**CIS-387 - Digital Forensics**

**Personal Notes – Demetrius Johnson**

**8-29-2022**

**With Professor Dr. Jinhua Guo**

# Lecture 1 (8-29-22) – introduction to Digital Forensics

* Course created with the help of the Criminal Justice Department since Digital Forensics is interdisciplinary
* No programming assignments: we will mainly do labs and analysis using software tools.
* The free open source software tools for digital analysis work about just as good as the licensed products; so we will just download and install these free tools and do analysis on out personal computers.
  + Then we can also always have the analysis tools on our own PC.
* Old version of the textbook is fine.
* There will be 10 labs.
  + Mostly straight forward
* There is actually a lot of fraud in the cryptocurrency world – so there is a lot of interesting forensics that can be conducted.
* There are many departments in government agencies that have a high demand for Digital Forensic analysis skillsets – so there is plenty of job opportunities for this.
* We will focus on data stored for this class; but we can study and investigate data transmitted.
* For example, we can download the RAM memory before a device is shutoff, because this is often where passwords are stored (they are often not stored on the hard disk space).
  + Often, applications need to do password checks for security periodically while apps run – but vulnerability is that it must store it in memory.
* Encrypted files are very hard to analyze.
  + No easy techniques to decrypt a file.
  + But this is the whole point of encryption – best way to handle it is with very strong computational power (i.e. quantum computers).
* Slack space – space allocated to a file but not used – still could contain old “deleted” (deallocated) memory information.
  + Free space – may also have old “deleted” (deallocated) memory that we can make use of by checking the contents – this is why we do bit stream copy (the stream of bits stored on the hard disk or memory exactly – bit for bit).
  + This is because computers don’t want to waste time deallocating and overwriting (takes a lot more time and makes life of a drive less because you have to do more writing to it).
* You often need to be familiar with many platforms (window, mac, linux, ios, android) as a digital forensic investigator.
  + Very difficult field; also you need to stay up to date with latest technology.
* At a crime scene, if device is on, you can copy its running memory.
  + Also get disk space bit stream copy.
  + Check current connections (network configuration) (IP addresses, ports, etc.).
  + Check all of the running processes; logs; etc.
* Often you need to be familiar with accounting practices in order to conduct accounting forensics as fraud often occurs through finance and inventory accounting.
* Onion ring networks make it hard to trace back to original network/device.
* After taking an image of a device and its memory, you do not touch the physical device anymore.
* Memory dump = copy volatile memory.
* Drive imaging = copy hard disk with a bit stream copy; does not pay attention to EOF = a logical end of memory space denoting end of the (root) file.
* FTK Imager is a free bitstream copy tool.
* Immediately hash a hard drive bitstream to make a unique identity so that we know that the file has not been altered.
* Sha1 command = secure hash 1 algorithm function
* Sha1(m) 🡪 checksum
  + If sha1(m’) 🡪checksum; then you know m and m’ are the same contents.
  + We can compute a checksum (hash output) using MD5 hash algorithm.
  + so for example: if you do a bitstream copy of a hard drive, and compute the checksum; if it has the same checksum value as the original evidence, then you know you have successfully create an exact copy of the evidence hard drive.
  + Usually hash output is a 160bit value (SHA-1.
* Linux hash command = Sha256 and Sha384 (for 256 bit or 384-bit hash value)
* Often checksum output as a hexadecimal value.
* For crypto currency mining: we want to make the first (leading) x-bits (for example, leading 30 bits as all 0’s; so you need to add data to the block chain, then compute checksum and hope that you get x-leading zeros; if you do , then you have successfully mined a cryptocurrency and now the block chain can be updated!).
  + Checksum/hash algorithms are designed to be “random”; so only way to find some checksum that meets some given constraints/requirements, you need to use brute force = a lot of computational power).
  + Also hash algorithms are designed to be irreversible (one-way encryption/hash), so that you cannot go from a desired output and reverse engineer it to get the necessary input needed to mine a currency = get the desired checksum from a hash input value (message).
* Also you can hash smaller files, so that if you can match a file (such as malware) checksum to the same checksum of a file on some other device, to prove that file also exists on other devices...or perhaps originated on another device or built on that device, etc.

