect lists, such as clearing the grown defect list (G-List);

- Using scripted programs that might impact Relo blocks, T-lists and other track or sector management firmware;
- Repairing or rebuilding a translator module;
- Modifying or substituting other firmware modules;
- Cutting a head in the case of media-damaged drives.

There is overlap between the best practices of data recovery (DR) and forensic digital examination (FDE). For example, a best practices principle for DR is to try to read and save all firmware modules and service tracks. The default in many situations is to ignore modules and service tracks that are corrupted or otherwise cannot be read in their entirety.

However, in a small percentage of cases, data may be deliberately hidden within modules and tracks that the drive does not normally access or use, such as the factory logs. Therefore, FDE practices encourage forced reading of defective firmware for retention and analysis. If a defective module is repaired or replaced, the original can be compared to the replacement. Likewise, if a corrupted G-List requires deletion or replacement, the original can be inspected for hidden data.

Cutting a head is an extreme measure in the case of media damage, in which continuing to operate the damaged head can cause additional contamination that will ruin recovery efforts. The missing data areas will show up as regular bands of null data, assuming the target device has been completely erased and filled with zeros. As a best practice, target drives should be checked in advance for successful erasure at multiple points.

Performing a *smart hot swap* is another technique used by data recovery technicians. This involves copying certain critical firmware modules to a donor hard drive and then using the donor to "jump-start" a defective drive.

If employed properly, the smart hot swap technique normally will not alter the data acquired. However, it might be prudent to document the procedure anyway.

The final best practice is that when in doubt, document anyway. It is a good idea to document every step internally and then produce a subset report to the client in order to preserve proprietary aspects of data recovery.

COORDINATING WITH A THIRD-PARTY DATA RECOVERY FIRM (OUTSOURCING)

Once a hard drive is delivered to a third party for data recovery, control over the process is relinquished. Although a chain of custody is normally established with the firm, what happens next can be a void if the contractor does not have established best practices and procedures.

This vulnerability can be mitigated in part by ensuring that the recovery firm understands the requisites of forensic data recovery and has established procedures in place for documenting pertinent parts of the recovery process.

CITATIONS

[1] National Institute of Standards and Technology. (2004). Disk Imaging: Digital Data Acquisition Tool Specification (Draft 1 of Version 4.0, October 4, 2004). Section 4.0 "Background", lines 124-125. *http://www.cftt.nist.gov/Pub-Draft-1-DDA-Require.pdf*

[2] National Institute of Standards and Technology. (2004). Disk Imaging: Digital Data Acquisition Tool Specification (Draft 1 of Version 4.0, October 4, 2004). Section 5.0 "Definitions", lines 197-202. *http://www.cftt.nist.gov/Pub-Draft-1-DDA-Require.pdf*

[3] Wikipedia. Data recovery. http://en.wikipedia.org/wiki/Data_recovery

[4] DeepSpar Data Recovery Systems. White Paper. Disk Imaging: A Vital Step in Data Recovery by Andrei Shirobokov. *http://deepspar.com/pdf/DeepSparDiskImagingWhitepaper3.pdf*

[5] op. cit.

[6] T13.org (2006). Working Draft Project American National Standard T13/1699-D. Information technology – AT Attachment 8 – ATA/ATAPI Command Set (ATA8-ACS) Section 7.52.2. *http://www.t13.org/documents/UploadedDocuments/docs2006/D1699r3f-ATA8-ACS.pdf*

BIBLIOGRAPHY

Carrier, Brian File System Forensic Analysis. Upper Saddle River, N.J: Pearson Education, Inc., 2005.

CONCLUSION

A defective or failing drive presents a challenge to the forensic data acquisition process in terms of maximizing the fidelity of the recovered data. The steps that are used during the data recovery process often are undocumented, leaving the investigator vulnerable to legal challenges in which insufficient or no documentation exists.

Investigators would be wise to develop their own sets of best practices when performing data recovery and insist the same from third-party contractors involved in the process.

Author's Bio

Jonathan Yaeger is the owner of Data Savers, LLC, a leading data recovery firm in Atlanta, Georgia, since 2005. Data Savers, LLC offers high customer service and affordable recovery rates (www.datasaversllc.com). After working for E-Tech in 1980 as a technical sales manager for energy conservation products, Yaeger started Atlanta Technical Specialists, Inc. (ATS) in 1987. ATS, which was sold in 1999, offered PC sales, manufacturing, and component-level service. He has also been involved in a number of high tech ventures including AppForge, Inc., PhysicianAssist, and Atlanta Mac Service. Yaeger holds a Bachelors Degree from Emory University. He is an ative member of the North Atlanta Rotary Club, and resides in Atlanta, GA. jon@datasaversllc.com, Data Savers, LLC

HOW TO USE MIRROR DRIVE FOR BACKUP

WITH ZERO-TIME RECOVERY!

by Dr. Wei Deng

With Mirror Drive technology, you can recover and replace a failed device with close to zero down time. The state-of-the-art technology first converts and compares all files, then clone only the changed data to the hard drive, providing you with a high-grade-speed to complete the Mirror Drive process.

What you will learn:

- You will learn how to use DriveClone's Mirror Drive.
- The article will teach you how to boot Mirror Drive on different computers, making migrate/replace new PC simple and easy

What you should know:

- Standarts of data recovery
- RTO
- Computer maintanance

The safest way to back up important data is to duplicate said data to an external storage device to achieve physically-isolated protection. However, the recovery process of traditional backup software is long and tedious, and can negatively affect your business operations.

With incremental cloning and other functions, DriveClone becomes an alternative and effective backup solution. Incremental cloning allows you to back up recently changed data quickly, and restoration only requires a boot up; business can operate without fear of down time in case of disaster.

PERFECT DEFRAG CLONING

During the process of Mirror Drive, DriveClone will re-organize the whole hard drive data. Consequently, the destination hard drive's performance is on average 20% faster than the original by keeping all blocks of files in order.

SMART CLONING

In order to optimize clone speed and reduce destination drive's size, DriveClone intelligently excludes temporary files, hibernation files, memory swap files, etc. from being cloned to the destination drive.

DISSIMILAR/UNIVERSAL BOOT CLONING

DriveClone keeps the cloned drive in a "Universal Bootable Format", allows it to boot on different computers.

TRUE UEFI SYSTEM SUPPORT

UEFI has more advantages, such as larger capacity, superior performance, GPT disk supported, etc.As such, nearly all new PC motherboards have adopt UEFI technology. Farstone have made joint efforts with motherboard producers to perfect the support of UEFI-based mainboard. Thus far, we are the only one who supports dissimilar recovery of UEFI motherboard.

MULTI-VERSION FILE RECOVERY

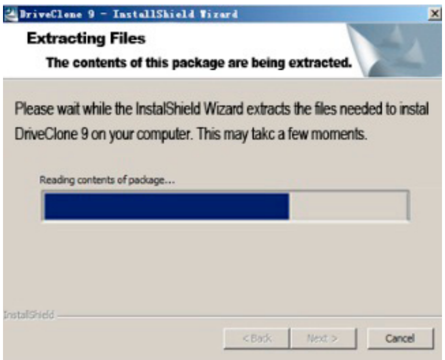
Farstone is the first and best firm that have incorporate cloning as true backup solution. Driveclone can clone iterations of files, automatically or specified, allowing restoration to any old data possible.

CREATE& USE A MIRROR DRIVE USING DRIVECLONE

Please insert a hard drive directly to a PC or through an USB enclosure.

1. Plug an USB hard drive

2. Install DriveClone



DriveClone 9 - InstallShield Wizard

Extracting Files

The contents of this package are being extracted.

Please wait while the InstalShield Wizard extracts the files needed to instal DriveClone 9 on your computer. This may takc a few moments.

Reading contents of package...

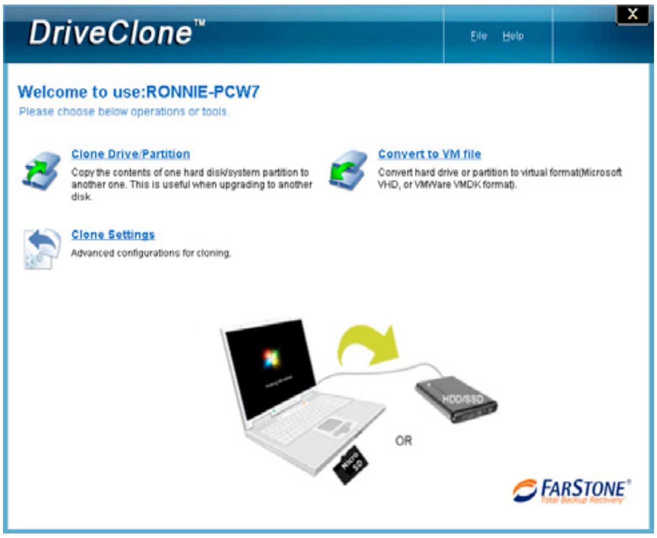
3. DriveClone will clone the computer's hard drive to an USB hard drive

4. Remove hard drive from USB enclosure

5. Replace hard drive on computer with the newly cloned hard drive

START UP YOUR PC AFTER LINKING THE HARD DRIVES

1. Run DriveClone, and click Clone Drive/Partition.



DriveClone™

Welcome to use:RONNIE-PCW7

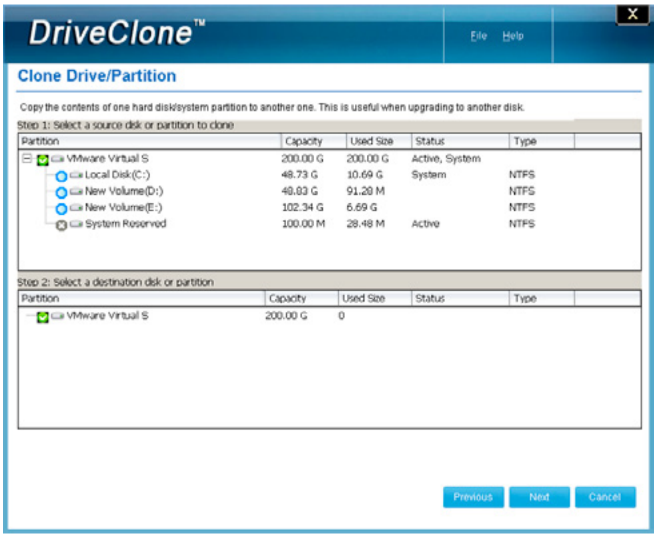
Please choose below operations or tools.

Clone Drive/Partition
Copy the contents of one hard disk/system partition to another one. This is useful when upgrading to another disk.

Convert to VM file
Convert hard drive or partition to virtual format(Microsoft VHD, or VMWare VMDK format).

Clone Settings
Advanced configurations for cloning.

2. Check source hard drive/partition and destination hard drive/partition.



Clone Drive/Partition

Copy the contents of one hard disk/system partition to another one. This is useful when upgrading to another disk.

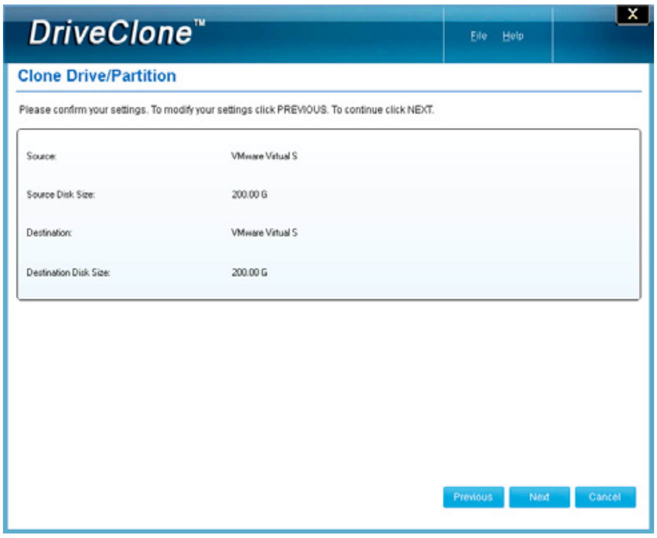
Step 1: Select a source disk or partition to clone

Partition	Capacity	Used Size	Status	Type
☑ VMware Virtual S	200.00 G	200.00 G	Active, System	
○ Local Disk(C:)	48.73 G	10.69 G	System	NTFS
○ New Volume(D:)	48.83 G	91.28 M		NTFS
○ New Volume(E:)	102.34 G	6.69 G		NTFS
☒ System Reserved	100.00 M	28.48 M	Active	NTFS

Step 2: Select a destination disk or partition

Partition	Capacity	Used Size	Status	Type
☑ VMware Virtual S	200.00 G	0		

3. Verify the information and click next.

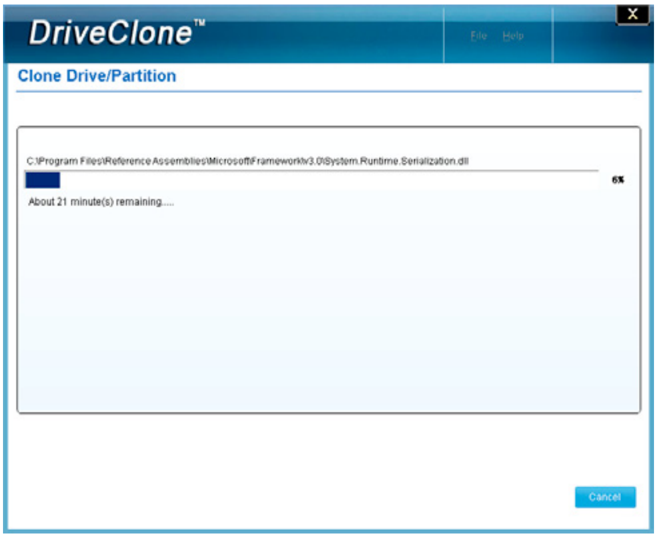


Clone Drive/Partition

Please confirm your settings. To modify your settings click PREVIOUS. To continue click NEXT.

Source:	VMware Vitual S
Source Disk Size:	200.00 G
Destination:	VMware Vitual S
Destination Disk Size:	200.00 G

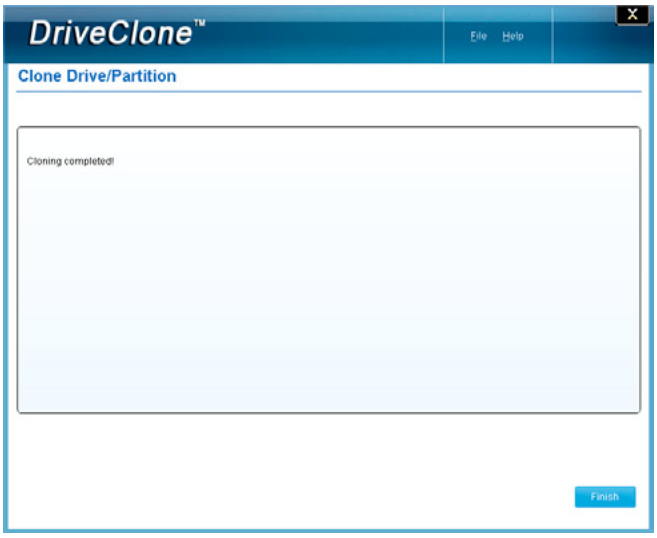
4. Cloning progress.



Clone Drive/Partition

```
C:\Program Files\Reference Assemblies\Microsoft\Framework\v3.0\System.Runtime.Serialization.dll
6%
About 21 minute(s) remaining....
```

5. Cloning completed.



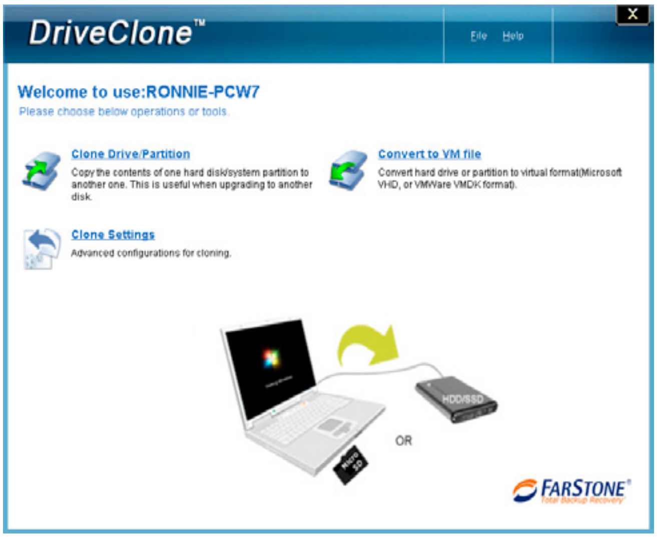
Clone Drive/Partition

Cloning completed!

USE DRIVECLONE TO CONVERT SYSTEM TO VM FILE

DriveClone's other main feature is the ability to convert a running OS to a virtual machine format. The resulting image can be launch directly by a virtual machine. At this moment, DriveClone supports VMWare and Hyper-V format.

1. Run DriveClone, and choose Convert to VM file.



Welcome to use:RONNIE-PCW7

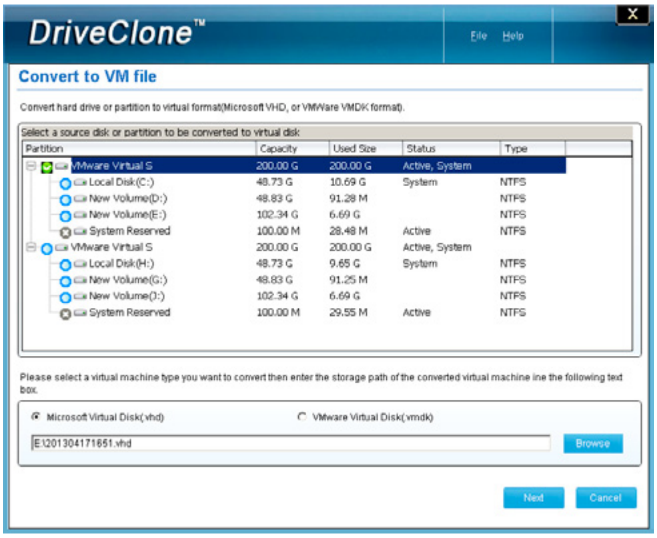
Please choose below operations or tools.

Clone Drive/Partition
Copy the contents of one hard disk/system partition to another one. This is useful when upgrading to another disk.

Convert to VM file
Convert hard drive or partition to virtual format(Microsoft VHD, or VMWare VMDK format).

Clone Settings
Advanced configurations for cloning.

2. Check the hard drive/partition that needs to be converted, choose a virtual machine format, and specify the storage path.



Convert to VM file

Convert hard drive or partition to virtual format(Microsoft VHD, or VMWare VMDK format).

Select a source disk or partition to be converted to virtual disk

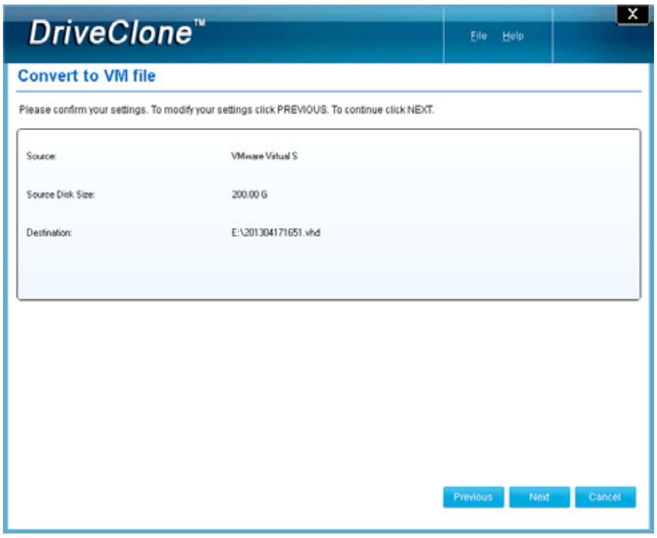
Partition	Capacity	Used Size	Status	Type
VMware Virtual S	200.00 G	200.00 G	Active, System	
Local Disk(C:)	48.73 G	10.69 G	System	NTFS
New Volume(D:)	48.83 G	91.28 M		NTFS
New Volume(E:)	102.34 G	6.69 G		NTFS
System Reserved	100.00 M	28.48 M	Active	NTFS
VMware Virtual S	200.00 G	200.00 G	Active, System	
Local Disk(H:)	48.73 G	9.65 G	System	NTFS
New Volume(G:)	48.83 G	91.25 M		NTFS
New Volume(J:)	102.34 G	6.69 G		NTFS
System Reserved	100.00 M	29.55 M	Active	NTFS

Please select a virtual machine type you want to convert then enter the storage path of the converted virtual machine ine the following text box.

● Microsoft Virtual Disk(vhd) ○ VMware Virtual Disk(vmdk)

E:\201304171651.vhd

3. Verify the information and click next.



Convert to VM file

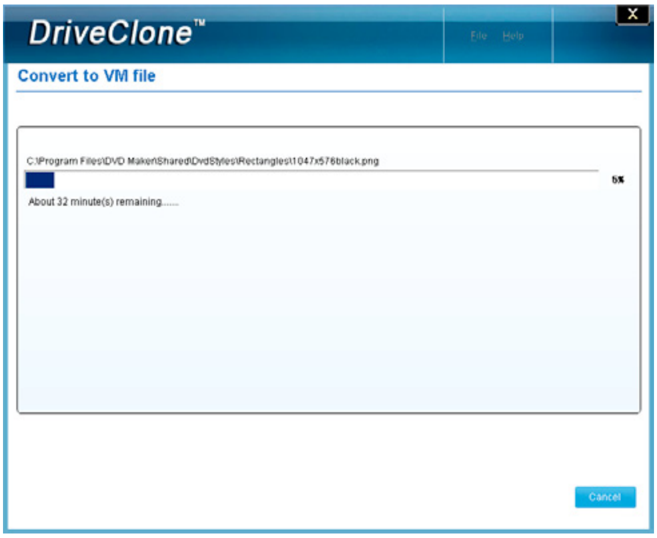
Please confirm your settings. To modify your settings click PREVIOUS. To continue click NEXT.

