Exploring the Intersection between Databases and Digital Forensics

Danilo B. Seufitelli[1], Michele A. Brandão[1,2], Mirella M. Moro[1]

[1] Universidade Federal de Minas Gerais, Brazil
{daniloboechat, michele.brandao, mirella}@dcc.ufmg.br
[2] Instituto Federal de Minas Gerais
michele.brandao@ifmg.edu.br

Abstract. Digital forensics has attracted attention from assorted researchers, who primarily work on predicting and solving digital hacks and crimes. In turn, the number and types of digital crimes have increased considerably, mainly due to the growing use of digital media to perform daily personal and professional tasks. Like most computer-related activities, data is at the center of such hacks and crimes. Hence, this work presents a systematic literature review of publications at the intersection between Digital Forensics and Databases. We discuss problems and trends of two main categories: Data Building and Database Management Systems. Overall, this research opens the doors for the communication between databases and an area with several exciting and concrete challenges, with great potential for social, economic, and technical-scientific contributions.

Categories and Subject Descriptors: H.2 [**Database Management**]: Miscellaneous

Keywords: Databases, Digital Forensic, Survey

1. INTRODUCTION

Digital transformation creates opportunities that change how information is consumed and produced. Different activities from leisure to work happen in the virtual environment, which contributes to increasing the flow and availability of data in digital media. Still, not everything is perfect, and such scenario has favored the propagation of cybercrimes that incur real risks to people. As a simple exercise, searching for Cyber Crimes on the *BBC News*[1] returns eight pages of results over the period between June 2019 and February 2022. Examples of news include "The spy app that helped the FBI disrupt an international criminal network", "Tiktok: Users use loophole to post videos of pornography and violence", and "3 new types of fraud and scams arising from the covid pandemic".

In this context, the term *forensic* means something related to judicial competence and plays a central role in crime solution [Bell 2013]. Specifically for cybercrimes, Computer Forensics and Digital Forensics come to play, given the need to find out the origin of such crimes and to protect the privacy and integrity of electronic information, their owners and users. Overall, investigating cyberattacks requires going after the data and back, from simple text files to huge DBMS, as criminals may even take data as hostage.[2,3] Indeed, as digital transformation has made most activities to become computer-related (e.g., through mobile devices and sensors), *data* is at the core of cybercrimes and acts as digital

[1]Cyber Crimes on BBC News: https://www.bbc.com/portuguese/topics/c95y3549y56t. Accessed on 02/13/2022
[2]US GAO: https://www.gao.gov/blog/ransomware-holding-it-systems-and-data-hostage. Accessed on 02/13/2022
[3]CBS News: https://www.cbsnews.com/news/ransomware-cyberattacks-60-minutes-2021-06-06/. Accessed on 02/13/2022

The authors would like to thank Ana Flávia, Ayane Cristina, Kayque Meira and Weverton Mata for all relevant insights and previous cooperation to this project. This work was supported by CAPES, CNPq, and FAPEMIG, Brazil.

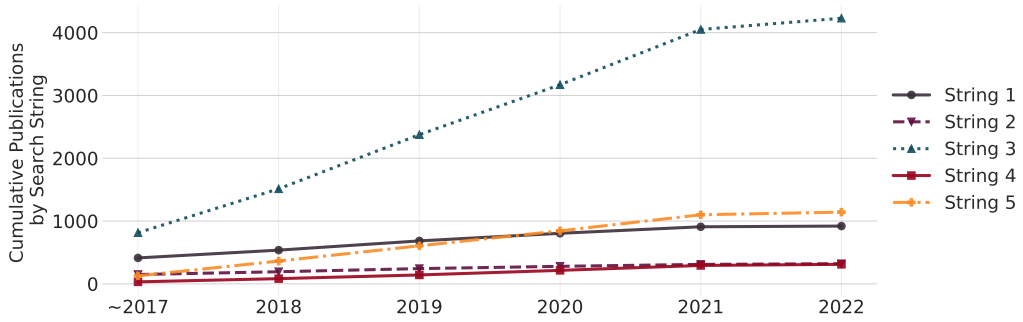

Fig. 1. Evolution of the number of publications in the interval 2006-2022 by each SLR search string, represented by String 1, String 2, String 3, String 4 and String 5. Note that we group publications until 2017 for better visualization.

evidence that feeds forensic investigations.

This work presents original research on databases (DB) and its intersections with Digital Forensics to explore such importance. Digital Forensics is an essential science in the cybercrime scenario, as it helps in the crime reconstruction during the investigation and in the development of approaches to prevent crime occurrences. Specifically, this science works in the search, analysis, identification, and categorization of data that can be crime evidence [Garfinkel 2010; van Beek et al. 2020]. For example, Digital Forensics may be used for preventing and detecting *SQL Injection* attacks [Xie et al. 2019].

Analyzing the research and technologies available in the area of Digital Forensics is challenging.[4] For example, its number of publications has increased over the years, as shown in Figure 1 (publications returned using search strings defined in Section 2). In particular, there are way more publications returned by String 3 than by the other strings; i.e., the keywords of String 3 are more frequent in the publications. Still, it does not mean the publications returned by String 3 are more relevant than the others. Dealing with such a large volume requires applying a Systematic Literature Review (SLR) methodology to identify the most relevant ones in the context of databases (Section 2). Next, we discuss the selected publications highlighting their objectives and group such publications in the main areas of interest: Data Building (Section 3) and DBMS (Section 4). As this article expands on a short paper published in the 36^{th} Brazilian Symposium on Databases [Seufitelli et al. 2021], we update our analysis to comprise the period between 2021 and 2022 in both previous Sections.[5] Finally, we summarize the SLR results by discussing new opportunities for research (Section 5), describing the related work (Section 6), and presenting final remarks to stimulate further studies at the intersection of Databases and Digital Forensics (Section 7).

2. METHODOLOGY

In this work, the methodology used is based on the execution of seven steps adapted from the protocol of Kitchenham and Charters (2007). In addition to the following explanations, Figure 2 summarizes the entire process, including the number of works remaining from Step 4 onwards.

Step 1: Define the research questions. We start this review by defining six exploratory research questions to obtain an overview of the state of the art in the area of Digital Forensics within the

[4]Although the terms computer forensics, digital forensics and cyber forensics are used interchangeably, they are different: computer forensics focuses on the investigation of crimes where computers are present, and the cyber and digital forensics refer to data from different digital devices.

[5]Besides such an update, as a new contribution, we considerably extend the discussion on the publications analyzed in the short version and describe more works than those chosen as representatives of the classes of the proposed taxonomy.

