DFRWS 2005 Workshop Report

Editor

Jessica Reust (Stroz Friedberg)

Contributors

Frank Adelstein (ATC-NY), Brian Carrier (Purdue University), Eoghan Casey (Stroz Friedberg), Golden G. Richard, III (University of New Orleans), Marcus Rogers (Purdue University), Vassil Roussev (University of New Orleans), Wietse Venema (IBM)

Overview

The 5th Annual Digital Forensic Research Workshop was held on August 17-19 2005 in New Orleans, bringing together researchers, developers, and practitioners from around the world to address the emerging challenges in our field. As always, the relaxed atmosphere and collaborative nature of the workshop generated fruitful discussion and exploration of many new ideas. Having a single track with everyone in the same room creates a lively and friendly setting. In addition, the workshop activities were designed around common themes to create continuity throughout the workshop. The sharing of ideas was facilitated by having lunchtime discussion sessions in nearby restaurants, with each group focusing on a particular theme. The proceedings for this workshop are available online at http://www.dfrws.org.

The primary themes of the workshop were: 1) Data Hiding and Counter Forensics, 2) Overload: Scalability & Automation, and 3) Digital Forensics Tools. Accordingly, the

Copyright © 2006 DFRWS. All Rights Reserved

workshop presentations were focused on four pertinent and central topics currently challenging the digital forensic community: 1) Digital Evidence Concealment & Analysis Techniques; 2) Digital Evidence Overload: Scalability and Automation; 3) Digital Forensic Tools; and 4) Digital Evidence Legal Issues. These four topics are not mutually exclusive, and the workshop discussions often encompassed issues and points from a number of the topics. The short presentation format was added to the schedule on the last day, allowing participants to sign up for a slot during the workshop and present works in progress and similar work and receive feedback from the audience.

Two extremely lively panel discussions added to the mix. The "Research Needs of Law Enforcement" echoed many of the main themes, calling for increased speed and intelligence of processing. The Legal panel exploded into a flurry of questions and debate that all participants felt could be expanded in future workshops to help address issues that are being raised in court. In addition, a survey was conducted among workshop participants to identify the most pressing issues facing the field in the coming years. Of the 28 people who participated in the survey, 12 were from academia, 10 from industry, 3 from military, and 3 from law enforcement.

As well as promoting new research, this workshop encouraged hands-on challenges and practical results. To foster the development of practical solutions to a difficult problem, the Forensic Memory Challenge was posted to the DFRWS web site a couple months before the workshop. This challenge comprised of memory images from a compromised machine. Given these memory dumps and very little contextual information, participants were tasked with determining what happened, what tools were used, and to find the answers to several specific questions. The solution to the Challenge was presented at the workshop, including the results of the two winning

Copyright © 2006 DFRWS. All Rights Reserved

entries which represented an impressive amount of work (see www.dfrws.org/2005/challenge/). DFRWS intends to continue this new tradition by setting challenges each year.

Furthermore, the Forensic Rodeo de-emphasized the competitive aspects in order to include more people and give non-experts some hands-on experience in the field. Several desktop computers were provided, courtesy of the University of New Orleans, with a variety of popular commercial and free tools (both Linux and Windows), giving participants a chance to try out different programs. Several of the organizers floated around to help steer the several participating teams in the right direction. In the end, the different groups presented their results, and solved the mystery of the illegal hippopotamus pictures.

The main reception was sponsored by Elsevier and the Journal of Digital Investigation, which features an award for Best Journal Paper of 2004. This award was won by Brian D. Carrier and Joe Grand for their paper "A hardware-based memory acquisition procedure for digital investigations." Several authors donated books that were given as prizes for the best workshop paper (Knut Eckstein), the Memory Challenge (Chris Betz, George M. Garner Jr., Robert-Jan Mora), and the Forensic Rodeo. The Forensic Rodeo had a special prize category for the most creative use of a hex editor and vi as forensic tools, which was jointly won by Phil Turner and Kulesh Shanmugasundaram.

The survey results, presentation highlights, and lunch discussions for this workshop are summarized in this report.

Copyright © 2006 DFRWS. All Rights Reserved

1) Keynote

Wietse Venema delivered the keynote "Forensic Discovery," describing the persistence of data at various levels of abstraction on computers, and speculating about future trends in technology that may impact digital forensics. He pointed out that deletion destroys file system structure not content, and even magnetic tracks that are partially overwritten leave shadow data that may be recoverable. Dr. Venema summarized his testing of data persistence on a dedicated UNIX server, noting that persistence of deleted file content is dependent on file system, activity and amount of free space (a complex relationship). This research is also covered in Dr. Venema's book "Forensic Discovery" (see Porcupine.org/forensics).

Although virtual memory in newer machines may not be as "full" of information as older systems, memory contents can persist. Dr Venema discussed how long data stays in main memory (e.g., processes, terminated processes, kernel memory). Chow's USENIX article demonstrated that there is a trail of data traveling through memory (persistence in kernel memory is longer than in process memory). Although process termination leaves memory after a few minutes (short term memory persistence), fragments can persist and it can be fruitful to dump memory even weeks after an incident.

Dr. Venema compared block cache versus virtual cache – block cache is inserted between disk blocks and file system, and virtual cache is inserted between application and file system. The virtual cache is smarter, is used in most modern operating systems, and the size of virtual cache is dynamic. The persistence of file cache can be high because it stays in memory until replaced. Even though WinXP zeroes out process pages periodically it does not clear the file system cache. RAM

Copyright © 2006 DFRWS. All Rights Reserved

stores cached file system data in block boundaries, identical to block boundaries on disk, which can simplify recovery.

Some actions that system administrators might consider good practice in response to a security breach are actually devastating from a forensic standpoint. Dr. Venema gave an example that making a system backup after an intrusion empties memory and changes last accessed dates/times. Therefore do not perform a system backup after an incident as valuable information is lost.

The value of memory as a source of evidence is quite high because it can contain passwords, decrypted data, and other data that may not be recoverable elsewhere on the system. For example, when a password is typed into Mozilla, there are 6 copies of that password throughout the system. Microsoft Encrypted File System (EFS) can encrypt by file or directory; by directory is superior as the files are encrypted BEFORE they are written to disk. A simple experiment of downloading a text file into an encrypted directory reveals that you cannot view the plain text in the disk image after the download. A question was raised about whether unencrypted EFS content is cached. The answer appears to be yes, with 99.6% of plain text found unchanged (i.e. unencrypted cached in memory). EFS stores multiple copies of data cached in main memory and will also make a plain text temporary backup during the encryption process, which will also be in cache. The fact that EFS encrypts data before writing to disk is good from a security standpoint but challenging from a digital forensics standpoint. Therefore, digital forensics needs to look more at memory where unencrypted data are still cached in main memory even after logging off. The fact that memory contents remain even after logoff is evident from the procedure to dump memory (Log off WinXP and hit "Ctrl Scroll Lock" twice).

