Where do Databases and Digital Forensics meet? A Comprehensive Survey and Taxonomy

Danilo B. Seufitelli[1], Michele A. Brandão[1,2], Ayane C. A. Fernandes[2],
Kayque M. Siqueira[2], Mirella M. Moro[1]
[1]Universidade Federal de Minas Gerais
[2]Instituto Federal de Minas Gerais
daniloboechat@dcc.ufmg.br,michele.brandao@ifmg.edu.br
{ayanecristinamb,kayque.siq}@gmail.com,mirella@dcc.ufmg.br

ABSTRACT

We present a systematic literature review and propose a taxonomy for research at the intersection of Digital Forensics and Databases. The merge between these two areas has become more prolific due to the growing volume of data and mobile apps on the Web, and the consequent rise in cyber attacks. Our review has identified 91 relevant papers. The taxonomy categorizes such papers into: Cyber-Attacks (subclasses SQLi, Attack Detection, Data Recovery) and Criminal Intelligence (subclasses Forensic Investigation, Research Products, Crime Resolution). Overall, we contribute to better understanding the intersection between digital forensics and databases, and open opportunities for future research and development with potential for significant social, economic, and technical-scientific contributions.

1. INTRODUCTION

Digital Forensics (DF) helps reconstruct cybercrimes and develop prevention mechanisms. Indeed, it searches, analyzes, identifies, and categorizes data that may become crime evidence [28, 30, 59]. For instance, some authors have used DF in preventing and detecting *SQL Injection* attacks [61], analyzing digital evidence manipulation [49], and others. Also, its number of publications has risen, reaching thousands of works indexed by Google Scholar when using search *strings* (see §3), which also challenges a person to start studying the field. Hence, we apply a Systematic Literature Review (SLR) protocol to extract Computing-related insights from DF, focusing on Database (DB) aware publications. Overall, the SLR covers works mainly from Digital Forensics that also use, explore, or advance DB topics.

The field of DF is widely recognized for its applications beyond the realm of Computing, including legal medicine plus civil, criminal, and corporation investigations. An extensive search and filtering process is essential to identify relevant publications specific to the computational context. However, an SLR can only solve half the problem by identifying such publications. The other half is organizing such articles comprehensively, promoting effective categorization and synthesis, and supporting future research on DF and DB.

For the second challenge (comprehensive organization), we can use a taxonomy to identify and classify approaches and concepts based on their purposes. For instance, Dave et al. [22] introduce a taxonomy for implicit requirements identification, a crucial part of requirements engineering. Through taxonomies, researchers can efficiently categorize their findings and develop a better understanding of the research landscape, ultimately advancing the field and supporting future research endeavors.

Then, our goal starts from the high availability of data on DF and tackles its *intersection* with DB. An initial study connects them on data infrastructure and availability [53]. The goal now is to cover works on how to identify and prevent *cyber attacks* as well as exploit data for *criminal intelligence*. This article aims to achieve two primary objectives: (*i*) identify how exploring data can aid in DF and identify the challenges faced by data engineers and scientists in doing so, and (*ii*) develop a better categorization of works in this area. As we note "data is the new oil" is not enough,[1] promoting a better understanding between databases and digital forensics is crucial as data plays a vital role in digital forensics. Hence, the contributions of this article include a survey and a corresponding taxonomy of publications at their intersection, as well as discussions on the various phases of forensics tasks related to data and the associated opportunities. Overall, researchers may better grasp how exploring data can aid to DF.

The rest of this article is structured as follows. We discuss the related work in §2. Then, we outline the SLR methodology in §3 and review concepts on the taxonomy classes in §4. Next, we focus the SLR on data from/to criminal intelligence and cyber attacks in §5. Finally, we map the taxonomy and its papers to digital forensic phases in §6, and discuss challenges for the survey questions in §7.

[1]Data as The New Oil: `https://bit.ly/f-dataoil`

Table 1: Search strings over digital libraries.

Search Strings
1."database forensic" OR "database forensics" OR "forensic database" OR "forensic databases"
2."criminal database" OR "criminal databases" OR "database auditing"
3.(database OR databases) AND ("forensic access" OR "forensic analysis" OR "forensic purpose" OR "forensic purposes")
4.(forensic OR forensics) AND ("database analysis" OR "database access")
5.(forensic OR forensics) AND ("database analysis" OR "database access")

2. RELATED WORK

There is no related work on the intersection of DB and DF besides ours [53], which goes over DBMS (Database Management System) and data building, solely. Others cover different aspects of digital forensics. First, Khan et al. [34] review its evolution between 1980 and 2020. Then, Sikos [55] focuses on the evolution of forensic packet analysis. Finally, Al-Dhaqm and collaborators cover Database Forensic Investigation (DBFI) from two aspects: (*i*) categorization of processes (planning, preparation and pre-response; acquisition and preservation; and analysis and reconstruction) in [3]; and (*ii*) common limitations (regarding investigation processes; concepts and terminologies; and lack of unified models) and their solutions in [4].

Our contributions are novel for focusing on cyber attacks and criminal intelligence linking databases and forensics (whereas [53] goes over DBMS and data building). They shall support youngsters and experts in forensics in quickly finding works as classified in a taxonomy (four classes and three subcategories) according to the forensics phase.

3. METHODOLOGY

In this SLR, we adapt the methodology of seven steps from Kitchenham and Charters' protocol [35] (which is the same applied in [53]).

Step 1: Define research questions. We first define questions to guide over the state of knowledge in Forensics and DB. Our questions and goals are:
– *When and where the studies were published?* Define interests and trends over time.
– *What kinds of research are there?* Classify into qualitative, quantitative or mixed.
– *What is the focus of data-driven digital forensics?* Define the sub-areas, themes and trends.
– *Which are the advances and potential challenges in the area?* Identify methods, models and tools.
– *What issues are still open?* Find new challenges.

Step 2: Define search strings. Such questions require looking for works that could answer them. First, we consider the most extensive Computing digital library – DBLP (Digital Bibliography and Library Project), and search for "data forens" within title, abstract and keywords. Its results serve as input for filtering the most relevant keywords and define the final search strings, see Table 1.

Step 3: Define inclusion criteria and general exclusion criteria. Keeping focus on our questions, we define the following criteria: *Inclusion criteria* verify if the paper is related to both DB and DF; and *General exclusion criteria* check if it has no abstract, is only an abstract, is an old version of another study already considered, is not a primary study, and is not possible to access its full content.

Step 4: Search for publications. Looking for works in one digital library may compromise coverage. Hence, we search for the pre-defined strings over: IEEE Xplore, Scopus, Science Direct, and Web of Science. All returned papers were collected – except from Scopus, as we consider only publications on Computing and Engineering. The numbers of papers found in each library are: IEEE 289; Scopus 3,836; Science Direct 2,042; and Web of Science 756; making a total of 6,923 records.

Step 5: Define specific exclusion criteria. We create exclusion criteria based on the publication titles. We discard those outside the domains of Computing and Engineering and those about other fields – e.g., biology, genetic forensics, biomedicine. By applying such criteria to titles and keywords, we identified 493 relevant publications for further analysis.