Source: VMware Vitual S

Source Disk Size: 200.00 G

Destination: E:\201304171651.vhd

4. Conversion progress.



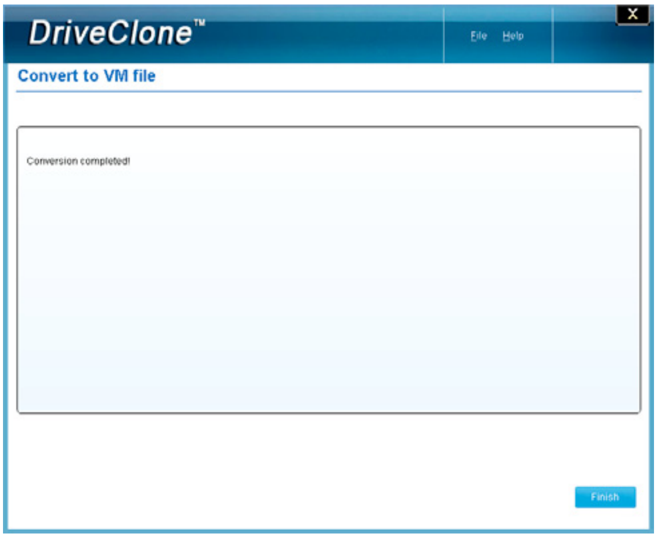
Convert to VM file

C:\Program Files\DVD Maker\Shared\DvdStyles\Rectangles\1047x576black.png

5%

About 32 minute(s) remaining......

5. Conversion complete.

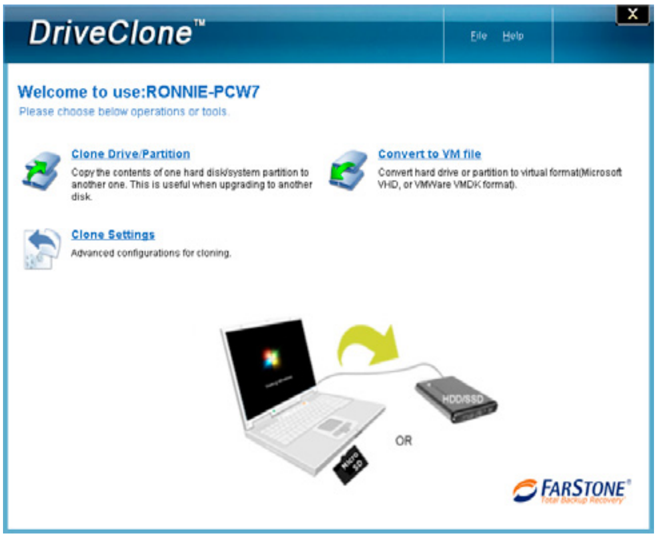


Convert to VM file

Conversion completed!

HOW TO RUN INCREMENTAL CLONING?

1. Run DriveClone, and choose Clone Setting.



Welcome to use:RONNIE-PCW7

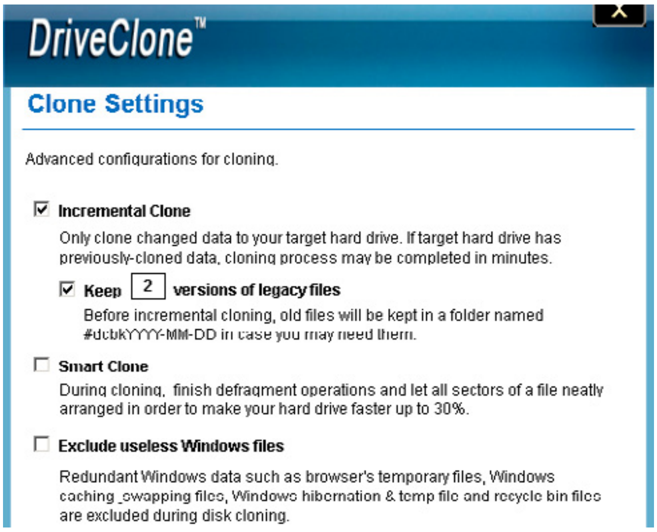
Please choose below operations or tools.

Clone Drive/Partition
Copy the contents of one hard disk/system partition to another one. This is useful when upgrading to another disk.

Convert to VM file
Convert hard drive or partition to virtual format(Microsoft VHD, or VMWare VMDK format).

Clone Settings
Advanced configurations for cloning.

2. Enable Incremental Clone; thereafter, you can check Keep X versions of legacy files.



Clone Settings

Advanced configurations for cloning.

☑ Incremental Clone

Only clone changed data to your target hard drive. If target hard drive has previously-cloned data, cloning process may be completed in minutes.

☑ Keep 2 versions of legacy files

Before incremental cloning, old files will be kept in a folder named #dcbkYYYY-MM-DD in case you may need them.

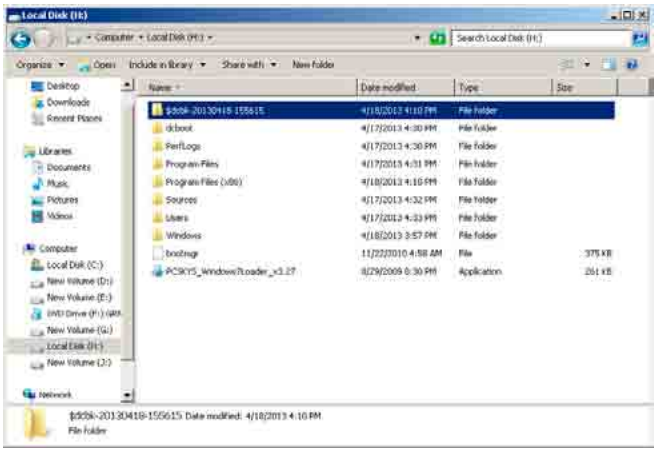
□ Smart Clone

During cloning, finish defragment operations and let all sectors of a file neatly arranged in order to make your hard drive faster up to 30%.

□ Exclude useless Windows files

Redundant Windows data such as browser's temporary files, Windows caching ,swapping files, Windows hibernation & temp file and recycle bin files are excluded during disk cloning.

3. Legacy files are saved in the corresponding partition of destination hard drive, and the folder name contains creation time.



Name	Date modified	Type	Size
$dcbk-20130418-155615	4/18/2013 4:10 PM	File folder	
dcboot	4/17/2013 4:30 PM	File folder	
PerfLogs	4/17/2013 4:30 PM	File folder	
Program Files	4/17/2013 4:31 PM	File folder	
Program Files (x86)	4/18/2013 4:10 PM	File folder	
Sources	4/17/2013 4:32 PM	File folder	
Users	4/17/2013 4:33 PM	File folder	
Windows	4/18/2013 3:57 PM	File folder	
bootmgr	11/22/2010 4:58 AM	File	375 KB
PCSKYS_Windows7Loader_v3.27	8/29/2009 0:30 PM	Application	261 KB

$dcbk-20130418-155615 Date modified: 4/18/2013 4:10 PM
File folder

A COMPARISON TABLE FOR SIMILIAR PRODUCTS

DriveClone is the best cloning tool in the market. Below are the comparison between DriveClone, Paragon, and Acronis cloning solutions (Table 1).

SUMMARY

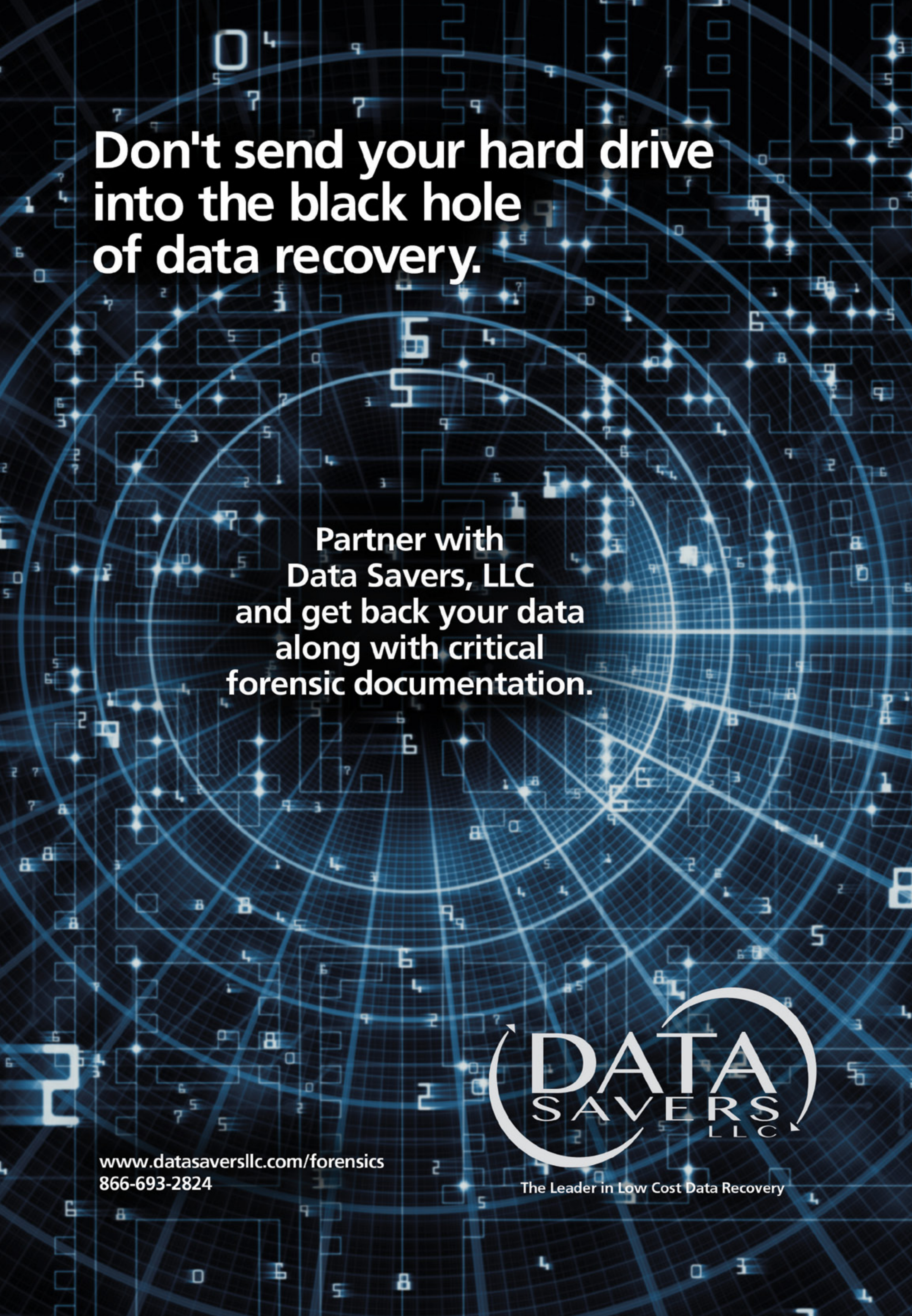
DriveClone provides two RTO standard solutions; Mirror Drive and Convert to Virtual Machine (VMWare and Hyper-V format). With Mirror Drive technology, you can replace a failed device with a Mirror Drive. On the other hand, you can also convert or load your data directly to a Virtual Machine format. Both solutions have the lowest RTO, thus allowing business to operate with no down time when disaster struck and also allowing migration to new hardware simple and easy.

Author's Bio

The author Dr. Wei Deng has been working in Backup & Storage field for the past 7 years, and also involved in developing database and cloud-based applications, intended primarily for improving data protection for customers and enterprises.

Table 1. *Comparison between DriveClone, Paragon, and Acronis cloning solutions*

Function		FarStone	Paragon	Acronis
		DriveClone 9	Drive Backup 11 Server	Disk Director 11 Home
Clone Mode	Automatic		Y	Y
	Manual		Y	Y
	Incremental clone	Y	Y	
	Remove free blocks between partitions		Y	
	Perfect Defrag Cloning	Y		
	Smart Cloning	Y		
	Dissimilar/Universal Boot Cloning	Y		
	Support UEFI System	Y		
	Multi-version file recovery	Y		
	Convert to VM file	Y		
	Mount VHD hard drive	Y		
	Map network drive	Y		
	Exclude useless Windows files	Y		



Don't send your hard drive into the black hole of data recovery.

Partner with
Data Savers, LLC
and get back your data
along with critical
forensic documentation.

DATA SAVERS LLC

The Leader in Low Cost Data Recovery

www.datasaversllc.com/forensics
866-693-2824

HARD DISK DRIVE DESIGN AND RECOVERY, DEMYSTIFIED

by John EDRL

Many in the Data Recovery industry know that HDD design is obviously not perfect, and can definitely use some improvement. Some Hard Drive brands and models are better than others. One flaw is the fact that most of the System Area (the Operating Instructions) for the Hard Drive reside on the Platters, where they are referenced. The Hard Drive needs this information in order to operate properly.

What you will learn:

- Intermediate – Advanced level Hard drive technology
- Bits and Bytes
- Some Hard Drive Design flaws
- Secret Data Recovery Techniques
- Head Stack Assembly (HSA) Design.
- Secrets to HSA Alignment for Recovery.
- Data Recovery Bible rules.

What you should know:

- Basic H/W Data Recovery
- Your limits

HDDs typically have two copies of these operating instructions but this isn't realistically helpful in preventing all failures or aiding all recoveries. When the Hard Drive powers up it goes to a designated location which it expects to find the required instructions and configuration information. If it successfully reads it, it comes to READY state and waits for instruction from the Computer.

[PERQUISITE]

The design of a Hard Disk Drive (HDD) incorporate some of the concepts and similar technology used in Record Players, DVD Drive, and a Cassette Tape Deck, integrated into one.

It works by reading and writing Data Bits to the internal Rewritable Platters, using Electro-Magnetic Heads at the Arm's Tip, instead of a Needle, like a Record Player.

Today's Hard Drive (Platter) *Tracks* are written at the Factory by a Low-Level Formating Servo-Writer.

These Tracks (which are used by the *Heads*) are not grooves like a Vinyl Record, but instead, circular Magnetic Tracks (*Technically: "Servo-Tracks & Sectors"*) that are read/followed by the Heads using a Closed-Loop Servo Circuit (Figure 1).

Without getting too technical, the Closed-Loop System constantly synchronizes and calibrates itself as it reads these special Servo Patterns on the Platters. These *(virtual-groove)* Tracks help position the Actuator Arm as it reads/writes data within them.

PLEASE NOTE THESE ACRONYMS

- HDD = Hard Disk Drive
- HSA = Head Stack Assembly

- PZT = Lead Zirconate Titanate (a.k.a: PieZoelectric Tranducer)
- LDV = Laser Doppler Vibrometer
- HW = Hardware
- SW = Software
- WD = Western Digital

NOW, FOR THE REAL-WORLD ISSUES

HDDs are designed and manufactured to be affordable to a degree. Predominately, efficiency comes first while reliability is last. Simply look up the definition of RAID. They are not built to be mobile even though they have been implemented into Laptops for years (without proper warning). HDDs want ideal conditions *(within its temperature range, while the Hard Drive is not moved about)* in order to operate as designed and marketed. Would you plug headphones into a modern Turntable and walk around the City while playing (extremely valuable and rare) Vinyl Records on it? Well that's what HDDs go through, constantly trying to dismount upon sensing movement (if that protection scheme even works). Other than the Heads Slapping/Scratching the Platters within the propagation delay of Shock-Detection to the Parking of Heads.

PHYSICS

The spinning platter basically create a Gyroscope effect. Where the Motor Shaft is the Spin Axis and the Platters are the Rotors. As we all know a Gyro wants to stay in its set axis. Any opposing force against it will apply stress on the Motor and its Spindle.

THE CANTILEVERS

1. The Motor's Shaft to the Platter's edge.
2. The Head Stack Assembly *(HSA)* Mount-Shaft to the Head's Tip.

These Cantilevers have their needs and limits. Basically shock and G-Forces will very likely cause them to flex, and in turn, fail. So in ideal conditions you: *1.)* Manufacture the Hard Drive, *2.)* Deliver it carefully to the hands of the Installer, and *3.)* have Him install it into a still Computer, while it operates with minimal *(if not any)* vibration.

Realistically, issues/failure are inevitable *(at some point in the future)* when you start moving the HDD around *(e.g. Laptops and external Hard Drives)*, especially when powered on. Hence the design of the (SSD) Solid-State Drive. But even Solid-State Drives have their flaws.

WESTERN DIGITAL HSA-ALIGNMENT FOR RECOVERY, DEMYSTIFIED

Even though Hard Drive Heads float on an air-bearing, and most modern HDDs use a PZT Actuated Suspension scheme *(Figure 4)*, they still rely on the arm being perfectly level *(to an extent)* at all positions of the Platter.

You would assume that if you assemble the Hard Drive back the right way and tighten everything thoroughly, that the Heads should be aligned since the Base, HSA Bearing-Seat and Platters are *(in theory)* true from the factory...

But typically, this is not the case. The Top-Lid of Hard Drives are bendable. Screw holes are not precise, nor is the HSA's Mounting shaft.

Some HDD Bodies *(Base Casting)* do have Top-Lid Guide-Pins *(Figure 2)* that help fix the Lid's position upon assembly, but obviously this does not

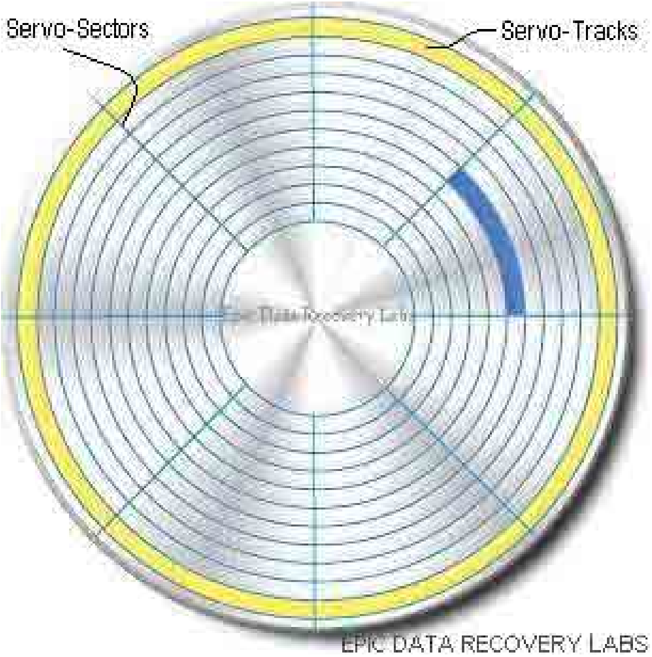


Figure 1. *HDD Servo Tracks / Sectors*

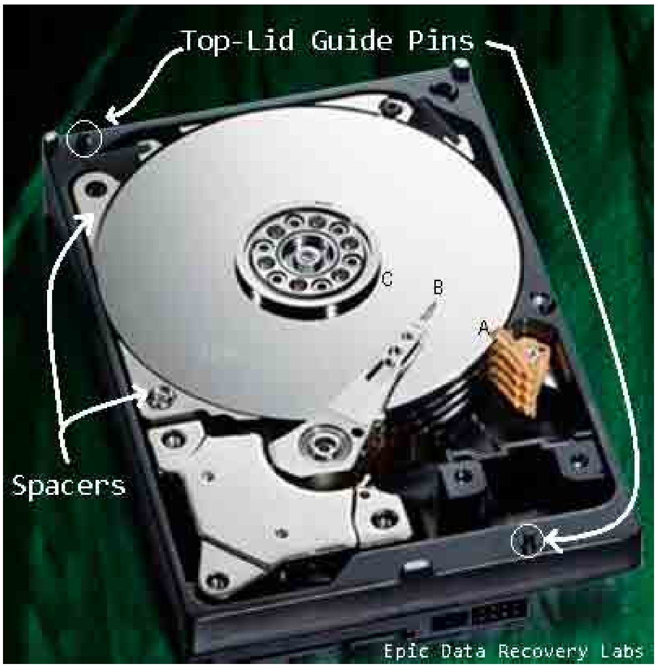


Figure 2. *Western Digital HDD inside view*

fix the WD HSA alignment issues upon re-assembly. *On a side note, the Platter Spacers (Figure 2) add another level of complexity to recoveries. We can delve into that another time.*

WD HDD DESIGN ISSUES

1. Aluminum Base and HSA Mounting shaft.
2. WD HSA Mounting shafts are Cone-shaped.

This design helps fast Production / Assembly and in turn keeps the Cost down, but is prone to failure and misalignment. Especially after the Hard Drive takes a fall or is opened for inspection / repair.

In theory, if the HSA Bearing-Seat *(Figure 3)* is truly parallel to the Platter's surface, a rigid Top-Lid would secure the HSA's Bearing to the *(Bearing-Seat of the)* Base. Ensuring alignment *(to a degree)* between the floating Heads and the Platters.