Copyright © 2006 DFRWS. All Rights Reserved

Dr. Venema had a few comments on the future of digital forensics – in particular that hardware will become softer due to complexity, which means that we may see more subversion of hardware in the future. For instance, reflashing firmware of hard drive or other peripherals could hide data (or make a hard drive unreadable, essentially making it into a "doorstop"). So, can we trust hardware to tell us what is on it? One question queried the forensic examination of protected devices like Direct TV which have a key that is checked in kernel and other places, and are therefore hard to hack. Dr. Venema responded that Digital Rights Management is good for privacy management, but gets in the way of hacking and forensics. A related question was raised about how secure computing would impact forensics. The answer was that you may encounter a password protected disk from which you are unable to determine the most basic information, such as size, rendering it effectively just a piece of electronics.

2) Digital Evidence Concealment & Analysis Techniques

The survey on data hiding techniques asked "What are the 5 most common data hiding techniques that you encounter?" The most common data hiding techniques as reported by the survey respondents are listed below with the percentage of respondents in parentheses:

- File Renaming (21%)
- Encryption (19%)
- Stegonography (12%)
- Wiping (11%)
- Rootkits (9%)
- File System (5%)
- Removable Media (5%)

Copyright © 2006 DFRWS. All Rights Reserved

2.1) Presentation Highlights

Knut Eckstein presented the research he did with **Marko Jahnke** on "Data Hiding in Journaling File Systems," providing an overview of the file system layers and how data can be hidden in each layer. Journaling file systems have more components with fixed positions on disk than other file systems. Some disadvantages to this type of data hiding are the volatility, the fact that the data can be easily discovered, and the low amount of hiding space. The "fsck" utility on UNIX performs consistency checks inside each metadata category and between categories. Having many hidden blocks in use with no inodes creates a large number of inconsistencies that "fsck" will detect. However, creating one inode for all blocks reduces the number of inconsistencies to one, which is harder to detect. The most effective approach to hiding data within the file system is to find a suitable area on the disk where there is available space and low operating system activity. A demonstration of this technique was presented. It may be possible to detect this type of data hiding by comparing the output of the "du" and "df" commands. Variations of the hiding techniques using inode space rather than data block were suggested but have not been tested.

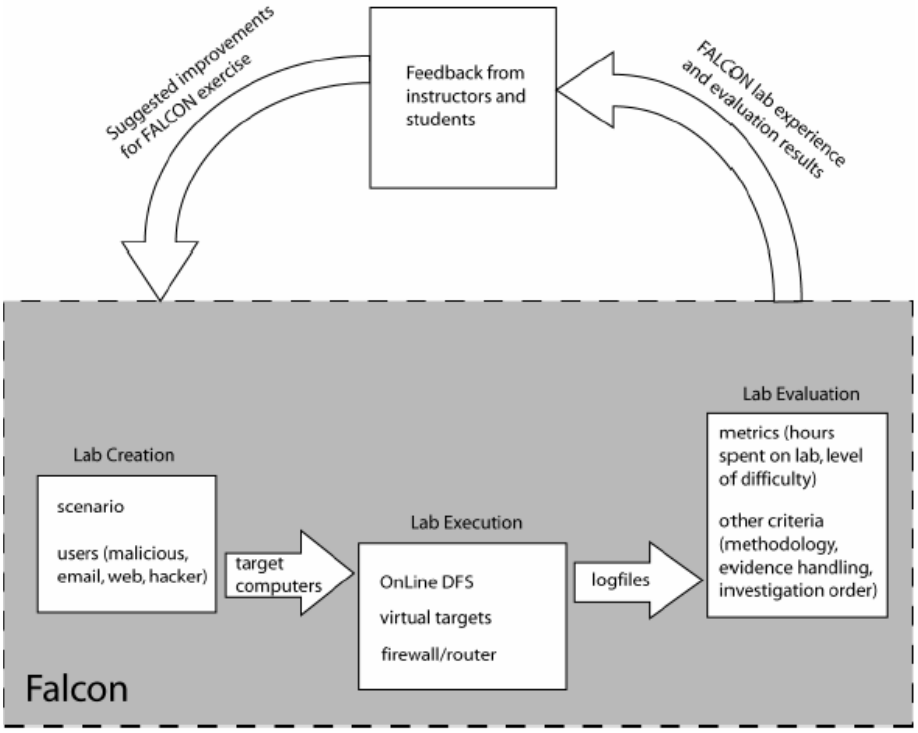
Matthew Geiger presented his research "Evaluating Commercial Counter-Forensic Tools," discussing tools that are designed to eliminate traces of specific records and files. All of the tools that were studied missed some data that could be useful as evidence. The testing of these tools highlighted two broad classes of failures: 1) flaws/bugs in the program; and 2) complexity failures - the inability to keep up with evolving systems and applications. In addition, each tool creates a distinct operational fingerprint on the file system, these fingerprints include the way the tool renames the files it is deleting and the log files the tool generates. Therefore, it may be possible to design a forensic utility to automatically scan a system for these

Copyright © 2006 DFRWS. All Rights Reserved

signatures to discover if these tools have been used. Such a tool is under development as part of this research.

Frank Adelstein presented his paper "Automatically Creating Realistic Targets for Digital Forensics Investigation" (co-authored with **Yun Gao** and **Golden Richard**). The main focus of the presentation was the introduction of a newly developed system, called FALCON (Framework for Laboratory Exercises Conducted Over Networks), designed to facilitate the lab component of teaching digital forensics. The quality of digital forensics classes depends heavily on the ability to present students with realistic cases where they can acquire the necessary skills. Creating a variety of such forensics targets, however, is a difficult and time-consuming task. FALCON is an extensible platform that considerably aids in this process. It provides a high-level specification mechanism (LCT—Lab Creation Tool) that can create lab cases. The specification language allows the instructor to define the target in terms of users, applications, and events of forensic significance. The tool then automatically configures a set of target machines for a lab exercise based on a specification that the instructor provides. Such a specification may include descriptions of both malicious and regular (non-malicious) activity on the target machines. Following that, the Lab Execution Environment (LEE) provides an interactive environment for students to conduct investigations on "live" machines. This part of the system is an extension of the OnLineDFS (developed by ATC-NY) presented in previous workshops. After students complete their work, the instructor can use the Lab Evaluation Tool (LET) to analyze the logs of student actions for certain patterns and to supplement the evaluation process. An important aspect is the ability, based on the activities log, to evaluate the suitability of the task on the target. The feedback cycle for implementing and refining FALCON is depicted below:

Copyright © 2006 DFRWS. All Rights Reserved



A prototype version of the FALCON system has been tested in an actual forensics course at the University of New Orleans with very encouraging preliminary results. Students found the created assignment moderately difficult with time requirements well fitting the needs of a 1-2 week deadline period. Future extensions call for more automation in creating multiple targets and adding more features in the specification language. A question was posed as to how traces of the creation process would be eliminated to give the students a realistic training scenario. Mr. Adelstein commented that the creation process would be designed to minimize such remnants.