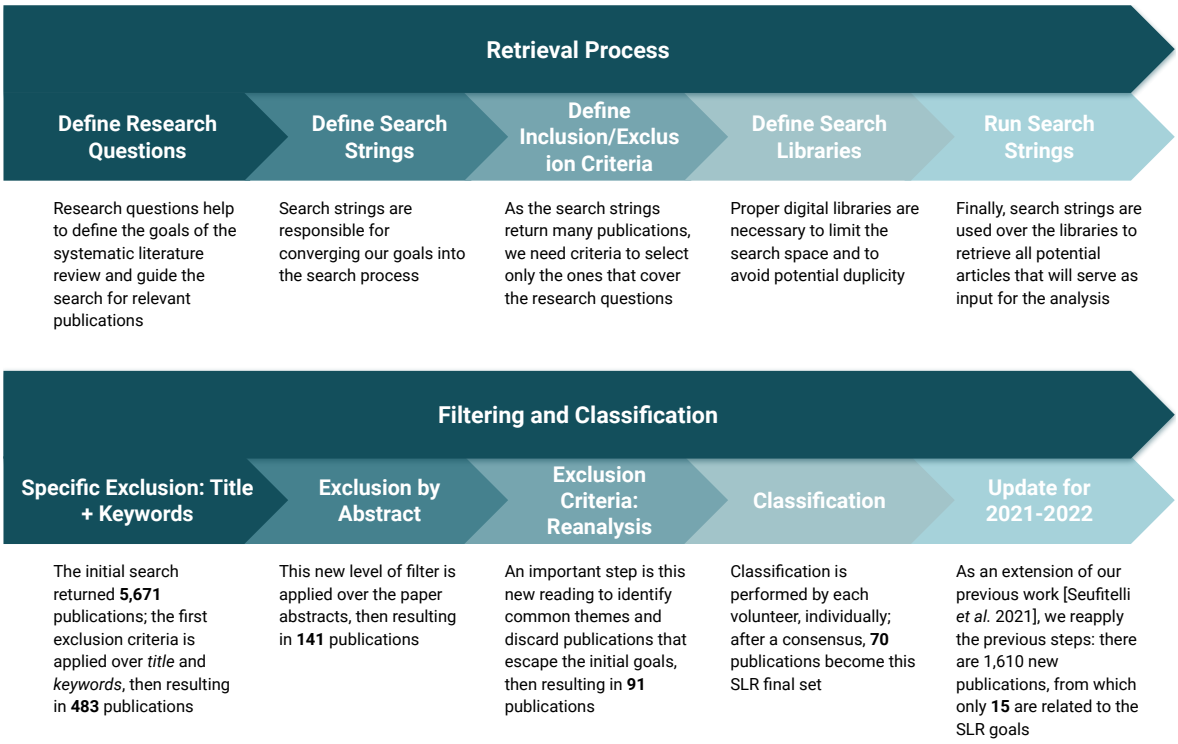

Fig. 2. We split the systematic literature review methodology into two parts: the process for retrieving articles from digital libraries; and the process of filtering, selecting, and classifying publications for the study.

context of Databases. Such questions help in the description and classification of studies selected for SLR, as defined in Table I.

Table I. Research questions and search strings.

Research Questions
What is the intersection between databases and digital forensics?
What types of research are most common in digital forensics: qualitative, quantitative or mixed?
What datasets are considered in digital forensics studies?
What sub-areas can be identified at the intersection of digital forensics and databases?
What are the challenges and opportunities in working at the intersection between DB and digital forensics?
Search Strings
String 1: “database forensic” OR “database forensics” OR “forensic database” OR “forensic databases”
String 2: “criminal database” OR “criminal databases” OR “database auditing”
String 3: (database OR databases) AND (“forensic access” OR “forensic analysis” OR “forensic purpose” OR “forensic purposes”)
String 4: (forensic OR forensics) AND (“database analysis” OR “database access”)
String 5: (forensic OR forensics) AND (SQL OR NoSQL)

Step 2: Define the search strings. From these questions, the search strings were defined. The bottom of Table I informs the final set of strings.

Step 3: Define the inclusion criteria and the general exclusion criteria for the data. These criteria help to define the set of publications: for inclusion, the content is related to the area of databases and discusses digital forensics; and for exclusion, the publication does not have an abstract,

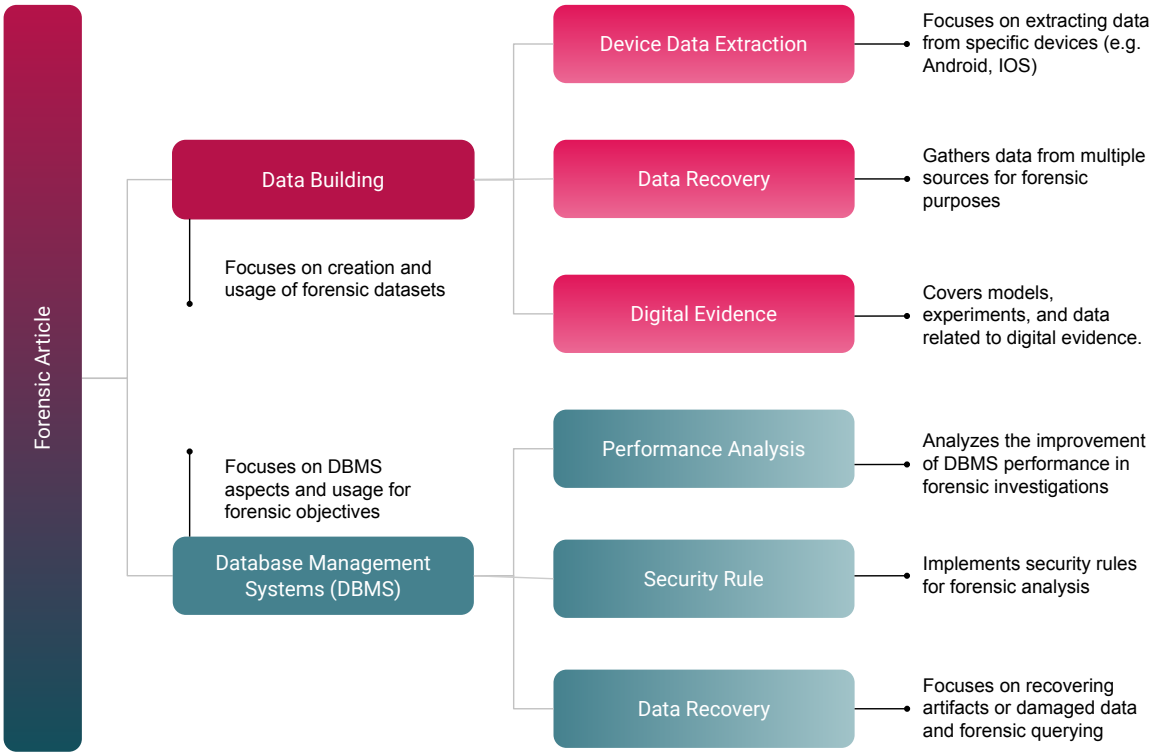

Fig. 3. Overview of the classification process of considered publications.

is only an abstract, is an old version of another study considered, is not a primary study or does not provide access to the full study, or is duplicated, and is not written in the English language.

Step 4: Look for publications. The first search for publications, conducted in May 2021, considered the following digital libraries: IEEE Xplore, Scopus, Science Direct and Web of Science. From each digital library, a different set of publications were collected, with a total of 5,671 articles: 278 articles from IEEE; 3,029 from Scopus; 1,665 from Science Direct; and 175 from Web of Science.

Step 5: Define specific exclusion criteria. Based on the titles, specific exclusion criteria were created: publications outside *Computing* and *Engineering*, and those dealing with other areas – e.g., biology works such as genetic forensics and biomedicine. After applying such criteria within both title and keywords, 483 publications remained.

Step 6: Select publications and identify common themes. By reading the abstract, publications outside the inclusion criteria were discarded, leaving 141 works. The volunteers then performed dynamic reading of all works and identified common themes that were used in the next step.

Step 7: Classify publications. The themes identified in the previous step were further studied for elaborating a classification (or taxonomy), which resulted in two classes: *data building* – a process that involves collecting, grouping, diversifying and segmenting data to obtain information and build knowledge; and *DBMS* – a system that helps to store, modify and extract data in a secure, concurrent, shareable and recoverable way, with or without replication through a set of interfaces and languages. Both classes are then divided into subclasses, as shown by Figure 3.