Step 6: Select publications and identify common themes. After reviewing abstracts of the 493 publications, we eliminate those outside the inclusion criteria; resulting in 151 articles, to which the exclusion criteria was reapplied. In the end, we have identified 101 articles for the next phase.

Step 7: Classify publications. We elaborate a taxonomy by using the identified themes in the previous step, with four major classes: DBMS, data building, cyber attacks, and criminal intelligence. Three volunteers manually labeled 101 publications (from step 6) by considering these four classes. We also evaluate the agreement of the classification through the Fleiss' Kappa coefficient [20], which achieved a value of 0.30 with 95% of confidence. The volunteers then discussed the content and reached a final verdict to the labels, which resulted in a better classification and the exclusion of 10 publications. Thus, we present the SLR for 91 publications; from which 52 fit the two categories used in this article: Cyber Attacks and Criminal Intelligence.

4. TAXONOMY OVERVIEW

Forensic science is the area interested in and concerned with finding the relationship among people, places, and things involved in criminal activities [46]. It is an essential science to assist in the investigation and judgment of civil and criminal cases. It also considers expertise from different areas; e.g., forensic chemistry, forensic biology, forensic physics, computer forensics (also called cyber forensics and digital forensics), and so on [46]. Here, we focus on *digital forensics*, a science whose goal is to identify,

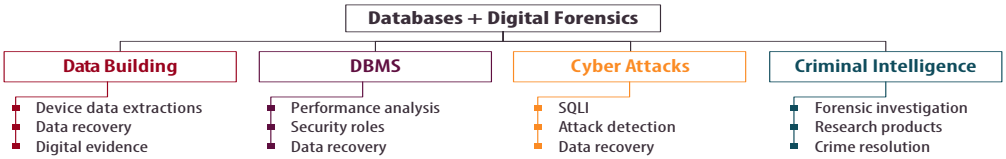

Figure 1: Proposed classifications and their guidelines. In this article, we focus on Cyber Attacks and Criminal Intelligence, as Data Building and DBMS are covered in [53].

preserve, retrieve, analyze and present digital data during a digital-related investigation [29]. Although one may use the terms computer forensics, digital forensics, and cyber forensics interchangeably, they are different: computer forensics is mainly about the investigation of crimes related to computers, whereas cyber and digital forensics are mainly about digital data from various digital devices [34].

Digital forensics uses data as evidence, which is often, but not only, associated with cybercrimes. DF helps to: support or refute assumptions, reconstruct criminal events, and predict unauthorized actions [29]. Studying DF also aids socially, by serving the criminal justice and defense systems [16]. More, investigating the solution to each crime is different. Thus, investigating human misbehavior on computers and digital devices creates new logic and technological challenges, especially when the offender tries to hide the evidence and related activities [16].

We now describe the main attributes of the publications for each class identified in the SLR. We present how we identify the topics of each class (§4.1) and then overview main definitions (§4.2).

4.1 Classification Summary

We propose a taxonomy with four classes: Data Building, DBMS, Cyber Attacks, and Criminal Intelligence, as shown in Figure 1. Note that we have empirically defined these classes based on the extensive literature analysis and by considering the link between topics on databases and digital forensics. Further, Table 2 provides an overview of each class by presenting LDA[2] topics. LDA considers the title and abstract of all papers in each class and returns 15 words to describe the classes. Terms in bold are more representative of each class defined as follows.
– **Data Building**: includes publications that enrich the forensic frameworks, with subcategories Device Data Extraction, Data Recovery, and Digital Evidence. Here, the publications involve the use of data to solve a problem in the digital forensics area.
– **DBMS**: has publications about forensic tools and data architecture, presenting subcategories for Performance Analysis, Security Rules, and Data Recovery. It is an important tool in the DB area; thus, this class includes publications on digital forensics whose main focus is related to DBMS usage.

[2]Latent Dirichlet Allocation – a generative probabilistic model of a collection of texts [13].

Table 2: Topics found by LDA for each class.

Class	LDA Topic
Data Building	forensic **cloud computing** analysis similarity deep incident **traces** learning response **reconstruction** monitoring **storage auditing** artifacts
DBMS	**database** internet analysis things systems **nosql** management iot control machine **recovery** clustering multi application networking
Cyber Attacks	data databases investigation analysis **detection intrusion** big sql dataset log file collection **injection** mining model
Criminal Intelligence	**forensics digital** security network **evidence crime** android applications computer **web cyber intelligence** based encryption center

– **Cyber Attacks**: has publications about digital attacks, with subcategories SQL Injection, Attack Detection, and Data Recovery (further discussed in §5.1). In other words, this class covers publications that address attacks on databases or sensitive data.
– **Criminal Intelligence**: for publications that improve systems and security, with subcategories Forensic Investigation, Research Products, and Crime Resolution (further described in §5.2). The publications in this class include those that apply data to investigate and solve digital crimes.

4.2 Classes and their Categories

This section overviews each class and their categories. The first two classes (DBMS and Data Building) focus on aspects of DB architecture and availability, being well described in [53]. This article highly improves the literature review by describing works related to cyber-attacks and criminal intelligence, whose intersections with DB have not appeared in any survey (to the best of our knowledge).

4.2.1 Data Base Management System (DBMS)

A DBMS allows creating, modifying and deleting a database, plus inserting, deleting and updating its data. Different DBMSs are studied in forensics. For example, Beirami et al [11] prepare DBMSs to optimize the results of forensic queries. We classify forensic-related publications as DBMS when they focus on: *Performance Analysis*, papers aim to improve the performance analysis of DBMS, compare different storage solutions, or propose frameworks for better operating forensic databases; *Security Rules*, papers present a formula that defines the conditions for granting access control, or access controls that specify the access permissions; and *Data*

Recovery, papers 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.

4.2.2 Data Building

Data building gets valuable content from DB through software, analysis tools, web forms, and so on. Data collection, analysis and layout define the work methodology and help researchers build the desired results. Then, anyone may also benefit from data building by using data science approaches [29]. Works here also study ethical and privacy issues related to using data posted online and complying with General Data Protection Regulation (GDPR) or its variations. We classify a publication as data building when it focuses on properly collecting data, describing, examining, and developing a dataset. Such publications are then categorized as: *Device Data Extraction*, works that analyze data stored in devices to explore their vulnerabilities or security issues; *Data Recovery*, papers focus on recovering deleted files that compose evidence; and *Digital Evidence*, works focus on models, experiments, and data directly related to digital proof.

4.2.3 Cyber Attacks

There are several examples of criminal hacking actions that endanger sensitive data. Cybercriminals try to find weaknesses in systems through different techniques, which become increasingly sophisticated as security systems improve their operation.

In this context, DF becomes an essential tool allied to companies and citizens, as it seeks to identify cybercriminals and their attacks in a well-structured way. Indeed, the main goal is to access sensitive data for identity theft, fraud, extortion, scams, and other criminal and malicious practices.

