Business continuity plan: overview

Business

Data Sources in Digital Forensics March 17, 2013 Joana Achiampong CSEC 650 Introduction Four sources of data that stand out for forensic investigators in most criminal investigations are files, operating systems, routers and network traffic, and social network activity. Each data source presents a variety of opportunities and challenges for investigators, meaning that the more reliable data collection and analysis activity typically involves examination of a variety of sources.

Digital forensics must cover the four basic phases of activity, which include: data collection, which describes the identification and acquisition of relevant data; data examination, which includes the processing of data through the use of automated and manual tools; analysis, which describes the evaluation and categorization of examined data into coherent groups, such as their usefulness in a court proceeding; and reporting, in which the results of analysis are described with careful attention paid to recommendations (Marcella & Menendez, 2009).

The viability of each data source to an investigation must be evaluated based on how they can contribute to each phase. For example, the ability of routers and switches as a data source to help investigators might be effective in one area, but not in the other three. An examination of router activity might yield a surfeit of observable data that fails to provide diverse analytical tools that cannot be relied upon in a forensic setting. Another example is network traffic, which may yield a large amount of data that is unreliable or has a high degree of volatility (Garfinkel, 2010).

Time is often essential for forensic investigators, and it is often important to know in advance the dynamics of each data source. This helps investigators

avoid wasted time, or spending time analyzing data that may of minimal help in a forensic setting. For these reasons, it is important to critically assess the pros and cons of each data source for their ability to provide contributions. A valid assessment of each data source should be made based on consistent factors such as costs, data sensitivity, and time investment.

The overall costs of each data source depend on the equipment that will be required to collect and analyze data without corruption. Costs also refer to the training and labor required during the course of the collection and analysis, which may be higher for uncommon sources that require a unique process and chain of command pattern. Data sensitivity is critical is a forensic tool, but may be more questionable depending on the source. For example, network activity can provide a wealth of information depending on the device and setting upon which data is moved.

However, a networkenvironmentwith many devices and multiple configurations may provide unreliable data that cannot be recognized in court proceedings. In addition, chain-of-command issues regarding the contribution of outside network analysts could compromise a source that would be otherwise valid. These issues have to be considered in any data source assessment. Data Files The most common data sources in a digital forensic examination are current and deleted files. Most forensic investigators in most data retrieval environments begin with an examination of the various media store on the hard drive of a computer, network, or mobile device.

The variety of types of stored data in current and deleted files, in addition to partitioned packet files and the slack space of a device’s memory, can be

massive and diverse. A typical first step in data retrieval is to shut down a system and create a data grab or forensic duplicate upon which collection and analysis can be made. This ensures the integrity of the original data, while allowing investigators the ability to manipulate data however they see fit. However, this process alone creates challenges for forensic investigators, including an inability to capture live system data.

This might prevent investigators from catching a perpetrator in the act of altering or adding data to a device or system. One of the primary benefits of files as a data source is the ability to separate and analyze the types of files, which creates a specific signature based on the content and user (Marcella & Menendez, 2008). Data can be pulled from deleted files, slack space on a system’s hard drive, or free space, all of which provides information that can be useful to investigators.

The directory location and allocation type for each file informs the data that has been collected, including a time stamp and whether tools have been used to hide the data. Each of these characteristics provides investigators easy-to-access information about a system. In addition, there are a variety of hardware tools that can be used to access data. Thistechnologyis fairly common, meaning that associated costs tend to be minimal when retrieving data from files (Purita, 2006). File examination can yield a variety of types of suspicious activity that tend to be helpful for investigators.

One example is the presence of hidden evidence on file systems. This type of data can be hidden in deleted file spaces, slack spaces, and bad clusters. File space is marked as deleted when it is removed from an active directory. This data will continue to exist within a cluster of a hard disk can be identified

and accessed by creating a file in Hex format and transferring the copied data. Data can also be hidden in many others ways, including by removing partitions that are created between data and by leveraging the slack space that exists between files.

