IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 21 September 2023, accepted 1 October 2023, date of publication 5 October 2023, date of current version 16 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3322162

RESEARCH ARTICLE

Ontology-Driven Digital Profiling for Identification and Linking Evidence Across Social Media Platform

ŠARŪNAS GRIGALIŪNAS, (Member, IEEE), RASA BRŪZGIENĖ, AND ALGIMANTAS VENČKAUSKAS

Department of Computer Sciences, Kaunas University of Technology, 51368 Kaunas, Lithuania

Corresponding author: Šarūnas Grigaliūnas (sarunas.grigaliunas@ktu.lt)

ABSTRACT In an era in which social media platforms are proliferating and becoming primary communication channels, the identification of evidence for crimes from such platforms is crucial for digital forensics and legal proceedings. This paper presents a novel approach for systematically structuring and categorising digital attributes that are interlinked across social media platforms using digital ontologies, as well as a method for user profiling using domain-specific digital artefacts. The ontology models consist of classes with subclass distinctions for text, image, and video types of evidence. These models are flexible and can be expanded to include various social media platforms and evidence categories. Simultaneously, the user profiling method employs mathematical formulas and visual representations to develop comprehensive profiles of individuals based on extracted social media data. This methodology evaluates the relevance of a set of digital artefacts and related attributes, such as interests, location, and activities, using their weights. Additionally, the research addresses the legal and ethical considerations pertinent to the collection of data from social media. Despite the approaches' immense potential for expediting evidence collection and developing insightful profiles, obstacles such as scalability, legal complexities, and data noise are identified. This work makes a substantial contribution to the development of digital forensics and cybercrime investigations involving social media platforms.

INDEX TERMS Social media, digital evidence, profiling, digital forensics, digital artefacts.

I. INTRODUCTION

Technology advancements, the proliferation of information, and the rise in popularity of social media platforms like Twitter, Facebook, Instagram, TikTok, Youtube and others have all contributed to the growing reliance on these sites in our everyday lives. Social media platforms have many positive effects, but it also poses risks to individuals. In this day and age of technology, there are a lot of crimes that may be linked to social media in some way. Back in the day, most criminals would leave behind some sort of trail of evidence in the real world. In today's world, criminals increasingly feel safe conducting their activities in cyberspace, with devastating consequences for society. They engage in illegal behaviours such as fraud, cyber stalking, cyber bullying, and many more by using online tools provided by social media [1], [2]. These days, even the most cynical among us have access to the most cutting-edge social media platforms, and the Internet's many advantages never cease to astound them. The anonymity and the capacity to establish a virtual world where people may connect digitally without ever having met each other face-to-face and share information, images, and other material are major factors in social media's popularity under the cybercriminals.

On the other way, social media platforms are a source of evidence for crimes because they include a great quantity of information that can be used to identify suspects, establish motivations, and recreate the events leading up to a crime. Given the fact that it is digital platform where individuals communicate their thoughts, feelings, and experiences with

The associate editor coordinating the review of this manuscript and approving it for publication was Chien-Ming Chen.

the general public or with a small group of people, social media may be valuable source for investigators of crimes. It may include a plethora of data that is useful for a criminal inquiry [3].

Social media platforms enable users to establish profiles that include personal information such as name, age, location, occupation, and level of education, in addition to photos and videos. These profiles can be used to identify and confirm the identities of suspects. Moreover, social media platforms permit users to post messages, comments, and status updates. It preserves a record of user activity, which includes the posts that users have liked, the comments that they have made, and the sites that they have visited. These pieces of information can be put to use in order to determine the suspects' interests, relationships, and connections in the criminal world. In addition, it includes private communication tools, such as direct messaging and discussion forums. The content of these conversations can be used as evidence in criminal investigations if they are recorded [4]. It's essential to keep in mind that acquiring data from social media platforms and using that data as digital evidence must be carried out in a manner that is compliant with applicable laws and privacy rules.

Digital evidence is critical in cases of cyber crime on social media, because it can help to establish the identity of the perpetrator, the scope and severity of the crime, and the extent of any damages or harm caused by the crime [5]. By identifying digital evidence from social media platforms, investigators can obtain important information such as IP addresses, login credentials, timestamps, communication records, *etc.*, that can help to identify the person or persons responsible for the cyber crime. This information can also be used to track the location of the perpetrator, determine the scope of the crime, and identify any other individuals who may have been involved.

Digital evidence can provide a clear and objective record of the events that took place, and can help to establish the authenticity and accuracy of other forms of evidence that may be presented. Nevertheless, it is essential to accurately depict the entire context of the crime by tracking the digital evidence that has been identified. In this way, the profiling the evidence in a case of cyber crime on social media can help investigators to identify patterns, connections, and other important information that may not be immediately apparent from individual pieces of evidence.

Profiling involves analyzing the digital evidence collected from social media platforms to identify key characteristics and trends that may be relevant to the case [6]. This involves looking for patterns in the timing, location, and content of social media posts, as well as analyzing metadata and other digital footprints to identify connections between different pieces of evidence. For instance, a suspect who frequently posts or comments on certain types of content, or who frequently interacts with certain individuals or groups, may be more likely to be involved in cyber crimes related to those topics or communities. Similarly, a suspect who frequently uses certain types of language or expressions in their posts or messages may be more likely to have specific motivations or psychological profiles. Profiling enables analysis of the timing and location of the suspect's digital activity, as well as their device and network characteristics, to identify potential leads and generate hypotheses about the suspect's movements and whereabouts at the time of the crime.

By profiling the evidence, investigators can puzzle out the answers to fundamental questions, which serve as the basis for the digital evidence object model [7]:

- **When** was the cyber crime committed? The time and date of the crime can be used to narrow down the list of suspects and identify individuals who were active on social media platforms around the time of the crime.
- **Why** was the cyber crime committed? The motive behind the crime can be uncovered by analyzing the content and context of social media posts and messages, as well as any previous interactions between the perpetrator(s) and the victim(s).
- **What** methods were used to carry out the crime? The type of cyber crime and the tools or techniques used can be identified through forensic analysis of digital evidence.
- **Where** did the cyber crime occur? The location of the crime can be determined by analyzing IP addresses and other digital data, which can help to identify potential suspects based on their geographic location.

Generally, by piecing together the information gathered from the previous questions, investigators can narrow down the list of important information such as:

1) the motive behind the crime;
2) the methods used to carry out the crime;
3) the extent and scope of the crime;
4) the potential impact on individuals or organizations affected by the crime.

Following that, the list of possible suspects may be reduced, allowing investigators to finally zero in on the person or people **Who** potentially committed the cyber crime.

Identification and profiling of digital evidence on social media require a combination of technical expertise, legal knowledge, and critical thinking skills to navigate the complex landscape of social media platforms and effectively gather and analyze relevant evidence. While social media platforms generate enormous amounts of data, it can be difficult to sift through all the information to find relevant evidence. Moreover, they are constantly evolving, with new features and interfaces being added frequently. This can make it difficult for investigators to keep up with the latest changes and understand how to navigate each platform effectively. The authenticity and reliability of digital evidence is difficult to verify, especially when it has been altered or manipulated in some way [8]. Furthermore, social media users expect a certain level of privacy, so investigators must be aware of the legal requirements and restrictions surrounding the collection

and use of digital evidence, especially when it comes to privacy laws and the admissibility of evidence in court.

Data acquired from diverse sources is used to identify and profile digital evidence on social media. The data gathered may be incomplete, and some crucial information may be overlooked. Users, for example, may remove their accounts or posts, and the data gathered prior to such deletion may not properly represent the user's behaviour. Users on social media networks behave differently on each platform. This means that narrowing down the potential suspect of the crime necessitates data from several platforms, and the data gathered may not be representative of the target community. This has the potential to inject bias into the analysis.

On the same note, social media networks impose limitations on the data that may be viewed and used. This can reduce the effectiveness of identification and profiling tools, especially when identifying anonymous individuals or examining private accounts. Identifying patterns and relationships in data is frequently required when profiling digital evidence on social media. This, however, can result in false positives, in which innocent users are incorrectly classified as suspicious or related to criminal activity.

Even though advanced analytical tools can be effective for analysing digital evidence from social media [9], [10], investigators should be aware of their limitations and weaknesses and use them in conjunction with other techniques to ensure accurate and ethical results. Typically, machine learning and artificial intelligence algorithms are trained on specific categories of data and may not be able to account for broader contextual factors that may influence the interpretation of digital evidence [11]. Depending on the data used to train an algorithm, it may produce inaccurate or unjust results. Even if a machine learning model performs well on its training data, it may perform inadequately when exposed to new data if it was not trained in a diverse environment or with context in mind [12], [13]. In the context of social media, where data can be influenced by social biases and algorithmic amplification, this is of particular importance. Complex and difficult-to-interpret machine learning and artificial intelligence algorithms can make it challenging to comprehend how the algorithm arrived at a particular result. This is particularly problematic in legal contexts, where transparency and explicability are essential. It is possible for machine learning and artificial intelligence algorithms to generate false positives and false negatives, causing investigators to overlook crucial evidence or draw erroneous conclusions from the data.

In the context of digital forensics and cyber crime investigations, digital ontology can be used to help organize and analyze digital evidence from sources like social media [15]. By providing a standardized way of representing relevant concepts and relationships, it can help investigators identify and prioritize key evidence more effectively, and can facilitate collaboration and information-sharing between different stakeholders involved in the investigation process [16].

Digital ontology help to ensure consistency and interoperability across different tools and systems used in the investigation process. This is especially important when working with multiple investigators or agencies that may be using different tools or systems. It reduces errors and inaccuracies in the investigation process by providing a clear and consistent representation of digital evidence. Even more important is that such a method can be customised and adapted to the specific needs and requirements of a particular investigation, making it a highly flexible and adaptable tool for analysing digital evidence from social media.

Motivated by this, the authors of this paper aims to explore the identification and profiling of digital evidence by examining domain-specific digital artefacts that are cross-linked between social media platforms. The process of justifying digital evidence involves the analysis of the attributes from various social media platforms to identify commonalities among different types of digital artefacts. Main contribution of the proposed approach includes:

- an ontology-based diagram for the digital profiling of attributes that are interlinked across social media platforms;
- a method for the identification of the digital evidence to narrow down the relationship with a potential suspect(s) of a cyber crime on social media;
- a user profiling based on domain-specific digital artefacts defined by set of digital attributes gained from social media.

Digital artefacts are data that is created or shared in a digital format, that are stored or transmitted on social media platforms. These artefacts may contain valuable information that can be used in legal or criminal investigations, such as evidence of a crime or a suspect's location or activities. The analysis of digital artefacts requires a systematic approach that involves the examination of different attributes associated with each artefact. These attributes can include technical details such as the file format, resolution, and metadata, as well as contextual information such as the location, time and date, and authorship of the artefact. By analysing these attributes across different types of digital artefacts, identification of patterns and commonalities can be done that may be relevant to a particular investigation. For instance, by analysing the metadata associated with an image or video, investigators can determine the location, date, and time of the original creation or upload of the artefact. This can provide clues as to the identity of the person who created or shared the artefact and may help establish a timeline of events. Additionally, the analysis of the file format and technical details of the digital artefact can help to identify the software or tools used to create or modify the artefact, which may provide further evidence to support the investigation.

The remainder of this paper is as follows. Section II reviews the current scientific methods and techniques, related to the analysis of digital evidence on social media. The proposed ontology-based model for the identification and

profiling of digital evidence is described in Section III. A validation of the proposed approach and analysis of the obtained results are discussed in Section IV. The significance of the proposed model and its limits are discussed in Section V. And finally, Section VI summarises the work by detailing the outcomes obtained and plans for future works.