Categories. Different papers address DF focused on cyberattacks, including domain hijacking [15], network intrusion [38], and dictionary attack [36]. Here, we classify a paper as a cyber attack when the authors address attacks related to databases in three different aspects: *SQL Injection*, papers focus on how the SQLi affects a DB or proposes tools for forensic analysis; *Attack Detection*, papers focus on detecting other kinds of cyber-attacks in forensic DB; and *Data Recovery*, papers lead to recovering data or systems after a cyber attack. Among those, SQLi is the most common one and explained next.

SQL Injection. SQLi is an attack approach based on manipulating SQL code, and is one common concern of DB professionals. A common example is injecting SQL code through the application login screen: "`select * from users where username = 'admin' -- and password = '1234';`". The issue is absence of verification and validation of data entered by a user, as the system concatenates the parameters to the query string. Thereby, when the attacker sends the comment symbol (−−), the DBMS ignores the remain of SQL code, disabling the password field.

4.2.4 Criminal Intelligence

Publications are classified as criminal intelligence when their goal is three-fold: *forensics investigation* to analyze DB contents, investigate DB incidents and build a timeline of illegal activities; *research products* to acquire a timely, valuable and accurate product from the logical processing of forensic case data (research products); and *crime resolution* to solve digital crimes by using forensic and computational techniques. Publications that address any of such branches are covered here.

5. LITERATURE REVIEW

This section discusses the publications filtered by the SLR. We now discuss the publications found in the methodology regarding Cyber Attacks (§5.1), and Criminal Intelligence (§5.2).

5.1 Cyber Attacks

Hackers hold criminal actions that put data about individuals or companies at risk. Here, papers mainly address *SQL Injection*, attack detection, and data recovery. We classify **nine** papers as Cyber Attacks, as they investigate digital crimes that expose, alter, disable, destroy, steal, gain access to or make unauthorized use of a system or device. Table 3 summarizes those publications: where **four** deal with SQLi, **three** perform attack detection methods, and **two** try to recover data from cyber attacks.

SQL Injection. Publications that address attackers who use SQLi to: obtain unauthorized access to a DB or read-protected data, corrupt DB, or grant access to unauthorized users. Pomeroy and Tan [44] highlight the challenges in recovering data after a SQLi attack. They develop a method for detecting SQLi attacks and recovering the excluded data. Likewise, Alam et al [7] explore SQLi vulnerabilities within the web applications of Bangladesh with *.bd* domain. They evaluate and analyze such vulnerabilities through a black-box penetration test. The results point SQLi exposure in over 600 web apps of the 900 considered in the study. The findings suggest experienced administrators do not acceptably maintain the web servers, as user input authentication and regular updates could prevent issues.

Kao et al [33] cover SQLi attacks with descriptive and investigative methods. They also propose a framework of SQLi Investigation Architecture (SIA) and prove its feasibility against SQLi attacks. Such solution can find hackers within the defined criteria of the SQLi attack, and detect issues or protect data against further attacks. Through machine learning, Xie et al [61] show a method for SQLi detection based on Elastic-Pooling Convolutional Neural Network (EP-CNN) and compare it with traditional

Table 3: Publications classified as Cyber Attacks, chronologically sorted.

	Reference, Keywords	Content
RECV	[38] Network Invasion	– Collect network data to support network forensic analysis, and store it in a MySQL DBMS
SQLI	[44] SQL Injection, Attack Reconstruction	– Propose a network recording solution to detect and capture SQLi. To validate it, they simulate an SQL injection attack on a local MySQL server as a data source
SQLI	[7] SQL Injection	– Evaluate about 900 public domain web apps searched on *Google* using a set of keywords
DETC	[15] Hijacking Detection	– Develop LUDIC (LookUp DIstributed Cache) for detecting domain hijacking attacks
DETC	[36] Wordpress	– Use *John the Ripper*, *Cain* and *Abel* dictionaries to decipher distinct types of passwords based on the password strengthening technique
SQLI	[33] SQL Injection	– Propose a framework (SQLi Investigation Architecture) to detect and combat SQLi attacks
SQLI	[61] SQL Injection, CNN	– Use 4.48 million real weblogs, of which 1/4 are SQLi logs and the others, common logs
RECV	[45] Industrial control	– Propose a forensic framework to analyze and retrieve data of control logic injection attacks
DETC	[57] Intrusion Detection	– List the different Intrusion Detection System (IDS) datasets used to evaluate IDS models

RECV: Data recovery. SQLI: SQL Injection. DETC: Attack detection.

ones. The method automatically extracts the hidden features (unrecognized by humans) of SQLi and identifies the attack traffic, bypassing the regular SQLi. Results show it to be effective, with high recognition accuracy compared to traditional ones.

Attack Detection. These publications focus on detecting many types of cyber attacks. For example, Borgwart et al [15] tackle the DNS Hijacking attack — a name service provided by the Domain Name System (DNS) that is essential for locating resources on the Internet, distributing security mechanisms in an authenticated manner, and facilitating future applications. The authors introduce LUDIC (LookUp DIstributed Cache) to detect and prevent outcomes of such attacks. LUDIC does not changes the existing infrastructure and can be easily integrated into an Intrusion Detection System (IDS) or a firewall while providing immediate benefits.

Kyaw et al [36] carry a dictionary attack experiment against WordPress handled by a fictional person. The attack broke the seven-character password by using well-known online dictionaries. The authors affirm applying password strengthening techniques can mitigate the attack. They also provide insights into implementing a forensic-ready WordPress system and investigating attacks on web applications, such as lockout systems, multi-factor authentication, strong passwords, and fake names.

Finally, Thakkar and Lohiya [57] overview Machine Learning and Data Mining techniques used for IDS, and discuss recent datasets that contain and organize network attack features and new attack categories. The types of network attacks changed over the years; hence, updating the datasets used for evaluating IDS is crucial. Then, the authors discuss recent advances in the attack detection datasets that are available for various research communities.

Data Recovery. These publications focus on recovering data or systems after a cyber attack or during forensics analyses. For example, Ming and Zhong [38] develop a network intrusion model as a forensic tool. The model performs intrusion detection while recovers all network data from the target system to simplify network forensic analysis. From another perspective, some cyberattacks inject malicious control logic into programmable logic controllers (PLCs) to sabotage physical processes (e.g., traffic light signals, nuclear plants). Hence, Qasim et al. [45] propose Reditus, a novel control-logic forensics framework for injection attacks that recovers control logic from suspicious industrial control systems (ICS) network traffic. Reditus assumes there is a built-in decompiler that can transform the control logic into its source code.

By analyzing the papers, we note three relevant aspects: *(i)* attacks collect and recover essential data from applications that use DB; *(ii)* having a DB with attack evidence is mandatory to perform forensic analyses; and *(iii)* when the target is a DB system, criminals prefer SQLi attacks.

5.2 Criminal Intelligence

Criminal intelligence include information compiled, analyzed, or disclosed to anticipate, prevent, or monitor illegal activity. A key point for criminal intelligence is DB, which store vast features on individuals, organizations, and transactions. Hence, gathering data is essential in any law enforcement agency. When acquired, the information from such data may anticipate or prevent crime by building a timeline of criminal activities.