Then, three volunteers[6] manually classified the 141 publications, independent from each other

[6]Professor/Doctor, Doctoral Student and Undergraduate in Computer Science with experience in databases.

to avoid bias. In parallel, an exclusion reanalysis was applied and out-of-scope publications were excluded. With the new filtering, only 91 publications were considered. Then, the coefficient *Fleiss' Kappa* [Fleiss et al. 2013] was applied to verify the agreement of the classifications. Thus, the coefficient reached 0.30, with 95% credibility. The volunteers then discussed the content of the works and reached a consensus that resulted in 70 publications.[7] From those 70, we use only 28, which are the publications that fit the two classifications used in this review.

Step 8: Update set of publications. As an extension of our previous work [Seufitelli et al. 2021], we reapply the previous seven steps for publications from 2021 to 2022. Thus, we update the analysis in the SLR. This step resulted in 1,610 new publications, from which only 15 are related to the SLR objectives. Then, we also classify such publications into data building or DBMS. In summary, we classify the first 28 papers selected (until 2020) and the 15 papers selected in the update (2021-2022), resulting in 43 classified papers in this article.

3. DATA BUILDING PUBLICATIONS ANALYSIS AND DISCUSSIONS

Data Building is a process that involves collecting, grouping, diversifying and segmenting data to obtain information and build knowledge. Works related to the Data Building section inspect a DB and can be divided into three groups: data extraction from devices, data recovery, and digital evidence. As a result of the methodology steps, **15** publications were classified as Data Building. Then, **seven** built their DB from data taken from applications and external devices, **two** perform data recovery, and **six** deal with digital evidence. Regarding the update of SLR, Ten publications were classified as Data Building. From those, **two** publications cover device data extraction, **five** present data recovery, and **three** discuss digital evidence. Table II summarizes such publications sorted by year.

Device data extraction. The publications in this group address data extraction from a specific device to analyze vulnerabilities and/or security issues in such a device. Starting with works that focus on mobile devices, Andriotis et al. (2012) propose a method for forensic experts to investigate crimes originated from a suspect Android smartphone. Such method allows to obtain the data and then to investigate any activity associated with wireless communications. Furthermore, Khobragade and Malik (2014) propose a methodology for collecting and analyzing data from cybernetic systems and browser logs, and Li et al. (2014) present a method of recovering operational logs from files. Satrya et al. (2016) take text and audio messages of the Telegram application from three Android mobile devices by using tools that allow the extraction and schematization of data, such as *Android Debug Bridge* and *SQLite Browser*. In the end, the authors note that the data remaining in the Telegram storage can be extracted and used as digital evidence of cybercrimes.

Changing to the context of computer networks, Awasthi et al. (2018) provide guidance for forensic data acquisition and analysis of artifacts extracted from user interactions in smart home hub environment. Also, Servida and Casey (2019) show how data traces from IoT devices (*Internet of Things*) can be useful for forensic purposes. Devices studied include *QBee Multi-Sensor Camera*, *Cube One & Accessories*, *Arlo Pro* and the *Nest Protect*. The authors extract data from such devices using multiple plugins developed by them (e.g., *Autopsy Framework*). Finally, Pessolano et al. (2019) review methods that allow forensic experts to obtain the following information from Nintendo 3DS: system activity, deleted images, Internet history items, relevant friends list information, the console's serial number and plaintext WiFi access point passwords.

One central task of forensic investigation is extracting data from physical devices, such as smartphones, hard disks, smart accessories, etc. A criminal investigation may use data stored on such devices and their linked applications as part of the evidence set. Concerning this possibility, Williams et al. (2021) analyze an intelligent fitness device (Fitbit Versa) using Cellebrite UFED and MSAB

[7]Worksheet with selected publications: `https://bit.ly/papersForense`

Table II. Data Building publications retrieved.

	Reference Keyword	Data Source
EVD	[Ming and LiZhong 2009] Network intrusion system	– A model whose goal is to detect and collect network intrusion data
EXTR	[Andriotis et al. 2012] Vulnerabilities in networks	– Data crawled from Android devices to analyze vulnerabilities in wireless connections
EXTR	[Khobragade and Malik 2014] Data mining/generation	– Data crawled from browser logs and data streams across networks and computers
EXTR	[Li et al. 2014] Android Recov. Methods	– Samsung Galaxy S3 SMS database considering data partition on device in three periods
RECV	[Liu et al. 2016] Recovery of deleted record	– Recovery of deleted records from SQLite3 through a new approach
EXTR	[Satrya et al. 2016] Android Telegram	– Database extracted from Telegram app text and audio artifacts from three Android mobile devices
EVD	[Grajeda et al. 2018] Digital forensic artifacts	– Data of digital forensic artifacts built from a proposed tool that stores artifacts, user account information and application logs activities
EXTR	[Awasthi et al. 2018] Smart home hub	– Data extracted from home hub (*Securifi Almond*), which allowed forensic analysis of artifacts related to user interaction in the hub
EVD	[Freiling and Hösch 2018] Evidence tampering	– Data on a disk image was manipulated by graduate students to study tampered digital evidence
RECV	[Atwal et al. 2019] Spotlight, Apple desktop	– Base with *Spotlight* metadata (search tool, Apple) to search for persistence of deleted data in your metadata
EVD	[van Zandwijk and Boztas 2019] iPhone Health App	– Investigates the use of data extracted from the iPhone Health application, which stores data about physical exercises performed by the user
EXTR	[Servida and Casey 2019] IoT, digital traces	– Data from mobile apps (e.g., Nest and Wink hubs) and from IoT devices (e.g., cloud data from QBee Camera and the Swisscom Home App); Plugins were developed for data extraction
EVD	[Hosler et al. 2019] Database for Video Forensics	– Data of digital forensic videos composed by 2,000 videos from 46 physical devices representing 36 unique camera models
EXTR	[Pessolano et al. 2019] Forensic analysis nintendo 3DS	– Data from nintendo 3DS to analyze methods of hacking and extracting data from that device's internal storage system
RECV	[Bahjat and Jones 2019] Deleted file fragment	– Framework for determining a time-window for file fragments considering the first moment that it was written to a media until it was deleted
EXTR	[Williams et al. 2021] Android/iOS recover	– Recover data from wearable smart fitness that run Android 9 and iOS 12
EVD	[Kumar and Karabiyik 2021] Instagram vanish mode	– Investigate the presence of vanished messages in the application database
EVD	[Afshar et al. 2021] Behavior detection	– Propose an Attribute/Behavior-Based Access Control for understanding and deriving users' behaviors from log files
RECV	[Deng et al. 2022] User profile dataset, ML	– Construct a dataset of users' profiles who have suffered from telecom fraud by crawling the Weibo website
RECV	[Yogameena et al. 2022] Face matching/recognition	– Construct a short face-video linked dataset, consisting of 100 three-minute duration videos
EXTR	[Son et al. 2022] Mobile devices, messengers	– Extract data from unrooted and rooted devices of three messenger apps: Signal, Wickr, and Threema
RECV	[Fernández-Fuentes et al. 2022] Web browser investigation	– Gather data from HD and memory of a browsing session from Google Chrome and Mozilla Firefox on four different Linux environments to obtain the hard disk's changes and a complete RAM dump
EVD	[Tolosana et al. 2022] DeepFakes detection	– Propose a novel approach based on the selection of specific facial regions as input to the fake detection system
RECV	[Davies et al. 2022] Ransomware, mixed dataset	– Propose a new cybersecurity dataset for ransomware detection and forensic analysis research
RECV	[Salim et al. 2022] IoT, privacy preservation	– Construct the SM-IoT dataset, which is generated by simulating SM users' data and the ground-truth of the IoT of the two major SM platforms, Facebook and the social IoT

EXTR: Device data extraction. RECV: Data recovery. EVD: Digital evidence.

