Towards Adapting Metamodelling Technique for an Online Social Networks Forensic Investigation (OSNFI) Domain

Aliyu Musa Bade
Department of Computer Science, Yobe State University
Damaturu, Nigeria

Siti Hajar Othman
School of Computing, Universiti Teknologi Malaysia
Johor Bahru, Malaysia

Abstract—With the ease of use of smart devices, most data is now kept and exchanged in digital forms such as images, diaries, calendars, movies, and so on. Digital forensic investigation is a new technology that emerged from criminals who extensively use computers and digital storage devices to commit different types of crimes. To address this issue, a new domain called Online Social Networks Forensic (OSNFI) was created to investigate these dynamic crimes perpetrated on social media platforms. OSNFI seeks to obtain, organise, investigate, and visualise user information as direct, objective, and fair evidence. Considering the millions of individuals using social media to share and communicate, they are becoming increasingly relevant for criminal investigations. In forensic investigation of online social network, there are currently major problems such as: lack of structured procedures, the lack of unified automated methods, and the lack of a theoretical context. The use of non-uniform and ad-hoc forensic techniques and procedures not only reduces the effectiveness of the process, but also affects the reliability and creditability of the proof in criminal proceedings. As a result, this paper will provide a method derived from the software engineering domain known as metamodelling, which will integrate OSNFI knowledge into an artifact known as a metamodel.

Keywords—Online social networks forensic; online social networks forensic investigation; metamodelling; metamodel; model

I. INTRODUCTION

Most organisations are now heavily reliant on digital media for information storage as digital media is used to create, process, store, and share the majority of information. Computer crimes and frauds are on the rise as the number of people using digital media grows and the law enforcement agencies faced several challenges as a result of the growing fraud and security threats. OSNs have received a lot of forensic investigation due to their widespread use and availability of supporting application interfaces. Furthermore, the number of criminal acts involving OSNs is increasing on a daily basis. It is vital to comprehend and evaluate online social network crimes and attacks in order to avoid illegal activity, discover malevolent users, and solve criminal cases. Furthermore, OSN users' safety should be enhanced to the greatest extent possible.

There are a variety of digital forensic investigation models available. However, most of them use similar methodologies and ignore the essential differences and special needs of online social networks. The main challenge is the use of a non-uniform and ad-hoc forensic techniques in the existing DFIM’s [1]. The employment of non-standard and ad hoc forensic techniques and processes not only diminishes the effectiveness of the process, but it also undermines the dependability and creditability of the evidence in criminal cases.

Furthermore, while most of the created models focused more on a certain platform or content than the entire OSN [2], some still required manual treatment and could reduce the dependability and trustworthiness of evidence in criminal proceedings [3].

Therefore, there is the need to create a metamodel (a conceptual framework) for a domain which will assist many newcomers and stakeholders of the domain to have clear understanding and views of the relationship between concepts in the domain. Therefore, this is where this research will contribute in unifying the knowledge and all the processes of the OSNFI domain.

The remaining sections of this paper are organized as follows. Section II discusses on the concept of OSNFI. The OSNFI issues and challenges are presented in Section III. Section IV discusses the metamodelling approach. The preliminary results of the OSNFI metamodel development steps are presented in Section V. Sections VI and VII contain the conclusion and future work respectively.

II. ONLINE SOCIAL NETWORKS FORENSIC INVESTIGATION

The Scientific Working Group (SWG) defines Digital Forensic as the application of systematically developed and established methodologies for the safeguarding, gathering, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence produced from digital sources [4]; to ensure the correct presentation of crime evidentiary data in court [5], primarily by protecting the data's credibility and ensuring a strict chain of custody. The ultimate aim of digital forensics is to gather evidence in order to answer questions like: What transpired, Who was involved, When did it take place, Where did it take place, Why did that happen, and How an incident occurred [6]. According to [7], the prevalence of computers and other devices has caused a proliferation in the quantity of incidents and amount of digital information in society. Previously, digital forensics was
According to [5], there are many reasons to justify why digital forensic investigation models need to be developed. Among which are: (i) the avoidance of future malicious events against the intended target, (ii) the successful tracing of the circumstances which lead to the crime and the identification of the parties involved, (iii) identifying and apprehending the culprits of the crime, (iv). Improving present prevention procedures in place to avoid such an event from happening again, (v). Improving corporate security experts’ usage of standards to secure their respective corporate networks, and (vi). How everyone connected into this digital world can become more aware of current vulnerabilities and preventive measures. However, multiplicity, measure cloud resources, secure of digital evidence, confidentiality, hiring, training, and development are some operational challenges faced by digital forensic [4].