II. RELATED WORKS

The field of digital forensics in social media is becoming more significant as social media platforms continue to play an essential role in various facets of our lives, such as personal communication, business, and politics. Legal processes are increasingly making use of evidence obtained from social media platforms, which highlights the need for developing digital forensics methods that are both resilient and dependable. Chat logs, posts, comments, likes, shares, direct messages, and other forms of online communication are some types of digital evidence that may be gathered during social media forensics investigations. The reconstruction of timelines and the piecing together of the events that led up to a specific occurrence may be accomplished with the help of metadata by professionals that specialize in digital forensics. Evidence relating to cyber crime, harassment, identity theft, fraud, and other forms of digital wrongdoing on social media may fall under this category. The review of the relevant scientific works focuses on the following:

- identification of target and their connections on social media;
- profiling of the user (potencial crime suspect) by the collected digital evidences.

A summary of the reviewed scientific papers is presented in Tables 1, 2 highlighting the elements involved in digital evidence identification and user profiling. The shortcomings of solutions proposed by other researchers are concluded as well.

The process of recognising, identifying and collecting important digital information that might serve as digital evidence in a forensic investigation is what plays a role in the identification of target on social media platforms equivalent to Twitter, Facebook, TikTok and others. In this initial phase, investigators define the specific elements they pursue on social media platforms in order to conduct an effective investigation. By establishing well-defined objectives, investigators can concentrate their efforts and optimise the investigation process.

Investigators utilize specific search techniques and keywords to identify relevant content on social media platforms. This includes searching for posts, comments, messages, profiles, groups, devices [16] or hashtags related to the investigation. For example, the identification of the authorship can be made by the analysis of a digital text, known as short comments on social media. Such analysis includes capturing of features of the author's writing style at multiple levels, such as the quantity of individual characters, the sort of words that an individual employs, the manner in which the writer organises the sentences, or the usage of distinct types of terms [17]. This strategy undermines confidence since it opens the door to the possibility that several people are producing

TABLE 1. Summary of reviewed papers.

Refer. no.	Specific elements in identification and profiling	Limitations
[13]	Searching for profiles or connections; Digital artefacts; User-generated content; Metadata Examination	Related just to the identification of the digital information and it verification.
[14]	Target search; Keyword techniques; Digital artefacts; User-generated content; Linguistic and Stylistic Analysis	The link between correctness and author set size reveals unreliable results, emphasizing differences in language term usage.
[15]	Searching for profiles or connections; Cross-Platform Analysis; Metadata Examination; Activity analysis; Temporal Analysis; Location Analysis	Described correlations cannot be characterised using durations and overlapping periods.
[16]	Digital artefacts; Metadata examination; Content analysis; Temporal Analysis; Location Analysis	The quality of the provided images may impact the accuracy of forgery detection, localisation.
[17]	Searching for profiles or connections; Cross-Platform Analysis; Cross-referencing; Social Network Analysis	The effectiveness depends on the availability and quality of data, potential data restrictions, and the presence of noise and misinformation.
[18]	Metadata examination; Content Analysis; Temporal analysis; Location analysis	Performance heavily relies on the availability of comprehensive provenance datasets for evaluation.
[19]	Searching for profiles or connections; Cross-referencing	Due to the unpredictable patterns of the random selection, little variance in the findings occurs.
[20]	Social Network Analysis; Searching for profiles or connections	Predicting social network interactions based on geographical linkages.
[21]	Searching for profiles or connections; Cross-Platform Analysis; Metadata Examination; Activity and Behavioral Analysis	Comparative analysis of AI models in a case of digital transactions lacking focus on profiling of the users .
[22]	Keyword techniques;Digital artefacts; user-generated content;Cross-referencing and Data Integration	Models of AI focus only on image processing in digital forensics.
[23]	Digital artefacts; user-generated content; Metadata Examination; Linguistic and Stylistic Analysis; Cross-referencing and Data Integration	Models of AI focus only on image forensics.
[24]	Digital artefacts; Metadata examination; Content analysis; Location Analysis	Missing connections between the discovered evidence item and any existing evidence pieces.
[25]	Digital artefacts; Metadata examination; Content analysis; Location Analysis	Study focuses solely on fraud detection and information extraction from video materials; lacks profiling techniques.
[26]	Digital artefacts; Metadata examination; Content analysis; Temporal Analysis	The model's resilience to recognise many sorts of modified face images.
[27]	Searching for profiles or connections; Activity and Behavioral Analysis; Content Analysis; Sentiment Analysis; Linguistic and Stylistic Analysis	The reliability of the approach relies on the presence of a large volume of noisy and short texts, limited availability of metadata, and potential user anonymity.
[28]	Digital artefacts; User-generated content; Metadata Examination; Cross-referencing	Does not take into consideration all the intricacies and variances in devices, evidence items, data formats that occur in real-world settings.
[29]	Cross-referencing; Temporal Analysis	IoT forensics focused only on the phases of the digital forensics.
[30]	User-generated content; Metadata Examination; Content Analysis	Focus on the custody chain while lacking identification of the digital evidence.

TABLE 2. Continuation of summary of reviewed papers.

Refer. no.	Specific elements in identification and profiling	Limitations
[31]	Digital artefacts; user-generated content; Cross-Platform Analysis; Content Analysis; Temporal Analysis	Focus on the preservation of digital evidence via the establishment and maintenance of a secure chain of custody is facilitated by using the capabilities of Ethereum and Hyperledger.
[32]	Digital artefacts; user-generated content; Metadata Examination; Content Analysis; Cross-referencing and Data Integration	The use of blockchain technology for the purpose of notarizing material generated inside social media networks.
[33]	Keyword techniques; Searching for profiles or connections; User-generated content; Social Network Analysis; Activity and Behavioral Analysis; Content Analysis; Sentiment Analysis	A system designed to recognise and analyse the sentiment expressed in social network opinions lacking focus on the identification of the digital evidence.
[34]	Searching for profiles or connections; Digital artefacts; Activity and Behavioral Analysis; Content Analysis	Not taken into account the sentiment analysis while users exchanging messages.
[35]	Target search; Keyword techniques; Activity analysis; Content Analysis; Cross-referencing and Data Integration	It may not be able to identify sophisticated bots fooling standard detection methods.
[36]	Keyword techniques; Digital artefacts; Activity and Behavioral Analysis; Sentiment Analysis; Linguistic and Stylistic Analysis	Missing investigation for the language and information of the profile.
[37]	Searching for profiles or connections; Digital artefacts; Activity and Behavioral Analysis; Content Analysis; Cross-Platform Analysis; Metadata Examination	The data gathering techniques depend on the premise that users actively publish their data openly, and the proposed approaches are tested using one particular social media site.
[38]	Target search; Searching for profiles or connections; Digital artefacts; Metadata Examination; Social Network Analysis; Activity and Behavioral Analysis	The provided solution does not provide a comprehensive analysis and interconnectedness of artefacts and attributes.

tweets while purporting to be the same person. It's possible that celebrities may pay marketing companies to boost their internet profile by tweeting on their behalf and helping them establish an audience. On the other hand assessing the credibility of media objects, particularly in conjunction with their associated textual information, complicates the forensic process. The association of text and media can either obfuscate or provide context, necessitating sophisticated verification algorithms.

Graph theory allows for the research of social networks by portraying social media platforms as graphs with users as nodes and connections as edges [18], [19], [20], [21]. Various approaches, such as centrality measures and community discovery algorithms, can be used to identify prominent individuals, locate clusters of users with similar behaviours, and uncover potentially dangerous actions. The extensive and heterogeneous characteristics of data on social media may be seen as a complex and interconnected graph with a multitude of nodes and edges. Conventional data mining techniques may encounter challenges in unraveling the complex interrelationships and patterns included in this network, thereby emphasizing the necessity for tailored forensic methodologies. The suggested approaches, which were mentioned above, places significant emphasis on identifying potential correlations and aims to show subgraphs or particular patterns within the broader network. Nevertheless, this particular emphasis may fail to consider other noteworthy graph topologies or linkages. Utilizing such models, automation would proficiently go over such a graph, actively searching for patterns or abnormalities. However, the dependability of the models depend on their capacity to accurately capture the intricate details of the graph and establish a reliable connection between each node (representing evidence) and its original source within the vast social media network.

Link analysis reveals hidden ties [22], powerful users, and partnerships, allowing the detection of networks involved in cyberbullying or the distribution of illegal information. This research acknowledges the occurrence of false-positive results and in the context of digital evidence identification, false-positives can lead to misinterpretation or incorrect conclusions. According to the authors [3], there is a theory of weak relations that observes that nodes with a large geodesic distance and a feminine account signature have a significant impact on malevolent activity due to a desire to comprehend the user. It can be stated that the theory's constrained scope might unintentionally overlook pertinent nodes or significant factors. Relying on a criterion characterized as a feminine account signature potentially introduces gender-related biases, particularly in light of the ambiguity surrounding this descriptor's specifics. Ascribing malevolent actions purely to the desire to comprehend the user may represent a reductive perspective, given that underlying motivations for maleficent behaviors can be complex and varied. An undue emphasis on geodesic distance may neglect the examination of more nuanced network interrelationships. Furthermore, the assumption that nodes adhering to this criteria exhibit an inherent propensity for malevolent actions could lead to inaccurate identifications. Due to this the broader applicability of this theory across a spectrum of network configurations warrants further exploration.

Advanced techniques as machine learning and AI are used to assist the identification of digital evidence on social media [23], [24] [25]. As digital cameras and mobile devices become more commonplace as sources of evidence on social media, a growing number of authors are looking at the identification of significant forensic evidence components, their piecing as evidence items, or the establishment of the connections between evidence items in specific cases [26], [27], [28]. Other authors in [29] proposed a method for grouping Twitter authors that combines feature extraction with the transformation of high-dimensional data into a kernel matrix. They suggested a system that automatically collects Twitter data, extracts features relevant to author clustering on Twitter, and employs an unsupervised learning module. This research emphasizes authorship analysis to counteract the inherent anonymity challenges in Internet services, particularly on the Twitter platform where users can control multiple accounts. However, relying solely on this method could overlook other crucial behavioral or technical

indicators. The study's dependency on an automated unsupervised learning approach for Twitter author clustering, while innovative, may not capture the nuances present in labeled data. Furthermore, the transformation of high-dimensional data to a kernel matrix might inadvertently lose significant information. The methodology's focus on data from an automatic Twitter collection module raises questions about data quality and representativeness. Moreover, while the method claims superior efficacy in clustering large numbers of accounts, precise performance metrics in comparison to existing methodologies remain unspecified. Lastly, the approach's adaptability to platforms other than Twitter remains unexplored.

Initial forensic study on the blockchain-based forensic investigation was published in [9]. The researchers tried to find out what links proof items, where they came from, how they can be traced, and how they can be checked, taking into account the variety of devices, evidence items, data types, and more in the complex IoT environment. Another study in [30] examines the advantages of utilizing blockchain technology in the field of digital forensics and provides an overview of the latest blockchain solutions designed for IoT forensic frameworks. Nevertheless, the studies may encounter difficulties stemming from the extensive range and continuous development of Internet of Things (IoT) devices and data types. The potential implementation of transferring analyzed data to a framework based on blockchain technology may give rise to problems regarding efficiency, security, and verification. Moreover, the research highlights the significance of establishing strong associations between evidence items and their origin, ability to be traced, and capacity to be audited. However, achieving consistency in these aspects across a wide range of devices poses a challenging endeavor. The scalability of the framework within a continuously increasing Internet of Things (IoT) environment and its capacity to quickly adapt to rapidly evolving technology are aspects that may raise concerns.

The immutability of Blockchain technology presents considerable promise in the field of digital forensics since it preserves the validity and integrity of data [31], [32] [33]. The decentralized structure of the system offers a level of resistance against both data loss and tampering. Additionally, the utilization of timestamps can effectively establish essential dates. Smart contracts have the capability to automate forensic procedures, hence facilitating investigations pertaining to cryptocurrencies [34]. Nevertheless, the integration of blockchain into digital forensics requires careful consideration due to persistent obstacles such as scalability, privacy issues, and acceptance limitations.

