TOWARDS THE DEVELOPMENT OF A FORENSICS MODEL FOR TRACKING CYBER CRIMINALS WITH REGARD TO THE EMERGENCE OF COVID-19 PANDEMIC

BY

Onyemauche, U.C.: Department of Computer Science, Federal University of Technology, Owerri, Imo State, Nigeria; E-mail: uchenna.onyemauche@futo.edu.ng/ osigwe.uchenna@yahoo.com

Abstract

The field of digital forensics is concerned with finding and presenting evidence sourced from digital devices, such as computers and other digital devices. The complexity of such digital evidence is constantly increasing as is the volume of data which might contain evidence. The exploration of digital evidence sourced from diverse devices becomes paramount in this era of Covid-19 as many Youths are idle looking for ways of bringing solace to their souls. The software engineering methods adopted for study are the Design Science Method and Object Oriented Analysis and Design Methodology with Resource Description Framework (RDF) data Model. The representational approach is used as a foundation of a novel analysis technique which uses knowledge based system to correlate related events into higher level events, which correspond to situation of forensic interest. The PHP/MYSQL programming language was used in developing the software and implementing the system. The new integrated digital forensics system for tracking cyber criminals will eliminate the problems observed in the physical and conventional digital forensics system by hypothesizing a general and formal representational approach which will benefit digital forensics by enabling timestamp which are pervasive to be attached to event logs. Overall, the steps to actualize reliable digital evidence follows the sequence: Event Sequence, Event Construction and finally Digital Tracking.

Keywords: Digital Evidence, Forensics, Forensics Toolkit, Sleuth kit

Introduction

Computer forensics emerged in response to the escalation of crimes committed by the use of computer systems and or other related digital devices either as an object of crime, an instrument used to commit a Crime or a repository of evidence related to a crime (Stephenson, 2017). This study has the tendency to give emphasis to highly efficient regimens in cyber threat, cyber related crimes, finger printing analysis, and criminal investigative analysis (profiling) that can result in capturing serial killers, fraudsters and other perpetrators of homicide most especially this time youth are idle because of the emergence of covid-19 pandemic. According to Sommer (2016), the first computer crimes were recognized in the 1978 Florida Computer Crimes Act which included legislation against the unauthorized modification or deletion of data on a computer system. In the time past, computer evidence means data from storage such as hard drives and floppy disks, captures of data transmitted over communication links, emails and log files generated by operating system. Forensic science is the scientific method of gathering and examining information about the past. This is especially important in law enforcement where forensics is done in relation to criminal or civil law. Johann (2000) stressed that forensics can also be carried out in other fields, such as astronomy, archeology Brown (2006) opined that computer forensic is the art and science of applying computer science to aid the legal process. Although plenty of science is attributable to computer forensics, most successful investigators possess a nose for investigations and a skill for solving puzzles, which is where the art comes in Computer Evidence Collection and Preservation. The central point of reference of every type of digital investigation is irrefutably the concept of digital evidence.

Research Objectives

The overall objective of this study is to develop an integrated digital forensics model for the provision of a reliable evidence that will help track criminals in this era of COVID-19 pandemic and thereafter.

The objectives are:

(a) To investigate the existing products and proffer solution to the inherent problems of volume and complexity.
(b) To examine artifacts from heterogeneous domains and assign roles to them.
(c) To design an integrated model with resilience to low memory condition using UML and RDF data model.
(d) To implement a novel model that can withstand hostile acts and influences from allied tools.

Review of Related Literature

Carrier (2012) sees digital evidence of an incident as any digital data that contain reliable information that supports or refutes a hypothesis about the incident. Thus, some definitions focus on the investigative process and the support that evidence can provide to hypothesis validation while others deal with the probative value of electronic data in the legal context. Schatz (2007) has identified three basic properties of digital evidence, namely latency, fidelity and volatility. Latency refers to the fact that a digital encoding in the form of binary data needs additional contextual information on how it should be interpreted. Fidelity is a property of digital data that allows a copy of it, assuming the verification of the integrity of such a process, to be equally treated as the original one. This is especially important in the Digital Forensics area where access of the original data must be restricted to exceptional circumstances only and be performed by competent personnel. More so, the volatile nature of digital evidence affects considerably the practice of acquisition and further processing of it due to the fact that its authenticity can easily be disputed except if proper and up-to-date procedures are always applied. Although, the focus of the current study is on the latent nature of evidence and how this can be enriched with semantic content, fidelity and volatility of evidence are quite important so as to be encapsulated into two commonly-referred forensic principles, Chain of Custody and Order of Volatility.

