Computer Forensics: Digital Forensic Analysis Methodology

Ovie L. Carroll
Director, Cybercrime Lab
Computer Crime and Intellectual Property Section
Criminal Division

Stephen K. Brannon
Cybercrime Analyst, Cybercrime Lab
Computer Crime and Intellectual Property Section
Criminal Division

Thomas Song
Senior Cybercrime Analyst, Cybercrime Lab
Computer Crime and Intellectual Property Section
Criminal Division

I. Introduction

In comparison to other forensic sciences, the field of computer forensics is relatively young. Unfortunately, many people do not understand what the term computer forensics means and what techniques are involved. In particular, there is a lack of clarity regarding the distinction between data *extraction* and data *analysis*. There is also confusion about how these two operations fit into the forensic process. The Cybercrime Lab in the Computer Crime and Intellectual Property Section (CCIPS) has developed a flowchart describing the digital forensic analysis methodology. Throughout this article, the flowchart is used as an aid in the explanation of the methodology and its steps.

The Cybercrime Lab developed this flowchart after consulting with numerous computer forensic examiners from several federal agencies. It is available on the public Web site at www.cybercrime.gov/forensics_gov/forensicschart.pdf. The flowchart is helpful as a guide to instruction and discussion. It also helps clarify the elements of the process. Many other resources are available on the section's public Web site, www.cybercrime.gov. In addition, anyone in the Criminal Division or U.S Attorneys' offices can find additional resources on the new intranet site, CCIPS Online. Go to DOJ Net and click on the "CCIPS Online" link. You can also reach us at (202) 514-1026.

II. Overview of the digital forensics analysis methodology

The complete definition of computer forensics is as follows: "The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal. . . ." A Road Map for Digital Forensic Research, Report from the First Digital Forensic Research Workshop (DFRWS), *available at* http://dfrws.org/2001/dfrws-rm-final.pdf.

Defining computer forensics requires one more clarification. Many argue about whether computer forensics is a science or art. *United States v. Brooks*, 427 F.3d 1246, 1252 (10th Cir. 2005) ("Given the numerous ways information is stored on a computer, openly and surreptitiously, a search can be as much an art as a science."). The argument is unnecessary, however. The tools and methods are scientific and are verified scientifically, but their use necessarily involves elements of ability, judgment, and interpretation. Hence, the word "technique" is often used to sidestep the unproductive science/art dispute.

The key elements of computer forensics are listed below:

- The use of scientific methods
- Collection and preservation
- Validation
- Identification
- Analysis and interpretation
- Documentation and presentation

The Cybercrime Lab illustrates an overview of the process with Figure 1. The three steps, Preparation/Extraction, Identification, and Analysis, are highlighted because they are the focus of this article. See Figure 1, page 5.

In practice, organizations may divide these functions between different groups. While this is acceptable and sometimes necessary, it can create a source of misunderstanding and frustration. In order for different law enforcement agencies to effectively work together, they must communicate clearly. The investigative team must keep the entire picture in mind and be explicit when referring to specific sections.

The prosecutor and forensic examiner must decide, and communicate to each other, how much of the process is to be completed at each stage of an investigation or prosecution. The process is potentially iterative, so they also must decide how many times to repeat the process. It is fundamentally important that everyone understand whether a case only needs preparation, extraction, and identification, or whether it also requires analysis.

The three steps in the forensics process discussed in this article come after examiners obtain forensic data and a request, but before reporting and case-level analysis is undertaken. Examiners try to be explicit about every process that occurs in the methodology. In certain situations, however, examiners may combine steps or condense parts of the process. When examiners speak of lists such as "Relevant Data List," they do not mean to imply that the lists are physical documents. The lists may be written or items committed to memory. Finally, keep in mind that examiners often repeat this entire process, since a finding or conclusion may indicate a new lead to be studied.

III. Preparation/Extraction

See Figure 2, page 5.