In social media forensics, user profiling methods entail mainly analysis of the identified and collected variety of data sources to obtain insight into user characteristics and behaviours. This includes examining account information, conducting social network analysis to identify connections and communities, analysing activity and behaviour patterns, examining content shared via text, images, and multimedia, conducting temporal and location analysis, conducting sentiment analysis, linguistic and stylistic analysis, and integrating data from multiple sources. By taking into account these factors, investigators can construct exhaustive profiles that disclose the identities, interests, affiliations, and intentions of users.

The authors in [35] presented a method for user profiling that analyses the non-volatile data that is still present on digital devices. It establishes that a user has a propensity and an impact, both of which point to patterns of applications use. The presented work primarily focuses on non-volatile data from digital devices, potentially overlooking the dynamic nature of evidence on diverse social media platforms. While previous research has been limited to specific applications or devices, this approach might not fully capture the complexities of social media, especially when considering volatile content like live streams or chats. The method's emphasis on ''tendency'' and ''impact'' might not encapsulate the multifaceted interactions on social media, and while it attempts to distinguish users, the vast and intricate nature of social media data can challenge accurate differentiation. Additionally, the approach static perspective may not account for the evolving behaviors of users over time, and the use of non-volatile data for profiling might raise significant privacy concerns.

To help identify social bots, a machine learning model with high prediction accuracy was presented in [6] that profiles users based on personal information gleaned from their online posts, including age, personality, gender, and level of education. This method leaks accuracy in the case of different user personal information. The authors in [36] sought to achieve a similar end, but offered a different approach; they looked at the feasibility of identifying automated users by use of a fingerprint of user activity and a collection of statistical measurements describing various facets of that conduct. The approach simply catches the surface level of conduct, without delving into the languages and profiles of those involved.

Using the ideas of Advanced Search Operators (ASOs), Social Aggregators (SAs), Cross-Platform Sharers (CPSs), Self-Disclosers (SDs), and Friend Finding Features (FFFs), the researchers in [37] suggested a framework for user profiling within social media forensics. With the use of ASOs, investigators may do more precise searches and collect more relevant data on people. Social media aggregators (SAs) compile information from several sites to provide a unified picture of a user's digital life. CPS displays cross-platform material, illuminating a user's passions, networks, and habits of expression. SD is the practise of examining user-provided data for clues about their mentality, goals, and security, among other things. When applied to a user's network, FFF is able to detect affiliations, communities, and even potentially malicious behaviours. Together, these ideas facilitate the collection, analysis, and integration of data from many sources, resulting in a holistic profile of a user's identity, behaviours, interests, and possible connections. While user profiling benefits from a combined approach,

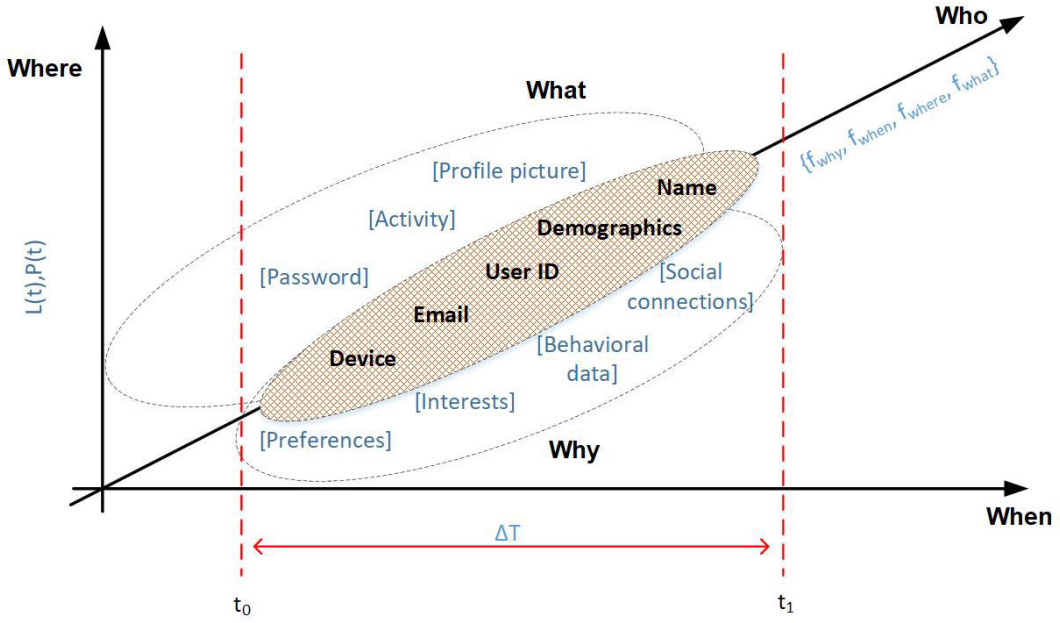

FIGURE 1. Interconnection between digital artefacts' domains.

there are several restrictions to be aware of. Potential data gaps might arise if ASOs don't account for all conceivable search conditions. Access to APIs and social media platforms is crucial for SAs, but they are not always freely available, which might limit or restrict data collection. Not all users will participate in cross-platform sharing, which will reduce the amount of data that can be collected via CPS. Users' purposeful provision of inaccurate or misleading information might undermine the accuracy of SD and its associated profiling. Inadequate or biassed information might underpin FFF suggestions, increasing the risk of incorrect inferences or missing important relationships. Furthermore, the accuracy and fairness of user profiling findings might be negatively impacted by depending entirely on these methodologies without taking into account contextual elements, data integrity, legal and ethical issues, and possible biases.

Another work proposes a framework and architecture that address the limits of data visualisation [38]. The proposed framework does this by resolving the fundamental 5W issues. However, there is a lack of links between artefacts and attributes for the identification of digital evidence, thus raising issues about its trust.

Despite the investigations done in user profiling, there are still gaps and limitations in this field. User profiling in social media forensics is a complex task, which lacks integration of multiple data sources, balance of collected information with fairness in profiling results.

III. PROPOSED METHOD FOR IDENTIFICATION AND PROFILING OF DIGITAL EVIDENCE

The identification of digital evidence begins with the collection of digital footprints left by users on social media platforms. This may be accomplished by browsing public profiles or via the sites' APIs (Application Programming Interfaces). The gathered digital footprints can be categorized in domains by time, place, reason and nature of the action (Fig. 1).

Within these domains, each footprint is considered as a separate digital artefact, some of which may have common links to one another and also share features with other digital artefacts. In a digital investigation, the process for the collection of digital artefacts refers to the specific starting point, t_1, which denotes the investigation's beginning. The investigation then collects digital artefacts over a period of time ΔT until it reaches t_0, the time at which the beginning of a potential digital crime is determined. Such details illustrate the time and/or duration of the action (*When*), which was done by a subject who is suspected. During this time period, investigators identify digital artefacts that may support or refute claims or suspicions related to the case. Supporting or refuting suspicions depends on collected digital artefacts, which represent: a) specific actions of the investigated subject (*What*); b) the reasons or motive behind the actions (*Why*); and c) the location L or place P of the action (*Where*). Based on that, the following interconnection is described (Eq. 1):

$$Who = f\{What, Why, Where, When\}, \ where$$
$$What \vee Why \vee Where \vee When \neq \emptyset \quad (1)$$

Such interconnection results in *Who* being functionally dependent on the combination of sets of artefacts and attributes in each of the domains, which are not empty sets. The *Who* indicates the combined influence of someone's activities, reasons or motives, location, and time. This provides a quantitative measure that shows the total impact

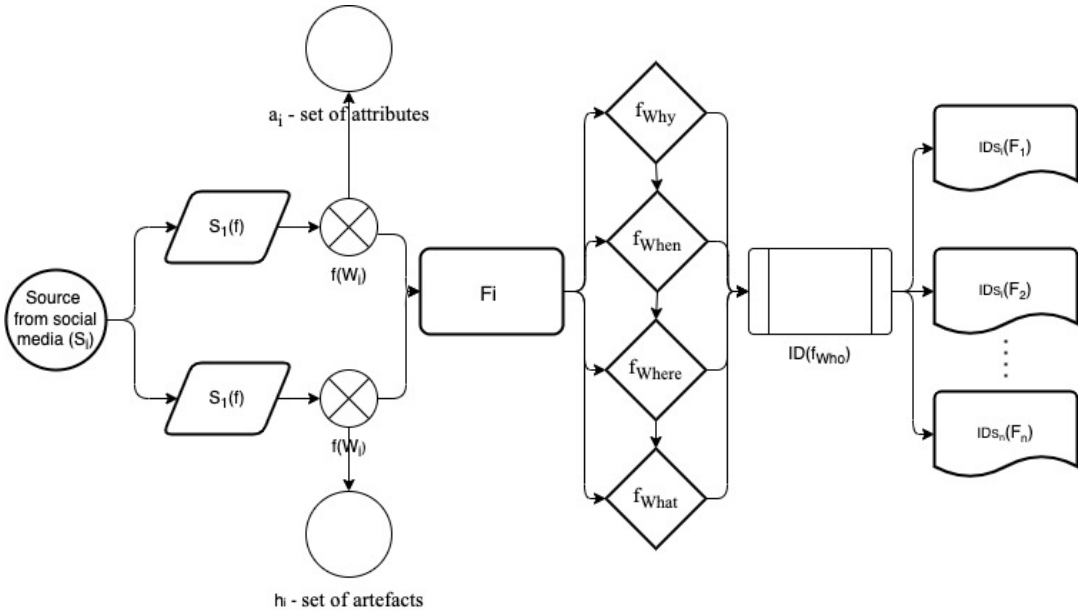

FIGURE 2. A method for identification of digital evidences.

or significance of these aspects in the profile or behavior of the investigated subject.

Once the artefacts are collected, the next step is identification of the digital evidence (Fig. 2). The identification of the digital evidence starts with the analysis of the metadata of the artefacts, which are collected from social media source(s) S_i for entry i. During metadata analysis, attributes are extracted and associated with the collected artefacts through a combination of manual examination and specialized tools. Manual examination involves inspection of the artefact and its associated metadata to identify relevant attributes such as timestamps, location data, author details, file formats, *etc*. Additionally, metadata extraction tools automatically parse and extract metadata from specific file types, saving time and effort. It is important to keep in mind that the set of artefacts and the set of attributes may have originated from the same source, or they may have arrived from distinct sources. Either way, this is a possibility.

Based on the extracted set of attributes a_i as well as set of artefacts h_i, the identification of the digital evidence F_i depends on the categorization for the domain of these sets in relation on their quantity and weight (Eq. 2):

$$F_i = (w_a \cdot \sum(a_i \cdot s_i)) + (w_h \cdot \sum(h_i \cdot s_i)) + (w_s \cdot S_i), \quad (2)$$

where a_i represents the set of attributes of entry i; h_i - the set of artefacts of entry i; s_i is the weight of entry i; w_a is defined as the weight assigned to attributes, w_h - the weight assigned to artefacts and w_s - the weight assigned to the source of social media.

The weights are considered to determine the importance or relevance of the attributes and artefacts from social media to the appropriate evidence. Also, $\sum(a_i \cdot s_i)$ represents the sum of all attributes of entry i each multiplied by its weight and $\sum(h_i \cdot s_i)$ represents the sum of all artefacts of entry i each multiplied by its weight.

It is important to mention, that the specific domain depends on the objectives of the profiling method. In this work, the domains are defined by actions, reasons, location and time (4W's). Following the categorization of the evidence, individual profiles for a possible suspect $ID(f_{Who})$ are subsequently developed based on the domain of digital evidence from the source of social media for entry i (Eq. 3):

$$ID(f_{Who}) \in (ID_{S_1}, ID_{S_2}, ID_{S_3}, ID_{S_4}) \quad (3)$$

These profiles include the relevant digital evidences that fall under each of domain (Eqs. 4-7):

$$ID_{S_1} = f_{Why} \cdot F_1 \quad (4)$$
$$ID_{S_2} = f_{When} \cdot F_2 \quad (5)$$
$$ID_{S_3} = f_{Where} \cdot F_3 \quad (6)$$
$$ID_{S_4} = f_{What} \cdot F_4 \quad (7)$$

The profiles offer a synopsis of the qualities, interests, and behaviors of a subject that falls under a specific domain. Here f_{Why} represents the interconnection between artefacts and relevant attributes that are included for profiling the reasons or motives of the suspected subject for the actions; f_{When} indicates the interconnection between the artefacts and the attributes that are included in profiling the time and/or duration of the actions that were carried out by the subject of suspicion; f_{Where} presents the connectivity between artefacts and associated attributes that are included for profiling the location and position of the suspected subject as well as the detected activities; f_{What} points out the link between artefacts and corresponding attributes that are used to profile the suspected subject's individual behaviors.