Realistically this isn't always the case with current WD HDD designs.

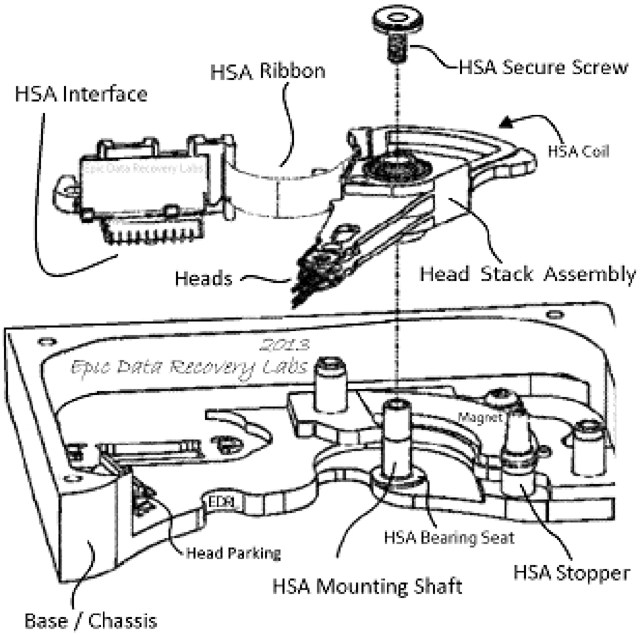


Figure 3. *Exploded view of the Head Stack Assembly*

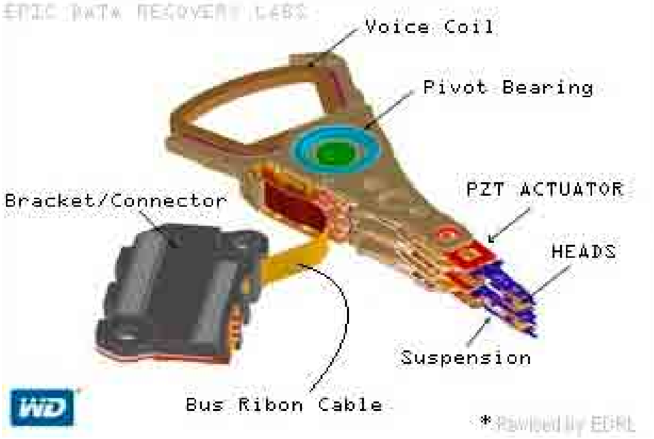


Figure 4. *Head Stack Assembly (HSA)*

TO RECOVER FROM SUCH HARD DISK DRIVES

You need to fully understand and respect the need for secure and accurate alignment of the two cantilevers.

There are tools that have surfaced in the market which help with this requirement but they are not perfect.

They work most of the time and when they don't, some people may assume that the issue is *(therefore)* something else, when in fact alignment is still the issue.

THE ISSUE WITH CURRENT HSA ALIGNMENT TOOLS

HSA Aligning on-the-fly has a high probability of damaging both the Platters and Heads, as the Technician tries to *(somewhat-blindly)* calibrate the Aligner. If the Platters suffer enough damage, standard recovery will not be possible even with new Heads. *e.g.: If you were a Pilot, would you put on a blindfold at the Gate and just use your other senses to fly an Airplane?*

HARD DRIVE ALIGNMENT CHECKLIST

1. You need a real Clean Room. Better yet, one with a pristinely clean *(Class 100/10)* Glove-Box.
2. Solid, Clean and Level surface.
3. A low-power Laser on a static mount *(LDV)*, which is used to check for Platter wobble or vibration.
4. Check the alignment of the HSA at 3 points *(A, B, and C)* before starting it *(Figure 2)*.

Ideally you need the means to float the Heads without having them touch the Platter's surface as they travel to both ends of the Platter *(that is not spinning since it is off)*.

There are tools that allow you to do this in the industry but they are not applicable to all situations. There is no need to go into details about this, since capable Technicians *(who are experienced in Data Recovery)* can figure this out. If you don't know how, or understand and accept the importance of this, it's a very good indicator that you shouldn't even attempt such a procedure.

SOME HINTS

Extremely smooth and sterile: Thin-Strips that are thick enough can be used as rails to allow the Heads to float over the Platter without causing (Drag, Scratches, Twisting, Harsh-Bending, etc.) damage to the Arms or Platters. This method is sometimes better than using expensive Head Holders that attach to the HSA, (especially the removal of stuck heads) because presently HSA Head-Floating Tools in the industry rely on, and as-

sumes that the Platter's Axis is True, when sometimes it isn't (especially after a fall). Using such a tool on a Platter that is Off-Kilter will very likely result in scratching the Platter's surface.

Once we are confident that the Platters are true, we need to check that the HSA Bearing is secured down. Then that the top of the HSA Bearing and arm is level.

TIP

Having a long rigid-rod attached to the HSA mounting hole can aid in fine positioning the HSA's Axis and in turn the levelness of the Head's on the Platters (Figure 5).

WARNING 1

Do remember that the HSA Mounting Shaft can snap-off if you pass its Threshold.

Having a custom-made *(Sterile)* see-through *glass (or Plexiglass™)* Top-Lid Cover will allow you to monitor the HDD's operation while calibrating and recovering. You need to understand that the Top-Lid is a critical part, since the Heads need the right amount of air pressure *(and steady air-flow)* to properly float over the Platters. Other than the Lid protecting the HDD from outside elements.

After all these checks… if you are certain that the Hard Drive's cleanliness has not been compromised, you can turn the Hard Drive ON for fine adjustments, until it comes to *Ready*.

Calibrate the HSA in a VERY QUIET SETTING, so that you're able to hear abnormal sounds, like scratching and scraping sounds.

One interesting thing to listen for, is the Head's PZT Actuators faint scream for help. A high-pitched *(roughly 16Khz)* squealing sound that it may produce, as it tries to compensate and adjust the Head fly-height, on-the-fly. Don't mistake it as the Heads Scratching the Platters. Basically, if the Heads and Platters are level you shouldn't hear any squeaks or abnormal sounds.

CRITICAL

With all that in mind and sorted, you need to have the means to communicate directly with the Hard Drive at the lowest level.

WARNING 2

Using a program in Windows, MAC, or Linux *(with default settings)* is not recommended since the Operating System *(other than the BIOS)* will try to communicate with the failing HDD *(while assuming it is operating normally)*. Low-Level *(yet limited)* Free Cloning/Imaging programs are MHDD or the Linux program "DDRescue" *(with Auto-Mount realistically disabled).* Note that these programs were designed to work with Hard Drives of several years back. I'm not too sure how they will behave with HDDs larger than 2TB. I fiddled with these programs a few years ago because they *were (and sometimes still are)* applicable and helpful. Especially with Scripts + Add-ons. *For the sake of not making this a Commercial, I will withhold what Professional Tools We currently (primarily) use.*

WARNING 3

There are several "Tools" out there that claim to provide low-level access but they have a lot of limitations and flaws that you should be aware of. Software based tools rely on the HW *(BIOS, Chip-sets, and Controllers)* and the Base Operating System it is using to communicate with the HDD. One issue that turns up, is something I call "Device Comm. Delusion" ***(D.C.D.)***, where the Software or Hardware you are using thinks it's still Cloning or Imaging Sectors off the Patient Hard Drive, when in fact it's getting nothing. The SW/HW is basically writing nothing but (HEX:) "FF" or "00" to the target Hard Drive or Image File.

For example, one way to produce this D.C.D. issue (at your own risk of damaging HW / Data), is get two working HDDs, that we will label as *HDD-1 Source*" obviously as the Source Hard Drive *(containing the sought-after data)*; and "*HDD-2 Target*" *(which has nothing on it but Hex-Zeros ("00 00 …")*, where we will try to Clone "HDD-1 Source" to.

All you need to do, is start the Cloning process and interrupt the process by *(either briefly or permanently)* removing HDD-1's SATA Cable *(or its Power Cable {which I DO NOT recommended, because you may cause a Short or Surge })* while in the process of Cloning it.

Ninety percent of the time, the Data Recovery (DR) Software will keep going… as if it thinks it's

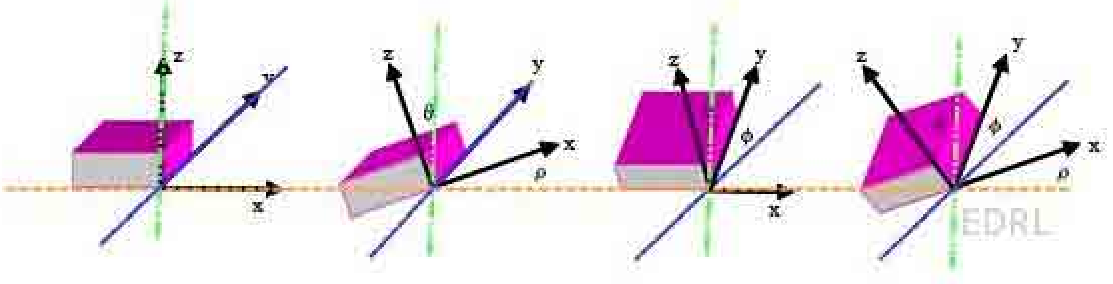


Figure 5. *Level read/write Heads are very important. You must ensure they are not off-kilter.JPG*

receiving Data from "*HDD-1 Source*"; or the DR-SW simply decides to ignore the fact that there is no transfer of data coming from "*HDD-1 Source*" and continues to write Zeros to the "*HDD-2 Target*".

Depending on what SW/HW you are using, most software will partially crash, where it still seems active and continues to write "00 00…", "FF FF…", or "40 40 …" *(pick your flavor) to the "HDD-2 Target".*

One way to determine this illusion is by using common sense. If you know the Patient Hard Drive is not expected to operate normally and some sectors should be bad; yet the Cloning/Imaging process is going too smoothly *(e.g.: No slow reads or slight pauses in the cloning process)*, then you should be concerned and stop the process for confirmation. Otherwise you are wasting precious/limited time.

Before Stopping the Cloning/Imaging process: Note which Sector you are about to stop at, so that (if it doesn't display the last Sector Cloned) you can set the start *(cloning from)* sector manually and continue where you basically left off, later *(if need be)*. Don't assume the Software or HW will run on *Auto-pilot* and do it for you *(even if it claims to)*.

So when you start the Clone process and run it for 5 minutes… write down the Last Sector cloned. Open up the Target with a Hex Editor and Jump-to the last sector Cloned, and roughly-check to see if any data has been written before that Sector. If it's good, continue and keep checking the Target *(as instructed above)* in increments if you suspect something is not going right.

WARNING 4

When working on a real Data Recovery case, stopping/interrupting the Cloning / Image Process is NOT recommended with HW failing HDDs *(e.g.: Bad Sectors; Scratched Platters; Corrupt, Deteriorating, Unreliable, or Unpredictable: Firmware / System Area Modules).* Especially if you reset the power to the failing HDD-1.

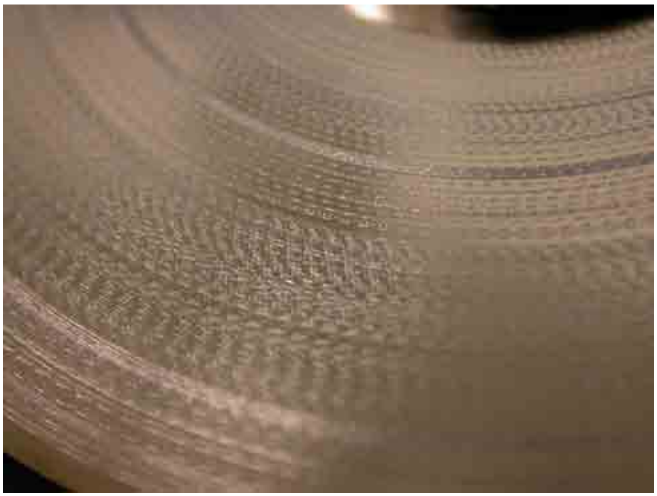


Figure 6. *Here is a picture of a completely scratched and unrecoverable Hard Disk Drive Platter*

It may never come to READY again. If it's not self-destructing and it seems to be working enough to give you the data, leave it ON and grab (Clone) as much as you can.

WARNING 5

Some consumer-grade S/W will delete the image file if you stop the Image process prematurely.

Because of that, Imaging is not the best idea for failing Hard Drives.

TIP

Monitor the Cloning/Imaging Process: Heads can fail at any point and if you leave it Cloning/Imaging overnight without some sort of watch *(or Auto-Power off scheme ;)* you will return to a permanently TRASHED *(unrecoverable)* Platter *(Figure 6)*.

IN CLOSING

I was going to entertain the Hacker-Mind with some DIY Alignment-Jig Hack to be used only for educational Purposes (not for real Data Recovery Cases). But since this is a serious publication for Professionals, I would rather keep it as such.

The last thing I want is for someone who is incapable to experiment with someone else's *(or even their own)* Hard Disk Drive which they wish to recover.

TAKE THIS AS MY FINAL WARNING

Hard Drive Data Recovery deals with an ever-changing secretive Technology and Industry. It takes several years of study and experience to gain the right Knowledge and Skills required. Please do not try this unless you are truly a Data Recovery Specialist with a Class 100/10 Clean Room Environment and have a great deal of experience. Good Luck!

Author's Bio

The Author has well over 20 years of Study and Experience with Computers, Electronics Engineering, Servos, Motors, Stepper Motors, Controllers, Lasers, Data Storage, etc. He has been working as a Data Recovery Specialist and Technology Consultant for Epic Data Recovery Labs (in New York) for several years. Where He leads a great deal of Research and Development in Electronics, Data Recovery, and Loss Prevention. One of his goals is to improve the performance and reliability of Electro-Mechanical Equipment and Mission-Critical Devices. As an Engineer, Consultant, Designer and Project Manager, He has lead several projects for major Corporations and Mission Critical Venues. He is passionate, meticulous and tenacious about Data Recovery, Technology and his Projects. Constantly researching the latest in: Nano-Technology; Data Storage Reliability/Recovery; Electronics / Electro-Mechanical Performance and Design, while focused on good Ethics and Integrity.

Epic Data Recovery Labs

Epic Data Recovery Labs

Real Hardware-Level Data Recovery Services.

We Recover Data & Files from virtually Anything!

Storage Medium: Hard Drives, Solid-State Drive (SSD), RAID Servers, USB Flash Sticks, Flash Memory Cards, DVDs, CDs, Mini-Discs, Floppy Disks, Vinyl, etc.

Electronic Devices: Mobile/Smart Phones, Tablets, Media Players, Digital Voice Recorders, Digital Cameras, CCTV/Security Digital Video Recorders, etc.

Analog / Digital Tape: Tape Backups, Cassette Tapes, Tape Reels, VHS, etc.

Audio / Video: Recovery, Restoration, Remastering & Conversion

Computer Services: Consulting - Design - Setup/Config - Repair - Optimization - Upgrades

Security Services: Virus/Rootkit Detection, Removal & Prevention - Network/WiFi Security Backup / Restore Schemes - Loss Prevention & Property Recovery

We Deliver Real Results and Expertise!

Our Data Recovery & Electronics Engineers have been conducting successful recoveries for well over 15 years. Our constant Research & Development in Data Recovery Technology, and Tenacity helps us succeed in many in-depth recovery Cases, where other Companies typically fail.

Our wide variety of professional services have been provided to a broad range of clientele Worldwide. From Mission-Critical recoveries, Events and Projects for Government Agencies, Fortune 100 Corporations, A-List Celebrities, to Residential type clientele. Your privacy and thorough recoveries are Top-Priority for us! We deliver real results and Expertise!

We are located in New York City.

Assistance is available 24 Hours / 7 Days a week!*

T: +1 . 800 . 691 . 8351
E: Info@EpicDataRecovery.com

For more information please visit us at:* **www.EpicDataRecovery.com

DATA RECOVERY INTRODUCE AND HOW TO RESCUE DATA

by Tommy Redden

A helpful and fast method for recovering data in a file system. This method includes the steps of performing a given way that scans the deleted and lost data files on a computer, and preview the recovered data before you decide to recover it back with professional data recovery software. To prevent the loss of data files, what tips you need to follow, and other knowledge of data recovery in details.

What you will learn:

- Intermediate – Advanced level Hard drive technology
- Bits and Bytes
- Data recovery

What you should know:

- A Basic Understanding of Hard drives
- Basics of hard drive recovery
- Basic computer maintenance

What is Data recovery? Here we give a definition: data recovery is one process of recovering or salvaging pieces of data from the disk drive or any other type of storage media when data can't be accessible using customary ways. Usually, these files are stored in hard drives and removable disks, such as CDs, DVDs, tape cartridges, flash memories and the alike. It's necessary to recover data files for users, because the damaged or corrupted storage media may cause a horrible loss for data users. Also, hard disks may crash because of mechanical failures or infestation of a virus; mere scratches on CDs and DVDs can provide rise to data reading problems; and as for the tapes, these can be broken easily. These are a few reasons why we need to do data recovery to rescue data files.

However, data file recovery is not limited to fix storage related failures, accidental data deleted or formatted hard disk, external devices and even mobile phones can also need file recovery.

DATA RECOVERY TECHNIQUES

For end users, once your computer hard disk is damaged or system has crashed without saving or backing up important data files, you may think that the data files were lost permanently. However, once you have had this kind of terrible experience, don't panic, just remember that, do not do anything on this computer, and do not overwrite any other files on the same hard disk, or you will possible lose your files forever.

Data recovery firms make use of various methods to salvage data to restore very important files or to make the electronic device function properly. One of these methods is using clean room facilities or controlled environment to minimize damage or corruption that may be brought by the ambient air. The use of these special rooms aims to protect the storage media while recovery is being made. But how do you self-rescue this lost data? Let's cover the fast method.

CASES OF HARD DRIVE DAMAGE

PHYSICAL DAMAGE

Data recovery can be achieved by leveraging different techniques. Data recovery from physical damaged hard drive can be done by replacing parts in the disk, but there may still be logical components damaged. A specialized disk-imaging procedure is used to recover every readable bit from the surface. If this image is acquired and saved on a reliable medium, the image can be safely analyzed for logical damage and will possibly allow much of the original file system to be reconstructed.

Generally, a number of physical damage cannot be repaired by end users, you need to take your hard disk to data recovery companies to get service and salvage important data perfectly.

Physical damage always causes at least some data loss, and in many cases the logical structures of the file system are damaged as well. Any logical damage must be dealt with before files can be salvaged from the failed media.

LOGICAL DAMAGE

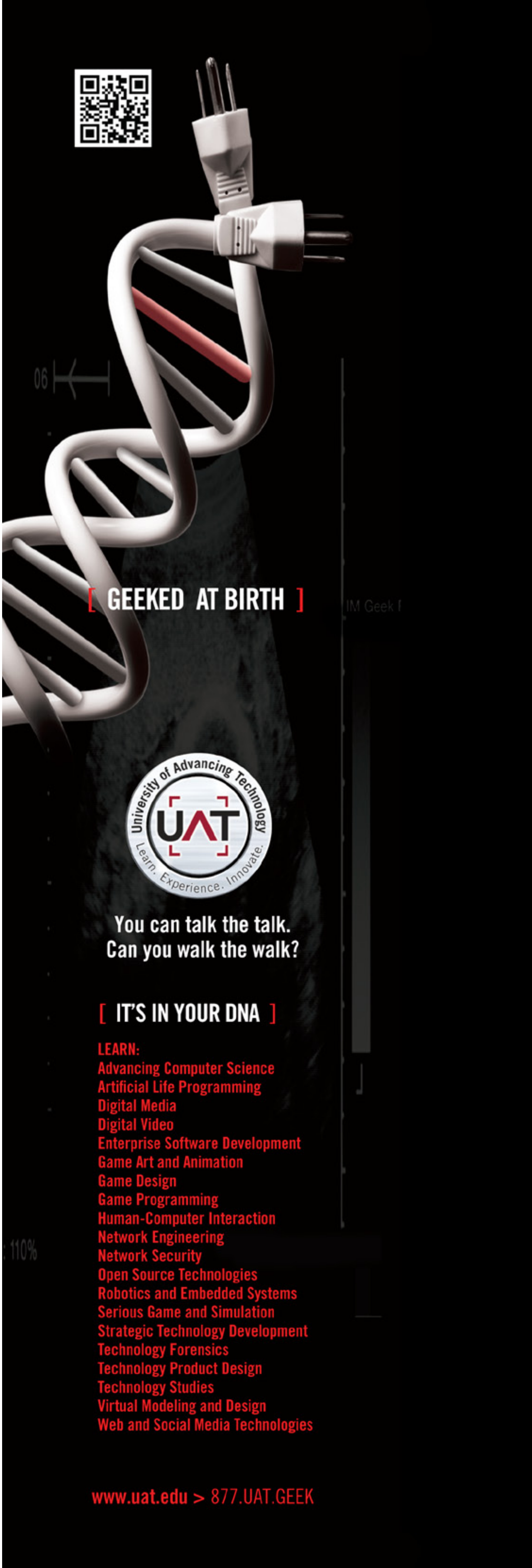
The data lost is not caused by a hardware failure, and it requires professional data recovery software or a software-level solution.

When some data is deleted by mistake or a hard disk is inaccessible, end users could effortlessly and unconsciously create further data loss instead: if the hard disk is overwritten, data files may not be recovered and, hence, lost forever. When data has been physically overwritten on a hard disk drive it is generally assumed that original content is no longer possible to recover.

