Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier

A Survey on Forensics and Compliance Auditing for Critical Infrastructure Protection

JOÃO HENRIQUES [1,2], FILIPE CALDEIRA [2,3], TIAGO CRUZ [1] (Senior Member, IEEE) and PAULO SIMÕES [1] (Senior Member, IEEE)

[1] University of Coimbra, CISUC, DEI, Pólo II - Pinhal de Marrocos 3030-290 Coimbra, Portugal
[2] CISeD – Research Centre in Digital Services, Polytechnic of Viseu, Viseu, Portugal
[3] Informatics Department, *Polytechnic of Viseu*, Viseu, Portugal

Corresponding author: Tiago Cruz (e-mail: tjcruz@dei.uc.pt).

ABSTRACT

The broadening dependency and reliance that modern societies have on essential services provided by Critical Infrastructures is increasing the relevance of their trustworthiness. However, Critical Infrastructures are attractive targets for cyberattacks, due to the potential for considerable impact, not just at the economic level but also in terms of physical damage and even loss of human life.

Complementing traditional security mechanisms, forensics and compliance audit processes play an important role in ensuring Critical Infrastructure trustworthiness. Compliance auditing contributes to checking if security measures are in place and compliant with standards and internal policies. Forensics assist the investigation of past security incidents. Since these two areas significantly overlap, in terms of data sources, tools and techniques, they can be merged into unified Forensics and Compliance Auditing (FCA) frameworks.

In this paper, we survey the latest developments, methodologies, challenges, and solutions addressing forensics and compliance auditing in the scope of Critical Infrastructure Protection. This survey focuses on relevant contributions, capable of tackling the requirements imposed by massively distributed and complex Industrial Automation and Control Systems, in terms of handling large volumes of heterogeneous data (that can be noisy, ambiguous, and redundant) for analytic purposes, with adequate performance and reliability. The achieved results produced a taxonomy in the field of FCA whose key categories denote the relevant topics in the literature. Also, the collected knowledge resulted in the establishment of a reference FCA architecture, proposed as a generic template for a converged platform. These results are intended to guide future research on forensics and compliance auditing for Critical Infrastructure Protection.

INDEX TERMS Critical Infrastructure Protection, Industrial Automation and Control Systems, Cybersecurity, Forensics, and Compliance Auditing.

I. INTRODUCTION

MODERN societies are increasingly dependent on essential products and services provided by Critical Infrastructures (CIs), supported by Industrial Automation and Control Systems (IACS) such as power plants, energy distribution networks, transportation systems, and manufacturing facilities. These IACS are becoming larger and more complex, due to the increasingly complex physical processes they manage and the increasing amount of (heterogeneous) data generated by a growing number of interconnected control and monitoring devices. These IACS are also heavily dependent on common IT systems whose security, management, and compliance must also be considered. This evolving scenario requires new strategies to improve the associated Critical Infrastructure Protection (CIP) frameworks.

A. THE CHALLENGE OF PROTECTING CRITICAL INFRASTRUCTURES

Critical Infrastructures (CIs) provide a series of essential services which are key to ensure the security, societal and economical activities of a country, thus constituting an

attractive target for cyber-attackers [1] [2]. Smart grids, water, oil, and gas distribution networks are becoming more complex due to the growing number of interconnected distributed devices, sensors, and actuators, often widely dispersed in the field, as well as the increasing amount of information exchanged among system components. Water-to-Wire generation, microgeneration, smart metering, oil, and gas distribution, or smart water management, among others, are pushing the boundaries of the classic Industrial Cyber Physical Systems (CPS) model, fostering a new generation of IACS and the Industry 4.0 paradigm [3]. Naturally, such developments have an impact on IACS cybersecurity requirements, due to a substantial increase in the scale and complexity of the protected infrastructure [4]. .

This increase in terms of interconnections has a direct impact in terms of the vulnerable attack surface, exposing the IACS to both traditional and new threats. For instance, according to IBM Managed Security Services data [5], attacks targeting IACS have increased over 110 percent in 2016. This is linked with the growing connectivity of industrial systems. Network-based attacks targeting Critical Infrastructure (CI) are also becoming a greater concern, as state-sponsored groups have become more active. Their activities comprise unauthorized access to government and corporate networks with the main purpose of gathering information, although they can be potentially disruptive for CIPs [6]. This trend is already a major concern, and is expected to further intensify in the future [7] [8].

Other IACS security threats come from their increasingly distributed nature, regarding both the physical processes under control, which have also become more widely dispersed and interconnected, and the associated control applications, which have also become increasingly distributed, for sake of scalability, elasticity, adaptability, resiliency, and fault-tolerance. Overall, this scenario makes it difficult to understand the nature of incidents and to assess their progression and threat profile. Moreover, defending against those threats is becoming increasingly difficult, requiring orchestrated and collaborative distributed detection, analysis, and reaction capabilities.

Continuously capturing live data from a running IACS system, that has an intrinsic volatile nature, presents important challenges to forensics investigators. For instance, volatile data in physical memory contains information about the current state of the system, such as process information, open network connections and encryption keys.

Another challenge comes from the amount of data to be collected, analyzed, and stored for detecting and profiling cyberattacks. According to IBM [9], the world produces over 2.5 quintillion bytes of data every day, and 80% of it is unstructured (and not analyzed). To improve decision making, enterprises are facing new challenges to collect a large amount of available data, retrieved from heterogeneous sources (including structured and unstructured data), and enriching it with the inclusion of additional contextualized data. In the specific scope of CIP, to face the tremendous growth of raw data being produced by sensors and process controllers, a Big Data approach is required to handle massive amounts of data in intensive online and offline processing flows. The growth of volume and heterogeneity of data sources, systems, workloads, and environment variability contributes to the complexity of data management. Traditional approaches, such as Relational Database Management Systems (RDBMS), might not be able to handle the deluge of industrial data they are experiencing, especially while addressing the need for improved performance, reliability, and user experience [10]. Gaining critical business insights by querying and analyzing such massive amounts of data is becoming a vital requirement [11].

B. THE NEED FOR BETTER FORENSICS AND COMPLIANCE AUDITING

Security incidents trigger a series of reactive activities, such as blocking access to and quarantining compromised systems, assessing the impact of the breach, mitigating the damage, and conducting forensics investigations to identify exploited vulnerabilities, identify the attackers, and enhance future defensive actions.

In 2020, it took an average of 207 days to identify a breach, and 280 days to contain it [12]. Such a scenario results from the current solutions demanding a multi-step process where security analysts goes to multiple systems to retrieve uncorrelated data and then correlate it manually. Moreover, the complexity, skillset, and costs required to deploy and operate those solutions present a significant number of obstacles to their adoption. Therefore, in addition to other security tools, such as specialized probes, intrusion detection platforms and firewalls, forensics tools are increasingly important for security professionals. Such tools provide the means to extract relevant insights and evidence from large volumes of heterogeneous data produced from the sources within the CI, which can be leveraged both for forensics and security analysis purposes.

Auditing compliance with applicable laws, regulations, policies, and standards processes also contributes to increase CI trustworthiness. However, such auditing processes are complex, since they need to use of a large number of tools, protocols and standards to correlate and enforce the audit compliance policies that may help to prevent future incidents.

Aggregating such tools in a unified platform can reduce the complexity, effort, and costs associated with investigating and connecting individual alerts to uncover potential threats. Such a proactive approach may avoid the disruption of operations and prevent evidence from being lost or corrupted. Moreover, it will also help deal with the evolving cyber threats affecting CIs, preparing the platforms for post-incident forensics analysis.

Due to the considerable overlap of functionalities associated with security forensics and compliance audit processes, it makes sense to consider them as unified platforms, which in this paper we generically designate as

FCA frameworks – even though many tools are applied only to one of these areas, they still share most requirements and technologies. In this work, we highlight the importance of both forensics and compliance auditing as high-priority topics for CIP.

C. SCOPE AND CONTRIBUTIONS

This paper surveys the research trends, challenges, and gaps in the field of FCA for Critical Infrastructures, exploring the most relevant approaches, methodologies, and technologies. To the best of our knowledge, this is the first survey specifically focused on FCA applied to the CIP domain.

Based on the lessons learned from the survey, this paper also presents:

- a classification taxonomy for the different aspects and technologies related with FCA systems.
- a reference architecture for converged FCA systems, identifying their key functional blocks.
- and a discussion of potential future directions for research on the subject of FCA for CIP.

As illustrated in Figure 1, the rest of the paper is organised as follows:

- The background of CIP and IACS security are introduced in Section II.
- The next three sections survey related works: Section III for forensics, Section IV for compliance auditing, and Section V for modern analytics applied to FCA, in the era of Big Data, AI and ML.
- Based on the lessons learned from the survey, the next two sections are devoted to classification and architectural models: Section VI proposes a new taxonomy for FCA systems for CIP, and Section VII introduces a reference architecture for FCA systems.
- Finally, Section VIII discusses achieved results and identifies open issues, while Section IX concludes the paper.

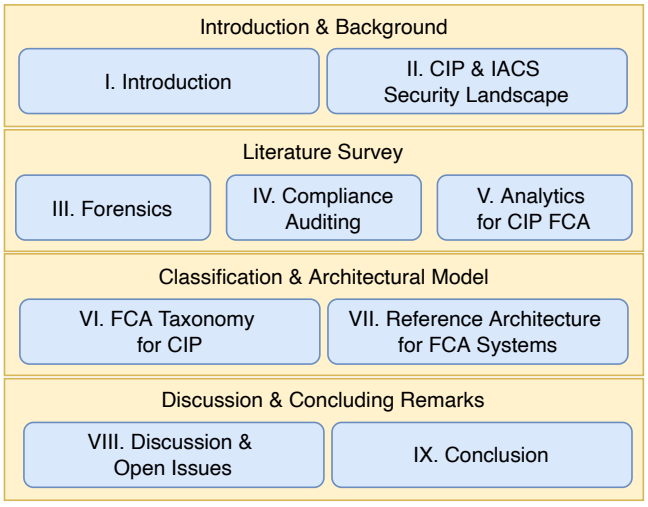

FIGURE 1: Structure of the paper.

II. CIP AND IACS SECURITY LANDSCAPE

In this section we provide an overview of CIP, with a more detailed perspective on IACS security – since most CIs are based on some sort of industrial control frameworks. The role of this section is to provide the reader with a more detailed perspective on how such systems are currently managed, from a security perspective, so that the associated needs, in terms of forensics and compliance auditing, become more clear. First, we discuss the role of IACS and Supervisory Acquisition and Data Control (SCADA) systems in CIP. Next, we introduce related security frameworks from NIST, ISO/IEC and other standards development organizations. Next, we address IACS security, introduce the concept of Security Information and Event Management (SIEM), and discuss other security analytics frameworks.

A. THE ROLE OF IACS AND SCADA SYSTEMS IN CIP

As already mentioned, a large number of CIs are based on large-scale IACS or Industrial Control Systems (ICS) which, traditionally, use SCADA systems to manage physical processes such as energy production and distribution, water and sewage treatment, traffic management and railways.

These SCADA systems can be roughly defined as a set of systems of command and control networks that control the operational sequence of the underlying physical processes. A typical SCADA system controlling CIs generally includes a control center and several field sites [13]. These sites are often distributed over a wide geographical area. Field sites are equipped with devices such as Program Logic Controllers (PLC)s or Remote Terminal Units (RTU)s [14], that control the on-site machines and periodically send information about the state of the field equipment to the control center. SCADA communications use a wide range of protocols, such as DNP3, Modbus, PCOM, ProfiNet, DeviceNet, ControlNet or Common Industrial Protocol [15].

In the early days, SCADA systems did not incorporate cyber-security mechanisms, since they were significantly resource-constrained and designed to run in isolated networks. They consisted of simple I/O devices transmitting signals between master and remote terminal units. Currently, SCADA systems can communicate over Internet Protocol (IP) networks, enabling its connection to the corporate network or even directly to the Internet, to integrate SCADA data with external systems such as Enterprise Resource Planning and Business Process Management tools. This interconnection of SCADA systems with wider networks brings new threats for which they were not originally designed, making them much more vulnerable. Moreover, CIs such as smart grids and water distribution networks have become increasingly complex due to the number of interconnected distributed devices, sensors and actuators, often widely dispersed in the field, and the larger amount of information exchanged both within the control system and between the control system and external systems.

As pointed out by Ahmed et al. [13], Cornelius and Fabro [16], and Eden et al. [17], the different nature of SCADA

systems also raises important challenges in the application of forensics, when compared to traditional approaches. Those classic forensics methodologies potentially interfere with the IACS operation, since they may introduce latency and cause critical processes to fail. Another challenge arises from the use of resource-constrained devices such as RTU and PLC, which often lack the storage and processing capabilities required by forensics tools. Also, SCADA logs might be not suitable for forensic investigation, as they are geared towards process management, not cybersecurity. Nonetheless, there is still a general lack of SCADA-specific forensics tools.

To prevent known and unknown attacks, including security vulnerabilities and threats, organizations are adopting a common set of defense solutions such as firewalls, antivirus, Intrusion Detection System (IDS)s, Intrusion Prevention System (IPS)s, and SIEM [18, 19]. Eden et al. [17] provided an overall forensic taxonomy of the SCADA system incident response model and discussed the development of forensic readiness within SCADA system investigations, including the challenges faced by the SCADA forensic investigator and suggested ways in which the process may be improved. van der Knijff [20] identified possible sources of evidence in the investigation process in CI. Some of them include engineering workstations, databases, historian, Human Machine Interface (HMI), application server, Field devices like PLC, RTU, Intelligent Electronic Devices (IED), firewall logs, web proxy cache, and ARP tables.

B. SECURITY FRAMEWORKS

Several security frameworks incorporate a series of documented processes used to define the policies and procedures around the implementation and management of information security controls in an enterprise environment. These frameworks are a blueprint for building an information security program to manage risk and reduce vulnerabilities by applying a function for identifying, protecting, detecting, and responding to activities. These frameworks can help information security professionals to define and prioritize the tasks required to manage their organizations' security. Examples of IT security frameworks include Control Objectives for Information and Related Technology (COBIT) [21], ISO 27000 series [22], NIST Special Publications 800-53 [23], 800-171 [24], NIST Cybersecurity Framework for Improving Critical Infrastructure Cybersecurity [25] and HITRUST CSF. The HITRUST CSF represents a certifiable framework that provides a comprehensive, flexible, and efficient approach to regulatory/standards compliance and risk management [26].

NIST SP 800-53 is the standard required by United States (US) federal agencies but could also be used by any company to build a technology-specific information security plan [27]. NIST 800-171 [24] provides federal agencies with recommended security requirements for protecting the confidentiality of controlled unclassified information.

The ISO/IEC 27000 series provide key information security frameworks applicable to any industry [28, 22]. For instance, ISO/IEC 27004:2016 provides guidelines supporting organizations in assessing security performance and effectiveness indicators [29] to fulfill the requirements of ISO/IEC 27001:2013, with ISO/IEC 27005:2018 providing guidelines for information security risk management. ISO/IEC 27037:2012 provides guidelines on the handling of digital evidence, including identification, collection, acquisition, and preservation of potential digital evidence [30]. ISO/IEC 27038:2014 covers the techniques for performing digital redaction on digital documents [31]. ISO/IEC 27042:2015 provides guidance on the analysis and interpretation of digital evidence keeping continuity, validity, reproducibility, and repeatability [32]. ISO/IEC 27050 represents a group of standards (27050-1 to 27050-3) addressing the discovery of Electronically Stored Information, a term coined to refer to*forensic evidence in the form of digital data* [33].

Also within the ISO/IEC 27000 series, ISO/IEC 27041:2015 provides guidelines on how to make sure that the methodologies and processes used to investigate information security events are suitable [34]. ISO/IEC 27043:2015 includes guidance for common incident investigation techniques across numerous incident investigation scenarios utilizing digital evidence, based on idealized models [35]. ISO/IEC 27006:2015 specifies requirements and guidance providing audit and certification of information security management systems [36]. ISO/IEC TS 27008:2019 provides guidance for evaluating the implementation and operation of information security controls, including their technical assessment, following an organization's established information security requirements, including technical compliance [37]. ISO/IEC 27040:2015 provides technical recommendations on how organizations can establish an appropriate level of risk mitigation by using a tried-and-true approach to data storage security strategy, design, documentation, and implementation.

Moreover, there are other relevant standards within the ISO/IEC frameworks, such as ISO 21043-1:2018 that introduces important terms and definitions in forensic sciences [38], also providing the requirements for the forensic process with a focus on the recognition, recording, collection, transport, and storage of potential forensic items [39]. Also, ISO/IEC 30121:2015 is a framework for helping organizations to be prepared for digital investigation processes [40].

Regarding forensics education and training, the ASTM standards are worth mentioning [41]. ASTM E2678 helps promoting computer forensics by developing model courses that are compatible with other forensic science programs. ASTM E2917 provides core standards for forensic science practitioners' training, continuing education, and professional development, including training criteria for competency, training documentation and implementation, and continual professional development. ASTM E2916 takes computer forensics, image analysis, video analysis, forensic audio, and facial identification are just some of the phrases

and definitions that are utilized in the study of digital and multimedia evidence.

Deciding upon the applicable regulatory or standardisation frameworks an organization must comply with must consider several factors such as the type of industry or country-specific compliance requirements. For example, US traded companies may start by complying with Sarbanes-Oxley [42] and COBIT. In case the company needs information security capabilities the option is ISO 27000 certification. NIST SP 800-53 is the standard required by US federal agencies but could also be used by any company to build a technology-specific information security plan. The HITRUST CSF integrates well with healthcare software or hardware vendors looking to provide validation of the security of their products. NIST 800-94 [43], was introduced in 2007 highlighting the challenges in the detection accuracy, extensive tuning, blindspots, and performance limits.

C. IACS SECURITY

Although the protection of CI is a topic not necessarily dependent on technology, this survey is driven by a technological approach focused on IACS protection. IACS incorporate Control Systems (CS) designed to manage and control physical processes, constituting one of the main targets for CIP activities. These CS can be defined as manual or automatic mechanisms used to manage dynamic processes by adjusting or maintaining physical quantities such as mass, temperature, or speed. CS are classified in two distinct categories: open- and closed-loop. Open-loop CS generate their output based on input only, while in a closed-loop the output is used as a feedback mechanism together with inputs to generate new output [20]. CS are generally used for monitoring and controlling industrial and infrastructure processes and dispersed assets supported by centralized data acquisition and supervisory control, often constituing a CPS. In the scope of the so-called essential services, these CPS are vital, often being highly interconnected and mutually dependent.

2010's Stuxnet [44] and 2015's BlackEnergy [45, 46] demonstrated that the so-called *security by obscurity* approach is no longer adequate for CIs. Stuxnet was the first known malware specifically designed to target automation systems, infecting between 50,000 to 100,000 computers worldwide. BlackEnergy was directly responsible for power outages for 250,000 customers in western Ukraine. Since then, many other attacks targeting IACS were recorded, such as Gauss, Havex, and Shamoon [47].

This situation has prompted the development of suitable mitigation mechanisms to deal with cyberthreats against IACS which may compromise integrity, information/control confidentiality or availability [48], such as unauthorised accesses, break-ins, penetration attempts, and other forms of abuse, to detect and secure the automation infrastructure perimeter from attacks [49].

Among these mechanisms, IDS provide the means to monitor the infrastructure, detecting security anomalies or suspicious behaviour by resorting to signature (rule-based) [50] or anomaly detection strategies [51]. Due to their nature, IDS often constitute one of the most relevant data sources for FCA purposes, detecting threats and recording incident-related valuable evidence for forensics analysis purposes, helping understand attacks and prevent them in the future.

While the IDS concept was borrowed from the Information and Communication Technologies (ICT) world, its deployment in IACS must obey a specific set of restrictions calling for the development of domain-specific approaches [52]. As a result, several proposals for IACS IDS have been presented over the past years, covering several levels of the automation infrastructure, from the field-level, as it is the case for the Shadow Security Unit (SSU) PLC security monitor [53], to higher levels, as it is the case for Rosa et al. [3], which presented a distributed security framework for IACS. IDS systems can be classified according to their targets, as it is the case for Network Intrusion Detection System (NIDS) and Host Intrusion Detection System (HIDS) [54].

IPS are the natural counterpart for IDS, providing active response capabilities, with Intrusion Detection and Prevention System (IDPS) combining both detection and response capabilities [54]. However, it must be said that automatic reaction mechanisms are often avoided by CI operators, due to the risk of a knowledgeable attacker abusing them for its own purposes.

Nevertheless, components such as IDS can not provide an encompassing level of protection for the infrastructure, a situation that requires the adoption of a structured approach capable of providing collection, analysis and storage for monitoring information coming from the entire IACS infrastructure. SIEM systems, which will be next presented, constitute one of the most popular approaches to consolidate diversified and relevant information, leveraging it for analytics purposes.

D. SECURITY INFORMATION EVENT MANAGEMENT

SIEM systems are designed to collect and correlate security log data (record of events that occurred on a computer or network device) from a wide variety of sources within organizations, including security controls, operating systems, and network infrastructure, systems and applications. Their data sources include log data and network telemetry data from flows and packets. Typically, their blocks include source device, log collection, parsing normalization, rule engine, log storage, and event monitoring [55]. Once the SIEM has the log data, data are normalized and further analysis generates alerts when suspicious activity is detected. Moreover, SIEM provides reports on the request of administrators. Some SIEM products can also act to block malicious activity, for instance by running scripts (e.g. triggering reconfiguration of firewalls and other security controls). Forensic investigations will benefit from correlating the collected data with the

information from the context, including assets, users, threats, and vulnerabilities.

As stated by Gartner [56], SIEM technology provides Security Information Management, log management, analytics, compliance reporting, and Security Event Management. They provide real-time monitoring and incident management for security-related events from networks, security devices, systems, and applications.

SIEM technology is typically deployed to support three primary use cases: advanced threat detection, basic security monitoring, and forensics and incident response. Forensics and incident response contributes with dashboards and visualization capabilities, as well as workflow and documentation support to enable effective incident identification, investigation and response. Basic security monitoring includes log management, compliance reporting, and basic real-time monitoring of selected security controls. In the case of advanced threat detection, it includes real-time monitoring and reporting of user activity, data access, and application activity, incorporation of threat intelligence, business context and ad hoc query capabilities. At the most basic level, a SIEM system can be supported by rules or employ a statistical correlation engine between event log entries. Pre-processing may happen at collectors, with only part of those events being moved to a centralized management component, reducing, in this way, the volume of information being communicated and stored. Notwithstanding, this approach can discard important events too early [57].

Sun et al. [58] presented an event-linked network model to query and organize big volumes of data. In this model, events are primary units in organizing the data, whereas links represent the association among them. This model is applied in Cloud or virtual-environment analysis, as a huge quantity of involved data, such as the case of the Internet service provider with SIEM solution having a huge quantity of data at centralized locations.

In 2021, Gartner Magic Quadrant for SIEM identified the following market leaders: Exabeam, IBM, LogRhythm, Rapid7, Securonix, Splunk, Splunk, HPE, and Intel Security [56]. A survey of those solutions, including an analysis of external factors affecting the SIEM landscape in the mid and long-term, can be found in [55]. Authors concluded that SIEM systems are slowly converging with Big Data analytics tools.

