Digital Investigation 10 (2013) 138–147

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

Honing digital forensic processes

Eoghan Casey*, Gary Katz, Joe Lewthwaite

Defense Cyber Crime Center (DC3), 911Elkridge Landing Rd., Linthicum, MD 21090, USA

ARTICLE INFO

Article history:
Received 26 April 2013
Received in revised form 30 June 2013
Accepted 10 July 2013

Keywords:
Digital forensic process
Digital investigation decision support
File system forensics
Malware forensics
Network forensics

ABSTRACT

The number of forensic examinations being performed by digital forensic laboratories is rising, and the amount of data received for each examination is increasing significantly. At the same time, because forensic investigations are results oriented, the demand for timely results has remained steady, and in some instances has increased. In order to keep up with these growing demands, digital forensic laboratories are being compelled to rethink the overall forensic process. This work dismantles the barriers between steps in prior digital investigation process models and concentrates on supporting key decision points. In addition to increasing efficiency of forensic processes, one of the primary goals of these efforts is to enhance the comprehensiveness and investigative usefulness of forensic results. The purpose of honing digital forensic processes is to empower the forensic examiner to focus on the unique and interesting aspects of their work, allowing them to spend more time addressing the probative questions in an investigation, enabling them to be decision makers rather than tool runners, and ultimately increase the quality of service to customers. This paper describes a method of evaluating the complete forensic process performed by examiners, and applying this approach to developing tools that recognize the interconnectivity of examiner tasks across a digital forensic laboratory. Illustrative examples are provided to demonstrate how this approach can be used to increase the overall efficiency and effectiveness of forensic examination of file systems, malware, and network traffic.

1. Introduction

Demands on digital forensic laboratories have grown in recent years, and this trend promises to continue. The number of forensic examinations being performed by digital forensic laboratories is also growing, and the amount of data received for each examination is increasing significantly. At the same time, because forensic investigations are results oriented, the demand for timely results has remained steady, and in some instances has increased. This is particularly true in cases that are heavily reliant on digital evidence; investigators need results as soon as viable.

In order to keep up with demand and ensure that customer needs are met, ongoing research and development is required to create tools and processes that increase efficiency in various aspects of forensic examinations. These include file system forensics, malware forensics, and the examination of network traffic. In addition to increasing efficiency within a digital forensic laboratory, one of the primary goals of these efforts is to enhance the comprehensiveness and investigative usefulness of forensic results. The purpose of these efforts is not to limit the forensic examinations to mass-produced, uniform products; but to empower the forensic examiner to focus on the unique and interesting aspects of their work. This allows forensic examiners to spend more time addressing the probative questions in an investigation, enabling them to be decision makers rather than tool runners, and ultimately increase the quality of service to customers. Throughout

* Corresponding author. Tel.: +1 202 670 2754.
E-mail address: eoghan@disclosedigital.com (E. Casey).

1742-2876/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.diin.2013.07.002

these efforts, care must be taken to ensure that all work is performed in a forensically sound manner.

This paper describes a method of evaluating the complete forensic process performed by examiners, and applying this data to developing tools that recognize the interconnectivity of examiner tasks across a digital forensic laboratory. The lessons learned from reengineering forensic processes, in order to increase efficiency and effectiveness, are presented in this paper. The goal is to assist other forensic laboratories and tool developers cultivate solutions for similar challenges in their environments.

2. Background

The efforts discussed in this paper aim to increase efficiency and consistency, and focus on processing data from three primary sources: 1) file systems 2) malware and 3) network traffic. In all instances, a more efficient forensic process can be developed, either by using parallel processing, strategic use of GPU support, distributed processing, or automatically extracting specific information that is commonly useful in a digital investigation. However, this approach only increases efficiency of a specific task, and concentrates on overcoming resource constraints. In fact, many tasks in the forensic process are not resource constrained and greater improvements can be achieved by rethinking the overall forensic process. As a result, with the number of examinations increasing, along with larger amounts of data, digital forensic laboratories are being compelled to rethink the overall forensic process in order to maintain and improve the quality of service delivered to customers.

Considering the full forensic process naturally leads to the realization that "timely results" do not simply refer to speeding up tasks, but also to delivering useful information at crucial decision points to support more effective case management and digital investigation. As such, improving the complete forensic process involves two general goals. First, increase efficiency of the full process, rather than dealing with subtasks separately. Second, provide useful information to support decisions that lead to more comprehensive and useful results. By reengineering the full forensic process to both support decisions and remove artificial barriers, digital forensics laboratories can simultaneously improve the efficiency and effectiveness of results.

In general terms, efforts to improve efficiency and effectiveness in forensic processes concentrate on the following core areas:

- Preparation and preservation: Before extracting and examining data, perform exam management and setup tasks, including evidence documentation and forensic duplication.
- Extraction and storage: Automatically pull common, useful information from available data and store it in a format that facilitates forensic examination and helps generate standard report output with results, in a consistent manner.
- Examination and reporting: Present automatically extracted information to forensic examiners and investigators at the beginning of their work on a case to jump start forensic examination and report creation, making the examiners decision makers rather than tool runners.
- Sharing, correlating and distributing: Provide the results of the examination to both the direct requestor and other interested parties in both a textual format and metadata format. The results can then be leveraged to be used in future examinations and allow organizations with similar needs to query the data.