METHOD FOR DATA RECOVERY

A method for recovering data objects stored in a data bucket in a computer system comprised of servers interconnected by a network, wherein each server includes a storage area, wherein data buckets are included in a plurality of the storage areas throughout the network. If your data files have not been overwritten, you can use professional data recovery software to rescue the lost data. Otherwise, you would be better to employ a computer service center staff to do physical data recovery service.

A data recovery system of a distributed transaction processing system with a two-phase commit scheme, in which data processing systems (server)



[GEEKED AT BIRTH]

University of Advancing Technology
UAT
Learn. Experience. Innovate.

You can talk the talk.
Can you walk the walk?

[IT'S IN YOUR DNA]

LEARN:
Advancing Computer Science
Artificial Life Programming
Digital Media
Digital Video
Enterprise Software Development
Game Art and Animation
Game Design
Game Programming
Human-Computer Interaction
Network Engineering
Network Security
Open Source Technologies
Robotics and Embedded Systems
Serious Game and Simulation
Strategic Technology Development
Technology Forensics
Technology Product Design
Technology Studies
Virtual Modeling and Design
Web and Social Media Technologies

www.uat.edu > 877.UAT.GEEK

systems are connected to a data processing system (client) through a communication line, and each of said server systems performs PHASE I processing and PHASE II processing in response to transaction processing requests from said client system, comprising:

- means for inquiring of said client system the transaction processing status data output from said two-phase processing means and stored in said storage means when an operation is resumed after a system failure of said server system; and
- means for performing data recovery processing when the status data indicates that the transaction completion processing of said server has not been completed.

A file allocation table (FAT) and associated directories are used for the convenience of users of a computer system, such as a communications switching system, that employs a file management system to manage large-capacity files. When a fatal error occurs in the course of processing files, the computer system must be restarted via software or hardware operation (by pressing a power-on switch twice or pressing a reset button, for example). A restart operation that occurs when the FAT or a directory is being changed may lead to an anomalous situation: the data processed before the restart operation remains available, but consistency is not assured for the data related to the FAT or the directory being changed.

When the computer system is restarted, the file system may fall out of consistency with respect to the storage state of management information maintained for controlling the overall file system. This inconsistency may result in the loss of file data and, most undesirably, in an interruption of services of the computer system. Such an interruption may amount to only an inconvenience in some situations involving only general personal computers. However, it presents a major problem for mission-critical systems and particularly for switching systems that are required to provide continuous service with high reliability. System reliability is substantially reduced if service may be suspended after a restart operation due to the loss of essential programs and data required for system operation.

A device for data recovery, comprising:

- flash memory coupled to a computer system and including an area for storing a control structure used by a file system of said computer system;
- nonvolatile memory coupled to said computer system for storing recovery data, with said recovery data including data contained in a recovery step flag and with said nonvolatile memory including a predetermined area for storage of said recovery step flag; and
- a processing unit coupled to said flash memory and to said nonvolatile memory and selectively storing in said predetermined area of said nonvolatile memory a mark indicating a position of said recovery step flag corresponding to a specified step of a file management task being executed by said file system, with a corresponding said mark representing completion of a corresponding said specified step of said file management task by said file system.

A method for recovering data, comprising the steps of:

- performing a specified step of a file management task for a file system of a computer system, with said file management task effecting a change to a control structure of said file system and being defined by a predetermined procedure including said specified step;
- storing in a predetermined area of a nonvolatile memory a mark indicating a position of a recovery step flag stored in said predetermined area, with said position corresponding to a corresponding said specified step, with a corresponding said mark representing completion of a corresponding said specified step, and with said nonvolatile memory being coupled to said computer system; and
- re-entering said predetermined procedure at a step subsequent to a last completed said specified step indicated by a corresponding said mark and completing said file management task to effect said change to said control structure when a restart event interrupts said file management task after a corresponding said mark has been stored.

Take 3rd part data recovery software to rescue lost data files include three steps: Install data recovery software -> Scan lost files automatically -> Recover the scanned data files.

Some of the software providers could separate one main function into several pieces, like recover data from hard disk, recover data from mobile, recover photo, recovery video, recovery songs, and even deeply recovery, raw recovery and so forth.

TIPS

Whatever data files you want to recovery, don’t do any operation on the hard disk or other devices, when the files are deleted by mistake or your hard disk was inaccessible, get professional data recovery software to rescue your important files as soon as possible. If you overwrite new data on the same device, then you might lost your data permanently.

DYNAMIC DATA RECOVERY

What is Dynamic Disk? The Dynamic Disk is a physical disk that manages its volumes through the use of LDM database. And what is the LDM database? LDM is short for of Logical Disk Manager, and it can be a hidden database which dimension is 1MB after the Dynamic Disk. The 1MB space records every piece of information on the volumes on one disk, as well as holds some related info on each dynamic disk. Such as Drive Letter, Volume Label, the begin sector of Volume, Volume size, the file system of Volume, along with the current dynamic disk is which and so on.

A computer storage system, includes two contents:

- a plurality of disk drives for storing parity groups, each parity group comprising storage blocks, said storage blocks comprising a number of data blocks as well as a parity block connected with said number of data blocks, every one of said storage blocks stored on the separate disk drive so that no two storage blocks from a given parity group reside about the same disk drive;
- and a recovery module to dynamically recover data lost when at least a portion of one disk drive in said plurality of disk drives becomes unavailable, said recovery module configured to produce a reconstructed block by using information in the remaining storage blocks of a parity group corresponding to an unavailable storage block, said recovery module further configured to split said parity group corresponding to an unavailable storage block into two parity groups if said parity group corresponding to said unavailable storage block spanned all of the drives in said plurality of disk drives.

REMOTE DISASTER DATA RECOVERY METHOD

Disaster recovery generally refers to a plan or strategy for duplicating computer operations, for instance, of a company, wherein copies of a volume or volumes of computer data and/or software of a primary location are established at a remote location thereby providing a redundant measure of protection in the event of a disruption of operations at the primary location. Disaster recovery thereby allows a company to resume operations in the remote location within days as opposed to, in certain instance, a permanent loss in certain aspects of the company's information infrastructure.

Disaster recovery systems appearing in the art provide companies with the ability to create remote backup copies of a volume or volumes of data and/or software. The information necessary to create the backup copies at the remote location is typically communicated to a remote server connected to a client computer over a communications network. Data recovery similarly entails receiving data over the communications network. Systems providing disaster recovery in this fashion, however, have numerous shortcomings with respect to creating backup copies of a volume or volumes having relatively large quantities of data and/or software. For instance, a large data transfer may increase network traffic and thereby consume a large portion of the network's capacity sufficient to slow the company's operations during the transfer. For example, creating a remote backup copy for a server computer having 100 gigabytes of data stored thereon over a company's network with multiple TI data transfer capability will tie up the company's network for months. This is particularly problematic for companies operating around the clock that may not otherwise limit data transfer to off-peak hours and companies having networks with limited bandwidth. There is therefore a need for remote disaster recovery systems and methods having a reduced impact with regard to network traffic over a company's network.

A disaster recovery volume is generally created at a local archival storage unit including therein at least one storage medium constituting the disaster recovery volume. The medium constituting the disaster recovery volume is associated with the primary volume thereby allowing the storage medium constituting the disaster recovery volume to be relocated to a remote archival storage unit at a remote location without compromising the association between the primary volume and the disaster recovery volume.

A disaster recovery computer system comprising at least one computer having programming associated therewith, the at least one computer communicatively connected to at least one local archival storage unit and at least one remote archival storage unit, wherein the computer programming when executed provides data transfer and control capability to create at the local archival storage unit a disaster recovery volume of a primary volume on at least one storage medium, which storage medium constitutes the disaster recovery volume, the computer programming associates the storage medium constituting the disaster recovery volume with the primary volume thereby allowing the storage medium to be relocated to a remote location without compromising the association between the primary volume and the disaster recovery volume.

The method of creating a disaster recovery volume may include the steps of identifying incremental changes to the primary volume, packaging data representing incremental changes to the primary volume, and communicating the packaged data over a communications network to the remote storage unit at a remote location. The step of packaging data representing incremental changes to the primary volume may include compressing the data representing incremental changes to the primary volume. The incremental changes to the primary volume may then be incorporated into the disaster recovery volume relocated to the remote location. The incremental changes to the primary volume may be identified in connection with at least one snapshot image of the primary volume. The disaster recovery volume of a primary volume may be made from at least one copy selected from the group consisting of a backup volume of the primary volume, a quick recovery volume of the primary volume, and a snapshot image of the primary volume.

DATA PROTECTION AND RECOVERY

A protected data file currently in use is duplicated as an authentic backup file, while changing the current file's data appearance and separating the location of the authentic backup file from the original and current file, to camouflage its identity from an unauthorized intruder intending to modify or destroy the original file. A series of indicate is generated and stored in a recovery address group or file. The indicate represent the original current file and is used to reconstruct the authentic backup file and to write a restored file into the current protected data file. The recovery process may be initiated on a schedule or whenever the original current file is accessed or whenever an unauthorized use of the current file is detected and a comparison of the authentic backup file indicates the original current file has been modified. In this way, the authentic data saved from the original current file may be used to restore the protected data file as originally written and saved in the authentic backup.

In a data processor, a system for making an authentic backup file from an authorized protected data file, with the data in said authentic backup file translated from said authorized protected data file to camouflage the source or identity of said authentic backup file or its relationship with said authorized protected data file and with indicia produced by said translation representing said translation, stored in a recovery address group for comparison with a test identifier produced from the current protected data file to determine if the current protected data file is the same or different from the authorized protected data file and for access and use of said recovery address group for translation of said authentic backup file to said authorized data file and restoration of said authorized protected file, comprising:

- for translation of an authorized protected data file to an authentic backup file, camouflaged to hide its relationship to said authorized protected data file, and for storing said authentic backup file;
- for producing an identifier from aid authorized protected data file and for storing said identifier;
- for producing a test identifier from a current protected data file for comparison with said identifier for determining if said current protected file is the same or is different from said authorized protected data file; and
- responsive to said comparison for translating said authentic backup file to said authorized protected data file for restoring said authorized protected file.

Data processing protection systems have been relying on encryption, personalization such as by passwords, or by scattering of the data through a data store randomly or by strict or intelligent algorithm, the intruder, once having reached all or part of the protected data, may have modified or destroyed the data without leaving an indication of the modification or the original and authorized authentic data. While data security systems or methods may detect the intrusion and determine whether the data modification or destruction was authorized, there is no method or system for safeguarding the authentic data or for verifying the data appearing in a protected file after an unauthorized intrusion is the same as the authorized data, or that an unauthorized modification has been made, or for recovery of the authentic data through an authentic backup file, or for camouflaging an authentic backup data file to hide it from access and destruction, using techniques to hide the data identity such as size change, content masking using encryption, name or location change or for using these data camouflaging techniques to reassemble the original authentic data to automatically recover the data after an intrusion.

Author's Bio

Tommy Redden, Product manager for 3 years at uMacsoft Studio. He is also freelance writer on any topic of Mac products and applications. Get more tutorials on data recovery at http://www.recover-data-mac.com, http://www.umacsoft.com

Dr.WEB®
20 years

Dr.Web SpIDer is 8-legged!

New Package Installer

Improved Technology for Protection against yet-unknown Threats

New Anti-Rootkit Module

Online Service Dr.Web Cloud

Improved Parental Control

Free Upgrade to New Version

Unified User Interface

Better than ever: Dr.Web Security Space version 8.0! 30-day trial available from https://download.drweb.com/?lng=en

New Version 8.0

Security Space and Dr.Web Antivirus for Windows

Get your free 60-day license under https://www.drweb.com/press/
to protect your PC and your smartphone with Dr.Web!
Your promo code: ***Hakin9***

Protect your mobile device free of charge!
https://support.drweb.com/free_mobile/

Doctor Web is a Russian anti-virus vendor with a software development record dating back to 1992.
www.drweb.com

© Doctor Web
2003 — 2013

IF YOU ARE STUPID AND YOU KNOW IT...: RAID 5 VMFS RECOVERY

by iCube Development

Basic hard drive data recovery is difficult, but a damaged RAID 5 Array containing VMWare ESX Virtual images, require an expertly trained technician and Lab to properly perform a recovery without causing further damage to a client's data. We will review the process to perform a recovery as well as steps to help prevent such a recovery in the future.

What you will learn:

- Basic Electrical Testing on a Hard Drive
- Intermediate – Advanced Recovery Steps for a RAID 5 Array
- Industry Standard Safe Guards for Data

What you should know:

- Basic to Intermediate level understanding of RAID
- Basic understanding of Virtual Environments
- Bits and Bytes
- Types of Hard Drives
- Basic Understanding of Linux commands

Our lab consistently encounters unqualified system administrators, and self-proclaimed experts which unknowingly complicate a data recovery case. Inaccurate Internet resources combined with simple "opinion" is a very dangerous combination. If you don't know what's wrong, or why it's not working don't try to recover the data yourself!

Let's be honest, I'm not about to fix my 2013 Ford F-350. I'm not a mechanic, I don't have the specialized tools, and I don't have any way of undoing damage I may cause in the repair process. The same can be said for data recovery. Short and simple, the *success of every data recovery case* is determined by the amount of variables. Keeping the amount of variables minimal always gives you the greatest chance of recovery. Adding more variables not only decreases the likelihood of a successful recovery, it also influences recovery cost.

The problem is getting worse; with the global adoption of virtualization and popularity of closed-source file systems, storage devices are ever-growing and are powering more than just simple file shares – they're powering an entire business.

Recently, we received a call from a client who had an HP Server operating in an ESXi Virtual Server Environment. Client data was stored on a RAID 5 array (VMFS formatted). The client stated the RAID array had failed, and despite his best attempts he was unable to recover the data or get the server to boot back into ESXi. The following day the client brought in the HP server for data recovery evaluation.

Upon initial review, we discovered the operating system and datastore were powered by single RAID 5 array configured with x4 300GB 320UW SCSI disks connected to an HP SmartStore controller. The client stated that during his diagnostics he noted an array member had failed and another member triggered a predictive failure warning. Upon rebooting the server the RAID POST GUI indicated the volume was offline and was not able to be rebuilt. To complicate matters, in his panic the client physical removed the array members from the server's backplane and not recorded or replaced the drives back in the correct order.

Clearly, this was not going to be a simple recovery.

THE DATA RECOVERY EVALUATION

After the client approved our RAID diagnostic fee we numbered and serialized each drive. The purpose of the serialization is later used to determine the drive member order by use of either the RAID controller BIOS values, or (if exists) the configuration and file system data.

Upon completion of drive serialization, we then use an ordinary desktop computer with a PCI slot to accommodate our next stage of diagnostics – *drive health and failure determination.* An ordinary desktop computer is setup with an eSATA PCI card (SiI 3231 chipset) and a DELL PCI-X PERC 5 card. What some people don't know is a DELL PCI-X PERC 5 card will work on an everyday 32BIT PCI bus. The PERC 3/4/5 series will also not write damaging RAID configuration values to a disk if configured as RAID 0. This ensures we don't tamper with any data located on the source SCSI hard drives.

This cost efficient setup allows our lab technicians to utilize non-server grade hardware to recognize and diagnose a wide variety of SCSI based devices. Our eSATA PCI card selection is chosen due to its wide variety of OS support, specifically with modern Linux distributions.

The first SCSI drive (1 of 4) was connected to our desktop setup, and by establishing its ability to be recognized by the PERC 5 RAID Controller we then started to determine the health and operating condition of the drive. We continued our diagnostic process by clearing the PERC 5 RAID controllers values and creating a standalone RAID 0 volume, *ensuring we do not perform any type of initialization*. The PERC 5 card does not support background initialization without the aid of software, so once the array is created it will *not modify the actual data on the media*. Once the single drive array is built, the BIOS will successfully identify the disk and allow us access volume contents (the data it held in the RAID 5 array). We can then create a 1:1 image of the array member using standard data recovery software tools. For the sake of simplicity we commonly use PartedMagic to perform most SCSI 1:1 clone operations.

DETERMINE DATA RECOVERY VARIABLES

At this point our lab takes a break to establish and confirm the best course of action. It's our policy to always stop mid-way through a complex data recovery to review the facts in effort to reduce the amount of variables present:

VARIABLE 1

We know evidence of drive order has been destroyed and was possibly overwritten by the client's recovery attempts, so we must attempt to establish the proper drive order before we can reconstruct a viable RAID 5 image. Because the HP SmartStore controller was a very early model - its lack of in-POST functions made it extremely difficult to obtain array configuration data. Further, the client powered up the server after changing the disk order, this caused the RAID controller to see the drives as both foreign and uninitialized disks. Our assessment concluded that due to the client's involvement the RAID controller no longer contained any information which would directly aid us in the recovery process.

VARIABLE 2

We don't know what the client did in his "self-recovery" attempts. We do know that a RAID 5 array without a hot spare and / or cold spare needs at least 3 out of the 4 drives to function. Given the client stated the server would not boot into the operating system we can accurately assume 2 out of the 4 drives are in an operable state. This means at least 1 drive will need to be repaired and /or recovered. However, because the client is not aware of what drive had failed initially and what drive generated the predictive failure message we can't accurately rebuild the most recent RAID 5 array without knowing the failure order (rebuilding the array from 2 working drives and the wrong 3rd drive will result in a corrupt array, or at the very least a partially constructed array). Because no determination can be made we have to perform a recovery both the failed and predicted failure array members.

VARIABLE 3

VMFS file system is a closed-source file system, all (at the time of writing this) commercially available VMFS data recovery tools utilize SSH to communicate to a working ESXi box. NTFS, EXTX and FAT32 software extraction tools will not work when dealing with VMFS. After solving Variables 1 and Variables 2 our end goal will be to obtain the working (non-corrupt) VDMK and VMX files needed to either boot or extract data from the Virtual Machine images. Possible complications may include snapshot data not yet integrated to the actual VMDK virtual disk file. Knowing this, we will be forced to use ESXi to extract or obtain client data.

After defining our data recovery variables, our initial goal is to obtain a working 1:1 image of each individual RAID member. In any circumstance (regardless of the data recovery steps taken) having a 1:1 image of each drive provides us a better chance of isolating client data and allows us to manipulate data without changing the source – something the client did not consider in his recovery attempts.

DATA RECOVERY PREPARATION

Using PartedMagic, we use dd_rescue to create an image or clone of the source drive to a more conventional SATA target hard drive. With the SATA drive connected to our eSATA card we can quickly bypass speed limitations of disk image creation over a 10/100/1000 network. Using the Sil 3132 chipset a native 2.6X Linux Kernel natively recognizes any SATA disk with AHCI support and speed.

We fire up terminal and establish the SCSI drive label:

```
#> fdisk -l (provide us a list of all physical
                        disks connected)
```

Then we use hdparm to identify the HPA value:

```
#> hdparm -N /dev/scsi_disk_id
```

For accurate RAID construction we prepared a SATA of greater size connected to our eSATA card to receive a 1:1 clone of the SCSI disk data. Now, we must modify our SATA drive to contain the exact amount of sectors as our source SCSI drive:

To do this, set the native HPA value of the SATA disk to that of the source SCSI disk:

```
#> hdparm -N p[HPAVALUE] /dev/sata_disk_id
```

Using dd_rescue we quickly grab as many error free areas of the drive, (not stressing the drive):

```
#> ddrescue -n /dev/scsi_disk _id /dev/sata_disk_id
```

WORKING WITH DUPLICATED DATA

Once the initial duplication process is complete we then determine the amount of erroneous sectors present based on the output of dd_rescue. Our next step is to obtain post-recoverable data from less unresponsive sectors, the following command will image the drive again, but this time it will instruct the computer to not skip damaged or non-responsive sectors. The idea behind the two commands is a pre-imaging and post-imaging process, whereas if the drive degrades to an inoperative state the original pre-image will still be resident and contain all recoverable sectors.

Our post-imaging command is as follows:

```
#> ddrescue -d -r1 /dev/scsi_disk _id
   /dev/sata_disk_id
```

Once the process completes we are then left with an exact copy of the SCSI disk on the SATA disk. Because we modified the HPA value on the SATA disk, the target disk will report itself as a 300GB HDD – an exact sector to sector match of our SCSI source. This is very important as a 300GB hard drive manufactured by one manufacturer will contain more or less sectors than another. A mismatched sector count will cause parity rotation faults during the assembly process resulting in a corrupted RAID 5 volume.

After attempting the above process on each array member the results were:

The first SCSI drive (1 of 4) = Successful 1:1 Duplication – No I/O Errors.

The second SCSI drive (2 of 4) = Successful 1:1 Duplication – Less than 1000 I/O Errors.

The third SCSI drive (3 of 4) = Partial 1:1 Duplication – Less than 100000 I/O Errors.

The forth SCSI drive (4 of 4) = Drive was not recognized by RAID controller, no disk activity.

Because of the previously determined *Variable 2* we cannot rebuild the array without knowing what drive actually failed and what drive generated the predictive failure message. Using the information above, we can safely assume:

- Drive 1 and 2 were seen as healthy and were active members of the current RAID 5 array.
- Drive 3 was most likely the drive which generated the predictive error message.
- Drive 4 was the drive that had initially failed.

To confirm the above findings, we looked at drive #4 in greater detail. We noted the RAID controller wouldn't power up or attempt to initialize the physical disk during POST – this suggests an electrical failure. To confirm the failure, we started performing some basic electrical tests.