Examiners begin by asking whether there is enough information to proceed. They make sure a clear request is in hand and that there is sufficient data to attempt to answer it. If anything is missing, they coordinate with the requester. Otherwise, they continue to set up the process.

The first step in any forensic process is the validation of all hardware and software, to ensure that they work properly. There is still a debate in the forensics community about how frequently the software and equipment should be tested. Most people agree that, at a minimum, organizations should validate every piece of software and hardware after they purchase it and before they use it. They should also retest after any update, patch, or reconfiguration.

When the examiner's forensic platform is ready, he or she duplicates the forensic data provided in the request and verifies its integrity. This process assumes law enforcement has already obtained the data through appropriate legal process and created a forensic image. A forensic image is a bit-for-bit copy of the data that exists on the original media, without any additions or deletions. It also assumes the forensic examiner has received a working copy of the seized data. If examiners get original evidence, they need to make a working copy and guard the original's chain of custody. The examiners make sure the copy in their possession is intact and unaltered. They typically do this by verifying a hash, or digital fingerprint, of the evidence. If there are any problems, the examiners consult with the requester about how to proceed.

After examiners verify the integrity of the data to be analyzed, a plan is developed to extract data. They organize and refine the forensic request into questions they understand and can answer. The forensic tools that enable them to answer these questions are selected. Examiners generally have preliminary ideas of what to look for, based on the request. They add these to a "Search Lead List," which is a running list of requested items. For example, the request might provide the lead "search for child pornography." Examiners list leads explicitly to help focus the examination. As they develop new leads, they add them to the list, and as they exhaust leads, they mark them "processed" or "done."

For each search lead, examiners extract relevant data and mark that search lead as processed. They add anything extracted to a second list called an "Extracted Data List." Examiners pursue all the search leads, adding results to this second list. Then they move to the next phase of the methodology, identification.

IV. Identification

See Figure 3, page 6.

Examiners repeat the process of identification for each item on the Extracted Data List. First, they determine what type of item it is. If it is not relevant to the forensic request, they simply mark it as processed and move on. Just as in a physical search, if an examiner comes across an item that is incriminating, but outside the scope of the original search warrant, it is recommended that the examiner immediately stop *all* activity, notify the appropriate individuals, including the requester, and wait for further instructions. For example, law enforcement might seize a computer for evidence of tax fraud, but the examiner may find an image of child pornography. The most prudent approach, after finding evidence outside the scope of a warrant, is to stop the search and seek to expand the warrant's authority or to obtain a second warrant.

If an item is relevant to the forensic request, examiners document it on a third list, the Relevant Data List. This list is a collection of data relevant to answering the original forensic request. For example, in an identity theft case, relevant data might include social security numbers, images of false identification, or e-mails discussing identity theft, among other things. It is also possible for an item to generate yet another search lead. An e-mail may reveal that a target was using another nickname. That would lead to a new keyword search for the new nickname. The examiners would go back and add that lead to the Search Lead List so that they would remember to investigate it completely.

An item can also point to a completely new potential source of data. For example, examiners might find a new e-mail account the target was using. After this discovery, law enforcement may want to subpoena the contents of the new e-mail account. Examiners might also find evidence indicating the target stored files on a removable universal serial bus (USB) drive—one that law enforcement did not find in the original search. Under these circumstances, law enforcement may consider getting a new search warrant to look for the USB drive. A forensic examination can point to many different types of new evidence. Some other examples include firewall logs, building access logs, and building video security footage. Examiners document these on a fourth list, the New Source of Data list.

After processing the Extracted Data list, examiners go back to any new leads developed. For any new data *search* leads, examiners consider going back to the Extraction step to process them. Similarly, for any new *source* of data that might lead to new evidence, examiners consider going all the way back to the process of obtaining and imaging that new forensic data.