Further work can enhance correlation and reconstruction processes, combining results from multiple sources to facilitate analysis, including a comprehensive timeline of events.

The paper presents an approach to honing the forensic processes in a digital forensic laboratory. This presentation begins by summarizing the primary lessons learned for simultaneously increasing efficiency and quality in forensic processes. Three illustrative examples are then provided to demonstrate how rethinking how the overall forensic process is engineered can be used to hone file system forensics, malware forensics, and network forensics, and therefore increase overall efficiency and effectiveness.

3. Dismantling process barriers

Traditionally, process models in digital forensics have separated individual tasks, such as forensic acquisition and extraction (Casey and Schatz, 2011; Kohna et al., 2013). Although treating each task separately can be useful for certain purposes, such as ensuring that a forensic process includes all necessary steps, this separation can create artificial barriers in the overall process. These artificial barriers become physical barriers when responsibility of each task is assigned to different individuals or groups within a digital forensic laboratory. As depicted in Fig. 1, such barriers can result in unnecessary redundancy, such as the same operation being run multiple times, because results of the operation are not shared among subtasks. Less obviously, such barriers can make it more difficult to develop solutions that merge subtasks or that take measures in one subtask that create efficiencies in subsequent subtasks.

When determining the most effective ways to speed up any process, it is necessary to understand the bottlenecks in the process (Goldratt, 1984). The first step is to study the full process (Radzicki and Taylor, 2008). In order to gain a comprehensive understanding of the entire forensic process in a given context, it is necessary to gather information from forensic examiners, digital investigators, and managers. In addition, when developing or refining a forensic process, it is important to use a methodical approach to documenting requirements and implementing capabilities

Fig. 1. Forensic process with separate subtasks being performed by different groups in a digital forensic laboratory.

(e.g., Agile development model). For example, take forensic examination of malware. Many organizations that perform forensic examinations of malicious code initially focus on automating the tools run by forensic examiners. While such efforts increase efficiency, greater improvements can be achieved by reconsidering the start and end points in the forensic processing of malware. Specifically, the potential for significant improvements in the overall process can be overlooked when the process is narrowly defined by the time it takes for the forensic examiner to identify the malware, extract the indicators, and perform any reverse engineering. It is also necessary to take into account the time spent by a case manager in a digital forensic laboratory, before analysis occurs. This includes evaluating the difficulty of a case and deciding who should be assigned each case. If it takes more than a week for the case manager to assign a case, this may be the longest delay in the overall forensic process, and attention should be devoted to facilitate the decision making task.

The full process must also include the time forensic examiners spend generating forensic reports and filling in required information. Storing extracted information in a standardized format and making it easily accessible to forensic examiners can speed up the report creation process by facilitating the creation of tables or exhibits in a consistent format. Since time spent by forensic examiners is much more expensive than time spent by computers, it is necessary to examine the full process, no matter how forensic examiners are spending their time.

Next, the engineer must examine any technical constraints that would slow down the system. Such evaluation involves measuring a computer system as it performs specific tasks to determine what is preventing the process from being performed faster. Are the issues caused by I/O speeds, processing power limitations, data organization, or other issues? These factors will play a large part in developing a final architecture for the system. For example, such an assessment of the full forensic process revealed that one of the main issues was not a processing power problem, but instead a disk I/O constraint. Specifically, using system monitoring tools to assess the performance of forensic workstations while they executed routine tasks showed that the processors were not fully utilized but that disk I/O was consistently at the maximum. Therefore, to reduce the impact of disk I/O speeds on the overall forensic process, it was more important to identify how multiple tasks could be performed while the data was loaded from the disk rather than concentrating on simply speeding up specific algorithms, or creating a tool runner.

Currently, when evidence is first received in many digital forensic laboratories, substantial resources are spent preparing the evidence, which can introduce substantial time in the workflow. These steps include evidence documentation and forensic duplication. Many digital forensic laboratories first acquire data from all evidential storage media, make a working copy of all forensic duplicates, and then extract specific data from the resulting working copies. This serial approach is not scalable because the process becomes less efficient as the size of storage media grows, particularly given the fact that disk I/O is the slowest operation in forensic processing.

One solution for making information available for review faster is to extract data from the original evidential media in read-only mode, prior to acquiring forensic duplicates. However, this approach may not be feasible for certain sources of digital evidence, such as smartphones. Furthermore, performing extensive processing on original evidence can be risky, particularly when dealing with old or damaged media that may only survive one forensic process. Perhaps more importantly, this approach maintains the separation of tasks, simply performing preservation and extraction operations in a different order, and actually adds a decision step (to create a forensic duplicate or not) potentially increasing the overall process time.