E. OTHER SECURITY ANALYTICS PLATFORMS

Active security and forensic capabilities are typically offered separately by different security systems [59]. While SIEM have pushed for the development of complementary approaches for collecting and analyzing event data to identify and respond to advanced attacks, several operators have found them to be somehow limited due to reasons such as the lack of orchestration capabilities, prompting the emergence of a new generation of security analytic technologies. Next, we introduce some of those tools and technologies.

Endpoint Detection and Response (EDR) is a complementary software to SIEM, extending detection and response capabilities by acting as an additional log source. According to Gartner [60], EDR is "a SaaS-based, vendor-specific, security threat detection and incident response tool that natively integrates multiple security products into a cohesive security operations system that unifies all licensed components." EDR solutions record and store system-level endpoint behaviors, and include several data analytics techniques to detect suspicious system behavior, provide contextual information, block malicious activity, and provide remediation suggestions to restore affected systems [61].

While the primary incident response tools for security teams are EDR platforms, emerging Extended Detection and Response (XDR) products integrate a set of security products into a cohesive security incident detection and response platform. Gartner defines them in a category aggregating and correlating telemetry from different sources to synthesize and draw conclusions to enable automated response actions. In comparison to XDR and SIEM and Security Orchestration, Automation and Response (SOAR) tools, XDR offers a higher level of integration of their products at deployment, with a focus on threat detection and incident response use cases. Moreover, while a SIEM can be be delivered in a Software as a Service (SaaS) model, most XDR products are developed using new cloud-native architectures, making them an emerging alternative or complement to existing SIEM tools. Despite such advantages, some of the SIEM use cases, such as generic log storage or compliance, are not replaced by XDR solutions [56].

The combination of Elasticsearch, Logstash, and Kibana from Elastic Stack, OpenSOC, Apache Metron, and other tools leveraged with or natively using Big Data platforms like Hadoop offers data collection, management, and analytics capabilities. Some security analytics platforms are available [62, 9, 63, 64, 65, 66], and many open-source solutions have been developed supporting a wide spectrum of security-based analysis [67, 68]. OpenSOC, for instance, was one of the open-source platforms incorporating scalable security analysis tools and providing, in many cases, an alternative to the expensive commercial SIEM-frameworks. It provides real-time security analysis and data analytics. The OpenSOC framework also integrates a great part of the Apache stack, such as Hadoop [69], Kibana [70] and Elasticsearch [71] to store, index, and enrich data sources, including network traffic and application log data. Apache Metron [68] is another example of those platforms, and the successor of OpenSOC. It also provides a full-stack software infrastructure for the analysis and detection of network intrusions, zero-day attacks, and advanced persistent threats. IBM QRadar is another security platform able to scale up in terms of performance and storage. It is designed to monitor, correlate and store large volumes of data. It includes searching capabilities over the indexed data and also provides key capabilities such as risk management, vulnerability

management, incident forensics, incident response, and application. It also includes incident forensics to enable visibility to the questions who, what, when, where, and how a security incident occurred [72].

There are examples of security platforms specifically designed for CIP, as it is the case for the platform proposed by Gonzalez-Granadillo et al. [73], where processes' events are received from multiple sources affecting a water CI to be correlated to generate security alarms accordingly, indicating the presence of a threat or an attack in the monitored systems. Another example is [52, 4], which presents an hierarchical two-level correlation architecture for electricity grids, which later evolved into a Big Data solution, presented in [3].

F. SUMMARY

This section was not intended to provide an exhaustive overview of the field, but rather to provide an encompassing perspective about the specifics of the CIP and IACS domains from a security standpoint, providing the reader with broad knowledge about the problems, limitations and the solutions being used by CI operators. These concepts are key for understanding the next two sections, which will be devoted to discussing the functions and role of forensics and compliance auditing capabilities.

III. FORENSICS

Forensics refers to the application of science and technology to an investigation process to find out the facts in criminal or civil litigation. It comprises collecting evidence of the occurred facts, records and digital trails that can be legally used for criminal prosecution [74]. Based on this data, backward tracing can be used to reconstruct the chain of events that led to an incident, with forward tracing helping understand the repercussions of that event. Moreover, such procedures are often undertaken for reasons other than legal, such as root cause analysis of system failures or incorrect procedures, based on operational traces.

This section will start with the definition of what a Forensic Process is, followed by a description of the associated investigation processes and a definition of digital and network forensics. Next, a brief survey on digital forensics is provided, followed by a discussion of the impact of cloud computing on forensics processes, data privacy aspects, forensics readiness. Forensic schemas and interoperability formats are also discussed, together with query and visualization tools. This section closes with an overview of CPS forensics and the impact of Internet of Things (IoT) and Industrial Internet of Things (IIoT) on the forensics domain.

A. WHAT IS A FORENSIC PROCESS?

Overall, the definition of what constitutes a forensic process is mostly coherent across different literature, regulatory and/or standardisation sources. For instance, Rani and Geethakumari [75] defined computer forensics as the science allowing to identify, extract preserve, and describe the digital evidence stored in digital devices and networks that can be legally admissible in court for any cyber-crime or fraudulent act. The National Institute of Standards and Technology (NIST) [76] defined digital forensics or computer forensics as a scientific method to identify, collect, examine and analyze data, also comprising a systematic investigation process of crimes in which evidence can be retrieved from the media contents found in the associated digital device. Casey [6] defined digital forensic investigation as a complex and time-consuming activity in response to a cybersecurity incident or cybercrime that should answer these questions: what happened, when, where, how, and who is responsible.

Attacks against ICS and SCADA systems, such as Stuxnet [44], Dragonfly [77] or Flame [78], highlighted the relevance of forensic investigations for post-mortem analysis. In many cases, this has prompted operators to design and implement defense and forensic readiness strategies, encompassing actions and procedures to provide the capabilities to diagnose incidents and support the identification and prosecution of attackers. Such capabilities can also be helpful to deal with harmful events such as natural disasters or hardware malfunctions, by providing the capabilities to analyse the underlying SCADA Information Technology (IT) system [13]. These approaches gain more significance as breaches in SCADA systems may cause dangerous consequences for both human life and the infrastructure, beyond significant monetary loss or service disruption [79, 80, 81].

While most cybersecurity tools are focused on detecting and monitoring, forensic tools are focused on collecting and recording traffic and events while, at the same time, providing feedback information to the security actors. Relevant operational events are monitored and recorded using a forensic approach akin to a system black box, providing the means to investigate and retrieve evidence. Also, it should be possible to trace the attack, prepare mitigation actions, adjust countermeasures, apply damage control policies or even recover from partial or total failure.

B. THE FORENSIC INVESTIGATION PROCESS

According to Hunt [59], the main purpose of intrusion analysis and collection of forensically sound data is to seek answers to the following questions:

- Who is responsible for the incoming intrusion or outgoing data transfer?
- What kind of equipment and services were involved?
- Were they able to do this because of limitations of incoming or outgoing security mechanisms?

As illustrated in Figure 2, according to authors such as Whitman and Mattord [54], the forensic investigation process follows the basic methodology:

1) Preparation, including the identification of the relevant items bringing value to evidence.
2) Acquisition of evidence with preservation, without alteration or damage.
3) Assure at every step the evidence is verifiably authentic and remains unchanged since the time it was seized.

4) Evidence examination and analysis of the data without risking modification or unauthorized access.
5) Report the findings to the proper authority and take the lessons learned.

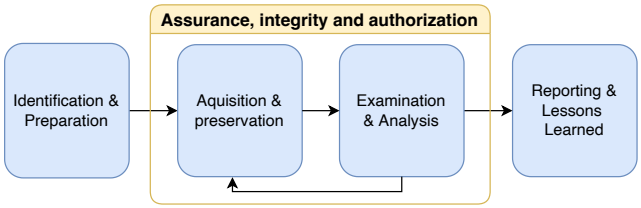

FIGURE 2: Forensic Investigation Process.

In this context, evidence may refer to a physical object or documented information about a past action that may help disclose the intent of a perpetrator [54], support an alibi [6] or provide legally admissible proof. It should be checked whether it was obtained legally as a result of a court order or by another order of an authorized institution or person.

The forensic investigation process should be able to capture evidence before processes or services on the running system overwrite useful volatile data [13]. This may be justified for a wide array for scenarios such as disputed transactions, allegations of employee misconduct, presenting legal and regulatory compliance, negligence and breach-of-contract charge avoidance, assisting law enforcement investigations, meeting disclosure requirements in civil claims, or supporting insurance claims when a loss occurs.

Digital evidence comprises the data stored or transmitted using computing means, which may be used for incident analysis and/or proof purposes. In the course of a forensic investigation, it should be assured that all available digital evidence is not only protected from deletion but also from modification without appropriate authorization [82], with all steps being recorded [83]. This is vital for integrity purposes, also protecting data from anti-forensics activities, which comprise the techniques aiming at hampering the forensics process, destroying or modifying any digital evidence [84].

For forensics applications, digital evidence integrity is a key property as its violation invalidates the admissibility of data for proof purposes. A cryptographic hash can be used to assess the integrity of the evidence, as well as the copies used along with the examinations and analysis results of compromised systems — this way, an examiner can rely on data he is working on, confident is exactly the one originally captured. A hash can be computed in the moment data is produced and used until the moment integrity is checked, allowing to detect abnormal situations, for instance, when an inconsistent data image does not accurately represent the state of the data acquisition [13].

Data provenance, which provides contextual information related to the origin of data, can support detailed explanations on how a specific state was reached, being included in evidence as a statement from the person carrying out the extraction. It specifies the source system, the acquired artifacts to denote the chain of custody as an audit trail of all activities, and a timestamp of data extraction [85]. Several approaches have been proposed to implement provenance tracking (e.g., ES3 [86], PASS [87], SPADE [88], Story Book [89], TREC [88]). Still in this scope, Zafar et al. [90] proposed a taxonomy of existing secure provenance schemes.

Data provenance analysis can be used to extract host events into provenance graphs that represent the entire system execution and help causal analysis of system actions. Some of the recent works focused on fidelity [91, 92, 93, 85, 94, 95], while others focused instead on efficiency [96, 95, 97, 98, 83, 99, 100, 101]. Data provenance can also help reducing alert fatigue [102] and identifying intrusions [103, 104, 105].

The highly volatile nature of digital evidence implies that a careful integrity safeguarding approach should be followed. This chain of custody process intends to help preserving the integrity of the information, providing a non-repudiable chronological trace [106] detailing how evidence was acquired, processed/analyzed, handled, stored, and protected, to be presented as admissible evidence in court [107]. A chain of custody ensures the collected evidence is not modified along the investigation process and from the moment it was collected until it is presented [108]. Prayudi and Sn [109] provided an overview of the state of the art about challenges in the digital chain of custody. Cosic and Baca [110] presented a digital evidence management framework aiming to improve the chain of custody of digital evidence in all stages of the digital investigation process, supported by the use of SHA-2 hash function for the digital fingerprint of evidence.

C. DIGITAL AND NETWORK FORENSICS

In 2008, the American Academy of Forensic Sciences (AAFS), one of the most widely recognized professional organizations for all established forensic disciplines, recognized forensic computer-related crime investigation as a legitimate area, for which a new Digital and Multimedia Sciences section was allocated [6]. This enabled the development of a common ground for the forensic science community to share knowledge and address current challenges [111].

Digital forensics deal with evidence extraction, preservation, identification, documentation, and analysis using well-defined law enforcement procedures, establishing clear lines within the chain of custody. According to McKemmish [112], digital forensics can be broadly considered as having four stages, namely: identification, preservation, analysis, and presentation. Several methods have been proposed in the literature, aiming to formally reconstruct the sequence of events executed during the incident using proven methods [113]. However, the significant growth in the volume of data and the number of evidence items coming from a wide range of sources raises new challenges when conducting digital forensic investigations.

Imaging, hashing, and carving are among the available techniques used by digital forensics investigations. Imaging consists of copying storage media to be examined as evidence. Such evidence can be compromised by modern Operating Systems (OS), due to the operations in the background on the file system, such as indexing or journal resolution [114].

Cryptographic hashing or signing is used to provide authenticity and integrity of files and other evidence. For instance, Afzaal et al. [115] presented an architecture aiming to overcome the limitations of the classic RSA algorithm to provide event integrity protection, allowing a group of n parties to participate in the digital signature process to enforce authenticity and non-repudiation. As for hashing techniques, while MD5 hashing was originally adopted by the forensics community [116], it was later superseeded by SHA-1 as a NIST federal standard, with a transition timeline towards SHA-2 or SHA-3 being announced in December 2022.

Carving refers to the forensic tools to scan unused disk blocks to find and recover deleted data. Carving uses known header and footer signatures to combine the non-used nodes into the original deleted files. Mikus [117] conducted an analysis supported by the use of carving techniques. Recent advances in carving included recovering capabilities of fragmented files with more accuracy [118].

Within the digital forensics field, network forensics is concerned with monitoring network traffic to assess anomalies and attacks. To investigate such attacks, several data sources are available, including packet filters, firewalls, intrusion detection systems, honeypots, sinkholes, surveillance and vulnerability scanning systems [59]. Software Defined Networking (SDN) was also leveraged by Bates et al. [119] to deploy capture points over the network to have a holistic view of network activity, which can be used for forensics purposes. [120] also discusses the challenges of executing network forensics investigations in virtual networking environments with tunneling and SDN. Nevertheless, one of the most important challenges in terms of network forensic has to do with the required data storage and computing capabilities [121]. For instance, even a moving window of some hours covering the duration of relevant real-time traffic may require a significant amount of storage from a computing cluster, something that may be aggravated in case of sustained attacks

D. A BRIEF SURVEY ON DIGITAL FORENSICS

There is a considerable corpus of related literature on digital forensics, whose focus is equally diverse. In this line, Casino et al. [122] reviewed several works in the field of digital forensics and identified their main topics and challenges.

Regarding methodological aspects, Sommer [123] raised awareness of the challenges involved in gathering, analyzing, and presenting digital evidence among directors, managers, and their professional advisers, with Williams [124] providing direction to those who assist in the investigation of cyber security incidents and crimes, not just for law enforcement. van Baar et al. [125] reported benefits and performance on processing digital forensic investigations on a particular case involving collaboration between different actors.

Regarding the subject of digital forensics frameworks and other architectural developments, Verma et al. [126] proposed a digital forensic framework that uses case information, case profile data and expert knowledge for automation of the digital forensic analysis process supported by Machine Learning (ML) for finding evidence. Hunt and Slay [59] advocate the need of a new forensic analysis approach requiring the implementation of forensic engines, supported by parallel processing while providing flexibility on customizing activities for the analysis of evidential data. Ahmadi-Assalemi et al. [127] presented a federated Blockchain model that achieves forensic-readiness by establishing a digital Chain-of-Custody and a collaborative environment to qualify as digital witness for post-incident investigations.

Specifically on the CIP scope, Ahmed et al. [128] highlighted that forensic analysis for ICS is still in its early development stages, due to its specialized nature, together with the prevalence of proprietary and poorly documented protocols. Nevertheless, Kilpatrick et al. [129, 130] and Chandia et al. [131] proposed an architecture allowing to capture and analyse sensor data and control actions in a SCADA network (using agents located at strategic positions within the network to capture the local traffic and forward a relevant portion of packets, called a synopsis, to a data lake). Also, Elhoseny et al. [132] proposed a conceptual framework for automated and secure forensic investigation in modern complex SCADA networks, intentionally designed to comply with green computing requirements. Eden et al. [17] suggested deploying forensic hardware instrumentation connected to field device artefacts as a wrapper implemented at physical level, in order to improve the availability and recovery of information for cases where SCADA devices have restricted physical access. Valli [133] created a framework that produces forensically verified signatures for the Snort IDS for known and published vulnerabilities of SCADA, enabling investigators to trace exploits during analysis.

There are also many works focused on identifying existing gaps and/or challenges, some which also proposing suitable solutions to address them. For instance, Huang [134] realized that the characteristics of big data complexity (e.g., volume and variety) make traditional data mining algorithms unsuitable to retrieve knowledge in forensics scenarios, something that Quick and Choo [135] also address, highlighting the challenges posed to digital forensic analysis (considering the ongoing growth in the volume of data seized and presented for analysis). These conclusions are reinforced by Koven et al. [136], who noticed a lack of suitable analysis tools for large datasets – despite the focus on email datasets, the findings are likely to be broadly

applicable to other types of sources. Stelly and Roussev [137] presented the concept and prototype implementation of the first domain-specific language aimed at providing a practical and formal description of digital forensic investigations as a computation.

Regarding causality analysis for attack investigation, several works have considered provenance graphs for tracking based on audit logs. Their approach is mainly related with the sub-topics based on causality, anomalies and learning analysis. As an example, Zipperle et al. [138] surveyed the literature on provenance-based IDS and proposed a taxonomy. Alsaheel et al. [139] proposed a framework to identify and reconstruct end-to-end cyber attack stories from unmodified systems and software audit logs. Kwon et al. [140] developed a model supported by causality-based inference for audit logging. Ma et al. [141] also proposed a provenance tracing system capable of alternating between logging and unit-level taint propagation, and event processing.

Considering digital forensics performance and assessment, Ayers [142] proposed several metrics for measuring the efficacy and performance of forensic tools, such as speed, accuracy, completeness, reliability, and auditability. Roussev and Richard [143] discussed the need of distributed forensics approaches, highlighting the performance benefits inherent to distributed computing and proposing a distributed digital forensic tool to centralize data and distribute processing over multiple devices, with background preprocessing capabilities of multiple concurrent searches. Daubner et al. [144] presented research towards verification of forensic readiness in software development, with a focus on produced digital evidence.

The topic of anti-forensics techniques and prevention is also addressed in the literature and has been the subject of research [145]. As an example, Rekhis and Boudriga [113] developed and demonstrated an anti-forensics aware theoretical digital investigation approach, with Noura et al. [146] proposing a solution to prevent anti-forensics techniques targeting log availability and integrity (such as wiping and injection attacks), using encryption, fragmentation and authentication for data distribution across several storage nodes.

E. CLOUD FORENSICS

The cloud computing paradigm, which shifts information from endpoint devices to a provider infrastructure [147], has become popular among many organizations due the potential cost and resource efficiencies it might entail, also offering several operational benefits for CI, including data redundancy, data availability and survivability when essential system components are isolated or lost [148]. Its introduction raises new and substantially different challenges for forensics, since the target environment is no longer isolated and data is no longer acquired under the investigator's control – thus, there is an evident need to go beyond traditional approaches [149]. In this scope, NIST identified 65 challenges of conducting digital investigation in cloud environments, also pinpointing existing technical gaps [150].

Forensics activities in the cloud present important challenges. Aspects such as the distribution of computing and storage (which have an impact in terms of increased attack surface), geographical storage dispersion across distinct jurisdictions with specific procedures and laws, privacy, or even the lack of norms on aspects such as Service Level Agreement (SLA) regulating the client and Cloud Service Provider (CSP), raise new complex challenges to forensics investigators [151]. Even if the CSP is compliant with the law enforcement agencies in its respective jurisdictions, cloud forensics may be a costly and time-consuming procedure [152], moreover considering how much storage may be used on the tenant storage pool.

The increased need for forensic investigations involving cloud-based scenarios has prompted the emergence of cloud forensics [153], a hybrid approach encompassing remote, virtual, network, live, and large-scale operations, geared towards the generation of digital evidence from cloud environments.

Moreover, cloud-based forensic architectures (which may still be used with private clouds) can be seen as an online solution to help remove any hardware dependency [154] [155]. This can enhance forensic experts and investigators activities with the tools and processes to be applied in the digital investigation of the collected evidence such as sorting, indexing, data recovery or bookmarking, among others. In this line, van Beek et al. [156] shared the lessons learned from providing Digital Forensics as a Service (DFaaS) implementations for almost 10 years, discussing the organizational, operational and development perspective, in a forensic and legal context. Zawoad et al. [157] presented an architecture for a secure cloud logging service, collecting information from different sources around the datacenter, both software (hypervisors) and hardware (network equipment), in order to create a complete landscape of the operations in a datacenter. Similarly, Zawoad et al. [158] proposed a Secure-Logging-as-a-Service (SecLaaS) to enhance forensic investigation in the cloud ecosystem that enables the acquisition of admissible log evidence in the cloud.

A literature review revealed several cloud forensics framework proposals. Manral et al. [159] surveyed the cloud forensic literature published between January 2007 and December 2018, categorized using a five-step forensic investigation process, and included a taxonomy of existing cloud forensic solutions as well. Ruan and Carthy [160] described the need for new forensic tools or to extend the existing digital forensic tools to make them fit into Cloud frameworks, also presenting a forensic tool for OpenStack Cloud which works through a daemon running in a compute node delivering network logs and the images of instances to the dashboard. Rani and Geethakumari [75] describe a snapshot-based approach to face the dynamic nature of

Cloud in which the CSP takes a snapshot of a suspected Virtual Machine (VM) when an anomaly is found by an IDS, isolating it from the network and storing it in permanent storage. A similar approach was suggested by Hibshi et al. [161], which presents a study highlighting a number of usability points that need to be taken into consideration when designing and implementing digital forensics tools, also proposing an efficient approach to forensic investigation in the cloud using VM snapshots. Yu et al. [162] presented a framework for automated detection of anomalies in a cloud environment including a module for cloud forensics with learning capabilities embedded in the management layer of the cloud infrastructure. Patrascu and Patriciu [163] claimed there should be a revision of the classic network forensic principles, and a reorganization of well-known workflows, taking in consideration tools such as ML or large scale computing.

A hypervisor-based approach has been considered for threat monitoring and forensic analysis in [164], where the hypervisor provides the means for examining VMs, by monitoring activities performed at a layer between the hardware and the virtual environment. The potential of this approach was demonstrated by Mishra et al. [165], which presented a taxonomy of hypervisor forensic tools and demonstrated how evidence that can be found in a VM, at the hypervisor and host system layers. Saibharath and Geethakumari [166] proposed a remote forensic evidence collection and pre-processing framework for cloud nodes that collects VM disk images, logs and network captures, pushed periodically into a Hadoop distributed file system. Huseinović and Ribić [167] evaluated the virtual machine memory dumps from Oracle VirtualBox and VMware VMs, with Cheng et al. [168] proposing a similar concept for a lightweight live memory forensic framework based on hardware virtualization that can build a virtualization environment on-the-fly. Also, Zhang et al. [169] and Guangqi et al. [170] proposed a KVM-based approach to acquire both data and VM meta-data, using the access and control privileges of a VM host to acquire VM-related information.

In alternative to VMs, the combination of containers and microservices can help improving isolation between components in an cloud-bative application, with a reduced overhead. However, the topic of forensic investigation in containerized environments is a complex task raising new challenges [120], due to the fact that instances can be started and stopped on different systems, which results in an ongoing change in the structure of the network, as well as their shorter life span which implies that container instances may not be available anymore when a investigation process is triggered. Which such environments in mind, Sharma et al. [171] presented a deep learning approach for containerized application runtime stability analysis, and an intelligent publishing algorithm that can dynamically adjust the depth of process-level forensics published to a backend incident analysis repository. Stelly and Roussev [172] presented a scalable containerized framework for forensic computations.

