Graduate Digital Forensics Education at the Air Force Institute of Technology

G. L. Peterson, R. A. Raines, and R. O. Baldwin
Center for Information Security Education and Research
Department of Electrical and Computer Engineering
Air Force Institute of Technology
2950 Hobson Way, Bldg 642
Wright Patterson AFB OH 45433-7765

Abstract

The Department of Electrical and Computer Engineering (AFIT/ENG) at the Air Force Institute of Technology (AFIT), currently offers a graduate-level introductory course in digital forensics. Students are introduced and exposed to several challenges and topics in the digital forensics course. The course addresses the ethical and legal procedures as well as basic forensic science principles in only the most general manner. A larger percentage of lecture and lab time is spent discussing the technical details of incident response and media analysis. The detail into the network forensics and digital device analysis topics start to breach technical details but not to the level of attempting mastery. This course provides our students with real world digital forensics experience to prepare them for the challenges they may face in post-graduate employment.

1. Introduction

The digital forensics course at AFIT is currently in its third year of existence. Enrollment has steadily increased over this period from a handful of students to approximately twenty for this year's offering. The course is offered as part of three different Masters of Science degree programs: Cyber Operations, Computer Science, and Computer Engineering. The digital forensics course is tightly integrated with our other computer security courses. The techniques that the students learn build on experiences from the Cyber Defense Exercise (CDX), in which the students administer a network and defend it against Red Team attacks for a week. Specifically, during the CDX students must determine what goes wrong after an attack. In the digital forensics course, one fourth of the course is spent on live network response. This exposes students to the tools needed when faced with these situations in the future.

Our course is actually entitled Cyber Forensics, but for this article, we use the term digital forensics as adopted by the National Center for Forensic Sciences, Scientific Working Group on Digital Evidence. The reason being that digital forensics is not limited to the computer holding the evidence, nor the network the computer is connected to but also all of the disparate digital devices which permeate our daily lives. Items such as mobile phones, GPS receivers, PDAs, and MP3 players all have the potential to store evidence.

The following section introduces background information discussing digital forensics in general and the material divisions possible, as well as those we used to drive our course content. The third section discusses the general course content by topic, and leads into the discussion of the lab contents. This is followed by minor details which we have found improve the course and future improvements we hope to implement.

2. Background

Digital forensics, similar to other forensic sciences, consists of the three parts: the science, the evidence, and the law [1]. All three parts are intimately tied together. The methods defining what can be used as evidence and how it is collected is governed by U.S. law. This governance in turn dictates what should be collected. The science drives the methods that generate further investigative leads and facts from the evidence. Science then is guided by what can be collected (evidence) and what can be done with it and still be admissible (law). Furthermore, the law provides the forum for which the evidence is collected and the science is used to present facts about a case. But, because these three are so interrelated and dependent, they don't allow for a good separation to teach by.

To gain a working insight into an approach for teaching digital forensics, an acceptable model for process breakdown needed to be found. One such approach was developed by Yasinsac, et al. [2]. This model presents the following divisions: collection, preservation, presentation,

and preparation. This work ties directly to the Digital Forensic Research Workshop (DFRWS) Investigative Process [3] which categorizes the essential steps as Identification, Preservation, Collection, Examination, Analysis, Presentation, and Decision. It can be seen from this guidance that the roles of the digital forensics scientist clearly centers on preservation collection, examination, and analysis. Several other process definitions have also been proposed [4-7].

From the above cited works, the presentation of digital forensics course material can come from different directions, either the component view (science, evidence, and law) or the process view (preservation, collection, examination, and analysis). Rather than blindly following one approach, we have adopted an approach that breaks the course material into five areas. This approach has a good balance between a) presenting the material from both views, and b) meeting our student population needs. The five topic areas are Ethics and Legal Procedures; Basic Forensic Science; Media Capture and Analysis; Network Forensics; and Digital Device Analysis. Table 1 shows the percentage of course time spent within each of the five topic areas.