Hence, papers classified as Criminal Intelligence show how DB can gather data and act as an investigation tool. By analyzing data within a DB, law enforcement agencies can better understand illegal activities and build a case to bring offenders to justice. This section then describes **40** papers classified as *Criminal Intelligence*, i.e., publications that analyze DB contents to investigate incidents and build a timeline of criminal activities (**13** in Forensic Investigation), process forensic case data to construct a general a product (**21** in Research Products) and solve forensic problems (**six** in Crime Resolution).

Forensic Investigation. Table 4 lists articles on forensic investigation processes that mainly use DB. Chang et al [17] describe how to use the Windows Registry for forensic analysis and investigation. Windows Registry is a central hierarchical DB with information for configuring system, applications and hardware devices. Hence, it is a signifi-

Table 4: Publications on Forensic Investigation, chronologically sorted.

Reference, Keywords	Content
[17] Windows Registry	– Use Windows registry database, which contains vital data used by windows and users
[39] Forensics on email evidence	– Use Enron Corp. E-mails, 8.70 GB, for 148 mailboxes with 517,431 messages and 3,299 folders
[60] Database Audit System	– Record and audit DB communication packets, which are mirrored by the switch on network traffic
[48] File Forensic Investigation	– Collect and store browser logs in a database considered as evidence of cybercrime
[2] IoT Investigation	– Consider different pieces of evidence that represent suspicious transactions
[9] CHAID, SQL	– SQL code for an automatic fraud-detection software application
[14] Digital traces, investigations	– Consider more than 200 real cases of cybernetics crimes registered by Geneva policy
[19] Database files, messaging	– User location and personal data of three message app in China and South Korea
[10] Privacy impact	– Assess privacy impact by considering DFaaS platform as a case study
[47] Decision making	– Discuss how scientific interpretation principles strengthen investigative process
[54] Mobile, Cloud Traceability	– Use WeChat app, which has data of photos and messages
[64] IoT Botnet Forensics	– Use the Mirai botnet server, available on GitHub
[40] Blockchain investigations	– Analyze blockchain to ensure data integrity on databases

cant forensic resource, and one may use it to prove the authenticity of judgment within the examination process and the forensic analysis phase.

Paglierani et al [39] share a systematic process for email forensics that integrates workflow into the normal forensic analysis and fits the distinct features of email evidence. They focus on detecting non-obvious artifacts related to email accounts, retrieving the data from the service provider, and describing email in a well-structured format.

Wu et al [60] design and implement a DB auditing system for distinct DB types. They use bypass mode to deflect any DB performance delays. Also, it provides a flexible audit system that decides which DB should be recorded or not.

Salunkhe et al [48] study how Decision Trees allow systems to quickly, easily, and affordably analyze log data on many formats for file forensic analysis. They propose an analysis strategy that aids investigators to detect criminal activities by collecting log files. When a crime occurs, the system investigates and stores the shreds of evidence in DB.

Al-Dhaqm et al [2] propose the Common Database Forensic Investigation Process (CDBFIP) to investigate cybercrime activities and cyber breaches over the Internet of Things. They propose a four-phase process: identification, artifact collection, artifact analysis, and documentation and submission process. Unifying these processes into an abstract diagram increases the knowledge available to users, newcomers, and professionals, and reduces the complexity and ambiguity of the investigation.

Another issue is to discern inexact or non-obvious similarities between cybercrimes. Hence, Bollé and Casey [14] propose a solution to finding links and repetitions between cases through the quasi-similarity calculation of distinct digital traits and the Levenshtein distance. Automatically detecting such similarities gives investigators a better understanding of the criminal context and the actual phenomenon, and can reveal many related crimes.

Bach et al [9] apply data mining techniques (CHAID decision tree) to discover patterns in fraud related to internal controls in a project-based organization. They increase the efficiency of internal fraud detection, which results in a SQL code used to develop an automatic fraud-detection application.

Choi et al [19] analyze the local and the formats of personal data files in three instant messaging apps (KakaoTalk, NateOn, and QQ) that use encryption. Thus, the authors apply reverse engineering to examine the encryption and decryption procedures of the internal databases of such applications.

Zhang et al [64] study a famous family of IoT bot malware – Mirai. They design a Mirai botnet network and run forensics analyses on its server. They identify and discuss forensic items left on the attacker's terminal command, control server, DB server, scan receiver, loader, and network packets.

Bas Seyyar and Geradts [10] explore and measure privacy risks specific to law enforcement activities that require processing large amounts of data. They assess privacy impact (PIA) on a big data forensics platform. They also answer the question of how a PIA should be performed for large-scale DF operations and describe privacy risks and threats. Finally, they articulate concrete privacy measures to demonstrate compliance with the Policy Directive.

Sharma et al [54] present a mobile cloud forensics process that incorporates inter and intra-application analysis and time synchronization allied to traditional forensics. Time synchronization enables the investigator to perform forensic analysis of the mobile cloud application concisely; then, inter and intra-application analysis process ensures the extraction of forensic evidence and enriches the performance of event traceability in the cloud, using the metadata of possible mobile evidence.

Collecting and analyzing digital and multimedia evidence require many decisions from forensic professionals. Then, Ryser et al [47] present a well-established logical framework for making structured decisions at all stages of an inquiry in DF, aiming to mitigate the risks of errors and establish adequate trust in digital and multimedia evidence.

Palanisamy and Nataraj [40] review the application of blockchain in an enterprise information management system (IMS) and explore how to incor-

Table 5: Publications that propose Research Products, chronologically sorted.

Reference, Keywords	Content
[26] Anti-Money Laundering	– Use a database with 100 fictitious records to emulate banking transactions
[42] Information Accountability	– Present a tool to detect tampering in high-performance DB.
[51] Enterprise Rights	– Microsoft RMS and Adobe LiveCycle
[43] Android Timestamps	– Introduce the Authenticity Framework for Android Timestamps (AFAT), tested over SQLite DB
[31] Education	– Interviews with forensic experts as judges, lawyers and prosecutors
[37] Forensic analysis on Android	– Present a tool to forensic analysis Android apps (Fordroid) tested over a 100 android app DB
[56] Digital forensics language	– Present a new digital forensics language called Nugget
[32] AI Speaker Ecosystems	– Use a DB with audio files from 4 models: Clova da NAVER, Kakao I da KAKAO, NUGU da SKT and GiGA Genie from KT
[5] Integrated Incident Database Forensic	– Use data from IEEE Xplore, Scopus, ACM, SpringerLink and Elsevier to develop an Integrated Incident Response Model
[30] Learning Efficacy of Digital Forensics Concepts	– Present an openly available virtual reality (VR) digital forensic education game via the Immersive VR Education ENGAGE platform
[50] Metadata-based storage reconstruction and carving	– Present the tool LAYR (available on GitHub) and use data from the Digital Forensics Tool Testing (DFTT) project to evaluate it
[49] Main Memory Images	– Perform experiments to build a dataset with adulterated images from main memory
[59] Digital forensics as a service	– Describe learned lessons of using DF as a Service ([58]) platform in a forensic and legal context
[62] True source of cyber crime	– Propose a framework (Root-Tracker) for identifying real source of cybercrime, and evaluate it over a network infrastructure built with different devices
[24] Cyber attacks	– Propose a framework for reviewing/ investigating cyber-attacks (D4I) and evaluate it over phishing
[6] Database Forensic	– Validate the Database Forensic Investigation Metamodel
[23] IA forensics	– Show how databases serve as valuable input for IA models supporting forensic analysis
[52] VISU criminal suspect	– VISU is connected to a criminal and suspect DB provided by law enforcement authority
[25] MORPH database	– Empirically evaluate on the academic MORPH database using a facial recognition system
[63] real-time image	– Provides YOLO v5, a new method for database auditing
[8] Unified Forensic Model	– Propose the Unified Forensic Model (UFM) for the database forensics analysis

porate blockchain technology in CampusStack, an integrated IMS, to audit the DB and ensure data integrity. They conclude decentralization must be considered and adopt the data auditing algorithm with the blockchain technology in critical DB.