A potential solution to this issue is to combine evidence acquisition with the automated extraction of information. Specifically, a forensic acquisition method can be augmented to simultaneously feed data into an extraction process while creating the forensic duplicate. Performing multiple extraction processes in parallel with forensic acquisition reduces the need to wait for the acquisition process to finish before performing further processing, thus increasing overall efficiency. In addition, to support further forensic examination of evidence without expending additional time making working copies, forensic duplicates can be stored on a network accessible storage system to provide all forensic processes with read-only access to acquired data in a single, centralized location.

When reengineering a forensic process, keep in mind that increasing the speed or capacity of a limited resource is not the only way to reduce its impact on the overall process. To get the most out of a bottleneck, it is also important to make sure that the constrained resources are rarely idle. In other words, to reduce the amount of time that other parts of the process have to stand idle until additional data passes through the bottleneck, make every effort to ensure that data is continuously passing through the constrained resource. One way to ensure that a constrained resource is rarely unused is to create a buffer of input waiting to be processed by that resource to ensure that the operation is seldom idle. For instance, having storage media queued up for processing in a way that is visually clear, when the buffer is low and risks leaving the process idle (e.g., color coded green/orange/red), can help keep the preservation and extraction process busy.

In addition, keep in mind that focusing on one bottleneck will ultimately cause changes in the system that may cause a new choke point in a different part of the process. Therefore, it is necessary to reevaluate the overall forensic process periodically to ensure that efforts to increase efficiency are addressing an actual, current constraint.

Lastly, it is important that the system architecture is confirmed and constantly reviewed with the end users by monitoring system performance, providing a user feedback mechanism, periodically meeting with users, and documenting any new requirements. The development organization is a services organization and must make sure to confirm requirements and discuss the system with their customers during each step of the development process. This ensures that the final system will be useful and successful in the eyes of forensic

examiners, case managers, and digital investigators it supports.

4. Enabling effective decisions

A primary goal of early extraction of information from digital evidence sources (commonly referred to as *triage*) is to support decisions about conducting forensic processing and advancing a digital investigation more effectively (Murphy et al., 2010). Delving into a forensic examination without first obtaining a high level overview of the terrain and forming a strategy can lead to delays when unforeseen but preventable problems are encountered. Worse, important evidence might be overlooked. Conducting an investigation without first reviewing available evidence and identifying missing evidence can undermine an investigation, resulting in missed opportunities and incorrect conclusions. The way that integrated preservation, extraction, and presentation of data can support decision making is depicted in Fig. 2.

At the beginning of the forensic process, information is made available to digital investigators and laboratory managers to support their decisions on how to proceed with a case. Subsequently, forensic examiners can be informed by the questions and priorities that digital investigators developed during their initial review of the information that was extracted from available evidence. Simply stated, the aim is to avoid "information we could have used yesterday" syndrome. Examples include:

- Information about potentially missing items and online sources of evidence they may want to obtain (e.g., Web-based e-mail, remote storage, cloud services). For instance, the sooner investigators know that an additional source of digital evidence exists, such as an external USB hard drive or online storage "dropbox", the greater the chance they have of obtaining that evidence.
- Information about the severity of an offense, additional criminal activities, or concealment behavior (e.g., encryption) that may influence investigative direction, interviews, or additional charges.
- Information about items and activities on evidential media that can help focus attention in the forensic examination. For instance, given the opportunity to review available digital evidence early in the process, investigators may see areas of particular interest that forensic examiners can focus on immediately.
- Information about the skill level of the offender or complex technical aspects of the evidence, such as encryption or other concealment behavior that may require that attention of an experienced forensic examiner. The sooner forensic examiners know that attempts were made to conceal or destroy data (e.g., encryption, file wiping, backdating), the more time they have to deal with the challenge and possibly salvage evidence that is difficult to recover.

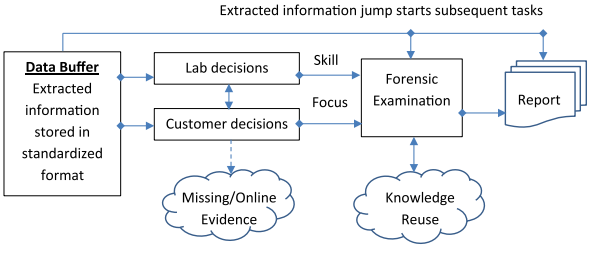

Fig. 2. Overall forensic process, including decision making and knowledge reuse.

In addition, information about previously encountered aspects of the evidence ("seen this before") that can benefit from reuse of forensic examination tools/techniques developed in prior cases to reduce duplication of effort.

Another important consideration is the user interface that enables an investigator, forensic examiner, case manager, or other decision maker, to review all extracted information with ease. Such an interface can provide useful context, such as highlighting files of potential interest and references to past forensic examination methods, that can be reused in the current case. An added design goal of the user interface is to format data in a consistent manner that is ready for immediate inclusion in a report, thus eliminating the time and potential inconsistencies previously introduced by manual data entry.

Three illustrative examples are presented in the remainder of this paper to demonstrate how reengineering the overall forensic process can greatly improve the efficiency and effectiveness of file system forensics, malware forensics, and network forensics.