Other works have proposed procedures and standards for forensics activities in the cloud. Saibharath and Geethakumari [173] developed a framework for cloud forensics in OpenStack, according to the Infrastructure-as-a-Service model and using existing forensic tools, which is able to take live snapshots, image evidence, packet captures and log evidence.

Banas [174] discussed the memory acquisition process to be placed in a kernel based virtual machine (KVM) storage and memory images in OpenStack without any CSP interaction in a self-service Cloud environment. The NIST Cloud Computing Forensic Science Working Group (NCC FSWG) [175] was established to research on Cloud forensic science challenges in the Cloud environment and to develop plans for measurements, standards and technology research to mitigate the challenges that cannot be handled with current technology and methods. Almulla et al. [152] proposed a forensic procedure based on the NIST model to examine private cloud VM snapshots, using existing digital forensic tools, being able to successfully acquire data without the need to transform the snapshot files.

There is also an emerging line of work regarding the use of Blockchain for FCA purposes, providing a tamper-resistant ledger mechanism which matches the needs for non-reputiation and chain of custody purposes. In this scope, Liang et al. [176] proposed a decentralized and trusted cloud data provenance architecture using blockchain technology. Also, Awuson-David et al. [177] and Ahmadi-Assalemi et al. [127] presented Blockchain-enabled methodologies and frameworks for keeping a chain of custody of the digital forensic log evidence from the cloud ecosystem, to ensure trustworthiness, integrity, authenticity and non-repudiation. Finally, [178] proposed a cloud forensics taxonomy and denoted the trend towards the implementation of digital provenance assurance using blockchain technology.

F. DATA PRIVACY PROTECTION IN DIGITAL FORENSICS

Privacy can be defined as the right to control who has information about someone, including activity tracking [179]. Some of the concepts raised in privacy laws intend to establish limits restricting data use or its correlation from multiple sources, often mandating anonymization or removal of personal data from records [179]. Such an example is the General Data Protection Regulation (GDPR), introduced in 2016 to bring protection to personal data [180], making it mandatory to obtain consent on the use of personal data.

The problem of balancing forensic investigation needs with privacy protection requirements is discussed by Aminnezhad et al. [181], with Dehghantanha and Franke [182] having established the foundations for the definition of privacy-respecting digital investigation as a new cross-disciplinary field of research, also reviewing the state of art in this field. Despite the large number of digital forensic models discussed in scientific literature, just a few of them are considering data privacy along the digital forensic

investigation process, many of which are either tailored for specific environments or included as an independent module [126].

van Staden [183] proposed a framework to protect privacy in multi-user environments that are subject to post-incident forensics investigation, supported by profiling and filtering mechanisms. Law et al. [184] described a way to protect data privacy using encryption, proposing the introduction of simultaneous data encryption processes by email servers and indexing of related keywords, allowing an investigator to give a keyword input to the server owner, who has the encryption keys, to get back the emails that contain the keyword. Also regarding encryption-based approaches, Hou et al. [185] proposed a mechanism to protect data privacy on a third-party service provider's storage center, using homomorphic and commutative encryption, with Hou et al. [186] describing a similar solution.

As for identity or knowledge-based approaches, Shebaro and Crandall [187] used an identity-based encryption mechanism to carry out a network traffic data investigation in privacy preserving setting. Croft and Olivier [188] proposed a mechanism where data is divided into layers of sensitivity, placing less private data on lower layers, and highly private data on higher layers. In this schema, access to private information is controlled by initially restricting investigator access to the lower layers, requiring further proof to get access to higher-level information.

G. DIGITAL FORENSIC READINESS AND FORENSICS-BY-DESIGN

For many, the possibility of a security incident should be regarded as a certainty rather than a possibility [189]. In fact, when incidents happen, the priority is often restoring normal operational levels, instead of making an effort to collect and preserve as much forensics evidence as possible, eventually to be admitted to a court. The generalised approach is mostly reactive: first restore operational capacity, and then carry out investigations and seek evidence. As a result, evidence might be lost or rendered unsuitable as proof.

Forensic readiness is a concept that contributes to minimise the aforementioned problems. It suggests taking proactive actions to capture evidence even before or during an incident and before investigations are started. This helps not only to save time and money, but also to mitigate potential incidents and ensure business continuity and compliance with minimal disruption and interruption of operations. Kruger and Venter [190] provided a systematic literature review to identify topics where digital forensic readiness is included. However, as denoted by Iqbal et al. [191], digital forensic readiness for CIP is still immature, judging by the lack of published research or industry reports.

Forensic readiness comprises planning activities to collect, preserve, protect and analyze digital evidence to be effectively used [192]. It can also assist in fulfilling the increasing demand for the implementation of security practices addressing compliance with organizational policies and regulatory requirements, providing the means to deploy continuous monitoring and review processes supported by the already collected forensic data. This approach can help fill that gap, since even common standards such as the ISO 9001 series and regulatory frameworks for B2B relationships (e.g. supply chain risk management) do not account for best practices in the CI and IACS security domains.

Forensic-by-design extends the concept of Digital Forensic Readiness. Similarly to Security-by-design, it advocates the integration of forensic requirements into the system's design and development stages. Ab Rahman et al. [189] proposes a system and software engineering driven Forensic-by-design framework, with an emphasis on Cloud computing systems. Akilal and Kechadi [193] investigated the potential adoption of Forensic-by-design in cloud computing systems, with [194, 195] suggesting the application of Forensic-by-design (FbD) strategy to enhance digital forensic readiness.

Moreover, several proposals for implementing digital forensics readiness are documented in the literature. For instance, Daubner and Matulevičius [196] proposed the introduction of forensic readiness mechanisms within security risk management to refine specific requirements on forensic-ready software systems, by re-evaluating the taken security risk decisions with the aim of providing trustable data when the security measures fail. Elyas et al. [197] presented a digital forensic readiness framework through a series of expert focus groups to discuss the critical issues facing practitioners in achieving digital forensic readiness. Also, De Marco et al. [198] proposed a reference architecture for a Cloud forensic readiness system. Mouhtaropoulos et al. [199] classified forensic investigation frameworks to expose gaps in proactive forensics research and reviewed prominent information security incidents with regard to proactive forensics planning. On a more network-focused scope, Endicott-Popovsky et al. [200] proposed a framework for operationalizing network forensic readiness, with Ngobeni et al. [201] proposing a wireless forensic readiness model designed to help monitor, log, and preserve wireless network traffic for digital forensic investigations.

Considering readiness maturity assessment, Ariffin and Ahmad [202] presented five indicators for the maturity and readiness of digital forensics, with Elyas et al. [203] describing an approach to identify the factors that contribute to digital forensic readiness and how these factors work together to achieve forensic readiness in an organization. Iqbal et al. [204] presented a study on the current support for forensic readiness of CI, highlighting the involved key challenges and providing a literature review on the subject. Also, Alenezi et al. [205] presented a framework to investigate the factors that facilitate the forensic readiness of organizations.

H. FORENSIC SCHEMAS AND INTEROPERABILITY

In a general way, interoperability is concerned with making it possible for components or systems coming from different vendors to easily communicate and interact with each other.

When investigation processes require evidence exchange between investigators, the use of different tools for the reconstruction of events or analytical purposes, the absence of standardised digital evidence formats can become a serious obstacle. Thus, it is particularly important to develop information interoperability mechanisms by means of common Forensic Schemas.

A standardized approach for representing and sharing digital forensic information is also useful to help investigators collaborate when incidents involve different jurisdictions. Similar challenges were also recognized in traditional investigations of violent crime and led to the development of the US Federal Bureau of Investigation's Violent Criminal Apprehension Program (ViCAP) and Royal Canadian Mounted Police's Violent Crime Linkage System (ViCLAS) programs. These programs enabled the correlation of all the available information from unsolved violent crimes in disparate regions, trying to find links between them.

There have been several schemas proposed in the past for representing digital forensic information, but these have not been widely adopted [206] [207] [208] [209] [210]. Also, Garfinkel [211] proposed a XML schema (DFXML) for easier interoperability between forensic extraction and visualization tools, primarily developed to represent the output from tools used to analyze storage media, including file system parsers, file carvers, and hash set generators.

Casey et al. [212] conducted a review of digital forensic data schemas, including DFXML, also proposing the CybOX schema for handling forensic data. CybOX is an open-source, community-driven effort to develop a standardized representation of digital observations led by the US Department of Homeland Security (DHS) office of cyber-security and communications.

The XML-based XIRAF system was created by the Netherlands Forensic Institute (NFI) to support digital forensic analysis, storing its data using a parent–child structure within a centralized database accepting structured output and searching tools [213]. Bhoedjang et al. [214] described the second generation of this analysis system and outlined the complexity of importing different file types and analyzing and preprocessing files before storing them in databases. van Baar et al. [125] outlined the latest iteration of this system, which incorporates Cloud features.

The Advanced Forensic Format (AFF4) has taken another approach for the representation of digital forensic information [215] [216], using the Resource Description Framework (RDF), a general purpose representational formalism for knowledge representation. Although the majority of digital forensic tools do not support AAF4, Google Rapid Response (GRR) uses the AFF4 data model to store information in a MongoDB database [217]. The AFF4 data model is flexible. However, the use of RDF requires the adoption of a shared supporting ontology. While there is still no community consensus on such ontology to exchange digital forensic information, the ontology proposed by Casey et al. [212] could be used as a basis for such consensus.

I. VISUALIZATION AND SEARCHING TOOLS

When it comes to the forensics practitioner toolset, usability is a crucial aspect not to be disregarded [161]. Specifically, Osborne and Turnbull [218] pinpointed the importance and need for tools incorporating adequate visualization capabilities for digital forensic data, claiming that there is a lack of algorithms to identify relationships, normalize data, incorporate multiple data sources, and provide effective visualization methods, all of which are important to retrieve further insights from evidence. Following this same line of thought, Osborne et al. [219] highlight the importance of considering architectures incorporating familiar visualization tools and algorithms that could be able to include distinct data sources, normalizing and correlating data, later proposing a conceptual framework able to Explore, Investigate, and Correlate (EIC) [218]. Tassone et al. [220] also highlighted the importance of visualization in forensic tools, pointing out that many existing solutions where just simple layouts to search and display basic tabular data, also presenting a proof of concept including a database schema designed for third-party forensic data storage and visualization.

Irfan et al. [221] describes a virtual cloud environment incorporating visualization capabilities designed to provide visibility for all security events, allowing to follow activities of cybercriminals, reproduce crude information identifying each respective incident, and execute proactive actions. Also, Aupetit et al. [222] presented a methodology and a tool for allowing the Internet Service Provider (ISP) to assess and visualize threats from an organization's network traffic, allowing them to deal for instance with Distributed Reflective Denial of Service (DRDoS) events. Another example is provided by Setayeshfar et al. [223], which presents a graphical forensic analysis system for efficient loading, storing, processing, querying, and displaying of causal relations extracted from system events to support computer forensics.

Tools such as the Elastic Stack have been widely adopted in industry and academia as a result of their capabilities and performance in terms of log handling. There are solutions for data visualization, including graph generation capabilities for analysis purposes, supported by frameworks such as Kibana [70], Grafana [224], and Prometheus (Prometheus.io), which retrieve data stored in indexed datastores like Elasticsearch [71]. Some of the tools built on ElasticStack are SOF ELK [225] and Plaso [226], that provide rich visualization and parsing capabilities. Despite their capacity for effective forensics and provenance tracking supported by queries, they lack information about the provenance models also don't provide users with many query abilities beyond filtering. Moreover, it should be stressed that while these tools can be used for multiple use cases without the incorporation of analytic inference mechanisms, that's not typically the case in cyber-security analytics or forensics [222].

J. FORENSICS CONSTRAINTS FOR THE CIP DOMAIN

Homem [227] identified a series of general challenges regarding digital forensics processes, namely: the rising volume of heterogeneous digital evidence involved in investigations, the evidence-centricity of industry-standard tools, a deficiency in the availability of a highly-skilled workforce, and the great effort required by the largely manual and time-consuming activities involved in the overall process. Besides, CI operational environments add further constraints related to aspects such as complexity, systems interdependency, dependency on ICT and components provided by third parties, or the deployment of heterogeneous technologies [228].

Typically, forensic investigation can rely on live or dead evidence aquisition. While the latter is performed offline on static data after a system is shutdown, the former collects data from live systems, such as the contents of physical volatile memory, and non-volatile data, such as the data maintained in a storage system. While dead forensics corresponds to the most traditional approach, there was a increasing emphasis on live forensics processes over the past years, as it is the case for network traffic analysis. More specifically, in the case of SCADA systems, the forensic investigator cannot turn it off to capture and analyze data, because this kind of system is supposed to be continuously operational [229] – in such cases, live forensics is a suitable digital investigation methodology [230]. However, since continuous availability of SCADA systems is a mandatory requirement, forensic investigators should strive to be minimally intrusive, in order to reduce the risks in critical operations while aiming at a rapid response time, to preserve evidence that may be overwritten by runtime processes [231].

It is known that SCADA and IT systems exhibit different behaviours and possess different characteristics, often requiring for IDS and other security mechanisms to be configured according to with the domain of operation [4]. For instance, in a SCADA system, network traffic is more deterministic than in IT networks, in the sense that a system component communicates to other system components following established patterns, frequently with bounded time restrictions. Thus, administrators may impose a set of rules for security purposes, with any non-deterministic behavior flagged as an anomaly – e.g., an IDS might be configured to consider a specific communication pattern as normal [232].

Moreover, the same restrictions regarding live network trace capture can also apply to SCADA stations and other process control or monitoring systems. Any evidence collection tool or technique must avoid imposing overheads that might degrade the system response, interfere with operational indicators or expand the vulnerable attack surface. Overall, a simple rule must be kept in mind: live (or, for that matter, any other) forensics processes must be designed to adhere to the least overhead principle, in line with the recommendations from standards such as NIST SP800-82 [233], which clearly identify the risks associated with intrusive security procedures.

K. IOT AND INDUSTRIAL IOT FORENSICS

IoT can be defined as a system of networked smart devices that can be identified, named and addressed [234]. IoT is attracting great attention not only for consumer applications but also in the IACS domain, where they are usually designated as Industrial IoT. Naturally, the introduction of these technologies has increased the amount of generated, transported and processed data, as well as the number of forensically relevant events in consequence of the increasing number of available sensor devices [235].

Considering the emergence of IIoT, organisms such as NIST have defined guidelines [236, 237] to ensure that IIoT infrastructures rely on adequate safety, security, privacy, consistency, dependability, resiliency, reliability, interaction and coordination measures. However, it's not always possible to apply traditional information security measures based on sophisticated encryption algorithms, multi-factor authentication, antivirus programs and firewalls (among others), due to the limited computational and energy resources of some sensor nodes [238], further reinforcing the need for the deployment of proper security monitoring and forensics capabilities.

Stoyanova et al. [239] identified and discussed the main issues involved in the process of IoT-based investigations, particularly all legal, privacy and Cloud security challenges. They also provided an overview of the past and current theoretical models in the digital forensics and frameworks aiming to extract data in a privacy-preserving manner or secure the evidence integrity using decentralized blockchain-based solutions. Vendors such as Infineon, NXP, and STMicroelectronics prepared a position paper for ENISA [240], stating the IoT market failure for cyber-security and privacy, and claiming that there were "no level zero defined for the security and privacy of connected and smart devices," no legal guidelines for IoT device and service trust, and no "precautionary requirements are in place". This paper also predicts that attacks will get more risky and threatening due to the rise of IoT enabled cars, CI, and health applications. In the same line of thought, Chehri et al. [241] identified the trends, problems, and challenges of cybersecurity in smart grid CI in Big Data and Artificial Intelligence (AI).

(I)IoT scenarios require the implementation of adequate forensic and compliance auditing approaches to improve security and privacy. In that regard, Yaqoob et al. [242] investigated studies on the topic of IoT forensics by analyzing their strengths and weaknesses. The authors categorize and classify the literature by devising a taxonomy based on forensics phases, enablers, networks, sources of evidence, investigation modes, forensics models, forensics layers, forensics tools, and forensics data processing. They also enumerate a few prominent use cases of IoT forensics and present the key requirements for enabling IoT forensics, identifying and discussing open research challenges as future research directions.

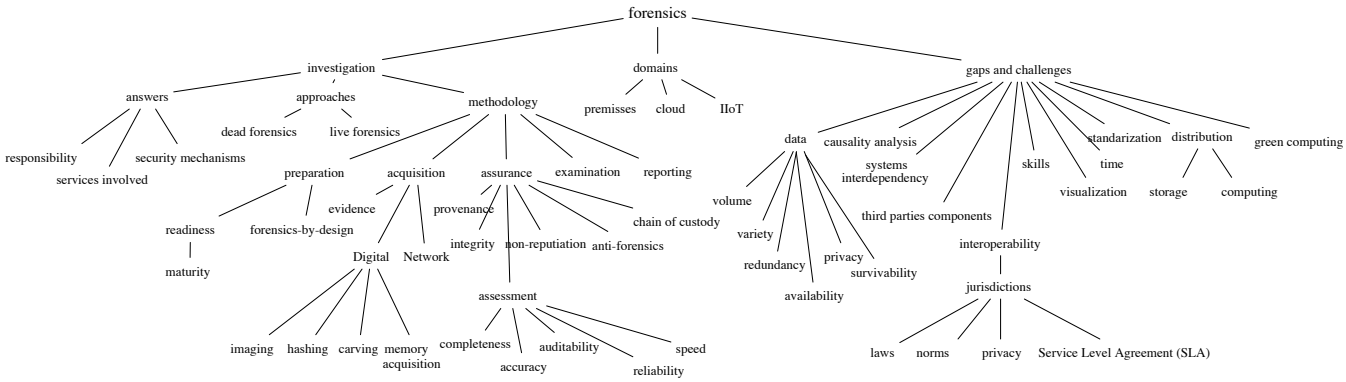

FIGURE 3: Key Forensics Concepts.

L. SUMMARY

The purpose of this section was to introduce and present a series of concepts and topics within the scope of digital forensics, with a view towards its application in the CIP domain. We started by conceptually introducing a definition of forensics activities, followed by a discussion about digital, network and cloud forensics, the latter constituting not only a challenge, but also an opportunity to implement innovative solutions tackling the issues of FCA. The implications of data privacy protection regulations in digital forensics activities were also discussed, followed by a review of the related subtopics of forensic readiness, interoperability, visualization and automation. Finally, we concluded with an overview of the current forensics constraints for the CIP domain. Figure 3 depicts how the key topics covered in this section relate with each other, and Table 1 summarises the reviewed literature on these topics. While some topics are addressed from a more neutral perspective, it must be noted that this is due to the fact that many are still valid in the CIP domain.

Scope	Works
Forensics Definitions	Morioka and Sharbaf [74]
	Rani and Geethakumari [75]
Forensic Investigation Process	Casey [6]
	Whitman and Mattord [54]
Data Provenance	Hassan et al. [91]
	Ma et al. [92]
	Bates et al. [93]
	Bates et al. [85]
	Hossain et al. [94]
	Pasquier et al. [95]
	Lee et al. [96]
	Pasquier et al. [95]
	Hassan et al. [97]
	Ma et al. [98]
	Hossain et al. [83]
	Tang et al. [99]
	Liu et al. [100]
	Xu et al. [101]
	Hassan et al. [102]
	Han et al. [103]
	Wang et al. [104]
	Bates and Hassan [105]
Chain of Custody	Giova et al. [106]
	Prayudi and Sn [109]
	Cosic and Baca [110]
	Awuson-David et al. [177]
Forensics for Smart Grid	Abdullah et al. [47]
Forensic Analysis of Intrusions	Hunt and Slay [59]
	Stelly and Roussev [137]
Evidence Definition	Whitman and Mattord [54]
Forensics Readiness	CESG [192]
	Daubner and Matulevičius [196]
	Iqbal et al. [204]
	Elyas et al. [197]
	Ariffin and Ahmad [202]
	De Marco et al. [198]
	Alenezi et al. [205]
	Mouhtaropoulos et al. [199]
	Elyas et al. [203]
	Endicott-Popovsky et al. [200]
	Ngobeni et al. [201]
	Kruger and Venter [190]
Stages of Digital Forensics	McKemmish [112]
Forensic Taxonomy in SCADA	Eden et al. [17]
Forensics for CPS	Cornelius and Fabro [16]
	Valli [133]
Surveys on CPS	Yaacoub et al. [243]
Anomaly Detection	Grubbs [244]
	Gogoi et al. [245]
	Chandola et al. [246]
	Fu et al. [247]
	Ten et al. [248]
	Yu et al. [162]
	Henriques et al. [249]
CIP in Cloud	Alcaraz et al. [148]
Forensic Green Computing	Elhoseny et al. [132]
Forensic Containerized Framework	Stelly and Roussev [172]
Hypervisor Forensics	Cheng et al. [168]
	Jackson et al. [164]
	Saibharath and Geethakumari [166]
	Huseinović and Ribić [167]
	Cheng et al. [168]
	Zhang et al. [169]
	Guangqi et al. [170]
Taxonomy of Hypervisor Forensics	Mishra et al. [165]
Forensics as a Service	van Beek et al. [156]
Data Provenance in Cloud	Liang et al. [176]
Scalable Microservice Forensics	Sharma et al. [171]
SDN and Virtual Network Forensics	Spiekermann and Eggendorfer [120]
	Bates et al. [119]

TABLE 1: Reviewed Works on Forensics for CIP.

IV. COMPLIANCE AUDITING

An audit process represents a systematic, independent, formal, structured, and documented process, usually performed by a certified professional on behalf of stakeholders, aiming to verify if certain criteria match internal policies, external formal standards, and/or legal requirements [250]. Auditing practices help organizations meet such requirements, also providing due diligence, certification, and stakeholder security. Compliance auditing expertise is closely related to and frequently overlaps with forensic processes, since both often share data sources, tools, and techniques.

This section will delve into the topic of Compliance Auditing, with a view towards its applicability in CIP environments. Starting with an overview of the motivation and context, it will next review existing audit models and proposals and standards, concluding with a discussion about logging systems compliance for audit purposes.

A. MOTIVATION AND CONTEXT

Policy definition and enforcement are cornerstones of modern security practices. For instance, Yaacoub et al. [243] describes a series of policies encompassing aspects such as employee screening processes before recruitment, privilege suspension outside working hours, or additional activity monitoring for people in charge of sensitive tasks, which contribute to enhance the security posture of an organization.

Compliance auditing checks whether workflows are compliant with organizational policies and rules – thus, each process or transaction may be checked to confirm whether it followed the applicable rules or policies. In case rules are violated, the auditor analyses relevant data to determine causes and recommends actions to prevent future deviations or non-compliance situations. Compliance audit frameworks can also help highlighting misconfigurations – for example, they can used for monitoring access security levels for individual and group accounts and help with detailed reports measuring the security progress.

The compliance auditing process ends up with a report that includes the conclusions and additional information about requirements that have been met and non-compliance situations (if found). It can also highlight the implications and risks of non-compliance, suggesting corrective actions to prevent future occurrences [251].