2.2) Lunch Discussion

This section summarizes some of the discussions during the breakout session on Digital Evidence Concealment and Analysis Techniques. The conversations were fairly free-form, and the topics represent areas of interest, rather than any group consensus or endorsement.

Copyright © 2006 DFRWS. All Rights Reserved

Because of the potential for hiding data in the NTFS journal, the NTFS journal should be reverse engineered and, more importantly, the metadata must be documented. The alternate data streams (ADS) of the NTFS is another area that needs work. Steganography can be used to hide information in long ACLs. There was some discussion about the difficulties of dealing with steganography and cryptography. The problem is that we do not know what we are looking for. We can detect its presence, but only sometimes. General detecting tools could be used, such as an entropy search tool, and then specific decoding tools used for known formats.

There are many other ways to hide data on a network. In addition to remote storage services like iDisk and Xdrive, free e-mail systems like Gmail and Hotmail provide offsite data hiding for a computer. It is relatively straightforward to write a mail file system that used a mail service to store, retrieve, and modify files, translating them to and from the e-mail format (e.g., GMail Drive – www.viksoe.dk/code/gmail.htm).

The discussion about tools for automatically detecting signatures of counter-forensic programs led into a discussion on the detection of Trojan horse programs and questions about the Trojan Defense. Is the Trojan horse defense ("my machine was compromised so anything you found on it wasn't from me") a problem for the prosecution or defense? In other words, is the burden of proof on the prosecutor to prove that no Trojan was involved or on the defense to prove that a Trojan was involved?

Some new forensic comparison tools would also be useful. One example is representing different disk attribute blocks by color to see if any visual patterns emerge. Lazarus in The Coroner's Toolkit does this to a certain degree, but the data categorization is quite general. Starlight (http://starlight.pnl.gov) can also be used

Copyright © 2006 DFRWS. All Rights Reserved

to model and visually represent file system data. Another example is to write a generic specification comparison tool (such as the JPEG specification) to see if something is violating the specification. This would address syntactic, as opposed to semantic, steganography i.e., if someone violates the protocol to embed information (corrupting the FCT tables in an image) as opposed to something embedded in the content (an image of someone holding a sign with coded information on it).

3) Digital Evidence Overload: Scalability & Automation

The survey on data analysis asked "What are the 5 biggest barriers to processing digital evidence in a timely manner?" The main causes of time-consuming processing as reported by the survey respondents are listed below:

- Data Size (25%)
- Poor Pre-planning (10%)
- Automation (7%)
- I/O Throughput (7%)
- Limited (Tech and Human) Resources (7%)
- Imaging Time (5%)
- Tools (5%)

The survey also asked "What are the 5 largest shortcomings of our data analysis capabilities?" The largest analysis shortcomings as reported by the survey respondents are listed below:

- Speed (14%)
- Automation (12%)
- Data Reduction (4%)

Copyright © 2006 DFRWS. All Rights Reserved

- Encryption (4%)
- Large Volume Capability (4%)
- Data Relationships Not Shown/Known (4%)
- Training for Analysis (4%)

3.1) Presentation Highlights

Erin Kenneally and **Christopher Brown** presented their research on "Risk Sensitive Digital Evidence Collection," which proposed standards for selective acquisitions via a transparent methodology. Risk sensitive digital evidence collection is essentially a shifting of filtering from analysis to collection, resulting in an overall reduction in time required for collection and analysis. This "front-end" filtering is already done in the physical forensic realm, and in e-discovery and intrusion investigations. This type of collection would result in a modification of current forensic techniques when time and resource costs becoming unmanageable.

The authors posited that current methodologies which focus on collecting entire bit-stream images of original evidence disks are increasing legal and financial risks. The aim of the research was to formalize an abstract methodology that the digital forensic community can get behind and support. The methodology would involve on-scene evidence identification and selective evidence extraction from "live" or static systems, where imaging and analysis of a subset of information is warranted due to the specific situation.

The possible legal challenges to this methodology, including defense claims that potentially exculpatory evidence was missed, violating 4th amendment and due process, and their counter-arguments, including the bad faith requirement, the standard of reasonableness, and privacy interests, were discussed. In the absence

Copyright © 2006 DFRWS. All Rights Reserved

of precedent, an e-discovery model analogy using a cost-benefit framework and the concept of "reasonableness" can be utilized.

Brian Carrier presented his and **Eugene Spafford's** research on "Automated Digital Evidence Target Definition Using Outlier Analysis." Target definition is the stage where the analyst formulates searches; the results are called target objects. Compared to other stages such as data preprocessing or data comparison, target definition is relatively human labor intensive, and can therefore benefit from automated support. An example of such support is a tool that takes a file search result and that makes suggestions for additional searches, such as other files with similar names, other files in the same directory, or files with similar content, time stamps, or application types. Outlier detection looks for objects that are hidden or that differ significantly from their direct environment. In a case study, the authors looked at outliers in file size, time stamps, disk block address and application type in the disk image of the 2000 Honeynet forensic challenge, and present many numbers. The authors also make comparisons with results from a two-year old Windows PC that they believe to be both normal and unaffected by spyware, viruses etc.

The presentation by **Chris Bogan** "Preparing for Large-Scale Investigations with Case Domain Modeling" (paper co-authored with **Dr. David Dampier**) addressed the issue of domain-specific modeling as it applies to digital forensics. The general approach is to utilize UML as a basis for describing domain concepts, attributes, and relationships. The paper presents a generic UML model and presents specific instances applicable to different domains. The model consists of concepts, relationships, and attributes. Candidate concepts may be identified by extracting nouns and verbs from case documents such as underlying facts and circumstances, warrants, subpoenas, arrest reports, incident reports, etc. In doing so, it is

Copyright © 2006 DFRWS. All Rights Reserved

important to keep the concept generic enough so that they can remain reusable within the domain as shown in the table below.