5. File system processing

Most cases submitted to digital forensic laboratories require some form of file system forensics, making this an obvious place to improve efficiency. As a result, various tools have been developed to provide investigators with rapid results from evidential computer systems to help them make investigative decisions and to enable them to direct forensic examination to address specific questions. These tools are sometimes referred to in terms of triage or forensic data extraction, and have capabilities to extract pictures, videos, instant messaging (IM), peer-to-peer (P2P) information, e-mail messages, web browsing history, including allocated files and unallocated space, automatically looking inside compressed files, and excluding known system files using NIST NSRL. For instance, Deepthought runs various processes from a bootable CD to extract information from computer systems for initial review by digital investigators (Shaw and Browne, 2013). As another example, XIRAF is a powerful system that automates various forensic processes and gives digital investigators access to information as quickly as feasible (Bhoedjang et al., 2012). These tools can be configured to run certain operations and not others, as fits the case and investigative focus. Generally, these tools perform operations sequentially, performing one task to completion before starting the next, as depicted in Fig. 3.

Fig. 3. Sequential processes performed by triage/forensic data extraction tools.

This section explains how triage or forensic data extraction tools can be reengineered to improve efficiency, support decision makers, and enhance the quality of results by digital forensic laboratories.

5.1. Integrating preservation, extraction and storage

As described in the "Dismantling process barriers" section above, assessing the overall forensic process revealed that the main bottleneck in this process was disk I/O speeds and that the greatest efficiencies could be gained by performing multiple tasks when the data was loaded from the disk and by reusing the extracted information to perform subsequent tasks. This realization leads to several reengineering decisions that can both break down barriers in the process and support decision making.

First, a triage or forensic data extraction tool can simultaneously feed data into multiple extraction operations while preserving digital evidence by creating the forensic duplicate as depicted in Fig. 4.

In addition to acquiring and extracting data simultaneously, significant increases in efficiency can be attained by performing multiple operations on extracted data in parallel, including file carving and bulk data extraction (Garfinkel et al., 2010). This capability has been demonstrated to show the ability to image, verify, extract unallocated space, and carve files with only a 32% increase in time over standard imaging, as shown in Table 1. This integrated acquisition and extraction can feed data into a data buffer for further processing and review to support decision making and forensic examination.

In addition to reducing the choke point created by I/O speeds, performing multiple operations on data at the same time has the added benefit of breaking down prior barriers between the forensic extraction and forensic examination tasks. Specifically, routine data extraction processes that were previously performed during forensic examination can now be performed during the forensic acquisition process, such as creating file listings, calculating hash values of files, exporting files (including unallocated space), and performing file carving. The extracted information is then available to forensic examiners without the need for additional operations being performed on a forensic duplicate, speeding up the overall process.

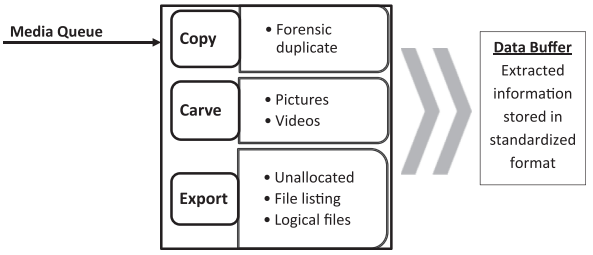

Fig. 4. Perform multiple operations simultaneously while acquiring and extracting data from I/O constrained evidential media.

By reengineering the overall process, additional operations can be performed on acquired and extracted data, such as keyword indexing and scanning for malware, as depicted in Fig. 5.

Another common barrier in forensic processing arises when a tool extracts information in a format that does not support subsequent, independent processing, using other tools. Instead, storing extracted information in a standardized database format supports further efficient processing, even enabling multiple tools to run simultaneously with queries to the database. To support this requirement, information extracted by a triage or forensic data extraction tool can be stored in both an XML format and SQLite database. This standardized output also makes it easier to incorporate new types of information and scale to larger systems as needed in the future. For instance, when it is necessary to store and process larger quantities of extracted information than can be accommodated by SQLite, a standardized format can be translated into another system such as a distributed NoSQL store.

5.2. Examination and reporting

By considering the full forensic process, tool developers can examine how the information gathered during initial triage or forensic data extraction can be leveraged by other areas of a digital forensic laboratory, increasing capabilities and decreasing the time of the overall forensic process. For instance, storing output from triage or forensic data extraction tools in a standardized format can enable systematic use of the output to support further forensic examination and to jump start reporting.

Table 1
Results showing only a 32% increase in time between straight imaging and imaging + verification + unallocated + carving in parallel. Testing was performed on a 4 GB Fat32 USB connected drive with 2.9 GB of Unallocated space. 4000 Files (740 MB of data) was carved.