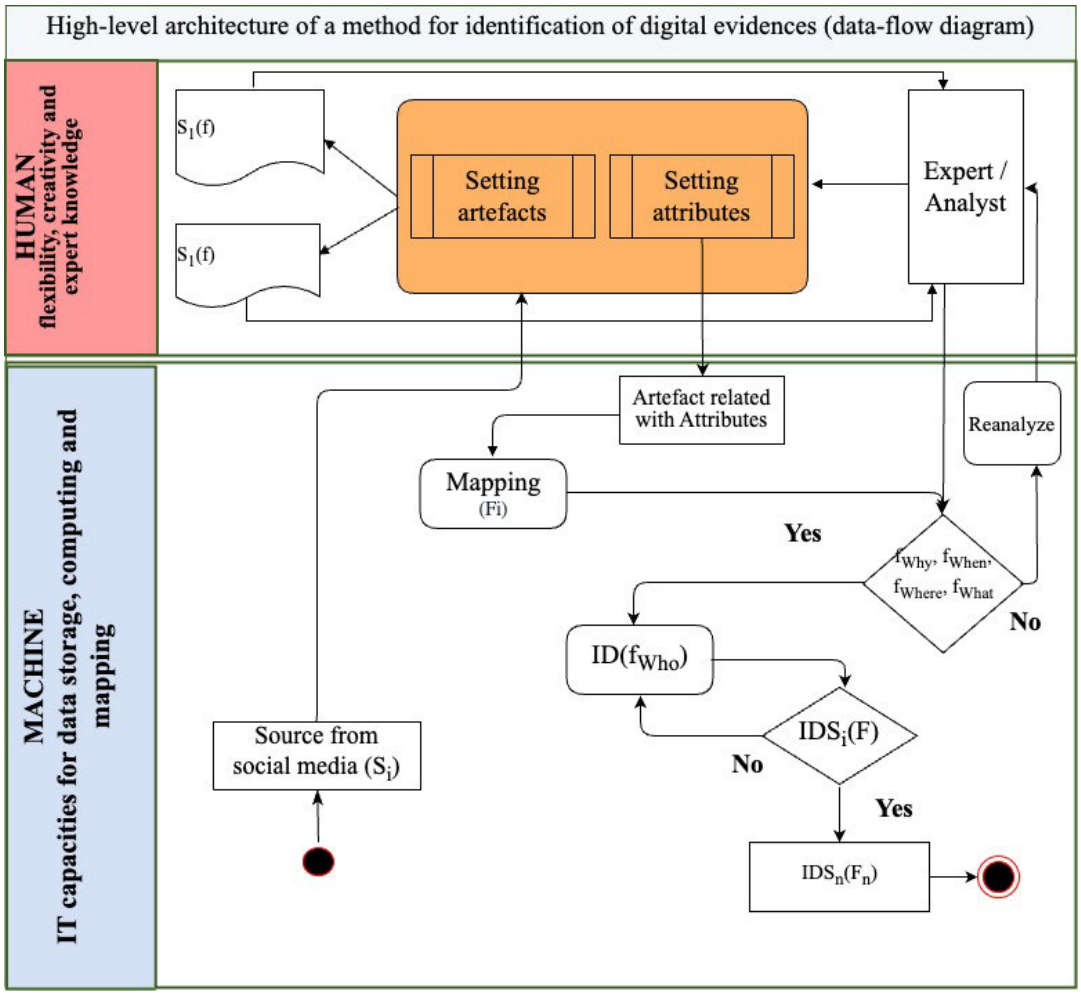

FIGURE 3. High-level architecture of the proposed method.

The artefacts, which serve as key features in profiling the suspected subject, are as follows:

$$f_{Why} = \{I, P, SC, B\}, \tag{8}$$

where: I is information about a subject's interests, and hobbies that can be used for personalization or targeted advertising; P is information about subject's preferences; SC presents information about a subject's relationships with other users within the social media platform or application, including friends, followers, or other social connections; B is information about a subject's behavior within the social media platform or application, including purchase history, search queries, or other actions.

$$f_{When} = \{Be, En, \Delta T\}, \tag{9}$$

where: Be is the time at which the beginning of a potential cybercrime is determined (t_0); En is the time at which the investigation starts (t_1) and ΔT is a period of time during that the investigators are identifying and collecting the digital artefacts related to the suspected subject.

$$f_{Where} = \{L, P, T\}, \tag{10}$$

where: L describes the absolute location from the geographical point of view; P is a suspected subject's physical location or the location of the device they are using to access social media platforms; T is a timestamp, which refers to a time and date when the suspected subject's location was recorded or updated on social media platforms.

$$f_{What} = \{A, Pas, Pro\}, \tag{11}$$

where: A refers to the activity that covers information about a suspected subject's interactions with social media platforms, including logins, clicks, views, and other actions; Pas describes all details related to the password that was used to access social media platforms; Pro is a suspected subject's chosen image that is used to represent them within the social media or application.

By following the above-mentioned steps,the following represents the process of constructing profiles of suspected

subjects by considering digital evidence across different domains and social media sources:

$$Who \in f(ID_{S_i}(F_1), ID_{S_i}(F_2), \ldots, ID_{S_n}(F_n)) \quad (12)$$

The objective of the profiling is to ascertain the inclusion of the suspected subject referred to as *Who* within the profile of a suspected subject derived from a particular social media source denoted as S_i. This determination is achieved by employing the profiling function to analyze the digital evidence $(F_1, F_2, \ldots, F_n)$ acquired from different domains (4W's) encapsulated within the particular profile on particular social media platform denoted as ID_{S_i}.

Figure 3 gives an in-depth representation of the proposed method, illustrating its high-level architectural structure. It is important to note that a high-level architecture highlights an interconnection between human and machine processing in the overall process of identifying digital evidence. An expert in the field of digital forensics is someone who has specialized knowledge and skills in the identification, collection, profiling, analysis and interpretation of digital evidence. They can assist with a variety of tasks, such as:

- identifying digital evidence;
- collecting digital evidence;
- profiling digital evidence;
- analyzing digital evidence;
- interpreting digital evidence;
- presenting digital evidence in court.

The reanalysis process is a method of revisiting digital evidence that has already been analyzed. This can be done for a variety of reasons, such as:

- to confirm the original analysis;
- to identify new evidence;
- to update the analysis to reflect new developments in technology;
- to address challenges to the original analysis.

The reanalysis process can be complex and time-consuming, but it can be an important tool for ensuring the accuracy and completeness of digital evidence analysis.

The first stage (see Figure 3) involves collecting digital artefacts from social media sources. This can be accomplished manually or automatically. The digital artefacts may consist of posts, communications, images, videos, and other types of content. Once the artefacts have been accumulated, their metadata is analyzed to extract pertinent attributes. This includes timestamps, location information, author details, and file formats. The metadata can be used to determine the origin of the artefact, its creation date and time, and the user who created it.

The extracted attributes are utilized to identify digital evidence. This is accomplished by categorizing the evidence into domains including actions, motives, location, and time. A post containing the location data. i.e., "London, UK" and the time signature, i.e., "2023-06-30 12:00:00" would be considered as digital evidence of the location domain.

Based on the digital evidence, individual profiles are then developed for potential suspects. These profiles contain details regarding the suspect's interests, motivations, activities, and location. For instance, a profile of a suspect who is interested in travel and has posted pictures of themselves in London, United Kingdom, would likely include the location domain as a prominent characteristic.

Determining whether or not a specific subject is included in a profile is the final phase. This is accomplished by analyzing the digital evidence in the profile and comparing it to the known attributes of the suspect. For instance, if a suspect's profile includes the location domain and the suspect has been spotted in London, UK, it is likely that the suspect will be included in the profile.

A complex data flow is required for the identification and profiling of digital evidence on social media. However, the scientific method outlined above can be utilized to identify and profile digital evidence in a systematic and rigorous manner.

The set of attributes associated with digital artefacts plays an important role in profiling suspects, as it not only enables the creation of a comprehensive and detailed profile but also facilitates the process of linking various social media sources. The attributes that are included in each of 4W's domain are following (see Fig. 4):

1) Domain Why:
   - Interest Category - the broad category of the suspect's interest, such as sports, music, or fashion.
   - Interest Subcategory - the specific subcategory or topic within the suspect's interest category, such as basketball, jazz music, or high fashion.
   - Interest Level - the level of interest the suspect has in the particular interest, which can be measured by the frequency of engagement or activity related to that interest.
   - Interest Source - the source of the suspect's interest data, such as user-generated content, likes, comments, or pages followed.
   - Interest Trends - the trends and patterns in the suspect's interest data over time, which can be used for predictive analysis and recommendation systems.
   - Interest Affinity- the level of affinity or similarity between the suspect's interests and the interests of other users or groups, which can be used for social network analysis and recommendation systems.
   - Interest Impact - the potential impact of the suspect's interests on their behavior or decision-making, such as the influence of political or ideological interests.
   - Preference Category - the broad category of the suspect's preference, such as food, travel, or entertainment.
   - Preference Subcategory - the specific subcategory or topic within the suspect's preference category,

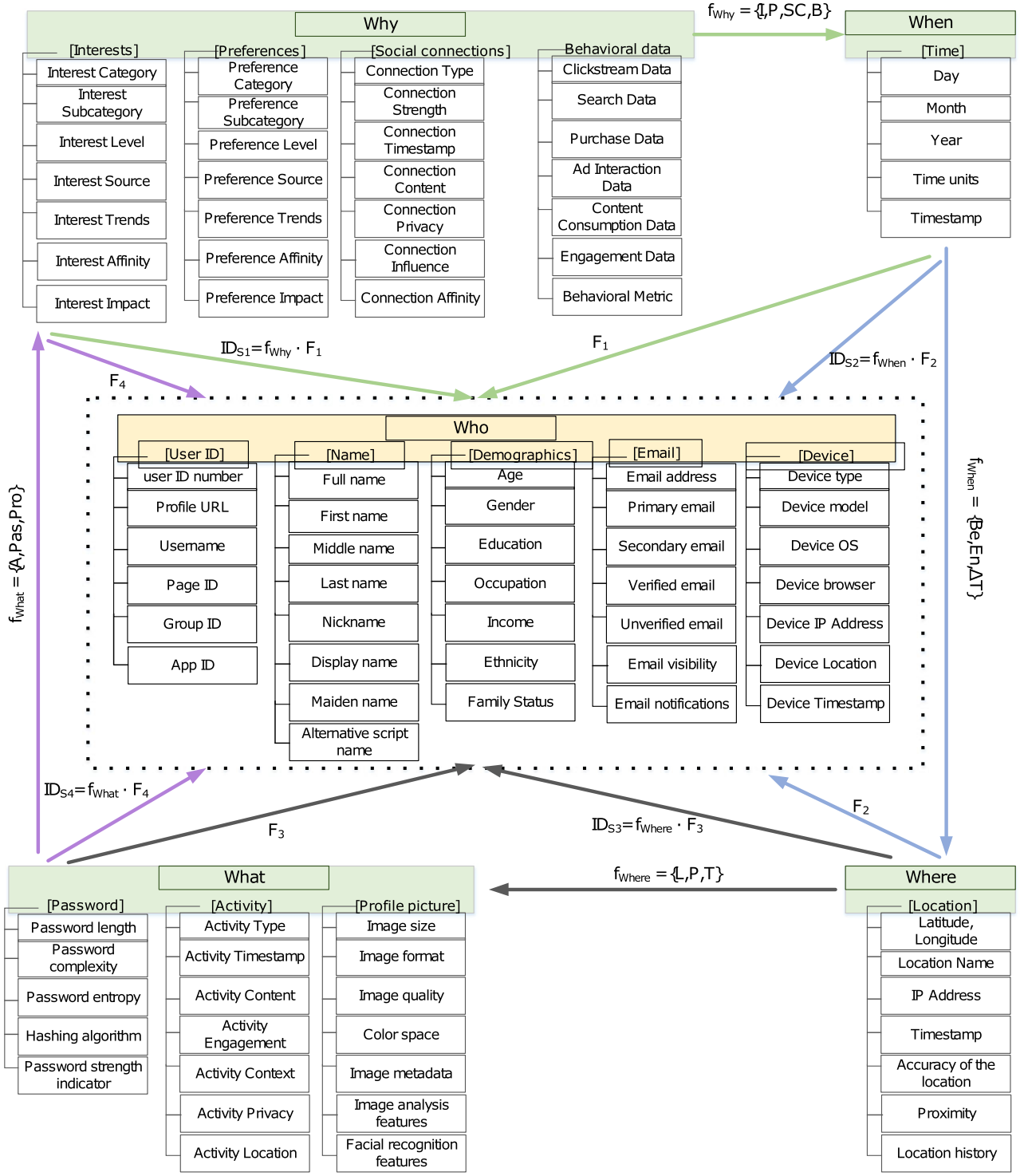

FIGURE 4. Profiling of a suspected subject.

such as Italian cuisine, adventure travel, or live music.