Attempts by users to hide data using these methods are quickly identifiable by investigators, who can then restore the data using a variety of inexpensive and efficient methods. For example, matching RAM slack to file slack identifies the size of a file and makes it easier to identify and retrieve (Sindhu & Meshram, 2012). This type of retrieval inherently emphasizes the importance of data integrity. This type of integrity is important in any forensic environment, and compromised data is usually rendered instantly unusable. The many opportunities for data retrieved from file space to be compromised are a drawback to this data source.

For example, data retrieval using bit stream imaging provides a real-time copy onto a disk or similar medium. However, this can be compromised based on the fact that re-imagining of data is constantly changing during re-writing. Investigators will typically choose the type of data copy system based on what they are looking for. However, changes to data can occur if the appropriate safeguards are not taken. Write-blockers are often used to prevent an imaging process from providing data that has been compromised by writing to that media. Sindhu and Meshram 2012) stated that computing a message digest will create a verification of the copied data based on a comparison to the original. A message digest is an algorithm that takes input data and produces an output digest. This comparison helps investigators ensure the integrity of data in many cases. There are additional pitfalls when

it comes to using files as data sources. Users have different resources for eliminating or hindering data collection. One example is overwriting content by replacing it with constant values. This type of wiping function can be performed by a variety of utilities.

Users can also demagnetize a hard drive to physically destroy the content stored there. Using files as a data source in this case will require a complex operation requiring different tools. Users can also purposefully misname files – for example, giving them . jpg extensions when they are not image content files – in order to confuse investigators. Investigators have to be familiar with strategies for circumventing these pitfalls, such as maintaining an up-to-date forensic toolkit and remaining committed to maintaining data integrity.

In the end, files are very highly relied upon by investigators and are a strong source forensic data. However, investigators must be experienced and have the appropriate tools to ensure the viability of collected data. Operating Systems Generally speaking, the data that can be collected from Operating Systems (OS) is more diverse and rich than file systems data, and has greater potential to uncover application-specific events or vital volatile data specific to a network operation (Sindhu, Tribathi & Meshram, 2012).

However, OS data mining can be more difficult and challenging, and often requires investigators to make quick decisions based on the type of data they are seeking. OS data mining is more case specific, in part because the retrieval of data is frequently connected to network configurations. Collecting volatile data can only occur from a live system that has not been shut down or rebooted (Marcella & Menendez, 2008). Additional activity that occurs over an individual network session is very likely to compromise the

OS data. For this reason, investigators have to be prepared and aware of what they are looking for.

Time is of the essence in this case, and it is important to decide quickly whether or not the OS data should be preserved or if the system should be shut down. Keeping a system running during data extraction can also compromise data files. This also leaves data vulnerable to malware that has been installed by a user with bad intentions, determined to undermine the operations of investigators. The types of data that can be retrieved from the OS include network connections, network configurations, running processes, open files, and login sessions.

In addition, the entire contents of the memory can be retrieved from the OS history, usually with little or no alteration of data when the footprint of retrieval activity is minimized. The order in which this data is collected typically runs in a standard succession, with network connections, login sessions, and memory collection sitting at the top of the list or priorities. These sources are more important because they tend to change over time. For example, network connections tend to time out and login sessions can change as users log in or out.

Network configurations and the files that are open in a system are less time-sensitive and fall further down the list of priorities for investigators. The forensic toolkit must be diverse to ensure that data retrieval is achieved with minimal alteration (Bui, Enyeart & Luong, 2003). In addition, the message digest of each tool should be documented, along with licensing and version information, and command logs. This careful documentation protects users from sudden loss of data or other disturbances during data retrieval.