According to Onyemauche (2017), some of the best known computer forensic tools, such as EnCase, the Forensic Toolkit (FTK), and the Sleuth Kit (TSK), are basically file analysis tools. They can be used by a forensic examiner to analyze individual files, as well as discover deleted and hidden files on a target file system. File analysis tools like EnCase, Forensics Toolkit (FTK) and the Sloth Kit (TSK) were all designed to facilitate an exhaustive interactive search of a computer hard disk. However, as hard disk sizes increase dramatically, an exhaustive interactive search may be too time and resource intensive to be practical. Recognizing this fact, Beebe et al., (2019) proposed the application of data mining techniques to reduce human processing time of large datasets. Such techniques could be incorporated into a file analysis tool, or perhaps into a computer profiling tool such as the one described here. The functionality provided by existing file analysis tools is distinct from the functionality provided by a computer profiling tool.

Kruse (2019) opined that the representation used to model events has a significant impact on the usability of correlation approaches, including conceptual expressiveness, extensibility, ease of integration of new information and maintainability. The MODEL language, a component of the DECS network management system, used an object oriented (OO) style model of classes of events related together in class/subclass relationship (which in this case was referred to as semantic generalization).

Methodology

Research methodology shows the procedure to be taken in order to achieve the objectives of the research study. This shows how relevant information about the study would be sourced and also the tools and techniques that would be applied in the analysis. Two key methodologies have been adopted in this study and a Data Model:

1. Design Science Methodology
2. Object Oriented Analysis and Design Methodology
3. RDF Data Model

The goal of the first step in Fig. 1 is the clarification of the practical problem situation, its precise definition as well as well-supported motivation on why a solution to this problem is needed. Research strategies such as case studies and action research and research methods such as questionnaires and observations can be applied for controlling the activity and its results. In the current study, domain knowledge obtained through extensive literature review and empirical experience has been the main resource for the formulation and elaboration of the practical problem. Case studies of digital investigations involving a variety of data sources (network captures, event logs, file-system forensics), in the context of international forensic contests and workshops, have been studied in order to better understand the problems rising due to the lack of advanced data integration and correlation techniques. The second step of the methodology has the goal to identify and outline an artifact as a possible solution to the previously defined explicated problem and further define the main requirements of the artifact to be developed. In the next step, the developed artifact is used to demonstrate if and how it can solve aspects of the previously stated problem in the context of an illustrative or real-life case. The proposed method along with the developed instantiation artifacts of this study, are validated through applying them for the digital investigation of representative cases as those used as case studies in the previous research steps when enough data are given or through experiments attempting to resemble realistic and probable cases where data and evidence are superficially generated. The goal of the next phase is the evaluation of the proposed artifact and the solution it provides to the original practical problem along with the level of fulfillment of the identified requirements. The final step is the communication of artifact knowledge where information about the proposed artifact is communicating to other researchers and judicial practitioners. Lastly, we are able to come up with a model of the solution domain that is a detailed description of how the system is to be built.