Concept Category	Examples
Physical or tangible objects	*Cell phone, Hard Drive, CDR disk*
Descriptions of things	*Marketing Report, Incident Report*
Places	*Home, Street*
Transactions	*Payment, Sale, Money Deposit, Email Transmission*
Roles of people	*Victim, Suspect, Witness*
Containers of things	*Databases, Hard Drives*
Things in a container	*Files, Transactions*
Computer or Electro-mechanical systems	*Internet Store, Credit Card Authorization System*
Abstract noun concepts	*Motive, Alibi, Insanity, Poverty*
Organizations	*Mafia, Corporate Department, Government Organization*
Events	*Robbery, Meeting, Phone Call, File Access*
Rules and policies	*Laws, Procedures*
Records of finance, work, contracts, legal matters	*Employment Contract, Lease, Receipt, Subpoena*
Services	*Internet Service Provider, Telephone Service, Cell Phone Service*
Manuals, Books	*Flight Manual, Explosives Manual*

Attributes are the defining characteristics of a concept, and they represent the information that is essential to the computer forensics examination. As a minimum, the list of attributes should be exhaustive enough to uniquely distinguish between instances of a concept. For the purposes of planning a forensics investigation, the concept names and attributes are the most important items of information; concepts and attributes are the relevant pieces of information that the technician will use to seed the examination plan. However, relating the concepts adds an additional layer of information that can help an outsider understand the background and circumstances of a case. Such a table may be used as a checklist for identifying potential relationships between selected case concepts.

Copyright © 2006 DFRWS. All Rights Reserved

Category	Examples
A is a physical part of B	DVD Drive – Workstation
A is a logical part of B	Network Mapping – Network Intrusion
A is physically contained in/on B	Used CDR Media – CD Case
A is a description for B	Readme file – Executable Program
A owns B	Suspect – Vehicle
A is a member of B	Suspect – Gang
A is an organizational subunit of B	Information Technology Division – Company
A uses or manages B	Systems Administrator – Company Network
A is a specialized version of the generalized B	Systems Administrator – Company Employee
A communicates with B	Suspect – Associates
A is known/logged/recorded/reported in B	Email Registration – Network Logs

In terms of practical implementation of this approach, this is still in the early stages of research. Digital forensics could benefit from other areas of computer science where extensive use of ontologies is the norm, such as medical informatics. Some initial (informal) testing on small cases has been performed with encouraging results. The presented experience suggests that this approach can be used to automate the modeling and create more targeted investigative techniques, such as domain-specific

3.2) Lunch Discussion

This section summarizes some of the discussions during the breakout session on Data Overload: Scalability & Automation. The conversations were fairly free-form, and the topics represent areas of interest, rather than any group consensus or endorsement.

When asked what they considered to be "big" in terms of digital evidence, most of the participant's responses ranged from 20GB to 2TB. However, one participant described a case involving processing 450TB worth of tapes. The task involved unique file discovery and keyword indexing. Ultimately only 6% of the tapes had relevant data. The process cost $250,000 in dedicated equipment, including 75 tape units, 32 workstations and 7 servers, and took 7 people driving 24/7 operation for 60 days.

Copyright © 2006 DFRWS. All Rights Reserved

The biggest overload problems are speed, case turnaround, and bewilderment not only with large drives, but also protected drives.

Approaches for dealing with data overload were discussed. There is a need to find better ways to use what you know before the investigation begins to allow you to focus in on relevant items. Developing the ability to more clearly define an investigation at the point of imaging could help retrieve only the relevant data and might be better able to be automated. In other words, better documentation from the outset delineating what is being sought, what are the priority questions, and what is the expected outcome could help the investigator focus on the most relevant data.

For multi-TB systems, reduce imaging volume could be achieved through the integration of hash sets with imaging tools. There is a need to move "smartness" upstream to the preservation phase. In network forensics, sampling techniques for network traffic capture, as well as better classification of what evidence is, could help reduce the amount of collected data. Another approach is to implement proactive, incremental imaging for corporate environments with distributed systems. Having images already in hand when an incident occurs can give investigators an advantage.

Another idea for dealing with data overload was Artificial Intelligence (AI) for results in niche domains in which it can be truly effective. The aim is not an "AI Digital Forensic Investigator" but rather tightly constrained goals for AI such as Natural Language Processing (NLP) for automated summaries, and "topic search" rather than "keyword search." Picture grouping (both AI and non-AI) could also help; picture

Copyright © 2006 DFRWS. All Rights Reserved

grouping would automatically place photographic images into related "piles" with the criteria for these piles being, for example, "pictures containing humans together" etc.

4) Digital Forensics Tools

The survey on tool capabilities asked "What are the 5 largest shortcomings of the existing tool capabilities?" The largest shortcomings in tool capabilities as reported by the survey respondents are listed below:

- Bugs (15%)
- Lack of Standards (10%)
- Interoperability (8%)
- Automation (6%)
- Formal testing (4%)

4.1) Presentation Highlights

Lei Pan presented a paper titled "Reproducibility of Digital Evidence in Forensic Investigations" (the paper was co-authored with **Lynn Batten**). The authors present a model for digital investigation consisting of: determining input and output layers, assigning read and write operations associated with forensic tools, and time-stamping the read and write operations. Examples were presented that showed how the model would be applied. The layers used are application program, file system, media management, physical media, and networking. The read and write actions span the different layers, and the examples included complex forensic operations, such as running "dd" to image a physical disk and sending the output through "netcat" on one machine, and then retrieving the image on another machine on the net. The different reads and writes are then time-stamped. Their model is designed

Copyright © 2006 DFRWS. All Rights Reserved

to help forensic analysts substantiate the reproducibility of the investigation (i.e., the forensic integrity).

Philip Turner proposed a standard Digital Evidence Bag (DEB) based on traditional evidence capture methods in forensic science (tag, bag, and seal). The proposed evidence unit is comprised of a tag, index and bag file. The tag contains a hash of index to date and other structure, and the application information could be added to the tag. This container would integrate chain of custody and continuity of possession, and would avoid the recalculation of the hash value each time the evidence is transferred. Using a standard format would also make it easier to import evidence into a new tool. The proposed format would be flexible enough to be used for disk images or "tcpdump" captures. The proposed DEB would also support intelligent or partial imaging, allowing practitioners to collect classes of data separately (e.g. text files in one DEB, and JPG in another). This type of selective acquisition is suitable for situations where there is limited imaging time on scene, or when investigators are only authorized to collect specific data from large systems. The discussions that developed as a result of this presentation led to the formation of the Common Digital Evidence Storage Format Working Group (see http://www.dfrws.org/CDESF/).

In presenting his research on "Monitoring Access to Shared Memory-Mapped Files," **Christian Sarmoria** (the paper was co-authored with **Steve Chapin**) addressed a weakness with intrusion reconstruction from system call logging. The problem is that read/write access of memory-mapped files does not show up in system call logging, because memory read/write access does not involve system calls. The authors monitor memory read/write access via the kernel-based memory page fault handler. Their implementation is subject to memory granularity: it does not know

Copyright © 2006 DFRWS. All Rights Reserved

the location within the memory page that the process attempts to access. In addition, the implementation is subject to time granularity: the CPU scheduler twiddles memory page protection bits when it schedules the process, and the page fault handler resets those bits, so that an "event" is triggered only for the first read and write operations of memory-mapped page within a CPU time slice. The paper discusses strategies to avoid race conditions.