Forensics is utilized on social networking networks such as Facebook, MySpace, Twitter, and LinkedIn. It is well known as social media forensics, and it’s a subset of digital forensics and network forensics [8]. OSNs are web-based services which allow users to create a public or semi-public profile within a limited system [3], articulate a list of other individuals that they share a link, display and navigate their list of connections as well as those created within the system by others [9]; [10]. Different SNs, like Facebook, Twitter, and LinkedIn, are used to connect people and enable them to communicate with one another [11]. People build personal profiles from various social networking sites to share their thoughts, photographs, images, emails, and instant messaging [12], as well as to find old friends or people with common interests or problems through various social networking sites [13].

Facebook is managing 2,895 million users around the globe, whatsapp managing 2,000 million users, Instagram managing 1,393 million users, and Twitter managing 436 million users as of October 2021 as shown in Fig. 1. With the proliferation of mobile phones, the use of social network services (SNS) has skyrocketed, this SNS stores a variety of data, including user conversations, user location information, personal networks, and user psychology which can be valuable evidence in a digital forensics investigation of an incident [15]. Other uses of social networking sites include, general chatting, broadcasting breaking news, setting up a date, tracking election results, planning disaster response, humour, and serious analysis [12].

III. OSNF Issues and Challenges

Investigators and legal practitioners are currently finding it difficult to examine social networking sites for evidence [16]. In order to discuss on the challenges associated to this explicitly, the method of the investigation are categorized into two: Conventional and Automated/Semi-automated OSN digital forensic investigation.

A. Conventional OSN Digital Forensic Investigation

Conventional OSN digital forensic investigation is a manual way of collecting and analysing all relevant information which can be considered as evidence. The following are some of the challenges associated with this technique:

1) Traditional digital forensic procedures include seizing and gathering everything that may be considered evidence. Nevertheless, this level of complete data extraction and preservation cannot be possible on online social network because of the extremely dispersed nature of social networks, their enormous scale, and mutual proprietorship of data.

2) Gathering data from people connected to the subjects (i.e., suspects or victims) lacking a reasonable suspicion of wrongdoing is virtually impossible and legally prohibited under confidentiality acts.

3) The evidence-collection process on social media sites is a process that is sometimes iterative. Therefore, the investigators will be required to gather more information if the examination subsequently identified other suspect [17].

B. Automated / Semi-automated OSN Digital Forensic Investigation

1) It can dumb down the profession because expert expertise cannot be derived solely from the field; it must also be derived from ongoing formal and informal research [13].

2) Digital forensics automation is more than just a technological issue; it also has legal and even political implications. Therefore, automation should be used only in specific phases and under expert supervision [13].

3) There are limited available automated tools that can be used in the investigation as a result of the heterogeneity of online social network and the non-existence of standardization [17].

4) Digital investigation training, practitioner training was a major concern, with 73 percent of respondents believing they don't get enough, especially in digital forensics, online investigations, and computer and network security as contained in a report by in a report by the High Technology Crime Investigation Association (HTCIA).
C. Current Issues of OSN Digital Forensic Investigation

Five relevant current issues of OSN forensic investigation as presented in Fig. 2 are discussed in this section as follows:

1) Increase in cyber-criminal activity in OSN: When creating a profile on many OSN sites, users can include their complete name, pictures, date of birth, up-to-date location, contact numbers, residential address, and office address (amongst others) [3]. Such profiles can aid in the connection of users. Users may choose to keep their profiles private, limiting their contacts, or make them public, allowing everyone to access and contact them.

Criminals may use all of the details available in these profiles to identify a specific person [18]. By their very nature, social networks have certain inherent properties that make them suitable for an adversary to manipulate. The following are the most critical of these characteristics: (i) a huge and extremely distributed user-base, (ii) clusters of users having the same social interests, developing trust with one another, and pursuing access to the same resources, and (iii) platform openness for deploying fraud resources and applications that trap users to install them.

Undoubtedly, these characteristics are the causes why cyber criminals believe there is a large chance to use online social network as a platform to commit crimes [3]. Among other illegal activities that have become a significant threat to OSN includes; Website phishing, Online sexual predators, OSN as vehicles for reaching an international audience, soliciting funding, recruiting new members, and disseminating propaganda.

2) Anti-Forensic techniques: Technical difficulties make obtaining data from the OSNs difficult. Criminals use anti-forensic techniques to hide evidence or to distract attention away from the investigation. Encryption, steganography, covert channelling, storage space data hiding, and residual data wiping are all methods used to conceal evidence [19].