XRY. The data extraction includes logical and physical operations using Android 9 and iOS 12, com-

paring Cellebrite and XRY capabilities. Likewise, instant messengers apps have become increasingly essential to obtain digital evidence as criminals use such apps. Nevertheless, due to instant messengers applying end-to-end encryption, obtaining digital evidence requires extracting data only from the mobile device. In such a context, Son et al. (2022) present a methodology for decrypting and extracting data from three messenger apps: Signal, Wickr and Threema.

Data recovery. The publications here mainly address the recovery of deleted files that can compose an evidence. Liu et al. (2020) present a method of data recovery and analysis on deleted DB files from SQLite3 (DB software used by cell phones). In this study, the location and size of data in the researched field are estimated from the analysis of SQLite3 formats. In a complementary way, Atwal et al. (2019) focus on analyzing the structure of the metadata store database, and the results show that records are no longer recoverable when deleted. Finally, Bahjat and Jones (2019) propose a dating framework for file fragments. The authors claim that dating file fragments is important for event reconstruction when deleted files compose an evidence.

Crimes in telecommunication networks are still frequent, although it is possible to detect the user's profile most likely to be defrauded. In this context, Deng et al. (2022) propose a warning fraud scheme based on users' profiles of telecom and Internet technologies. They build a dataset of users' profiles who have suffered from telecom fraud by crawling the Weibo website. Another source for recovering forensic data is the web browser. Most browsers frequently update their tools and features for privacy. In this context, Fernández-Fuentes et al. (2022) propose a method to explore the effectiveness of the private mode. The method recovers data from private browsing sessions on Mozilla Firefox and Google Chrome running on four different Linux environments. Still on privacy but within a different perspective, Salim et al. (2022) propose a new framework for privacy-preserving the interaction between social media and Internet of Things (IoT) services. As the main part of such a framework, they propose a new relational dataset called SM-IoT.

Regarding the digital forensic for suspect recognition, the task of identifying a suspect face from videos during a criminal investigation is challenging, mainly when the suspects' image is unavailable. Then, the common alternative is to use a face sketch drawn based on eyewitness's memory recollection to search over photo databases. In this context, Yogameena et al. (2022) propose the SpyGAN, a sketch-face matching that includes profile faces with different illumination and poses. They created a short-face-video linked dataset consisting of 100 three-minute videos. Lastly, Davies et al. (2022) propose NapierOne, a modern cybersecurity mixed file dataset inspired by the Govdocs1 dataset, as its a complement. The main objective of the dataset is to serve as input for ransomware detection and forensic analysis research. Also, the dataset focuses on addressing the deficiency in reproducibility and improving consistency by facilitating research replication.

Digital evidence. The publications here cover models, experiments and data that are directly related to digital evidence. In this sense, Ming and LiZhong (2009) develop a network intrusion model as a forensic tool. The model performs intrusion detection and, in the meantime, collects all network data from the target system to facilitate network forensic analysis. Grajeda et al. (2018) explore building, implementing and maintaining the Artifact Genome Project (AGP), digital forensics repository. Also, the paper presents the impact of AGP in professional and academic realms of digital forensics. On the other hand, Freiling and Hösch (2018) describe a set of experiments performed to tamper with a disk image and differentiate tampered and unadulterated evidence. van Zandwijk and Boztas (2019) present experiments on three iOS devices on the application *Health*, which stores data on the number of steps and time a person moves. Finally, Hosler et al. (2019) aim to propose a new DB composed of forensic videos and develop a camera model identification results on these videos using the deep-learning techniques.

Recent works extract and analyze digital evidence from many forms. For example, Kumar and Karabiyik (2021) explore the vanish mode, a new feature by an Instagram update that allows people to have secure and private chats by sending messages that soon disappear. However, such a feature

might interest criminals (child predators, drug dealers, and cyberbullies) who find ways to clear their activity trace. Therefore, the authors focus on identifying any artifacts useful in a forensic investigation of such an app. Afshar et al. (2021) provide another way to derive digital evidence by proposing an Attribute/Behavior-Based Access Control (ABBAC) for understanding and deriving users' behaviors from log files. The authors construct a feature dataset based on user logs from the UCI Machine Learning Repository database. Finally, Tolosana et al. (2022) provide an in-depth analysis of both first and second generations of DeepFakes databases regarding fake detection performance as digital evidence. They consider two distinct methods: the state-of-art using the entire face as input to the fake face detection and a novel approach based on selecting specific facial regions. They conclude that using specific facial areas achieves results above 99% Area Under the Curve (AUC). They also highlight the need for more efforts to analyze inter-database scenarios to improve the fake detectors.

The analysis of these publications reveals the focus on getting data from mobile devices and applications. Also, Android is the most used system in these studies. For database researchers, there are opportunities on specializing extraction and recovery approaches to mobile devices, as well as proposing mining techniques to aid on digital evidence identification and collection.

4. DBMS PUBLICATIONS ANALYSIS AND DISCUSSIONS

A DBMS helps to store, modify and extract data in a secure, concurrent, shareable and recoverable way, with or without replication through a set of interfaces and languages. In our taxonomy, this class considers publications in the area of digital forensics that focus on three aspects: performance analysis, security rules and data recovery. Only **13** of set of publications were classified as DBMS. From those, **five** publications present performance analysis, **six** cover topics related to security rules, and **two** deal with data recovery within the DBMS. Regarding the update of SLR for such a classification, five were labeled as DBMS. Among then, **three** present results related to security rules, and **two** cover data recovery. We did not find new publications on performance analysis. Table III summarizes such publications sorted by year.

Performance analysis. This group of publications aims at improving performance analyses of DBMS, comparing different storage solutions, or proposing frameworks for better operating forensic databases. More broadly, Qi (2014) reviews the four main types of NoSQL DBs (key-value, document, column family, and graph based), but focuses on MongoDB and Riak performance analysis. Similarly, Qi et al. (2014) analyze two techniques to improve scalability in data management, comparing MongoDB, Riak and MySQL. Switching to relational DB, Khanji et al. (2015) focus on MySQL and Oracle, and propose a framework with auditing features to assist in performing forensic database analysis. Note that Khanji et al. (2015) define the important concept of **Forensic DB**, which may be a subarea of digital forensics with little focus on literature. These forensic databases are logically identical to the regular ones, but they differ in terms of physical file structure, security and concurrency mechanisms, query optimization, among others. In particular, the area of forensic databases requires the development of forensic analysis tools that can be used in different DBMSs.

In the mobile storage context, Schmitt (2018) evaluates a collection of tools that analyze SQLite files, mainly, in relation to forensic extraction and recovery routines. They apply different manipulations to every single file and thereby introduce anti-forensic aspects that are tested against distinct performance analysis tools. Finally, Liebler et al. (2019) present a review article that analyzes publications on the topic, either as a citation, proposal, or review of the evaluation performance methods.

Security rules. This class typically comprises two aspects: a formula that determines the conditions for granting access controls, or access controls that specify the access permissions. Nevertheless, there are works that propose diverse approaches to improve their security rules. Specifically, motivated by audit log requirements, Azemović and Mušić (2009) describe a model that ensures an effective and efficient tamper detection based on SQL triggers to collect audit logs. They implement hashing

Table III. Database Management Systems publications retrieved.