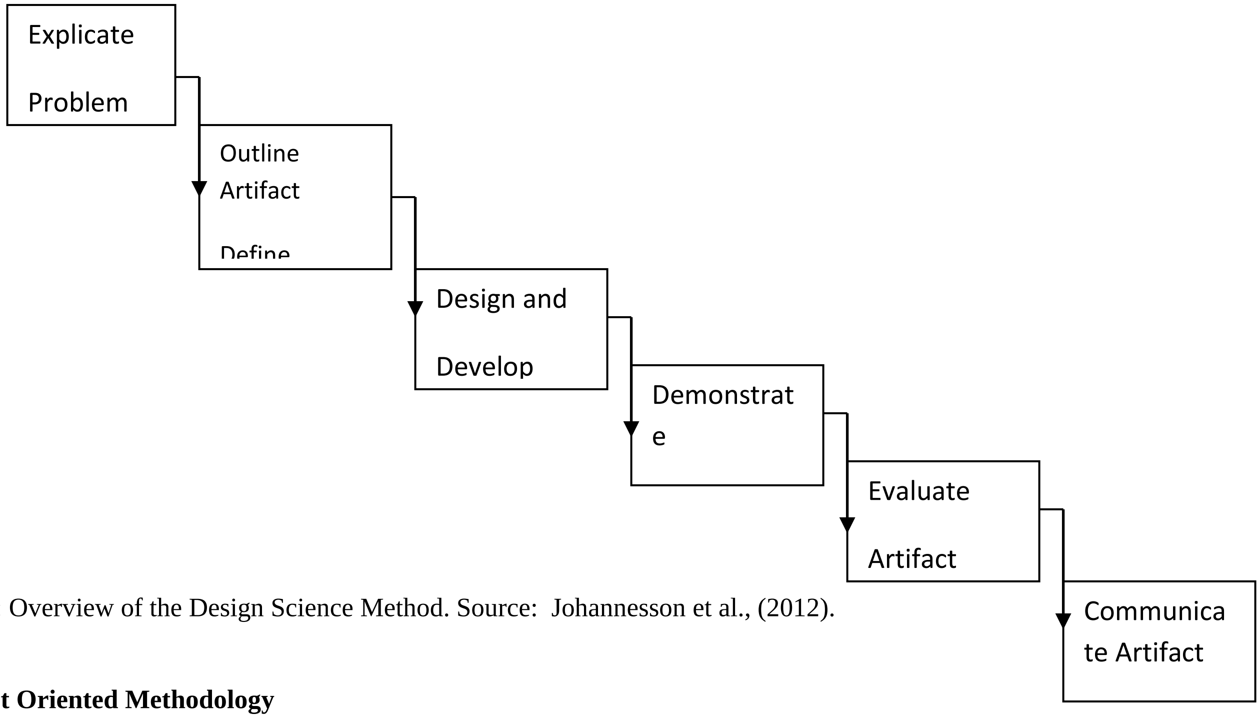

Fig. 1: Overview of the Design Science Method. Source: Johannesson et al., (2012).

Object Oriented Methodology

This model makes use of six attributes: class, object, state transition, interaction, module and process. Object Oriented Programming (OOP) enables one to consider a real life entity as an object. In OOP, one creates a basic structure of a program and keep extending the functionality of the program. The object programming models the real world more accurately than the conventional or procedural approach. In OOP, objects are independent of each other and maintained separately. One can make modifications on the required objects without affecting the functionality of others. Examples of OOP languages are C++, Visual Basic, .Net languages for example PHP. The Semantic Web has evolved over the last decade to a complex aggregation of different technologies each one responsible for different aspects of the Semantic Web framework. Antoniou and Van (2014) have depicted the semantic web architecture using a layered approach as shown in Figure 2.. Such an approach enables better understanding on what are the main functions of the different technologies as well how the layers relate to each other and share results. Since the Semantic Web is built on top of the existing Web architecture, URIs provide the foundation on which the other layers are based upon. Berners-Lee et al., (2005) sees URI as a compact sequence of characters that identifies an abstract or physical resource. URIs are extensively used in the Semantic Web so as each resource described can be uniquely identified and referenced. URIs enable unique identification of a resource under a global scope and a consistent interpretation thus alleviating the problem of different resources represented by the same name in different contexts thus introducing interoperability problems. It is important to note that a URI does not necessarily imply access of the resource as in the more familiar Uniform Resource Locator (URL) scheme but can simply be used for denoting a resource. Unicode is a standard promoting a consistent encoding and representation of text, supporting most of modern writing systems, thus enabling support for multi-lingual environments. Internationalized Resource Identifier (IRI) is a form of URI that can support characters out of the ASCII character set and thus are more useful in modern Web's internationalized context.