Table 1. Digital Forensics Course Material Breakdown

Course Subject Area	Percentage of Course
Ethics and Legal Procedures	10
Basic Forensic Science	10
Media Capture and Analysis	40
Network Forensics	25
Digital Device Analysis	15

The Ethics and Legal Procedures include material on ethical behavior as it relates to computer usage. We discuss where individuals learn computer ethics (at home, school, and/or from the community) and how ethical behavior translates into a networked environment. The digital forensics side of these issues emphasizes the criminal mind and how some individuals reject ethics. The legal procedures then address the definition of cyber crime, concerns about search and seizure rights, the Fourth Amendment, and the large base of legal precedent being developed. This also extends into the question of the validity of analysis tools. That is, what are the standards, practices and/or precedence for use that must take place prior to a tool being "validated" and its results admissible in a court of law? An excellent introduction to many of these topics can be found in Eoghan Casey's "Digital Evidence and Computer Crime" [8].

Basic Forensic Science is concerned with both the law enforcement view of forensics as well as general lab policies. Some of these topics include: Locard's Principle

(Figure 1), Inman & Rudin Forensic Science Paradigm, as well as questions of what can and should be seized at a crime scene, what needs to be included in a warrant's text to ensure that the seizure is legal, once items are seized what happens with them, and how are items treated in the lab. Some of these questions are addressed via a general overview and guide by the Department of Justice (DOJ) on Search and Seizure of digital media [9]. The American Society of Crime Laboratory Directors (ASCLD) has provided a means by which forensics and digital forensics labs can be certified, and this is discussed as well [10].

Media Capture and Analysis is concerned with the correct and accurate handling of media which includes proper techniques for acquiring and verifying an image of the media, and analyzing the media's physical and logical structure to extract evidence. The data analysis portion includes some of the most difficult problems that forensics investigators encounter, that of information hiding in the logical structure of the media and in the network traffic itself. This includes such topics as steganalysis, Domain Name Service (DNS) messaging, document metadata, and encryption.

Network Forensics investigates the situation from a network standpoint. When viewed from that perspective, evidence can be contained within network log files. Questions can be raise about the type of logging information available and how from this log can additional information about the network traffic itself be extracted.

Because many of our graduates fill network support positions, we also include a significant discussion of incident response with live machines under both the Media Capture and Analysis and Network Forensics topics. The reason is, and this is also true for some corporations, it is more important to restore the systems operational status than to provide the evidentiary validity for a legal action.
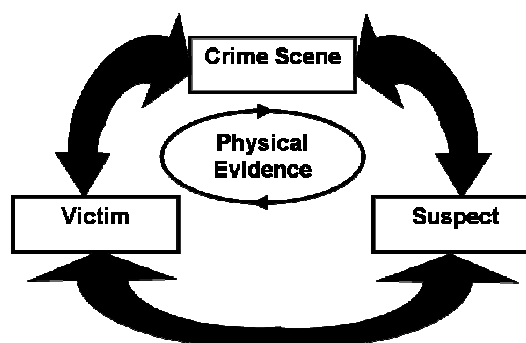

Figure 1. Locard's Principle.

Digital Device Analysis looks at all of the disparate devices that may confront investigators. In our course, we look at the storage and extraction of information from USB flash drives, and MP3 players, and introduce the topic of mobile phone forensics, but, not to the detail level of media and data analysis.

Any introductory course in digital forensics should introduce all of these topics. The depth to which each is covered can vary depending on the program. The text that we use for the course is Mandia and Prosise's "Incident Response and Computer Forensics, 2nd Edition" [11], supplemented with several documents on best practices and search and seizure, and course notes. Of note for search and seizure best practices, we make use of the National Institute of Justice's "Electronic Crime Scene Investigation: A Guide for First Responders" [12], and also the Air Force Office of Special Investigations crime scene training manual [13].

3. AFIT's Digital Forensics Course Content

Our graduate course is offered within the Graduate School of Engineering and Management and can be used to partially fulfill the requirements for the Masters of Science degrees in Computer Science, Computer Engineering, or Cyber Operations. We specifically focus the course on the technical details of digital forensics rather than legal, law enforcement, and policy issues.