In addition, a number of accessibility issues can be implemented by users, including the placement of screen saver passwords, key remapping and log disabling features, all of which can disrupt the work by investigators, either providing unworkable obstacles or time-consuming hurdles that make complete transfer impossible. Ultimately, the use of OS as a data source is a case-by-case tool dependent on the availability of other sources and the specific needs and tools of investigators. Routers and Network Traffic

Among network configuration data sources, router activity and network sourcing has the potential to provide the most specific amount of incriminating activity for forensic use. Forensic equipment should have time stamping capabilities activated to provide an accurate time signature of network interaction between an end-user and a router or switch (Schwartz, 2011). Importantly, firewalls and routers that are tied to a network often provide network address translation which can offer additional information by clarifying configuration or additional IP addresses on a network (Huston, 2004).

There are a number of tools available to people seeking an analysis of network activity, including packet sniffers and intrusion detection systems (Marcella & Menendez, 2008). These tools help investigators examine all packets for suspicious IP addresses and special events that have occurred across a network. This data is usually recorded and analyzed so that investigators can compare unusual events to evaluate network weaknesses and special interests of would-be attackers.

This is of great interests to security agents determined to identify and stop potential network intrusions. A number of technical, procedural, legal and

ethical issues exist when examining and analyzing network data. It is imperative that investigators be sure to avoid disconnected from a network or rebooting a system during data retrieval. They should also rely on live data and persistent information. Finally, it is important to avoid running configuration commands that could corrupt a network or its activity (Gast, 2010).

Issues such as storage of large amounts of data over a highly trafficked network and proper placement of a decryption device along a network can impact how data is available and whether or not it maintains integrity. It is also important to consider the ethical and legal issues of data retrieval along a network when it involves sensitive data, such as financial records and personal information like passwords. In many cases, ethical issues can be circumvented with careful documentation and the publication of organizational policies and procedures that are strictly followed.

However, these are all issues that must be considered in the analysis of network trafficking as a data source. Social Network Activity The sheer volume of social network activity – such as that on Facebook, Twitter, and Instragram – makes examining it as a data source great potential as a forensic tool. To this point, the little available research on social network data has failed to come up with a comprehensive framework or set of standards for investigators. Social network tools across mobile platforms invariably have geolocation services.

However, the use of these as a data source has been questioned from ethical and legal perspectives (Humaid, Yousif, & Said, 2011). Thecommunicationlayer ofsocial mediaapplications on mobile devices can

yield rich data, such as a browser cache and packet activity. Packet sniffing can expose unencrypted wifi use and third party intrusion across a social network. However, these tools are highly limited when they are restricted to social network activity. The best tools may be the ability to create a social footprint, which includes all friend activity, posted pictures and videos, communication habits, and periods of activity.

For most people, this information is only available on social network websites and is not stored on a user's hard drive. A certain climate of permissibility tends to apply to social network use, in which users are prone to making data available online that they would not otherwise expose. All of this strengthens the use of social networks as a data source. The greatest pitfall to social network activity is the malleability of the material. Users frequently change their habits, including the times of the day and the users with whom they connect.

Cumulative social network data can be used to create a graph of all activity across a variety of factors, including time, space, usage, and devices (Mulazzani, Huber, & Weippl). But this is a rapidly changing field. There is little doubt that the cloud computing data storage and continued growth of social networks will change this field quickly, which could quickly undermine past data that has been retrieved. Potential Usefulness in Specific Events The usefulness of a data source is strictly tied to the event it is intended to investigate.

It is imperative that investigators are clear on theirgoalsprior to selecting a source to retrieve and analyze data from. For example, a network intrusion would be best tackled with an examination of network traffic, followed by

social network analysis, Operating Systems, and data file systems. Network analysis is less prone to attacking strategies that can compromise file and OS data. It can observe network traffic to find anomalous entities and their entry point within a network. It can also identify source and destination data by data recovery and access to routers r other network access points (Aquilina, Casey & Malin, 2008). This is critical information for network intrusion investigations. Operating Systems enable access to volatile data, but this is limited by single-time use and data integrity issues. Most OS examinations look at network connections first, which is often another way of accessing the same data. File storage and social network analysis tend to offer peripheral views of the same material. Operating systems are the most helpful data source in malware installation investigation, followed by network traffic, data files, and social network activity.