3) Standard models: A scarcity of standardisation, as well as a theoretical structure for the field of digital forensics, is one of the most serious issues that investigators are currently facing [20]. A such, the use of ad hoc methods and tools for eliciting digital evidence may limit the evidence's reliability and legitimacy, particularly in criminal cases where both the evidence and the processes used to gather it can be contested [3].

4) Legal challenges: The complications in investigations as a result in scarcity of commonly established rules and criteria governing the field are referred to as legal challenges. There is no unified legal system that applies to all jurisdictions. Despite the fact that internet use has risen globally. Many countries formulate policies in accordance with their regulatory structure, which differs from country to country [19].

5) Resource challenges: Advances in cybercrime, strategies in propagating them and evading investigations are made possible by technical development. The volume of data accessible for assessments for OSNs is normally enormous. The DFIs must classify the most relevant data without jeopardising the evidence's consistency. As a result, there is a pressing need for the development of new technologies and mechanisms for fighting cybercrime, as well as qualified personnel to carry them out [19].

D. Crimes Involving OSNs

Online Social Network is a database of information that criminals can use to commit various forms of crimes such as malware distribution, fraud, harassment, grooming, Assault, burglary, domestic violence, kidnapping and so on [21]. Therefore, OSN-related crimes are divided into two categories: Classical crimes and Digital crimes as presented in Fig. 3.

---

Fig. 2. OSNFI Challenges and Issues.

---
Fig. 3. Crimes Involved in OSNs [21].

1) Classical crimes: Online social network is referred to as a location where criminals can commit traditional criminalities. For example, social media users can update their status by posting their current whereabouts, the time they will be absent from home, and what they will be doing, giving possible criminals enough time to break into their home. This is one of several incidents that have been reported in the media [21]. Other crimes may include: vandalism, domestic violence, terrorism organised crime, fraud, sexual threat, kidnapping, rape sexual assault, sex trafficking, murder and stalking.

2) Digital crimes: Any criminal activity involving an information technology infrastructure, such as unauthorised or illegal entry, surveillance, and so on, is referred to as digital crime. The most common OSNs digital crimes are: cyber-based, and social engineering which is one tool that can be used to commit these crimes [21]. Other crimes may include: drug network, street gang, bombing, harassment, child phonography, cyberstalking, scammer, grooming, cyberbullying, identity theft, and malware distribution as presented in Fig. 4.

Fig. 4. Global Map of Sample Crime Analysis [22].

E. Necessity for OSN Digital Forensic Investigation Metamodel

SNS stores a variety of information about its users which in an investigation case, the suspect's personal data stored on a social media site will support in a variety of ways, including identifying living habits, determining geographical location, and assessing ideas and mental state. However, by simply collecting data from an SNS user's personal information section, it is difficult to gather useful data for a case investigation [15]. In order to address these flaws, OSN requires the establishment of a standardised forensic investigation procedure which can help investigators in an investigation. These procedures are presented in Fig. 5.

IV. METAMODELLING APPROACH FOR ONLINE SOCIAL NETWORKS FORENSIC

Metamodelling is the process of developing metamodels [23]. Modelling and metamodelling are both talking about model creation but the only difference between them is in how they are interpreted.

There are numerous digital forensic investigation models available. However, the majority of them employ comparable approaches and disregard the crucial distinctions and unique requirements of online social networks.

In [17], proposed a seven phase semi-automated forensic investigation model for online social networks. These phases are: Pre-Investigation, Incident specification, Extraction, Preservation, Analysis, Iteration and Presentation. A WhatsApp Messenger Smartphone Forensic Investigation Analysis Against Web-Based WhatsApp was proposed by [24] which comprises of six phases, namely: Identification, Preservation, Collection, Examination, Analysis and Presentation. A four-phase Framework Analysis of IDFIF V2 in WhatsApp Investigation Process on Android Smartphones was proposed by [25]. The phases are: Preparation, Incidence response, Laboratorium process and Presentation.

A Framework for the Forensic Analysis of User Interaction with Social Media was proposed by [26] which has four phases: Acquisition, Triage, Analysis and Presentation. A four-phase Digital Forensic Investigation Model for Online Social Networking was proposed by [21] which has: Preliminary, Investigation, Analysis and Evaluation phases.

The fundamental issue is the employment of ad hoc and non-uniform forensic techniques in the current DFIMs. The use of ad hoc and non-standard forensic procedures reduces not only the efficiency of the procedure but also the dependability and credibility of the evidence in criminal proceedings. Consequently, this is where the research will help to standardise the OSNFI domain's operations.

A. Models and Metamodels

A model is a representation of real-world phenomena [27]. It is a description of something [28], which are used to reason about a problem domain and design a solution in its domain [29]. Models are essential for comprehending and disseminating information about complex systems [30]; [31]. They have been and continue to be crucial in many scientific...
contexts as a large branch of philosophy of science is centered on models.