Research Products. Here, papers propose new frameworks, tools, models resulting from research, as listed in Table 5. By analyzing **21** papers in this group, we found: three tools – DRAGOON, Fordroid, LAYR [37, 42, 50]; three frameworks – AFAT, Root-Tracker, D4I) [24, 43, 62]; and one digital forensics language – Nugget [56]. The search strings returned works since 2006, but the oldest one classified as Criminal Intelligence that proposes a tool is from 2012, indicating a recent research field in forensics. Also, the three tools have distinct goals: DRAGOON detects tampering in DB; Fordroid analyzes Android mobile apps; and LAYR combines methods to optimize storage reconstruction techniques. The frameworks have also distinct goals: AFAT investigates the authenticity of timestamps on Android smartphones; Root-Tracker aims to identify the source of network security attacks; and D4I improves the digital forensics process, especially the examination of cyber-attacks. Then, Nugget is a domain-specific language to specify the data flow of a forensic inquiry, execute forensic computation, and return a log of the inquiry.

There are also two works on services. Van Beek et al [59] provide Digital Forensics as a Service (DFaaS) implementations to agencies and share lessons learned within a forensic and legal context. Then, Henseler and van Loenhout [31] depict standards and requirements of the Dutch Register of Judicial Experts (NRGD), with requirements for a person to qualify as an NRGD Registered Specialist.

The remaining works tackle other current issues. First, in Internet of Things (IoT) and Digital Forensic, an AI speaker is a cloud-based IoT system built by merging an AI speaker and IoT devices. Such AI speakers are continuously operating and may provide vital evidence for digital forensics. However, privacy issues may arise. Hence, Jo et al [32] propose five digital forensic analysis models for four distinct AI speakers. They introduce a forensic tool for collecting user command history for NAVER Clova (Korean IA Speaker) as a research product. Likewise, Al-Dhaqm et al [5] present an Integrated Incident Response Model (IIRM) to recognize, respond, mitigate and recover from a potential database incident. IIRM is a hybrid model with four main goals: establish a plan to prevent DB disasters, investigate and seek possible evidence, recover DB operations, and share DB disaster knowledge.

Hassenfeldt et al [30] present an openly available virtual reality (VR) DF education game via the Immersive VR Education ENGAGE platform. They investigate methods to hack and extract data from the Nintendo 3DS storage system, the NAND.

Flores et al [26] use data mining and data warehousing to assist digital forensic investigations related to money laundering in compliance with the Know-Your-Customer 'KYC' policies defined inside an organization. They prove BI tools could support the analysis of money laundering evidence by using simple DB transactional logs to present the investigation results more comprehensively than using extensive written Suspicious Activity Reports (SARs).

Table 6: publications that propose Crime Resolution, chronologically sorted.

Reference, Keywords	Content
[18] Mining Criminal Databases	– Build and use a data warehouse with records of 378,000 cars stolen in 11 years in Taiwan
[21] Crimes from Web	– Consider over 33k crime reports in the "Onde Fui Roubado" platform within [2012,2016]
[27] Cryptocurrency investigations	– Provide an empirical analysis of CoinJoin transactions
[41] Vulnerabilities, Forensic Issues	– Discuss various types of cloud attacks and inform how to mitigate them
[1] Crime-scene image	– Apply machine learning to a subset of 60,520 images from the illicit drug database
[12] Suspect Identification	– Train a model for suspect id over 100k images of CelebaFaces

Schrittwieser et al [51] explore forensic techniques for Enterprise Rights Management (ERM) systems and develop application-specific guidelines for forensic investigations, targeting Microsoft Active Directory Rights Management Services (RMS) and Adobe LiveCycle Rights Management. Such ERM systems heavily use databases to store keys and relevant forensic investigation data. Also, they show the critical role of database forensics for inquiries in ERM systems. They conclude that MySQL InnoDB storage engine's data and log files store enough data to reveal older versions of LiveCycle policies and even allow recovering deleted cryptographic keys.

Schneider et al [49] study how to manipulate main memory copies obtained during a digital investigation based on controlled experiments. Note, tampered digital evidence may compromise its interpretation. Handling prominent memory dumps is problematic, as they detect most spoofs. Overall, tampering with main memory dumps seems more difficult than tampering with hard disk images, but the likelihood of misleading an analyst is also higher.

Al-Dhaqm et al [6] aim to validate the Database Forensic Investigation Metamodel (DBFIM) by using the qualitative method of face validity. DBFIM proposes resolving interoperability, heterogeneity, complexity, and ambiguity in a database forensics investigation, where various models were identified, collected, and reviewed to develop DBFIM.

Delgado et al [23] argue that AI connects scientists and forensic investigators. Indeed, they highlight the growth of data analytics in all fields of life, starting from enterprises to public health solutions during the COVID-19 pandemic. Such examples are relevant instances of data analytics for society and are suitable for forensic intelligence.

Sethuraman et al [52] present VISU, a prototype of a three-dimensional printed robot for crowd surveillance. It captures the surround with a built-in camera and sends it to a cloud DB. Then, VISU connects to the criminal and suspect DB provided by an authority to identify any suspicious activity. The authors suggest VISU works for crowd surveillance purposes in both crowded and non-crowded locations with minimal production cost.

On facial recognition analysis, Drozdowski et al [25] analyze the watch list imbalance effect, which can cause (unintentional) discrimination based on individual demographic properties; i.e., the impacts of an unbalanced DB regarding a specific demographic group (e.g., containing many more men than women). They conduct an empirical evaluation of the MORPH DB by using a facial recognition system. The authors show unbalanced samples influence the fairness of biometric identification systems. Further, it indicates demographically equitable biometric verification systems do not necessarily guarantee demographically fair identification.

Zhang et al [63] propose a real-time DB operation recognition method based on the YOLO algorithm. It performs real-time image recognition on the DB operation and maintenance interface. The instruction image converts into a character sequence, and the DB audit is complete according to the rules.

Alhussan et al [8] propose the Unified Forensic Model (UFM): initialization, acquisition, investigation, restoration and recovery, and evaluation (each with processes and activities). UFM aims to collect, preserve, identify, analyze, reconstruct, and document DB incidents in the forensic field.

