Digital Investigation 12 (2015) S102–S110

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

DFRWS 2015 Europe

Leveraging CybOX™ to standardize representation and exchange of digital forensic information

Eoghan Casey*, Greg Back, Sean Barnum

The MITRE Corporation, 7525 Colshire Drive, McLean, VA 22102-7539, USA

Keywords:
Digital forensics
Standard representation
Digital forensic ontology
Digital forensic XML
CybOX
DFXML
DFAX

A B S T R A C T

With the growing number of digital forensic tools and the increasing use of digital forensics in various contexts, including incident response and cyber threat intelligence, there is a pressing need for a widely accepted standard for representing and exchanging digital forensic information. Such a standard representation can support correlation between different data sources, enabling more effective and efficient querying and analysis of digital evidence. This work summarizes the strengths and weaknesses of existing schemas, and proposes the open-source CybOX schema as a foundation for storing and sharing digital forensic information. The suitability of CybOX for representing objects and relationships that are common in forensic investigations is demonstrated with examples involving digital evidence. The capability to represent provenance by leveraging CybOX is also demonstrated, including specifics of the tool used to process digital evidence and the resulting output. An example is provided of an ongoing project that uses CybOX to record the state of a system before and after an event in order to capture cause and effect information that can be useful for digital forensics. An additional open-source schema and associated ontology called Digital Forensic Analysis eXpression (DFAX) is proposed that provides a layer of domain specific information overlaid on CybOX. DFAX extends the capability of CybOX to represent more abstract forensic-relevant actions, including actions performed by subjects and by forensic examiners, which can be useful for sharing knowledge and supporting more advanced forensic analysis. DFAX can be used in combination with other existing schemas for representing identity information (CIQ), and location information (KML). This work also introduces and leverages initial steps of a Unified Cyber Ontology (UCO) effort to abstract and express concepts/constructs that are common across the cyber domain.

Introduction

In the modern age, any type of investigation can have a digital dimension, ranging from computers as a source of information in homicides and terrorist attacks, to computers as instrumentalities of fraud and cyber-attacks. As a result, digital forensics supports decision makers in various domains, including law enforcement, incident response, malware analysis, cyber threat intelligence, and situational awareness. To combat crime effectively in the modern age, digital forensic information needs to be represented and shared in a form that is usable in any of these contexts.

When investigating a single incident, being able to combine the results from multiple tools that are used to extract information from the digital evidence supports forensic reconstruction, including timeline creation and link analysis. In addition, being able to automated the comparison of similar results from multiple tools facilitates dual-tool verification. When crime spans borders, sharing of information between investigative agencies in multiple jurisdictions

* Corresponding author. Reception Desk ext. 3-6004.
E-mail address: ecasey@mitre.org (E. Casey).

http://dx.doi.org/10.1016/j.diin.2015.01.014

is crucial for a successful resolution. A fundamental requirement in digital forensics is to maintain information about evidence provenance as it is exchanged and processed, to help establish authenticity and trustworthiness.

Furthermore, without a standardized approach to representing and sharing digital forensic information, investigators in different jurisdictions may never know that they are investigating crimes committed by the same criminal. A similar challenge was recognized in traditional investigations of violent crime, and led to the development of the U.S. Federal Bureau of Investigation's Violent Criminal Apprehension Program (ViCAP) and Royal Canadian Mounted Police's Violent Crime Linkage System (ViCLAS). These programs collect distinctive details about unsolved violent crimes in disparate regions, and correlate this information to find links between related crimes.

Current efforts to manage and exchange digital forensic information are typically ad hoc, inconsistent, and limited in sophistication and expressivity. Combining results from different tools into a consistent format can be a laborious process that can result in errors or omissions. For example, using Excel to import and format data from various sources can result in items such as date-time stamps being altered, entries not being imported, and other problems that negatively impact forensic analysis.