- Preference Level - the level of preference the suspect has for the particular preference, which can be measured by the frequency of engagement or activity related to that preference.
- Preference Source - the source of the suspect's preference data, such as user-generated content, likes, comments, or pages followed.
- Preference Trends - the trends and patterns in the suspect's preference data over time, which can be

used for predictive analysis and recommendation systems.
- Preference Affinity - the level of affinity or similarity between the suspect's preferences and the preferences of other users or groups, which can be used for social network analysis and recommendation systems.
- Preference Impact - the potential impact of the suspect's preferences on their behavior or decision-making, such as the influence of political or ideological preferences.
- Connection Type - the type of connection between the suspect and their social connection, such as friend, follower, group member, or page fan.
- Connection Strength - the strength or intensity of the connection between the suspect and their social connection, which can be measured by the frequency of interaction or engagement.
- Connection Timestamp - the date and time when the connection was established or last updated, which can be used for temporal analysis and trend detection.
- Connection Content - the content or topic of the connection, such as the common interest, hobby, or activity shared by the suspect and their social connection.
- Connection Privacy - the privacy settings of the connection, which can be used for privacy analysis and data protection compliance.
- Connection Influence - the level of influence or impact that the social connection has on the suspect's behavior, decisions, or opinions.
- Connection Affinity - the level of affinity or similarity between the suspect and their social connection, which can be used for social network analysis and recommendation systems.
- Clickstream Data that is related to the suspect's clicks and navigation on social media, including pages visited, buttons clicked, and search queries.
- Search Data - the data related to the suspect's search queries on social media, including keywords searched, search history, and search results clicked.
- Purchase Data - the data related to the suspect's online purchases on social media, including transaction history, product categories, and purchase frequency.
- Ad Interaction Data - the data related to the suspect's interaction with social media ads, including ad views, clicks, and conversions.
- Content Consumption Data - the data related to the suspect's consumption of social media content, including posts viewed, shared, and commented on.
- Engagement Data - the data related to the suspect's engagement withsocial media content, including likes, comments, and shares.
- Behavioral Metrics that are used to measure the suspect's behavior on social media, including session duration, frequency of visits, and time of day activity.

2) Domain When includes data related to timing factors such as a day, a month, a year, time zones and timestamps related to a suspect's actions on social media platforms.
3) Domain Where:
   - Latitude and Longitude - the coordinates of the suspect's location on the earth, typically expressed in degrees.
   - Location Name of the suspect's location, such as a city, state, or country.
   - IP Address associated with the suspect's location.
   - Timestamp - the time and date when the suspect's location was recorded or updated on social media.
   - Accuracy of the suspect's location data, which can vary based on the source of the data and the method used to determine the location.
   - Proximity of the suspect's location to other users or landmarks, which can be used for social network analysis or location-based marketing.
   - Location History - a record of the suspect's past locations, which can be used for location-based analysis and recommendation systems.
4) Domain What:
   - Password length - the number of characters in the password.
   - Password complexity - the level of complexity of the password, such as the use of upper and lower case letters, numbers, and special characters.
   - Password entropy - a measure of the randomness and unpredictability of the password.
   - Hashing algorithm used to convert the password into a hashed value for storage and comparison.
   - Brute-force resistance - the ability of the password to resist attacks that attempt to guess or crack the password through trial and error.
   - Password strength indicator - a value or metric indicating the strength of the password.
   - Activity Type performed by the suspect, such as posting, commenting, liking, sharing, or following.
   - Activity Timestamp - the date and time when the activity was performed by the suspect, which can be used for temporal analysis and trend detection.
   - Activity Content - the content of the activity, such as the text, image, video, or link that was posted, commented, or shared.
   - Activity Engagement - the level of engagement or interaction of the suspect's activity, such as the number of likes, comments, or shares received.

- Activity Context - the context or topic of the suspect's activity, such as the group, page, or hashtag associated with the activity.
- Activity Privacy settings of the suspect's activity, which can be used for privacy analysis and data protection compliance.
- Activity Location or geotag of the suspect's activity, which can be used for geospatial analysis and location-based marketing.
- Image size - the dimensions of the profile picture, which can be expressed in pixels or inches.
- Image format - the file format of the profile picture.
- Image quality - the level of compression used when saving the profile picture, which can affect the image quality and file size.
- Color space - the color space used in the profile picture, such as RGB, CMYK, or grayscale.
- Image metadata - additional information about the profile picture, such as the date and time it was uploaded, the device used to capture the image, and any editing or processing that was applied.
- Image analysis features extracted from the profile picture using image processing techniques, such as edge detection, color histogram, and texture analysis.
- Facial recognition features extracted from the profile picture using facial recognition algorithms, such as facial landmarks, expression, and gender.

By leveraging these attributes, which encompass a wide range of information derived from digital evidence, investigators can construct a thorough and multifaceted depiction of the suspect. Furthermore, these attributes serve as valuable connecting points that allow for the integration and correlation of data obtained from different social media platforms. By considering attributes in 4W's domains, investigators can establish meaningful links between diverse sources of social media content. This linkage greatly enhances the effectiveness and depth of the profiling process by providing a more comprehensive understanding of the suspect's activities, connections, and behaviors across multiple social media platforms.

IV. CASE STUDY

A. A DIGITAL ONTOLOGY-BASED IDENTIFICATION OF DIGITAL EVIDENCE

A visual model such as a digital ontology-based diagram aims to capture and illustrate the relationships between various digital artefacts and attributes shared and connected across different social media platforms in order to identify common digital artefacts related to the suspect that are interlinked across social media platforms. The digital artefacts that are common and can be used to draw a part of a suspect's profile ontology (Fig. 5) over Twitter, Facebook, Instagram, and other social networks, are the following:

- user ID - a unique identifier for each user on the social media platform;
- username - the name that the user has chosen for their profile;
- profile picture - the user's chosen profile picture or avatar;
- bio/about me - a short description of the user, usually written in the first person;
- location - the user's current location or the location they have chosen to share on their profile;
- website - a link to the user's personal website or blog;
- followers/following - the number of other users who follow the user and the number of users the user follows;
- posts/activity - the user's activity on the social media platform, such as the number of posts they have made, likes, shares, comments, and engagement with other users;
- interests - the user's interests, likes, and dislikes, often expressed through their posts and engagement with content on the social media platform;
- connections - the user's connections to other users on the social media platform, such as friends, family, colleagues, or interest groups;
- privacy settings - the user's chosen privacy settings for their profile, posts, and activity on the social media platform;
- analytics - data on the user's engagement with their own content and the content of others, such as views, clicks, and shares.

These artefacts may vary slightly depending on the particular social network and its specific features, but they basically serve as a starting point for building a suspect's profile ontology for social networks.

Another ontology (Fig. 6) includes a super-class of "Social Media Platform", with "Twitter" and "Facebook/Instagram" as sub-classes. Each social media platform has a "User Profile" class (that is considered as a suspect profile) with specific attributes gained from social media.

To provide a mathematical model, the "User Profile" class and its attributes can be represented as a set of variables. For example, the following notation can be used:

$$\begin{aligned} UserProfile = \{&id, name, screen\_name, location, url,\\ &description, protected, followers\_count, friends\_count,\\ &listed\_count, created\_at, favourites\_count, verified,\\ &statuses\_count, is\_translator, profile\_image\_url\_https,\\ &default\_profile\_image, translator\_type, email\} \end{aligned} \quad (13)$$

Each variable represents a specific attribute of a suspect profile on a social media platform, with the "User Profile" set representing the entire user profile. Depending on the specific use case or research query, the mathematical model may include additional functions or operations to analyse the data in greater depth.

An ontology model for the identification and preservation of evidence on social media is presented in Fig. 7. In this

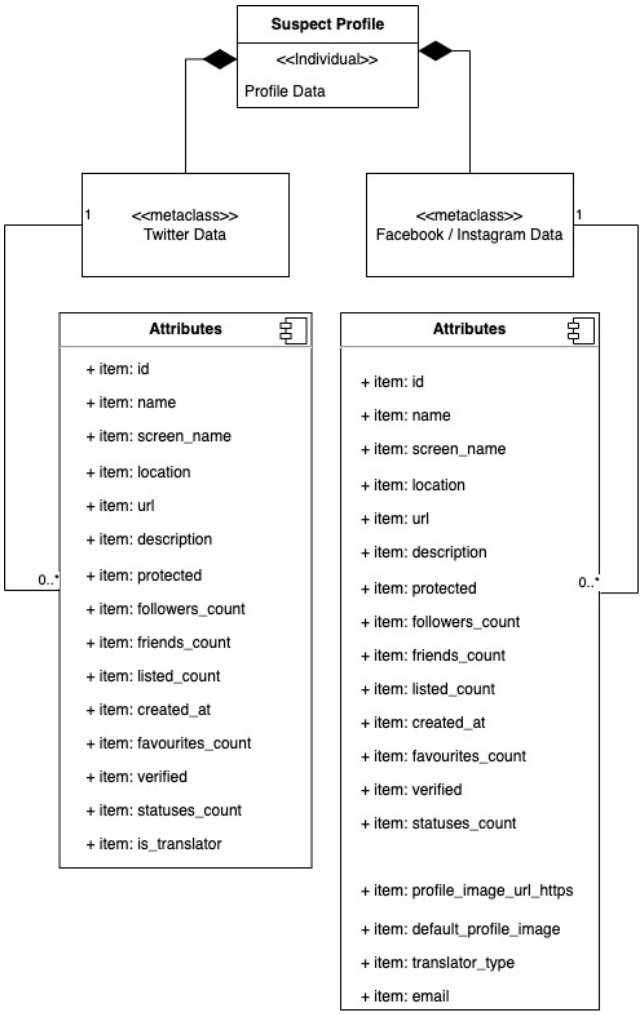

FIGURE 5. Suspect profile's ontology for various social media platforms.