Crime Resolution. Here, the publications discuss new computational solutions that may help in crime resolution, as listed in Table 6. Chen [18] applies data mining techniques (such as classification, clustering, association rule, and prediction) to manipulate data on stolen automobiles in Taiwan. He explores information hidden in such data and provides knowledge to transportation, insurance, and police agencies for decision support. The results allow understanding of automobile theft, finding stolen automobiles, questioning theft claims, etc.

Silveira and Brandão [21] collect data from websites with crime occurrences and apply clustering analysis to discover crime patterns. They highlight that more than 41% of the crimes were not reported, most of which are thefts and robberies occurring at night and dawn. Moreover, minor offenses present different patterns of serious crimes, and crime patterns differ in rich and poor neighborhoods.

Analyzing cryptocurrency payment flows has become a critical forensic method in law enforcement and is used to investigate a broad spectrum of criminal activities. Then, Fröwis et al [27] identify internationally accepted standards and rules for supporting suspicions and providing evidence in court, and project them onto current cryptocurrency forensics.

Pandi et al [41] discuss the STRIDE threat model and the misuse of cloud services for many malicious

Table 7: Proposed classification considering the phases of digital forensics and their guidelines.

Phase / Timeline	Goal	Methodology
Prevention Before crime	Prevent digital crimes	Vulnerability analysis and planning
Detection During a crime	Identify crimes – where how when	Find malicious activities among normal ones
Recovery After a crime	Restore data & processes	Uncover deleted, lost data; restore damaged evidence

purposes, as most criminals flee for lack of evidence. The article addresses some forensic issues in a cloud computing environment, such as collecting and analyzing evidence in the digital world.

Abraham et al [1] present two proof of concept for crime-scene image classification models. They evaluate two ML classifiers using images from an illicit drug database as ground truth for automatically cataloging and classifying crime scene images.

Bhajaj et al [12] propose a system to generate a face image by using a technology called Conditional Generative Adversarial Networks (C-GANs). The system considers gender, hair color, face shape, etc., as input data. It then shows a real-time face preview after inputting the data. Such a face-matching feature is evaluated over criminal databases to check whether the suspect already exists.

Here, publications reveal digital forensics has been proposing and using increasingly robust computational tools, models, or frameworks to identify and track crimes committed through digital devices. Also, exciting research presents methods or processes to guarantee that collected digital evidence serves as proof in a court of law. Hence, some works seek to solve or prevent crimes from occurring based on prior knowledge through data analysis.

Overall, criminal intelligence plays a critical role in DF. Indeed, our work may inspire or open paths for future research in integrating AI with DF. Combining such areas can provide potential solutions for forensic purposes. AI can help automate simple and repetitive tasks, freeing up time for investigators to focus on more complex issues. This integration may also inspire new advanced tools and techniques that can aid in preventing and detecting cybercrime. Integrating databases, digital forensics, and AI holds tremendous promise for the future of criminal intelligence and law enforcement, and further research in this area may yield significant benefits.

6. PUBLICATIONS BY PHASE

Simply put, within digital forensics, experts can act over three critical goals: prevention, detection, and recovery; as listed in Table 7. After explaining the taxonomy, we now draw a parallel between selected works according to each phase in Table 8. Next, we discuss the main aspects of such perspectives and how DB may relate to them.

Prevention. Preventing cybercrime requires digital forensic experts proactively identify vulnerabilities intrinsic to an organization's processes and systems. Such vulnerabilities may be understood as possible flaws in systems or processes that allow cybercriminals to exploit them to access data or even control the system. Thus, actions aimed at identifying and reducing the risks of vulnerabilities may include: implementing new information security technologies; acquiring or developing computational tools to perform periodic expert examinations; implementing methods of ensuring that systems are traceable (i.e., using logs); and constant monitoring and tracking any suspicious activity.

By drawing a parallel between this phase and our goal (identifying works that act direct or indirectly with DB), we classify as Prevention those papers whose goals (primary or not) include working to assist in: identifying vulnerabilities in DB systems and minimizing their risks; and ensuring data is traceable after a potential intrusion (Table 8).

Detection. Once a cyber attack is detected, DF experts must identify all evidence of such an incident and which systems, data and processes were affected. It is also crucial to keep all digital evidence found intact and unaltered to avoid inconclusive results. Further, tracking and analyzing all related data are needed to identify the attack causes.

We classify as *Detection* articles that propose models or methods to identify attacks in database environments. In addition, we also include works that aim to avoid violating the integrity of digital evidence found for ensuring its use as evidence in court. One of the most common methods of detecting and tracing attacks is the use of logs.

Recovery. After proper identifying and tracking all data and systems that have suffered a digital attack, recovering such losses is paramount. The damage levels in data and computational systems depend on the episode degree, i.e., detection time (to end such an attack asap), number of systems and volume of data involved, and so on. We classify as *Recovery* articles that propose efficient methods of recovering data or data systems to return to their total capacity. Recovery can include restoring corrupted, altered, or deleted data.

7. DISCUSSION AND CHALLENGES

Overall, Digital Forensics primarily collects, preserves and analyzes digital evidence for crime and legal related scenarios; whereas Databases are responsible for managing and storing data in an organized and efficient system. The intersection between these areas is vast, with many studies on device data extraction, data recovery, and digital evidence as instances of data building. Hence, this section answers each SLR question. We note such discussions may guide research in the intersection

Table 8: Papers on prevention (P), detection (D) and recovery (R), sorted by taxonomy class and year.

Ref.	Class	Forensic Phase			Ref.	Class	Forensic Phase			Ref.	Class	Forensic Phase			Ref.	Class	Forensic Phase		
[38]	CA		D		[51]	CI			R	[19]	CI	P			[41]	CI	P		
[44]	CA		D		[39]	CI			R	[5]	CI	P	D	R	[1]	CI		D	
[7]	CA		D		[60]	CI	P			[64]	CI		D		[6]	CI	P	D	
[15]	CA		D		[43]	CI		D		[10]	CI	P			[23]	CI	P	D	
[36]	CA	P			[48]	CI		D		[30]	CI		D		[52]	CI		D	
[33]	CA		D		[2]	CI		D		[49]	CI	P			[25]	CI	P		
[61]	CA		D		[21]	CI		D		[54]	CI		D		[63]	CI		D	
[45]	CA		D		[31]	CI	P			[59]	CI		D		[40]	CI		D	R
[57]	CA		D		[37]	CI			R	[47]	CI	P	D		[8]	CI	P		
[17]	CI	P			[56]	CI			R	[27]	CI			R	[12]	CI		D	
[18]	CI		D		[14]	CI	P	D		[62]	CI		D						
[26]	CI		D		[9]	CI		D		[24]	CI		D		**CA:** Cyber Attacks				
[42]	CI		D		[32]	CI		D		[50]	CI			R	**CI:** Criminal Intelligence				

between databases and digital forensic, reducing the gap between both knowledge areas.

When and where the studies were published? The search strings retrieved relevant works since the year 2006. However, only from 2017 onwards there was a growth in the number of publications that address digital forensics and databases. In general, *String 3* retrieved more papers, which may indicate trending keywords. Also, Scopus and Science Direct are the digital libraries that returned more publications.

What kinds of research are there? Works use different statistical strategies to prove or validate their hypotheses; i.e., mostly quantitative. However, we also find exceptions that use a qualitative approach and present reviews or analyses of forensic methods or issues in a non-numerical way, e.g., [31].