	Reference Keyword	Data Source
SEC	[Azemović and Mušić 2009] Data tempering	– Proposes a model to detect authorized and unauthorized modification of database schema and data itself
SEC	[Han et al. 2009] Database server detection	– Introduces new client-based methods to analyze information about the database server connection and detect specific databases that target forensic investigation
RECV	[Lindauer et al. 2012] Forensic traces	– Introduces a description model for different forensic trace types and applies such model to a well established database schema
SEC	[Pavlou and Snodgrass 2012a] Information accountability	– Develops a prototype audit system for information accountability in high-performance databases
SEC	[Hommes et al. 2013] Source Code, Debug	– Proposes a framework for network data security, and the records are stored in a NoSQL database
PERF	[Qi 2014] NoSQL Databases	– Analyzes four NoSQL DBMSs, focusing on MongoDB and Riak performance analysis
PERF	[Qi et al. 2014] Big Data	– Compares MongoDB, Riak and MySQL for data management
PERF	[Khanji et al. 2015] Database Auditing	– Uses Oracle and MySQL DBMS to test the performance of a proposed search framework
PERF	[Schmitt 2018] SQLite corpora	– Evaluates a collection of tools that analyze SQLite files, mainly, in relation to forensic extraction and recovery routines
SEC	[Hauger and Olivier 2018] NoSQL databases	– Examines whether NoSQL databases have security features that leave relevant traces so an accurate forensic attribution can be conducted
RECV	[Wagner et al. 2019] Database forensics artifacts	– Develops a framework for representing and searching database forensic artifacts and introduces a new storage format, called Database Forensic File Format
PERF	[Liebler et al. 2019] Artifact lookup strategies	– Analyzes articles previously written on the topic, either as a citation, proposal or review of the methods
RECV	[Choi et al. 2021] Forensic Recovery	– Introduces a method of recovering deleted data for use in MySQL
SEC	[Alfadli et al. 2021] Database Forensics	– Propose a unified identification model applicable to the database forensic field, which integrates and harmonizes all exiting identification processes into a single abstract model
SEC	[Bašić et al. 2021] SQL trigger, auditing	– Propose an effective in-database auditing subsystem used as a base for access control, intrusion detection, fraud detection, and others
SEC	[Andrade et al. 2021] Internal changes	– Present a new system that allows to track the changes of internal databases used by Universal Windows Platform
RECV	[Zhang et al. 2022] MySQL binlog	– Introduce a method for data recovery using the binary log of the MYSQL database
RECV	[Gupta et al. 2022] Database as a Service	– Propose a communication-efficient and information-theoretically secure system for aggregation queries on Database as a Service

PERF: Performance analysis. SEC: Security rules. RECV: Data recovery.

algorithms to improve data integrity and prevent audit falsification from providing authentication codes on collected data. Also to avoid data tampering and to mediate access to valuable databases (even for insider access), Pavlou and Snodgrass (2012b) propose DRAGOON. Such a prototype aims to improve the security rules of DBMSs by implementing tampering detection and a forensic analysis system designed to determine when tampering(s) occurred and what data were tampered with. Similar to log analysis, Han et al. (2009) propose client-based methods for detecting specific databases in target companies by analyzing connection information of the database server, which DBMSs store on the client-side. They introduce a tool that can automatically extract and analyze this information to collaborate with forensic investigations and improve security rules.

Moving to NoSQL solutions, also using log files, Hommes et al. (2013) propose a framework that can modify controller programs transparently by using graph transformation, enabling online fault management through logging of network parameters in a NoSQL database. Such a database acts

as a storage system for flow entries and respective parameters, which can be leveraged to detect network anomalies or to perform forensic analysis. Searching for the lack of default security measures such as access control and logging, Hauger and Olivier (2018) specifically investigate whether NoSQL databases have security features that leave relevant traces to conduct accurate forensic attribution. They explore the most common NoSQL databases to establish what authentication and authorization features are available. They also evaluate logging mechanisms since access control without auditing would not aid forensic objectives.

Database Forensics is a digital forensics field of study that inspects DBMS content to verify database crimes. Each DBMS has a different infrastructure with a distinct identification investigation model for security purposes. Alfadli et al. (2021) discuss such identification process from four perspectives: identification, collection, analysis, and presentation. They propose a unified identification six phases model that integrates all exiting identification processes into a single abstract model, called Common Identification Process Model (CIPM). The phases comprise: notifying an incident; responding to the incident; identifying the incident source; verifying the incident; isolating the database server; and providing an investigation environment. As a result, the CIMP helps practitioners and newcomers to the forensics domain to control database crimes.

Then, Bašić et al. (2021) present other primary aspects concerning security rules: a simple, secure, configurable, role-separated and effective in-database auditing subsystem, which can serve as a base for access control, intrusion detection, fraud detection, and other security-related analyses and procedures. Such auditing subsystem is capable of keeping the entire audit trail (data history) of a database and all the executed SQL statements. In turn, all this information enables different security applications, from ad hoc intrusion prevention to complex a posteriori security analyses.

From an application perspective, an update generated by the life cycle may silently disappear (or change) the internal data structure of a DBMS, hindering a digital forensic artifact. On the other hand, new releases may also create new opportunities for digital forensics. Nonetheless, searching for digital forensics artifacts often requires analyzing and exploring internal data structures. As a promising solution, Andrade et al. (2021) presents UWPscanner. This open-source system allows tracking the changes of internal databases used by Universal Windows Platform (UWP) applications to guarantee data consistency for digital forensic tool developers. The authors perform a case study by tracking Microsoft Skype (Skype App) and Your Phone evolution with UWPscanner. As result, the study illustrates the high rate of changes of UWP applications and thus the appropriateness to rely on an automated system to detect modifications and new data structures.

Data recovery. This subclass covers publications that promote the search for digital or digitized artifacts stored in a DBMS, introduce optimized storage format for digital evidence, or reconstruct damaged or deleted data. During a forensic investigation, the number of digitally captured traces may increase fast. Every trace has its properties (e.g., minutiae for fingerprints or raking traces for locks), but they share metadata, such as location, time, and other crime scenes information-related. Lindauer et al. (2012) introduce a standard description model for different forensic trace types: fingerprint, lock traces, micro traces, and ballistic traces. They apply a well-established database schema development process, the phases of transferring expert knowledge in the corresponding forensic fields into an extendable, database-driven, generalized forensic description model. The main objective is to propose a single database capable of handling and retrieving all possible forensic evidence acquired (in a digital way) at a crime scene, facilitating the analysis of forensic experts.

In the same direction but focusing on the digital traces, Wagner et al. (2019) present a new storage format for database forensic artifacts named Database Forensic File Format (DB3F). They also include a toolkit to view and search data stored in DB3F. After all, the uniqueness of each DBMS storage solution and the different format of forensic artifacts require new patterns for storage and view tools to carry out a better and more efficient forensic analysis. Finally, Choi et al. (2021) investigate the operation storage engine of Microsoft SQL Server (MSSQL). The goal is to understand the internal

structure of MSSQL data files. As a result, they propose a method to recover deleted tables and records, and make the solution available as an open-source tool.

Zhang et al. (2022) analyze the structure of the DBMS MySQL binary log and use the binlog to restore the database. They compare the mysqlbinlog tool and its defects, use Python to write a better binlog analysis tool, generate standard SQL, and have a rollback function to better recover data. In a different perspective, database-as-a-service (DaS) allows authenticated users to execute queries on an untrusted public cloud. Although there is progress on cryptography, secure and efficient query processing over outsourced data is still a challenge. In this context, Gupta et al. (2022) propose the Obscure, a communication-efficient and information-theoretically secure system for aggregation queries on DaS with conjunctive or disjunctive predicates, using secret-sharing. As a result, the Obscure prevents the network and adversarial servers from learning the user's queries, the results, the database, and hiding access and query patterns.