As the surrounding environment evolves, infrastructure and service operators are often forced to adapt to an increasingly complex and constantly changing regulatory landscape. Thus, an organization aiming to implement specific regulatory or standardisation measures should depart from the identification of the entities with relevant technical and/or legal jurisdiction over its domain of activity. In this line, the GDPR [180] regulations constitute an example of a mandatory framework for privacy protection, which applies to organizations within the European Union (EU).

Besides generic or sectorial standards, CI-specific regulations may also be imposed by organizations such as North American Electric Reliability Corporation Critical Infrastructure Protection (NERC-CIP) [252], which publishes a set of security guidelines, as it is the case for Electronic Security Perimeters (CIP-005) and System Security Requirements (CIP-007).

B. CYBERSECURITY AUDIT MODELS

Businesses are being increasingly pressured to undergo periodic audits and inspections as part of legal and regulatory compliance certification requirements. While such certifications processes are important to reinforce trust at the B2B and B2C levels, it should not be forgotten that their ultimate aim is to ensure that adequate preventive and reactive security mechanisms are implemented, as well as proper handling of sensitive data. Ultimately, it all comes to the establishment and maintenance of suitable levels of data confidentiality, integrity and availability within an organization, which may vary accordingly to the type of applications, data to be stored or processed (e.g., the case of sensitive healthcare data), or geographical location (e.g., regional requirements for data privacy and protection).

From the industry standpoint, an organization may be required to comply with regulations such as Payment Card Industry Data Security Standard (PCIDSS) [253], Health Insurance Portability and Accountability Act (HIPAA) [254], Federal Information Security Modernization Act (FISMA) [255], GDPR, FedRamp, and SOC2. These are examples of compliance drivers prescribing the application security activities. The Institute of Internal Auditors (IIA) also provides guidance in the form of the International Professional Practices Framework Standard 2420 (Quality of Communications) [256], whose aim is to establish guidelines for objective, clear, concise, constructive, complete and timely reporting.

Three different types of cybersecurity audits were described by Donaldson et al. [257]. The first category corresponds to threat audits targeting cyber threats, aiming to search for evidence in IT environments. The second one evaluates the cybersecurity controls mapped against frameworks, regulatory requirements, standards or a specific cyberthreat. The last one comprises validation assessments against cybersecurity controls measuring their effectiveness against designed and documented requirements.

The assessment of access control policies is one of the aspects typically resorting to formal reasoning mechanisms to verify application control expressed at design time (for instance with eXtensible Access Control Markup Language, XACML) to dynamically enforce authorization by externalizing access controls. Fisler et al. [258] proposed Binary Decision Diagrams and custom algorithms to check access-control policies. Ahn et al. [259] used answer set programming (ASP) and leverage existing ASP reasoning models to conduct policy verification. Arkoudas et al. [260] proposed a Satisfiability Modulo Theory policy analysis framework.

Sabillon et al. [261] proposed an audit model for conducting cybersecurity audits in organizations and nation-states. Agrawal et al. [262] introduced an auditing framework for determining whether a database system is adhering to its data disclosure policies by allowing users to formulate audit expressions to specify the data subject to disclosure review. Kaaniche et al. [263] proposed the usage of hierarchical ID-based encryption and signature schemes. Noura et al. [146] presented a security and protection audit that can be done by using an audit management system to collect and store logs in a distributed system. Bouet and Israël [264] presented a security assessment framework including an off-line tool enabling security and vulnerability audits of information systems to be used by system architects to assess the security of the system they are designing during the planning phase. The patent "Critical function monitoring and compliance auditing system" [265] describes a system and method for monitoring, auditing, and flagging compliance issues or other user-defined exceptions. Finally, Slapničar et al. [266] analyzed the effectiveness of internal audit of cybersecurity by developing a Cybersecurity Audit Index composed of three dimensions: planning, performing and reporting.

In the scope of compliance auditing cloud computing platforms, Ullah et al. [267] proposed an architecture to build automated security compliance tools, focusing on auditing remote administration and on diagnosing port protection and clock synchronization. Also, Henze et al. [268] presented a practical approach enforcing data compliance in key-value-based Cloud storage systems. Doelitzscher [269] implemented an on-demand audit architecture for Infrastructure as a Service (IaaS) clouds, based on software agents for identifying anomalies for auditing purposes. Finally, there is also SecGuru, designed to audit Azure datacenter network policies [270].

Figure 4 summarizes the main regulations, security frameworks and auditing models applicable to this domain.

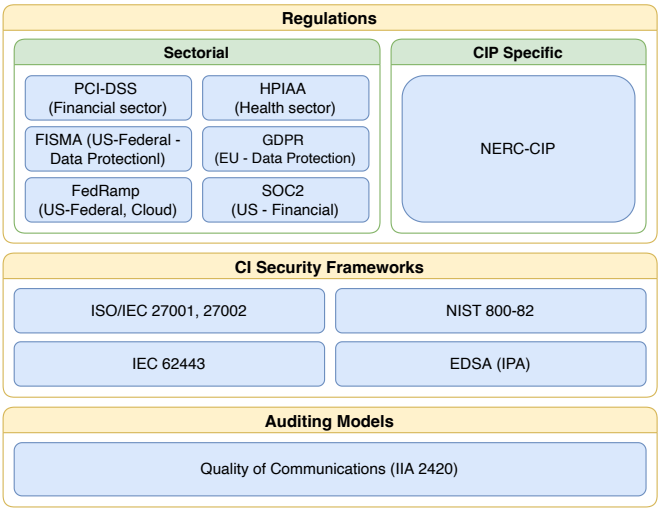

FIGURE 4: Regulations, Frameworks and Auditing Models.

C. STANDARDS FOR COMPLIANCE AUDITING

The development of cyber Information Security Management Systems (ISMS) is guided by standards such as ISO/IEC 27001 and ISO/IEC 27002. As already mentioned, these standards cover the protection of an organization from cyber-attacks [271]. Domain-specific initiatives were also launched to develop and implement IACS standards to secure SCADA environments, including the ones from NIST, which presented the Special Publication 800-82 and International Electrotechnical Commission (IEC) 62443 [272] [273].

Another example is ISO/IEC 62443-1-1 (Security for industrial automation and CS: Terminology, concepts, and models), which constitutes an ongoing effort towards the improvement of cyber-security, robustness, and resilience design. The ISO/IEC 62443 series standard elements are arranged in four groups, namely: Policies and Procedures, System, and Component Requirements. The Policies and Procedures group is focused on the policies and procedures associated with IACS security, with the Systems group addressing the requirements at the system level. Systems and Component Requirements provide information about specific and detailed requirements associated with the development of IACS products [273]. The Japanese Information-Technology Promotion Agency (IPA) also implemented the Embedded Device Security Assurance Certification (EDSA) Program for provisioning SCADA devices [274].

Nevertheless, and despite these efforts from the academia and industry, there is still a lack of standards for compliance auditing techniques in Cloud domains [275].

D. LOGGING SYSTEMS COMPLIANCE

Logs constitute key data sources to acquire visibility and obtain insights from the operational infrastructure processes, with log analysis being recognised as vital for collecting evidence and retrieving the necessary insights to understand the behaviour of a whole system, as well as its individual components, regardless of the deployment type. For instance, Amazon suggests the use of AWS CloudTrail and CloudWatch [276] for auditing purposes, as a web API offering logs and metrics data to their clients.

Being important for administrators, developers and security operators alike (albeit for different reasons), log handling and processing components often have to comply with suitable availability, resiliency and continuous operation requirements – such systems should be sized and ready for possible high-demand situations where the overall system becomes unstable or overloaded, triggering a large number of events.

It is important to rely on a logging system to acquire and deliver information, but also to intelligently process it using insight and analytics. A logging system should provide visibility over its behavior to enable correct predictions. From a security standpoint, log analysis must be reliable and accurate, especially in circumstances involving security incidents or critical situations. Thus, using a inadequate or non-compliant logging system may

have several consequences, such as hampering monitoring, diagnosis or forensics procedures, up to the point of potentially voiding the possibility of gathering legally admissible evidence.

E. SUMMARY

This section presented the key definitions, topics, and related work about compliance auditing standards and regulations. Figure 5 depicts the relationship between the key topics that were addressed. Moreover, the related reviewed literature is summarised in Table 2.

Together with Section III (Forensics), this section also emphasises to which extent forensics procedures and requirements intersect with the regulatory frameworks and standards for compliance auditing, often with mutual benefit. This is one aspect among the many contact points that characterise the relationship between the FCA contexts, for which analytic procedures and tools constitute another cornerstone relationship, which will be further discussed in the next section.

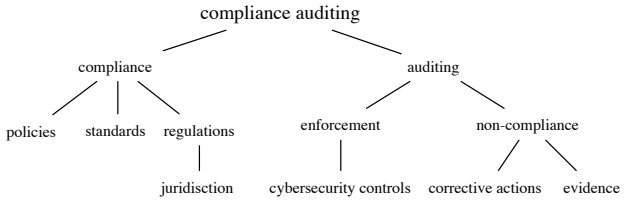

FIGURE 5: Key Concepts of Compliance Auditing.

Scope	Works
Enterprise Cybersecurity	Donaldson et al. [257]
Cybersecurity Audit Model	Sabillon et al. [261]
CI Security Model	Torres et al. [277]
Forensics for incident response	Kent et al. [251]
Auditing Framework	Agrawal et al. [262]
Data Compliance	Henze et al. [268]
Data Privacy Compliance	Kaaniche et al. [263]
	European Union [180]
Anti-Forensic Approaches	Noura et al. [146]
Compliance Audit System Patent	Lee et al. [265]
Security Assessment Framework	Bouet and Israël [264]
Formal Reasoning Mechanisms	Fisler et al. [258]
	Ahn et al. [259]
	Arkoudas et al. [260]
Cloud-based Compliance Auditing	Henze et al. [268]
	Ullah et al. [267]
	Doelitzscher [269]
	Bjørner and Jayaraman [270]

TABLE 2: Reviewed Literature on Compliance Auditing.

V. ANALYTICS FOR CIP FCA: THE ROAD AHEAD

Analytics corresponds to the set of activities focused on how to extract insights from data, correlating evidence to provide security-related capabilities to system administrators, security analysts, and network and application engineers. Analytics leverage FCA capabilities to improve CIP because they help identifying anomalies and their root cause and then extract evidence.

This section will delve into the benefits and challenges of modern analytics in the era of Big Data, AI and ML, starting with a motivation and following with a discussion about the impact of Big Data technologies on CIP. The intersection of Big Data technologies, AI and ML is also discussed, followed by the topic of Forensics Automation. This section closes with a discussion of anomaly detection techniques for log analytics.

A. MOTIVATION AND CONTEXT

Modern protected infrastructures are becoming increasingly complex, a situation for which CIP is no exception, with Industrial IoT infrastructures spreading on a massive scale, both geographically and in terms of components [278]. This has the side effect of generating considerable amounts of operational data and evidence, that cannot be properly handled by traditional analysis techniques. This poses a challenge to FCA, requiring the introduction of scalable techniques able to transport, store and process large amounts of data, thus calling for the adoption of Big Data techniques, designed to handle large amounts of data whose volume is beyond the ability of typical vertical approaches [279].

These circumstances also deem unfeasible to manually analyze large amounts of data, requiring practitioners to resort to automated techniques [280], often supported by AI-based techniques (with a particular focus on ML [241], a branch of AI geared towards automating pattern recognition or classification tasks to analyze vast amounts of data to predict or detect certain behaviors, which in the case of forensics, may consist of discovering or detecting malicious activity). These ML and information retrieval techniques have significantly improved in the last years, enabling the extraction of deeper insights from data [281, 282], with many of these analytic frameworks being able to perform effective and efficient data analysis supported by ML models implemented from a few lines of code, also supporting the automation of time-consuming tasks.

B. THE IMPACT OF BIG DATA TECHNOLOGIES ON CIP

One of the most pressing issues when handling large data volumes is the implementation of efficient distributed storage and retrieval technologies. Big Data NoSQL databases address such challenges with technologies such as MongoDB, HyperTable, Cassandra, and Amazon Dynamo offering scalability and performance predictability that is suitable for storing and indexing real-time streams of big datasets [283]. Kalakanti et al. [10] evaluated different NoSQL datastores as a solution to the data and knowledge management challenges to meet the requirements of performance, reliability and scale imposed by the next generation of data historians as a central repository of SCADA systems.

The need to deal with increasingly big data volumes also calls for an increase in the required amount of computational resources, which must be balanced with the need to contain query latencies within acceptable thresholds. To address this

problem, Google developed the Google File System [284], as well as MapReduce [285], that was designed to address computational challenges. Several efforts were also made to have those technologies available as open source software, resulting in tools such as Apache Hadoop and the Hadoop File System [286].

As already explained, Big Data technologies are especially suitable for CIP and particularly IIoT, where large volumes of data are produced devices from distributed CPSs, for time series analysis. Specialized Time Series Management Systems (TSMS) have been developed to overcome the limitations of general purpose Database Management System (DBMS) for times series management [287]. For instance, Jensen et al. [287] surveyed the field of TSMSs developed bt the academy and the industry, and organized them into categories. Finally, Wang et al. [288] surveyed TSMSs in industrial and IoT fields addressing the new demand such as large amount and real-time analysis of industrial data.

Big Data also poses significant challenges and stresses out privacy requirements, especially those related to privacy regulation emanated from the EU [289]. In that regard, Gartner predicted that by 2018, 50 percent of business ethics violations will be related with data [290].

C. BIG DATA ANALYTICS IN THE AGE OF IA AND ML

In FCA applications, handling large volumes of data is only half of the equation, with analysis being the other half. Extracting insights and patterns from evidence calls for methods other than manual analysis, thus constituting a natural fit for AI and particularly ML techniques, something that was investigated by Brighi et al. [291], that tried to bridge these technologies with the substantive and procedural rules to be observed during investigation activities.

Regarding forensics applications, Hoon et al. [292] reviewed the literature by addressing the challenges and opportunities of employing Big Data in Distributed denial-of-service (DDoS) forensics, implementing and comparing the performance of multiple supervised and unsupervised learning models, according to their efficiency and accuracy. They found that Naïve Bayes, Gradient Boosting and Distributed Random Forest are the most suitable models for DDoS detection, due to their accuracy and time taken on training.

As for network forensics, Yavanoglu and Aydos [293] reviewed the most commonly used datasets in AI and ML techniques, as primary tools for analyzing network traffic and detecting anomalies. Usman et al. [294] proposed a ML approach supported by Decision Tree algorithms to predict IP reputation in zero-day attacks, categorized via behavioral analysis to highlight forensic issues in big datasets. Wiyono and Cahyani [295] presented classification algorithms for network forensics based on the identification of network flows that could track suspected botnet activity in the infected network.

Other tools presented by Hassan et al. [102], Setayeshfar et al. [223] implemented models based on AI to assist forensics experts in monitoring the system and detecting malicious behaviors based on known patterns – however, these tools are not designed for manual forensics tasks such as whole system provenance tracking, being often bound to a single proprietary data stream scheme.

In the scope of Compliance Auditing, Moore and Childers [296] presented a ML solution to automatically generate program affinity policies that consider program behavior and the target machine. Similarly, Quiroz et al. [297] relied on unsupervised algorithms to capture the dynamic behavior of systems and the hidden relationship between the high-level business attribute space and the low-level monitoring space. Similarly, Pelaez et al. [298] used supervised models to capture dynamic behavior. Johansen et al. [299] proposed a mechanism for expressing and enforcing security policies for shared data expressed as stateful meta-code operations defined in scripting languages interposed in the filesystem. Gheibi et al. [300] reviewed the state of the art on the use of ML in self-adaptive systems based in the traditional Monitor-Analysis-Planning-Executing (MAPE) [301] feedback loop. Weyns et al. [302] also presented an approach combining MAPE and Control Theory to produce better adaptive systems.

D. FORENSICS AUTOMATION

Organizations often check whether their security and forensic controls are actually in place as intended using manual assessment procedures. Forensic processes are often no different, being typically time-consuming activities dependent on humans. From this perspective, the lack of qualified human skills and resources can hamper investigation and compliance auditing processes [291].

The use of technology to implement automated processes can streamline forensic investigation tasks fed by large volumes of data. The adoption of automation is therefore seen as an effective strategy to implement forensic processes while reducing the costs and operational errors resulting from human intervention [303], also constituting an emergent field of interest in the research community.

Regarding the introduction of automated procedures, Hayes and Kyobe [304] reviewed the existing research in the field of cyber forensics, identifying current practices and associated challenges that could be tackled by the adoption of automation, as well as the relevant technology that could be leveraged to address such needs. Asquith and Horsman [303] provided an introductory discussion on robotic process automation, a form of service task automation that can improve efficiency in the field of forensics, with Moffitt et al. [305] discussing the automation of repetitive and manual rule-based tasks.

From a more practical perspective, Verma et al. [126, 306] proposed a digital forensic framework that uses case information, case profile data, and expert knowledge for automation of the digital forensic analysis process supported by ML for finding evidence. Also, Patrascu and Patriciu [307] discussed the issues threatening CI systems and proposed

an automated learning framework based on ML algorithms to protect such systems that, despite not being focused on forensics applications, can be leveraged for such purpose.

Finally, recent contributions on the use of ML models supporting the automation of self-adaptive IT operations have been focusing on topics such as observability and AIOps [308, 309] – Notaro et al. [310] has compiled several contributions in this scope.

E. ANOMALY DETECTION FROM LOG DATA SOURCES

An anomaly corresponds to an outlying observation that appears to deviate significantly from a nominal state or a statistical data distribution [244]. Anomalies are often classified into three types: point anomalies, contextual anomalies, and collective anomalies contexts [245]. Anomalies can be expressed by scores or labels [246].

While anomaly detection techniques can be applied for all sorts of data sources, logs are of special importance for FCA applications, due to their almost pervasive and non-invasive nature, playing a vital role in case of a breach or incident analysis as they provide detailed information about activities. Nevertheless, the use of anomaly detection mechanisms using application and service log data for forensics and compliance auditing raises important challenges, due to factors such as the abundance of unstructured plain text contents and heterogeneous formats, redundant runtime information (which sometimes may change, as it is the case for certain IP addresses), and the existence of a significant amount of unbalanced data (a direct consequence of the prevalence of a normal operation mode). Moreover, with the increasing scale and complexity of distributed systems in the CI environment, monitoring, correlating and analysing logs is a time-consuming task that takes considerable effort, making it increasingly unfeasible to manually sort out trough evidence to detect anomalies.

Event correlation can be also categorized into different categories: temporal, spatial, or hybrid, whose combined use allows to capture both local (subsystem level) or global (IACS level) abnormalities [248]. After anomalies have been identified, is important to take forensic efforts in the analysis to determine the root causes and collect evidence, which will help to elaborate on the definition and application of countermeasures.

Some proposals have addressed the usage of log analysis as one of the input sources for anomaly detection. Chen and Li [311], for instance, proposed an improved version of an algorithm for detecting anomalies from audit data while updating the detection profile along with its execution.

Clustering techniques, such as the k-means algorithm, are often used by intrusion detection systems for classifying normal or anomalous events, having also application in the forensics analysis field. For instance, Asif-Iqbal et al. [312] correlated logs from different sources, supported by clustering techniques, to identify and remove unneeded logs. Syarif et al. [313] compared five different clustering algorithms and identified those providing the highest detection accuracy, also concluding that those algorithms were not mature enough for practical applications. Hoglund et al. [314], as well as Hajamydeen et al. [315], classified events in two different stages supported by the same clustering algorithm.

Münz et al. [316] applied the k-means clustering algorithm to feature datasets extracted from raw records, where training data are divided into clusters of time intervals for normal and anomalous traffic. Tian and Jianwen [317] improved traditional means clustering algorithm, to improve efficiency and accuracy when classifying data. Eslamnezhad and Varjani [318] proposed a detection algorithm to increase the quality of the clustering method based on a MinMax k-means algorithm, overcoming the low sensitivity to initial centers in the k-means algorithm. Ranjan and Sahoo [319] proposed a modified k-medoids clustering algorithm by presenting a new strategy to select the initial medoids, overcoming the means in anomaly intrusion detection and the dependency on initial centroids, number of clusters, and irrelevant clusters. Also, a k-nearest neighbor classifier for intrusion detection was explored by Liao and Vemuri [320].

Other authors adopted hybrid solutions for log analysis, combining the use of the k-means algorithm with other techniques for improving detection performance. They realized that despite the inherent complex structure and high computational cost, hybrid classifiers can contribute to improving accuracy. Mohammed et al. [321] proposed a clustering approach based on Fuzzy C-Means (FCM) and K-means algorithms to identify the evidential files and isolate the non-related files based on their metadata. Makanju et al. [322] took advantage of an integrated signature-based and anomaly-based approach to propose a framework based on frequent patterns. Varuna and Natesan [323] introduced a hybrid learning method integrating k-means clustering and Naive Bayes classification. Muda et al. [324] proposed k-means clustering and Naive Bayes classifiers in a hybrid learning approach by splitting instances into potential attacks and normal clusters.

Hybrid approaches have indeed proven to be quite interesting. However, in general, they still take a considerable amount of time to generate models for particular datasets, aggravated by the growth patterns normally associated with log sources in production systems. Elbasiony et al. [325] used data mining techniques to build a hybrid framework for identifying network misuse and detecting intrusions through the use of random forests algorithm to detect misuses, with k-means as the clustering algorithm for unsupervised anomaly detection. Fu et al. [247] presented an algorithm to convert free-form text messages in log files to log keys without heavily relying on application-specific knowledge. Du et al. [326] proposed the use of a Long Short-Term Memory (LSTM) to model a system to automatically learn log patterns from normal execution, and detect anomalies when log patterns deviate from the model trained from log data under normal execution. Henriques et al. [249] proposed an integrated scalable framework for efficiently detecting

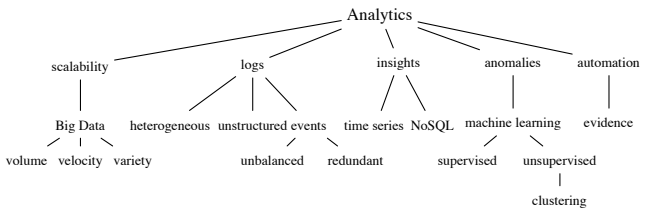

FIGURE 6: Analytics key concepts

anomalous events on large amounts of unlabeled data logs through the use of clustering and classification methods supported by a parallel computing approach.

F. SUMMARY

This section addressed the opportunities and challenges in the use of advanced analytics based on Big Data technologies, with AI and ML support, in the field of FCA. Figure 6 depicts how the key topics addressed in this section are related.