Operation	Processing time (parallel)	Processing time (serial)	Throughput (MB)
Image	210	210	18.4
Image/verify	210	408	18.4
Image/verify/ unallocated	273	569	13.9
Image/verify/ unallocated/ carve	279	806	16.9

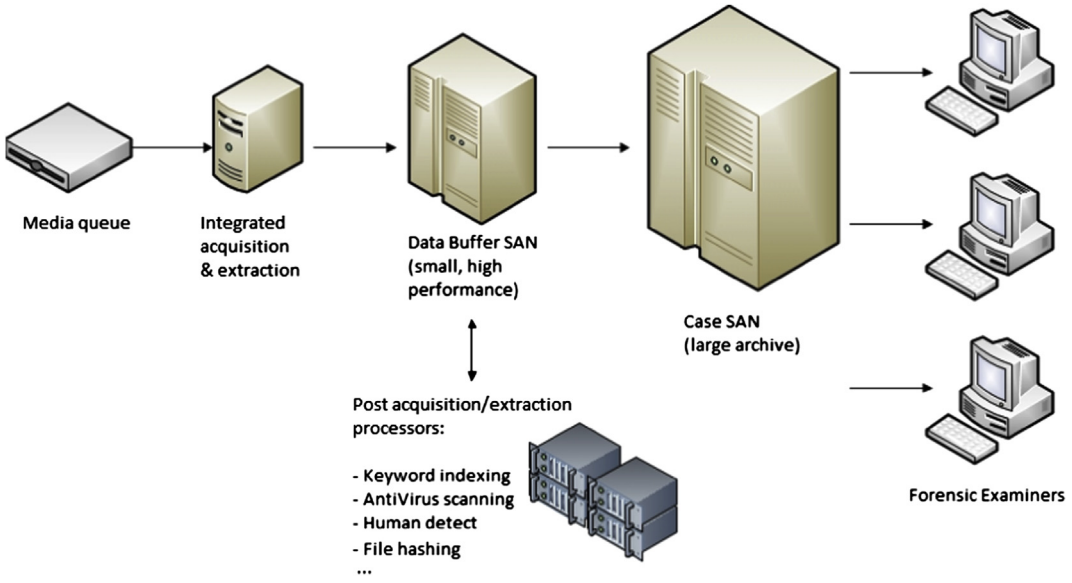

Fig. 5. Flow of acquired and extracted data in a reengineered forensic process.

Furthermore, providing a user-friendly front end can facilitate viewing, sorting, and classification of the files by the investigative, medical, and legal communities.

Some forensic data extraction tools highlight potential child pornography by comparing MD5 hashes with known child pornography hash sets. For example, the graphical user interface shown in Fig. 6 is a front end to the Defense Cyber Crime Center (DC3) forensic data extraction process that helps investigators inspect and tag items of interest. In addition to utilizing hash sets, DC3 developed and integrated a tool called Human Detect to prioritize the review of digital pictures in child exploitation investigations, and a tool called DC3 Video Validator to facilitate the review of digital videos (http://sourceforge.net/projects/video-validator/).

6. Malware processing

A common approach to automated malware processing is to extract information from executable files using a series

Fig. 6. Screenshot of GUI developed at DC3 to review results of forensic data extraction.

of utilities. The output is then made available to forensic examiners for review, and results for unique malware samples can be saved for future reference. In this way, when an MD5 match reveals that files in a new case had already been encountered and processed in a previous case, the results from the prior file instances can be reused, saving time on both processing and forensic examination.

This section explains how such an automated malware processing (AMP) system can be reengineered to improve efficiency, support decision makers, and enhance the quality of results. Significant improvements can be achieved by taking into account the entire process of conducting a forensic examination of malicious code. Digital investigations involving malware are often time sensitive, and the value of a malware analysis report is dependent upon the speed at which the malware has been processed. Therefore, time spent while the malware sits idle waiting to be processed, time spent assigning the examination, time spent while the report is being generated, and time spent between when the report is generated and it is used by a customer, are as important and valuable as time spent performing the forensic examination. The stages in Fig. 7 are defined by how forensic examiners view malware analysis occurring and do not reflect the order in which each task is completed. For example, the automated processing of malware generally occurs prior to preprocessing, and data fusion occurs at all stages.

6.1. Preparation and preservation

As shown in Fig. 7, forensic processing of malware begins with the preprocessing decisions, which include case assignment. Factors, such as the current case load of an examiner, the difficulty of the particular sample, the operating system the sample runs on, and other factors all contribute to the case assignment. Advanced correlation algorithms can be used to determine the uniqueness of a particular sample. This allows the case manager to determine whether this malware sample is a variation on a known piece of malware or something new. Slight variations to existing malware can be handed to less experienced forensic examiners, while unique and 'interesting' malware are handled by more experienced reverse engineers. Correlation is also an important aspect in ensuring work is not repeated by a forensic examiner. The automated malware process can examine and compare each individual file, determining whether it has been seen before and returning the previous analysis reported on that file.