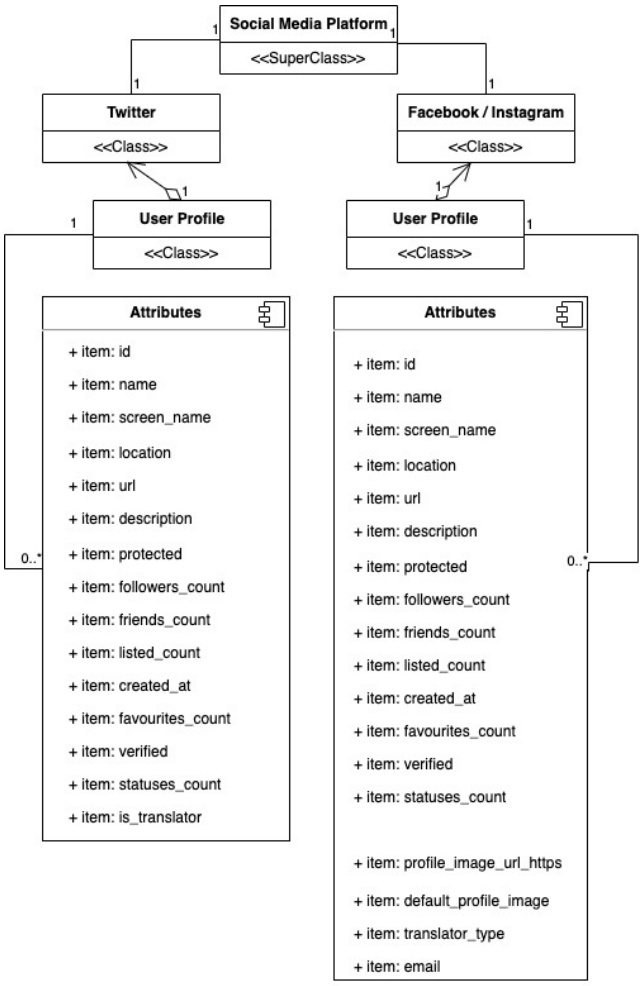

FIGURE 6. Profiling of suspect in particular social media platform.

ontology, ''Evidence'' is the main class, which has three sub-classes: ''Text'', ''Image'', and ''Video''. These sub-classes represent the different types of evidence that can be found on social media platforms. The ''Text'' subclass has two sub-classes: ''Tweet'' and ''Comment'', which represent the two main types of text-based evidence on social media. The ''Image'' subclass has one subclass ''Photo'', which represents images that are uploaded to social media platforms. The ''Video'' sub-class has one subclass ''Video Clip'', which represents short video clips that are uploaded to social media platforms. In addition to the sub-classes, each class and sub-class has an attribute that describes the specific evidence, such as the content of a tweet or the image file for a photo. Such ontology can be used to identify and preserve digital evidence on social media platforms by providing a structured framework for organizing and categorizing different types of evidence. It can also be used to help researchers and investigators analyze social media data by providing a consistent and standardized way to classify evidence.

The combined digital ontology model for social media platforms is presented in Fig. 8. In this ontology, ''Social Media'' is the main class, which has three subclasses: ''Twitter'', ''Facebook'', and ''Instagram''. These subclasses represent the different social media platforms. Each social media platform has its own specific types of digital evidence, such as ''Tweet'' and ''Retweet'' for Twitter, ''Post'' and ''Comment'' for Facebook, and ''Photo'' and ''Video'' for Instagram. Each type of evidence has its own specific attributes that describe the evidence, such as the content of a tweet or the file for a photo. This ontology can be used to help identify, categorize, and analyze evidence on different social media platforms in a structured and standardized way. It can also be extended to include additional social media platforms and types of evidence as needed.

B. AN EXPERIMENTAL SUSPECT PROFILING

A suspect profiling method is based on domain-specific digital artefacts defined by set of digital attributes gained from social media. The above provided mathematical background were used for a visual representation of the suspect's

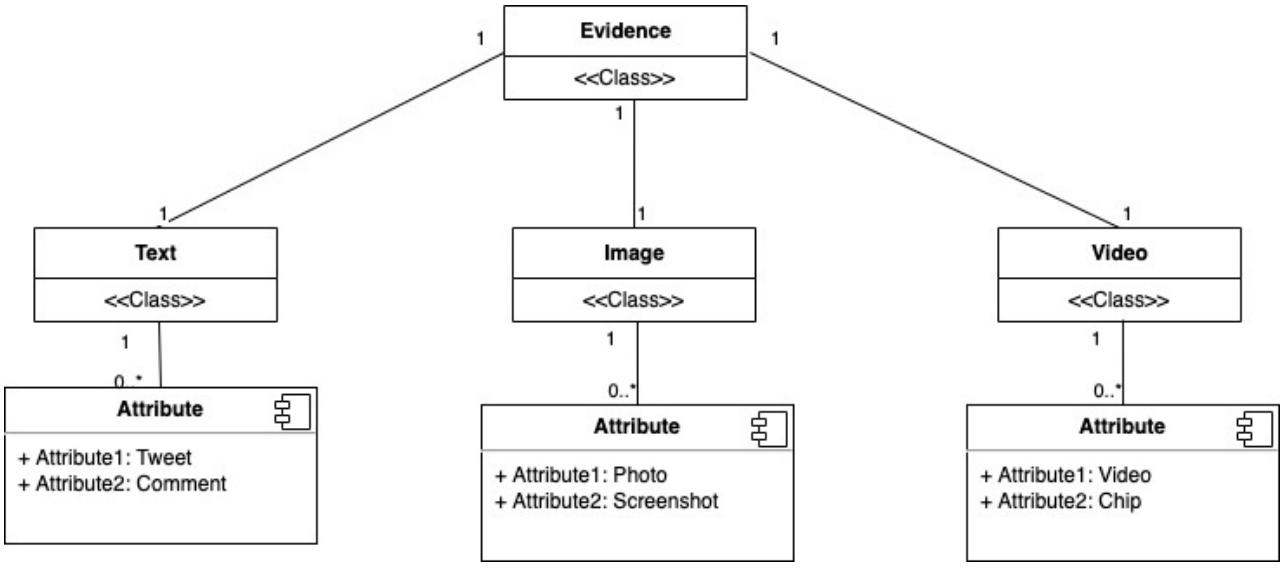

FIGURE 7. Ontology for digital evidence identification.

social network, including their friends, followers, and other connections. Such methodology can help law enforcement agencies streamline the process of collecting and analyzing digital evidence, thereby contributing to more efficient and effective cyber-crime investigations.

The first step in the process is to collect relevant data from social media platforms. This involves:

1) Identifying the target social media accounts associated with the suspects;
2) Acquiring necessary legal permissions and following platform-specific guidelines for data collection;
3) Extracting all available information from the accounts, including profile information, posts, comments, likes, shares, friends, followers, and other connections.

To create the synthetic dataset, MATLAB software was used to generate social media profiles and connections for a set of suspects and their contacts. The dataset will include fields such as "id", "name", "screen_name", "location", "url", "description", "protected", "followers_count", "friends_count", "listed_count", "created_at", "favourites_count", "verified", "statuses_count", "is_translator", "profile_image_url_https", "default_profile_image", "translator_type", and "email". This synthetic data will serve as the basis for this case study.

It is important to note that the data gathered must be pre-processed before it can be analysed to verify its relevance and correctness. It covers filtering away irrelevant or noisy data, detecting and resolving data inconsistencies, such as duplicate entries or missing information, and formatting the data into an appropriate structure for future analysis.

Using the information provided and the method outlined, the artefacts that are presented in Table 4 were identified from various social media sources. The attributes related to the artefacts in the Table 4 are derived from the methodology given above, that covers f_{Who}, f_{What}, f_{Why}, f_{When}, and f_{Where} (Table 3). These attributes are essential in creating an understanding of data gathered from social media platforms.

TABLE 3. Description of the artefacts in a specific domains.

Domain	Description of artefacts
Entry ID	The unique identification number assigned to each social media profile.
f_{Who}	The name of the subject associated with the social media profile.
f_{Why}	Interests and social connections associated with the individual.
f_{When}	The time when the social media profile was created.
f_{Where}	The location information provided in the social media profile.
f_{What}	The activity, behavior, or role of the individual based on profile data.

For simplicity, the arbitrary values for weights were assumed to be as follows:

$$w_a = 0.3;\ w_h = 0.5;\ w_s = 0.2;\ s_i = 1 \tag{14}$$

Also, only the first two entries from the gathered data of subjects and specific attributes were considered:

$$\begin{aligned} h_i &= \{Name, Location\};\\ a_i &= \{followers\_count, friends\_count\};\\ S_i &= \{profile\_image\_url\_https\} \end{aligned} \tag{15}$$

The results for the digital evidence F_i calculated as a weighted score are presented in Table 5.

Using the provided data, the Table 4 has been updated considering that f_{Who} corresponds to different subjects from various sources across social media platforms. As it was mentioned earlier, the individual profiles for suspected subjects are based on the domains of identified evidence: ID_{S_1}, ID_{S_2}, ID_{S_3} and ID_{S_4} (see Table 6). It's worth noting

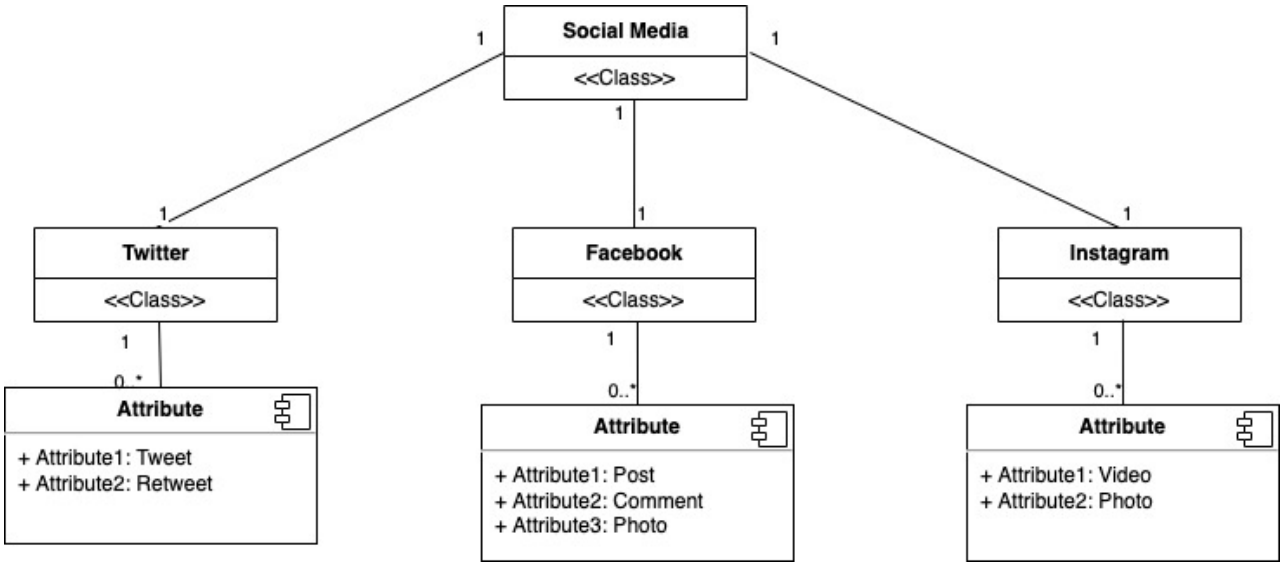

FIGURE 8. Combined digital ontology.

that the data provided doesn't contain explicit information regarding some attributes for domain f_{When}, which has a delta time. For simplicity and illustration, it will be excluded from further research.

For ID_{S_1} only "Interests" has been included as other artefacts such as "Preferences", "Social connections" and "Behavioral data" were not present in the sample data. Similarly, for ID_{S_2} only "Begin time" has been included as the "End time" and "ΔT" were not present in the data. For ID_{S_4} activities were included as they could be inferred from the descriptions of the profiles of social media users.

In order to find a single suspect "Who" from the data, it is needed to evaluate it against certain criteria. For this example, a combination of the artefacts and related attributes were considered:

- interests (f_{Why}) - the variety and depth of interests can indicate engagement and knowledge sharing potential;
- location (f_{Where}) - depending on the context, location might be a key factor;
- activity (f_{What}) - the activity or professional background can indicate expertise and potential for valuable contribution.