What is the focus of data-driven digital forensics? We identified four classes of works, each with three categories: *Data Building* – Device Data Extraction, Data Recovery, and Digital Evidence; *DBMS* – Performance Analysis, Security Roles, and Data Recovery; *Cyber Attacks* – SQLi, Attack Detection and Data Recovery; and *Criminal Intelligence* – Forensic Investigation, Research Products, and Crime Resolution. SLR results for the first two are in [53]. Still, almost half of the articles focuses on Criminal Intelligence, i.e., digital forensics follows the current trend of intelligent solutions.

We have also mapped all articles according to the phases of DF covered: Prevention, Detection, and Recovery. Grouping publications on stages allows companies and researchers to quickly find current solutions for one or a combination of phase(s).

What are the advances and potential challenges in the area? It is unquestionable that forensic investigation is data-driven. Indeed, DF and DB have an almost interdependent relationship as both areas have data as a central element. On the one hand, research on databases (mostly) aims to discover how to optimize data storage and retrieval (querying) as well as improve internal tools and mechanisms, e.g., logging and transaction control. On the other hand, research on data-oriented digital forensics acts on at least two main points: organizing/storing data for further investigation or using the data stored for investigations. We also note advances in using machine learning techniques (mostly mining) over forensics data. Chen [18] applies data mining over stolen automobile data and uses classification, clustering, association, and prediction algorithms.

What issues are still open? In 2010, Garfinkel [28] pointed out some technological challenges for digital forensics, including: the increase in device storage capacity, which makes processing data promptly difficult; the diversity of hardware that complicates data access standardization; and the proliferation of different data, operating systems and file systems that increase the complexity and cost of developing tools to exploit data. Technology has since changed a lot, but so have the issues. For example, one challenging opportunity is to keep in sync forensics and database updates on data formats, systems programming languages, etc. Such changes may favor security as technologies evolve and possible failures are already known (especially for cybercrimes). Still, new technologies also mean new forms of cyberattacks. Especially for databases, another challenge is maintaining compatibility between systems.

Further, new technologies demand preparing specialists and building approaches to forensic investigation. For example, many works focus on extracting data from devices with specific systems, such as android, IOS, cameras and games. One 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 to store and share knowledge involving these two areas.

Amidst the challenges that new technologies represent, cyberculture promotes a series of social changes with significant impacts on human relationships. The main reason is that it integrates individuals, companies, devices, networks, artificial intelligence, and the internet of things without the barriers of time or geographic limitations. The constant

exchange of information means the virtual environment needs to be regulated in the ethical field (using and treating personal and private data) within private and public settings.

Further, privacy has changed through the years. Still, individual guarantees and freedoms (e.g, the right to privacy, intimacy, and the inviolability of the home) are protected by the legal system of states and countries. Hence, a key open issue is dealing with data during investigations with proper ethical conduct, which is still fuzzy worldwide.

In summary, the intersection of databases with digital forensics is essential for investigating cybercrimes and legal disputes, as well as preserving data security. Professionals in such areas must know the techniques and tools available to ensure that data can be collected, preserved and analyzed correctly during investigations while adopting ethical behavior. After all, having ethics is crucial when conducting forensic database analyses to ensure proper, responsible work (and research).

8. CONCLUSION

This article presented a systematic literature review on the intersection between digital forensics and DB areas. One goal was to promote a better categorization and synthesis and, hence, to simplify the search and access to related work. We proposed a new taxonomy (focusing on cyber-attacks and criminal intelligence), and uncovered a specific pattern of publications – i.e., the three phases: prevention, detection, and recovery. Besides classifying all papers within the taxonomy, we have also mapped them to the patterns. Such a mapping enables researchers and organizations to quickly find leading solutions grouped by forensic purpose.

Based on the SLR, we may conclude: the number of publications on digital forensics has increased, given the evolution of computing and the relevance of solving digital crimes; forensic expertise has expanded in various computational and digital contexts; and there are distinct digital forensics functions related to data usage and development. Advances in DB are crucial to forensic investigations as well, not only in the search for better processing speed (through indexes and specialized access methods) but also in the accuracy of results (through new query techniques and correlation, for example).

Future work shall expand the coverage of publications over other areas related to Databases, such as Data Mining and Machine Learning, which may aid forensics analysis. Also, we plan to study further privacy and ethical issues uncovered here.

Acknowledgments. Funded by Research Scholarship Programs - IFMG, CNPq, and CAPES, Brazil.

9. REFERENCES

[1] J. Abraham et al. Automatically classifying crime scene images using machine learning methodologies. *Forensic Sci Int'l: Dig Investigation*, 39, 2021.

[2] A. Al-Dhaqm et al. Cdbfip: Common database forensic investigation processes for internet of things. *IEEE Access*, 5:24401–24416, 2017.

[3] A. Al-Dhaqm et al. Categorization and organization of database forensic investigation processes. *IEEE Access*, 8:112846–112858, 2020.

[4] A. Al-Dhaqm et al. Database forensic investigation process models: A review. *IEEE Access*, 8:48477–48490, 2020.

[5] A. Al-Dhaqm et al. Towards the development of an integrated incident response model for database forensic investigation field. *IEEE Access*, 8:145018–145032, 2020.

[6] A. Al-Dhaqm et al. Face validation of database forensic investigation metamodel. *Infrastructures*, 6(2):1 – 19, 2021.

[7] D. Alam et al. A case study of sql injection vulnerabilities assessment of .bd domain web applications. In *CyberSec*, pages 73–77, 2015.

[8] A. A. Alhussan et al. A unified forensic model applicable to the database forensics field. *Electronics (Switzerland)*, 11(9), 2022.

[9] M. P. Bach et al. Internal fraud in a project-based organization: Chaid decision tree analysis. *Procedia Computer Science*, 138:680–687, 2018.

[10] M. Bas Seyyar and Z. Geradts. Privacy impact assessment in large-scale digital forensic investigations. *FSI: Dig. Investigation*, 33:200906, 2020.

[11] A. Beirami et al. Trusted relational databases with blockchain: design and optimization. *Procedia Computer Science*, 155:137–144, 2019.

[12] P. Bhajaj et al. Figsi—facial image generation for suspect identification. *LNNS*, 351:877 – 891, 2022.

[13] D. M. Blei et al. Latent dirichlet allocation. *JMLR*, 3:993–1022, 2003.

[14] T. Bollé and E. Casey. Using computed similarity of distinctive digital traces to evaluate non-obvious links and repetitions in cyber-investigations. *Dig. Investigation*, 24:S2–S9, 2018.

[15] A. Borgwart et al. Detection and forensics of domains hijacking. In *GLOBECOM*, 2015.

[16] E. Casey. *Digital evidence and computer crime: Forensic science, computers, and the internet*. Academic press, 2011.

[17] K. Chang et al. Initial case analysis using windows registry in computer forensics. In *FGCN*, 2007.

[18] P. S. Chen. Discovering investigation clues through mining criminal databases. In H. Chen and C. Yang, editors, *Intelligence and Security Informatics: Techniques and Applications*, pages 173–198. Springer Berlin Heidelberg, 2008.