At this point in the process, it is advisable for examiners to inform the requester of their initial findings. It is also a good time for examiners and the requester to discuss what they believe the return on investment will be for pursuing new leads. Depending on the stage of a case, extracted and identified relevant data may give the requester enough information to move the case forward, and examiners may not need to do further work. For example, in a child pornography case, if an examiner recovers an overwhelming number of child pornography images organized in user-created directories, a prosecutor may be able to secure a guilty plea without any further forensic analysis. If simple extracted and identified data is not sufficient, then examiners move to the next step, analysis.

V. Analysis

See Figure 4, page 7.

In the analysis phase, examiners connect all the dots and paint a complete picture for the requester. For every item on the Relevant Data List, examiners answer questions like who, what, when, where, and how. They try to explain which user or application created, edited, received, or sent each item, and how it originally came into existence. Examiners also explain where they found it. Most importantly, they explain why all this information is significant and what it means to the case.

Often examiners can produce the most valuable analysis by looking at when things happened and producing a timeline that tells a coherent story. For each relevant item, examiners try to explain when it was created, accessed, modified, received, sent, viewed, deleted, and launched. They observe and explain a sequence of events and note which events happened at the same time.

Examiners document all their analysis, and other information relevant to the forensic request, and add it all to a fifth and final list, the "Analysis Results List." This is a list of all the meaningful data that answers who, what, when, where, how, and other questions. The information on this list satisfies the forensic request. Even at this late stage of the process, something might generate new data search leads or a source of data leads. If this happens, examiners add them to the appropriate lists and consider going back to examine them fully.

Finally, after examiners cycle through these steps enough times, they can respond to the forensic request. They move to the Forensic Reporting phase. This is the step where examiners document findings so that the requester can understand them and use them in the case. Forensic reporting is outside the scope of this article, but its importance can not be overemphasized. The final report is the best way for examiners to communicate findings to the requester. Forensic reporting is important because the entire forensic process is only worth as much as the information examiners convey to the requester. After the reporting, the requester does case-level analysis where he or she (possibly with examiners) interprets the findings in the context of the whole case.

VI. Conclusion

As examiners and requesters go through this process, they need to think about return on investment. During an examination, the steps of the process may be repeated several times. Everyone involved in the case must determine when to stop. Once the evidence obtained is sufficient for prosecution, the value of additional identification and analysis diminishes.

It is hoped that this article is a helpful introduction to computer forensics and the digital forensics methodology. This article and flowchart may serve as useful tools to guide discussions among examiners and personnel making forensic requests. The Cybercrime Lab in the Computer Crime and Intellectual Property Section (CCIPS) is always available for consultation. CCIPS personnel are also available to assist with issues or questions raised in this article and other related subjects.❖

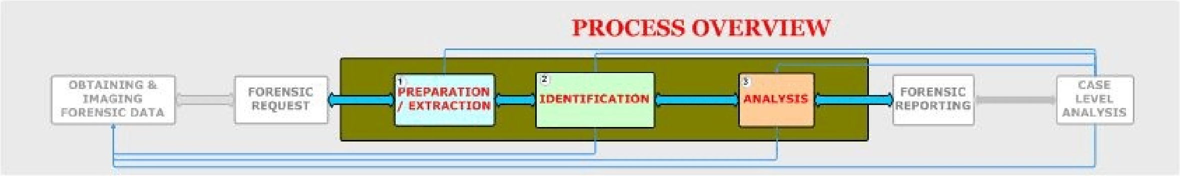
PROCESS OVERVIEW
OBTAINING & IMAGING FORENSIC DATA
FORENSIC REQUEST
1 PREPARATION / EXTRACTION
2 IDENTIFICATION
3 ANALYSIS
FORENSIC REPORTING
CASE LEVEL ANALYSIS