Recall from Table 1 above, the five broad topic areas provide the focus for our digital forensics course. As shown in Table 1, we address the ethical and legal procedures and basic forensic science principles in only the most general manner. A larger percentage of time is spent discussing media analysis. The details of network forensics and digital device analysis topics start to breach technical details but not to the level of attempting mastery.

We have found that the principles, methods, and science are best understood and learned in a joint lecture and lab setting. During the lectures, we discuss the science and technical hardware details, as most of our students come from a Computer Science (CS) or Computer Engineering (CE) background. We have found that while most of our CS/CE students have had courses in operating systems and computer architectures, they are predominantly unaware of how the interfaces between the components in today's desktop PC are really put together. Because of this, the technical portion of the lecture is spent discussing these items. We reason that a good forensics analyst needs to know how to manually do everything that a tool they use does automatically. For example, this includes locating a file, and/or undeleting it. The student should be able to describe the process in both general and technical terms.

The remainder of the lecture time is spent discussing the processes and procedures for the labs themselves. In the labs, the students have the opportunity to run experiments and learn about the systems as well as all of the facets of digital forensics. Because of the heavy lab component of this course, we address the bulk of the course contents in terms of the labs the students complete.

3.1 Lab structure, requirements and type: What works and what didn't

AFIT is on an academic quarter system. This means there are ten weeks of instruction time available for a course. Most AFIT courses are four quarter credit hours. This allows us to interact with the students for a minimum of forty hours over a term. Typically, AFIT student-instructor interactions are increased by close to fifty percent (sixty hours) for laboratory courses.

Over the ten weeks there are seven labs and a class project, the syllabus is shown in Table 2. The labs themselves are structured similarly to those at the University of Tulsa, and cover the range of topics shown in Table 1. The students work in teams of three. The group process provides two noticeable benefits in this course. The first is associated with student background and experience level. Since the majority of AFIT students are military, a great wealth of operationally diverse experience is brought into the lab. The different experiences and ideas come together when solving the labs. This improves each student's opportunity to complete the lab. The second benefit is from the group lab team structure. Because student schedules vary, they are forced to maintain a chain of evidence as it is not always possible for the student group to collectively meet in the lab at the same time.

Our digital forensics laboratory setup includes 16 machines. One of the machines serves as the victim/evidence computer. This machine is a 2.8 GHz Pentium 4 with 1 GB of RAM, and a 20 GB HD running Windows XP with Service Pack 2. The machine is disconnected from our university network during the live response as we install two rootkits. To perform the live response, the machine is connected to a laptop that has it's hard drive wiped and OS reinstalled after the lab. All of our labs make use of Windows XP as it is the operating system mandated for use by the Air Force, and as such will be the OS most frequently encountered.

The students are issued their own hard drive for imaging, analysis, and retention of chain-of-evidence. The machines the students use during analysis are 3.0 GHz Pentium 4s with 1 GB of RAM, and 40 GB hard drives, all running Windows XP with Service Pack 2. The software is a mixture of freeware and commercial. We use Helix, and Penguin Sleuth bootable CDs, both of which

include the dcfldd imaging tool and the Autopsy analysis tool suite. The commercial tools range from the forensics professional version of Winhex, which allows the students the lowest level view of the media, to EnCase and FTK which provide a GUI with advanced recovery and analysis tools.

Table 2. Digital Forensics Syllabus

Week	Class Activities	Lab Due
1	Introduction, Forensic Principles, Legal and Policy Issues, Best Practices, Investigation Guidelines	
	Guest Speaker (Law Enforcement)	
2	First Response	
	Live Response	Policy Creation
3	Live Response (con't)	
	SHA/MD5 BIOS Password	First Response
4	SAM Database (NTFSDOS)	
	Bios and the Hard Disk	Live Response
5	History and Partitions, Disk Storage Introduction	
	EnCase Demo (RM 2011)	Password Cracking
6	Disk Storage (FAT 12, 16 and 32, NTFS, RAID)	
	Disk Storage (cont'd)	Imaging a drive
7	IE, Netscape, E-Mail	
	Searched and Recent Files, Slack	
8	Guest Speaker (Malicious Code Analysis)	
	Network Traffic	HD Analysis
9	Routers, Web Attacks	
	Steganography, Image Authentication	Network Tracking
10	Project Presentations	
	Project Presentations	Final Project

The first lab, Policy Creation, is part of the Ethics and Legal Procedures topic, and develops a first responder's policy for search and seizure. This starts the students thinking about the different situations that could confront them when they hit the second lab. An added twist is each team must use another team's policy when conducting the second lab, First Response. For this assignment, the students must perform a digital first response knowing they will be responsible for the search and seizure of other items present at the scene. The case is information theft via an insider, but does not include a murder or witnesses.