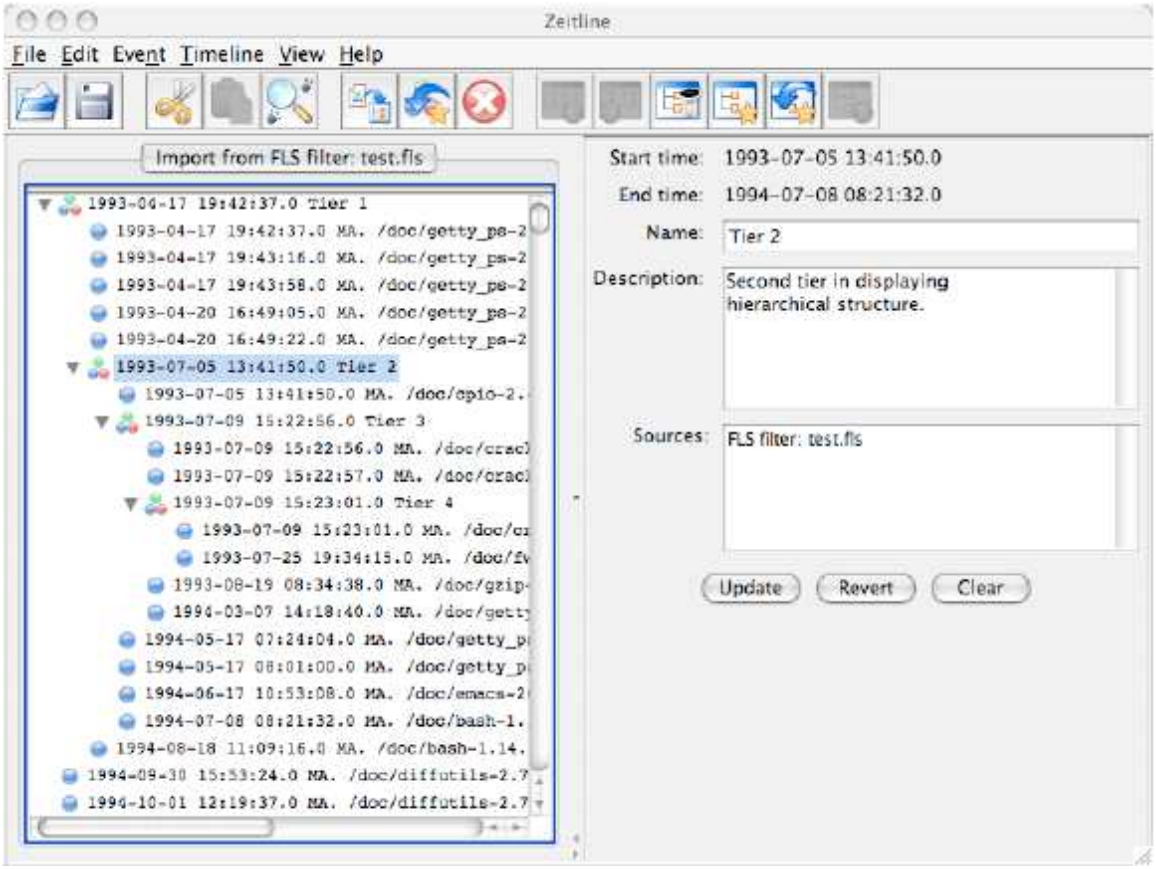
The intention to read or write is recorded when the process traps into the kernel, but the actual read/write operation happens later when the process gets to the CPU. In the mean time, a different process may read or modify that memory page.

Wei Wang presented a paper titled "Network Forensics Analysis with Evidence Graphs" (the paper was co-authored by **Thomas Daniels**). In the paper, the authors present a prototype analysis tool for network forensics that takes primary evidence (such as IDS alerts) and uses a graph-based model to represent the evidence. The analysis modules use local and global reasoning; local reasoning defines state of individual machines (attacker, victim, stepping stone, affiliated), and global reasoning defines the relations among the different machines in terms of the scenario. An example of a complex attack scenario was presented. Their tool was able to take a large number of Snort events (7500) and aggregate them to 4 "hyper alerts" relating to port scans and 17 events relating to exploits. The reasoning module was able to take the events and nodes and determine which machine was the original attacker, and which were stepping stones. This represents the beginning of automatic analysis of network-based forensic evidence.

Florian Buchholz and **Courtney Falk** presented "Zeitline, a Forensic Timeline Editor." This tool focuses on event correlation and reconstruction, providing a

Copyright © 2006 DFRWS. All Rights Reserved

framework for importing and normalizing data from many different sources. Importing and normalization of data is handled using filters. Filters are loaded dynamically so that new filters can be developed by the user without needing to recompile the program. This tool can be used to query events based on a keyword and/or a time range. Events are organized in a hierarchical structure as shown below.



Golden Richard and **Vassil Roussev** presented a tool named Scalpel for carving files based on headers and footers. Scalpel aims to quickly carve files of arbitrary size on machines with modest resources. Experiments indicate that this tool is considerably faster than other file carvers such as Foremost, and somewhat faster than WinHex and FTK. Scalpel achieves this high performance by performing an initial scan of data to be carved, and creating an index of header and footer