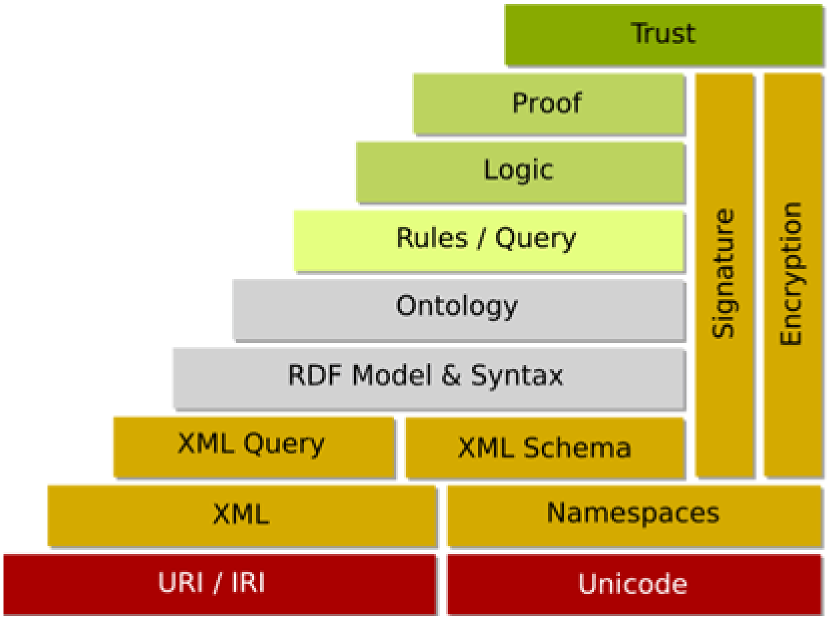

Fig.2: RDF Data Model

Results and Discussion

Steps to actualize Digital Evidence Reality in Criminal Tracking during COVID-19 pandemic and hereafter.

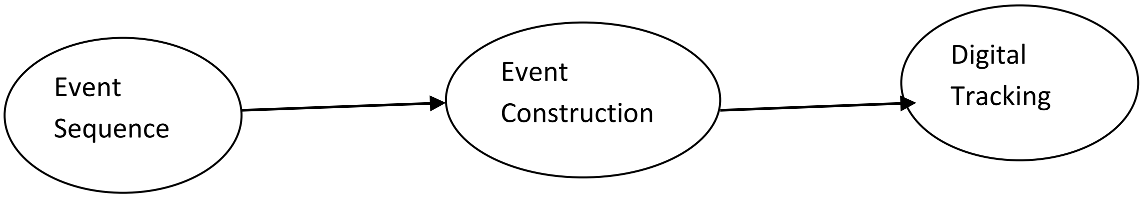

Fig 3: Steps to Actualize Digital Tracking

One can construct and exchange directed graph models of arbitrary complexity as depicted in Fig.3. One began by saying very simple things, such as "John Smith is the Author of the document whose URL is http://www.bar.com/some.doc". (We use a notation where Nodes are represented as ellipses, arcs as arrows, and strings are given in rectangles). As an example of sequences, one will look at some of the works written by John Smith. Since John is rather prolific, one could use sequences to keep lists of his works sorted by publication date, or according to the alphabetical order of the subject of the article. This is illustrated in Fig. 4.

This assertion can be modeled with the directed graph:

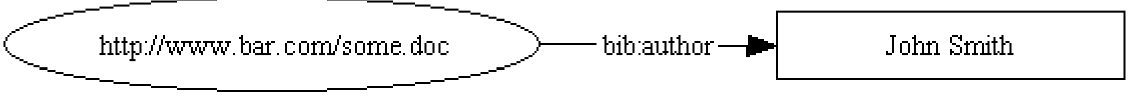

Fig.4: RDF Data Model (Event Sequencing)

