CuFA: a more formal definition for digital forensic artifacts

Vikram S. Harichandran, Daniel Walnycky, Ibrahim Baggili, & Frank Breitinger

Graduate Research Assistant & UNHcFREG Member

Presenting @ DFWRS 2016, Seattle, WA, USA

University of New Haven
Cyber Forensics Research & Education Group

UNHcFREG

UNHcFREG

?

Problem

- Usage of the term “artifact,” or “artefact,” varies in the digital forensics and cybersecurity domains. No work thus far has attempted to formalize this.

- This is important for the community to have efficient communication and sharing.

Previous work: Linguistic definitions

- Sources:
  - Merriam-Webster Dictionary, Oxford Dictionary, CybOX™ project, SWGDE/SWGIT, and papers with explicit definitions

- Commonalities:
  - Artificiality/external force
  - antecedent temporal relation
  - Exceptionality (accidentally procured, rare, or an individual's particular interest)
  - Legality and science (forensics)

Previous work: Perspectives & usage

UNHcFREG

Items	Category	Paper & perspective
User credentials, personal details, activities, location; Activity timestamps Images;	Databases Media	Azfar et al., 2015 Researcher
Opened/saved files; Email attachment; Skype log (chat & transfer); Index.dat (downloads);	Files	Goh, 2014 Researcher
User assist (program location); Last executed files by app; Run command executed; App compatibility cache; Taskbar jump list; Prefetch/service event logs;	Program execution	
Opened/saved fields; Last executed files by app; Recently opened files; Shellbags; Shortcut files (LNK); Taskbar jump list; Prefetch files; IE history files;	Files created & opened timeline	
Search assistant/history; Keywords search from Start Menu; Last executed files by app; Hidden files in dir (Thumbs.db); Recycle bin; IE history files;	Deleted files	
Current system timezone; Network history, IE cookies; Time website visited;	Physical location	

Previous work: Ontologies & schemas

- Ontologies
  - STIX/DFAX *[1]*: Criminal, investigator, machine, etc.
  - DESO *[2]*: Superclasses/subclasses; Location vs. type
  - UCO

- Schemas
  - XIRAF
  - DFXML
  - CybOX™

[1] Casey, E. Back, G. & Barnum, S. (2015). Leveraging CybOX™ to standardize representation and exchange of digital forensic information. *Digital Investigation*, 12, S102-S110.
[2] Brady, O. Overill. R. E., & Keppens, J. (2014, September). Addressing the Increasing Volume and Variety of Digital Evidence Using an Ontology. *JISIC* (pp. 176-183).

Previous work: Archival science

- Cyber forensics and archival science have similarities:
  - Procedures for acquiring, authenticating, and preserving items
  - Minimize alterations and document unavoidable changes
  - Easy retrieval of items for future analysis

- Current inadequacy with cyber forensics:
  - Often no long-term data preservation and maintenance policy beyond physical storage
  - Solution to this is the concept of curation

Survey: Methodology & demographics

- 54 Likert scale, 12 free response, and 4 multiple choice.