The First Response lab, which is part of the Basic Forensic Science topic, provides the students experience in following a policy that they have not written, and also a differing view on the search and seizure procedure. The second lab consists of conducting a search and seizure, supposing that Dr. X has stolen bomb making secrets and is planning to sell them to an overseas competitor. In this lab, the students must locate and seize all media and other physical evidence related to this fictitious case. Figure 2 shows a crime scene sketch from a typical setup for this search and seizure. In addition to locating the evidence and creating a crime scene sketch, the students also must tag, photograph, and retain the evidence for their chain of custody documentation.

Labs three and four focus on incident response. This is because many of our graduates fill network support positions at military installations around the world. For these networks, 100 percent availability (or as close as possible) is an absolute must. The third lab Live Response, part of the Media Analysis and Network Analysis topics, addresses a live network response, where the machine must remain on and they must determine what has gone wrong and reverse it without loss of service. Specifically, the students must open a secure command line interface and create a network connection to another machine. The students must transfer as much volatile information from the machine as possible as well as logs, registry keys, and anything else they feel is relevant. After the transfer, the results are analyzed. In the past, we have focused on having the students detect more issues with the computer itself rather than as a part of a network. Specifically, we have installed various Trojans, viruses, and rootkits that the students are responsible for locating.

In the fourth lab, Password Cracking (Data Analysis), the students have been locked out of the victim machine, which has been turned off, and must gain reentry by circumventing the computer's security. This includes gaining access to both BIOS and login passwords. In doing so, the students must not only circumvent the computer's security, but they must also provide the least amount of evidentiary harm. In the lecture, we demonstrate and discuss how the BIOS and POST function as well as the mechanisms available to bypass user and administrator BIOS passwords. For the computer itself, we discuss the Windows XP authentication mechanism and the different mechanisms by which to bypass it. The most successful of the different methods has been for the students to use pwdump during the Live

Response and capture the password hashes prior to this lab.

The fictitious scenario, about Dr. X, continues into the fifth lab, Imaging a Drive (Media Analysis), where the students must seize the machine and image the hard drive. For this lab, the students prepare their own drive, and image the machine twice, the first time by pulling the plug, which is the recommended method, while the second imaging occurs after restarting and shutting the machine down. This shows students the number of files an OS touches on startup and shutdown and why pulling the plug is the recommended method.