FIGURE 1

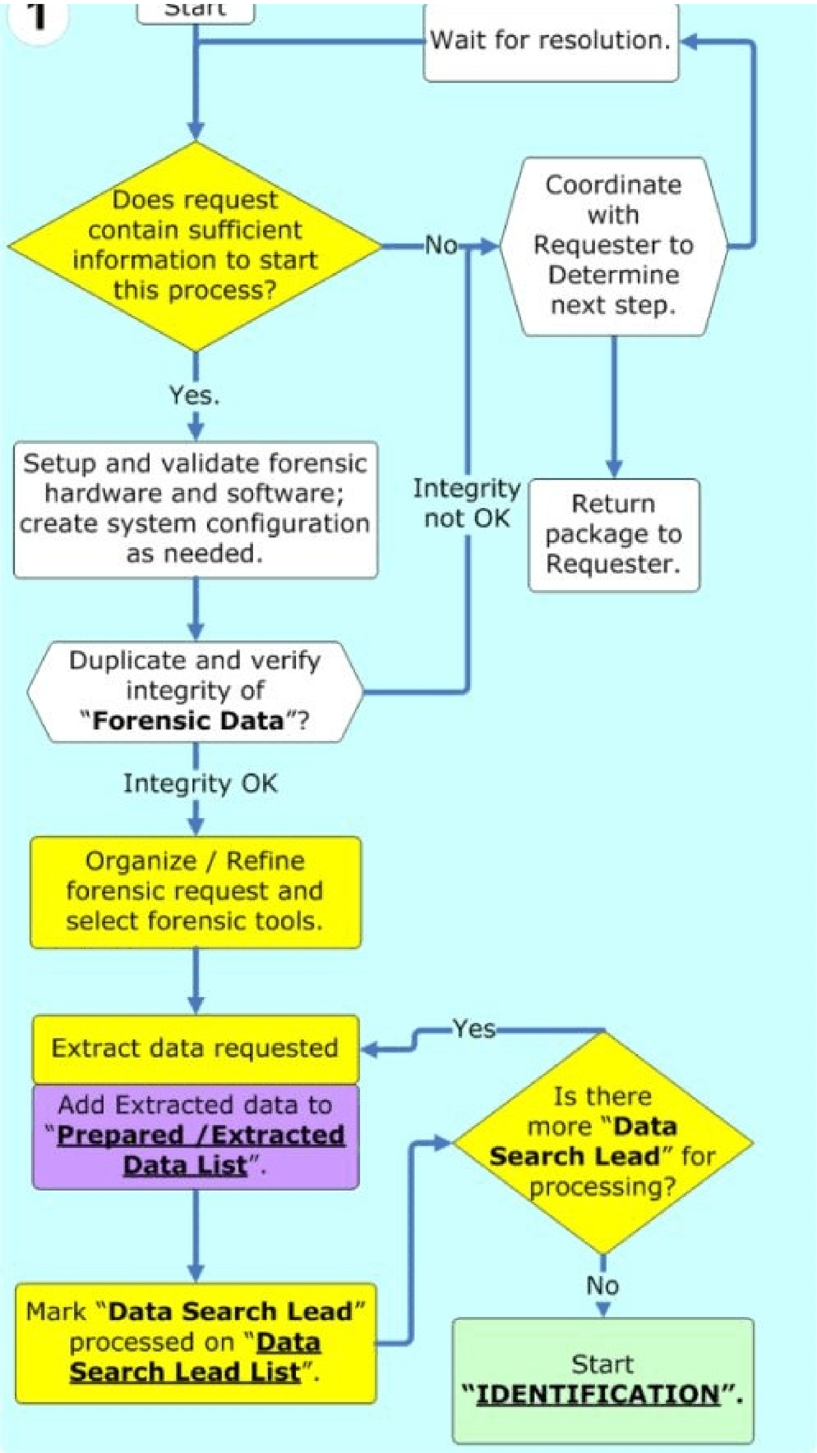
1
Start
Wait for resolution.
Does request contain sufficient information to start this process?
No
Coordinate with Requester to Determine next step.
Yes.
Setup and validate forensic hardware and software; create system configuration as needed.
Integrity not OK
Return package to Requester.
Duplicate and verify integrity of "Forensic Data"?
Integrity OK
Organize / Refine forensic request and select forensic tools.
Extract data requested
Yes
Add Extracted data to "Prepared /Extracted Data List".
Is there more "Data Search Lead" for processing?
No
Mark "Data Search Lead" processed on "Data Search Lead List".
Start "IDENTIFICATION".