Copyright © 2006 DFRWS. All Rights Reserved

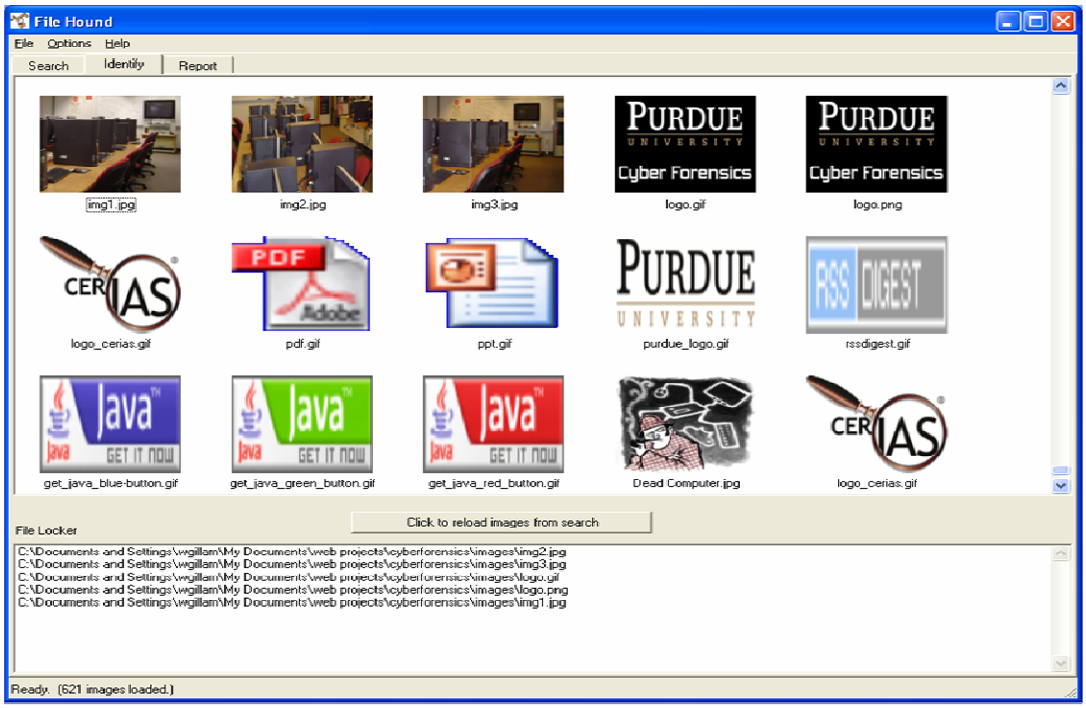
locations. Scalpel then creates work queues and makes a second pass of data to carve out files according to options in the configuration file. The Scalpel configuration file is compatible with Foremost, but the command line options are different as shown here:

```
scalpel version 1.53
Written by Golden G. Richard III, based on foremost 0.69.
Carves files from a disk image based on file headers and footers.

Usage: scalpel [-b] [-h|V] [-v] [-s num] [-i <file>] [-o <outputdir>]
          [-n] [-c <config file>] <imgfile> [<imgfile>] ...

-b  Carve files even if defined footers aren't discovered within
    maximum carve size for file type [foremost 0.69 compat mode]
-c  Choose configuration file
-h  Print this help message and exit
-i  Read names of files to dig from a file
-o  Set output directory for carved files
-n  Don't add extensions to extracted files
-r  Find only first of overlapping headers/footers [foremost 0.69 compat mode]
-s  Skip n bytes in each disk image before carving
-V  Print copyright information and exit
-v  Verbose mode
```

One outcome of the comparisons of file carving tools in this paper is that different tools may produce significantly different output. This problem indicates that further work is needed to create more consistency in file carving methods.

Copyright © 2006 DFRWS. All Rights Reserved

Blair Gillam and **Marc Rogers** presented File Hound, a free, user-friendly tool for law enforcement that enables first responders to search a hard drive for headers of common graphics files (PNG, GIF, JPG, TIF, WMF, BMP). Once the file search is complete, the first responder can use File Hound to view thumbnails of the images, and select images for further review as shown below.



Additional functionality planned for File Hound includes a timeline viewer, hash comparisons, and skin color analysis algorithm.

4.3) Demonstrations and Short Presentations

Sudhir Aggarwal presented an anti-cyberstalking system called "Predator and Prey Alert System" that monitored a victim's system for signs of cyberstalking, alerted the victim of potential concerns, and captured associated evidence.

Copyright © 2006 DFRWS. All Rights Reserved

Eoghan Casey presented a utility developed with **Steve Mead** for scanning media for fragments of known files. This utility is used to locate fragments of deleted files wherever they may be scattered on a piece of media, and is useful in cases involving intellectual property theft, spoliation, and civil discovery.

John Ward presented "STRIKE! - Real-Time Extraction of Actionable Intelligence," which is a tool developed by I.D.E.A.L. Technology Corporation to automate basic forensic examination tasks. This tool is used to quickly determine whether a system contains information that is being sought, effectively performing triage to identify the relevant few systems out of hundreds or thousands of computers.

Kulesh Shanmugasundaram presented work he and **Nasir Memon** did on evidence reassembly of images from raw data blocks on a disk. The data blocks are not in order and the blocks from different files are intermingled. They compared a number of different techniques, with the result that many failed on certain test images. Their final technique succeeded in properly reconstructing all images in most of their test sets. Interestingly, the only place where it failed was in reconstructing a set of "head shots" of soccer players. The algorithm chose the wrong hair (top part of the picture) for a player. The error was not readily noticeable by the audience as the picture "looked" correct; only those who knew what the player's real hairstyle was would notice. And the algorithm chose that match because it "fit." There is much promise to being able to make sense out of the jigsaw puzzle of unallocated disk space. Regardless of whether the results are used as evidence, it can provide valuable clues to aid in the investigation.

Copyright © 2006 DFRWS. All Rights Reserved

4.4) Lunch Discussion

This section summarizes some of the discussions during the breakout session on Digital Forensics Tools. The conversations were fairly free-form, and the topics represent areas of interest, rather than any group consensus or endorsement. Having developers of commercial forensic tools at the same table with users was very productive, and there was discussion of forming a consortium of vendors and practitioners to improve forensic applications.

Bugs in forensic tools are a major concern. There are software design and programming techniques to reduce problems, such as isolating problematic areas of code from running programs. Hiring more experienced software developers may help improve forensic software. It was generally agreed that reverse engineering the file system is challenging, resulting in some error in interpretation.

There was significant discussion about the need for more quality assurance and testing. The NIST CFTT program is critical but testing of software by vendors and NIST have different goals. NIST tests whether the tool does what it is supposed to do in relation to admissibility of evidence. Vendors test whether their tool does what it is supposed to do in different situations (i.e. does it work given a specific function of performance). There was general agreement that formal testing done by vendor is much more extensive than testing performed by NIST. ProDiscover has a set of 500 test cases for every platform they roll out.

Although extensive prerelease testing by the vendor is critical, QA is expensive for vendors. There was some discussion of having grants for vendors to enhance internal testing and quality assurance. Although such grants would help improve QA and testing, one developer pointed out that software is sophisticated in nature, and

Copyright © 2006 DFRWS. All Rights Reserved

they cannot anticipate all uses in all situations, and there are only so many permutations of systems that can be tested by a vendor. Given the diversity of environments in which customers run the software creates issues, it is impossible to fully anticipate and test all combinations of hardware and software. Ultimately, there is only so far tool developers can go to catch bugs before releasing their software.

From the vendor standpoint, once a bug is identified, it is easy to fix. For this reason, communication/participation between consumer and vendor is very important. From a vendor perspective, the users' expectation of performance of the tool is always interesting and helps refine the software. One of the biggest barriers to this communication is that customers cannot provide vendors with the evidence files they are having trouble with, usually due to privacy considerations. Another barrier is that users do not have sufficient information about the problem because of poor error handling in forensic software. There is a need for forensic tools to be more transparent with better error handling - if a tool fails, it should provide the user with details about the failure whenever possible. Although this does not fix bugs, it enables practitioners to observe problems that might otherwise be overlooked. It might be feasible to implement an error log that provides details of the problem to the vendor but does not require the user to disclose evidence or other sensitive information.