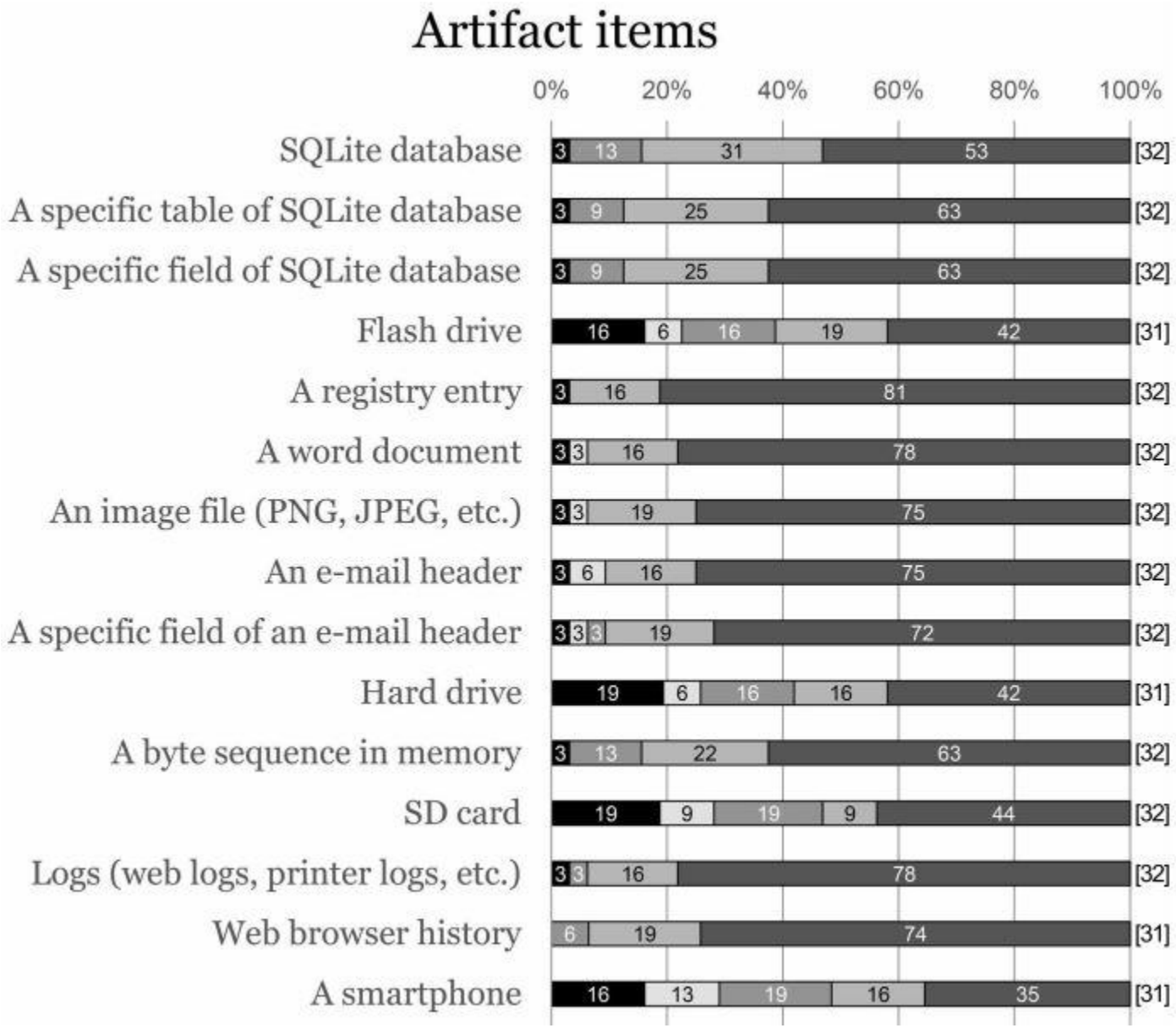
- 87 – 50 (only answered demographics questions) = 37 respondents.

- More than half of respondents were:
  - Americans
  - 7 or more years of experience
  - Older than 34