Where standardized representations of digital forensic information are used, they are typically focused on an individual portion of the overall digital forensic process (Flaglien et al., 2011). Such focused efforts have benefits, supporting in-depth exploration of specialized domains such as file systems, but do not support broader representation and analysis. In addition, existing formalized representations of digital forensic information do not integrate well with each other, or lack coherent flexibility and semantic structure. Existing information sharing activities are often human-to-human exchanges of unstructured or semi-structured descriptions of digital forensic artifacts and analysis, and often require conversion from proprietary formats. For instance, individual forensic examiners document their findings on personal blogs, and share parsers in proprietary formats such as EnCase's EnScript.

To address these issues, this work aims to formalize and extend the management and direct machine-to-machine exchange of progressively more expressive sets of digital forensic information using fully-structured data. Specifically, this paper describes a community-driven solution to address this problem, which leverages the Cyber Observable eXpression (CybOX) language (http://cybox.mitre.org). CybOX is an open-source, community-driven effort to develop a standardized representation of digital observables led by the U.S. Department of Homeland Security (DHS) office of Cybersecurity and Communications. CybOX is designed to represent digital actions and objects along with their context, which can be leveraged in a wide variety of use cases, including incident response, intrusion detection, and digital forensics. Development of CybOX has occurred under the coordination of the DHS-funded and MITRE operated Systems Engineering and Development Institute (SEDI), a Federally Funded Research and Development Center (FFRDC). Thus, MITRE manages the CybOX website, supports community engagement, and oversees its discussion lists to enable open and public collaboration around CybOX with all stakeholders.

This paper proposes a new standard for representing and exchanging digital forensic information called Digital Forensic Analysis eXpression (DFAX) that leverages CybOX for representing the purely technical information. DFAX incorporates its own structures to represent the more procedural aspects of the digital forensic domain, including those for chain-of-custody, case management, and forensic processing. A related effort has already been accomplished in the development of the Structured Threat Information eXpression (STIX) language to represent cyber threat information (Barnum, 2012). STIX makes use of CybOX to represent technical cyber threat details, e.g., malicious IPs, domains, and file hashes, and adds other constructs to represent domain-specific information such as campaigns and threat actors.

The capture of general criminal justice related information has been considered in other efforts such as the National Information Exchange Model (www.niem.gov), EVIDENCE Project (2013), and FIDEX (NFSTC, 2010). However, there is a need in this space to accommodate more than just the criminal justice application, and as such, DFAX proposes general elements to cover all use cases.

The ontological view of DFAX is depicted in Fig. 1, showing where CybOX fits. As is clearly shown, at a high-level DFAX covers information about various roles involved in digital forensics, various actions these roles take, evidence records resulting from forensic actions, and domain specific concepts such as authorizations as well as various abstractions to lend context to roles and actions. Actions in particular play a significant role in DFAX. A `Forensic Action` is defined as any action performed on or resulting in an `Evidence Record`. DFAX also defines `Subject Action` and `Victim Action` that can describe associated digital traces.

This work also introduces and leverages initial steps of a Unified Cyber Ontology (UCO) effort to abstract and express constructs that are common across the cyber domain, and that can be leveraged for consistency and broad-scope interoperability by various domain specific languages, including DFAX and STIX. Two examples of these abstractions leveraged in DFAX are `Action Pattern` and `Action Lifecycle`. The `Action Pattern` construct enables the contextualization of a given `Action` instance as to what sort of behavior it may represent. The `Action Lifecycle` construct can be adapted to define phases of a forensic investigation (e.g., documentation, preservation, examination, analysis, presentation) and criminal activities such as a sexual predator's grooming of victims or a network intruder's method of operation (e.g., kill chain phases). This generalized approach can be used to classify each action in a case, which provides context to support further analysis.

This paper starts with an overview of existing work related to representing digital forensic information, and then describes how CybOX can be leveraged and extended to represent digital evidence, relationships between objects, and actions associated with digital forensic information. Use cases for structured digital forensic information are discussed, and examples are presented to demonstrate how DFAX provides a layer of domain specific information overlaid on CybOX. This paper includes links for community