The analysis of publications in the DBMS class unveils the main interest in dealing with DBMS log files to explore improvements in security rules and recover data for forensic purposes. Also, there are challenges and solutions proposed for both SQL and NoSQL systems. There are opportunities for researchers and database enthusiasts to propose new security techniques using log files, offer new solutions for implementing log in DBMS as a powerful forensic tool, and further use machine learning algorithms to detect user behavior (authorized or not) in DBMSs.

5. DISCUSSION AND OPPORTUNITIES

This section summarizes the RSL results with responses to each survey question. We emphasize that such discussions are essential to show possible directions to research in the intersection between the database and digital forensic areas. Also, such research may minimize the gap between both areas.

What is the intersection between databases and digital forensics? The main intersection is the need that Digital Forensics has regarding its data, which often needs to be stored during investigation and for future studies. This need is generally adequate in view of the research objective. In summary, the publications present the use of data to perform forensic investigation itself [Ming and LiZhong 2009; Al-Dhaqm et al. 2020a; 2020b], to build a new dataset [Awasthi et al. 2018; Atwal et al. 2019; Davies et al. 2022; van Zandwijk and Boztas 2019] and to test the performance of a proposed approach (methodology, framework or language) [Andrade et al. 2021; Choi et al. 2021; Khanji et al. 2015]. Another novel theme in the intersection between DB and digital forensics is the study of deep fake databases [Tolosana et al. 2022].

It is important to emphasize that there is not a common sense regarding the data model used, since works such as Hommes et al. (2013); Qi (2014); Qi et al. (2014) use NoSQL databases, and Khanji et al. (2015); Choi et al. (2021); Zhang et al. (2022) the relational model. Also, the SLR reveals that MongoDB and MySQL are the DBMSs most considered in the aforementioned publications.

What are the most common types of research carried out in digital forensics: qualitative, quantitative or mixed? Most of the research in the selected works is quantitative, that is, they use different statistical strategies to validate a given hypothesis. Also, these works use a forensic data system to analyze their approach. Few exceptions include [Freiling and Hösch 2018; Henseler and van Loenhout 2018; Liebler et al. 2019; Servida and Casey 2019], which use a qualitative approach (i.e., they present reviews or analyses of approaches or forensic issues in a non-numeric way).

Which datasets are considered in digital forensics studies? The selected works propose several new datasets and also use existing datasets. As their goals are quite diverse, there are works that use data from packet flows in computer networks [Sikos 2020], stolen car records [Chen 2008], web browser [Fernández-Fuentes et al. 2022; Khobragade and Malik 2014; Salunkhe et al. 2016], logs files [Afshar et al. 2021], and users' profiles of telecom who have suffered a fraud [Deng et al. 2022].

Which sub-areas can be identified at the intersection between digital forensics and databases? We identified two main classes of research, each one with three categories: *data building* – data extraction from devices, data retrieval, and digital evidence; *Database Management System (DBMS)* – performance analysis, security rules, and data recovery. Note that although the two classes (data building and DBMS) have the category of *data recovery*, the focus on each one of them is different: in data building, retrieval focuses on data building and studies associated with it (e.g., analysis of data persistence when deleted and development of plugins for data extraction); and in DBMS, data retrieval associates with DBMSs (e.g., data retrieval in MySQL).

At the intersection between DB and digital forensics comes the task of data mining. For example, it is in the methodology of Khobragade and Malik (2014) to aid in forensic investigation, and it is applied by Chen (2008) in the analysis of stolen cars. Other works use machine learning algorithms. For example, Khobragade and Malik (2014) uses clustering algorithms, and Chen (2008) uses classification algorithms, clustering, association rules and prediction. More recently, to compare the performance of models that detect user behavior in a system, Afshar et al. (2021) use three different machine learning techniques, which are linear regression, Random Forest, and k-Nearest Neighbors (kNN). Similarly, to assess the accuracy/quality of the proposed dataset to train a classifier, Salim et al. (2022) employ four standard classification models, including Gradient Boosting, Random Forest, Naive Bayes, and Feed Forward learning models.

What are the challenges and opportunities in working at the intersection between DB and Digital Forensics? In general, one of the main challenges is the change in technologies, as more and more data formats, systems, programming languages, physical devices, among others. On the one hand, such a change can favor security, as they are technologies that evolve from old ones, that is, there is already knowledge about possible flaws, especially for cyber crimes. On the other hand, new technologies can also mean new forms of cyberattacks. In relation to databases, a challenge is to maintain compatibility between systems. In addition, new technologies also demand the training of more specialists and the creation of new approaches to forensic investigation. For example, many works focus on extracting data from devices with specific systems, such as android, IOS, cameras, video games, and others. A research opportunity is to define a way to standardize: extracting data from different devices; and the terminologies and concepts most used in the intersection between the areas. Another opportunity is to define a knowledge base that enables the storage and sharing of knowledge involving these two areas.

An interesting point of view according to Khanji et al. (2015) is that seeking a balance between the performance of a DBMS and the audit resources for forensic investigation is a challenge. Specifically, it is important to make the database scalable for such analyses by changing the audit settings in the DBMS during the development phase. Finally, Tolosana et al. (2022) warn that a significant public concern is: fake images and videos, including facial information generated by digital manipulations and DeepFakes methods. The trendy term "DeepFakes" refers to a deep learning-based technique able to create fake videos by swapping the face of a person by another person's face. There are open software and mobile applications that automatically generate fake videos by anyone without prior knowledge of the task. Such a possibility opens doors to discover such a yet unexplored world: identifying, developing, and analyzing deep fake databases as input of forensic analyses.

6. RELATED WORK

During the SLR presented here, the search strings returned seven surveys [Ahmad et al. 2022; Al-Dhaqm et al. 2020b; Kanta et al. 2020; Liu et al. 2022; Mahapatra and Khan 2012; Prajapati and Shah 2020; Sikos 2020]. However, their main focus is different, as they address: cyber security in IoT-based cloud computing [Ahmad et al. 2022], database forensic investigation process models [Al-Dhaqm et al. 2020b], open source intelligence for smarter password cracking [Kanta et al. 2020], intrusion detection systems in the cloud computing [Liu et al. 2022], existing techniques against SQL injection attacks

[Mahapatra and Khan 2012], secure data deduplication [Prajapati and Shah 2020], and packet for network forensics [Sikos 2020].

Other surveys related to this article, but not found by the search strings of the SLR, focus on digital forensics tools to extract data from databases or to assist in the recovery of database [Cankaya and Kupka 2016], the last ten years of research related to forensic analysis of relational and NoSQL databases [Chopade and Pachghare 2019] and the evolution of the digital forensic (its origins, current position, and future directions) [Damshenas et al. 2014]. However, to the best of our knowledge, none of this work broadly focuses on the intersection between the database and digital forensic areas.

7. CONCLUSION

The number of publications on digital forensics has increased, and the performance of forensic expertise has expanded in various contexts. One of this work's main goals is to promote a better classification synthesis and facilitate the search and access to the main researches on digital forensics and databases. To do so, we followed a methodology formed by seven steps, from defining research questions, search strings, inclusion/exclusion criteria to searching the publications in digital libraries, selecting publications and identifying common themes, and classifying the publications. Finally, we updated the SLR by running such steps to filter new publications in the period of 2021-2022.