Scope	Works
Automation in Forensics	Asquith and Horsman [303] Hayes and Kyobe [304] Homem [227] Brighi et al. [291] Verma et al. [306]
Big Data Properties	Demchenko et al. [278]
Big Data & AI challenges	Chehri et al. [241]
IoT Definition	Minerva et al. [234]
IoT Cyber-security and Privacy	Infineon et al. [240]
IoT-based Investigations	Stoyanova et al. [239]
IoT Forensics Analysis	Yaqoob et al. [242]
Big Data SCADA Historians	Kalakanti et al. [10]
Time Series Databases	Jensen et al. [287] Wang et al. [288]
Mapreduce for Big Data Analysis	Hegazy et al. [280]
Review on Cybersecurity Datasets for ML	Yavanoglu and Aydos [293]
ML Algorithms to Predict Zero-day Attacks	Usman et al. [294]
Network Forensics Classification Algorithms	Wiyono and Cahyani [295]
DDoS Forensics with ML Big Data Analytics	Hoon et al. [292]
Compliance Auditing Platforms	Ullah et al. [267] Doelitzscher [269] Bjørner and Jayaraman [270]

TABLE 3: Reviewed works related to Big Data-supported FCA.

We surveyed the research in the field of advanced Big Data analytics taking into account the increased softwarization trend in terms of computing and network resource usage, as well as the benefits of leveraging advanced learning algorithms for improved automation. This has allowed to unveil a series of emerging development and evolution paths for FCA practices which are expected to have a profound change across the entire domain. Table 3 summarizes the relevant literature in Big Data for CIP.

VI. A FORENSICS AND COMPLIANCE AUDITING TAXONOMY FOR CIP

To the best of our knowledge, there is no specific taxonomy in the domain of FCA for CIP in the surveyed literature. To fill this gap, we devised a taxonomy covering the scopes as well as the functional and non-functional dimensions of the FCA practice, inspired by forensic investigation and compliance practices. The proposed taxonomy is depicted in Figure 7, being organised along seven major dimensions, inspired by the methodology proposed by [327]. These are the following:

- **Critical Infrastructures:** this dimension characterises the scope and environment to be protected, including SCADA and IACS core systems. Moreover, specific attacks targeting CIs, SIEM, and other security platforms and systems providing protection capabilities are also considered
- **Governance:** gathers the orientations that can support the decisions in the application of FCA processes. It comprises the investigation processes, guidelines, agencies, standards and regulations, training, directives, and existing specific security frameworks.
- **Preparedness:** this dimension comprises the proactive aspects that may be considered to safeguard, support and prepare in advance the execution of FCA processes. It encompasses readiness, forensic by design, forensic frameworks, anti-forensics, and auditing frameworks.
- **Data Acquisition:** this dimension deals with the challenges of gathering digital and network forensics covering aspects such as volume, live forensics and data provenance, while safeguarding the need to protect information about evidence.
- **Evidence Identification:** covers the models, algorithms and approaches helping to identify evidence and non-compliant events. It comprises IDS, detection techniques, causality, and learning, in this last case by using approaches supported by clustering and hybrid approaches algorithms.
- **Reporting:** this category covers communication and interoperability-related aspects, encompassing topics such as privacy concerns, visualization and searching, interoperability, and Chain of Custody.
- **Deployment:** this encompasses non-functional aspects, which relate to platform and infrastructure-related aspects, such as cloud computing, virtualization support, scalability, automation, and quality.

This taxonomy aims at presenting FCA-related topics in a convenient way, using a set of criteria covering both functional and non-functional aspects while striving to provide a convenient organization for the most significant developments.

VII. A REFERENCE ARCHITECTURE FOR FCA SYSTEMS

As already mentioned, even though Forensics and Compliance Auditing are different activities, both in terms of purpose and expected outcomes, there is a considerable

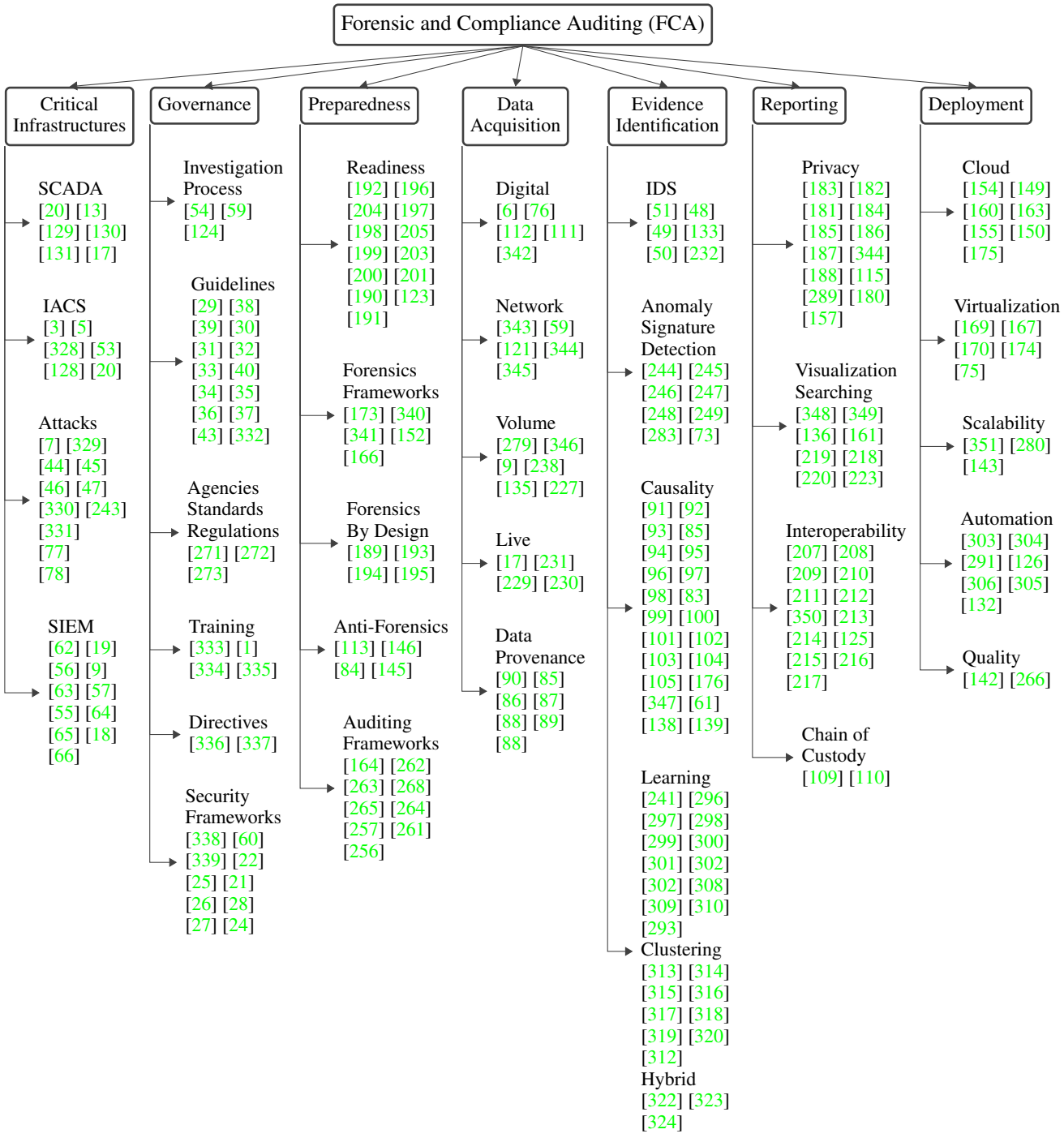

FIGURE 7: Proposed FCA Taxonomy for CIP

amount of proximity between them, since they often resort to the same data sources and similar information and context extraction techniques to gather and process evidence. This hints at the possibility of building both capabilities on top of a shared reference architecture, providing data acquisition, transport and processing pipelines, as well as persistence capabilities.

In this section, we provide such a reference architecture, in order to better identify the various functional blocks typically found in FCA systems. It should be noted that this is an abstract architecture. Real-world FCA tools will usually map into subsets of this architecture.

The main functional requirements to be met by FCA solutions include identifying, extracting, preserving and presenting digital evidence. Table 4 highlights how the architecture's functional blocks typically required for Forensics operations and for Compliance Auditing activities largely overlap.

Figure 8 presents proposed reference architecture. The first stage of FCA systems includes the collection of heterogeneous data from internal and external sources to be gathered into a single logical store. That data can include a vast amount of structured and unstructured heterogeneous data from a large number of sources widely dispersed across the CI, including those from the associated IACS and the ICT infrastructures.

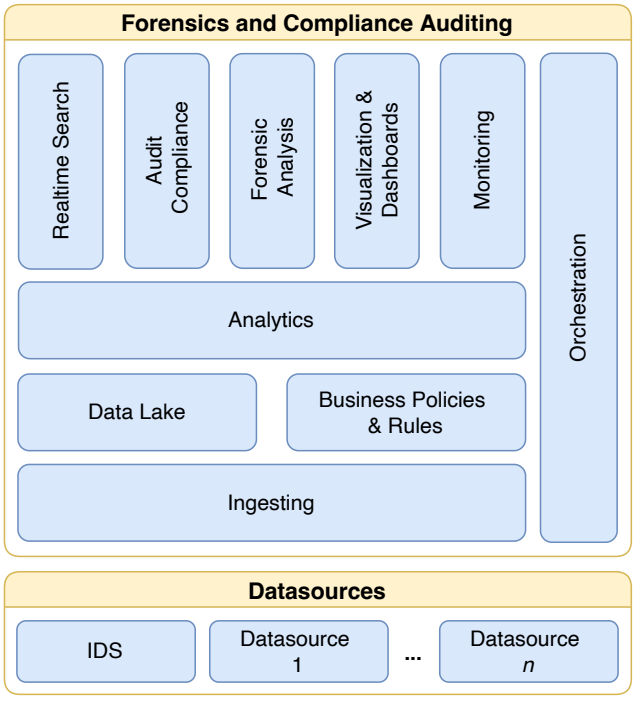

FIGURE 8: Reference Architecture for FCA Systems.

The second stage incorporates forensic analysis and third-party continuous auditing capabilities for the identification of *post mortem* security events, foreseeing, tracking, and tracing possible anomalies. Such objectives can be achieved by correlating the features retrieved from a seemingly disparate class of events that usually are not considered in terms of CI. Thus, beyond the forensics activities, the auditing layer checks compliance with standards, policies and rules. An example of such verifications is the cross-check of past system logs with the registration of physical access to remote facilities, to indirectly detect unauthorised accesses.

Next, we discuss the key components of this architecture.

Functional Block	Forensics	Compliance Auditing
Data Lake	Yes	Yes
Analytics	Yes	Maybe
Business Policies & Rules	No	Yes
Real Time Search	Yes	Yes
Monitoring	No	Yes
Ingesting	Yes	Yes
Orchestration	Yes	Yes
Visualization & Dashboards	Yes	Yes
Platform as a Service	Yes	Yes
Security	Yes	Yes
Cloud-native	Yes	Yes
Scalability	Yes	Yes

TABLE 4: Relevance of Functional Blocks vs. Forensics and Compliance Auditing.

A. DATA SOURCES & DATA INGESTION

The Ingesting Module acts as a set of probes capturing data from a large number of heterogeneous data sources from the surrounding environment, including applications such as Authentication Authorization and Accounting (AAA), ICT security logs (e.g., anti-virus, IDSs), internal personnel activities, physical access control logs (door switches and surveillance cameras), maintenance activities (physical and logical systems), interactions with third-parties (e.g., general documents, emails) and incident logs (e.g. ICT trouble tickets).

Integration of third-party sources within the Ingesting Module is usually accomplished by using custom data adapter components. Such modules ingest data from IDSs, third-party applications or triggered alerts from monitoring processes, also including trust and reputation data, all of it being integrated using pull or push-based approaches.

The ultimate goal of the Ingesting Module is to acquire, parse, enrich and normalize incoming data (which may be structured or unstructured, depending on its nature and sources) into a common format suitable to be stored in the Data Lake (DL) and later used for analysis purposes, while ensuring consistent timestamp synchronization across several sources in order not to compromise event timelines. This means that incoming raw data needs to be handled in a streamlined way, in order to optimize its transport, storage and processing, thus implying the deployment of data processing pipelines akin to Extract, Transform and Load (ETL) workflows.

These Ingesting Module workflows, which may also include filtering, normalization, indexing, enrichment, and aggregation steps, must be capable of dealing with high volumes of heteregeneous data later to be fed into the DL,

which constitutes the central repository component in the reference architecture. Persisted data from different sources (including enriched data) may be used for several purposes, such as training learning models or to feed visualization tools helping to identify threats.

B. DATA LAKE

The DL provides a repository to store data in different formats. This repository centralizes logs and other different sorts of data collected by the Ingesting Module (IM), to be made avaliable to FCA activities (and possibly other applications). Aditionally, the DL also persists the correlation and/or classification results of such data feeds, helping streamline higher-dimensional analytic procedures.

The DL often assumes a distributed nature, to horizontally scale in order to fit increasing volumes of data and/or to increase the performance of data searching and correlating activities. It usually provides the automation capabilities to manage how indexes and queries are distributed across the cluster to accommodate large amounts of data and transactions, including support for automated scaling. This is important since high availability, resiliency, throughput, and low latency when querying large volumes of data are important non-functional requirements for the DL.

The DL may also provide integration mechanisms to plug-in common authentication systems such as Active Directory, Lightweight Directory Access Protocol (LDAP), and Security Assertion Markup Language (SAML).

C. ANALYTICS

After the data is captured and stored in the DL, the Analytics Module takes the responsibility for extracting relevant insights. Supported by state-of-the-art analytic methods, this module provides the capabilities to classify threats with potential impact on the systems' integrity, confidentiality, or availability. It starts by individually identifying unusual behaviors in past events, logged in computers or networks, correlating them in order to identify the compromised systems from the chain of events. For instance, this can be used to correlate the sequence of past executed shell commands with the list of files that have changed, to discover threats. The outcomes of this component also provide an important input to trigger automated rapid response actions.

Within the Analytics Module, the use of ML techniques can help discover new behaviors and patterns to define and/or reveal the policies and business rules used to classify threats, from a vast amount and variety of data. Thus, it is expected that taking such a proactive approach to classify events in advance (before the forensic investigation has even started) may contribute to improve the readiness of forensic and compliance auditing processes. This is further reinforced by the fact that the resulting classified data will also be stored in the DL as input for further forensic analysis processes.

The nature of its role requires Analytics Module to be flexible, allowing models to be updated "on the fly" between retraining, but also to offer a good performance/efficiency balance. The latter can be achieved by decoupling the training and classification processes and running in parallel, thus reducing the time devoted to event classification while increasing the chances of automatically recognizing new threats. Improving the time spent on training can also be achieved by dividing the dataset and even the model, assigning parts to different processes. Thus, even when the training model is too large, it can be trained in the background without disturbing the live system.

Taking advantage of its scale-out properties, the reference Analytics Module architecture is designed to simultaneously train and run different models. Some of them can be used for training, while other ones can be used for classification purposes. Update or introduction of models into production after training should follow best practices, eventually pursuing a MLOps-like lifecycle management approach.

D. FORENSIC ANALYSIS

Forensic analysis is a key step in the investigation process to identify the traces of malicious activity and extract evidence. Additionally, this may also encompass the establishment of a causality path between classified anomalies, oriented towards identifying the root cause and progression path of an incident.

Forensic analysis capabilities can be leveraged by using ML models in the context of the Analytics Module. These can help forensic investigators efficiently find out the relevant events from large amounts of data, coming from diversified sources. Technically, evidence can be collected with queries entailing a set of rules to be run against the events previously stored in the DL.

The adoption of a common standardized forensic schema assumes particular importance in collecting and exchanging relevant information or evidence between different entities and even jurisdictions, along the investigation chain. To ensure that evidence is legally admissible while safeguarding authenticity and integrity, schemas may adopt techniques such as cryptographic hashing.

E. AUDIT COMPLIANCE

The audit compliance component provides the capability to assess conformity with standard practices and defined policies, as part of an ongoing CIP strategy. Such standards may encompass regulatory requirements and/or industry guidelines that the infrastructure operator must comply with for certification, security and/or safety reasons. In case an audit trail is available, an expert can return to the source material to check the quality of the analysis and processing.

Beyond the policies resulting from the need to comply with regulatory or standardization frameworks, organizations can establish custom rules based on their own internal processes and procedures, such as corporate laws, plans, and procedures.

The Audit Compliance module takes business rules and regulatory policies to identify violations and trace the path of non-conforming events. This process assesses the

compliance of the facts denoted by the ongoing events with the defined business rules and policies, providing an outcome that includes scores computed by quantifying the aspects regarding security and the level of risk. Both the Forensic Analysis and the Audit Compliance modules leverage the outcomes from correlating data at the Analytic component.

F. VISUALIZATION AND DASHBOARDS

Visualization capabilities are key for forensics activities, providing the means to display information in a manner that may evince the presence of suspicious or anomalous patterns. Such capabilities can be key to help understand and analyze specific domain datasets by applying histograms, scatter and box plots, tree maps, surface pots, parallel coordinate plots, and radar charts [348, 349].

This module is fed by the data persisted in the DL repository, which is used for analysis purposes. In a typical arrangement, dashboard panels are used to highlight a variety of indicators which may be directly generated from agent feeds, or as the result of enrichment (providing contextual information), aggregation or analytics/analysis sources. For instance, panels may provide information about the total number of received events, their variety, or a histogram depicting when events were received, just to name a few. Moreover, this data may be exported for integration with third-party tools.

Visualization and Dashboards provide operators with suitable graphical tools to explore and analyze contextual information – such tools must provide querying and summarization capabilities adequate for dealing with large volumes of data in repositories, computing metrics and applying specific functions against some attributes.

G. BUSINESS POLICIES AND RULES

Beyond the mandatory regulatory, legal and standardization frameworks, organizations often define specific procedural or workflow rules based on their own internal processes and needs, based on corporate laws, plans or roadmaps.

A repository of CI business policies and rules may be used to support organizational-wide compliance assessment. If those events trigger some of the rules describing policies, then the associated alerts will also be triggered. Such rules can be tuned according to specific thresholds and can help prioritizing and score events. For example, a company policy may impose constraints on their employees on the use of resources, thus, any login attempt violating this rule should be reported. Physical access control is another example: alerts can be triggered when the doors in a given department or physical installation are opened out of the authorized period. Formally, those CI Business rules will assess the compliance of processes accordingly to the business norms.

H. MONITORING

A Monitoring component provides the capabilities to look at things as they happen, helping operators to identify anomalies from data. It can either trigger alerts or highlight information resulting from such a continuous assessment, matching CI audit compliance rules against persisted events in DL. Moreover, it will also check the level of trust and reputation risks to classify eventual threats and trigger alerts to the operators.

Such a Monitoring component may also offer the ability to set up automatic response rulebooks or human-supported actions, as well as triggering alerts and notifications, providing information to help the operator become an effective link of a human-in-the-loop decision chain. Necessarily, models and rules used for alerting purposes must be fine tuned to provide adequate accuracy and low false positive rates.

I. REAL-TIME SEARCH

A Real-Time Search component provides high-performance query capabilities from large amounts of stored data in the DL to support the extraction of relevant FCA information. The component is able to run queries against the indexed data in the DL. Because every second counts when looking up for quick responses, the process for indexing data can be executed in advance to improve the query performance. This component also includes an interface to integrate third-party components to run lookup actions for data.

J. ORCHESTRATION

The Orchestration component offers capabilities for managing and coordinating the different FCA components. Such capabilities comprise automation, self-healing, and service discovery. For cloud-native implementations, the use of containerized services integrated within a microservice architecture may improve scalability, eventually allowing for the deployment of multiple instances of the same architectural functional block to scale capabilities, as needed.

A set of high-performance services requires a management and maintenance subsystem to coordinate the configuration settings supporting distributed FCA service synchronization and operation. While the capabilities of this framework are provided independently by different modules and components and made available to third parties through the use of an Application programming interface (API), the overall system can be depicted as being akin to an atomic structure. This paves the way to its possible provisioning and deployment in the form of a SaaS model.

K. FCA COMPONENT INTEGRATION

Figure 9 depicts the component integration view, providing a perspective that is complementary to the previous discussion. Here we can identify the ingestion stage, with the information coming from the several data sources being admitted by corresponding layer, where it can be also formatted and preprocessed with the help of the Ingestion core to perform filtering, normalization, indexing, enrichment, and aggregation.

Data streams can be persisted to a data lake (for batch processing purposes and forensics evidence persistence), also being sent over a fast path to feed stream processing mechanisms. The data lake can also store the results and outcomes of automatic and manual analytical processes.

Next, the analytics layer provides the primitives and mechanisms for data analysis (by means of the Analytics engine), supporting the core FCA modules, as well as the monitoring, visualization and search components. Moreover, the results of FCA analysis tasks can potentially be used to update the Business Policies and Rules repository, which provides the knowledge base assisting the Analytic engine.

Finally, the Orchestration layer performs a function that is orthogonal to the entire FCA framework, monitoring, managing and coordinating its components.

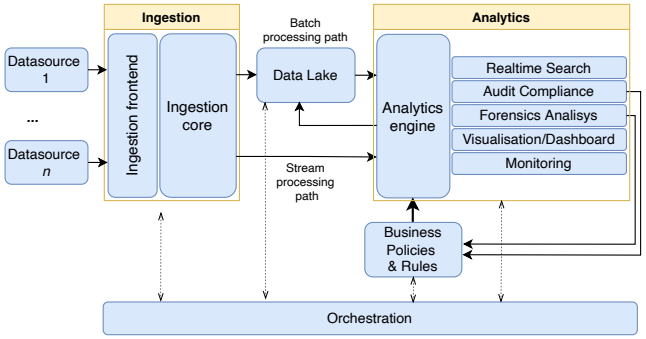

FIGURE 9: FCA Component Integration View.

L. A CLOUD-NATIVE PLATFORM AS A SERVICE

The implementation of a framework designed according to the reference architecture hereby described can also benefit from adopting a Cloud-native architecture in which features are decoupled in microservices designed to improve scale-out capabilities, eventually hosted in containers. Taking this approach makes it possible to have wrap-up FCA solutions supporting a large number of customers.

Providing independent external interfaces to the available functions of the cloud hosting or orchestration platform can provide instrumentation mechanisms for third parties, allowing them to tailor and deploy custom scenarios according to their needs. This approach increases the opportunity to integrate custom third-party solutions with the FCA, resorting to APIs or queuing mechanisms to enable effective integration between third-party applications and the FCA reference architecture. Such capabilities can enable the integration of specific policies and business rules, also providing the means to customize data sources, disable some components or extend their core capabilities, among other options. Moreover, this makes it possible, for instance, to customize solutions to integrate this reference architecture with solutions such as SIEM, SOAR, EDR and XDR.

M. PLATFORM SECURITY

Incorporating security in the FCA reference architecture allows to develop, deploy and operate each component safely, following the best practices in the field. It is also important to protect communication channels, providing secure inter-module integration. For this purpose, the adoption of Zero trust principles [352] may be a key design feature – while those principles are primarily focused on data and service protection, they can and should be expanded to include all enterprise assets and subjects.

Complementary to active protection characteristics, Authentication and Authorization mechanisms should also be properly implemented and continuously assessed to check compliance with the defined access rules.

N. SUMMARY

This section proposed a reference architecture for FCA and its functional building blocks, according to the identified requirements, also detailing roles and interactions. Moreover, non-functional aspects comprising the implementation of Cloud-native Architecture, Platform as a Service, and Platform Security were also addressed, in order to demonstrate the plasticity of the proposed concept in terms of deployment and operational options.