Following imaging the drive, the students' sixth lab, HD Analysis, requires them to perform the entire process. The students enter a crime scene, perform a first response, seize all evidence in the area and image and analyze the drive and the file system for hidden information (Media Analysis and Data Analysis). The first time the course was run, the drive that the students analyzed was the same as the one found in the evidence machine. A few files were planted on the drive, the network was logged onto and off of under different user names, and a few other normal user behaviors were performed. The analysis of the drive takes the students two weeks to complete rather than one week. The extended time was due to the speed of the searching a 20GB drive. Additionally, the size of the drive made it difficult for the instructor to tailor the image and include enough 'evidence' to make for an interesting search for the students. On the up side, the time required to analyze the 20GB drive did provide the students with the very real experience of how time consuming media analysis is. The feedback from one of the students was that he considered starting up his own Digital Forensics firm until he did this lab in which he learned how much work it can really be.
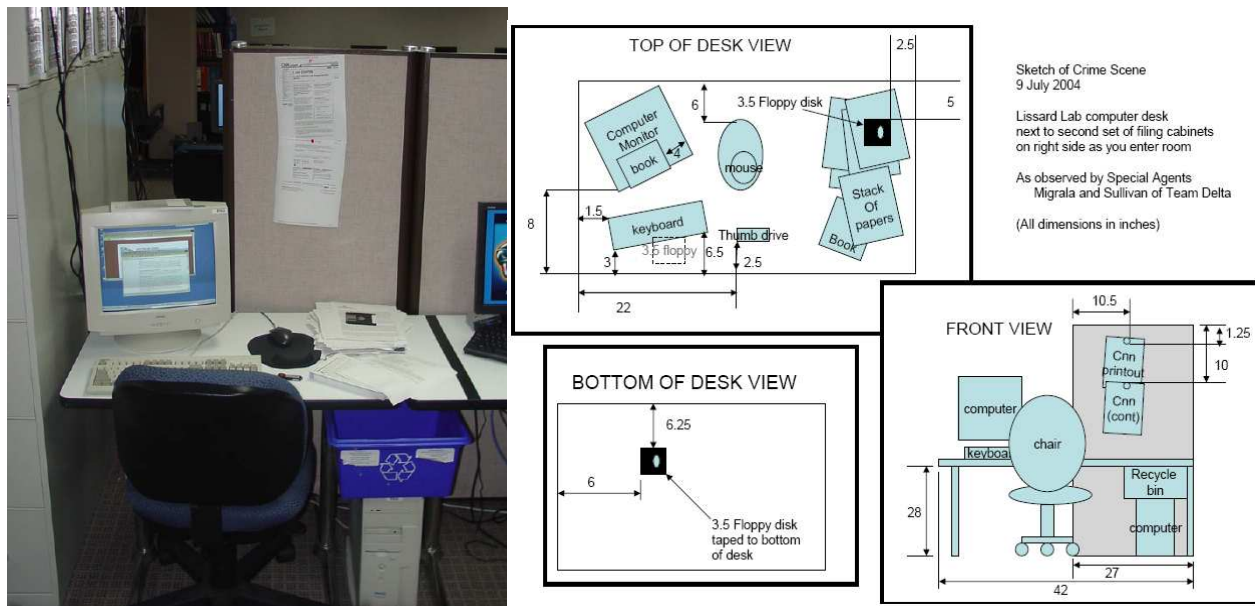

Figure 2. Lab 2 Setup and Crime Scene Sketch.

In subsequent years, students imaged a hard drive and a USB flash drive, and performed the analysis on the USB flash drive alone. Due to the much smaller scale (128MB vs 20GB) it was much easier to hide a larger number of items in different ways and still almost fill the drive space. The resulting image was designed so that no one tool would find absolutely everything. The image itself contained information hidden in all of the different slack spaces, in the boot cylinder, a hidden partition, deleted files, bridging sectors in a reverse order (i.e. the keyword is only locatable by searching in the file), steganography, and very simple cryptography. There was also a compression bomb. If the students were not paying attention to the analysis tools settings while searching the file, the bomb causes the machine to freeze. This approach worked much better, but the students indicate that it is still not perfect because although some of the hidden items point to other items, they are not all set up as a set of 'clues' that lead to some really incriminating piece of evidence.

This year, based on student input, we generated a more in depth back story and sequence of clue breadcrumbs that lead one to the other. The back story is

based on the game of 'CLUE™', for which the objective is to determine who killed Mr. Boddy:

"CEO John Boddy has been found dead in his estate after a friend; Ms. Scarlet called the police department reporting she had not heard from Boddy in some time. It is believed that foul play is involved, and it occurred in the library. We know that there is a computer present as well as a USB thumb drive. It is your task to collect evidence from the library including imaging the drives and examine the USB thumb drive for clues that may lead to information pertaining to the events surrounding his death, i.e. who did it and with what."

The setup for the lab is show in Figure 3. In this setup, the USB thumb drive is in the machine. Underneath the candlestick is a note which says that "I know what you two are up to, you better double check your last message again." On the computer screen is a scrolling message about Mr. Boddy and his company which starts the train of clues, and inside the computer case is a note indicating where the last message is hidden and the password to get through it's encryption, the key to the encryption is found on the thumb drive in file slack. The candlestick, wrench, and the rope (CAT5 cable) are present just to keep the students guessing.