Figure 2

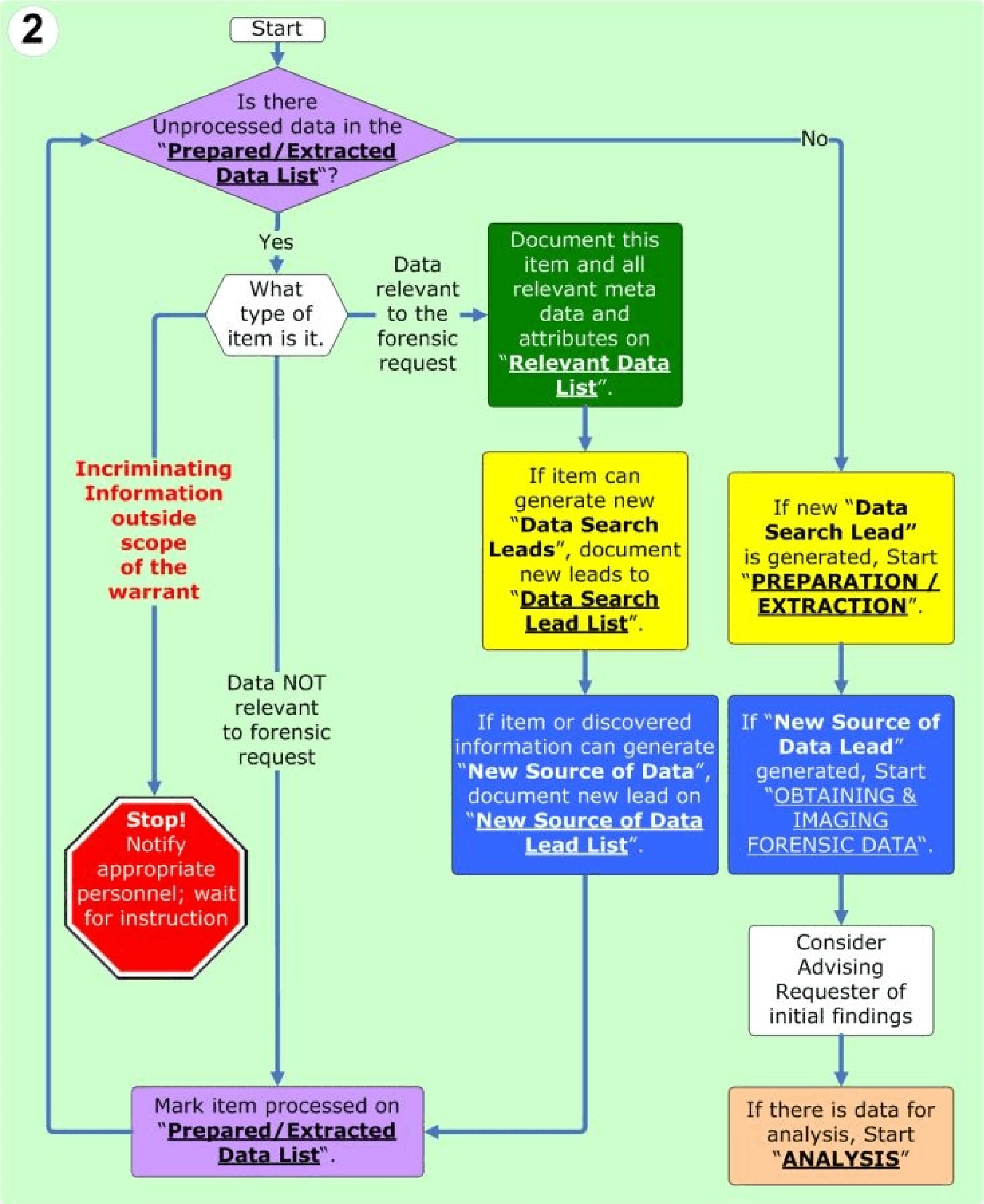
2
Start
Is there Unprocessed data in the "Prepared/Extracted Data List"?
No
Yes
What type of item is it.
Data relevant to the forensic request
Document this item and all relevant meta data and attributes on "Relevant Data List".
If item can generate new "Data Search Leads", document new leads to "Data Search Lead List".
If new "Data Search Lead" is generated, Start "PREPARATION / EXTRACTION".
Incriminating Information outside scope of the warrant
Data NOT relevant to forensic request
If item or discovered information can generate "New Source of Data", document new lead on "New Source of Data Lead List".
If "New Source of Data Lead" generated, Start "OBTAINING & IMAGING FORENSIC DATA".
Stop! Notify appropriate personnel; wait for instruction
Consider Advising Requester of initial findings
Mark item processed on "Prepared/Extracted Data List".
If there is data for analysis, Start "ANALYSIS"

