Challenges to Consensus and Consistency in DF Education

Garfinkel, Simson L.
Monterey, California. Naval Postgraduate School

http://hdl.handle.net/10945/44329
Challenges to Consensus and Consistency in DF Education

15 minutes

Simson L. Garfinkel
Associate Professor, Naval Postgraduate School

Michael McCarrin,
Research Associate, Naval Postgraduate School
Challenges to
Consensus
and
Consistency
in DF Education
“Consensus” — general agreement?
Who is setting the digital forensics education standards?

Are standards set by the educators?
- Challenge — Undergraduate vs. Graduate programs
- Challenge — Background of teachers

Are standards set by employers?
- Challenge — There are many different employers with many different missions
  - Government DF law enforcement investigator (Federal, State, Local)
  - Government Information Assurance employee
  - Government Intelligence employee
  - E-discovery professional
  - DF “researcher.” (Reverse engineering malware? Decoding data structures?)
  - Consulting Firms
“Consistency” — Conformity to a standard
There are many different ways for programs to conform…

High Level / Strategic:
• DF concepts
• Domain being examined
• Purpose — what are the exercises supposed to teach?

Low Level / Tactical:
• Specific DF skills
• Specific tools
• Specific techniques and methodologies
• Exercises
  — Use the same exercises?
  — How current are the exercises?
  — How complex are the exercises?

Challenge: DF is a huge domain
The NSA Center for Academic Excellence requirements list 42 topics in 5 forensic categories

**Digital Forensics**
- Legal Compliance, Applicable Laws, Affidavits, How to Testify, Case Law, Chain of custody, Digital Investigations, E-Discovery, Authentication of Evidence, Chain of Custody Procedures, Metadata, Root Cause Analysis, Using Virtual Machines for Analysis

**Host Forensics**

**Device Forensics**
- Mobile Device Analysis, Tablets, SmartPhones, GPS (must include hands-on activities)

**Media Forensics**
- Drive Acquisition, Authentication of Evidence, Verification and Validation, Hashes, Metadata, Live vs. Static Acquisition, Sparse vs. Full Imaging, Slack Space, Hidden Files/clusters/partitions (must include hands-on activities)

**Network Forensics**
- Packet Capture and Analysis, Intrusion Detection and Prevention, Interlacing of device and network forensics, Log-file Analysis, Forensic Imaging and Analysis (must include hands-on activities)

**Challenge: This is a lot of stuff!**

Source: [https://www.iad.gov/NIETP/](https://www.iad.gov/NIETP/)
The NSA guidelines focus on the low level/tactical employer-centric skills:
- Operating system structures
- Tool use.

As a community, do we need to teach other skills?
- Tool testing (e.g. understanding tool bias & error)
- Tool extending (e.g. seeing data that the tool misses)

What about “research?”
- Decoding new file formats
- Finding new vulnerabilities
- Discovering new approaches for forensic analysis
  — *Research is necessary for many investigations.*

Does consensus and consistency mean all programs must have the same rigor?
Example tool choice challenge: Use proprietary or open source tools?

Employers want students *trained to use proprietary tools*

- Graduates should be able to start work on day 1
- *Favors proprietary tools in classroom (FTK, EnCase)*

- Proprietary tools have challenges:
  - *Cost of tools*
  - *Installation & licensing*
  - *Maintaining relevance (tools & exercises go out of date fast)*

Some educators favor *open source tools*

- Teaching underlying concepts
- Approaches that are vendor agnostic
- No license hassles
- *Natural gateway to research*
- *But few employers use open source tools...*
Tool Choice Challenge #2: Teach lots of tools, a few tools, or how to build tools?

There are hundreds of current digital forensics tools.
- Open source & proprietary

Do we teach students:
- How to use a few tools well / How to know them deeply?
- How to find the “right” tool with a web search?
  - How to learn to use a tool on their own?
  - How to validate the tool?
  - How much trust to put in the tool?
- How to extend a tool?
  - e.g. write RegRipper plug-ins
- How to build a tool that they need for a particular case?

Challenge: different DF customers have different needs & approaches.
Exercise challenges:
It’s hard to make exercises and keeping them current.

What’s the goal of exercises?
- Learn how to use the tool?
- Learn to “think like an investigator”
- Learn about real-life complexity?

What’s in a typical exercise?
- Running a tool and reporting the result?
  - *Getting a specific result*?
- Describing a forensic process?
- Exploration? Telling a story based on the evidence?

How do you validate an exercise?
- Exercise difficulty?
- Is there a trick? Or just a lot of work?
- Does each step build on the previous step?
- Does the student need to understand what the tool is doing to succeed?
Exercise Challenge:
How much “extraneous” data do we include?

In real forensic investigations, most of the data are irrelevant.

Forensic case data evolves over time.

Forensic case data has depth other than the incident.

- Good exercises are time-consuming to build.
- Does extraneous data detract from learning objectives?
These are even harder to build.
But they reflect actual cases.
Creating Exercise Challenge: Keeping the exercise Current

Exploits:
- Need to prevent security patches from “fixing” the VM
- Students object to learning exploits that don’t work anymore
- Are we teaching how to run the exploit or to analyze?
  — *Current exploits are too difficult to teach in most courses*

Encryption:
- Current systems are trending towards encrypting everything.
  — *Increasingly the encryption can’t easily be cracked.*
- Do we turn off encryption?
- Do we make encryption the focus of digital forensics?

THIS

Not this
Challenges in distributing corpora and problem sets

Size — Corpora are big
  • Hosting the data
  • Downloading — some schools are challenged to download large data sets.
  • Durability (persistence) of the download location

Support
  • Many of the educators who download the exercises need help!

Problem sets:
  • Are problem sets for a specific tool?
  • How do we restrict answers to teachers?
    — *We used a password-protected PDF file… But distributing the password is a burden.*
What do we want to teach from related fields?

Forensics increasingly involves traditional computer science:

• Machine learning (clustering & classification)
• Natural language processing
• Large data set analysis
• Visualizations

Should these skills be part of the “consensus” digital forensics curriculum?
Or should different programs have specializations?