The analyses conducted in this SRL reveal that data is essential in digital forensics. In addition to compose digital evidence, data is the base to develop and evaluate different approaches that help to solve problems in digital forensic areas. Still, as an area, Databases has much to contribute to forensic investigations, not only in the search for improving processing speed (through indexes and specialized access methods) but also in the accuracy of the results (through advanced query and correlation, for example). Furthermore, Databases (relational or not) are the natural choice to store data resulting from criminal investigations. Here, we highlight that such resulting data include, but are not limited to: digital evidence for crimes committed on digital scenarios; or digitized traces from crimes in non-digital scenarios. To the former, examples include log traces, browsing history, network traces, and user behavior capture. To the latter, we highlight items that usually are digitized (e.g., photographs) in a crime scene, such as fingerprint traces, ballistic analysis, audio/video capturing, and any types of objects found at crime scenes. By reading the publications, it is also clear that more advanced data organization and processing techniques are still necessary and work further study within the context of criminal investigation. The next step of this research is to expand the coverage of publications considered for areas related to Databases, such as Data Mining and Machine Learning, among others.

Acknowledgment. This work was partially funded by CNPq, CAPES, FAPEMIG and resources of IFMG Edital 087/2019.

REFERENCES

Afshar, M., Samet, S., and Usefi, H. Incorporating behavior in attribute based access control model using machine learning. In *2021 IEEE International Systems Conference (SysCon)*. IEEE, Vancouver, BC, Canada, pp. 1–8, 2021.

Ahmad, W., Rasool, A., Javed, A., Baker, T., and Jalil, Z. Cyber security in iot-based cloud computing: A comprehensive survey. *Electronics (Switzerland)* 11 (1): 11–16, 2022. cited By 0.

Al-Dhaqm, A. et al. Categorization and organization of database forensic investigation processes. *IEEE Access* vol. 8, pp. 112846–112858, 2020a.

Al-Dhaqm, A. et al. Database forensic investigation process models: A review. *IEEE Access* vol. 8, pp. 48477–48490, 2020b.

Alfadli, I. M., Ghabban, F. M., Ameerbakhsh, O., AbuAli, A. N., Al-Dhaqm, A., and Al-Khasawneh, M. A. Cipm: Common identification process model for database forensics field. In *2021 2nd International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*. IEEE, Cameron Highlands, Malaysia, pp. 72–77, 2021.

Andrade, L. M., Domingues, P., and Frade, M. Keeping track of uwp application changes for digital forensic purposes. In *2021 Telecoms Conference (ConfTELE)*. IEEE, Leiria, Portugal, pp. 1–5, 2021.

Andriotis, P., Oikonomou, G., and Tryfonas, T. Forensic analysis of wireless networking evidence of android smartphones. In *2012 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, Costa Adeje, Spain, pp. 109–114, 2012.

Atwal, T. S. et al. Shining a light on spotlight: Leveraging apple's desktop search utility to recover deleted file metadata on macos. *Digital Investigation* vol. 28, pp. S105–S115, 2019.

Awasthi, A. et al. Welcome pwn: Almond smart home hub forensics. *Digital Investigation* vol. 26, pp. S38–S46, 2018.

Azemović, J. and Mušić, D. Efficient model for detection data and data scheme tempering with purpose of valid forensic analysis. In *Proc. Int. Conf. Comput. Eng. Appl.(ICCEA)*. IACSIT Press, Singapore, pp. 83–89, 2009.

Bahjat, A. A. and Jones, J. Deleted file fragment dating by analysis of allocated neighbors. *Digital Investigation* vol. 28, pp. S60–S67, 2019.

Bašić, B., Udovičić, P., and Orel, O. In-database auditing subsystem for security enhancement. In *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*. IEEE, Opatija, Croatia, pp. 1642–1647, 2021.

Bell, S. *A Dictionary of Forensic Science*. Oxford University Press, Online, 2013.

Cankaya, E. C. and Kupka, B. A survey of digital forensics tools for database extraction. In *2016 future technologies conference (ftc)*. IEEE, IEEE, San Francisco, CA, USA, pp. 1014–1019, 2016.

Chen, P. S. Discovering investigation clues through mining criminal databases. In *Intelligence and Security Informatics: Techniques and Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 173–198, 2008.

Choi, H., Lee, S., and Jeong, D. Forensic recovery of SQL server database: Practical approach. *IEEE Access* vol. 9, pp. 14564–14575, 2021.

Chopade, R. and Pachghare, V. K. Ten years of critical review on database forensics research. *Digital Investigation* vol. 29, pp. 180–197, 2019.

Damshenas, M., Dehghantanha, A., and Mahmoud, R. A survey on digital forensics trends. *International Journal of Cyber-Security and Digital Forensics* 3 (4): 209–235, 2014.

Davies, S. R., Macfarlane, R., and Buchanan, W. J. Napierone: A modern mixed file dataset alternative to govdocs1. *Forensic Science International: Digital Investigation* vol. 40, pp. 301330, 2022.

Deng, W., Liang, G., Zhang, X., and Shi, Y. An early warning model of cybercrime based on user profile. In *Proceedings of the 11th International Conference on Computer Engineering and Networks*, Q. Liu, X. Liu, B. Chen, Y. Zhang, and J. Peng (Eds.). Springer Singapore, Singapore, pp. 751–757, 2022.

Fernández-Fuentes, X., F. Pena, T., and Cabaleiro, J. C. Digital forensic analysis methodology for private browsing: Firefox and chrome on linux as a case study. *Computers & Security* vol. 115, pp. 102626, 2022.

Fleiss, J. L., Levin, B., and Paik, M. C. *Statistical methods for rates and proportions*. John Wiley & Sons, New Jersey, 2013.

Freiling, F. and Hösch, L. Controlled experiments in digital evidence tampering. *Digital Investigation* vol. 24, pp. S83–S92, 2018.

Garfinkel, S. L. Digital forensics research: The next 10 years. *digital investigation* vol. 7, pp. S64–S73, 2010.

Grajeda, C., Sanchez, L., Baggili, I., Clark, D., and Breitinger, F. Experience constructing the artifact genome project (agp): Managing the domain's knowledge one artifact at a time. *Digital Investigation* vol. 26, pp. S47–S58, 2018.

Gupta, P., Li, Y., Mehrotra, S., Panwar, N., Sharma, S., and Almanee, S. <sc>obscure</sc>: Information-theoretically secure, oblivious, and verifiable aggregation queries on secret-shared outsourced data. *IEEE Transactions on Knowledge and Data Engineering* 34 (2): 843–864, 2022.

Han, J., Lee, K., Choi, J., Lim, K., and Lee, S. Analysis of connection information for database server detection. In *2009 2nd International Conference on Computer Science and its Applications*. IEEE, Jeju, Korea (South), pp. 1–5, 2009.

Hauger, W. K. and Olivier, M. S. NOSQL databases: Forensic attribution implications. *SAIEE Africa Research Journal* 109 (2): 119–132, 2018.

Henseler, H. and van Loenhout, S. Educating judges, prosecutors and lawyers in the use of digital forensic experts. *Digital Investigation* vol. 24, pp. S76–S82, 2018.

Hommes, S. et al. Automated source code extension for debugging of openflow based networks. In *CNSM*. IEEE, Zurich, Switzerland, pp. 105–108, 2013.

Hosler, B. C., Zhao, X., Mayer, O., Chen, C., Shackleford, J. A., and Stamm, M. C. The video authentication and camera identification database: A new database for video forensics. *IEEE Access* vol. 7, pp. 76937–76948, 2019.