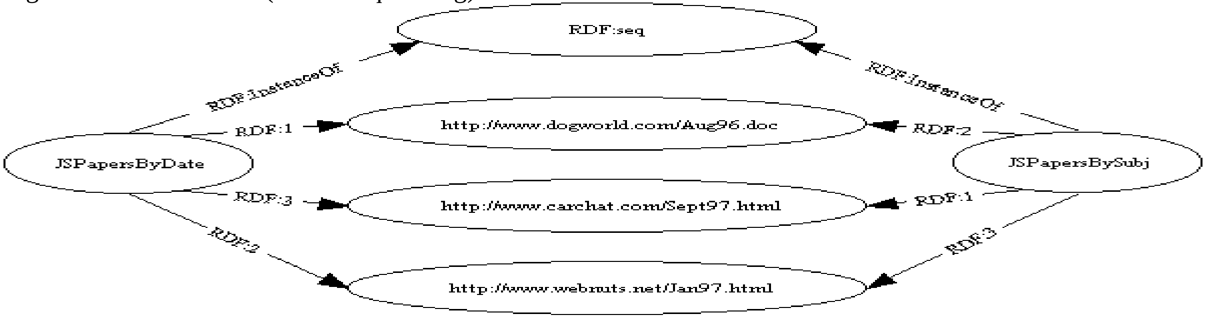
RDF:seq
RDF:InstanceOf
RDF:InstanceOf
RDF:1
http://www.dogworld.com/Aug96.doc
RDF:2
JSPapersByDate
JSPapersBySubj
RDF:3
http://www.carchat.com/Sept97.html
RDF:1
RDF:2
RDF:3
http://www.webnuts.net/Jan97.html

Fig.5: RDF Data Model (Example of Sequence)

RDF Data Model (Event Construction)

One needs to create a more elaborate model as presented in Fig.6 in order to say additional things about John Smith, such as his contact information. One might construct the model which could be exchanged using the XML serialization representation.

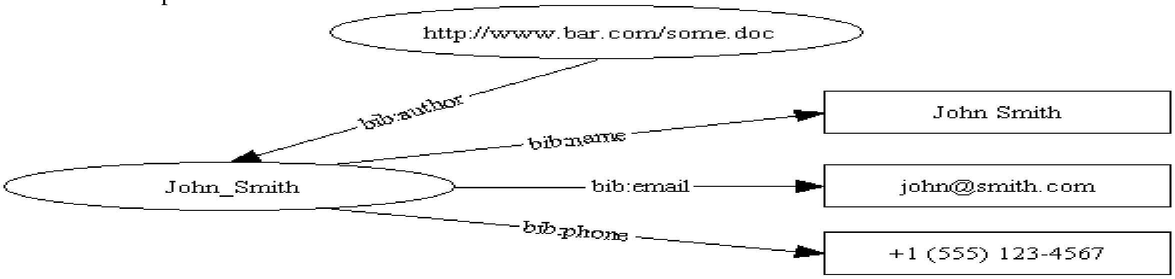

Fig.6: RDF Data Model (Event Construction)

RDF Data Model (Digital Tracking)

As an example of making a statement about a statement, that is meta data, one will consider the case of wanting to compute a digital signature on an RDF assertion. The signature is computed over a concrete XML rendition of the assertion rather than over an internal representation. Fig.7 shows a box containing a small graph. This is a convention to indicate that the XML content whose ID (Identity) is foo and is a concrete representation of the graph it contains. What we want to say in the model is expressed by the pair of graphs at the top of the figure, that we have an XML encoding of some assertion, and that there is some other XML content that is a digital signature over that encoding, then the model could be built at the bottom of the image.