Examination of volatile data offers a range of data, including network connections and login sessions, which are primary tools for finding the source of malware installation (Aquilina, Casey & Malin, 2008). Maintaining the integrity of data through quick retrieval and minimal footprints helps ensure its usefulness. At the same time, monitoring network traffic in a pro-active manner is often the surest way of pinpointing time signatures and matching them with network activity (Marcella & Menendez, 2008). The best data sources for identifying insider file deletion are data files, network traffic, social network activity and OS.

Each source offers benefits for this type of investigation, but data file collection and analysis yields bad clusters and slack space, both of which pinpoint the likelihood of deleted files. Recovery can begin from this point.

Network activity and OS data retrieval can lead investigators to unusual login attempts and anomalous activity in order to pinpoint the location of deleted files along a network. At the same time, social network examination can help investigators understand reasons for deleted files and even learn more about the habits and lifestyle of a likely perpetrator.

In the end, a collection of each of these sources provides a rich, revealing glimpse at deleted file activity. Conclusion Network traffic, data files, operating systems, and social network activity are four common data sources in digital forensic. Each provides a unique opportunity and set of risks for investigators, and the source should be chosen based on clear objectives and awareness of all circumstances. In many cases, the best choice is a combination of sources to provide multiple opportunities to arrive at the relevant evidence.

Another factor is whether the data search is reactive or pro-active, with network traffic often providing the best source of evidence in a pro-active, forward-thinking environment. The variable of time must also be considered, specifically withrespectto how investigators approach volatile data. Each of these issues must be considered when evaluating data sources. References Aquilina, J. , Casey, E. & Malin, C. (2008). Malware forensics: Investigating and Analyzing Malicious Code. Burlington, MA: Syngress Publishing. Bui, S. , Enyeart, M. & Luong, J. (2003, May). Issues in Computer Forensics. Retrieved ttp://www. cse. scu. edu/~jholliday/COEN150sp03/projects/Forensic %20Investiga tion. pdf Garfinkel, S. (2010). Digital forensics research: The next 10 years. Digital Investigation, 7. 64-73. Gast, T. (2010). Forensic data handling. The Business Forum. Retrieved from http://www. bizforum.

org/whitepapers/cybertrust-1. htm Humaid, H. , Yousif, A. & Said, H. (2011, December). Smart phones forensics and social networks. IEEE Multidisciplinary EngineeringEducationMagazine, 6(4). 7-14. Huston, G. (2004, September). Anatomy: A look inside network address translators. The Internet Protocol Journal, 7(3).

Retrieved from http://www. cisco. com/web/about/ac123/ac147/archived_issues/ipj_7- 3/anatomy. html Marcella, A. & Menendez, D. (2008). Cyber Forensics: A Field Manual for Collecting, Examining, and Preserving Data. Boca Raton, FL: Auerbach Publications. Mulazzani, M. , Huber, M. & Weippl, E. (n. d.). Social network forensics: Tapping the data pool of social networks. SBA-Research. Retrieved from http://www. sba- research. org/wp-content/uploads/publications/socialForensics_preprint. pdf Purita, R. (2006). Computer Forensics: A valuable audit tool. Internal Auditor. Retrieved from http://www. theiia. rg/intAuditor/itaudit/archives/2006/september/computer- forensics-a-valuable-audit-tool-1/ Schwartz, M. (2011, December). How digital forensics detects insider theft. InformationWeek Security. Retrieved from http://www. informationweek. com/security/management/how-digital-forensics- detects-insider-t/232300409 Sindhu, K. & Meshram, B. (2012). A digital forensic tool for cyber crime data mining. EngineeringScience and Technology: An International Journal, 2(1). 117-123. Sindhu, K. , Tripathi, S. & Meshram, B. (2012). Digital forensic investigation on file system and database tampering. IOSR Journal of Engineering, 2(2). 214-221.