VIII. DISCUSSION AND OPEN ISSUES

This section discusses the findings of this survey and highlights the open issues and the research opportunities to be considered in the topic of FCA in the scope of CIP.

A. DISCUSSION

When it comes to CIP, most literature references are focused on conventional cybersecurity prevention, detection and mitigation techniques. However, and given the considerable overlap of functionalities associated with security, forensics and compliance audit contexts, it makes sense to consider some proposals and technologies as candidates for application in FCA contexts.

In fact, the lack of proper FCA capabilities within CIs may not be attributed to any sort of technological obstacles, but rather to a chronic lack of readiness. For instance, this can lead to situations where forensics procedures are undertaken on an *as needed* basis, long after incidents have occurred, in an offline basis. This can restrict the forensics process in a decisive way, hampering the establishment of a clear perspective about incidents, their root causes and implications.

Moreover, and on the compliance auditing side, there is an ongoing trend requiring CI operators to comply with a growing body of standards and regulations while, at the same time, having to keep up with increasingly complex and interconnected infrastructures with a proliferation of control, sensory and endpoint devices.

The implementation of adequate FCA mechanisms can assist in the prevention as well as in the mitigation of the potential consequences of incidents or adverse events,

improving the CI resiliency. In fact, it is worth noticing that forensically reconstructing past events and highlighting disrupted compliance events in CI environments can make it possible to discover potential vulnerable vectors and hidden threats whose correction can be decisive to avoid future consequences.

When it comes to FCA, proactivity is key. But operators need to understand the added value of adding such capabilities before committing to invest to adapt infrastructures, for instance to deploy and customize adaptor agents to extract the significant amounts of data living in silos (e.g. ICT systems surrounding the CI environment) into a single homogeneous coherent dataset, whose existence can help overcome the complexity arising from the use of a large number of forensic tools, protocols, and standards.

With proper collection mechanisms in place, it becomes possible to correlate data by applying models, algorithms, architectures, and solutions to effectively classify and predict behaviors and extract evidence from large amounts of data or automatically support data-driven decision-making. Moreover, results from correlation can also help enforce auditing compliance on security policies, regulations, recommendations, applicable laws, and standards processes to increase the security and trust in CIs that may help to prevent future security incidents.

Also, FCA need to keep up with times and adapt, as the trend towards resource consolidation also reaches CIs, with the adoption of virtualization technologies within private, public or hybrid clouds. For instance, while the adoption of a cloud-native setup with containers can bring significant challenges in terms of forensics integration, it can also provide net benefits in terms of management, monitoring, and control of FCA frameworks for CIP, providing elasticity to accommodate transient requirements from analysis processes.

Another significant trend with impact in FCA processes is the emergence of IIoT and Big Data, which tend to go hand-in-hand in modern CIs, due to the considerable data handling requirements for massively distributed infrastructures. However, while such developments pose challenges to FCA solutions, it should also be noted that Big Data technologies also provide a technological basis enabling the development of sophisticated forensic data and evidence transport, processing and storage mechanisms that can take advantage of the elasticity of virtualization and cloud technologies.

All the aforementioned aspects have been considered to devise a comprehensive and easy-to-deploy FCA framework template which was designed to be neutral from a deployment standpoint and decoupled from the end-user infrastructure to be protected. This reference design gathers the capabilities to collect and continuously monitor and correlate data from diversified data sources, being able to support decision-makers and forensics practicioners alike, also enabling the definition of responsive actions from large amounts of data. This approach can help track past events to perform evidence extraction and incident root cause analysis, also allowing to detect non-compliant events in near real-time, for example, from logs collected before, during, and after incidents.

B. OPEN ISSUES

This survey also identified a series of open issues and research gaps in terms of FCA capabilities for CIP. Probably one of the most important findings of this survey has to do with realising that, in most cases, existing security tools are missing the integration means for a full-stack FCA solution. This is due to the fact that many of these tools are not embracing open standards on maintaining an effective chain of custody or plug-and-play capabilities to increase their interoperability and reduce the need for collaborative work between tool owners and end-users. Also, many of these tools lack flexible FCA capabilities, decoupled from the applications they aim to protect (e.g. applied to 5G vertical applications taking advantage of cloud-native approaches).

Other identified handicaps that equally affect SIEMs and forensics tools for CIP include: the absence of custom connectors and parsers for data source integration, incomplete data, lack of basic correlation rules, elemental storage capabilities, reliance on manual operation, basic reaction and reporting capabilities, limited data visualization, or deployment, and management complexity [55]. Other missing aspects comprise the lack of GDPR privacy compliance [353], as well as the absence of high-level security risk metrics. Also regarding metrics, there are no well-defined KPIs for FCA tools, for example to assess the Quality of Service (QoS) and Quality of Experience (QoE), reliability, availability, and resiliency.

Also, the availability of open standards, languages, and data abstractions for sharing and exchanging evidence are key to enhance FCA tools and improve their interoperability while enhancing the processes devoted to discovering forensic evidence in an automated, effective, and efficient manner. That includes, for example, the adoption of open standards for sharing evidence and keeping an effective chain of custody.

With the emergence of IIoT scenarios, the requirements to capture, transport and process of large volumes of data become more demanding. Thus, the lack of adequate computational and storage resources may impose limits on the application of FCA methodologies for gathering and analyzing data. Overcoming them is instrumental to achieve a near real-time data correlation latency from multiple physical sources, also enabling the deployment of effective alerting mechanisms for non-compliance incidents. Equally important is the lack of automated and dynamic orchestration capabilities, adaptation systems, and tools supporting FCA activities and managing their entire life cycle, which are key for implementing efficient and resource-effective FCA capabilities.

Another key concern in FCA activities is their eventual impact on performance and efficiency on systems being

secured, such as in the case of collecting large amounts of data for forensic purposes and preserving data privacy [306]. For instance, in systems with specific determinism and real-time requirements, special care must be taken to avoid imposing any kind of undesirable overhead or creating potential points of failure.

IX. CONCLUSION

This work highlights the importance of considering both forensics and compliance auditing (FCA) as high-priority topics for CIP, contributing with guidance in the design and implementation of security processes by considering policies, standards, guidelines and procedures and evidence analysis techniques. For this purpose, we surveyed the latest developments, methodologies, challenges, and solutions addressing FCA in the scope of CIP, focusing on contributions capable of tackling the requirements imposed by massively distributed and complex IACS which handle large volumes of heterogeneous, noisy, redundant and even ambiguous data, for analytic purposes.

We started by highlighting the need for addressing modern security challenges and requirements to improve the security of CI by considering FCA capabilities. With that in mind, a survey of the the relevant literature was undertaken, focused on the intertwined topics that may stress the benefits and value brought by FCA approaches. From this survey it was also noticed the lack of specific FCA approaches and taxonomies for CIP. One of the reasons for this relates to the misleading perception that CIP requirements for FCA may be fulfilled resorting to generic solutions with integrator customisation. However, that may prove difficult due to the domain-specific standardisation and regulatory frameworks which often deviate from more generic recommendations, due to the limitations imposed by the often rigorous CI service continuity, reliability, security and safety requirements. Moreover, aspects such as broad heterogeneity of data sources, and the geographic and administrative dispersion of the CIs, also preclude a straightforward application of mainstream solutions.

The surveyed literature resulted in a taxonomy gathering the major identified categories, such as CIs governance, preparedness, data acquisition, evidence identification, reporting, and data. Together with the lessons learned from the literature analysis, this taxonomy was instrumental to help identify the most relevant FCA capabilities, resulting in the identification of a series of key functional blocks later organized as part of a reference FCA architecture template, designed to provide a strong foundation to support the implementation of future solutions aiming to protect CIs.

ACKNOWLEDGEMENTS

This work was partially funded by National Funds through the FCT – Foundation for Science and Technology, I.P., and the European Social Fund, through the Regional Operational Program Centro 2020, within the scope of the projects UIDB/05583/2020 and CISUC UID/CEC/00326/2020. It was also co-funded by FEDER, in the context of the Competitiveness and Internationalization Operational Program (COMPETE 2020) of the Portugal 2020 framework, in the scope of the Smart5Grid (POCI-01-0247-FEDER-047226) project and also by the "Agenda Mobilizadora Sines Nexus" project (ref. No. 7113), supported by the Recovery and Resilience Plan (PRR) and by the European Funds Next Generation EU, following Notice No. 02/C05-i01/2022, Component 5 - Capitalization and Business Innovation - Mobilizing Agendas for Business Innovation. Furthermore, we would like to thank the Research Center in Digital Services (CISeD) and the Polytechnic of Viseu for their support.

REFERENCES

[1] N. Chowdhury and V. Gkioulos, "Cyber security training for critical infrastructure protection: A literature review," Computer Science Review, vol. 40, 2021.

[2] T. Cruz, J. Proença, P. Simões, M. Aubigny, M. Oedraogo, A. Graziano, and L. Yasakhetu, "Improving cyber-security awareness on industrial control systems: The cockpitci approach," Journal of Information Warfare, vol. 13, no. 4, pp. 27–41, 2014.

[3] L. Rosa, T. Cruz, M. B. de Freitas, P. Quitério, J. Henriques, F. Caldeira, E. Monteiro, and P. Simões, "Intrusion and anomaly detection for the next-generation of industrial automation and control systems," Future Generation Computer Systems, vol. 119, pp. 50–67, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X21000431

[4] C. Foglietta, D. Masucci, C. Palazzo, R. Santini, S. Panzieri, L. Rosa, T. Cruz, and L. Lev, "From detecting cyber-attacks to mitigating risk within a hybrid environment," IEEE Systems Journal, vol. 13, no. 1, pp. 424–435, 2019.

[5] IBM, "IBM Managed Security Services - United States," https://www.ibm.com/security/services/managed-security-services, 2017, visited on 2016-07-02.

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

[7] L. Martin, "5th Annual edition of Cyber Defense Magazine - 2017 Predictions," https://www.tradepub.com/free-offer/cyber-defense-magazine--2017-predictions, 2017, visited on 2017-04-01.

[8] The New York Times, "Biden signs an executive order aimed at protecting critical American infrastructure from cyberattacks." https://www.nytimes.com/2021/07/28/us/politics/cyber-security-biden-executive-order.html, 2021, visited on 2021-10-19.

[9] IBM, "SIBM security intelligence with big data," http://www-03.ibm.com/security/solution/intelligence-big-data, 2016, visited on 2022-12-01.
[10] A. K. Kalakanti, V. Sudhakaran, V. Raveendran, and N. Menon, "A comprehensive evaluation of nosql datastores in the context of historians and sensor data analysis," in Big Data (Big Data), 2015 IEEE International Conference on. IEEE, 2015, pp. 1797–1806.
[11] A. Fernandez, S. Rio, V. Lopez, A. Bawakid, M. Jesus, J. Benitez, and F. Herrera, "Big Data with Cloud Computing: an insight on the computing environment, MapReduce, and programming frameworks," Wiley Interdisciplinary Reviews (WIREs) Data Mining and Knowledge Discovery, December 2014.
[12] Chaossearch, "The threat hunter's handbook: Using log analytics to find and neutralize hidden threats in your environment white paper," National Institute of Standards and Technology, Tech. Rep., 2020.
[13] I. Ahmed, S. Obermeier, M. Naedele, and G. G. Richard III, "SCADA Systems: Challenges for Forensic Investigators," Computer, vol. 45, no. 12, pp. 44–51, Dec. 2012.
[14] K. Stouffer and J. Falco, Guide to supervisory control and data acquisition (SCADA) and industrial control systems security. National institute of standards and technology, 2006.
[15] L. Rosa, M. Freitas, S. Mazo, E. Monteiro, T. Cruz, and P. Simões, "A Comprehensive Security Analysis of a SCADA Protocol: From OSINT to Mitigation," IEEE Access, vol. 7, pp. 42 156–42 168, 2019.
[16] E. Cornelius and M. Fabro, "Recommended practice: Creating cyber forensics plans for control systems," Idaho National Laboratory (INL), Tech. Rep., 2008.
[17] P. Eden, P. Burnap, A. Blyth, K. Jones, H. Soulsby, and Y. Cherdantseva, "A forensic taxonomy of scada systems and approach to incident response," in 3rd International Symposium for ICS & SCADA Cyber Security Research 2015. BCS Learning &Development Ltd, 2015, pp. 42–51.
[18] F. Ullah, M. Edwards, R. Ramdhany, R. Chitchyan, M. A. Babar, and A. Rashid, "Data exfiltration: A review of external attack vectors and countermeasures," Journal of Network and Computer Applications, vol. 101, pp. 18–54, 2018.
[19] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," Journal of Network and Computer Applications, vol. 60, pp. 19–31, 2016.
[20] R. van der Knijff, "Control systems/scada forensics, what's the difference?" Digital Investigation, vol. 11, no. 3, pp. 160 – 174, 2014, special Issue: Embedded Forensics. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1742287614000814
[21] G. Ridley, J. Young, and P. Carroll, "Cobit and its utilization: A framework from the literature," in 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the. IEEE, 2004.
[22] G. Disterer, "ISO/IEC 27000, 27001 and 27002 for information security management," 2013.
[23] NIST, "Security and privacy controls for information systems and organizations SP 800-53 Rev. 5," National Institute of Standards and Technology, Tech. Rep., 2020.
[24] ——, "Protecting controlled unclassified information in nonfederal systems and organizations SP 800-171 Rev. 2," National Institute of Standards and Technology, Tech. Rep., 2020a.
[25] ——, "Framework for improving critical infrastructure cybersecurity - version 1.1," 2018.
[26] HITRUST Alliance, "Hitrust csf framework," 2021, visited on 2022-12-15. [Online]. Available: https://hitrustalliance.net/product-tool/hitrust-csf/
[27] NIST, "Sp 800-53 rev. 5 security and privacy controls for information systems and organizations," National Institute of Standards and Technology, Tech. Rep., 2020. [Online]. Available: https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final
[28] ISO, "ISO/IEC 27000:2018 information technology — security techniques — information security management systems — overview and vocabulary," 2018.
[29] ——, "ISO/IEC 27004:2016 information technology — security techniques — information security management — monitoring, measurement, analysis and evaluation," 2016.
[30] ——, "ISO/IEC 27037:2012 information technology — security techniques — guidelines for identification, collection, acquisition and preservation of digital evidence," 2012.
[31] ——, "ISO/IEC 27038:2014 information technology — security techniques — specification for digital redaction," 2014.
[32] ——, "ISO/IEC 27042:2015 information technology — security techniques — guidelines for the analysis and interpretation of digital evidence," 2015.
[33] ——, "ISO/IEC 27050-1:2019 information technology — electronic discovery — part 1: Overview and concepts," 2019.
[34] ——, "ISO/IEC 27041:2015 information technology — security techniques — guidance on assuring suitability and adequacy of incident investigative method," 2015.
[35] ——, "ISO/IEC 27043:2015 information technology — security techniques — incident investigation principles and processes," 2015.
[36] ——, "ISO/IEC 27006:2015 information technology — security techniques — requirements for bodies providing audit and certification of information security management systems," 2015.
[37] ——, "ISO/IEC TS 27008:2019 information

technology — security techniques — guidelines for the assessment of information security controls," 2019.
[38] ——, "ISO 21043-1:2018 forensic sciences — part 1: Terms and definitions," 2018.
[39] ——, "ISO 21043-2:2018 forensic sciences — part 2: Recognition, recording, collecting, transport and storage of items," 2018.
[40] ——, "ISO/IEC 30121:2015 information technology — governance of digital forensic risk framework," 2015.
[41] ASTM International, "ASTM Standards and publications," https://www.astm.org/products-services/standards-and-publications/standards.html, 2016, visited on 2023-08-04.
[42] R. R. Moeller, Sarbanes-Oxley internal controls: effective auditing with AS5, CobiT, and ITIL. John Wiley & Sons, 2008.
[43] NIST, "Guide to Intrusion Detection and Prevention Systems (IDPS) SP 800-94," National Institute of Standards and Technology, Tech. Rep., 2017.
[44] D. Kushner, "The real story of stuxnet," IEEE Spectrum, vol. 50, no. 3, pp. 48–53, 2013.
[45] ICS-CERT, "Cyber-Attack Against Ukrainian Critical Infrastructure," https://ics-cert.us-cert.gov/alerts/IR-ALERT-H-16-056-01, Feb 2016, visited on 2022-12-01.
[46] R. Khan, P. Maynard, K. McLaughlin, D. Laverty, and S. Sezer, "Threat Analysis of BlackEnergy Malware for Synchrophasor based Real-time Control and Monitoring in Smart Grid," in 4th Int'l Symposium ICS & SCADA Cyber Security Research. BCS, 2016, pp. 53–63.
[47] H. I. M. Abdullah, Z.-A. Ibrahim, F. A. Rahim, H. S. Fadzil, S. A. S. Nizam, and M. Z. Mustaffa, "Digital forensics investigation procedures of smart grid environment," International Journal of Computing and Digital System, 2021.
[48] H. Debar, "An introduction to intrusion-detection systems," Proceedings of Connect, vol. 2000, 2000.
[49] V. V. Phoha, Internet security dictionary. Springer Science & Business Media, 2007.
[50] K. Scarfone and P. Mell, "Guide to intrusion detection and prevention systems (IDPS)," NIST special publication, vol. 800, no. 2007, p. 94, 2007.
[51] W. Wang, S. Gombault, and T. Guyet, "Towards fast detecting intrusions: using key attributes of network traffic," in Internet Monitoring and Protection, 2008. ICIMP'08. The Third International Conference on. IEEE, 2008, pp. 86–91.
[52] T. Cruz, L. Rosa, J. Proença, L. Maglaras, M. Aubigny, L. Lev, J. Jiang, and P. Simões, "A cybersecurity detection framework for supervisory control and data acquisition systems," IEEE Transactions on Industrial Informatics, vol. 12, no. 6, pp. 2236–2246, 2016.
[53] T. Cruz, J. Barrigas, J. Proença, A. Graziano, S. Panzieri, L. Lev, and P. Simões, "Improving network security monitoring for industrial control systems," in 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), 2015, pp. 878–881.
[54] M. E. Whitman and H. J. Mattord, Principles of information security. Cengage learning, 2011.
[55] G. González-Granadillo, S. González-Zarzosa, and R. Diaz, "Security information and event management (siem): Analysis, trends, and usage in critical infrastructures," Sensors, vol. 21, no. 14, p. 4759, 2021.
[56] Gartner, "Magic Quadrant for Security Information and Event Management," https://www.gartner.com/doc/reprints?id=1-26Q3T88Y&ct=210706&st=sb, August 2021, visited on 2021-11-26.
[57] A. S. Gillis, "Security information and event management (siem)," http://searchsecurity.techtarget.com/definition/security-information-and-event-management-SIEM, Jun 2017.
[58] Y. Sun, H. Yan, J. Zhang, Y. Xia, S. Wang, R. Bie, and Y. Tian, "Organizing and Querying the Big Sensing Data with Event-Linked Network in the Internet of Things," in International Journal of Distributed Sensor Networks, 2014.
[59] R. Hunt and J. Slay, "Achieving critical infrastructure protection through the interaction of computer security and network forensics," in 2010 Eighth International Conference on Privacy, Security and Trust, Aug 2010, pp. 23–30.
[60] Gartner, "Magic Quadrant for Endpoint Protection Platforms," https://www.gartner.com/doc/reprints?id=1-27NCOQKK&ct=211014&st=sb, May 2021, visited on 2021-11-29.
[61] W. U. Hassan, A. Bates, and D. Marino, "Tactical provenance analysis for endpoint detection and response systems," in 2020 IEEE Symposium on Security and Privacy (SP). IEEE, 2020, pp. 1172–1189.
[62] Securonix, "Securonix security analytics platform," http://www.securonix.com/security-intelligence, 2016, visited on 2022-12-01.
[63] RSA, "Rsa security analytics," http://www.emc.com/collateral/data-sheet/security-analytics-overview-ds.pdf, 2016, visited on 2022-12-01.
[64] LogRhythm, "Logrhythm security analytics," https://logrhythm.com/products/security-analytics, 2016, visited on 2022-12-01.
[65] Pravail, "Pravail security analytics," https://www.pravail.com, 2016, visited on 2022-12-01.
[66] Alienvault, "Alienvault: A integrated solution with real-time threat intelligence," http://www.alienvault.com, 2016, visited on 2022-12-01.
[67] Cisco, "OpenSOC: Big Data Security Analytics Framework," http://opensoc.github.io, December

2016, visited on 2022-12-01.
[68] Apache Foundation, "Apache metron: Real-time big data security," http://metron.incubator.apache.org, December 2016, visited on 2022-12-01.
[69] C. Lam, Hadoop in Action. Manning Publications, 12 2010. [Online]. Available: http://amazon.de/o/ASIN/1935182196/
[70] Kibana, "Kibana: Explore and Visualize Your Data," https://www.elastic.co/products/kibana, 2016, visited on 2022-12-01.
[71] Elastic, "Elasticsearch: Search and Analyze Data in Real Time," https://www.elastic.co/products/elasticsearch, 2016, visited on 2022-12-01.
[72] IBM, "Security Thought Leadership White Paper," http://www.redbooks.ibm.com/redpapers/pdfs/redp4956.pdf, September 2016, visited on 2017-02-01.
[73] G. Gonzalez-Granadillo, R. Diaz, J. Caubet, and I. Garcia-Milà, "CLAP: a cross-layer analytic platform for the correlation of cyber and physical security events affecting water critical infrastructures," Journal of Cybersecurity and Privacy, vol. 1, no. 2, pp. 365–386, 2021.
[74] E. Morioka and M. Sharbaf, "Cloud Computing: Digital Forensic Solutions," in International Conference on Information Technology-New Generations, ser. 12, Las Vegas, April 2015, pp. 589–594.
[75] D. R. Rani and G. Geethakumari, "An efficient approach to forensic investigation in cloud using VM snapshots," in 2015 International Conference on Pervasive Computing (ICPC), Jan 2015, pp. 1–5.
[76] NIST, "Guide to Integrating Forensics Techniques into Incident Response," http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-86.pdf, 2017, visited on 2017-06-01.
[77] J. T. Langill, "Defending against the dragonfly cyber security attacks," Retrieved, vol. 11, p. 2015, 2014.
[78] M. Fillinger and M. Stevens, "Reverse-engineering of the cryptanalytic attack used in the flame super-malware," in International Conference on the Theory and Application of Cryptology and Information Security. Springer, 2015, pp. 586–611.
[79] D. Dzung, M. Naedele, T. P. Von Hoff, and M. Crevatin, "Security for industrial communication systems," Proceedings of the IEEE, vol. 93, no. 6, pp. 1152–1177, 2005.
[80] F. Köster, M. Klaas, H. Nguyen, M. Braendle, S. Obermeier, and W. Brenner, "Collaborative Security Assessments in Embedded Systems Development," in International Conference on Security and Cryptography (SECRYPT 2009), 2009.
[81] E. Levy, "Crossover: online pests plaguing the off line world," IEEE Security & Privacy, vol. 99, no. 6, pp. 71–73, 2003.
[82] K. Sindhu and B. Meshram, "Digital Forensic Investigation Tools and Procedures," in International Journal of Computer Network and Information Security, ser. IJCNIS, April 2012.
[83] M. N. Hossain, J. Wang, O. Weisse, R. Sekar, D. Genkin, B. He, S. D. Stoller, G. Fang, F. Piessens, E. Downing et al., "{Dependence-Preserving} data compaction for scalable forensic analysis," in 27th USENIX Security Symposium (USENIX Security 18), 2018, pp. 1723–1740.
[84] A. Kumar, G. Singh, A. Kansal, and K. Singh, "Digital image forensic approach to counter the JPEG anti-forensic attacks," IEEE Access, vol. 9, pp. 4364–4375, 2021.
[85] A. Bates, W. U. Hassan, K. Butler, A. Dobra, B. Reaves, P. Cable, T. Moyer, and N. Schear, "Transparent web service auditing via network provenance functions," in Proceedings of the 26th International Conference on World Wide Web, 2017, pp. 887–895.
[86] R. Bose and J. Frew, "Composing lineage metadata with XML for custom satellite-derived data products," in Proceedings. 16th International Conference on Scientific and Statistical Database Management, 2004. IEEE, 2004, pp. 275–284.
[87] K.-K. Muniswamy-Reddy, D. A. Holland, U. Braun, and M. I. Seltzer, "Provenance-aware storage systems." in Usenix annual technical conference, general track, 2006, pp. 43–56.
[88] A. Gehani and D. Tariq, "Spade: Support for provenance auditing in distributed environments," in ACM/IFIP/USENIX International Conference on Distributed Systems Platforms and Open Distributed Processing. Springer, 2012, pp. 101–120.
[89] R. P. Spillane, R. Sears, C. Yalamanchili, S. Gaikwad, M. Chinni, and E. Zadok, "Story book: An efficient extensible provenance framework." in Workshop on the Theory and Practice of Provenance, 2009.
[90] F. Zafar, A. Khan, S. Suhail, I. Ahmed, K. Hameed, H. M. Khan, F. Jabeen, and A. Anjum, "Trustworthy data: A survey, taxonomy and future trends of secure provenance schemes," Journal of network and computer applications, vol. 94, pp. 50–68, 2017.
[91] W. U. Hassan, M. A. Noureddine, P. Datta, and A. Bates, "Omegalog: High-fidelity attack investigation via transparent multi-layer log analysis," in Network and Distributed System Security Symposium, 2020.
[92] S. Ma, J. Zhai, F. Wang, K. H. Lee, X. Zhang, and D. Xu, "MPI: Multiple perspective attack investigation with semantic aware execution partitioning," in 26th USENIX Security Symposium (USENIX Security 17), 2017, pp. 1111–1128.
[93] A. Bates, D. J. Tian, K. R. Butler, and T. Moyer, "Trustworthy Whole-System provenance for the linux kernel," in 24th USENIX Security Symposium (USENIX Security 15), 2015, pp. 319–334.