Figure 3

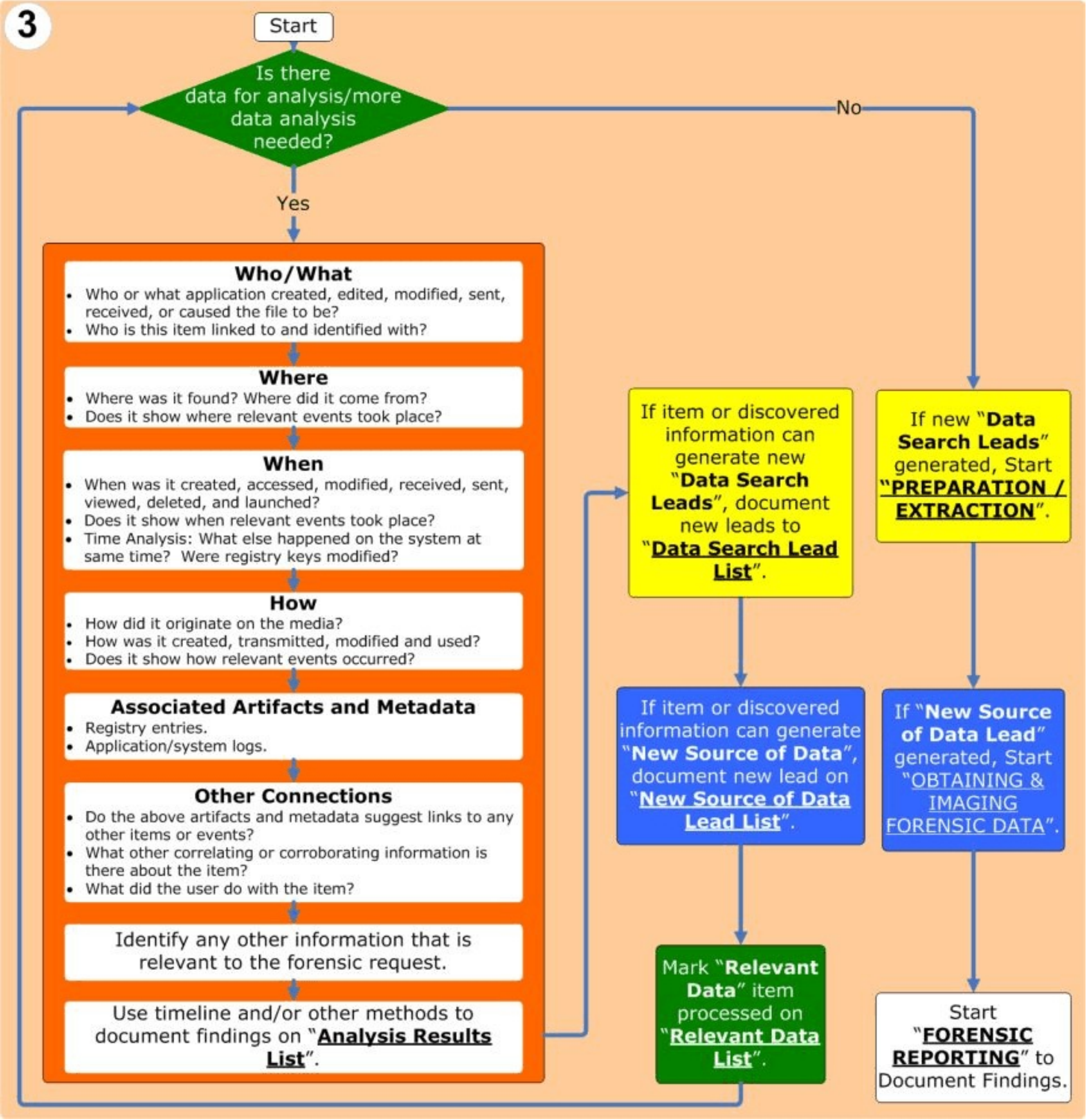
3
Start
Is there data for analysis/more data analysis needed?
No
Yes
Who/What
Who or what application created, edited, modified, sent, received, or caused the file to be?
Who is this item linked to and identified with?
Where
Where was it found? Where did it come from?
Does it show where relevant events took place?
When
When was it created, accessed, modified, received, sent, viewed, deleted, and launched?
Does it show when relevant events took place?
Time Analysis: What else happened on the system at same time? Were registry keys modified?
How
How did it originate on the media?
How was it created, transmitted, modified and used?
Does it show how relevant events occurred?
Associated Artifacts and Metadata
Registry entries.
Application/system logs.
Other Connections
Do the above artifacts and metadata suggest links to any other items or events?
What other correlating or corroborating information is there about the item?
What did the user do with the item?
Identify any other information that is relevant to the forensic request.
Use timeline and/or other methods to document findings on "Analysis Results List".
If item or discovered information can generate new "Data Search Leads", document new leads to "Data Search Lead List".
If item or discovered information can generate "New Source of Data", document new lead on "New Source of Data Lead List".
Mark "Relevant Data" item processed on "Relevant Data List".