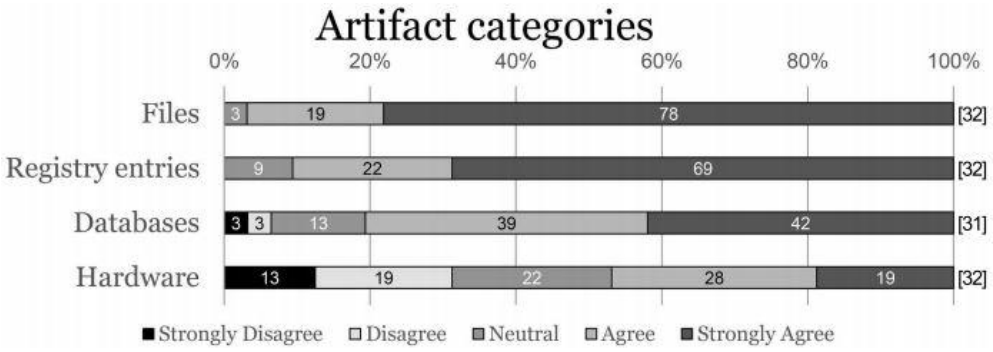
Survey: Themes

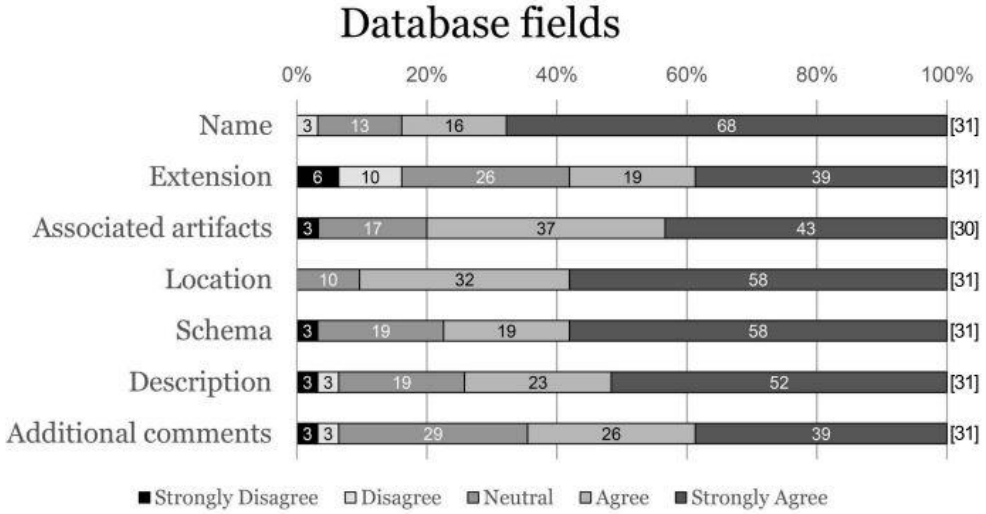
- Frequent responses when asked to define “artifact”:
  - “Evidentiary value”
  - “Applying digital forensic (analysis) techniques”
  - Not hardware

- Most common categories respondents mentioned:
  - Files
  - Network packets
  - Memory/memory dumps
  - Application data


Artifact items
0%
20%
40%
60%
80%
100%
SQLite database 3 13 31 53 [32]
A specific table of SQLite database 3 9 25 63 [32]
A specific field of SQLite database 3 9 25 63 [32]
Flash drive 16 6 16 19 42 [31]
A registry entry 3 16 81 [32]
A word document 3 3 16 78 [32]
An image file (PNG, JPEG, etc.) 3 3 19 75 [32]
An e-mail header 3 6 16 75 [32]
A specific field of an e-mail header 3 3 3 19 72 [32]
Hard drive 19 6 16 16 42 [31]
A byte sequence in memory 3 13 22 63 [32]
SD card 19 9 19 9 44 [32]
Logs (web logs, printer logs, etc.) 3 3 16 78 [32]
Web browser history 6 19 74 [31]
A smartphone 16 13 19 16 35 [31]

UNHcFREG


Artifact categories
0%
20%
40%
60%
80%
100%
Files 3 19 78 [32]
Registry entries 9 22 69 [32]
Databases 3 3 13 39 42 [31]
Hardware 13 19 22 28 19 [32]
Strongly Disagree
Disagree
Neutral
Agree
Strongly Agree


Database fields
0%
20%
40%
60%
80%
100%
Name 3 13 16 68 [31]
Extension 6 10 26 19 39 [31]
Associated artifacts 3 17 37 43 [30]
Location 10 32 58 [31]
Schema 3 19 19 58 [31]
Description 3 3 19 23 52 [31]
Additional comments 3 3 29 26 39 [31]
Strongly Disagree
Disagree
Neutral
Agree
Strongly Agree

Survey: Investigative procedure

1. Acquire (identify which tool the item of interest came from).
2. Backup.
3. Check database to see if encountered before (compare hashes or fields).
4. If familiar, do quick search in database to see if methods used previously are still applicable/effective. If they are, use them, then jump to step 8. If not, continue to next step.
5. Classify into a category using ontological model and catalog/extract taxanomic fields (used in schemas).