BASIC ELECTRICAL TESTING

We remove the PCB from the hard drive and using a volt meter attached positive and negative probes to the PCB 3 pin motor output. By turning in the server and letting the RAID controller initialize the PCB it causes the motor to receive power and spin up the platters to read the firmware. Our volt meter read 0 volts – meaning an electrical problem was present. To combat this problem, we removed the PCB from SCSI Drive #1 (a drive we successfully obtain a 1:1 copy without any errors). Our plan is to utilize the PCB from SCSI drive #1 (a drive we knew had a working PCB) to replace the damaged PCB of drive #4. As the drives were HP white labeled we had to look at the model number and design of the PCB to determine the actual manufacturer of the hard drive.

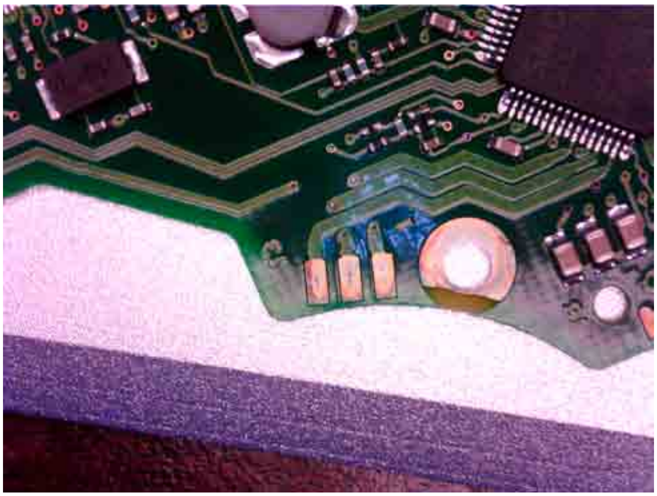


Figure 1. *3pin motor connector*

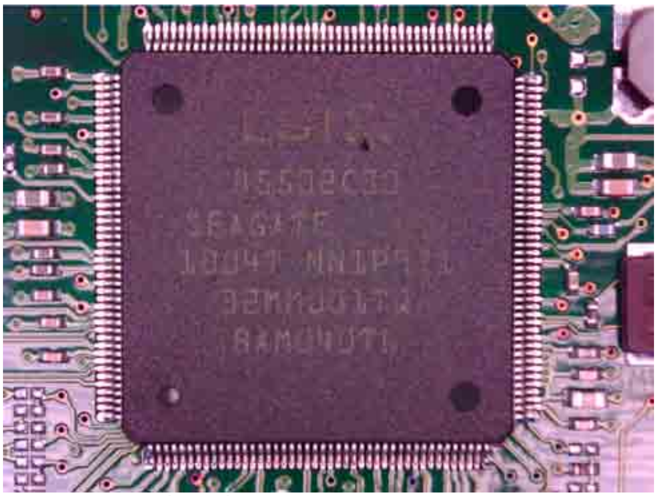


Figure 2. *mcu brandname*

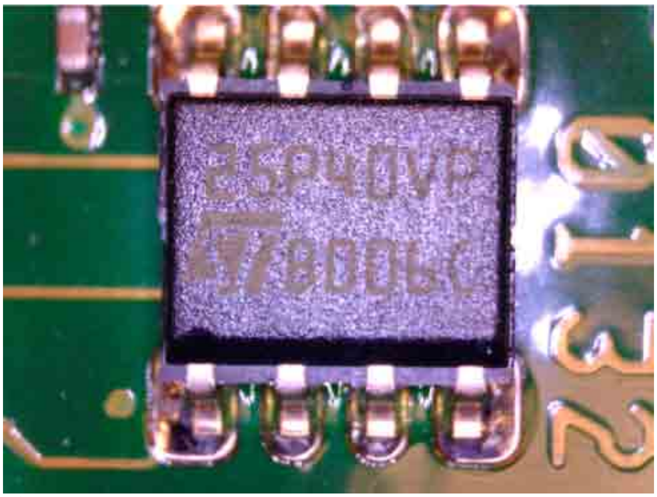


Figure 3. *ROM*

By referencing the MCU (the biggest black chip located on the PCB we could then determine the drives were manufactured by Seagate.

The rules on a PCB swap for a Seagate drive are fairly simple, its best to ensure the MCU brand and version are the same but the critical requirements is the transfer of ROM. Most Seagate hard drives utilize an 8 pin ROM module that starts with “ST25K”.

We easily identified the ROM chip on Drive #1 PCB and de-soldered it; the same process was performed on the PCB of Drive #4. Once the ROM module from Drive #4 PCB was soldered onto the PCB of Drive #1 we assembled our donor PCB to Drive #4 and retested using our voltmeter. Success, the motor was now receiving power during and was recognized during POST RAID controller LUN initialization.

We then repeated the pre and post imaging on a working Drive #4 which left us the following:

The first SCSI drive (1 of 4) = Successful 1:1 Duplication – No I/O Errors.

The second SCSI drive (2 of 4) = Successful 1:1 Duplication – Less than 1000 I/O Errors.

The third SCSI drive (3 of 4) = Partial 1:1 Duplication – Less than 100000 I/O Errors.

The forth SCSI drive (4 of 4) = Successful 1:1 Duplication – No I/O Errors.

Now that Variable #2 has been determined and proven, we can start solving Variable #1.

Variable 1 involves two processes, the drive order and stripe size. To find the stripe size, we can simply install a single healthy SCSI Disk member back into the server. During POST, we are told the RAID 5 array has failed (which is expected) and are then given the option to access the HP SmartStore RAID GUI. Once inside the GUI, the RAID controller did not provide us the drive order but did tell us the total size of the RAID volume and the stripe size (64K).

FIXING THE CLIENT’S MISTAKE

To determine the drive order, we use two different techniques:

The first technique is a suggestive tecnique – meaning drive order is suggested based on the RAID card writing RAID configuration data to the physical SCSI members. The other technique in evidential – where a given sample of data is randomly selected from array and then scanned for parity evidence calculation using a XOR operation. Combined, both discovery techniques allow us to accurately determine drive order regardless of third party intervention (specifically the client self-repair attempts).

The suggestive technique is performed by using a HEX editor and depending on the RAID controller manufacturer reviewing either the first few or last thousand sectors on the drive. Model specific, we are viewing the sectors looking for sequential ASCI values which either reference the controller model number (itself) or the drive serial number. There are other tricks to establish the drive order; the most common is to look for the NTFS partition table reference (which starts on sector 63 within the sectors allocated within the defined array). The drive which contains the start of an NTFS partition record is typically the first drive in the array.

Parity Calculation is performed by data recovery software and works by taking samples of the data from each disk then scanning these samples for calculation data (those used for either data or data redundancy calculation). Parity in itself is a simple Boolean operation operating at the binary level to create RAID parity. This operation is called the Exclusive Disjunction operation also known as Exclusive – abbreviated as XOR. Using XOR, raw binary data is processed through an operation that results in a binary result, which can then be used for redundancy and error correction.

After establishing the stripe size, and drive order we then create an assembled RAID 5 disk image. It should be noted that HP and Compaq RAID controllers have something called a *Delayed Parity* – meaning two different block sizes. An initial delay block and a typical RAID parity delay. Free and commercially available data recovery software will provide you the option to set your initial delay block size, stripe block size, and then ask you to input the drive correct drive order.

At our lab, we commonly use image files in place of physical disk images, but for this case we used a StarTech 4 Port SATA disk “toaster” to handle our physical disks.

THE DATA RECOVERY

Using a suggestive technique we used the device order of: SATA DISK3 + SATA DISK1 + FAILED (MISSING) + SATA DISK 2.

The result was an 838.19GB RAW disk image file. To ensure we have a copy of this working image file, we then make a copy of the file so we can manipulate the image in R/W mode.

It is important to note that when we created a RAID 0 volume using the PERC 5 RAID controller, both our individual and combined disk image will contain some unnecessary overhead. The HP SmartStore RAID controller natively stores information like drive health, individual disk probe values, and other RAID sensitive data in the first few hundred-thousand sectors. Because of this, the VMFS partition table will not start on the sector ESXi expects it too. So, we need to determine the VMFS partition starting sector.

Using one of our data recovery workstations we created a VMWare Workstation VM running the newest version of ESXi v4. We choose the newest version of ESXi v4.0 to ensure that we don't try to assemble the VMFS volume with an older file system interpreter (which may corrupt the assembled RAID 5 volume). During the configuration of the ESXi Virtual Machine we add the assembled RAID 5 as a physical disk. Finally, we start ESXi and enable Tech Support (SSH) under the Firewall Tab:

By use of the SSH console we first need to determine the label of the RAID 5 physical disk (this is the physical RAID 5 assembled image mounted as a RAW disk in VMWare Workstation):

```
#> fdisk -l
```

Your physical disk start with `naa.[uniqueDeviceIDString]`:

Then we run the command:

```
#> fdisk -lu /vmfs/devices/disks/[DiskString]
```

Fdisk will report (as expected) that there are no partitions on this LUN. To establish the starting sector of a VMFS partition we can instruct ESXi to search for the HEX value “0d d0 01 c0” within the first 5 million sectors of the drive.

```
#> hexdump -C -n 5000000 /vmfs/devices/disks/
   [DiskString] | grep -m 1 "0d d0 01 c0"
```

The result will be the offset within the RAID 5 image file mounted as a physical disk:

```
0x13ecd0000 0d d0 01 c0 05 00 00 00 15 00 00 00 02
16 05 00 |................|
```

With the correct offset of `0x13ecd0000` we can use this to find where the partition should begin. As per VMFS whitepapers, we know that the VMFS volume begins 1MB from the beginning of the partition we can subtract 1MB, (0×100000):

Our Calculation:

```
0x13ecd0000 - 0x100000 (1MB) = 5347540992 /
   512 (Sectors) = 10444416
```

We just need to write the new partition values to the RAID 5 disk image:

```
#> fdisk -u /vmfs/devices/disks/[DiskString]
```

When prompted, create a new partition: “n”.

```
When prompted, confirm the new partition is a
                 primary partition: "p".
When prompted, confirm the new partition as
```

```
partition 1: "1".
Provide the first sector (X-X, default 63):
"10444416"
Use the default ending sector (Last sector or
+size or +sizeM or +sizeK): "enter"
Now define the partition type as VMFS:
Command (m for help): t
Selected partition 1
Hex code (type L to list codes): fb
Changed system type of partition 1 to fb (VMFS)
Now, simply write the partition data:
Command (m for help): w
The partition table has been altered.
Calling ioctl() to re-read partition table
Lastly; instruct ESXi to rescan all available
LUN's for partition records:
#> vmkfstools -V
```

Within ESXi, you we can now add the DataStore. Click on the configuration tab within ESXi, and navigate to Storage. Then simply click Add Storage and select Disk/LUN and click next. Our VMFS volume with the corrected partition record will now show up in the list of available storage devices and will be referenced by its Datastore name. We then complete the addition of the Datastore then we can move the recovered Datastore items by browsing the Datastore and copying them to a local disk.

For our recovery, our lab was able to recover all but a single VM from the corrupted RAID 5 volume. Subsequent recovery attempts on the damaged VM were successful by using WinImage to mount and manually extract the Data from the VM image file.

WHAT WENT WRONG?

Any type of RAID array (regardless of its configuration) serves as a form of active data redundancy. If you are using a single storage array to power your business, you're practicing a form of stupidity.

The best advice a data recovery lab can provide someone who wants to self-recover data is to never work with the source. The first goal of any data recovery lab is to duplicate any recoverable data (in a corrupt or partial form) and manipulate the duplication. Data Recovery Software does not have an "*undo button*", changes made the source data not only complicate a data recovery case, but can irreversibly destroy data. During this recovery many critical errors executed by the client complicated our data recovery efforts. The client failed to utilize a hotspare, changed the order of active members within the array, wrote new configuration data to unhealthy drives, and failed to keep physical records.

Other avoidable complications involve the way the client initially setup the RAID array. Following industry standard safeguards would have reduced chances of failure and the diagnostic investment needed to recover data:

RESOURCES

- PartedMagic *http://partedmagic.com/doku.php*
- dd_rescue: *http://www.forensicswiki.org/wiki/Dd_rescue*
- Hdparm: *http://en.wikipedia.org/wiki/Hdparm*

INDUSTRY STANDARD SAFEGUARDS

- Never store your operating system and datastore on the same array.
- Never store you data and backups (regardless of form) on the same array.
- Perform a SMART or Health Check on all members of the array before your attempt to rebuild.
- Utilize a hotspare; having an active hot spare greatly reduces the chances of failures and data loss during a RAID array rebuild. Remember, when a RAID array is rebuilding remaining members of the array are stressed until the rebuild process is completed. The chance of another drive failing during the rebuild process is high.
- Never under any circumstances remove active members from a RAID array.
- Record disk order and RAID configuration (type, stripe, and brand) in a technical document kept outside the datastore.

Author's Bio

iCube Development, is a data recovery lab operating in Calgary Alberta. We've been the recipient of many industry awards and are a recognized leader for data recovery in Western Canada. Our industry contributions and data recovery services have been featured in mainstream media receiving Global news coverage.

[i]CUBE DEVELOPMENT

GREP AND REGEX, THE OVERLOOKED FORENSIC TOOLS

by Dr Craig S Wright GSE GSM LLM MStat

This article takes the reader through the process of learning to use GREP and Regular Expressions (RegEx). GREP May not seem to be a tool that relates to the process of data recovery, but we will show that this is an essential tool in recovering data. If you cannot find data, how can you recover it?

What you will learn:

- Intermediate – Advanced level use of Linux command line tools for Digital Forensics (GREP & RegEx)
- Windows Data Recovery

What you should know:

- Basics of Digital Forensics
- Basics of Linux Command line use
- Basic Windows system files

Using the GREP command we can search through a variety of information sources. For the forensic analyst, incident handler or system administrator, this means a simplified method of searching for information. Coupled with the use of regular expressions grep is a powerful tool for the IT investigator. In this paper, we look at some uses of grep and regular expressions.

INTRODUCTION

We are going to approach data recovery from a different perspective in this article. Windows partitions can be found as they have the same values as a final marker. This is 0x55AA for partitions. Most files have special markers that allow you to determine at least the start of a file if not the end [1]. The issue we have is in finding these values on our media. Whether we have a captured pcap trace or a drive image or even if we are looking at a damaged hard drive or USB key, we cannot start to carve data using tools such as DD unless we know where the start and end of the file is.

In some cases (such as with Word documents) the length of the file does not matter greatly. As long as we have captured the entire file, the extra data will be overlooked. In others, it is critical to find both the start and end markers.

We start by asking what GREP is. GREP is a *nix [2] command that allows you to search for a pattern in a list of files. This article takes the reader through the process of learning to use GREP and Regular Expressions (RegEx). Using the GREP command we can search through a variety of information sources. For the forensic analyst, incident handler or system administrator, this means a simplified method of searching for information. Coupled with the use of regular expressions grep is a powerful tool for the IT investigator. In this paper, we look at some uses of grep and regular expressions.

GREP is available for Windows hosts as well. We will not detail all of the different variants in this paper but leave the reader to try a ported version of GREP on their Windows machine themselves.

- In the following video, you can see a little of how dd is used to carve files. *http://www.youtube.com/watch?v=mnhzltE3G68*
- In the following, you can see a little more of Grep and RegEx in action. *http://www.youtube.com/watch?v=a7OkqhcmCSg*
- Finally, before we start to delve into GREP, you can read a little more on DD and how this can help you. *http://gse-compliance.blogspot.com.au/2008/09/next-tool-dd_02.html*

For more information on how you can use these offsets to carve data and partitions, see my earlier article, "DRIVE AND PARTITION CARVING PROCEDURES" in eForensics Free 1/2012 [3].

USING GREP

In this article I have used the SANS Forensic Workstation [4]. This is a free Linux based Virtual Machine that is pre-configured for forensics. It has a number of tools setup and ready to be run and is a good introductory forensics workstation for those wanting to learn how to conduct an investigation.

You use GREP in the following manner:

```
grep [pattern] <file-name1> <file-name2>
```

For example, to look for the string "password file" in any directory under the `/usr/local/` base directory we enter:

```
bash % grep 'password file' /usr/local/*
```

USING GREP ALONE

GREP can be used alone or in conjunction with regular expressions. In Figure 1 we have issued the following command:

```
$ grep 'Apr 3 07' /var/log/messages
```

This command basically shows how you can use grep to extract lines containing a particular string from a text file. The above command could be used to find out all information occurring in the messages log that took place at a particular time. In this instance, the command will return any logs in the "messages" log from the time and date 03rd of April between 07:00 and 07:59:59.

This is also possible to reverse. Using the following command we can display all of the log entries that DID NOT occur on the 03rd Apr between 07:50:00 and 07:50:59:

```
$ grep -v 'Apr 3 07:50' /var/log/messages
```

This will print all the entries in the file other than that which matches the selected pattern (Figure 2).

Notice, we are searching for all except the search string. This ability to select a pattern or the entries outside the pattern makes GREP an extremely powerful search tool.

RETURNING THE FILES

Grep is also of use in seeking files that contain a pattern and returning the names of the files.

```
$ grep -l ' kernel: ' /var/log/*.1
```

In Figure 3 we see the results of this command.

The above command searches for those files that end with a '.1' (within the `/var/log directory`) and in which the text 'kernel: ' is present. This variant of the command will only return the names of these files and not the lines where it found the string.

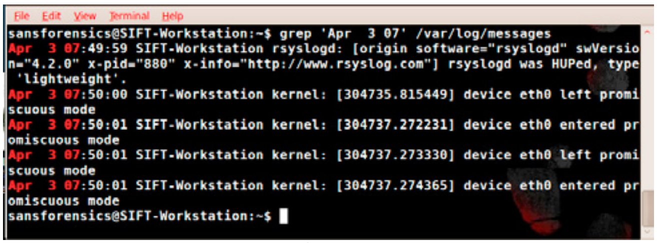
Compare this to the following command (also in Figure 3):

```
grep ' kernel: ' /var/log/*.1
```

The "-l" flag allows us to return just the filename and not the actual line where the match occurred.

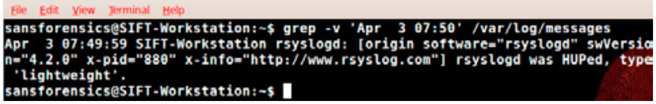
REFINING TEXT SEARCHES

The following commands search for text in a more refined way.



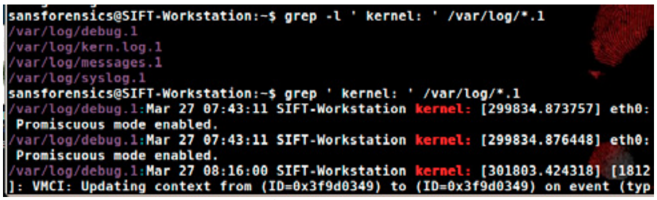
```
sansforensics@SIFT-Workstation:~$ grep 'Apr  3 07' /var/log/messages
Apr  3 07:49:59 SIFT-Workstation rsyslogd: [origin software="rsyslogd" swVersio
n="4.2.0" x-pid="880" x-info="http://www.rsyslog.com"] rsyslogd was HUPed, type
 'lightweight'.
Apr  3 07:50:00 SIFT-Workstation kernel: [304735.815449] device eth0 left promi
scuous mode
Apr  3 07:50:01 SIFT-Workstation kernel: [304737.272231] device eth0 entered pr
omiscuous mode
Apr  3 07:50:01 SIFT-Workstation kernel: [304737.273330] device eth0 left promi
scuous mode
Apr  3 07:50:01 SIFT-Workstation kernel: [304737.274365] device eth0 entered pr
omiscuous mode
sansforensics@SIFT-Workstation:~$
```

Figure 1. *GREP and the log file*



```
sansforensics@SIFT-Workstation:~$ grep -v 'Apr  3 07:50' /var/log/messages
Apr  3 07:49:59 SIFT-Workstation rsyslogd: [origin software="rsyslogd" swVersio
n="4.2.0" x-pid="880" x-info="http://www.rsyslog.com"] rsyslogd was HUPed, type
 'lightweight'.
sansforensics@SIFT-Workstation:~$
```

Figure 2. *GREP to select all but the pattern*



```
sansforensics@SIFT-Workstation:~$ grep -l ' kernel: ' /var/log/*.1
/var/log/debug.1
/var/log/kern.log.1
/var/log/messages.1
/var/log/syslog.1
sansforensics@SIFT-Workstation:~$ grep ' kernel: ' /var/log/*.1
/var/log/debug.1:Mar 27 07:43:11 SIFT-Workstation kernel: [299834.873757] eth0:
 Promiscuous mode enabled.
/var/log/debug.1:Mar 27 07:43:11 SIFT-Workstation kernel: [299834.876448] eth0:
 Promiscuous mode enabled.
/var/log/debug.1:Mar 27 08:16:00 SIFT-Workstation kernel: [301803.424318] [1812
]: VMCI: Updating context from (ID=0x3f9d0349) to (ID=0x3f9d0349) on event (typ
```

Figure 3. *GREP to select files with our pattern*

```
$ grep -w '\<password' *
$ grep -w 'password\>' *
```

The first command searches for those lines where any word in that line begins with the letters 'password'.

The second command searches for those lines where any word in that line ends with the letter 'password'.

In this way, we can also start to set the position of the strings we seek and to differentiate from selected strings.

REDIRECTING OUTPUT AND INPUT

Like most command line tools in *Nix, we can pipe output to GREP. For example, we can send the results of a string command to GREP and search for selected ASCI patterns.

```
strings ./hackin9.pcap | strings ./hackin9.pcap
  | grep 'User-Agent:'
```

In the command above and in Figure 4 we have returned any line from the "strings" command that contains the pattern "User-Agent: ". This is useful in searching binary files for patterns.