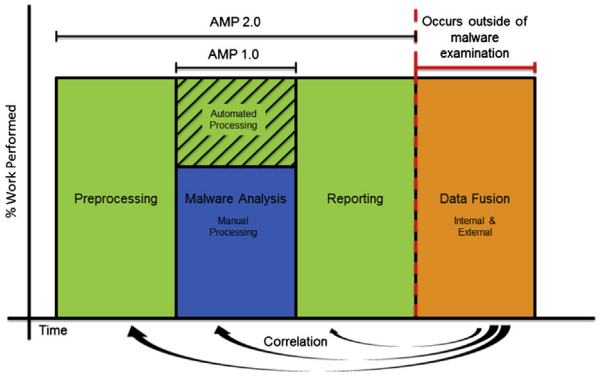

Fig. 7. Stages of the overall automated malware process. The horizontal axis shows tasks performed as time progresses. The vertical access shows the percentage of work performed within AMP verses through external tools.

6.2. Extraction and storage

In the second stage, depicted in Fig. 7, the actual malware analysis occurs. Malware analysis is a combination of running tools, performing tasks, and making decisions based on the results provided by those tools (reverse engineering is a separate task and is outside the scope of this paper). An effort is made to automate as many tasks and tools as possible, converting the forensic examiner from a tool runner to a decision maker. For example, the automated malware process developed by DC3 leverages a combination of COTS, GOTS, and open source tools to provide the necessary information to the forensic examiners.

The green (in the web version) in the second stage of Fig. 7 represents the amount of work in this stage that has been automated; the blue represents work that still needs to be performed by the forensic examiner. Over time, as new tools are developed and integrated, the area of the blue portion decreases, reducing the time a forensic examiner needs to spend on each malware sample. The information generated in this stage provides the data used to correlate future samples.

In addition to extracting information about a processed executable, it is important to store the information in a structured manner to support forensic analysis, reporting, correlation, and data sharing.

6.3. Examination and reporting

A graphical user interface should provide forensic examiners with powerful ways to interact with information that has been extracted from malware samples. In addition, a GUI can be designed to enable reuse of analysis techniques and tools from past cases by providing links to past cases with the same/similar malware characteristics. For example, Fig. 8 shows the GUI for the DC3 automated malware processing system. In this way, the GUI increases knowledge reuse and reduces redundant or inconsistent work across many different cases that are submitted at various times, but which involve similar malware.

The next stage shown in Fig. 7, reporting, helps forensic examiners create a common document for each analysis product. Providing forensic examiners with a reporting module can help jump start the report creation process by automatically populating the report with information from the previous stage and previous reports, identified through correlation. Furthermore, by requiring each analysis product to be generated within the reporting module, the malware processing system can capture information created by the forensic examiners, as well as information created by the tools, and reuse that information in future analysis products.

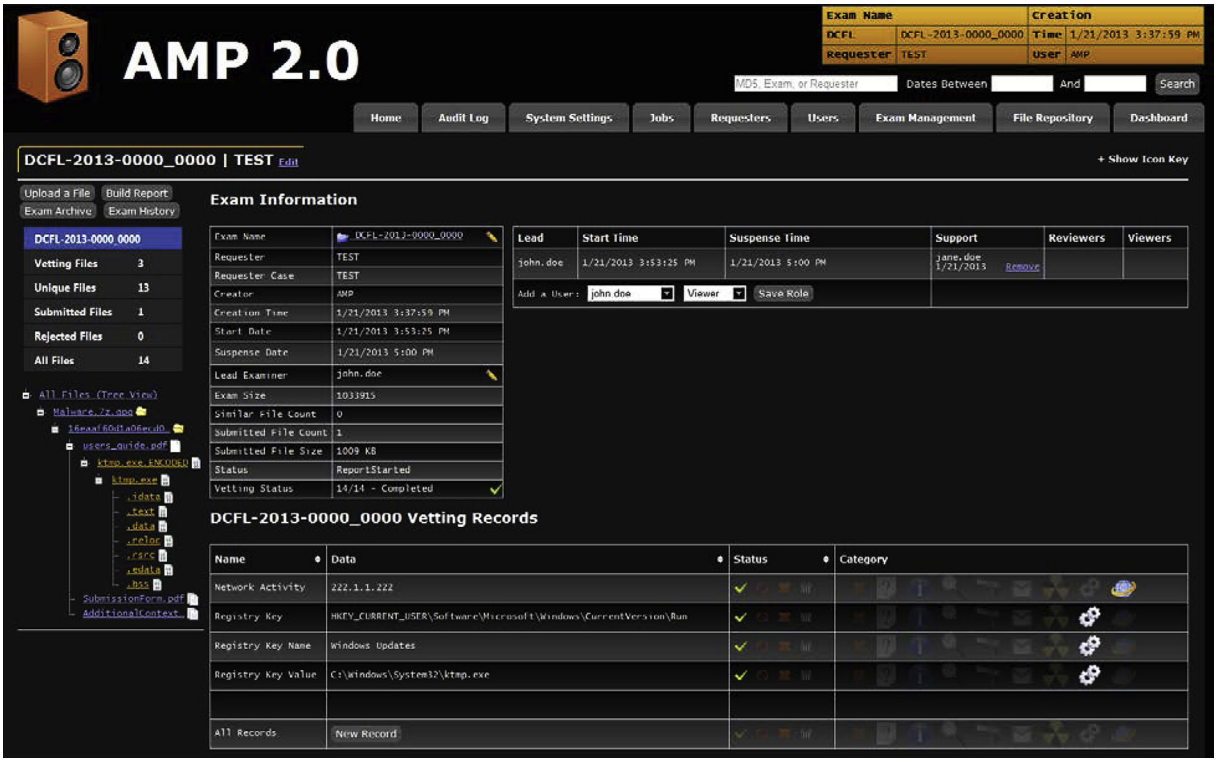

Fig. 8. Screenshot of AMP 2.0 user interface.