Survey: Investigative procedure (cont.)

6. Attempt to use techniques effective for the category. If ineffective, repeat steps 4-6 until so.
7. If no effective techniques are encountered try reconstruction to see if item can be recreated or reverse engineered.
8. After a technique is successful in analyzing, repairing, isolating, or rendering item harmless, document the process (with fields) and create a report.
9. Examine the system for associated items (type or follow pointers).
10. Prepare reports of each (type of) item to support legal case.

Proposal

- Definition:
  - Must be curated via a procedure which uses forensic techniques
  - Must have a location in a useful format (when applicable)
  - Must have evidentiary value in a legal proceeding
  - Must be created by an external force/artificially
  - Must have antecedent temporal relation/importance
  - Must be exceptional (based on accident, rarity, or personal interest)

- Why is location “optional”?

Location type (original source of creation)

- User (e.g. using a text editor application to create a text file)
- Application (e.g. log/database file created by an application to store user information)
- System (e.g. registry file or altercation created by the system via a process/application)
- Download (e.g. package of files or executable in stand-alone form before installation)
- Network (e.g. packet in transit which has been captured)

CuFA requirements

- Name
- Description
- Comments
- MD5/SHA1/MRSHv2
- Person(s)/time of entering into database
- Person(s)/time of discovery
- Location type (original source of creation)
- Enabled/disabled
- Location (specific source, inherited from CybOX if applicable)
- Pointers to other related artifacts found because of this artifact (implemented as linked list)
- Object type (inherited from CybOX)
- Device
  - Manufacturer
  - Model
  - OS
  - Type (PDA, mobile, laptop, server, don't know/external)

CybOX object (examples below)