[94] M. N. Hossain, S. M. Milajerdi, J. Wang, B. Eshete, R. Gjomemo, R. Sekar, S. Stoller, and V. Venkatakrishnan, "SLEUTH: Real-time attack scenario reconstruction from COTS audit data," in 26th USENIX Security Symposium (USENIX Security 17), 2017, pp. 487–504.

[95] T. Pasquier, X. Han, T. Moyer, A. Bates, O. Hermant, D. Eyers, J. Bacon, and M. Seltzer, "Runtime analysis of whole-system provenance," in Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 2018, pp. 1601–1616.

[96] K. H. Lee, X. Zhang, and D. Xu, "Loggc: garbage collecting audit log," in Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security, 2013, pp. 1005–1016.

[97] W. U. Hassan, L. Aguse, N. Aguse, A. Bates, and T. Moyer, "Towards scalable cluster auditing through grammatical inference over provenance graphs," in Network and Distributed Systems Security Symposium, 2018.

[98] S. Ma, J. Zhai, Y. Kwon, K. H. Lee, X. Zhang, G. Ciocarlie, A. Gehani, V. Yegneswaran, D. Xu, and S. Jha, "Kernel-supported cost-effective audit logging for causality tracking," in 2018 USENIX Annual Technical Conference (USENIX ATC 18), 2018, pp. 241–254.

[99] Y. Tang, D. Li, Z. Li, M. Zhang, K. Jee, X. Xiao, Z. Wu, J. Rhee, F. Xu, and Q. Li, "Nodemerge: Template based efficient data reduction for big-data causality analysis," in Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 2018, pp. 1324–1337.

[100] Y. Liu, M. Zhang, D. Li, K. Jee, Z. Li, Z. Wu, J. Rhee, and P. Mittal, "Towards a timely causality analysis for enterprise security." in NDSS, 2018.

[101] Z. Xu, Z. Wu, Z. Li, K. Jee, J. Rhee, X. Xiao, F. Xu, H. Wang, and G. Jiang, "High fidelity data reduction for big data security dependency analyses," in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 504–516.

[102] W. U. Hassan, S. Guo, D. Li, Z. Chen, K. Jee, Z. Li, and A. Bates, "Nodoze: Combatting threat alert fatigue with automated provenance triage," in Network and Distributed Systems Security Symposium, 2019.

[103] X. Han, T. Pasquier, A. Bates, J. Mickens, and M. Seltzer, "Unicorn: Runtime provenance-based detector for advanced persistent threats," arXiv preprint arXiv:2001.01525, 2020.

[104] Q. Wang, W. U. Hassan, D. Li, K. Jee, X. Yu, K. Zou, J. Rhee, Z. Chen, W. Cheng, C. A. Gunter et al., "You are what you do: Hunting stealthy malware via data provenance analysis." in NDSS, 2020.

[105] A. Bates and W. U. Hassan, "Can data provenance put an end to the data breach?" IEEE Security & Privacy, vol. 17, no. 4, pp. 88–93, 2019.

[106] G. Giova et al., "Improving chain of custody in forensic investigation of electronic digital systems," International Journal of Computer Science and Network Security, vol. 11, no. 1, pp. 1–9, 2011.

[107] J. R. Vacca, Computer Forensics: Computer Crime Scene Investigation (Networking Series). Charles River Media, Inc., 2005.

[108] M. M. Houck and J. A. Siegel, Fundamentals of forensic science. Academic Press, 2009.

[109] Y. Prayudi and A. Sn, "Digital chain of custody: State of the art," International Journal of Computer Applications, vol. 114, no. 5, 2015.

[110] J. Cosic and M. Baca, "A framework to (Im)prove" chain of custody" in digital investigation process," in Central European Conference on Information and Intelligent Systems. Faculty of Organization and Informatics Varazdin, 2010, p. 435.

[111] S. Bosworth and M. E. Kabay, Computer security handbook. John Wiley & Sons, 2002.

[112] R. McKemmish, What is forensic computing? Australian Institute of Criminology Canberra, 1999.

[113] S. Rekhis and N. Boudriga, "A system for formal digital forensic investigation aware of anti-forensic attacks," IEEE transactions on information forensics and security, vol. 7, no. 2, pp. 635–650, 2011.

[114] G. Forman, K. Eshghi, and S. Chiocchetti, "Finding similar files in large document repositories," in Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. ACM, 2005, pp. 394–400.

[115] M. Afzaal, C. Di Sarno, L. Coppolino, S. D'Antonio, and L. Romano, "A resilient architecture for forensic storage of events in critical infrastructures," in 2012 IEEE 14th international symposium on high-assurance systems engineering. IEEE, 2012, pp. 48–55.

[116] R. Rivest, "RFC 1321: The MD5 message-digest algorithm," 1992.

[117] N. Mikus, "An analysis of disc carving techniques," Master's thesis, Naval Postgraduate School, 2005, visited on 2017-03-01.

[118] A. Pal and N. Memon, "The evolution of file carving," IEEE Signal Processing Magazine, vol. 26, no. 2, pp. 59–71, 2009.

[119] A. Bates, K. Butler, A. Haeberlen, M. Sherr, and W. Zhou, "Let SDN be your eyes: Secure forensics in data center networks," in Proceedings of the NDSS workshop on security of emerging network technologies (SENT'14), 2014, pp. 1–7.

[120] D. Spiekermann and T. Eggendorfer, "Challenges of network forensic investigation in virtual networks," Journal of Cyber Security and Mobility, pp. 15–46, 2016.

[121] D. McPherson, R. Dobbins, M. Hollyman, C. Labovitzh, and J. Nazario, "Worldwide infrastructure security report," www.arbornetworks.com/report, January 2010, visited on 2022-12-01.

[122] F. Casino, T. K. Dasaklis, G. P. Spathoulas, M. Anagnostopoulos, A. Ghosal, I. Borocz, A. Solanas, M. Conti, and C. Patsakis, "Research trends, challenges, and emerging topics in digital forensics: A review of reviews," IEEE Access, vol. 10, pp. 25 464–25 493, 2022.
[123] P. Sommer, "Digital evidence," Digital Investigations and E-Disclosure: A Guide to Forensic Readiness for Organizations, Security Advisers and Lawyers, The Information Assurance Advisory Council (IAAC),, 2012.
[124] J. Williams, "Acpo good practice guide for digital evidence," Metropolitan Police Service, Association of chief police officers, GB, pp. 1556–6013, 2012.
[125] R. van Baar, H. van Beek, and E. van Eijk, "Digital forensics as a service: A game changer," Digital Investigation, vol. 11, pp. S54–S62, 2014.
[126] R. Verma, J. Govindaraj Dr, S. Chhabra, and G. Gupta, "DF 2.0: an automated, privacy preserving, and efficient digital forensic framework that leverages machine learning for evidence prediction and privacy evaluation," Journal of Digital Forensics, Security and Law, vol. 14, no. 2, p. 3, 2019.
[127] G. Ahmadi-Assalemi, H. M. Al-Khateeb, G. Epiphaniou, J. Cosson, H. Jahankhani, and P. Pillai, "Federated blockchain-based tracking and liability attribution framework for employees and cyber-physical objects in a smart workplace," in 2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3). IEEE, 2019, pp. 1–9.
[128] I. Ahmed, S. Obermeier, S. Sudhakaran, and V. Roussev, "Programmable logic controller forensics," IEEE Security & Privacy, vol. 15, no. 6, pp. 18–24, 2017.
[129] T. Kilpatrick, J. Gonzalez, R. Chandia, M. Papa, and S. Shenoi, "An architecture for SCADA network forensics," Advances in digital forensics II, pp. 273–285, 2006.
[130] ——, "Forensic analysis of SCADA systems and networks," International Journal of Security and Networks, vol. 3, no. 2, pp. 95–102, 2008.
[131] R. Chandia, J. Gonzalez, T. Kilpatrick, M. Papa, and S. Shenoi, "Security strategies for SCADA networks," in International Conference on Critical Infrastructure Protection. Springer, 2007, pp. 117–131.
[132] M. Elhoseny, H. Abbas, A. E. Hassanien, K. Muhammad, and A. Kumar Sangaiah, "Secure automated forensic investigation for sustainable critical infrastructures compliant with green computing requirements," IEEE Transactions on Sustainable Computing, vol. 5, no. 2, pp. 174–191, 2020.
[133] C. Valli, SCADA forensics with Snort IDS. CSREA Press, 2009.
[134] J.-J. Huang, "Two Steps Genetic Programming for Big Data - Perspective of Distributed and High-Dimensional Data," in 2015 IEEE International Congress on Big Data (BigData Congress). IEEE, 2015, pp. 753–756.
[135] D. Quick and K.-K. R. Choo, "Impacts of increasing volume of digital forensic data: A survey and future research challenges," Digital Investigation, vol. 11, no. 4, pp. 273–294, 2014.
[136] J. Koven, E. Bertini, L. Dubois, and N. Memon, "InVEST: intelligent visual email search and triage," Digital Investigation, vol. 18, pp. S138–S148, 2016.
[137] C. Stelly and V. Roussev, "Nugget: A digital forensics language," Digital Investigation, vol. 24, pp. S38–S47, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1742287618300380
[138] M. Zipperle, F. Gottwalt, E. Chang, and T. Dillon, "Provenance-based intrusion detection systems: A survey," ACM Computing Surveys, vol. 55, no. 7, pp. 1–36, 2022.
[139] A. Alsaheel, Y. Nan, S. Ma, L. Yu, G. Walkup, Z. B. Celik, X. Zhang, and D. Xu, "Atlas: A sequence-based learning approach for attack investigation." in USENIX Security Symposium, 2021, pp. 3005–3022.
[140] Y. Kwon, F. Wang, W. Wang, K. H. Lee, W.-C. Lee, S. Ma, X. Zhang, D. Xu, S. Jha, G. F. Ciocarlie et al., "MCI: modeling-based causality inference in audit logging for attack investigation." in NDSS, vol. 2, 2018, p. 4.
[141] S. Ma, X. Zhang, D. Xu et al., "Protracer: Towards practical provenance tracing by alternating between logging and tainting." in NDSS, vol. 2, 2016, p. 4.
[142] D. Ayers, "A second generation computer forensic analysis system," digital investigation, vol. 6, pp. S34–S42, 2009.
[143] V. Roussev and G. Richard, "Breaking the performance wall: The case for distributed digital forensics," in Proceedings of the 2004 digital forensics research workshop, vol. 94, 2004.
[144] L. Daubner, M. Macak, B. Buhnova, and T. Pitner, "Towards verifiable evidence generation in forensic-ready systems," in 2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020, pp. 2264–2269.
[145] R. Harris, "Arriving at an anti-forensics consensus: Examining how to define and control the anti-forensics problem," Digital Investigation, vol. 3, pp. 44–49, 2006, the Proceedings of the 6th Annual Digital Forensic Research Workshop (DFRWS '06). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1742287606000673
[146] H. N. Noura, O. Salman, A. Chehab, and R. Couturier, "Distlog: A distributed logging scheme for IoT forensics," Ad Hoc Networks, vol. 98, p. 102061, 2020.
[147] S. Almulla, Y. Iraqi, and A. Jones, "A state-of-the-art

review of cloud forensics," Journal of Digital Forensics, Security and Law, vol. 9, no. 4, p. 2, 2014.
[148] C. Alcaraz, I. Agudo, D. Nunez, and J. Lopez, "Managing incidents in smart grids a la cloud," in 2011 IEEE Third International Conference on Cloud Computing Technology and Science. IEEE, 2011, pp. 527–531.
[149] Martini and Choo, "Cloud forensic technical challenges and solutions: A snapshot," IEEE Cloud Computing, vol. 1, no. 4, pp. 20–25, 2014.
[150] NIST Cloud Computing Forensic Science Working Group and others, "NIST cloud computing forensic science challenges," National Institute of Standards and Technology, Tech. Rep., 2014.
[151] L. M. Kaufman, "Data security in the world of cloud computing," IEEE Security & Privacy, vol. 7, no. 4, pp. 61–64, 2009.
[152] S. Almulla, Y. Iraqi, and A. Jones, "Feasibility of Digital Forensic Examination and Analysis of a Cloud Based Storage Snapshot," Journal of Digital Information Management, vol. 15, no. 1, 2017.
[153] K. Ruan, J. Carthy, T. Kechadi, and M. Crosbie, "Cloud Forensics," in 7th IFIP Advances in Digital Forensics VII, G. Peterson and S. Shenoi, vol. 361, 2011, pp. 35–46.
[154] H. Abbas, C. Magnusson, L. Yngstrom, and A. Hemani, "Addressing dynamic issues in information security management," Information Management & Computer Security, vol. 19, no. 1, pp. 5–24, 03 2011.
[155] M. P. Mohite and S. B. Ardhapurkar, "Design and Implementation of a Cloud Based Computer Forensic Tool," in 2015 Fifth International Conference on Communication Systems and Network Technologies, April 2015, pp. 1005–1009.
[156] H. van Beek, J. van den Bos, A. Boztas, E. van Eijk, R. Schramp, and M. Ugen, "Digital forensics as a service: Stepping up the game," Forensic Science International: Digital Investigation, vol. 35, p. 301021, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666281720300706
[157] S. Zawoad, A. K. Dutta, and R. Hasan, "SecLaaS: Secure Logging-as-a-Service for Cloud Forensics," CoRR, vol. abs/1302.6267, 2013. [Online]. Available: http://arxiv.org/abs/1302.6267
[158] ——, "Towards building forensics enabled cloud through secure logging-as-a-service," IEEE Transactions on Dependable and Secure Computing, vol. 13, no. 2, pp. 148–162, 2015.
[159] B. Manral, G. Somani, K.-K. R. Choo, M. Conti, and M. S. Gaur, "A systematic survey on cloud forensics challenges, solutions, and future directions," ACM Comput. Surv., vol. 52, no. 6, Nov. 2019. [Online]. Available: https://doi.org/10.1145/3361216
[160] K. Ruan and J. Carthy, "Cloud computing reference architecture and its forensic implications: A preliminary analysis," in Digital Forensics and Cyber Crime. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 1–21.
[161] H. Hibshi, T. Vidas, and L. Cranor, "Usability of forensics tools: A user study," in 2011 Sixth International Conference on IT Security Incident Management and IT Forensics, May 2011, pp. 81–91.
[162] L. Yu, L. Chen, Z. Cai, H. Shen, Y. Liang, and Y. Pan, "Stochastic Load Balancing for Virtual Resource Management in Datacenters," IEEE Transactions on Cloud Computing, vol. PP, no. 99, pp. 1–1, 2016.
[163] A. Patrascu and V.-V. Patriciu, "Logging system for cloud computing forensic environments," Journal of Control Engineering and Applied Informatics, vol. 16, no. 1, pp. 80–88, 2014.
[164] C. Jackson, R. Agrawal, J. Walker, and W. Grosky, "Scenario-based design for a cloud forensics portal," in 2015 IEEE International Symposium on Technologies for Homeland Security (HST), April 2015, pp. 1–6.
[165] A. K. Mishra, M. Govil, and E. Pilli, "A taxonomy of hypervisor forensic tools," in IFIP International Conference on Digital Forensics. Springer, 2020, pp. 181–199.
[166] S. Saibharath and G. Geethakumari, "Cloud forensics: Evidence collection and preliminary analysis," in 2015 IEEE International Advance Computing Conference (IACC), June 2015, pp. 464–467.
[167] A. Huseinović and S. Ribić, "Virtual machine memory forensics," in 2013 21st Telecommunications Forum Telfor (TELFOR), Nov 2013, pp. 940–942.
[168] Y. Cheng, X. Fu, X. Du, B. Luo, and M. Guizani, "A lightweight live memory forensic approach based on hardware virtualization," Information Sciences, vol. 379, pp. 23–41, 2017.
[169] S. Zhang, L. Wang, and X. Han, "A KVM virtual machine memory forensics method based on VMCS," in Computational Intelligence and Security (CIS), 2014 Tenth International Conference on. IEEE, 2014, pp. 657–661.
[170] L. Guangqi, W. Lianhai, Z. Shuhui, X. Shujiang, and Z. Lei, "Memory dump and forensic analysis based on virtual machine," in Mechatronics and Automation (ICMA), 2014 IEEE International Conference on. IEEE, 2014, pp. 1773–1777.
[171] P. Sharma, P. Porras, S. Cheung, J. Carpenter, and V. Yegneswaran, "Scalable microservice forensics and stability assessment using variational autoencoders," arXiv preprint arXiv:2104.13193, 2021.
[172] C. Stelly and V. Roussev, "SCARF: a container-based approach to cloud-scale digital forensic processing," Digital Investigation, vol. 22, pp. S39–S47, 2017.
[173] S. Saibharath and G. Geethakumari, "Design and Implementation of a forensic framework for Cloud in OpenStack cloud platform," in Advances in Computing, Communications and Informatics