# Lecture 2 (8-31-22) – introduction to Digital Forensics (continued)

* Continuing basic, high-level overview of digital forensics, next week we will begin the technical discussion.
* Collision = when two different messages map to the same hash (hash function gives same output).
* Hash algorithms for cryptography need to be collision-resistant (actually, needs to be collision-free in practice).
* A computer scientist was able to take MD5 hash function (128 bits) and find messages that produced the same output (collision).
* We can use hashing to make sure that the original message was sent and not modified along the way to the destination.
* MD5 is still acceptable for digital forensics investigations, but not for security systems in the modern day.
  + We need to use SHA-1 (160 bits), SHA-2 (256 or 512 bits), or SHA-3 for the security of a digital system.
* Analysis for digital forensics
  + Start analysis by looking at all logical partitions of a drive (sometimes, people hide other partitions to make it appear like only 1 partition of the drive exists).
  + Find most recently accessed, created, and modified files; create a timeline based on the three fields in every file: last accessed, date modified, date created.
  + Recover deleted files (if the deleted block has not already been overwritten; thus you may not be able to recover a file or only fragments of deleted files).
  + Do a hash analysis: we hash every file on the drive; then we can do a hash table lookup to see if some file is on the drive we are investigating.
    - This is better and faster than a bit-by-bit search.
  + Do keyword search: do regular expression searches (phone number, sentence, words, URL, etc.)
  + Do file signature analysis: i.e., you can check file extensions (but people can easily change this), so better to check first few bytes of a file, because that is especially what defines the file type (and it is much harder to manipulate the binary).
  + Glean OS registers (check all devices connected to the device you are investigating; there is a register usually stored on the drive that gives a log of all connected devices past and present and tells how those devices interacted with the host device).
  + Digital forensics tools are usually a few years behind, so often you need to be prepared to an analysis more manually.
* Reporting
  + Need to support or refute some hypothesis
* Challenging aspects of digital evidence
  + Hard to analyze large data sets
  + When you convert bits into useful information, each layer up the abstraction ladder may introduce errors as you try to put fragments together.
  + For digital crimes, you can rely totally on digital evidence; but for non-digital accusation, digital evidence can only be used as circumstantial.
  + Digital evidence = easy to manipulate and erase
    - Can be hard to prove authenticity (multimedia forensics)
  + Every file has 3 times: creation time, last accessed, and last modified; if those dates are changed upon making a digital copy, then the evidence can be much harder to analyze and rendered useless in court.
* Evidence dynamics and error introduction
  + Rootkit: some malware and different tools used by criminals that allows all traces of the malware to be untraceable by deleting the evidence upon a regular shutdown.
* Following cyber trail
  + We can do IP Traceback (a traceroute) to trace the IP of a perpetrator in order to potentially find their machine and their location and other information
    - But sometimes attacks are launched from different places where US does not have jurisdiction, or they launch attack using TOR tool which uses a ring of devices and IP addresses so that it is hard to trace it to one device.
  + TOR uses a ring: many layers of devices that the original data goes through to make it hard to trace that source from where the data originates.
  + Cloud makes it hard to find source of data; so we have such a thing as cloud forensics.
  + We can actually serve search warrants to cloud companies and search a user’s data without their knowledge (cloud company must turn over users data to government agency to be analyzed and investigated).
* Digital Forensics Profession (Discipline) and a History
  + Digital forensics is unique and different to data recovery
    - Data recovery is more about accidental or mistaken power and data lost of a system.
  + Digital forensics focuses on all the methods needed for mistaken and purposefully data loss
  + Also for data recovery, we normally know exactly what we are looking for.
  + For digital forensics, we often don’t know if there is evidence on the investigating drive; it is a lot less straight forward.
  + Often three teams:
    - Team 1: Do threat and vulnerability assessments
    - Team 2: Check for intrusions and mitigate them quickly (using tools)
    - Team 3: Conduct digital investigations
  + Relatively new field
    - 1980s, PC hit market for average person, so cyber crime became a big problem.
  + Cliff Stoll: caught Markus Hess who was selling US data from US machines that he hacked into to the KGB (Soviets).
    - He wrote a best selling book on how he caught the perpetrator after first feeding him false information after he realized system was compromised.
  + New, big Area of interest for attackers: cryptocurrency – because there is a very large financial incentive.
  + Companies have developed (and also hobbyist) digital forensic investigation tools to keep up with the latest tech and to be approved of and used in court cases.
  + For 4th amendment: not clear if you need a separate search warrant for digital evidence (“might not be necessary”); so if it is not specified or not clear (which today the laws are still developing) then searches may be conducted without a warrant.
  + If you convince a judge with sworn statements and or facts (affidavit) then the judge may grant search warrant.
  + Problem: existing laws cannot keep up with rate of technological change.
    - Temporary solution: use case law (similar cases/situation) and apply it to current situation.
  + If procedure is not followed, digital evidence (like any evidence) can be thrown out.
  + There are many specializations in digital forensics.
  + For this class we will mainly focus on technical aspect, but we need to know about some of the legal aspects.
  + Companies mostly try to settle investigations in-house due to expensive court fees.
  + Https uses encryption; so for our school, we can see traffic in the network but for users we cannot see what the actual contents are.

# Lab 1 Review

* Make sure when making shared folder that user profile is a part of the vboxsf (shared folder) group with rwx access (read write execute) to the shared folder.
  + If not a part of the group, need to do the command to add user to that group.
* Access Control for a file or folder:
  + OWNER (R-W-X)
  + GROUP (R-W-X)
  + OTHERS (- - -) //(SET TO: NO READ WRITE OR EXECUTE PERMISSIONS)
* By default, output for commands go to console window: but you can also use an option to redirect output from console window (technically a file in the memory buffer) to (another) file or program.
  + Use > or <
  + Also you can use pipe command | to redirect output to be the input for next command (program/function)
  + Ex: ps -eaf | more | grep
  + The pipe | is implemented as an anonymous file (so the file serves as a way to pass data from one function/process to another).
* Manual for any command for Linux: man “command”
  + Ex: man ps
* sudo command= run at the root (and with root user access).
* Use “strings” command to search the file to which you sent memory dump to find passwords.
  + ^forensics indicates that beginning of a string should “forensics”
* Most processes that use usernames and passwords leave them in main memory as plain text (usually, developers apparently do not have security in mind in terms of a digital forensics tool such as a memory dump application).
* Goal of this lab is just to get us familiar with basics of linux.
* If you set a port to promiscuous (using linux command) then the port will listen to and obtain all LAN traffic even if it is not addressed to that port’s MAC address.
  + Command to set mode is: “promisc”
  + Hackers can use this to launch more attacks once one machine in a network is compromised.
  + So, we want to check ports for this mode bit as on, then we can know that probably something very suspicious is going on with the machine.
* lsof -i command: show all transport layer connections
* you can also do the open file table command to show all files loaded in main memory
* also lsof command can show deleted files but that have not been overwritten.