A concern was raised that error/audit logs would create fodder for confusing cross-examination. Even if an examiner did not do anything wrong, an attorney could use detailed logs to prolong cross-examination and create confusion and doubt. Optional auditing and debugging logs can facilitate bug reports but can be disabled if there is a concern that too much auditing will complicate testimony unnecessarily.

Copyright © 2006 DFRWS. All Rights Reserved

Everyone was in favor of standards, defining the behavior we expect from the tools. For instance, there is a need for a standard evidence storage format, import/export format (XML, CSV), sharing of bookmarks and hash sets, and definitions of terms and processes. There was debate as to who should develop standards (users, vendors, or a consortium). The consensus was that the forensic community should set the standards, not NIST; NIST could help and then set up a test lab. Vendors raised concerns about the costs of implementing standards and pointed out that most vendors want to do more than they have the resources to pay for. A debate arose about who would pay for the functions that a subset of users requires. The idea of grants for forensic tool developers was raised again – providing a source of external funding for standards implementation or for functions that the government wants.

Data interoperability between forensic applications is key, and there is a need to foster that as a community. Standardization in evidence storage formats could facilitate interoperability. The practitioner perspective is that they like the different views provided by the different tools. One forensic tool developer (AccessData) stated their philosophy – a tool will be used as long as it provides the best information. A tool needs to be interactive with other products, and be consistent with the standards of core elements of tools, and the standards developed within the industry.

Finally, the need for more automation was discussed, including high-level scripting of tools to automate routine processing and queuing up tasks for a tool to run without user input. More advanced tools (expert systems) are also needed.

Copyright © 2006 DFRWS. All Rights Reserved

5) Panel – Law Enforcement Needs

This panel was made up of **Marc Ortiz** (Kenner PD & Gulf Coast Computer Forensics Lab), **Todd Shipley** (SEARCH), and **Phil Turner** (QinetiQ), and was moderated by **Eoghan Casey**. The panel was asked to describe the most pressing needs of law enforcement to encourage developers and researchers to create solutions and tools.

The need for speed is the most pressing requirement – getting what is useful out of large datasets in a timely manner. There are more computer-related cases with growing amounts of data, making it more difficult to obtain results in a timely manner. There are also more diverse digital devices like GPS, pen drives, and handhelds. On top of that, the more we educate law enforcement to preserve digital evidence, the more requests digital forensics labs receive. As a result, delays are increasing and law enforcement is becoming less inclined to collect digital evidence – what is the point if it will just delay their work? This trend raised concern among participants, leading to a discussion about how to deal with the increasing case load of computer examiners. The use of distributed systems to process data more quickly and support collaboration on a case could help deal with the increasing amount of data. Also, establishing a pyramid approach could alleviate the strain on forensic labs – have local LE do the basic data processing, and push the more advanced analysis to the labs.

It was emphasized that extensive use of the command line does not work for law enforcement or court, and it is necessary to put decent front-ends on tools. There was general consensus that tools needed to be more user-friendly, and operate with fewer hiccups (i.e. be more stable and reliable). Interoperability between tools was also identified as a critical need. There is also a need for tools to aid the presentation of what is found, and a lot of effort needs to go into the GUI /

Copyright © 2006 DFRWS. All Rights Reserved

presentation for the public to understand. One participant asked what kinds of presentation tools/techniques would help. Ideally, law enforcement needs CSI-like capabilities – show everything important to the case in an hour. Tools to aid presentation of evidence need to be flexible as different jurisdictions might want different visualizations and capabilities. The ensuing discussion indicated that developers do not understand specifically what type of presentation of complex data and digital forensics issues would be useful - they need better goals and training on what is needed and acceptable in presentations.

Additional tools for special purposes are needed, such as Registry analysis. Also needed are more intelligent data carving tools to recover deleted and partially overwritten data, including when headers and footers are not available (recovering and analyzing fragments of files). The need for live system analysis tools that are more geared to criminal investigation was hammered home by Todd Shipley's case example of a drug deal gone bad with RAM capture and analysis – a homicide with chat logs in RAM showed who the victim went to meet to get drugs. Because encryption is increasing in use, there is a need for ways to identify whether whole disk encryption is being used, and perhaps databases of Rainbow tables for more applications. In addition, because criminals are increasingly utilizing cell phones (e.g., drug dealers, trophy photos of murders and rapes), there is a need for improved tools for cell phone data extraction that support forensic investigation of devices in a uniform and reliable manner.

Marc Ortiz highlighted the need for more research on Trojan horse programs to help investigators understand their capabilities. Also needed are automated tools for detecting signatures left behind by counter-forensic programs, and perhaps a way to supplement typical virus scan with information on the virus (e.g., library links).

Copyright © 2006 DFRWS. All Rights Reserved

Phil Turner pointed out the need for more intelligent language processing in unallocated space. The need is for more than strings, the need is for tools that identify whether there is anything in a block of data worth analyzing further. Mr. Turner also stated that audit features are needed to track examiner activities and find better ways to perform examinations and provide training (e.g. FALCON). A tool developer mentioned that some investigators are resistant to logging their actions because they do not want an audit of their mistakes. One participant opined that forensic examiners should have nothing to hide, even if it is embarrassing that it took a long time or mistakes were made, and that audit helps show that the examiner is beyond reproach. In response, Todd Shipley pointed out that auditing everything has risks in the US criminal justice system which tends to attack the person and process rather than the evidence. However, Mr. Turner pointed out that automated auditing takes the load off of examiners note taking. It was agreed that the digital forensics tools should have optional audit functions to satisfy both needs.

Todd Shipley added that there is a need for more forensic research and knowledge sharing about application and OS analysis. Currently, detailed knowledge about operating systems, file systems, and applications is restricted to individual examiners. Law enforcement can benefit from an understanding of how applications work, and what remnants are left on computers, and tools for investigating them as well as automated detection tools (e.g., VoIP, Skype, wireless artifacts, counter-forensic tools, P2P tools, Trojan horse programs, etc.). One participant pointed out that it is difficult for tool developers because vendors are reluctant to release this information to the public. One suggestion was to have vendors disclose information to law enforcement, and develop separate tools for law enforcement.

Copyright © 2006 DFRWS. All Rights Reserved

6) Legal and Certification Issues

This panel was made up of **Mohamed Battisha** (University of Louisville), **Mark Wettle, Esq.** (Attorney at Law), and **Joseph K. West, Esq.** (Entergy Corporation), and was moderated by **Michael Losavio** (University of Louisville) and **Marcus Rogers** (Purdue University). The panel was asked to provide input on issues related to legal and certification issues surrounding digital forensics in general, and more specifically, scientific evidence considerations.

To bound the discussion each panelist was asked to comment in the context of the Federal Rules of Evidence Rule 702 and *Daubert* considerations. The discussion began with Mr. West reviewing the basics of the *Daubert* considerations to ensure that everyone in the audience was familiar with the topic and context of the discussion. Mr. West further expanded his discussion to cover the Kumho ruling that extended the Daubert considerations to include engineering and technical evidence.