Given these criteria, here is the following analysis in regard to ID 544 (subject 4):

- interests are into payments, statistics, entrepreneurship, and are also an angel investor;
- location is based in San Francisco, a hub for technology and startups;
- activity is involved in business development and strategy and has led Visa Ventures.

This combination indicates that Subject 4 is likely to have a wealth of knowledge and expertise, especially in the domains of payments, investments, and entrepreneurship. Their location also places them in a key area for technological innovation. Therefore, based on the data provided and considering the attributes, ID 544, Subject 4 could be the "Who" for his potential to offer valuable insights and contributions in a technology and entrepreneurship context.

From the sample data provided, it's important to understand that unusual activity can mean a variety of things and doesn't necessarily imply wrongdoing. For the purpose of this example, it's assumed that unusual activity refers to an unusually high number of posts, followers, or other social interactions that stand out in comparison to the other data samples. Subject 2 is defined as having a relatively high number of followers (5766) compared to the others and has also made a very high number of posts (25942). Additionally, they have a high number of favourites (13606). This subject seems to be highly active on social media, which may be considered unusual compared to the average user. It's important to note that high activity on social media is not inherently suspicious and could be normal for someone who uses social media extensively for professional or personal reasons. In practise, detecting unusual activity that might be indicative of something nefarious requires a more sophisticated analysis and should be done carefully, taking into account context, behavioural patterns, and other factors.

From the analysed data, several relationships and attributes of the social media users were identified:

1) geographical proximity: several subjects are from California (Subject 1, Subject 2, Subject 3, and Subject 5). This could suggest that they might have a higher likelihood of being connected or having mutual friends or interests based on their geographical location.
2) professional background and interests: the description fields suggest some professional background and interests. For example, Subject 1 is a scientist and designer; Subject 2 has a background in music and design and has worked at notable companies like

TABLE 4. Identified artefacts.

Entry ID	f_{Who}	f_{Why}	f_{When}	f_{Where}	f_{What}
220	Subject 1	Interests: scientist, designer, artist	Begin time: Mon May 01 05:51:29 +0000 2006	Location: Gualala, CA	Activity: scientist at shield.ai, designer, artist
294	Subject 2	Interests: music, design; Social connections: Ex-Google, Amazon	Begin time: Tue May 23 23:55:27 +0000 2006	Location: Walnut Creek, CA	Activity: Tweets on Sundays; Link figūra
388	Subject 3	Interests: Food, design, technology, Agriculturalist	Begin time: Fri Jun 30 07:15:52 +0000 2006	Location: San Francisco, CA	Activity: coder at Airbnb
544	Subject 4	Interests: Payments, stats; Behavioral data: entrepreneur	Begin time: Wed Jul 12 23:34:02 +0000 2006	Location: San Francisco, CA	Activity: BD and strategy at VGS, angel investor; Link figūra
690	Subject 5	Interests: None specified	Begin time: Thu Jul 13 22:43:59 +0000 2006	Location: Pleasure Point, CA	Activity: None specified
1006	Subject 6	Interests: Politics, history, cats, baseball; Social connections: Teacher, Texan, Israeli	Begin time: Sat Jul 15 09:32:43 +0000 2006	Location: Houston, TX	Activity: Politics, history, baseball
1047	Subject 7	Interests: Knows a little bit about a lot of things	Begin time: Sat Jul 15 21:30:46 +0000 2006	Location: Not specified	Activity: Not specified
1313	Subject 8	Interests: Not specified	Begin time: Sun Jul 16 15:07:20 +0000 2006	Location: Not specified	Activity: Not specified

TABLE 5. Identified digital evidence.

Entry ID	Name	Location	Followers	Friends	Profile	F_i
1	Subject 1	Gualala, CA	1217	580	1	899.9
2	Subject 2	Walnut Creek, CA	5766	948	1	3360.9

Google, Amazon, *etc*. This could be used to establish professional relationships or identify potential areas of common interest.

3) social media engagement: users like Subject 2 and Subject 4 have high social media engagement in terms of the number of followers, friends, and statuses. This could indicate that they are either public figures, influencers, or active networkers.

TABLE 6. Individual profiles for suspected subjects based on the domains.

$ID(f_{Who})$	ID_{S_1}	ID_{S_2}	ID_{S_3}	ID_{S_4}
220	Interests: scientist, designer, artist	Begin time: Mon May 01 05:51:29 +0000 2006	Location: Gualala, CA	Activity: scientist, designer, artist
294	Interests: papa, music, design nerd;	Begin time: Tue May 23 23:55:27 +0000 2006	Location: Walnut Creek, CA	Activity: music, design
388	Interests: Food, design, technology	Begin time: Fri Jun 30 07:15:52 +0000 2006	Location: San Francisco, CA	Activity: Food, design, technology
544	Interests: Payments, stats, entrepreneur	Begin time: Wed Jul 12 23:34:02 +0000 2006	Location: San Francisco, CA	Activity: BD and strategy, entrepreneur, angel investor
690	Interests: None specified	Begin time: Thu Jul 13 22:43:59 +0000 2006	Location: Pleasure Point, CA	Activity: None specified
1006	Interests: Politics, history, cats, baseball	Begin time: Sat Jul 15 09:32:43 +0000 2006	Location: Houston, TX	Activity: Politics, history, baseball
1047	Interests: General knowledge	Begin time: Sat Jul 15 21:30:46 +0000 2006	Location: Not specified	Activity: General knowledge
1313	Interests: Not specified	Begin time: Sun Jul 16 15:07:20 +0000 2006	Location: Not specified	Activity: Not specified
1318	Interests: design, ux, product	Begin time: Sun Jul 16 17:09:00 +0000 2006	Location: Köln	Activity: design, ux, product
1380	Interests: Product development	Begin time: Sun Jul 16 10:23:15 +0000 2006	Location: San Francisco via Toronto	Activity: Product development, CEO

4) creation dates: the "created_at" attribute can be used to understand how long these subjects have been on the social media platform. Subjects with older accounts might have been early adopters of the platform.
5) verified status: none of the subjects in the sample data are verified. Verified subjects usually have some public significance. This could mean these are regular private individuals or professionals without public figure status.
6) protected status: the "protected" field tells whether a subject has a private account. None of the provided subjects have a protected account, suggesting they all have public profiles.

Such structural analysis provides some insights into the relationships and attributes of the subjects; making any definite conclusions would require a more in-depth analysis and additional data from social media sources. In generally, the proposed user profiling method holds great promise for improving the investigation and prosecution of cybercrimes.

As social media platforms continue to evolve, it is essential for law enforcement agencies to adopt advanced tools and techniques to stay ahead of emerging threats and protect society from the negative impacts of cybercrime.

C. COMPARATIVE ANALYSIS

While both the proposed method and ''Answering to 5W Using Digital Forensics Data'' (DF5W) [31] offer unique approaches to handling digital forensic data, they serve slightly different purposes. The authors' proposed approach provides a more in-depth and interconnected view of digital attributes, especially from social media, making it more trustworthy. On the other hand, DF5W offers a quick and intuitive visualization of data, making it more user-friendly, especially for non-experts. Depending on the specific needs of a digital forensic investigation, one might be preferred over the other, or they could potentially be used in conjunction for a more comprehensive analysis.

Proposed method's primary goal is to systematically structure and categorize digital attributes interlinked across social media platforms and develop user profiles using domain-specific digital artefacts.

DF5W work's main aim is to visualize digital forensic data by answering the 5W questions (Who, What, When, Where and Why) to facilitate quick identification and deeper analysis of data.

Main advantages. Proposed method provides a comprehensive profile of individuals based on a wide range of digital attributes from social media. It also addresses legal and ethical considerations in data collection.

DF5W offers a quick and intuitive visualization of data, making it easily understandable even for non-experts. The narrative display can help in the rapid identification of data groups that need deeper analysis.

Limitations. The proposed method faces challenges such as scalability, legal complexities, and data noise.

While DF5W provides a convenient visualization, it might not offer the depth and interlinking of artifacts and attributes that article provides.

The authors' proposed method presents several advantages over DF5W. The method employs an ontology-based diagram, allowing for a more structured and interconnected profiling of digital attributes across various social media platforms. This interconnectedness ensures a comprehensive view of evidence, enhancing its reliability. In this paper the authors introduces a specific method to identify digital evidence, which can precisely narrow down the relationship with potential suspects of cybercrimes on social media. This targeted approach can streamline investigations and improve the accuracy of suspect identification. The user profiling in article is rooted in domain-specific digital artifacts, defined by a rich set of digital attributes extracted from social media. This depth in profiling provides a more detailed and holistic view of users, making it a more robust tool for digital forensics and cybercrime investigations. In contrast, DF5W emphasizes rapid visualization of digital forensic data by answering the 5W questions, making it more accessible to non-experts. DF5W offers a user-friendly, narrative display for quick data identification. The choice between them hinges on the specific needs of a digital forensic investigation, with potential for their combined use in certain scenarios.

V. DISCUSSION

The paper outlines a structured approach for handling digital evidence identification on social media platforms through digital ontologies. Additionally, it also describes a suspected subject profiling method which utilizes domain-specific digital artefacts defined by a set of digital attributes gained from social media. In this discussion, we will delve into the implications, limitations, and possible improvements of the mentioned approaches.

The digital ontology models for social media evidence identification can have far-reaching implications for digital forensics and legal proceedings. By defining classes such as ''Social Media Platform'', ''User Profile'', and ''Evidence'' and establishing relationships among them, investigators can structure and categorize the vast array of information found on social media platforms. This structure is invaluable for the systematic collection of evidence, whether it be text, images, or videos. The subclass distinctions like ''Tweet'' and ''Comment'' under the ''Text'' class, for example, allow for more fine-grained data retrieval. This specificity can greatly enhance the efficiency of evidence collection and make sure that no critical piece of information is overlooked.

The suspect profiling methodology enables investigators to build a comprehensive profile of suspects by extracting and analyzing domain-specific digital artefacts and attributes from social media. Through mathematical formulas and visual representations, this approach not only helps in evidence collection but also aids in understanding the context around the evidence, such as the social networks of the suspects. This context can sometimes be crucial in legal proceedings.

Some limitations of the presented work should be discussed also. The digital ontology model outlined in the given material is noted to be not comprehensive. In practice, social media platforms are continually evolving, introducing new features and data types. Therefore, the term ontology-based is used. The scalability and adaptability of the ontology model to these rapid changes may be a challenge. Moreover, maintaining an updated ontology which encompasses all potential data forms is not trivial.

Besides this, the collecting data from social media platforms involves various legal and ethical considerations. The material mentions acquiring necessary legal permissions as one of the steps, but does not go into detail about how complex and time-consuming this process can be. Moreover, different jurisdictions may have different laws regarding data collection, which could further complicate the process.

Finally, the user profiling method necessitates data pre-processing to filter out irrelevant or noisy data. This step is crucial for the accuracy of the profiles generated.

Determining what data is relevant or noise can sometimes be subjective and prone to error.

VI. CONCLUSION AND FUTURE WORKS

The use of ontologies for identifying digital evidence on social media, as well as the characterization of suspects through digital artefacts and attributes, is a promising strategy for improving the efficiency of cybercrime investigations. The difficulties with scalability, legal compliance, and data relevance require ongoing refinement and adaptation of these methods. In this case, future work will be focused on improving the proposed solution by:

- integrating it with real-time data and machine learning - could allow for more proactive monitoring and evidence collection, which can be particularly beneficial in time-sensitive investigations; the integration of machine learning algorithms can improve the user profiling method by automating the process of identifying relevant attributes and artefacts, and by creating more robust and dynamic profiles that evolve as more data is collected;
- developing an ethical and legal framework for data collection in consultation with stakeholders and legal experts could streamline the process of acquiring permissions for data collection and ensure that the methodologies adhere to the laws and respect user privacy.

These future implementations would increase the value of these digital forensics approaches.

REFERENCES

[1] T. K. H. Chan, C. M. K. Cheung, and R. Y. M. Wong, "Cyberbullying on social networking sites: The crime opportunity and affordance perspectives," *J. Manage. Inf. Syst.*, vol. 36, no. 2, pp. 574–609, Apr. 2019, doi: 10.1080/07421222.2019.1599500.