6.4. Sharing, correlation and distributing

The final stage of the automated malware analysis process is Data Fusion. One of the major advantages of the AMP system developed at DC3 is its design to allow for information reuse wherever possible. This updated malware processing system enables more advanced correlation (considering similarities beyond just MD5 value) to find related malware that has already been processed in prior forensic examinations. AMP does not simply look for direct malware matches, but searches for matches at the individual component layer and reuses the analysis performed during previous cases to support new products. Currently, this search is limited to DC3's own internal malware metadata database, but in the near future, the system will also correlate with information developed by other cyber centers. AMP will also be a provider of information, allowing other cyber centers to query its database for their own correlation activities. The data fusion activities are not limited to correlation. The DC3 Intrusions group mainly serves internal customers. These organizations each have their own custom systems for organizing information and creating products. In the next stage of AMP development, these organizations will receive the formal intrusions reports but will also have the results of the malware analysis automatically ingested as metadata and integrated into their products.

The stages discussed above outline the complete process, from ingestion of new malware to the dissemination of the malware report and the associated metadata. The system addresses all four of the core areas mentioned in the efficiency and effectiveness section. To achieve these goals, the AMP development team worked closely with forensic examiners and managers in the Intrusions division to develop a complete overview of what the team did and how they spent their time. This requirements gathering resulted in designing the first 3 stages of AMP. Further analysis included stage 4, data fusion, after taking a holistic view of the DC3 processes and understanding the needs of our external partners and customers.

7. Network traffic processing

DC3 performs forensic examination of network traffic, most commonly in PCAP format as part of investigations into network intrusions. To improve the efficiency of processing network traffic, DC3 developed a suite of tools called PCAPFAST. The use cases that formed the requirements of this tool are summarized in this section.

7.1. Preparation and preservation

When DC3 receives network traffic as part of the corpus of evidence in a case, it commonly contains multiple terabytes of data collected using any number of different methods from networks of unknown typology. In some circumstances, forensic examiners have received network traffic that was corrupt or not from the time period of interest. Therefore, before expending resources performing a

forensic examination of network traffic, it is necessary to determine basic aspects of the network traffic, including whether the data is from the time period of interest in the investigation, what network segments it covers, and if data is actually usable or corrupt.

To support initial assessment of network traffic, the first information that PCAPFAST provides to forensic examiners is a general overview of the network traffic, including number of packets, time period, and whether any corruption or other errors were found.

In addition, PCAPFAST provides references to the location of information in the original data set to enable forensic examiners to easily locate specific items or issues. This connection to the original evidence also helps verify important findings by going back to the original evidence.

7.2. Extraction and storage

Most network forensic analysis tools (NFAT) cannot deal with multiple terabytes in an efficient manner. As a result, DC3 initially created a set of in-house Perl scripts to perform processing of large quantities that ran across the entire data set to extract specific features that were relevant to the investigation. This was a time consuming procedure that usually had to run overnight each time a new set of indicators needed to be extracted. The sequential approach to examining network traffic was ultimately inefficient and limited the amount of information that could be extracted in a timely manner.

PCAPFAST was developed to create an "index" of network traffic that could be used to search for information more efficiently and with greater flexibility. Specifically, to support more efficient and flexible searching, PCAPFAST stores header information from each packet along with session information in a SQLite database. Forensic examiners can use PCAPFAST to query the database using syntax similar to tcpdump and Wireshark, which many forensic examiners are already familiar with. The database components were designed to make it possible to move to more powerful SQL databases if needed in the future.

7.3. Examination and reporting

When run against a quantity of network traffic, even when dealing with multiple files over an extended period, PCAPFAST creates a comprehensive report with items that forensic examiners commonly look for in network traffic. The output of PCAPFAST provides a high level report that gives forensic examiners an overview of network traffic and brining potentially interesting features to their immediate attention. The report includes histograms showing spikes in network activity, an overview of conversations and protocols, DNS lookups, country codes, top talkers, top destinations, and can be configured to highlight traffic to any known hosts of interest.

In addition, storing information in a database enables forensic examiners to perform searches and more complex queries faster than before. In addition to speeding up individual queries, PCAPFAST supports simultaneous queries running in parallel, which can significantly improve efficiency of examinations.

Finally, PCAPFAST enables forensic examiners to extract a subset of network traffic that matches specific criteria of interest. This smaller PCAP file can be examined in more depth using other tools, such as Wireshark.