Mr. Wettle discussed his experiences in the courtroom and indicated that currently it is a struggle to educate the judiciary on the idea that digital evidence testimony is scientific in nature and therefore *Daubert* should apply. Mr. Wettle indicated that in criminal law, judges are not comfortable with computer forensics and digital evidence.

Mr. West indicated that civil law was starting to see more digital evidence and that judges were starting to deal with the question of admissibility but were still not well versed in scientific issues.

There were several very interesting questions tabled from the floor. Of particular interest was a question related to timelines; if an examiner on the stand says Person

Copyright © 2006 DFRWS. All Rights Reserved

X logged on at this time, how would the opposing attorneys approach that statement?

The answer provided by the panel was to determine: 1) Is it relevant (most of the time this is 'yes') and 2) Is it reliable, and for this, it is the methodology that would be important.

It was noted by many in the audience and by the panel, that examiners are currently making statements on the stand and getting away with them without having conducted enough of an investigation to support these statements. In general, judges and counsel tend not to challenge the veracity of the statements related to the accuracy of computer and digital evidence.

This led to an animated discussion related to the need to conduct an investigation using the scientific method (how did the data get there etc.) to come to conclusions. This need to investigate may therefore make the testimony into a "science" (move from an artifact to an opinion) hence the courts should filter the testimony through the Daubert considerations.

Additional discussion by the panel members focused on the requirement that when science is involved, it would be useful to be able to calculate the "probability" of an activity having occurred using the technical evidence.

The panel then further expanded on the notion of challenges to digital evidence. From a *Daubert* challenge perspective, potential methods to attack evidence and expert witness testimony were categorized as attacking the:

Copyright © 2006 DFRWS. All Rights Reserved

- Science,
- Technique
- Methodology
- Analyst

The panel agreed in principle that digital evidence and computer forensics as a forensics science need to be treated by the courts as equal to the other forensic sciences. Mr. Battisha spoke to the requirements of the other forensic sciences and indicated that specific DNA techniques have been challenged; similar challenges could occur in digital forensics.

The topic of data persistence was raised and how the courts, both civil and criminal, would deal with this kind of evidence and testimony. The panel responded that while it has been proven that software can recover deleted data, the challenge, if any, would and should focus on the completeness of recovery. This provided a smooth transition into a discussion surrounding the "blackbox" nature of proprietary vendor tools and issues related to the understanding of what is happening under the proverbial "hood." What are the underlying assumptions of the software? If an examiner doesn't know, is it really the software that is doing the analysis as opposed to the examiner? One possible solution is ensuring that the findings of one tool are correlated with the findings from another tool or tools.

On a side note Mr. Wettle related anecdotally that one judge he had argued before indicated digital evidence was related to computers and therefore the business records exception applies; hence Daubert does not, therefore there can be no challenge to the admissibility.

Copyright © 2006 DFRWS. All Rights Reserved

A further question from the audience related to whether electronic evidence alone was sufficient to acquit or convict an individual? The panel's response was again a consensus and that the answer was no. All agreed that supporting evidence was needed and digital evidence was relegated to a corroborative role. This response was qualified though, as judges and jurors develop more of an interest in electronic evidence, they may be more willing to assign it greater weight.

By way of concluding the discussion, the panel was asked to consider and respond to one final question - "Is *Daubert* an issue in digital forensics going forward?"

The general consensus was yes, there is always going to be a debate. Law evolves slower than technology; the CSI effect is and will continue to play a role in judge's and juror's assumptions and expectations. However the evolution of law is slow and laborious and all panelists doubted it will keep up with the dynamic nature of technology.

Mr. Wettle spoke specifically to criminal law and stated that at the state level, if the state introduces the digital evidence, it will continue to be admitted regardless of FRE 702/*Daubert.*

Mr. West indicated that with civil law digital evidence is becoming more commonplace, therefore Daubert would be a factor.

The panelist further agreed that the formulation of an independent body should be a priority for the scientific community.

Copyright © 2006 DFRWS. All Rights Reserved

Overall the panel provided some excellent insight into the legal domain of digital forensics and challenged the attendees to not only consider the science, technology, engineering, and mathematical issues of the field, but also those issues related to the law and judiciary.

Copyright © 2006 DFRWS. All Rights Reserved

7) Closing Comments and Future Plans

Attendee feedback was overwhelmingly positive, lauding the relaxed atmosphere and multidisciplinary nature of the workshop. Having an audience that was supportive rather than hostile encourages discussion and exploration of new ideas. Having people from diverse backgrounds was interesting and productive. The panels were effective, and the legal panel in particular was fantastic. It is important for the legal viewpoint to be strongly represented at a workshop dedicated to forensic science, and the topics raised by the legal panel could be expanded in future workshops to delve into certain aspects more deeply. One suggestion was to expand the legal aspects to include a mock trial with an actual judge presiding.

A single track is preferred to multiple tracks because having everyone in the same room encourages multidisciplinary discussion and exchange of ideas. This avoids stove piping, and increased connections and synthesis of presentations and ideas as the workshop progresses. The common themes in papers helped create continuity throughout the workshop. The Forensic Rodeo was fun and educational, providing hands-on experience and giving participants a chance to see how others reason through an investigation. Having teams is better because it enables knowledge sharing between more and less experienced individuals. The breakout lunch discussions were enjoyable and productive, allowing people to get to know each other, exchange ideas, and discuss issues relating to the theme of the workshop.

Based on suggested improvements, there are several plans for improving future Digital Forensics Research Workshops. Having the proceedings published by a "quality" established publisher, even if only in soft form, is necessary for academics. It is still desirable to have the papers/proceedings freely available online without

Copyright © 2006 DFRWS. All Rights Reserved

subscription. Forensic software developers should be encouraged to participate in conferences such as DFRWS to improve communications between users and vendors, and set realistic expectations on both sides. More hands-on activities like the Forensic Rodeo, some earlier in the day rather than late at night. Several weeks before workshop, announce tools needed for hands-on activities like Forensic Rodeo to allow participants time to try the tools out. Release Forensic (Memory) Challenge much earlier (10 months) to allow time for people and classes to work on problem, and allot more time during the workshop to review the Forensic (Memory) Challenge results. More dinners or have a hospitality suite to encourage people to stay together in the evenings to develop conversations.

Logistics:

- Schedule workshop earlier so that it is not right before the semester.
- Adding a day to allow more time for papers and activities.
- Better audio/visual setup with clip-on microphone and larger projection screen.
- Prompt speakers to finish up when they are going over allotted time, but allow some flexibility to remain informal. Allow 10 minutes before lunches for post session/panel discussion and questions.
- More laptop friendly but not too laptop friendly.

Copyright © 2006 DFRWS. All Rights Reserved