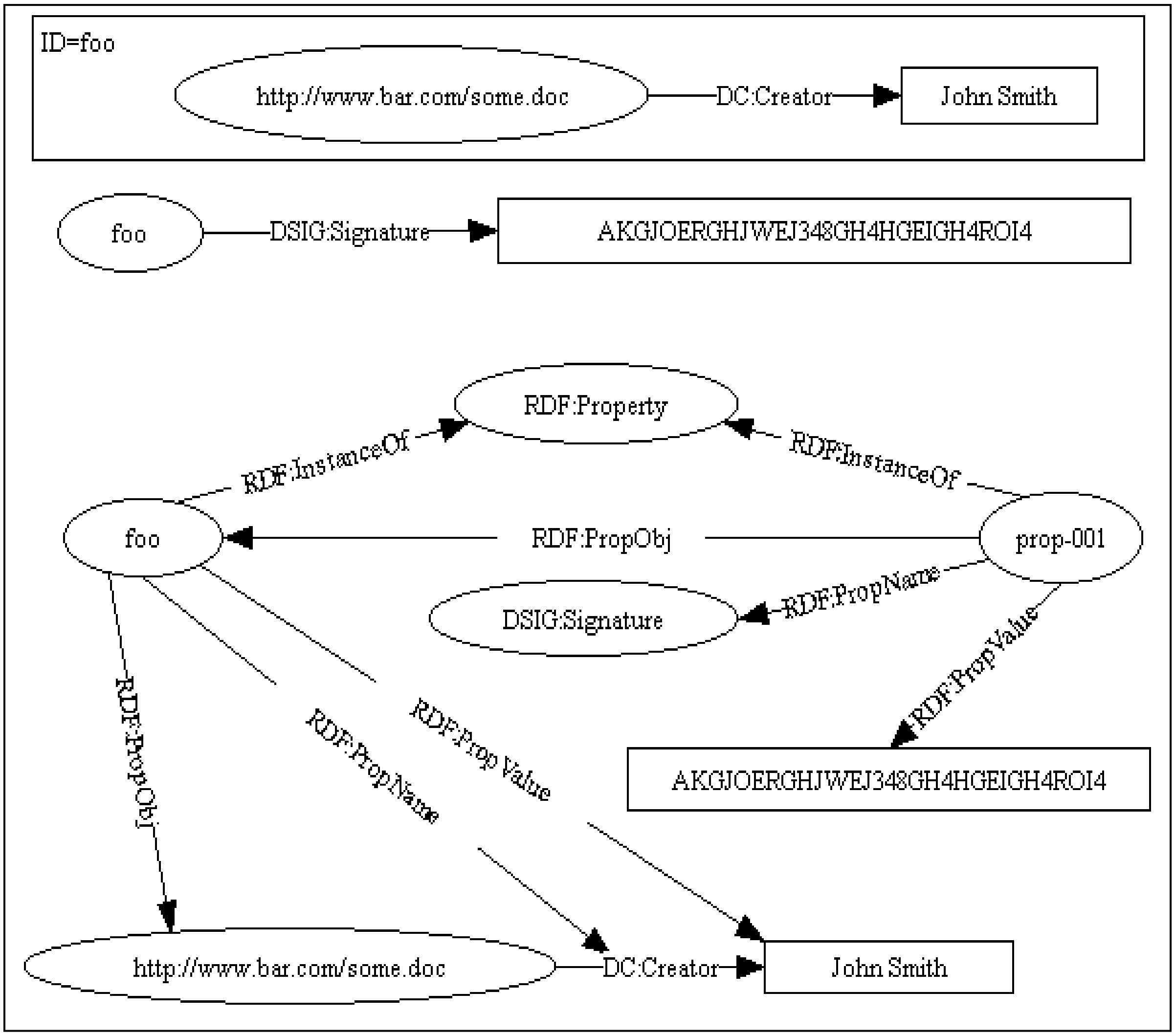

Fig 7: RDF Data Model (Digital Tracking)

Conclusion

The intention of this study was to identify and describe the potential benefit that Semantic Web based technologies and ideas can offer in the provision of a reliable digital evidence as well as tackle some of its most prominent problems. Such problems pertain to the ever-increasing amount and complexity of data, the heterogeneity and incompatibility of various disparate tools and techniques as well as the lack of automation and advanced forms of analytical capabilities. The study started with an extended background research on the field of digital evidence, semantic web, the link data initiative and the state of the art of digital investigations regarding conceptual foundations as well as concrete problems and limitations that they currently face. It further explored different digital evidence representation frameworks with its stack of cross – complementing technologies and standard along distinctive capabilities in automated reasoning, rule evaluation and expressive querying. The study continued with evaluation of recent approaches on merging data from heterogonous domains along with promising results and shortcomings. Based on this background, the study proposed and developed an adaptable method based on semantic representation, integration and correlation of digital evidence describing a hybrid framework bridging these fields as well as describing a proof – of- concept implantation of it. This study continued with a demonstration of a framework utilizing the Network Forensics Design method based on information system upon two experiments that closely resemble a quite common contemporary method of compromising a system with malicious payloads over the internet. The demonstration showed how sources of data of different origin and nature (disk images, network captures, firewall logs) can be automatically semantically represented according to respective ontologies and similar or identical entities be integrated and correlated upon various factors such as ip addresses, network captures, blocks and time. The ability of such integrated and correlated data to provide a fast and meaningful insight to the investigator has been showcased through a number of relevant queries of combinatorial nature providing results in a much effortless and analytically rich approach. The study concluded with an evaluation of the proposed system using FTK and highlighted some of the strong points such as increased automation, improved analytical capabilities, decoupled implementation and ability to use user defined concepts, rules and queries. The evaluation pinpointed also some of the merits of the proposed system with respect mostly to its performance and scalability capabilities.