Models can be used for different purposes. They can come in form of graphical, mathematical or textual. Models can be used for descriptive, prescriptive or for defining the method by which a system will be implemented. Descriptive is simply describing a system's or a context's reality while Prescriptive is to determine the extent and depth of a problem's investigation [32]. On a philosophical level, one can concur that "everything is a model," because nothing can be managed by human mind unless it is "modelled." As a result, it's not astonishing that models have become crucial in technical fields like mechanics, civil engineering, and, eventually, computer science and engineering [33]. According to [34], models can be categorised into; conceptual versus data models, viewable and executable models, active and passive models, static and dynamic models in the study of information systems. Identifying concepts terminologies, meaning, definition and interconnections are some of the challenges encountered in model development; this is because concepts terminologies are too inconsistent and too ambiguous.

Metamodel is a model that describes/prescribes models [35]; [23]; [32]. It is a collection of concepts and relations that define the syntax of a model, and they are described using a model description language [36]. Metamodel is an abstraction that highlights properties of the model itself just as how a model is an abstraction of real-world phenomena [27]. It describes a collection of concepts and their relevant relationships, and it used as an abstraction filter in a specific modelling activity [37]. A modelling language is used in the creation of a metamodel and this language is termed as Metamodeling language [38].

In Model-Driven Engineering, metamodels are widely used to specify available model elements and structures. Nevertheless, metamodels are likely to evolve during development for a variety of reasons, such as changing requirements or evolving domain expertise [39]. Metamodelling is among the significant components of MDD; its main importance include the creation of a modelling language that will be utilized to accurately define metamodels [40] and is used to ensure the consistency of models during transformation [41]. Metamodelling is the study of processes, development of frames, production rules, constraints, models, and theories that can be used to extended models of intelligent software and information systems [23]. The basic goal of metamodelling is to represent data in such a way that it becomes self-contained and allows for the extension and alteration of its structure. Metamodelling must adhere to its own set of rules, which are as follows: Suitability, Completeness, Dynamic, Openness, Compatibility, Consistency, Reusability, and Simplicity.

B. Step-by-Step of OSNFI Metamodel Development.

A uniform representation of language is frequently required to reflect shared understanding in a consistent manner that fulfills the needs of the various parties involved. This method discovers specific domain characteristics, gathers domain concepts, and divides domain problems into sub-domain problems. The steps in the procedure are as follows:

1) Step 0: Identification of common phases of domain.
2) Step 1: Models collection and classification.
3) Step 2: Extraction of concepts.
4) Step 3: Identification of common concepts.
5) Step 4: Short-listing and reconciliation of definitions.
6) Step 5: Classification of common concepts: Concepts are assigned to one of the OSNFI phases: preliminary, acquisition/preservation, analysis, or presentation.
7) Step 6: Identification of relationships.
8) Step 7: Metamodel validation (is not covered in this paper).

To achieve the goal of creating a universally applicable OSNFI, a broad coverage across concepts is required. Looking at the coverage measure alone, as shown in Table I, it gives a rapid indication of the supplied model's applicability. If a model can cover all of the stages of OSNFI, it is said to have a high coverage value. If the model just describes a single OSNFI phase, the coverage value of the model is lower.

| Models | Coverage | Coverage of Model (according to phases) |
|--------|----------|----------------------------------------|
| 1 A Digital Forensic Investigation Model for Online Social Networking [3] | 0.2 | Preliminary, Analysis and Presentation |
| 2 Online Social Networks As Supporting Evidence: A Digital Forensic Investigation Model and Its Application Design [9] | 0.2 | Preliminary, Analysis and Presentation |
| 3 A Framework for the Forensic Analysis of User Interaction with Social Media [26] | 0.2 | Acquisition, Analysis and Presentation |
| 4 Forensic Imaging for Online Social Networks [11] | 0.2 | Preliminary, Analysis and Presentation |
| 5 A Review of Using Online Social Networks for Investigative Activities [21] | 0.2 | Acquisition, Analysis and Presentation |
| 6 A comprehensive digital forensic investigation process model [42] | 0.3 | Preliminary, Acquisition, Analysis and Presentation |
| 7 WhatsApp Messenger Smartphone Forensic Investigation Analysis Against Web-Based WhatsApp [24] | 0.2 | Acquisition, Analysis and Presentation |
| 8 Framework Analysis of IDFIF V2 in WhatsApp Investigation Process on Android Smartphones [25] | 0.2 | Acquisition, Analysis and Presentation |
| 9 Forensic investigation of cross platform massively multiplayer online games: Minecraft as a case study [2] | 0.2 | Preservation, Analysis and Presentation |
| 10 A semi-automated forensic investigation model for online social networks [17] | 0.3 | Preliminary, Acquisition, Analysis and Presentation |
V. The Resultant Online Social Networks Forensic Investigation Metamodel (OSNFIM)