Kanta, A., Coisel, I., and Scanlon, M. A survey exploring open source intelligence for smarter password cracking. *FSI: Digital Investigation* vol. 35, pp. 301075, 2020.

Khanji, S. I. R., Khattak, A. M., and Hacid, H. Database auditing and forensics: Exploration and evaluation. In *AICCSA*. IEEE, Marrakech, Morocco, pp. 1–6, 2015.

Khobragade, P. K. and Malik, L. G. Data generation and analysis for digital forensic application using data mining. In *2014 Fourth International Conference on Communication Systems and Network Technologies*. IEEE, Bhopal, India, pp. 458–462, 2014.

Kitchenham, B. and Charters, S. Guidelines for performing systematic literature reviews in software engineering. Tech. Rep. EBSE-2007-01, Keele University and Durham University Joint Report, 2007.

Kumar, S. T. and Karabiyik, U. Instagram forensic analysis revisited: Does anything really vanish? In *2021 International Symposium on Networks, Computers and Communications (ISNCC)*. IEEE, Dubai, United Arab Emirates, pp. 1–6, 2021.

Li, Q., Hu, X., and Wu, H. Database management strategy and recovery methods of android. In *ICSESS*. IEEE, Beijing, China, pp. 727–730, 2014.

Liebler, L., Schmitt, P., Baier, H., and Breitinger, F. On efficiency of artifact lookup strategies in digital forensics. *Digital Investigation* vol. 28, pp. S116–S125, 2019.

Lindauer, I., Schäler, M., Vielhauer, C., Saake, G., and Hildebrandt, M. A first proposal for a general description model of forensic traces. In *Optics, Photonics, and Digital Technologies for Multimedia Applications II*. Vol. 8436. International Society for Optics and Photonics, SPIE, Brussels, Belgium, pp. 84360U, 2012.

Liu, T.-M., Kao, D.-Y., and Chen, Y.-Y. Loocipher ransomware detection using lightweight packet characteristics. *Procedia Computer Science* vol. 176, pp. 1677–1683, 2020. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 24th International Conference KES2020.

Liu, X., Fu, X., and Sun, G. Recovery of deleted record for SQLite3 database. In *IHMSC*. IEEE, Hangzhou, China, pp. 183–187, 2016.

Liu, Z., Xu, B., Cheng, B., Hu, X., and Darbandi, M. Intrusion detection systems in the cloud computing: A comprehensive and deep literature review. *Concurrency and Computation: Practice and Experience* 34 (4): 1–23, 2022.

Mahapatra, R. and Khan, S. A survey of sq1 injection countermeasures. *International Journal of computer Science and engineering survey* 3 (3): 55, 2012.

Ming, H. and LiZhong, S. A new system design of network invasion forensics. In *ICCEE*. IEEE, Dubai, UAE, pp. 596–599, 2009.

Pavlou, K. E. and Snodgrass, R. T. Achieving database information accountability in the cloud. In *2012 IEEE 28th International Conference on Data Engineering Workshops*. IEEE, Arlington, VA, USA, pp. 147–150, 2012a.

Pavlou, K. E. and Snodgrass, R. T. Dragoon: An information accountability system for high-performance databases. In *2012 IEEE 28th International Conference on Data Engineering*. IEEE, Arlington, VA, USA, pp. 1329–1332, 2012b.

Pessolano, G., Read, H. O., Sutherland, I., and Xynos, K. Forensic analysis of the nintendo 3ds nand. *Digital Investigation* vol. 29, pp. S61–S70, 2019.

Prajapati, P. and Shah, P. A review on secure data deduplication: Cloud storage security issue. *Journal of King Saud University - Computer and Information Sciences* vol. 1, pp. 1–12, 2020.

Qi, M. Digital forensics and NoSQL databases. In *FSKD*. IEEE, Xiamen, China, pp. 734–739, 2014.

Qi, M. et al. Big data management in digital forensics. In *CSE*. IEEE, Chengdu, China, pp. 238–243, 2014.

Salim, S., Turnbull, B., and Moustafa, N. Data analytics of social media 3.0: Privacy protection perspectives for integrating social media and internet of things (sm-iot) systems. *Ad Hoc Networks* vol. 128, pp. 102786, 2022.

Salunkhe, P., Bharne, S., and Padiya, P. Data analysis of file forensic investigation. In *SCOPES*. IEEE, CParalakhemundi, India, pp. 372–375, 2016.

Satrya, G. B., Daely, P. T., and Nugroho, M. A. Digital forensic analysis of telegram messenger on android devices. In *ICTS*. IEEE, Surabaya, Indonesia, pp. 1–7, 2016.

Schmitt, S. Introducing anti-forensics to SQLite corpora and tool testing. In *2018 11th International Conference on IT Security Incident Management IT Forensics (IMF)*. IEEE, Hamburg, Germany, pp. 89–106, 2018.

Servida, F. and Casey, E. Iot forensic challenges and opportunities for digital traces. *Digital Investigation* vol. 28, pp. S22–S29, 2019.

Seufitelli, D., Moura, A. F., Fernandes, A., Siqueira, K., Brandão, M., and Moro, M. Forense digital e bancos de dados: um survey. In *Anais do XXXVI Simpósio Brasileiro de Bancos de Dados*. SBC, Porto Alegre, RS, Brasil, pp. 307–312, 2021.

Sikos, L. F. Packet analysis for network forensics: A comprehensive survey. *FSI: Digital Investigation* vol. 32, pp. 200892, 2020.

Son, J., Kim, Y. W., Oh, D. B., and Kim, K. Forensic analysis of instant messengers: Decrypt signal, wickr, and threema. *Forensic Science International: Digital Investigation* vol. 40, pp. 301347, 2022.

Tolosana, R., Romero-Tapiador, S., Vera-Rodriguez, R., Gonzalez-Sosa, E., and Fierrez, J. Deepfakes detection across generations: Analysis of facial regions, fusion, and performance evaluation. *Engineering Applications of Artificial Intelligence* vol. 110, pp. 104673, 2022.

van Beek, H. et al. Digital forensics as a service: Stepping up the game. *FSI: Digital Investigation* vol. 35, pp. 301021, 2020.

van Zandwijk, J. P. and Boztas, A. The iphone health app from a forensic perspective: can steps and distances registered during walking and running be used as digital evidence? *Digital Investigation* vol. 28, pp. S126–S133, 2019.

Wagner, J., Rasin, A., Heart, K., Jacob, R., and Grier, J. Db3f & df-toolkit: The database forensic file format and the database forensic toolkit. *Digital Investigation* vol. 29, pp. S42–S50, 2019.

Williams, J., MacDermott, A., Stamp, K., and Iqbal, F. Forensic analysis of fitbit versa: Android vs ios. In *2021 IEEE Security and Privacy Workshops (SPW)*. IEEE, San Francisco, CA, USA, pp. 318–326, 2021.

Xie, X. et al. SQL injection detection for web applications based on elastic-pooling cnn. *IEEE Access* vol. 7, pp. 151475–151481, 2019.

Yogameena, B., Jakkamsetti, G., and S., A. Spygan sketch: Heterogeneous face matching in video for crime investigation. *Journal of Visual Communication and Image Representation* vol. 82, pp. 103400, 2022.

Zhang, Z., Yuan, M., and Qian, H. Research on mysql database recovery and forensics based on binlog. In *Proceedings of the 11th International Conference on Computer Engineering and Networks*, Q. Liu, X. Liu, B. Chen, Y. Zhang, and J. Peng (Eds.). Springer Singapore, Singapore, pp. 741–750, 2022.