Figure 3: Lab 6 HD Analysis Setup.

In addition to the stronger back story for this lab, we have also refined the steganalysis component, by requiring the students to find the passphrase instead of having it present on the drive. For this, we used JPHide to perform a double hide of data, the first being part of a regular message being passed, the second having the killer (Ms. Scarlet) say they were blaming Col Mustard and Mr. Green, who were swindling Mr. Boddy's company, for Mr. Boddy's death. The detection can be done quite easily using StegDetect and broken using StegBreak.

In addition to all of the ways we hid data from the previous incarnations, and the added requirement to break the steganography, we also added evidence inside of Mozilla Thunderbird.

The seventh lab (Network Forensics) addresses network forensics. Students analyze two days of network capture logs and track individuals attacking the system as far as their ISP. In the past, the network traffic logs have been pulled from the Lincoln Labs Intrusion Detection System Dataset [14]. Due to the datasets statistically normal behavior [15], we are switching to the dataset that the students captured during this year's Cyber Defense Exercise. This dataset provides a much richer, more realistic environment for forensics analysis, and draws from the students experience during the CDX as well. For both of these data sets, only two day's worth of network traffic are analyzed again because of the sheer amount of data. The lab requires the students to use multiple tools to prune the search space before performing a packet by packet analysis to track down the exact attack and exploit packets. The commonly used toolset consists of Snort, Ethereal, and EagleX.

From the Lincoln Labs dataset we use a tcpdump capture of traffic from Monday and Tuesday to and from the machine named 'marx'. In this traffic, there are two attacks, a denial of service and a port scan, and one CGI exploit, back.

The overall lab structure of the course has been changed in two ways since the first offering. These changes provide a better flow as well as challenges for the students. The first change was to lab six. This lab was originally two labs: restoring deleted files, and an analysis. These two labs were converted to one lab at the same time as the move from the 20GB image to the 128MB flash drive. Originally, with the available tools, the undelete process took very little time while the analysis took twice the allotted time.

The second alteration included removing the requirement that the students execute the entire process from the search and seizure, through live response, collection, and analysis as a final lab. This is instead replaced with a final project, and with lab 6, HD Analysis, encompassing the search and seizure, collection and analysis process. The final project allows the students to explore an area of forensics that interests them but may not have been covered in the course. Some of the topics the students have investigated include steganalysis, analyzing anonymous routing networks, such as TOR, wireless network penetration, and compression file cryptography.

This year, the students investigated modifications made to the MS Office 2007 file structure, the applicability of Helix for the entire forensic process, the applicability and challenges that VMWare may provide an

investigator, the detectability of rootkits during a live response, and a drive analysis.

The close integration of the labs and the lectures is one of the most important lessons we have learned, and it's success is evident in the student's comments. Specifically, that the "labs were very good" and that they "learned more in this quarter from this course than I did in my other three courses combined."

4. Other Lessons Learned

Besides the tight lecture and lab integration, we have found that it is very important to invite in speakers who are subject matter experts (SME). These SMEs assist in filling specializations that the instructors may not have. They are also typically actual practitioners who give the students a real life perspective on digital forensics. This year, the FBI Miami Valley Regional Computer Forensics Lab Director came to discuss the law enforcement view on digital forensics. We also had a detective from one of the local cybercrime child pornography unit come and discuss what he does and how it differs from straight media analysis. For the past two years, we have had an expert in malware analysis come and speak about catching and reverse engineering viruses, Trojans, and other software security risks. In the future, we hope to bring in a legal expert to present and conduct a mock trial.

Another aspect that has been very rewarding is our outreach to local law enforcement. Since the course's inception, we have offered to the local law enforcement the opportunity to attend the course without charge. The officers that attend have enjoyed the course, commenting that the level of difficulty with the USB image analysis required more of them than most of the cases that they work on a regular basis. The other benefit is that while they are in the class much like the SMEs, they provide a real world view of the topics in the course.