[19] J. Choi et al. Digital forensic analysis of encrypted database files in instant messaging applications on windows operating systems: Case study with kakaotalk, nateon and qq messenger. *Dig. Investigation*, 28:S50–S59, 2019.

[20] J. Cohen. A coefficient of agreement for nominal scales. *EPM*, 20(1):37–46, 1960.

[21] M. da Silveira and W. Brandão. Characterizing crimes from web. In *BraSNAM*, 2017.

[22] D. Dave et al. Management of implicit requirements data in large SRS documents: Taxonomy and techniques. *SIGMOD Rec.*, 51(2):18–29, 2022.

[23] Y. Delgado et al. Forensic intelligence: Data analytics as the bridge between forensic science

and investigation. *FSI: Synergy*, 3, 2021.
[24] A. Dimitriadis et al. D4i - digital forensics framework for reviewing and investigating cyber attacks. *Array*, 5:100015, 2020.
[25] P. Drozdowski et al. The watchlist imbalance effect in biometric face identification: Comparing theoretical estimates and empiric measurements. In *ICCVW*, pages 3750–3758, 2021.
[26] D. A. Flores et al. Combining digital forensic practices and database analysis as an anti-money laundering strategy for financial institutions. In *EIDWT*, 2012.
[27] M. Fröwis et al. Safeguarding the evidential value of forensic cryptocurrency investigations. *FSI: Dig. Investigation*, 33:200902, 2020.
[28] S. L. Garfinkel. Digital forensics research: The next 10 years. *Dig. Investigation*, 7:S64–S73, 2010.
[29] A. Guarino. Digital forensics as a big data challenge. In *ISSE*, pages 197–203, 2013.
[30] C. Hassenfeldt et al. Exploring the learning efficacy of digital forensics concepts and bagging & tagging of digital devices in immersive virtual reality. *FSI: Dig. Investigation*, 33:301011, 2020.
[31] H. Henseler and S. van Loenhout. Educating judges, prosecutors and lawyers in the use of digital forensic experts. *Dig. Investigation*, 24:S76–S82, 2018.
[32] W. Jo et al. Digital forensic practices and methodologies for AI speaker ecosystems. *Dig. Investigation*, 29:S80–S93, 2019.
[33] D. Kao et al. A framework for sql injection investigations: Detection, investigation, and forensics. In *SMC*, pages 2838–2843, 2018.
[34] M. Z. Khan et al. Cyber forensics evolution and its goals. In *Critical Concepts, Standards, and Techniques in Cyber Forensics*, pages 16–30. IGI Global, 2020.
[35] B. Kitchenham and S. Charters. Guidelines for performing systematic literature reviews in software engineering. Technical report, Un of Durham, 2007.
[36] A. K. Kyaw et al. Dictionary attack on wordpress: Security and forensic analysis. In *InfoSec*, pages 158–164, 2015.
[37] X. Lin et al. Automated forensic analysis of mobile applications on android devices. *Dig. Investigation*, 26:S59–S66, 2018.
[38] H. Ming and S. LiZhong. A new system design of network invasion forensics. In *ICCEE*, volume 2, pages 596–599, 2009.
[39] J. Paglierani et al. Towards comprehensive and collaborative forensics on email evidence. In *CollaborateCom*, pages 11–20, 2013.
[40] A. M. Palanisamy and R. V. Nataraj. A novel methodology to ensure data integrity in enterprise information systems using blockchain technology. In *ICEEICT*, pages 1–5, 2022.
[41] G. S. Pandi (Jain) et al. Exploration of vulnerabilities, threats and forensic issues and its impact on the distributed environment of cloud and its mitigation. *Procedia Computer Science*, 167:163–173, 2020.
[42] K. E. Pavlou and R. T. Snodgrass. Dragoon: An information accountability system for high-performance databases. In *ICDE*, 2012.
[43] H. Pieterse et al. Playing hide-and-seek: Detecting the manipulation of android timestamps. In *ISSA*, 2015.
[44] A. Pomeroy and Q. Tan, Effective sql injection attack reconstruction using network recording. In *IEEE CIT*, pages 552–556, 2011.
[45] S. A. Qasim et al. Control logic forensics framework using built-in decompiler of engineering software in industrial control systems. *FSI: Dig. Investigation*, 33:301013, 2020.
[46] Q. Rossy et al. Integrating forensic information in a crime intelligence database. *FSI*, 230(1-3):137–146, 2013.
[47] E. Ryser et al. Structured decision making in investigations involving digital and multimedia evidence. *FSI: Dig. Investigation*, 34:301015, 2020.
[48] P. Salunkhe et al. Data analysis of file forensic investigation. In *SCOPES*, pages 372–375, 2016.
[49] J. Schneider et al. Tampering with digital evidence is hard: The case of main memory images. *FSI: Dig. Investigation*, 32:300924, 2020.
[50] J. Schneider et al. Unifying metadata-based storage reconstruction and carving with layr. *FSI: Dig. Investigation*, 33:301006, 2020.
[51] S. Schrittwieser et al. Digital forensics for enterprise rights management systems. In *iiWAS*, 2012.
[52] S. C. Sethuraman et al. Visu: A 3-d printed functional robot for crowd surveillance. *IEEE Consumer Electronics Mag.*, 10(1):17–23, 2021.
[53] D. B. Seufitelli, M. A. Brandão, and M. M. Moro. Exploring the intersection between databases and digital forensics. *Journal of Information and Data Management*, 13(3), Sep. 2022.
[54] P. Sharma et al. Enhanced forensic process for improving mobile cloud traceability in cloud-based mobile applications. *Procedia Computer Science*, 167:907–917, 2020.
[55] L. F. Sikos. Packet analysis for network forensics: A comprehensive survey. *FSI: Dig. Investigation*, 32:200892, 2020.
[56] C. Stelly and V. Roussev. Nugget: A digital forensics language. *Dig. Investigation*, 24:S38–S47, 2018.
[57] A. Thakkar and R. Lohiya. A review of the advancement in intrusion detection datasets. *Procedia Computer Science*, 167:636–645, 2020.
[58] R. Van Baar et al. Digital forensics as a service: A game changer. *Digital Investigation*, 11:S54–S62, 2014.
[59] H. van Beek et al. Digital forensics as a service: Stepping up the game. *FSI: Dig. Investigation*, 35:301021, 2020.
[60] K. Wu et al. The design and implementation of database audit system framework. In *ICSESS*, 2014.
[61] X. Xie et al. Sql injection detection for web applications based on elastic-pooling cnn. *IEEE Access*, 7:151475–151481, 2019.
[62] P. R. Yogesh and D. S. R. Backtracking tool root-tracker to identify true source of cyber crime. *Procedia Computer Science*, 171:1120–1128, 2020.
[63] L. Zhang et al. Research and implementation of database operation recognition based on yolo v5 algorithm. In *CISAI*, pages 367–372, 2021.
[64] X. Zhang et al. Iot botnet forensics: A comprehensive digital forensic case study on mirai botnet servers. *FSI: Dig. Investigation*, 32:300926, 2020.