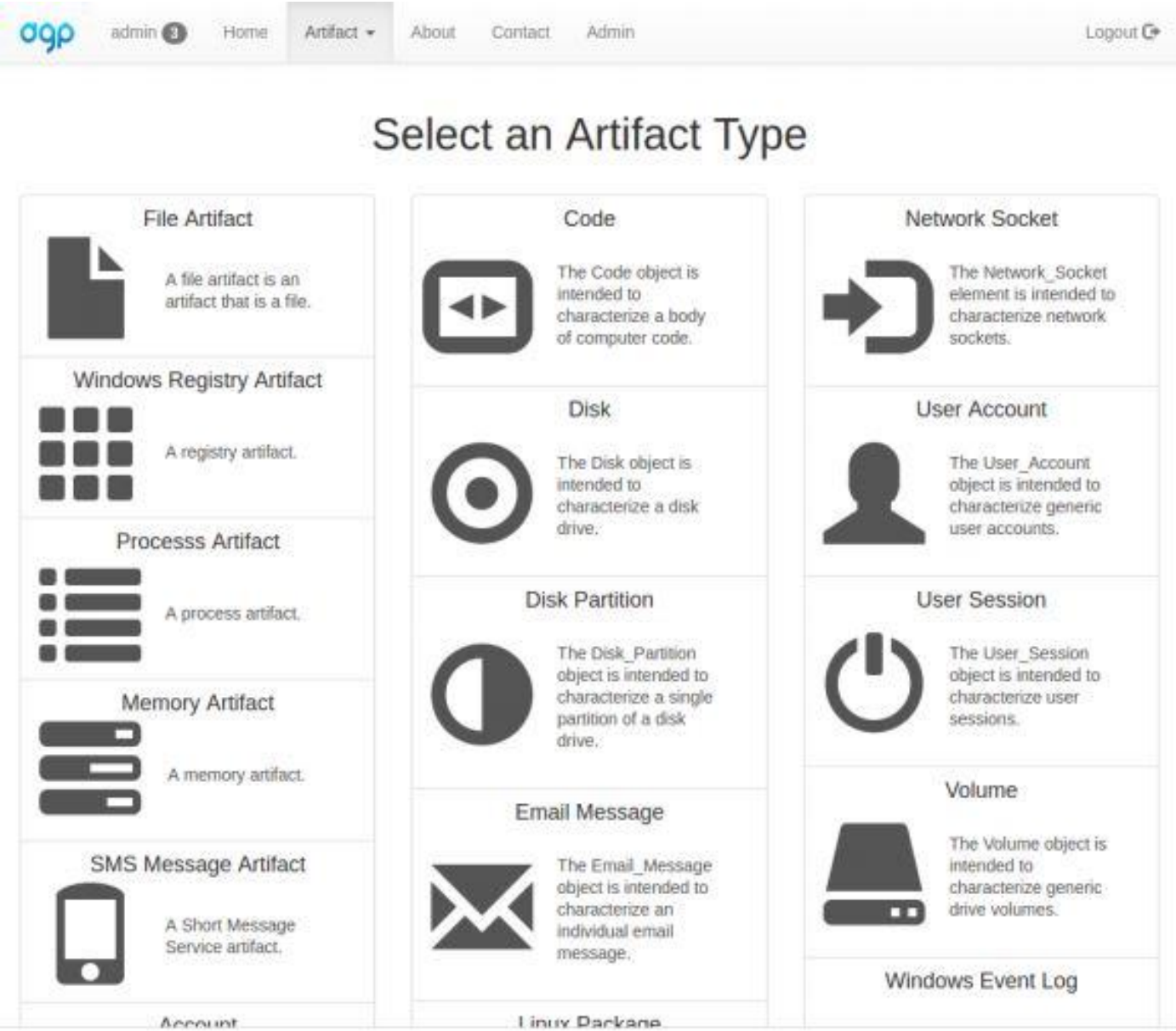
- File
  - Device_path
  - Full_path
  - File_extension
  - File_format
  - Modified_time
  - Accessed_time
  - Created_time
  - File_attirubutes_list
  - ...
- Process
  - Name
  - PID
  - Parent_PID
  - Child_PID
  - Username
  - User_time
  - Start_time
  - Status
  - ...
- Win registry
  - @object_references
  - Key/hive
  - Number_values
  - Creator_username
  - Handle-_list
  - Subkeys
  - Byte_runs
  - Custom_properties
  - ...
- Archive file
  - Version
  - Encryption_algorithm
  - Full_path
  - File_extension
  - Size_in_bytes
  - File_format
  - Digital_signatures
  - Hashes
  - ...
- Network socket
  - Address_family
  - Domain
  - Local_address
  - Protocol
  - Remote_address
  - Type
  - @is_blocking
  - @is_listening
  - ...

Contribution

- Our work proposes:
  - A linguistic definition that emphasizes curation
  - An ontological model that can be paired alongside one or more other ontologies/schemas

- Our work on the Artifact Genome Project (AGP), a database allowing the study of the evolution of CuFAs, similar to the Human Genome Project, helped inspire this work.

AGP



UNHcFREG

Conclusions

- Most importantly, we need discussion! The community should have one mind. Three-letter agencies should be involved.
  This should allow people involved in different steps of the process to communicate better (e.g. judicial, police, investigators, etcetera).

Future work

- Survey limitations:
  - Small sample size
  - Unclear scope for questions allowed different interpretations (disregarded)
  - “Decline to respond” option

- Tools to support curated databases, e.g. OSXAuditor/OSXCollector, in addition to plugins that could automatically triage a system through plugins.

Questions

Vikram S. Harichandran

(vhari2@unh.newhaven.edu)

Thanks to Google for the scholarship,

Department of Homeland Security for funding (Award Number 2009-ST-061-CCI001-05),

the other authors,

and the AGP developers for additional input.