An additional outreach that has benefited from the digital forensics course is our collaboration with Sinclair Community College (SCC) in Dayton Ohio. Through a grant sponsored by the National Science Foundation, we are partnering with SCC to develop courseware appropriately structured for first-responders attending classes at the community college level. We are in the process of sending a survey to several Chief Information Officers of large corporations in the Miami Valley to garner information on their preparedness to deal forensically with a computer security problem as well as their interest in a course at the community college level. Working closely with SCC faculty, we will use this information to tailor their course to best meet the needs of the corporate and first-responder communities. Our vision is to assist in the preparation of SCC students that will be hired by the area corporations to deal with and understand the ramifications of mistreating possible evidence and how to interface with local law enforcement.

5. Conclusion

Currently, we offer our digital forensics course once a year. We continue to seek ways to improve the course content and make the laboratories as relevant and realistic as possible. Our students' feedback indicates positive learning and a feeling of high value for the course content exposure. We believe education and research in digital forensics is critical to our national security. Our graduates will face many of the issues presented in class in future Air Force and DoD assignments. We hope their digital forensics exposure gives us a distinct advantage over our adversaries be they nation states or malicious hackers.

This impression is echoed by the students themselves who have commented that the course "promotes OPeration SECurity (OPSEC) and COMPuter SECurity (COMPUSEC) which can be passed to even non-technical people." And feel that "as long as crimes can be committed on computers, courses like this will not lose their purpose but may in fact grow in importance."

In our three years of offering the digital forensics course, we are convinced that the integration of the labs and the lecture material is integral to it's success. In addition, having the subject matter experts come and speak provides a real world grounding that isn't always possible in an educational setting.

In the future we will extend the digital forensics offerings at AFIT, adding courses that offer more depth in both the Network Forensics, Digital Device Analysis, and even in the Data Analysis topic areas. Some of the topics in Data Analysis that can be expanded are a more depth coverage of information hiding and its role in steganalysis, metadata, and network protocols.

6. References

[1] Pollit, M. "What is Digital Forensics?', Presentation at DFEWG, Feb. 2004.

[2] Alec Yasinsac, Robert F. Erbacher, Donald G. Marks, Mark Pollitt, Peter M. Sommer: Computer Forensics Education. IEEE Security & Privacy 1(4): 15-23, 2003.

[3] A Road Map for Digital Forensic Research, DFRWS Technical Report, DTR-T001-01, 2001.

[4] Carney, M., and Rogers, M., The Trojan Made Me Do It: A First Step in Statistical Based Computer Forensics Event Reconstruction, *International Journal of Digital Evidence,* 2(4), 2004.

[5] Carrier, B.D., and Spafford, E.H., Defining Digital Crime Scene Event Reconstruction, Journal of Forensics Science 49(6), 2004.

[6] Gladyshev, P., and Patel, A., Finite State Machine Approach to Digital Event Reconstruction, *Digital Investigation,* 1(2), 2004.

[7] Stephenson, P., Modeling of Post-Incident Root Cause Analysis, *International Journal of Digital Evidence*, 2(2), 2004.

[8] Casey, E., *Digital Evidence and Computer Crime*, Academic Press, Cambridge University Press, Cambridge, 2000.

[9] Department of Justice, computer crime and intellectual property section, http://www.usdoj.gov/criminal/cybercrime/searching.html, June 2006.

[10] The American Society of Crime Laboratory Directors, http://www.ascld.org, June 2006.

[11] Mandia K. and C. Prosise, "Incident Response and Computer Forensics, 2nd Edition, MacMillan Publishing, 2005.

[12] National Institute of Justice, http://www.ojp.usdoj.gov/nij/, June 2006.

[13] U.S. Air Force Office of Special Investigations, http://public.afosi.amc.af.mil/, June 2006.

[14] Lippmann, R.P. and J. Haines, Analysis and Results of the 1999 DARPA Off-Line Intrusion Detection Evaluation, *Recent Advances in Intrusion Detection, Third International Workshop*, *RAID 2000*, 162-182.

[15] Mahoney, M.V., and Chan, P.K., An Analysis of the 1999 DARPA/Lincoln Laboratory Evaluation Data for Network Anomaly Detection, *Recent Advances in Intrusion Detection, RAID 2003*.