[2] K. Faust and G. E. Tita, "Social networks and crime: Pitfalls and promises for advancing the field," *Annu. Rev. Criminol.*, vol. 2, pp. 99–102, Jan. 2019, doi: 10.1146/annurev-criminol-011518-024701.

[3] R. Rawat, V. Mahor, S. Chirgaiya, and A. S. Rathore, "Applications of social network analysis to managing the investigation of suspicious activities in social media platforms," in *Advances in Cybersecurity Management*, K. Daimi and C. Peoples, Eds. Cham, Switzerland: Springer, 2021, pp. 315–335.

[4] A. S. Putra, N. Aisyah, and V. H. Valentino, "Analysis of NIST methods on Facebook messenger for forensic evidence," *J. Innov. Res. Knowl.*, vol. 1, no. 8, pp. 695–702, Jan. 2022.

[5] A. Powell and C. Haynes, "Social media data in digital forensics investigations," in *Digital Forensic Education* (Studies in Big Data), vol. 61, X. Zhang and K. K. Choo, Eds. Cham, Switzerland: Springer, 2019, pp. 281–303.

[6] M. Heidari, J. H. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on Twitter," in *Proc. Int. Conf. Data Mining Workshops (ICDMW)*, Sorrento, Italy, Nov. 2020, pp. 480–487, doi: 10.1109/ICDMW51313.2020.00071.

[7] S. Grigaliunas, J. Toldinas, A. Venckauskas, N. Morkevicius, and R. Damaševičius, "Digital evidence object model for situation awareness and decision making in digital forensics investigation," *IEEE Intell. Syst.*, vol. 36, no. 5, pp. 39–48, Sep. 2021, doi: 10.1109/MIS.2020.3020008.

[8] M.-H. Maras and A. Alexandrou, "Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos," *Int. J. Evidence Proof*, vol. 23, no. 3, pp. 255–262, Jul. 2019.

[9] S. Li, T. Qin, and G. Min, "Blockchain-based digital forensics investigation framework in the Internet of Things and social systems," *IEEE Trans. Computat. Social Syst.*, vol. 6, no. 6, pp. 1433–1441, Dec. 2019.

[10] N. M. Karie, V. R. Kebande, and H. S. Venter, "Diverging deep learning cognitive computing techniques into cyber forensics," *Forensic Sci. Int., Synergy*, vol. 1, pp. 61–67, Jan. 2019.

[11] I. Rahwan et al., "Machine behaviour," *Nature*, vol. 568, no. 7753, pp. 477–486, 2019.

[12] A. Gotmare, N. S. Keskar, C. Xiong, and R. Socher, "A closer look at deep learning heuristics: Learning rate restarts, warmup and distillation," 2018, *arXiv:1810.13243*.

[13] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru, "Model cards for model reporting," in *Proc. Conf. Fairness, Accountability, Transparency*, Jan. 2019, pp. 220–229.

[14] H. Arshad, A. Jantan, and E. Omolara, "Evidence collection and forensics on social networks: Research challenges and directions," *Digit. Invest.*, vol. 28, pp. 126–138, Mar. 2019.

[15] P. V. Srimukh and S. Shridevi, "Ontology-based crime investigation process," in *Advances in Smart Grid Technology*, vol. 687, P. Siano and K. Jamuna, Eds. Singapore: Springer, 2020, pp. 497–509.

[16] C. Pasquini, I. Amerini, and G. Boato, "Media forensics on social media platforms: A survey," *EURASIP J. Inf. Secur.*, vol. 2021, no. 1, pp. 1–19, May 2021.

[17] F. Alonso-Fernandez, N. M. S. Belvisi, K. Hernandez-Diaz, N. Muhammad, and J. Bigun, "Writer identification using microblogging texts for social media forensics," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 3, no. 3, pp. 405–426, Jul. 2021, doi: 10.1109/TBIOM.2021.3078073.

[18] H. Arshad, A. Jantan, G. K. Hoon, and I. O. Abiodun, "Formal knowledge model for online social network forensics," *Comput. Secur.*, vol. 89, Feb. 2020, Art. no. 101675.

[19] O. Mayer and M. C. Stamm, "Exposing fake images with forensic similarity graphs," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 5, pp. 1049–1064, Aug. 2020.

[20] O. Elezaj, S. Y. Yayilgan, and E. Kalemi, "Criminal network community detection in social media forensics," in *Proc. Int. Conf. Intell. Technol. Appl.* Cham, Switzerland: Springer, 2021, pp. 371–383.

[21] X. Zhang, Z. H. Sun, S. Karaman, and S.-F. Chang, "Discovering image manipulation history by pairwise relation and forensics tools," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 5, pp. 1012–1023, Aug. 2020.

[22] N. A. Khan, S. Zhang, W. Zhou, A. Almogren, I. U. Din, and M. Asif, "Inferring ties in social IoT using location-based networks and identification of hidden suspicious ties," *Sci. Program.*, vol. 2020, pp. 1–16, Nov. 2020.

[23] N. Pocher, M. Zichichi, F. Merizzi, M. Z. Shafiq, and S. Ferretti, "Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics," *Electron. Markets*, vol. 33, no. 1, p. 37, Dec. 2023.

[24] T. Nayerifard, H. Amintoosi, A. G. Bafghi, and A. Dehghantanha, "Machine learning in digital forensics: A systematic literature review," 2023, *arXiv:2306.04965*.

[25] E. Nowroozi, A. Dehghantanha, R. M. Parizi, and K.-K.-R. Choo, "A survey of machine learning techniques in adversarial image forensics," *Comput. Secur.*, vol. 100, Jan. 2021, Art. no. 102092.

[26] J. Xiao, S. Li, and Q. Xu, "Video-based evidence analysis and extraction in digital forensic investigation," *IEEE Access*, vol. 7, pp. 55432–55442, 2019, doi: 10.1109/ACCESS.2019.2913648.

[27] A. R. Javed, Z. Jalil, W. Zehra, T. R. Gadekallu, D. Y. Suh, and M. J. Piran, "A comprehensive survey on digital video forensics: Taxonomy, challenges, and future directions," *Eng. Appl. Artif. Intell.*, vol. 106, Nov. 2021, Art. no. 104456.

[28] J. Hu, X. Liao, W. Wang, and Z. Qin, "Detecting compressed deepfake videos in social networks using frame-temporality two-stream convolutional network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1089–1102, Mar. 2022.

[29] S. Shao, C. Tunc, A. Al-Shawi, and S. Hariri, "Automated Twitter author clustering with unsupervised learning for social media forensics," in *Proc. IEEE/ACS 16th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Abu Dhabi, United Arab Emirates, Nov. 2019, pp. 1–8, doi: 10.1109/AICCSA47632.2019.9035286.

[30] S. Brotsis and N. Kolokotronis, "Blockchain-enabled digital forensics for the IoT: Challenges, features, and current frameworks," in *Proc. IEEE Int. Conf. Cyber Secur. Resilience (CSR)*, Jul. 2022, pp. 131–137.
[31] A. H. Lone and R. N. Mir, "Forensic-chain: Blockchain based digital forensics chain of custody with PoC in hyperledger composer," *Digit. Invest.*, vol. 28, pp. 44–55, Mar. 2019.
[32] K. Kaushik, S. Dahiya, and R. Sharma, "Role of blockchain technology in digital forensics," in *Blockchain Technology*. Boca Raton, FL, USA: CRC Press, 2022, pp. 235–246.
[33] G. Song, S. Kim, H. Hwang, and K. Lee, "Blockchain-based notarization for social media," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2019, pp. 1–2.
[34] T. Hsien-De Huang, P.-W. Hong, Y.-T. Lee, Y.-L. Wang, C.-L. Lok, and H.-Y. Kao, "SOC: Hunting the underground inside story of the Ethereum social-network opinion and comment," 2018, *arXiv:1811.11136*.
[35] H. Kwon, S. Lee, and D. Jeong, "User profiling via application usage pattern on digital devices for digital forensics," *Expert Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114488.
[36] D. Kosmajac and V. Keselj, "Twitter user profiling: Bot and gender identification: Notebook for PAN at CLEF 2019," in *Proc. Int. Conf. Cross-Lang. Eval. Forum Eur. Lang.* Cham, Switzerland: Springer, Sep. 2020, pp. 141–153.
[37] R. Kaushal, V. Ghose, and P. Kumaraguru, "Methods for user profiling across social networks," in *Proc. IEEE Int. Conf. Parallel Distrib. Process. Appl., Big Data Cloud Comput., Sustain. Comput. Commun., Social Comput. Netw.*, Dec. 2019, pp. 1572–1579.
[38] C. Ferrante and B. Habibnia, "Answering to 5W using digital forensics data," in *Proc. Int. Symp. Comput. Sci. Intell. Controls (ISCSIC)*, Nov. 2021, pp. 185–193.

ŠARŪNAS GRIGALIŪNAS (Member, IEEE) received the Ph.D. degree from the Kaunas University of Technology. He was a Sr. Cyber Security Consultant at Business for six years. He is currently a Senior Consultant with Transcendent Group Baltics. He is also an Associate Professor with the Kaunas University of Technology. His primary competencies lie within the IT sector, where he started as an IT Security Auditor. He is an ISECOM Certified Penetration Tester and a Certified ISO 27001 ISMS Auditor. He has more than 15 years of experience in IT auditing and security, interested in cybersecurity. He is an Analyst of social network security, cyber-attacks, psychology and practical implementations of digital forensics evidence, and new gamification methods of learning security awareness, building a culture of security, or managing insider threats. As the Chairperson of TK 79 information security with the Lithuanian Standardization Department, led the development and harmonization of information security standards. Keep abreast of the latest financial industry regulations and standards (such as PCI-DSS, ISO 27001, and ICT requirements from BoL) and ensure that the company complies with these legal and regulatory requirements.

RASA BRŪZGIENĖ is/was a Lecturer in integrated cyber security training for employees of state and municipal institutions on control and management of systems in critical, high-incertitude situations as well as on risk management in the electronic resources managed by the institution. She is currently an Associate Professor with the Department of Computer Sciences, Kaunas University of Technology. She is also a Researcher with the Scientific Group of Cyber Security. She is the leader of five international Nordplus study projects, related to excellence and best practice in the fields of cybersecurity and intelligent communication. She is the author of 31 scientific publications, four scientific monographs and parts thereof, and two educational books. Her research interests include cyber security; communication networks, their resilience and security; security of critical infrastructure and cyber-physical systems; reliability and efficiency of communication systems; and cyber-sustainability.

ALGIMANTAS VENČKAUSKAS is currently the Head of the Computer Science Department and the Research Group of Cyber Security, Kaunas University of Technology. He was the Leader and a Principal Investigator of a number of projects, such as the 2019–2022 Project "Strategic programs for advanced research and technology in Europe" (SPARTA), H2020; the 2020 Project "Model for the organization of the remote work and training process and recommendations for the extreme and transitional period (LMT, No. S-COV-20-20); and the 2020–2021 projects on security awareness and training "A comprehensive cyber security training programs for employees of state and municipal institutions and organizations" (CVPA procurement No. 458424. 2020, 2021). He is an Organizer and the Manager of the Information and IT Security Master's Program, which has been running for 12 years in a blended learning way using distance learning technologies. He is the author or coauthor of more than 65 articles in various international journals, including 34 publications in scientific periodicals, cited in the Clarivate Analytics Web of Science database with impact factor, his H-index is 10, and he has written 11 textbooks. His research interests include information technology, the Internet of Things, cyber security, applied cryptography, the application of artificial intelligence methods to cyber security, and the application of distance learning technologies. He is a member of the Cyber Security Council of the Ministry of National Defense of the Republic of Lithuania. He represents the university in international and national institutions that create and implement science and innovation policy: Ministry of Economy and Innovation, Industry 4.0 Platform, and MOSTA smart specialization.