Recommendations

There are a number of areas which require future research in the field of Digital Forensics. An obvious research problem lies in the automated characterization of Content objects. In future work, we intend to encourage researchers to investigate the application of KFF known-file filter technology to characterize and categorize hard disk contents.

1. Existing forensics techniques typically employ technology to eliminate operating system files and common application files from an investigation, allowing an investigator to focus his or her efforts on the remaining files.
2. We believe that known File Filter (KFF) technology could be used in computer profiling to allow an “educated guess” to be made about the contents of a file system or subdirectory on the basis of positive identification of files belonging to suspicious categories (e.g. pornography, copyrighted music, etc) using a KFF database. In investigations of distribution rings involving multiple computers, some of the same files would be found on each computer in the ring.
3. We believe that some theoretical aspects of computer profiling, especially relationships between objects in the model described in have been inadequately defined. One basic approach that was adopted during this study was that since common domain ontologies have not been yet developed and standardized, such method can still be implemented even in such a multi-ontological environment. One of the main issues was the establishment of relations between individuals representing the same concept, for example the same IP address.

References

Antoniou, G. K., Van H, F. (2014). *A Semantic Web Primer* M. P. Papazoglou & J. W. Schmidt, eds., The MIT Press. Retrieved on 12th October, 2015. From: http://doi.wiley.com/10.1002/asi.20368

Beebe, N.L., Morgan, M.P. (2019). Dealing with Terabyte Datasets in Digital Investigations, Journal of Research Advances in Digital Forensics, Vol.2, Issue 3. Norwell: Springer, Pp. 3-16.

Berners-Lee K.P, Van H, F. Kruse, M. N. (2012). A Retrospective on Semantics and Interoperability Research. In D. Fensel, ed. Foundations for the Web of Information and Services. Springer Berlin Heidelberg, Pp. 3-27. Retrieved on 10th October, 2015. From http://cs.univie.ac.at/research/research-groups/multimedia-information-systems/publikation/infpub/2921/.

Brown, F.P. (2006). Overcoming Reasonable Doubt in Computer Forensic Analysis. A handbook on Digital Forensics evidence representation. Vo1. 2 Pp 12-16.

Carrier, B.P. (2012). An Event Based Digital Forensics Investigation Framework. Proceedings in the 4th Digital Forensics Research workshop. Baltimore, MD. Pp 45.

Johann, N.K. (2000). Method ontology for intelligent network forensics analysis. In Privacy Security and Trust {(PST)}, 2010 Eighth Annual International Conference on Pp. 7-14.

Johannesson, N.K. (2011). Design Science approach for digital forensic investigation in communication Networks. *Computers Security*. Retrieved on 12th September, 2018 From: http://www.sciencedirect.com/science/article/B6V8G-52bpjwh-62893b83cb7ceffcc14c

Kruse, M.N. (2019). Crime Films: Investigating the Scene. Wallflower press, London, Pp 252.

McBride, N.M. (2016). What is forensic computing? *Trends &Issues in Crime and Criminal Justice* Canberra: A Journal by Australian Institute of Criminology, Vol.2, Issue 3, Pp 1345 -1355.

Onyemauche, U.C. (2017). *Development of an Integrated Digital Forensics Model for the Provision of a Reliable Evidence.* A Published PhD Dissertation submitted to Nnamdi Azikiwe University Awka Nigeria.

Parker, K.P. (2000). An Event based Approach to Digital Evidence Representation. Digital Forensics Research Part C. Emerging Technologies pp 231 – 300.

Sommer, P. (2016). Computer Forensics: An Introduction. Retrieved on: 10^{0h}, December, 2018. From: http//www.virtualcity.co.uk/vcaforensics.htm

Stephenson, L.P. (2017). Digital droplets: Microsoft Skydive forensic data remnants. *A Journal of Future Generation Computer Systems, Issue* 29, Vol. 6, Pp 1378–1394.