If new "Data Search Leads" generated, Start "PREPARATION / EXTRACTION".
If "New Source of Data Lead" generated, Start "OBTAINING & IMAGING FORENSIC DATA".
Start "FORENSIC REPORTING" to Document Findings.

Figure 4

ABOUT THE AUTHORS

❑**Ovie L. Carroll** is the Director of the Cybercrime Lab in the CCIPS. He has over twenty years of law enforcement experience. He previously served as the Special Agent in Charge of the Technical Crimes Unit at the Postal Inspector General's Office and as a Special Agent with the Air Force Office of Special Investigations.

❑**Stephen K. Brannon** is a Cybercrime Analyst in the CCIPS's Cybercrime Lab. He has worked at the Criminal Division in the Department of Justice and in information security at the FBI.

❑**Thomas Song** is a Senior Cybercrime Analyst in the CCIPS's Cybercrime Lab. He has over fifteen years in the computer crime and computer security profession. He specializes in computer forensics, computer intrusions, and computer security. He previously served as a Senior Computer Crime Investigator with the Technical Crimes Unit of the Postal Inspector General's Office.

The Cybercrime Lab is a group of technologists in the CCIPS in Washington, DC. The lab serves CCIPS attorneys, Computer Hacking and Intellectual Property (CHIP) units in the U.S. Attorneys' offices, and Assistant U.S. Attorneys, by providing technical and investigative consultations, assisting with computer forensic analysis, teaching, and conducting technical research in support of Department of Justice initiatives.✠