Piping allows us to take any number of other commands and sent the output to another command. In this case, we are stripping all of the unreadable characters (those which we cannot display on screen) and searching through a binary pcap (network capture) file for User Agent strings.

MORE PIPES

This is just the start of what we can do using piped commands. This is of use to security personal, network administrators and more. Even checking the permissions on files is simplified. For instance, if we wish to list all of the files that any user can write to and alter, we could issue the following command:

```
ls -la / | grep 'rwxrwxrwx'
```

Here we are returning all chmod 777 files in the root directory.

PIPING PIPES

There is also no reason to limit the command stream to a single pipe. We can also pipe the output from more than on command.

```
sudo du /usr | grep 'gz' | more
```

The command above will return a list of filenames and their sizes containing the string '*gz*' in the `/usr` directory.

In Figure 5 we see that the use of nested pipes can be used to return output from one command to become input to another.

LOOKING AT PROCESSES

As an incident handler, it will be necessary to find rogue processes. GREP allows us to simplify this search. For instance, if we are seeking a running process, we can search for only selected processes. In this example, we have the sshd and hald daemons in our search and instead of scrolling through all of the running processes, we can return only the required information.

```
ps aux | grep "sshd\|hald"
```

In the command above and in Figure 6 we are looking at the output of ps for any process including hald OR sshd.

The characters '\|' act as a logical OR in grep.

OTHER USES OF GREP

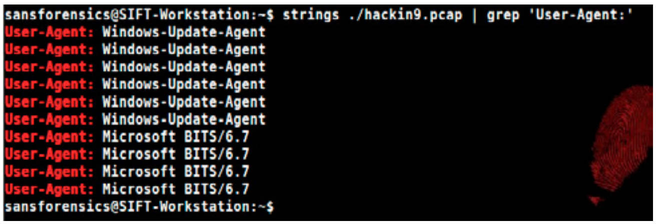
The GREP command is also handy for counting the number of times a string occurs in a file. For example, the following command is used to count patterns:

```
grep -c false /etc/passwd
```

To be exact, we have counted how many times the string '*false*' has occurred within the file `/etc/passwd`.

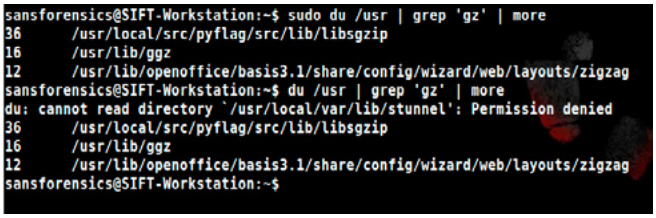
SOME GREP OPTIONS

The following are some of the main options used in the GREP command.



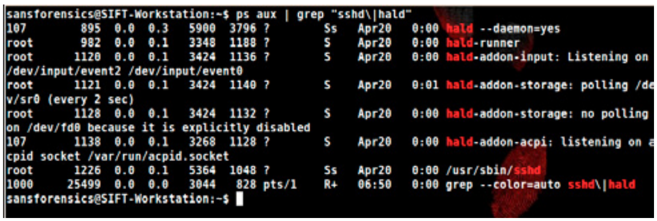
```
sansforensics@SIFT-Workstation:~$ strings ./hackin9.pcap | grep 'User-Agent:'
User-Agent: Windows-Update-Agent
User-Agent: Windows-Update-Agent
User-Agent: Windows-Update-Agent
User-Agent: Windows-Update-Agent
User-Agent: Windows-Update-Agent
User-Agent: Windows-Update-Agent
User-Agent: Microsoft BITS/6.7
User-Agent: Microsoft BITS/6.7
User-Agent: Microsoft BITS/6.7
User-Agent: Microsoft BITS/6.7
sansforensics@SIFT-Workstation:~$
```

Figure 4. *GREP to search for output in piped commands*



```
sansforensics@SIFT-Workstation:~$ sudo du /usr | grep 'gz' | more
36      /usr/local/src/pyflag/src/lib/libsgzip
16      /usr/lib/ggz
12      /usr/lib/openoffice/basis3.1/share/config/wizard/web/layouts/zigzag
sansforensics@SIFT-Workstation:~$ du /usr | grep 'gz' | more
du: cannot read directory `/usr/local/var/lib/stunnel': Permission denied
36      /usr/local/src/pyflag/src/lib/libsgzip
16      /usr/lib/ggz
12      /usr/lib/openoffice/basis3.1/share/config/wizard/web/layouts/zigzag
sansforensics@SIFT-Workstation:~$
```

Figure 5. *Many Pipes*



```
sansforensics@SIFT-Workstation:~$ ps aux | grep "sshd\|hald"
107        895  0.0  0.3   5900  3796 ?        Ss   Apr20   0:00 hald --daemon=yes
root       982  0.0  0.1   3348  1188 ?        S    Apr20   0:00 hald-runner
root      1120  0.0  0.1   3424  1136 ?        S    Apr20   0:00 hald-addon-input: Listening on
/dev/input/event2 /dev/input/event0
root      1121  0.0  0.1   3424  1140 ?        S    Apr20   0:01 hald-addon-storage: polling /de
v/sr0 (every 2 sec)
root      1128  0.0  0.1   3424  1132 ?        S    Apr20   0:00 hald-addon-storage: no polling
on /dev/fd0 because it is explicitly disabled
107       1138  0.0  0.1   3268  1128 ?        S    Apr20   0:00 hald-addon-acpi: listening on a
cpid socket /var/run/acpid.socket
root      1226  0.0  0.1   5364  1048 ?        Ss   Apr20   0:00 /usr/sbin/sshd
1000     25499  0.0  0.0   3044   828 pts/1    R+   06:50   0:00 grep --color=auto sshd\|hald
sansforensics@SIFT-Workstation:~$
```

Figure 6. *Logical Operators*

-v Reverses the normal behaviour of the grep command – Instead of selecting lines, it rejects the lines that match the given criteria.
-c It suppresses the normal output and only prints the total count of matching lines instead of the actual lines.
-i Ignores the case of the text when matching the given pattern. For example it would treat "the" and "The" as the same word
-w Checks if the given pattern is a word by itself and not a part of another word. Thus if you search for 'pass' and the word 'password' is present in a file, the particular line containing that word would not be returned in the result.
-l Only returns the names of the files in which the given pattern was found.
-r Checks for the given pattern, recursively within the directory that you specify after the -r option
-n precedes each line with the line number where it was found

STARTING TO LOOK AT REGULAR EXPRESSIONS

If we are to look at what we can really do with GREP, we need to use Regular Expressions. To take a definition from Regular-Expressions.info [5]:

> *"A regular expression (regex or regexp for short) is a special text string for describing a search pattern. You can think of regular expressions as wildcards on steroids".*

GREP works with RegEx in several ways:

- grep -e Use a Posix based search pattern
- grep –E egrep – Use extended Regular Expressions
- grep –P RegEx as with Perl (for those Perl lovers amongst us)
- grep –o By Default, grep will display the entire line which matches the search pattern.
  "-o" is the option that allows us to only return the part of the line matched.
- grep –n Prefix each matched line or part of a matched line with the line number that the match was found at
- grep -x Forces a match of the entire line and not a part of a line

The following command would display those lines (from the files ending with an extension ".conf" in the `/etc directory`) that start with a '#'.

```
grep ‘^#’ /etc/*.conf | less
```

The term '^#' means that # should be present as the first character on a line. The piped command "Less" basically displays the output a page at a time and allows you to scroll the results where the output exceeds one page. This is like "more" as we used above, but is generally more of a Linux command than a UNIX command.

Alternatively we can search for patterns with Regular Expressions.

```
grep -v ‘^[0-9]’ /var/log/* | more
```

This command above searches for lines within the files in the `/var/log` directory having any of the numbers from 0-9 in them as the first character on the line. The command then displays all the lines except the ones it found initially. See how we have selected any number as the search pattern.

Using Regex, we can make some complex searches. For instance, we can find IP addresses in a disk image:

```
strings ./Image.dd | grep -E ‘\b(25[0-5]|2[0-4]
[0-9]|1[0-9][0-9]|[1-9]?[0-9])\.(25[0-5]|2[0-4]
[0-9]|1[0-9][0-9]|[1-9]?[0-9])\.(25[0-5]|2[0-4]
[0-9]|1[0-9][0-9]|[1-9]?[0-9])\.(25[0-5]|2[0-4]
[0-9]|1[0-9][0-9]|[1-9]?[0-9])\b’
```

Or for instance, we can find RFC-2822 emails in a disk image for selected domain extensions:

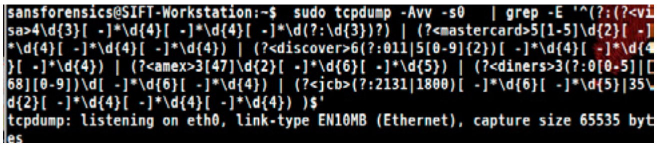
```
./Image.dd | grep -P ‘[a-z0-9_]+(?:\.[a-z0-9]+)*@
(?:[a-z0-9](?:[a-z0-9-]*[a-z0-9])?\.)+(?:[A-Z]
{2}|asia|com|org|net|gov|mil|biz|info|mobi|name|ae
ro|jobs|museum|travel|au)\b’
```

For instance, we can find and name credit cards in a pcap image:

```
strings ./hackin9.pcap | grep -E
‘^(?:(?<visa>4\d{3}[ -]*\d{4}[ -]*\d{4}[ -]*\d(?:\
d{3})?) | (?<mastercard>5[1-5]\d{2}[ -]*\d{4}
[ -]*\d{4}[ -]*\d{4}) | (?<discover>6(?:011|5[0-9]
{2})[ -]*\d{4}[ -]*\d{4}[ -]*\d{4}) |
(?<amex>3[47]\d{2}[ -]*\d{6}[ -]*\d{5}) |
(?<diners>3(?:0[0-5]|[68][0-9])\d[ -]*\d{6}
[ -]*\d{4}) | (?<jcb>(?:2131|1800)[ -]*\d{6}
[ -]*\d{5}|35\d{2}[ -]*\d{4}[ -]*\d{4}[ -]*\d{4})
)$’
```

In this more complex example, we return the values based on the type of credit card (mastercard, visa, etc.) returned from the strings command run on our network capture. We could even pipe a live capture using tcpdump as we can see in Figure 7.

Grep can become an extremely powerful search tool with the simple addition of Regular Expressions.



```
sansforensics@SIFT-Workstation:~$  sudo tcpdump -Avv -s0   | grep -E '^(?:(?<vi
sa>4\d{3}[ -]*\d{4}[ -]*\d{4}[ -]*\d(?:\d{3})?) | (?<mastercard>5[1-5]\d{2}[ -]
*\d{4}[ -]*\d{4}[ -]*\d{4}) | (?<discover>6(?:011|5[0-9]{2})[ -]*\d{4}[ -]*\d{4
}[ -]*\d{4}) | (?<amex>3[47]\d{2}[ -]*\d{6}[ -]*\d{5}) | (?<diners>3(?:0[0-5]|[
68][0-9])\d[ -]*\d{6}[ -]*\d{4}) | (?<jcb>(?:2131|1800)[ -]*\d{6}[ -]*\d{5}|35\
d{2}[ -]*\d{4}[ -]*\d{4}[ -]*\d{4}) )$'
tcpdump: listening on eth0, link-type EN10MB (Ethernet), capture size 65535 byt
es
```

Figure 7. *more complexity*

Learning these simple tools will make your role as a forensic analyst far easier and faster.

SO WHAT IS REGEX ANYWAY?

Regular Expressions (or Reg Ex) is a special text string that is used to describe a search pattern. Regular expressions are in effect wildcards. Wildcard notations such as *.txt are commonly used to find all text files (or at least files with a .txt extension). The regex equivalent is `.*\.txt$`.

LITERAL CHARACTERS

The easiest regular expression consists of a single literal character (e.g. c which will match the c in Jack). RegEx will only match the first c in the previous example.

METACHARACTERS

There are several characters with special meanings for RegEx and GREP. These are:

1. [ ] Match anything inside the square brackets for ONE character position once and only once, for example, [12] means match the target to 1 and if that does not match then match the target to 2 while [0123456789] means match to any character in the range 0 to 9.
2. \ The \ is an escape character. If you have \) the \ will treat the) as a literal.
3. ^ The ^ (caret) inside square brackets negates the expression, [^Zz] means anything except upper or lower case Z and [^a-z] means everything except lower case a to z. The ^ (circumflex or caret) outside square brackets means look only at the beginning of the target string, ^Mic will not find M1cros0ft but ^Mic will find Microsoft.
4. $ The $ tells the regex to match look only at the end of the target string. E.g. $dog will match black dog but not doggone.
5. . The dot matches a single character, except line break characters. It represents [^\n] .og matches dog and bog
6. | The | (pipe) is referred to as alternation. It refers to a logical XOR of the values on either side. For example, gr(a|e)y will return gray or grey
7. ? The ? matches the preceding character 0 or 1 times. That is ? Is used if a character exists once or not at all. Th? Will match with both Th, The or Tha but not That
8. * The * returns a match on the preceding character 0 or more times. E.g., Stre*t will return Street (2 matches) and Streat (1 match) and Streight (0 matches).
9. + The + returns a match on the previous character 1 or more times (like * with only matches) E.g., Stre+ will return Street (2 matches) and Streat (1 match) but NOT Straight (0 matches).
10. ( ) The parenthesis, (and) are used in order to group search expression together. E.g. ((4\.[0-3])|(2\.[0-3])) could be used to return the string 4.0 in Mozilla/4.0.
11. - The – inside square brackets is the range separator. It allows us to define a range, [0123456789] could be rewritten as [0-9].

If you want to use a metacharacter as a literal in a regex, it is necessary to escape it using the backslash. For instance, in order to match the string 1+3=4 you would use the regex is 1\+3=4.

The plus sign has a special meaning (as we saw above) and needs to be escaped.

CHARACTER CLASSES OR CHARACTER SETS

A character class matches one of a set. For example, the string [HS]ack will match both Hack and Sack but it will not match Shack.

A range of characters can be selected with the – [0-9a-fA-F] will match a hex character (0 to 9 or A-F with or without capitalisation).

The ^ after a [will match anything NOT in the class. For example d[^a] will match do but will not match lad or da. It will not match lad as there is nothing following the d. It will not match da as the a is in the character class and is excluded. That is any character except a will match.

SHORTHAND CHARACTER CLASSES

You can also match characters using shorthand symbols:

- \d matches a single digit character. These are the values (0 to 9)
- \w matches an alphanumeric characters. This also includes an underscore or _
- \s matches any whitespace. Whitespace includes tabs and line breaks.

NON-PRINTABLE CHARACTERS

In any search, there are always non-printable characters to be accounted for as well. This means that you can also find format characters and other non-printable characters using RegEx.

\t a tab character ASCII 0x09
\r a carriage return ASCII 0x0D
\n a line feed ASCII 0x0A

Some of the far less common non-printable characters include

\a bell ASCII 0x07
\e escape, ASCII 0x1B
\f form feed ASCII 0x0C
\v vertical tab ASCII 0x0B

We could for instance Match Line Terminators:

Windows text files use \r\n to terminate lines
*NIX text files use \n.

We can extend this to search for non-ASCI characters and Hex.

`\xFF` matches a specific character using the hexadecimal index value for the character set.

E.g. `\xA9` will find the copyright symbol within the Latin-1 character set.

In Unicode you would use `\uFFFF` to match a Unicode character.

E.g. `\u20AC` matches the euro currency sign.

Any and All non-printable characters can become a part of the regular expression or they can be used as a part of a selected character class.

ANCHORS

An Anchor matches a position within a string. For instance:

`^` will match values that are at the beginning of a string
`$` will match values that occur at the end of a string
`\A` only has a match at the start of the string
`\Z` only has a match at the end of the string

A word boundary is a position between a character that can be matched by \w and a character that cannot be matched by \w.

`\b` matches at a word boundary.
`\b` matches at the start and/or end of the string where the first and/or last characters in the string are word characters.
`\B` will return a match for each position where \b cannot

For more information see: *http://www.regular-expressions.info/anchors.html.*

REPETITION

Then of course we can also look for repeated characters.

- The * character will attempt to match the preceding token zero or more times.
- The + character will attempt to match the preceding token one or more times.
- {} are used in order to specify a select number of repetitions.

  E.g. \b[1-9][0-9]{3,5}\b matches a number between 1,000 and 999,999. That is [1-9] for the first character and [0-9]{3,5} for between 3 and 5 more numerals.

ITERATION

As noted, {n} matches up with the preceding character, or character range exactly n times. So as an example, we could match an Australia international phone number (format +61-2-4478-1000) as:

```
\+61-[0-9]-[0-9]{4} -[0-9]{4}
```

ALL GOOD, BUT WHAT ABOUT FORENSICS?

Now that we know how to use GREP and RegEx, we will start to see how these commands can aid us in a forensic investigation.

Firstly, we can conduct simple searches for files based on their extensions. The following GREP command will search for Microsoft office files:

```
grep -iE '\.(xlsx|xls|doc|docx|ppt|pptx)\b'
```

In this command we are seeking a file with the selected extensions where the patters occur at the trailing end of the file name (the \b switch sets the pattern to the end of the file name). This allows us to quickly search for a set of files and list these.

We can also seek URLs within the files or streams we are searching. The following patter will return a set of selected URLs for either http or SSL sites (the s? in the term https? Makes the s optional and returns both http:// or https://):

```
grep - E '\bhttps?://.+\.(edu|edu.au|org|org.
            au|csiro.au)'
```

In this example, we have a pattern that looks for URLs within the EDU, CSIRO and ORG domains internationally and within the AU country domain. We can of course set any domain name pattern and even look for partial matches.

Using a modification if this command in Figure 8, we see that our image file contains 52 URL entries that match our search term.

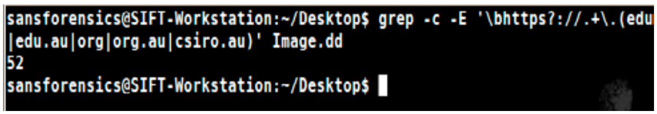
CARVING FILES

Of course, a discussion of using GREP for forensics is not complete without a demonstration on how we can use this command in carving files from a disk image.

We will start with looking at a table (Table 1) of common Hex File headers. More can be found on sites such as *http://www.file-extensions.org/.*

Using the commands below (displayed in Figure 10) we have searched a drive image named Image.dd for word documents using the hex file header we listed in Table 1. The first command counts how many ".doc" format word files we have on the image.

```
grep -aP -c "\xD0\xCF\x11\xE0" Image.dd
```



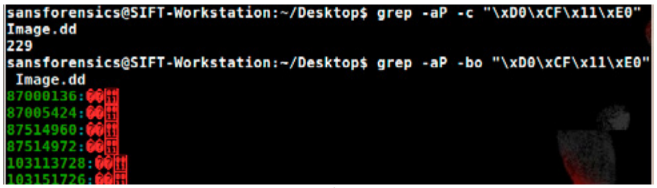
```
sansforensics@SIFT-Workstation:~/Desktop$ grep -c -E '\bhttps?://.+\.(edu
|edu.au|org|org.au|csiro.au)' Image.dd
52
sansforensics@SIFT-Workstation:~/Desktop$ 
```

Figure 8. *The RegEx Bible*

We see from Figure 9 that the drive has 229 files that we can attempt to extract. The hex offset in these commands is loaded using the '-P' or Perl format. Next, using the '-b' flag to only return the matched value and the '-o' flag to return the offset, we have collected a set of the initial offset values for each of the word documents we have located on the drive Image.

Table 1. *Common Hex File Header values*

File Extension	Hex Value of file start
bmp	42 4D F8 A9
bmp	42 4D 62 25
bmp	42 4D 76 03
cab	4D 53 43 46
dll	4D 5A 90 00
Excel	D0 CF 11 E0
exe	4D 5A 50 00
exe	4D 5A 90 00
gif	47 49 46 38 39 61
gif	47 49 46 38 37 61
jpeg	FF D8 FF E1
jpeg	FF D8 FF E0
jpeg	FF D8 FF FE
mp3	49 44 33 2E
mp3	49 44 33 03
PDF	25 50 44 46
Word	D0 CF 11 E0
zip	50 4B 03 04



```
sansforensics@SIFT-Workstation:~/Desktop$ grep -aP -c "\xD0\xCF\x11\xE0"
Image.dd
229
sansforensics@SIFT-Workstation:~/Desktop$ grep -aP -bo "\xD0\xCF\x11\xE0"
 Image.dd
87000136:
87005424:
87514960:
87514972:
103113728:
103151726:
```

Figure 9. *Finding a starting offset*

```
grep -aP -bo "\xD0\xCF\x11\xE0" Image.dd
```

This process can also be used in carving mobile data as well as on memory images.

Some files have both a starting hex value as well as an end. Word files do not. This makes it a little more difficult to determine the end point, but luckily, as long as we capture the entire file word will read the document correctly and ignore the extra input.