The PCAPFAST system addresses three of the core areas mentioned in the efficiency and effectiveness section. To achieve these goals, the PCAPFAST development team worked closely with the Intrusions team at DC3 to gain a comprehensive understanding of what the team did and how they spent their time examining network traffic. This collaboration led to the creation of detailed software requirements specification and design documentation for PCAPFAST. In addition, the Intrusions team was given regular opportunities during the development process to assess PCAPFAST results and performance, and provided feedback that refined the system. Future work is planned to support sharing and distributing specific network traffic analysis capabilities, such as detecting certain command and control mechanisms, and decrypting certain network traffic.

8. Conclusion

There is increasing demand for timely results from digital forensic processes. In order to meet these needs, digital forensic laboratories must look across barriers within their organization to discover inefficiencies. These inefficiencies may result from one section of the organization not leveraging the full work of other sections, compute or I/O intensive operations being repeated, or other waste areas that slow the process. This paper documents examples of how to examine the complete workflow of an organization for process improvement and develop systems that provide solutions across that workflow. By automating across the process, forensic examiners are able to spend a greater portion of their time as decision makers rather than managing resources, running tools, or documenting results. Given the importance of timely information for decision making, the success of digital forensic processes depends heavily on working with investigators, forensic examiners, and other decision makers to ensure that their needs are being met.

When looking for inefficiencies and delays in the digital forensic processes, it is important to consider the complete process. Following existing digital forensic process models can lead to separation of tasks, creating barriers that hinder efficiency and effectiveness of the overall process. It is not sufficient to speed up a single task; it is also necessary to consider how the task can be performed in a way that "jump starts" subsequent tasks. If a constraint is managed successfully, future assessment of the overall process may reveal a new constraint elsewhere in the process. Therefore, assessment and reengineering of digital forensic processes is an ongoing effort for improvement. Ultimately, the goal of these efforts is to make the most of limited resources, support decisions at key points in an investigation, and increase the quality of forensic findings by enabling brain power to concentrate on probative questions.

In addition to honing the overall forensic process, further work is planned to enhance correlation and reconstruction processes at DC3, combining results from multiple sources to facilitate analysis, including

comprehensive timelines of events. In addition, future efforts are focusing on knowledge sharing and reuse in order to reduce redundant effort and reinventing the wheel during a digital investigation, giving forensic examiners access to methods and tools from prior work that can be applied to the current and future situations.

Acknowledgments

Special thanks to AMP Development Team: Chanpreet Julka, Jason Agurkis, Keith Bertolino, Matt Kowalski, Andrew Havens, FDE Development Team: Maurice Calhoun, Joe Lewthwaite, and PCAPFAST Development Team: Richard Cordovano, Andrew Medico, Matt Nolan.

About the Defense Cyber Crime Center (DC3)

DC3 provides digital and multimedia (D/MM) forensics, cyber investigative training, research, development, test and evaluation (RDT&E), and cyber analytics for the following DoD mission areas: information assurance (IA), critical infrastructure protection (CIP), law enforcement and counterintelligence (LE/CI), document and media exploitation (DOMEX), and counterterrorism (CT).

DC3 was established as an entity within the Air Force Office of Special Investigations in 1998.

DC3 is a national cyber center, as recognized in NSPD 54/HSPD 23 and serves as the operational focal point for the Defense Industrial Base Cybersecurity and Information Assurance Program (DIB CS/IA Program).

DC3 is located in Linthicum, MD with a staff of approximately 400, including DoD civilians, military, and contract partners. DC3 also hosts 23 liaisons/detailees from other agencies, including the Department of Homeland Security, OUSD (AT&L) Damage Assessment Management Office (DAMO), National Security Agency, Federal Bureau of Investigation, Defense Criminal Investigative Organizations, U.S. Army Military Intelligence, and U.S. Cyber Command.

References

Bhoedjang RAF, van Ballegooij A, van Beek HMA, van Schie JC, Dillema FW, van Baar RB, et al. Engineering an online computer forensic service. Digital Investigation 2012;9(2).

Casey E, Schatz B. Conducting digital investigations. In: Digital evidence & computer crime: forensic science, computers and the internet. 3rd ed. Academic Press; 2011.

Garfinkel S, Nelson A, White D, Roussev V. Using purpose-built functions and block hashes to enable small block and sub-file forensics. In: Proceedings of DFRWS 2010; 2010. Digital Investigation;7(Suppl.).

Goldratt E. The goal: a process of ongoing improvement. North River Press; 1984.

Kohna MD, Eloffb MM, Eloffa JHP. Integrated digital forensic process model. Digital Investigation 2013;38:103–15.

Murphy S, Flynn DP, McAleer TJ, Medico A, Raygoza D, Salyards MJ. Forensic data extraction in computer child pornography investigations. In: AAFS 2010 2010.

Radzicki MJ, Taylor RA. Origin of system dynamics: Jay W. Forrester and the history of system dynamics. In: U.S. Department of Energy's introduction to system dynamics. http://www.systemdynamics.org/DL-IntroSysDyn/start.htm; 2008.

Shaw A, Browne A. A practical and robust approach to coping with large volumes of data submitted for digital forensic examination. Digital Investigation 2013;10(3).