In Fig. 5, the preliminary results of developing OSNFIM are represented by twenty-two key concepts that were collected from nineteen and filtered based on the four phases (Preliminary, Acquisition/Preservation, Analysis, and Presentation). Disparities between definitions are addressed in this process. All of the definitions listed in this process are taken into account when picking or synthesizing the common concept definition to be used as presented in Table II.

Although OSNFI is no longer widely used in research, the precise meaning of important terminology and phrases used in its concepts can often still differ among the few academics who are familiar with it. This can be a result of the researchers having different backgrounds and viewpoints. If two or more sources employ concept definitions in conflict, a technique to harmonize and fit the definition in the metamodel is necessary. In some circumstances, some models don't participate in the reconciliation procedure because they don't describe some of their concepts explicitly.
| S/No. | CONCEPT                          | DEFINITION                                                                                           |
|-------|----------------------------------|------------------------------------------------------------------------------------------------------|
| 1     | Operational readiness            | This is a proper initial policy and procedure documents before a successfully launch of an effective digital forensic investigation. |
| 2     | Infrastructure readiness         | Infrastructure readiness component is determined by elements external and internal to the organization which is critical to test them before implementation so that they are ready for use when needed. |
| 3     | Incident notification            | Explicitly specifying the incident under investigation.                                               |
| 4     | Authorization                    | Stages to gain access to evidence and legal status of the inquiry process.                           |
| 5     | Acknowledgment                   | This is the first step of an online social network forensic investigation where a case or an audit is requested from an external organisation such as the police, customs, or a company. |
| 6     | Identify Social Net sources      | A way of identifying the social network involved by Initialize the SN source.                        |
| 7     | Identification                   | A process of identifying any evidence or supporting information that might be available in an online social network. |
| 8     | Searching                        | This is a process of discovering relevant data automatically based on the relevant data gathered from the investigation process. |
| 9     | Filtering                        | An activity which will scale down and focus the investigation on relevant information and discard any irrelevant information. |
| 10    | Capturing                        | Information collected through filtering will be captured in the best way to ensure the integrity of the data is sustained. |
| 11    | Transport                        | This is the process of moving digital evidence from the scene to the forensic digital laboratory.     |
| 12    | Storage                          | A process of keeping potential digital evidence which might be needed if the analysis cannot be performed right away or if there is a legal requirement to keep digital evidence for a certain period of time. |
| 13    | Preserve a forensic copy of Data Set | Safeguarding the integrity of the original digital evidence.                                       |
| 14    | Sort and filter the data relevant to the inquiry | The investigators must examine the results in the context of a given incident.                     |
| 15    | Conclusion                       | Investigators will conclude their examination in this stage. They may confirm their hypothesis, or they might need to find more information related to the entities involved in an investigation to credit or discredit a theory. |
| 16    | Select Relevant Evidence         | The investigators will select the evidence that is relevant and appropriate for presenting in court.  |
| 17    | Present the Evidence             | To show anything you see, experience, read, or hear that leads you to assume something is true or has actually happened |
| 18    | Decision                         | A decision or resolution reached after careful consideration                                          |
| 19    | Interpretation                   | To use scientifically proven methods to explain facts discovered throughout the analysis process within the context of the investigation. |
| 20    | Documentation                    | Documentation of all activities that have been done from the beginning of the investigation process to the end of the analysis process in the forensic laboratory. |
| 21    | Investigator                     | A person who conducts a formal investigation or inquiry                                               |
| 22    | CourtOfLaw                       | A group of individuals presided over by a judge or judges who operate as a tribunal in civil and criminal proceedings |

VI. CONCLUSION

Online social networks forensic investigation domain is a new, but extremely important and a high demand domain. The number of crimes increases on daily basis due to the advancement in technology and the use of smart devices. The problems with the OSNF domain which are addressed in this paper are: lack of uniformity in the procedures for the OSN investigation, increase in cyber-criminal activity, anti-Forensic techniques, standard models, legal and resource challenges. Hence, the OSNF investigation metamodel proposed will aid in the proper investigation of digital crimes through various processes.

VII. FUTURE WORK

The future work will be to validate the proposed Metamodel based on two relevant metamodel validation technique: (i) Comparison to other models and, (ii) Frequency-based Selection validation techniques.