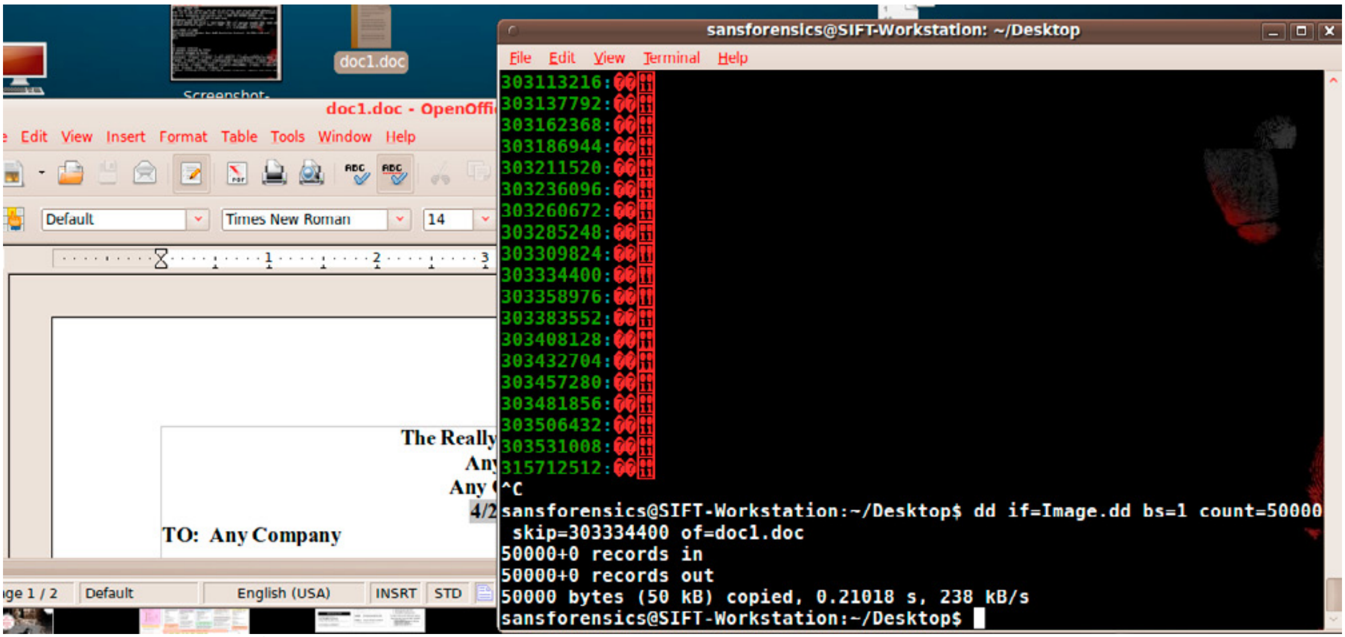
In this example (Figure 10) I have randomly selected one of the 229 possible files returned in our initial search. In this case the one starting at offset 303334400. Using 'dd' it is easy to carve this file using a command such as that below:

```
dd if=Image.dd bs=1 count=50000 skip=303334400
               of=doc1.doc
```

In this command, I have set the block size (bs) to equal 1. This just simplifies the process of jumping to the start of the file we want to carve. In the command, 'if' defines the file we use as input and 'of' defines the name of the file we are carving as output.

The command has used 'count=50000' to carve 50000 1 bit sized blocks starting from our chosen start point which was set using 'skip=303334400'. As can be seen in Figure 10, the file we have created opens with formatting as a fully functional Word document.

To get rid of the extra data, we can save the file from word. This is not actually necessary and from



```
303113216:
303137792:
303162368:
303186944:
303211520:
303236096:
303260672:
303285248:
303309824:
303334400:
303358976:
303383552:
303408128:
303432704:
303457280:
303481856:
303506432:
303531008:
315712512:
^C
sansforensics@SIFT-Workstation:~/Desktop$ dd if=Image.dd bs=1 count=50000
 skip=303334400 of=doc1.doc
50000+0 records in
50000+0 records out
50000 bytes (50 kB) copied, 0.21018 s, 238 kB/s
sansforensics@SIFT-Workstation:~/Desktop$
```

Figure 10. *Carving the file*

a forensic perspective adds little, but it does mean we can save a little disk space.

In Figure 11, the file has been opened and saved from Word as “doc1.new.doc”. The ‘ls’ command shows the size of these files and we can see the original carved file with a file size of 50,000 and the saved file with the actual size minus all the extra data of 19,456 bytes.

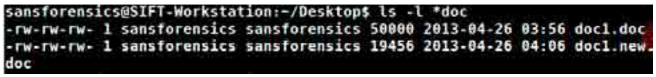
Taking this further, it is possible to script the output from GREP and create a simple process that runs through each of the offset values found and feeds these directly to “dd” with an incremental value returned as a name for the carved files.

I will leave this exercise for the reader to try…

TO CONCLUDE…

GREP when coupled with Regular Expressions is one of the most powerful tools in the incident handler and Forensic analyst’s toolkit. When used well and coupled with other commands, GREP creates a simple method to be able to quickly search files and data for signs of intrusions, for selected data (such as credit card numbers, email addresses Etc.) or just to investigate the contents of files.

There are MANY more RegEx Strings and one way to learn these well is to read and try the *Regular Expressions Cookbook* [6].



```
sansforensics@SIFT-Workstation:~/Desktop$ ls -l *doc
-rw-rw-rw- 1 sansforensics sansforensics 50000 2013-04-26 03:56 doc1.doc
-rw-rw-rw- 1 sansforensics sansforensics 19456 2013-04-26 04:06 doc1.new.
doc
```

Figure 11. *The saved file*

Detailed Solutions in Eight Programming Languages
Regular Expressions Cookbook
O'REILLY
Jan Goyvaerts & Steven Levithan

Figure 12. *The RegEx Bible*

RESOURCES

[1] A good site to see a list of these values is *http://www.garykessler.net/library/file_sigs.html*
[2] In this paper we refer to Linux and Unix systems collectively as *Nix.
[3] See *http://eforensicsmag.com/oracle-forensics-detection-of-attakcs-through-default-accounts-and-passwords-in-oracle-eforensics-free/*
[4] *http://computer-forensics.sans.org/community/downloads*
[5] *http://www.regular-expressions.info/*
[6] See Regular Expressions Cookbook Jan Goyvaerts (Author), Steven Levithan for the source

Also, have a look at the website: *http://www.zytrax.com/tech/web/regex.htm*. Both are great ways to start learning Regular Expressions and to make your Incident handling and forensic work easier.

GREP and RegEx are some of the most powerful and also most overlooked forensic and incident handling tools available. Maybe it is time to have a look at these and to make your searches faster. More, when you know where a file is, what the offsets are and the length, you can carve the file or partition and recover your lost data.

Author’s Bio

Dr Craig Wright (Twitter: Dr_Craig_Wright) is a lecturer and researcher at Charles Sturt University and executive vice –president (strategy) of CSCSS (Centre for Strategic Cyberspace+ Security Science) with a focus on collaborating government bodies in securing cyber systems. With over 20 years of IT related experience, he is a sought-after public speaker both locally and internationally, training Australian and international government departments in Cyber Warfare and Cyber Defence, while also presenting his latest research findings at academic conferences. In addition to his security engagements Craig continues to author IT security related articles and books. Dr Wright holds the following industry certifications, GSE, CISSP, CISA, CISM, CCE, GCFA, GLEG, GREM and GSPA. He has numerous degrees in various fields including a Master’s degree in Statistics, and a Master’s Degree in Law specialising in International Commercial Law. Craig has just completed working on his second doctorate, a PhD on the Quantification of Information Systems Risk and is mad enough to be planning his third doctorate.

RAID 5 DATA RECOVERY – A GUIDE FOR THE RAID OWNER

by Wayne Horner

This article is a guide for RAID owners. Its purpose is to educate the owner of the potential dangers faced when attempting to restart a failed RAID array. Common pitfalls are described. DO's, DON'Ts and safe practices are explained.

What you will learn:

- Some techniques to identify devices
- As behaves a system emulated

What you should know:

How to protect yourself from data loss:

- What is a RAID, how it works.
- Parity – a simple explanation.
- 4 Rebuild traps to avoid..
- How to safely restart the RAID.
- Why you should not reset parity to recover data.
- How to properly rebuild a missing drive.

Your business stores data in a RAID storage array and the RAID has failed. Most RAID failures are recoverable but many times data is lost during the recovery process. IT consultants may mean well but be unaware of the traps. Tech support call centers can be more interested in quickly closing the support ticket. This article will outline some common traps and safe procedures you need to know to protect your data.

With 30 years of RAID recovery experience we have seen many instances of data being lost when the 'rescue' begins. There are 2 types of rescuers that you need to be careful of: IT consultants and tech support departments. IT consultants often think that RAIDS are redundant and can't be harmed. They may start swapping drives and boards and performing irreversible rebuilds – without experience they may not recognize the irreversible consequences. Your RAID manufacturers' tech support is interested in closing the support ticket in the shortest time. From their point of view they are successful when the RAID is functioning again – even if your data is lost. Besides you were supposed to back it up. The safest advice is to take your array to a reputable data recovery expert with lots of RAID experience. The next best thing is to backup all the drives one-by-one – but this will take time and disk drives.

BACKGROUND: WHAT IS A RAID?

A RAID is a disk storage system where multiple drives are blended together and presented to the OS virtually as a big single disk drive. To the OS it's a big disk drive but it's really a group of smaller drives. This virtual transformation magic is made possible by the RAID controller. This is usually an intelligent controller board or software drivers in the OS. Redundancy means that you can lose 1 drive and still not lose your data. This redundancy is made possible by RAID 5 parity. How does this work?

RAID-5 PARITY: A SIMPLE EXAMPLE OF HOW IT WORKS

Addition... it's just addition. How can you lose 1 drive from a 4 drive array (or 10) and not lose any data? Whats the trick behind this? It can be illustrated with a simple example:

Visualize a tiny 4-drive RAID-5 array composed of 4 disk drives – 3 data drives and 1 for parity:

- Drive 1 contains only the number 3.
- Drive 2 contains only the number 6.
- Drive 3 contains only the number 2.
- Drive 4 is used for parity and is a backup of all the other drives.

For this illustration, parity is the numeric total of the 3 data drives: 3 + 6 + 2 = 11. So let's say that Drive 2 fails! Now we have a missing drive array: 3 + ?? + 2 = 11. It's a degraded array because 1 drive is failed. Its easy to calculate that the missing data must be 6, because only 3 + 6 + 2 =11. From this example you can see some of the features of RAID-5 parity.

- All the drives must be equal in size.
- You lose 1 the capacity of 1 drive to store parity.
- You can lose 1 drive and still not lose data.
- If you lose 2 drives then you lose data.
- Resetting parity loses the previous backup that parity provides.

RAID-5 DEGRADED MODE: WHAT IS IT?

A RAID-5 array can run with 1 drive failed – this is called degraded mode. The RAID controller is using parity to simulate the missing drive. This is bad – it runs slow, there is no safety and if another drive fails then you will lose data. Sometimes an array can run for months degraded, meanwhile the offline drive is becoming more stale. Eventually a second drive fails and the RAID stops.

OUT-OF-SYNC: WHAT IS THAT?

Out-of-sync means that the parity and the data don't agree. In our parity example this could occur if you replaced failed drive #2 with a new blank drive full of 0's.

- Out-of-sync raid, data corrupt: 3 + 0 + 2 = 11 <<<< Parity doesn't agree with the data.
- Now in-sync sync raid, data corrupt: 3 + 0 + 2 = 5 <<<< After resetting parity.
- Now in-sync sync raid, data correct: 3 + 6 + 2 = 11 <<<< After proper rebuild.

Your raid controller can detect this situation and tell you that you need to reset the parity. There are 2 ways to fix this, either reset the parity or rebuild the missing drive.

- Reset the parity ONLY if you don't care about the data.
- Reset the parity ONLY if you are absolutely sure that the filesystem and data are correct.
- Resetting the parity is irreversible – you cannot get the lost data back!

RAID FAILURE FIRST STEP: ASSESSING THE SITUATION

First stop and document the situation with notes and pictures. Physically examine your RAID and make notes of the following conditions:

- How many drives are in the array?
- What sizes are the drives?
- What lights are on and what colors, are they blinking?

RED could be a fault. GREEN blinking usually means activity.

If your RAID-5 has failed then there should be 2 drives in an off-line condition – are 2 different?

– Is there a display panel? What messages is it showing?

Label your drives, being careful to note which drives are in which slots. Be aware that RAIDS often number drives starting with 0 while humans number things starting with 1. So a 5 drive RAID array may be numbered 0-1-2-3. But humans will label them 1-2-3-4. This can lead to confusion when the array says drive 3 is faulty – which drive 3?

Examine your system console and look for any error messages.

Does your RAID software send email to an account? What about the event log?

Most RAID controllers have an administration program which is available when the system is powering up. For example if you have an Adaptec controller then pressing CONTROL-A will let you enter the RAID administration program. From there you can examine and document how your array is configured. How many drives are there? Are there any hot-spares? Is it RAID 5? 6? 10? Serial numbers and dates are important.

CALCULATE THE SIZE AND TYPE OF RAID

One of the first things that data recovery is going to want to know is what size volume does your OS see? That is, how much total disk space did your system think the RAID provided? This is important to determine the size and configuration of your RAID. RAID-5 loses 1 drive worth of space for parity. So a RAID-5 with 4 100 GB drives would have

3 x 100 = 300 GB of usable space for your OS to use. This calculation is complicated by type of RAID, presence of hot-spares etc – its best to have this documented before failure.

POWER DOWN AND TEST THE DRIVES

So at this point you should have a good idea of the size and configuration of your RAID. Now you are ready to power it down. This is usually where your IT support person will take over. But you need to know what's going to happen so that you can guard your data. Most IT techs at this point are going to power off and then eject and reinstall each drive. They are testing that it's just an intermittent glitch – and that reseating all the drives will clean the electrical contacts and maybe the RAID will restart all by itself. This can be dangerous. Some RAID controllers have a 'feature' called auto-rebuild. This feature lets the RAID decide that it needs to rebuild a drive – so simply powering up with all drives can start the rebuild. But if the RAID controller gets confused it could rebuild the wrong way and actually destroy data. So the safest approach is to power off, reseat the drives and then eject all but one drive. Now you can power up and test the single drive by itself without danger of the RAID array auto-rebuilding. Now that you are in the bios with only 1 drive you can examine its status. Is it clean? Is it part of the array? What date was it last a part of the array? This last question is very important! Pay very close attention to the dates that are stored for each drive. You want to look for a drive that is stale. While in the RAID administration program you should look for and turn-off any auto-rebuild feature.

CLASSIFYING THE RAID DRIVES

At this point you have tried to test and examine all the drives. You need to classify the drives into 2 groups, current drives and stale drives:

- *Current drives* are the ones that were functioning together the last time the RAID was in use. They should all have the same date stamp. For RAID-5 you can only have 1 drive that's missing. For our 4-drive RAID example you need at least 3 current drives in order to recover.
- *Stale drives*. A stale drive is one that's been offline – that is the RAID controller determined it was faulty and removed it from the array. The RAID continues functioning in the degraded mode. Sometimes when you repower the array the stale drive reappears. You don't want to use it to rebuild the array. Its data is old and doesn't fit with your current data – using a stale drive will create a corrupt file system. You may not find a stale drive – because its faulty and missing. As long as you have a current set minus 1 then you can rebuild.

REBUILDING A LOST DRIVE

If you have a complete set of current drives with only 1 missing then you can regenerate the missing drive by doing a rebuild. Rebuilding itself is not dangerous, but there are some traps.

REBUILD TRAP #1: REBUILDING THE WRONG DRIVE

This shouldn't happen – your RAID controller should be smart enough to prevent this but... One way to do this wrong is to get confused about drive numbering. For example, your RAID controller numbers the drives 0-1-2-3. Your chassis has slots numbered 1-2-3-4. So just make sure that you are writing to your new blank drive and not overwriting a current data drive. Some raid controllers have an IDENTIFY function which will blink the light on the selected drive.

REBUILD TRAP #2: REBUILDING WITH A STALE DRIVE

Rebuilding with a stale drive is not in itself dangerous – you aren't overwriting current data. The problem comes when you reboot the system and a filesystem check is run. It will find that the filesystem needs repairs. The OS will make these repairs and in the process create lots of permanent damage. You need to follow the safe-restart process outlined below.

REBUILD TRAP #3: FAILURE TO FINISH

In order to rebuild a new drive the RAID controller must read every single sector of every drive, calculate the parity and write the regenerated data to the new drive. What can happen though is that the rebuild process can abort early. It's frequent that other bad sectors will have developed on the other drives. So if the rebuild process stops at a bad sector then you will have a partially rebuilt drive and a file system that is corrupt. It will also be out-of-sync. A mistake at this point is to reboot and let the OS 'repair' the file system. This will cause more damage.

REBUILD TRAP #4: RESETTING PARITY

An out-of-sync condition can be caused by replacing a drive or after a failed or partial rebuild. The raid controller can detect that it is out-of-sync and want you to reset the parity. This will just destroy the correct data from the correct file system. The proper solution is to clone the bad second failing drive and try the rebuild again. This is getting advanced and should be done by a data recovery professional.

What if you don't have a current set of drives to rebuild with?

In this case you will need to have the faulty drives repaired and cloned first.

What if you can't tell which are the current drives?

In this case you will need to rebuild the array with different combinations and figure out which gives the best outcome (Figure 1).

Picture of 2-drives-failed scenario. 6 drive, 300gb, SAS RAID-5 array. 2 Drives are completely crashed- filled with metal filings, media ground to a powder. These are high-speed 15K RPM drives. They run super-fast, super-hot and crash super-fast. Last week we had the exact same – only it was 2 crashed in a 5 drive array.

THE 2-DRIVES FAILED SCENARIO – THE MOST COMMON RAID-5 FAILURE

Its a frequent occurrence with RAID-5 to have 2 drives failed at once. The reason is simple. What really happens is that first 1 drive fails – and nobody notices. The RAID-5 continues to operate in degraded mode – simulating the missing drive. It can run this way for months. We once recovered a Vmware raid that ran degraded for 18 months! The system runs a little slower than usual but users don't notice until the second drive fails. Now the RAID is forced to stop and it appears as if 2 drives failed at once.

THE 2-DRIVES FAILED TRAP

The danger and risk of the 2-drives failed scenario is best illustrated with an example: You have a 4 x 100 GB RAID-5, one drive fails but the RAID continues to function in degraded mode with the 3 remaining drives. 6 months later, a second drive fails and the RAID halts. So let's say the IT tech comes in and repowers all the drives. So he needs the 3 *current* drives working for a perfect recovery. The *current* failed drive has a bad board – it won't restart without data recovery. But the one from 6 months ago only had a bad sector glitch and restarts. The smart RAID array is saying its stale and doesn't want to use it. The IT tech can force the RAID to accept the drive. He puts in a new blank drive and uses the stale drive to rebuild. On rebooting the system is when the second bad thing happens. The OS says that the file system is damaged and needs to be repaired. So the IT tech lets it run disk repair. This creates all kinds of damage and lost files. We have seen cases where the filesystem repair ran for 10 hours. When it's done it's a mess of missing and corrupt files.

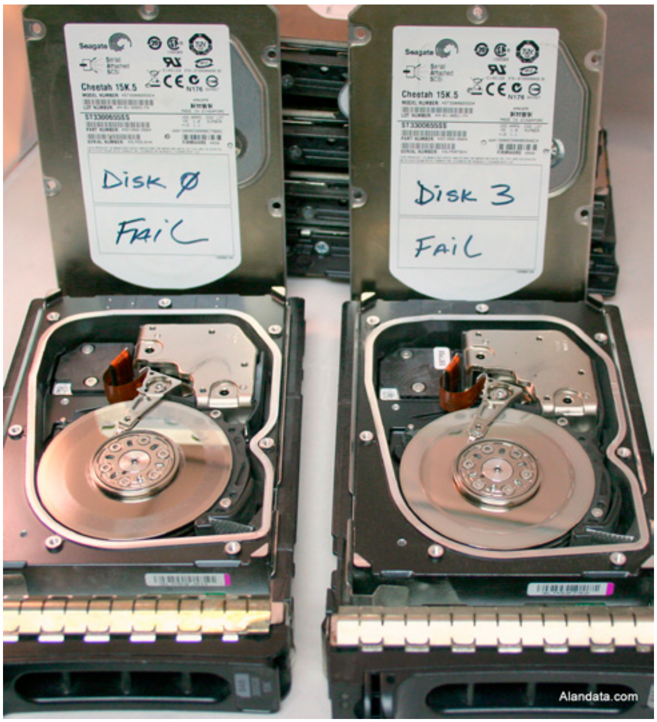


Figure 1. *Two failed hard drives after RAID*

True story: I was diagnosing an array failure and I asked aloud how nobody noticed the giant red error message on the boot screen? The secretary that came in early every day to start the system said that "Oh yes – that message had appeared one day – but if she just pressed return, then the computer would continue and boot just like always"...

How to avoid the 2-drives-failed-simultaneously disaster.

Rebuilding the new blank drive didn't cause any damage. But you have to boot the system carefully and safely and check it for damage first.

DO: SAFELY-RESTARTING – HOW TO BOOT SAFELY AND CHECK FOR DAMAGE

After a RAID rebuild or reconfiguration, you need to bring the system up in a safe mode where nothing is allowed to write to the RAID. You need to test that the files are healthy without writing to the disk. First disconnect the users. Disconnect the network cables. When the system starts to boot – be ready to press escape and interrupt CHKDSK or other offers to repair file system inconsistencies. When booting windows it will try to run CHKDSK and repair the volumes – don't let it. Its better to press reset than to let repairs run for hours. Macs will do the same. Once you are booted into windows, you should run a readonly test of the RAID filesystems. Do this by getting into a command prompt and run "chkdsk X:". Look for error messages. A few errors are not unusual but if it spews a stream of bad files then something is wrong. "CHKDSK /F" will try to write repairs so don't repair. Something may be configured wrong. Stop. Get an expert. Going further at this point could cause irreversible file system damage. For macs run "diskutil verifyvolume" and verify but don't repair.

DO: SAFE INTEGRITY TESTING

A second test is to search the drive for large files that can be tested for integrity. One of the best ways is to search the drive for large ZIP files and then do a zip integrity test. You want to find the most recent zips that you can. If the drive was stale and you test a zip that is 6 months old then it's likely it will pass undamaged because it hasn't moved or changed in 6 months. You want to test zips that are at least megabytes in size.

DON'T: REUSE THE DRIVES TO MAKE A NEW ARRAY

Its a bad idea to reuse the array drives and create a new array and then restore the data from backups. Why? Because what if your backups are no good? You don't want to destroy your only copy of the data and then restore backups and find out that the backup was setup wrong or not working at all. This happens a lot.

DON'T: RESET THE PARITY

Parity means backup – it's a backup of the last time the RAID was running. When you reset the parity then you backup the current state of the file system and you LOSE THE PREVIOUS STATE. So if its configured wrong, and the file system is damaged or corrupt then you lose the chance of going back to a previous version of the file system.

True story: We recently did a RAID recovery for a major healthcare where their major tech support center experts told them that their disk problems would go away if the reset the parity. They spent 4 hours resetting parity and 10 hours watching CHKDSK 'repair' their files. It made a mess.

DON'T: FORCE A DRIVE ONLINE

Explanation: the RAID controller marks each of its disks so that it knows which disk belongs where. If a disk loses this mark then it goes offline – and the raid controller ignores it. So forcing a drive online means that you tell the raid controller to accept this drive as part of the array. Why it's a bad idea: this is risky. Doing this wrong could cause the array to reset parity and cause further damage. Get professional help.

ADVICE ON FINDING A RAID DATA RECOVERY

There are 2 rescuers that you need to keep a careful eye on.

- The IT tech support person. They may not have a lot of experience repairing RAIDs. They often believe that RAIDs are invulnerable, that they can't be damaged because they are smart and redundant right? We have seen them swap drives and swap boards on drives and rebuild and regenerate parity...
- The manufacturers' tech support team. You need to understand the motivations here. The manufacturer views support as an expense. They measure a good support call by how *quickly* it is ended. They are not responsible for your data – after all it says that you should have backed up. So be very wary of their advice. They may tell you to reset the parity, or force a drive online. This can make the RAID array functional, the support ticket closed and your data is all gone. So be careful.

Features

- Free evaluation.
- No recovery, no charge.
- Instant walk-in evaluation.
- Estimate prior to work.
- File list for approval upon completion.
- IBM preferred data recovery vendor.

Capabilities

- Complete drive cleanroom rebuilds – heads and platters.
- Fried board repair.
- Repair corrupt firmware – (Seagate's frequently go dead).
- RAID, VMware, NAS, Linux, MAC experts and custom utilities.T
- Unlock password locked drives.
- Handling encrypted drives – PGP, SafeBoot, Bitlocker etc.
- Forensic certified for forensic analysis and handling.

- The internet is filled with low-cost data recovery programs and companies. We do many RAID recoveries for other data recovery companies. RAID recovery requires expertise and experience to get it right. We wrote our own RAID tools and don't rely on the manufacturers utilities which got you where you are now. You should think twice about a recovery company that wants the whole server and not just the bare drives. This suggests that they are going to try to use the manufacturers' built-in tools to fix your raid. Be careful.

Author's Bio

Wayne Horner is president of Alandata Data Recovery. He graduated from University of California at Irvine in 1982 with a Bachelors degree in Computer Science. His first 'impossible' data recovery was in 1986 when he developed a program to 'undelete' deleted UNIX databases. Since then he has developed many more custom recovery innovations. Data recovery has been Alandata's only business for 30 years and 90% of our work comes from other data recovery companies. We have state of the art tools for diagnosing and repairing disk drives including a clean bench. We replace heads, swap platters, repair corrupt firmware and service area issues and repair boards. We even wrote our own suite of RAID and VMware recovery utilities. Alandata has expertise in all the major filesystems, tape formats, forensics, disk drive repair and recovery.

Alandata Data Recovery – (949)287-3282
"Cleanroom Data Recovery of RAID, VMware, NAS, Linux, Tape, Disk, Forensics"
www.AlanData.com www.AlanDataRecovery.com

F.S.S.C.

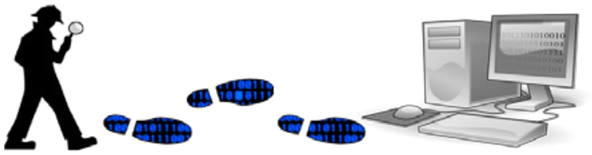
Forensic Security Solutions Co.

A Computer Forensics and Network Security Consulting Co.

- Forensic Imaging & Preservation of Digital Data
- Forensic Analysis & Investigations
- E-Discovery Collections
- Targeted & Multi-User Collections
- Risk & Threat Analysis
- Vulnerability Assessment
- Penetration Testing
- Forensic Wiping of Digital Data Sources (Hard Drives, Thumb Drives, etc.)

Forensic Security Solutions Company is geared toward providing their customers with extraordinary project management and client interfacing that can be utilized for any size matter. Feel free to check us out at www.ForensicSSC.com

F.S.S.C.



Tel: (908) 917-1482 Email: Contact@ForensicSSC.com

www.ForensicSSC.com

INTERVIEW WITH BRIAN GILL, CEO AT GILLWARE, INC.

by Kishore P.V. and Richard C. Leitz Jr.

I'm a computer scientist and entrepreneur. I started writing basic programs when I was in elementary school and am a life-long programmer. Before I founded Gillware I was an engineer at many IT startups and later was an IT consultant. I have also co-founded 3 other start-ups dealing with cloud storage, fusion accelerators and medical isotope production. I live with my wife Kara and our son Charlie in Middleton, WI.

I run one of the best data recovery labs on planet Earth. We've got more than 100,000 engineering man-hours spent recovering data from some of the worse possible scenarios. These folks are some of the highest IQ computer scientists, electrical engineers, mechanical engineers, and physicists to ever grace our little industry. We partner with Dell, Intel and Western Digital just to name-drop a few of our partners that utilize us to help their clients out of a jam now and again. We have an active group of many thousands of computer services and managed service providers worldwide that utilize our affiliate program when they have a client in need.

How do the services offered by your company differ from that offered by other companies?

We always offer a free evaluation and attempt, deliver an electronic listing of validated files before any payment is received and our costs tend to be a little more reasonable due to our significant investment in highly efficient processes. This ensures that unless we can recover what is critical to our client at a price that makes sense for them, they won't pay us anything. If we can't help we literally do not want a penny from our customers, we'll even ship the devices to Gillware on our dime if they are in the United States.

Our engineering staff is world class. We never outsource any type of engineering to any third parties. We are a fully comprehensive lab, capable of recovering data from anything that has had data on it, whether it be 30 year old floppy disks, a crashed helicopter or airplane (we've done both), HDD or modern SSD drives. We recover data from single units, RAID-5/6/1/0 arrays, all the way up to 100+ drive SAN units. We utilize our computer scientists to produce unique software for any file system on the planet and are not reliant on any off the shelf toolkits.

What best practices does your company use when performing data recovery?

All storage equipment is checked in under HD cameras and every aspect of a device is labeled and bar coded to ensure nothing is ever co-mingled. Drives are evaluated by a senior staff member that has gone through extensive training. Temporary repairs are made if necessary, whether they are electrical, mechanical, firmware related. Write-blocking hardware is utilized to intelligently clone to a unique storage device that has been previously zero filled to Department of Defense standards. Data is then logically analyzed to find file system geometry and eventually find and test individual files. Quality assurance procedures are in place to insure customer's data is never co-mingled and always treated as confidential. Recovered data can be delivered however our customers require. If a client wishes we can encrypt the returned data. The number one rule of recovery is "do no harm" and never alter the logical nature of the device.

If you do how does your company handle the chain of custody procedures?

Our standard procedures, forensic or not, always track a case and track who works on a case and track where data lives. So it's simply a matter of providing additional paperwork that the forensic company provides us.

What ways do have for sending recovered data back to the customer? (Hard drive, cloud access)

We'll deliver data however the client needs. Drives, optical, cloud, clones, VMs, whatever they require. Recovering data is significantly more challenging than converting it into whatever the client needs. We had one lady recently actually have our staff print out her word documents onto paper for her because she was done with computers.

Do you have employees that are certified in digital forensics? Possible certifications are from the Digital Forensics Certification Board, GIAC Certified Forensic Analyst through the SANS Institute or the Certified Computer Examiner certificate.

No, not because we lack expertise but because we want that expertise focused on what we do best. Many Gillware staff members are frequently called upon to lecture to forensic examiners about the aspects of various file systems, proper handling of storage devices, etc. We have never pursued outside certification as we don't ever want to be primary investigators in any forensics case. We want our engineers recovering data from failed devices or nightmare logical scenarios. If our engineers are testifying in court they aren't fixing storage devices or rebuilding failed Raid arrays. We are most often utilized for our expertise by forensics companies in broken storage devices due to malicious behavior, then providing the forensic clones so those investigators can do their jobs. Those folks love being in court.

Do you do data recovery for forensic investigations? How do these practices differ when performing a forensic recovery?

Yes. Our standard practices are basically unchanged, with the exception that there's typically additional chain of custody paperwork and sometimes our client does not want us doing any file-level investigation, especially in criminal cases; they just want the devices fixed and cloned. Also, whereas on a non-forensic case we will typically hold the clients data for a couple weeks we typically are instructed not to do this in forensic cases. In forensic cases we often are requested to provide hard drive image files with an MD5 hash of that image file which is transmitted separately at our client's request.

In a forensic recovery do you perform extra steps versus a typical recovery?

The only extra steps that are common are providing whatever hashing the client requires (like MD5). If the case is a logical tampering we may provide additional assistance with writing customized file carving utilities, additional analysis of volume snapshots if Windows Volume Shadow Copy Service was enabled, or more detailed analysis of historical file system shrapnel outside the active file system.

Take the case of a typical forensic scenario. What are the methods by which we can know if data(evidence) has been tampered?

Well of course you will always take an immediate baseline clone and hash it and hash any individual files if necessary so you can mathematically prove they have been altered moving forwards. After that baseline if you are trying to prove that data was maliciously tampered before it got here, there are a variety of ways, typically involving file system meta-data analysis, listing transactions in the file system journal, comparing the volume at previous points in time through snapshots taken by Windows Volume Shadow Copy Service, detailed analysis of the unused area of the disks, etc.

Are there any particular instances of data recovery that initially gave you trouble, but you managed to perform a full recovery at the end of the day?

Every day that ends in "Y". Storage manufacturers pump out hundreds of new models every year. These devices are state of the art and are not built to be serviceable. They come with no manual that would detail repair or troubleshooting. Some manufacturers provide very little if any support for our industry.

It is a constant investment in R&D to be always reverse engineer these devices, *before they fail*. You need to understand how they work, to understand how to troubleshoot, to ultimately understand how to eventually repair and recover. At Gillware we tend to see the worst-of-the-worst cases, as about 20% of our revenue actually comes directly from *other data recovery labs* that are looking for a second opinion on a case they were unable to recover. So the short answer is, literally every single day (other than Christmas) we have this scenario occur.

What do you think are the common mistakes users do when trying to recover data from their devices?

I could talk for an hour on this subject. Some big ones: If a drive isn't detecting in the BIOS within 10 seconds it's likely troubled and may not warrant a software attempt. If it is not detecting in the BIOS with the correct S/N or capacity it is not eligible for a software recovery attempt. If it is clicking it needs to be turned off so as to not damage the platters. Modern drives should never, ever be put into an oven or freezer or tapped with hammers or other various caveman tactics that the internet tends to promote.

Never re-initialize a drive or format a drive, never install data recovery software on the device you are trying to recover, never restore data to the same device you are recovering. If a piece of data recovery software is telling you it has an estimated 20+ hours remaining turn it off and do not let it run un-attended overnight. Modern HDD drives are typically fully calibrated and that unique information lives on the printed circuitry so the days of simply swapping the PCB with a PCB from a similar drive are over. Never throw any original equipment (like a PCB or external USB chassis) away as it may be unique or performing unique encryption/decryption/adaptive-calibration roles. Never open a drive outside a cleanroom. Amateur data recovery horror stories are a daily occurrence here.

The short story is this: Is the data critical for a professional, litigation, or personal reasons? Hire an expert and don't do anything that's going to cost you the data, permanently. Most consumers don't understand that they aren't in a "zero-sum game". The more you mess with it trying to avoid utilizing folks with actual expertise, the worse it's going to get. Failed data recovery attempts will often cause harm.

Do you shut off the SMART logging when performing a recovery?

Ideally. Some of the drives that we see are very unstable post repair and we use curious techniques to get them to "boot" so it is not always possible. It would be a good thing if you could do it as you don't want a storage device logging a ton of events trying to help the end user predict failure. The device has already failed and that is why it's here. Stop logging, drive! But, if you attempt to mess with the core firmware of the drive when it's in an unstable R/W state you can sometimes brick a drive, just ask Seagate. These things are judgment calls based on available toolsets and expertise.

Do you shut off drives defect auto reallocation updates to the grown defect list?

This is the same train of thought as the SMART table manipulation/disabling. Can you successfully do it in RAM? Do it. If a drive is very defective it would be a good idea as again the point of a Glist is preserving the lifetime of a drive. It is at Gillware so it's already dead, I don't want the data manipulated on the drive; I just want to read it off. As long as you can do so safely without risk of killing the HDD OS we do it.

What methods do you suggest for removing data from storage devices completely?

Full drive encryption is a great way to do this; this is how the SSD manufacturers do it. Need all the data gone? Kill the key, 5 seconds later all the data is gone. Without that? 0 or F fill the device, even one pass will typically be sufficient on modern equipment, but there are a billion software suites out there to do this.

Can you tell us how chip level recovery can be carried out on any flash storages(USB/SD) in case of heavy physical damage?

Not only physical damage, solid-state storage devices fail due to firmware corruption just like their spinning magnetic brethren. The major difference between the two types of storage, from a recovery standpoint, is direct access to the underlying medium. Most flash storage devices are comprised of discrete components, including one or more NAND flash memory chips having an industry-standard interface. The ability to read the raw storage media independent of the device that originally wrote it is incredibly powerful and frees us from burden having to fix the drive in order to recover data.

We wrote a whitepaper on these techniques a few years back. *http://gillware.com/docs/SSD_whitepaper.pdf*.

Chip-level recoveries, naturally, begin with the removal of each NAND flash memory chip. Specialized hardware is then used to dump the contents of each chip a file. They layout of these "dumps", however, is very different from what you'd see looking at the raw device in a sector editor. Interspersed with user data are things like error-correcting codes (ECC) and other pieces of information used internally by the device itself. The first obstacle to recovery is separating user data from this device metadata. Once the general layout is established, each sector's ECC must be applied to correct any bit errors that would otherwise lead to file corruption. From there, we must determine how the available storage was spread across each memory chip. For reasons including both performance and endurance, there is huge disconnect from where the host PC thinks its writing data and where it actually gets stored. The mechanics of this logical-to-physical translation can vary wildly from device to device and reassembling the disk image as it would be seen by the host PC is the most difficult part of flash drive recovery.

Can we trust only data encryption as a sound security measure for our storage devices?

I certainly hope so as many billions of dollars are made encryption software/hardware and hundreds of thousands of organizations are depending on it. Strong encryption must be coupled with strong passwords and strong security on an organization's key-store.

If data is encrypted with something like Windows NTFS EFS, the answer is NO. Is the data encrypted? Yes. Can a savvy hacker determine the user it was encrypted with? Easily. Can a savvy hacker brute-force the user's password over a period of 2-4 weeks if it isn't strong? Every time. So bye bye encryption, hello data. Some forms of encryption just plain suck. They are a deterrent but not a protectant.

The truth is if your data is unique encrypted with a 256 bit key and whatever mechanism is protecting that key is equally strong, it would take most of the computing power on Earth millennia to hack. So there is the key to the whole thing, will quantum computing ever come around with infinite amounts of computing power? Probably not, I hope so though. So if you are a hacker faced with big encryption what do you do? You go after the mechanism protecting those 256 bits. We do this type of thing occasionally.

Do you recommend the readers use open source or commercial software for data recovery,and what software do you use?

Full disclosure: We don't use third party data recovery software. We write data recovery software. We're armed with some of the best computer scientists in the world. We distribute a dumbed-down version of our internal platform in our consumer software *RecoverBot* that we distribute through Dell computer. RecoverBot does a lot of what our in-lab software does, but with a front-end where an end user just pushes the "Go Button". You don't need to know anything about file systems or boot sectors to get really decent results; it was built for non-computer savvy OEM customers. As your readers are likely highly intelligent forensic folks, they should probably be using something like EnCase or FTK.

What are the latest technological advancements that are making data recovery an easy/difficult task?

Everything is getting harder, trending towards encryption, virtualization and various other levels of data-indirection. So this is adding additional steps and in some cases preventing recovery entirely if the organization has lost something like unique encryption keys. SSDs fail and the latest generations of SSD encrypt data written to the individual memory chips to thwart unauthorized access. This, of course, also thwarts conventional recovery techniques. Subsequently recoveries from these devices will typically require manufacturer support. Hard Drives are getting higher capacity and more complicated, less and less compatible with their internal parts. HDDs have unique calibration-adaptives stored on their platters and their PCBs which makes recoveries harder.

Virtualization makes recovery really insane when things go logically south. Sometimes we're dealing with more than 5 levels of indirection between us and the data. Troubled hard drives, RAID cards with their unique striping patterns, rotations, offset. Physical volume groups, logical volume groups, innate or proprietary file systems housing highly fragmented virtual files of yet another set of logical partitions that actually finally render the user's data. Now have some poor IT guy push the wrong button and accidentally re-initialize the whole physical volume and restore from a backup only to realize the backup hasn't been working for 6 months and he just annihilated everything. Sound fun trying to untwist all that? Apply here at Gillware. It's a rare engineer that has the extreme IQ and temperament to deal with this stuff.

TrustSphere

Global
Reputation

TrustCloud

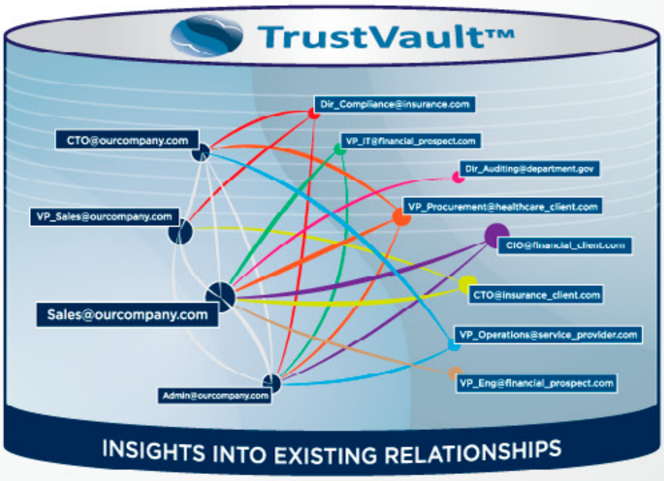
Industry's Most Comprehensive Real Time Dynamic Reputation List

+

Local
Relationships

TrustVault™

Restoring Security, Integrity & Reliability to Messaging Systems


TrustVault™
Dir_Compliance@insurance.com
CTO@ourcompany.com
VP_IT@financial_prospect.com
Dir_Auditing@department.gov
VP_Procurement@healthcare_client.com
VP_Sales@ourcompany.com
CIO@financial_client.com
CTO@insurance_client.com
Sales@ourcompany.com
VP_Operations@service_provider.com
VP_Eng@financial_prospect.com
Admin@ourcompany.com
INSIGHTS INTO EXISTING RELATIONSHIPS

TrustSphere
Tel: +65 6536 5203
Fax: +65 6536 5463
www.TrustSphere.com

3 Phillip Street
#13- 03 Commerce Point
Singapore 048693

FORENSICS EUROPE EXPO

24 – 25 April 2013
Olympia, London
ForensicsEuropeExpo.com

The Premier International Forensics Event for Police, Military, Intelligence Agencies, Lawyers, Corporate Forensic Analysts, Laboratories, Government Bodies and Agencies together with leading suppliers, services, equipment and practitioners from across the world.

Conferences – Workshops – Training – Networking – Exhibition

REGISTER FOR FREE ENTRY TODAY

www.ForensicsEuropeExpo.com/digital

Co-located with COUNTER TERROR EXPO

Sponsored by LGC Forensics

In Collaboration with The Forensic Science Society

Organised in Partnership with the Investigator

Organised by CLARION EVENTS

Technology is a double sided sword.
Internet makes you naked online!
Get Secured & Get Certified!

Welcome to the world of Certified Ethical Cracker
with Hands-on practical sessions.

CERTIFIED
ETHICAL
CRACKER

An Advance **Information Security** Course

For more detais, visit:
http://www.infysec.com/training/courses/certified-ethical-cracker

infySEC UK :

145-157, St.John Street,
London, EC1V 4PW
England, UK

Phone: +44-7405190001

infySEC India :

#37/45, P.H Road,
Arumbakkam,
Chennai- 600106
TamilNadu, INDIA

Phone: +91-44-42611142,43

infySEC
Demystifying Innovations

www.infysec.com

enquiry@infysec.com