(ICACCI, 2014 International Conference on. IEEE, 2014, pp. 645–650.
[174] M. Banas, "Cloud Forensic Framework For IaaS With Support for Volatile Memory," Master's thesis, School of Computing, National College of Ireland, 2015.
[175] NIST, "NIST Cloud Computing Forensic Science Challenges," http://csrc.nist.gov/publications/drafts/nistir-8006/draft_nistir_8006.pdf, note = visited on 2016-11-01, November 2016.
[176] X. Liang, S. Shetty, D. Tosh, C. Kamhoua, K. Kwiat, and L. Njilla, "Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability," in 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID). IEEE, 2017, pp. 468–477.
[177] K. Awuson-David, T. Al-Hadhrami, M. Alazab, N. Shah, and A. Shalaginov, "BCFL logging: an approach to acquire and preserve admissible digital forensics evidence in cloud ecosystem," Future Generation Computer Systems, vol. 122, pp. 1–13, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X21000807
[178] P. Purnaye and V. Kulkarni, "A comprehensive study of cloud forensics," Archives of Computational Methods in Engineering, vol. 29, no. 1, pp. 33–46, 2022.
[179] P. O'Callaghan, Refining privacy in tort law. Springer Science & Business Media, 2012.
[180] European Union, "General data protection regulation - REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Official Journal of the European Union, 2016.
[181] A. Aminnezhad, A. Dehghantanha, and M. T. Abdullah, "A survey on privacy issues in digital forensics," International Journal of Cyber-Security and Digital Forensics, vol. 1, no. 4, pp. 311–324, 2012.
[182] A. Dehghantanha and K. Franke, "Privacy-respecting digital investigation," in 2014 Twelfth Annual International Conference on Privacy, Security and Trust. IEEE, 2014, pp. 129–138.
[183] W. van Staden, "Protecting third party privacy in digital forensic investigations," in IFIP International Conference on Digital Forensics. Springer, 2013, pp. 19–31.
[184] F. Y. Law, P. P. Chan, S.-M. Yiu, K.-P. Chow, M. Y. Kwan, K. Hayson, and P. K. Lai, "Protecting digital data privacy in computer forensic examination," in 2011 Sixth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering. IEEE, 2011, pp. 1–6.
[185] S. Hou, T. Uehara, S. Yiu, L. C. Hui, and K.-P. Chow, "Privacy preserving multiple keyword search for confidential investigation of remote forensics," in 2011 Third International Conference on Multimedia Information Networking and Security. IEEE, 2011, pp. 595–599.
[186] S. Hou, T. Uehara, S.-M. Yiu, L. C. Hui, and K.-P. Chow, "Privacy preserving confidential forensic investigation for shared or remote servers," in 2011 Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE, 2011, pp. 378–383.
[187] B. Shebaro and J. R. Crandall, "Privacy-preserving network flow recording," digital investigation, vol. 8, pp. S90–S100, 2011.
[188] N. J. Croft and M. S. Olivier, "Sequenced release of privacy-accurate information in a forensic investigation," Digital Investigation, vol. 7, no. 1-2, pp. 95–101, 2010.
[189] N. H. Ab Rahman, W. B. Glisson, Y. Yang, and K.-K. R. Choo, "Forensic-by-design framework for cyber-physical cloud systems," IEEE Cloud Computing, vol. 3, no. 1, pp. 50–59, 2016.
[190] J.-L. Kruger and H. Venter, "State of the art in digital forensics for the internet of things," in International Conference on Cyber Warfare and Security. Academic Conferences International Limited, 2019, pp. 588–596.
[191] A. Iqbal, M. Ekstedt, and H. Alobaidli, Digital Forensic Readiness in Critical Infrastructures: A Case of Substation Automation in the Power Sector. Springer, 01 2018, pp. 117–129.
[192] CESG, "CESG Good Practice Guide No. 18," http://www.nationalarchives.gov.uk/documents/information-management/forensicreadiness.pdf, 2009, visited on 2021-10-14.
[193] A. Akilal and M.-T. Kechadi, "An improved forensic-by-design framework for cloud computing with systems engineering standard compliance," Forensic Science International: Digital Investigation, vol. 40, p. 301315, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666281721002407
[194] G. Grispos, W. B. Glisson, and K.-K. R. Choo, "Medical cyber-physical systems development: A forensics-driven approach," in 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017, pp. 108–113.
[195] N.-A. Le-Khac, D. Jacobs, J. Nijhoff, K. Bertens, and K.-K. R. Choo, "Smart vehicle forensics: Challenges and case study," Future Generation Computer Systems, vol. 109, pp. 500–510, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X17322422
[196] L. Daubner and R. Matulevičius, "Risk-oriented design approach for forensic-ready software systems," arXiv preprint arXiv:2106.10336, 2021.
[197] M. Elyas, A. Ahmad, S. B. Maynard, and A. Lonie, "Digital forensic readiness: Expert perspectives on a theoretical framework," Computers & Security,

vol. 52, pp. 70–89, 2015.
[198] L. De Marco, M.-T. Kechadi, and F. Ferrucci, "Cloud forensic readiness: Foundations," in International Conference on Digital Forensics and Cyber Crime. Springer, 2013, pp. 237–244.
[199] A. Mouhtaropoulos, P. Dimotikalis, and C.-T. Li, "Applying a digital forensic readiness framework: Three case studies," in 2013 IEEE International Conference on Technologies for Homeland Security (HST). IEEE, 2013, pp. 217–223.
[200] B. Endicott-Popovsky, D. A. Frincke, and C. A. Taylor, "A theoretical framework for organizational network forensic readiness." J. Comput., vol. 2, no. 3, pp. 1–11, 2007.
[201] S. Ngobeni, H. Venter, and I. Burke, "A forensic readiness model for wireless networks," in IFIP International Conference on Digital Forensics. Springer, 2010, pp. 107–117.
[202] K. A. Z. Ariffin and F. H. Ahmad, "Indicators for maturity and readiness for digital forensic investigation in era of industrial revolution 4.0," Computers & Security, vol. 105, p. 102237, 2021.
[203] M. Elyas, S. B. Maynard, A. Ahmad, and A. Lonie, "Towards a systemic framework for digital forensic readiness," Journal of Computer Information Systems, vol. 54, no. 3, pp. 97–105, 2014.
[204] A. Iqbal, M. Ekstedt, and H. Alobaidli, "Digital forensic readiness in critical infrastructures: A case of substation automation in the power sector," in International Conference on Digital Forensics and Cyber Crime. Springer, 2017, pp. 117–129.
[205] A. Alenezi, H. F. Atlam, and G. B. Wills, "Experts reviews of a cloud forensic readiness framework for organizations," Journal of Cloud Computing, vol. 8, no. 1, pp. 1–14, 2019.
[206] P. Turner, "Unification of digital evidence from disparate sources (digital evidence bags)," Digital Investigation, vol. 2, no. 3, pp. 223 – 228, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1742287605000575
[207] ——, "Selective and intelligent imaging using digital evidence bags," Digital Investigation, vol. 3, pp. 59 – 64, 2006, the Proceedings of the 6th Annual Digital Forensic Research Workshop (DFRWS '06). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S174228760600065X
[208] R. Eaglin and J. Craiger, "Data Sharing and the Digital Evidence Markup Language," in 1st Annual GJXDM Users Conference, Atlanta, GA.(not peer reviewed), 2005.
[209] S. S. Lee, T.-S. Park, S.-U. Shin, S.-K. Un, and D.-W. Hong, "A new forensic image format for high capacity disk storage," in Information Security and Assurance, 2008. ISA 2008. International Conference on. IEEE, 2008, pp. 399–402.
[210] B. N. Levine and M. Liberatore, "DEX: digital evidence provenance supporting reproducibility and comparison," Digital Investigation, vol. 6, pp. S48 – S56, 2009, the Proceedings of the Ninth Annual DFRWS Conference. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1742287609000395
[211] S. Garfinkel, "Digital forensics XML and the DFXML toolset," Digital Investigation, vol. 8, no. 3–4, pp. 161–174, 2012.
[212] E. Casey, G. Back, and S. Barnum, "Leveraging CybOX™ to standardize representation and exchange of digital forensic information," Digital Investigation, vol. 12, pp. S102–S110, 2015.
[213] W. Alink, R. Bhoedjang, P. A. Boncz, and A. P. de Vries, "XIRAF–XML-based indexing and querying for digital forensics," digital investigation, vol. 3, pp. 50–58, 2006.
[214] R. A. Bhoedjang, A. R. van Ballegooij, H. M. van Beek, J. C. van Schie, F. W. Dillema, R. B. van Baar, F. A. Ouwendijk, and M. Streppel, "Engineering an online computer forensic service," Digital Investigation, vol. 9, no. 2, pp. 96–108, 2012.
[215] B. Schatz, "Digital evidence: representation and assurance," Ph.D. dissertation, Queensland University of Technology, 2007.
[216] M. Cohen, S. Garfinkel, and B. Schatz, "Extending the advanced forensic format to accommodate multiple data sources, logical evidence, arbitrary in formation and forensic workflow," digital investigation, vol. 6, pp. S57–S68, 2009.
[217] A. Moser and M. I. Cohen, "Hunting in the enterprise: Forensic triage and incident response," Digital Investigation, vol. 10, no. 2, pp. 89 – 98, 2013, triage in Digital Forensics. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1742287613000285
[218] G. Osborne and B. Turnbull, "Enhancing computer forensics investigation through visualisation and data exploitation," in Availability, Reliability and Security, 2009. ARES'09. International Conference on. IEEE, 2009, pp. 1012–1017.
[219] G. Osborne, B. Turnbull, and J. Slay, "The 'explore, investigate and correlate' (EIC) conceptual framework for digital forensics information visualisation," in Availability, Reliability, and Security, 2010. ARES'10 International Conference on. IEEE, 2010, pp. 629–634.
[220] C. F. Tassone, B. Martini, and K.-K. R. Choo, "Visualizing digital forensic datasets: A proof of concept," Journal of forensic sciences, 2017.
[221] M. Irfan, H. Abbas, Y. Sun, A. Sajid, and M. Pasha, "A framework for cloud forensics evidence collection and analysis using security information and event management," Security and Communication Networks, vol. 9, no. 16, pp. 3790–3807, 2016. [Online]. Available:

http://dx.doi.org/10.1002/sec.1538
[222] M. Aupetit, Y. Zhauniarovich, G. Vasiliadis, M. Dacier, and Y. Boshmaf, "Visualization of actionable knowledge to mitigate drdos attacks," in Visualization for Cyber Security (VizSec), 2016 IEEE Symposium on. IEEE, 2016, pp. 1–8.
[223] O. Setayeshfar, C. Adkins, M. Jones, K. H. Lee, and P. Doshi, "Graalf: Supporting graphical analysis of audit logs for forensics," Software Impacts, vol. 8, p. 100068, 2021.
[224] Grafana, "Grafana 3.1.0 released," http://grafana.org/blog/2016/07/12/grafana-3-1-released.html, 2016, visited on 2022-12-01.
[225] Lewes Technology Consulting, "SOF-ELK Virtual Machine Distribution," https://github.com/philhagen/sof-elk/blob/master/VM_README.md,, 2019, visited on 2021-07-21.
[226] Plaso, "plaso," https://plaso.readthedocs.io/en/latest/,, 2019, visited on 2021-07-21.
[227] I. Homem, "Advancing automation in digital forensic investigations," Ph.D. dissertation, Department of Computer and Systems Sciences, Stockholm University, 2018.
[228] C. Alcaraz and S. Zeadally, "Critical infrastructure protection: Requirements and challenges for the 21st century," International Journal of Critical Infrastructure Protection, vol. 8, pp. 53–66, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874548214000791
[229] M. Naedele, "Addressing IT security for critical control systems," in System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on. IEEE, 2007, pp. 115–115.
[230] F. Adelstein, "Live forensics: diagnosing your system without killing it first," Communications of the ACM, vol. 49, no. 2, pp. 63–66, 2006.
[231] P. Taveras, "Scada live forensics: real time data acquisition process to detect, prevent or evaluate critical situations," European Scientific Journal, vol. 9, no. 21, 2013.
[232] H. Hadeli, R. Schierholz, M. Braendle, and C. Tuduce, "Leveraging determinism in industrial control systems for advanced anomaly detection and reliable security configuration," in Emerging Technologies & Factory Automation, 2009. ETFA 2009. IEEE Conference on. IEEE, 2009, pp. 1–8.
[233] K. Stouffer, M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, and S. Lightman, "NIST SP800-82 R3 (draft) guide to operational technology (OT) security," apr 2022. [Online]. Available: https://doi.org/10.6028%2Fnist.sp.800-82r3.ipd
[234] R. Minerva, A. Biru, and D. Rotondi, "Towards a definition of the Internet of Things (IoT)," IEEE Internet Initiative, Torino, Italy, 2015.
[235] K. Isoyama, Y. Kobayashi, T. Sato, K. Kida, M. Yoshida, and H. Tagato, "A scalable complex event processing system and evaluations of its performance," in Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems, ser. DEBS '12. New York, NY, USA: ACM, 2012, pp. 123–126. [Online]. Available: http://doi.acm.org/10.1145/2335484.2335498
[236] Z. Drias, A. Serhrouchni, and O. Vogel, "Analysis of cyber security for industrial control systems," in 2015 international conference on cyber security of smart cities, industrial control system and communications (SSIC). IEEE, 2015, pp. 1–8.
[237] L. Monostori, B. Kádár, T. Bauernhansl, S. Kondoh, S. Kumara, G. Reinhart, O. Sauer, G. Schuh, W. Sihn, and K. Ueda, "Cyber-physical systems in manufacturing," CIRP Annals, vol. 65, no. 2, pp. 621–641, 2016.
[238] K. Gai, L. Qiu, M. Chen, H. Zhao, and M. Qiu, "SA-EAST: security-aware efficient data transmission for ITS in mobile heterogeneous cloud computing," ACM Transactions on Embedded Computing Systems (TECS), vol. 16, no. 2, p. 60, 2017.
[239] M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E. K. Markakis, "A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues," IEEE Communications Surveys & Tutorials, vol. 22, no. 2, pp. 1191–1221, 2020.
[240] Infineon, NXP, STMicroelectronics, and ENISA, "Common Position On Cybersecurity," https://www.enisa.europa.eu/publications/enisa-position-papers-and-opinions/infineon-nxp-st-enisa-position-on-cybersecurity, Dec 2016.
[241] A. Chehri, I. Fofana, and X. Yang, "Security risk modeling in smart grid critical infrastructures in the era of big data and artificial intelligence," Sustainability, vol. 13, no. 6, p. 3196, 2021.
[242] I. Yaqoob, I. A. T. Hashem, A. Ahmed, S. A. Kazmi, and C. S. Hong, "Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges," Future Generation Computer Systems, vol. 92, pp. 265–275, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X18315644
[243] J.-P. A. Yaacoub, O. Salman, H. N. Noura, N. Kaaniche, A. Chehab, and M. Malli, "Cyber-physical systems security: Limitations, issues and future trends," Microprocessors and Microsystems, vol. 77, p. 103201, 2020.
[244] F. E. Grubbs, "Procedures for detecting outlying observations in samples," Technometrics, vol. 11, no. 1, pp. 1–21, 1969.
[245] P. Gogoi, D. Bhattacharyya, B. Borah, and J. K. Kalita, "A survey of outlier detection methods in network anomaly identification," The Computer Journal, vol. 54, no. 4, pp. 570–588, 2011.
[246] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly

detection: A survey," ACM Comput. Surv., vol. 41, no. 3, Jul. 2009. [Online]. Available: https://doi.org/10.1145/1541880.1541882

[247] Q. Fu, J.-G. Lou, Y. Wang, and J. Li, "Execution anomaly detection in distributed systems through unstructured log analysis," in 2009 ninth IEEE international conference on data mining. IEEE, 2009, pp. 149–158.

[248] C.-W. Ten, G. Manimaran, and C.-C. Liu, "Cybersecurity for critical infrastructures: Attack and defense modeling," IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 40, no. 4, pp. 853–865, 2010.

[249] J. Henriques, F. Caldeira, T. Cruz, and P. Simões, "Combining K-means and XGBoost models for anomaly detection using log datasets," Electronics, vol. 9, no. 7, p. 1164, 2020.

[250] I. . Security, "ISO27k Toolkit, ISMS Auditing Guideline, Version 2, 2017 ," http://www.iso27001security.com/html/27007.html, 2017, visited on 2017-06-01.

[251] K. Kent, S. Chevalier, T. Grance, and H. Dang, "Guide to integrating forensic techniques into incident response," NIST Special Publication, vol. 10, pp. 800–86, 2006, visited on 2017-06-01.

[252] North American Electric Reliability Corporation, "NERC Cyber Security Standards," https://www.nerc.com/pa/Stand/Pages/Cyber-Security-Permanent.aspx, 2016, visited on 2023-08-04.

[253] E. A. Morse and V. Raval, "PCI DSS: payment card industry data security standards in context," Computer Law & Security Review, vol. 24, no. 6, pp. 540–554, 2008.

[254] H. Benefield, G. Ashkanazi, and R. H. Rozensky, "Communication and records: Hippa issues when working in health care settings." Professional Psychology: Research and Practice, vol. 37, no. 3, p. 273, 2006.

[255] E. Hulitt and R. B. Vaughn, "Information system security compliance to fisma standard: a quantitative measure," Telecommunication Systems, vol. 45, no. 2, pp. 139–152, 2010.

[256] Institute of Internal Auditors Research Foundation, "International professional practices framework - implementation guide 2420 / quality of communications," Institute of Internal Auditors. Research Foundation, Tech. Rep., 2013. [Online]. Available: https://www.theiia.org/en/content/guidance/recommended/implementation/2420-quality-of-communications/

[257] S. Donaldson, S. Siegel, C. K. Williams, and A. Aslam, Enterprise cybersecurity: how to build a successful cyberdefense program against advanced threats. Apress, 2015.

[258] K. Fisler, S. Krishnamurthi, L. A. Meyerovich, and M. C. Tschantz, "Verification and change-impact analysis of access-control policies," in Proceedings of the 27th international conference on Software engineering, 2005, pp. 196–205.

[259] G.-J. Ahn, H. Hu, J. Lee, and Y. Meng, "Representing and reasoning about web access control policies," in 2010 IEEE 34th Annual Computer Software and Applications Conference. IEEE, 2010, pp. 137–146.

[260] K. Arkoudas, R. Chadha, and J. Chiang, "Sophisticated access control via smt and logical frameworks," ACM Transactions on Information and System Security (TISSEC), vol. 16, no. 4, pp. 1–31, 2014.

[261] R. Sabillon, J. Serra-Ruiz, V. Cavaller, and J. Cano, "A comprehensive cybersecurity audit model to improve cybersecurity assurance: The cybersecurity audit model (CSAM)," in 2017 International Conference on Information Systems and Computer Science (INCISCOS). IEEE, 2017, pp. 253–259.

[262] R. Agrawal, R. Bayardo, C. Faloutsos, J. Kiernan, R. Rantzau, and R. Srikant, "Auditing compliance with a hippocratic database," in Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30, ser. VLDB '04. VLDB Endowment, 2004, pp. 516–527. [Online]. Available: http://dl.acm.org/citation.cfm?id=1316689.1316735

[263] N. Kaaniche, M. Laurent, and C. Levallois-Barth, "Id-based user-centric data usage auditing scheme for distributed environments," Frontiers in Blockchain, vol. 3, p. 17, 2020.

[264] M. Bouet and M. Israël, "Inspire ontology handler: automatically building and managing a knowledge base for critical information infrastructure protection," in 12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops. IEEE, 2011, pp. 694–697.

[265] M. Lee, B. Hatfax, and J. Wingad, "Critical function monitoring and compliance auditing system," https://www.google.com/patents/US20070136814, Jun. 14 2007, uS Patent App. 11/299,049.

[266] S. Slapničar, T. Vuko, M. Čular, and M. Drašček, "Effectiveness of cybersecurity audit," International Journal of Accounting Information Systems, vol. 44, p. 100548, 2022.

[267] K. W. Ullah, A. S. Ahmed, and J. Ylitalo, "Towards building an automated security compliance tool for the cloud," in 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications. IEEE, 2013, pp. 1587–1593.

[268] M. Henze, R. Matzutt, J. Hiller, E. Mühmer, J. H. Ziegeldorf, J. v. d. Giet, and K. Wehrle, "Practical Data Compliance for Cloud Storage," in 2017 IEEE International Conference on Cloud Engineering (IC2E), April 2017, pp. 252–258.

[269] F. Doelitzscher, "Security audit compliance for cloud computing," Ph.D. dissertation, Plymouth University, 2014.

[270] N. Bjørner and K. Jayaraman, "Checking cloud contracts in microsoft azure," in International Conference on Distributed Computing and Internet Technology. Springer, 2015, pp. 21–32.
[271] IEC, "IEC 62443-2-1: Industrial communication networks – Network and system security Part 2-1: Establishing an industrial automation and control system security program," 2009.
[272] NIST, "Guide to Industrial Control Systems (ICS) Security," http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82.pdf, April 2013, visited on 2016-11-01.
[273] IEC, "IEC 62443 - IEC Technical Specification - Industrial communication networks - Network and system security - Part 1-1: Terminology, concepts and models," 2017.
[274] ISASecure, "Establishment of ISASecure Japanese Scheme and Publication of ISASecure Embedded Device Security Assurance Certification Program Specifications in Japan," http://www.isasecure.org/en-US/News-Events/Establishment-of-ISASecure-Japanese-Scheme-and-Pub, April 2013, visited on 2022-12-01.
[275] S. Majumdar, T. Madi, Y. Wang, Y. Jarraya, M. Pourzandi, L. Wang, and M. Debbabi, "Security compliance auditing of identity and access management in the cloud: Application to openstack," in 2015 IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom). IEEE, 2015, pp. 58–65.
[276] Amazon AWS, "Security at scale: Logging in aws," https://d1.awsstatic.com/whitepapers/compliance/AWS_Security_at_Scale_Logging_in_AWS_Whitepaper.pdf?did=wp_card&trk=wp_card, note = visited on 2022-01-17, 2022.
[277] J. M. Torres, F. O. Sveen, and J. M. Sarriegi, "Security strategy analysis for critical information infrastructures," in International Workshop on Critical Information Infrastructures Security. Springer, 2008, pp. 247–257.
[278] Y. Demchenko, P. Grosso, C. de Laat, and P. Membrey, "Addressing big data issues in Scientific Data Infrastructure," in 2013 International Conference on Collaboration Technologies and Systems (CTS), 2013, pp. 48–55.
[279] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation, 2011, visited on 2023-04-12.
[280] O. Hegazy, S. Safwat, and M. El Bakry, "A mapreduce fuzzy techniques of big data classification," in SAI Computing Conference (SAI), 2016. IEEE, 2016, pp. 118–128.
[281] A. E., Introduction to Machine Learning. 2nd ed. Cambridge, MA: MIT Press, 2010.
[282] H. M. Witten IH, Frank E, Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Series in Data Management Systems. Amsterdam: Morgan Kaufmann, 2011.
[283] R. Ranjan, "Streaming big data processing in datacenter clouds," IEEE Cloud Computing, vol. 1, no. 1, pp. 78–83, May 2014.
[284] S. Ghemawat, H. Gobioff, and S.-T. Leung, "The Google file system," in ACM SIGOPS operating systems review, vol. 37, no. 5. ACM, 2003, pp. 29–43.
[285] J. Dean and S. Ghemawat, "MapReduce: a flexible data processing tool," Communications of the ACM, vol. 53, no. 1, pp. 72–77, 2010.
[286] T. White, "Hadoop: the definitive guide: the definitive guide:"o'reilly media, inc."," 2009.
[287] S. K. Jensen, T. B. Pedersen, and C. Thomsen, "Time series management systems: A survey," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 11, pp. 2581–2600, 2017.
[288] M.-q. Wang, K. Wei, and C.-y. Jiang, "Survey of time series data processing in industrial internet," in 2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS). IEEE, 2019, pp. 736–741.
[289] European Commission, "Protection of personal data," https://commission.europa.eu/aid-development-cooperation-fundamental-rights/your-rights-eu/know-your-rights/freedoms/protection-personal-data_en, 2017, visited on 2013-07-28.
[290] Iamwire, "Big Data: 17 Predictions Everyone Should Read," http://www.iamwire.com/2016/11/big-data-17-predictions-everyone-should-read/145040, 2016, visited on 2022-12-01.
[291] R. Brighi, M. Ferrazzano, and L. Summa, "Legal issues in ai forensics: understanding the importance of humanware," on Applications of AI to Forensics 2020 (AI2Forensics 2020), p. 13, 2020.
[292] K. S. Hoon, K. C. Yeo, S. Azam, B. Shunmugam, and F. De Boer, "Critical review of machine learning approaches to apply big data analytics in ddos forensics," in 2018 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2018, pp. 1–5.
[293] O. Yavanoglu and M. Aydos, "A review on cyber security datasets for machine learning algorithms," in 2017 IEEE international conference on big data (big data). IEEE, 2017, pp. 2186–2193.
[294] N. Usman, S. Usman, F. Khan, M. A. Jan, A. Sajid, M. Alazab, and P. Watters, "Intelligent dynamic malware detection using machine learning

in ip reputation for forensics data analytics," Future Generation Computer Systems, vol. 118, pp. 124–141, 2021.

[295] R. T. Wiyono and N. D. W. Cahyani, "Performance analysis of decision tree c4. 5 as a classification technique to conduct network forensics for botnet activities in internet of things," in 2020 International Conference on Data Science and Its Applications (ICoDSA). IEEE, 2020, pp. 1–5.

[296] R. W. Moore and B. R. Childers, "Automatic generation of program affinity policies using machine learning," in Compiler Construction, R. Jhala and K. De Bosschere, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 184–203.

[297] A. Quiroz, M. Parashar, N. Gnanasambandam, and N. Sharma, "Autonomic policy adaptation using decentralized online clustering," in Proceedings of the 7th international conference on Autonomic computing, 2010, pp. 151–160.

[298] A. Pelaez, A. Quiroz, and M. Parashar, "Dynamic adaptation of policies using machine learning," in 2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), 2016, pp. 501–510.

[299] H. D. Johansen, E. Birrell, R. van Renesse, F. B. Schneider, M. Stenhaug, and D. Johansen, "Enforcing privacy policies with meta-code," in Proceedings of the 6th Asia-Pacific Workshop on Systems, ser. APSys '15. New York, NY, USA: Association for Computing Machinery, 2015. [Online]. Available: https://doi.org/10.1145/2797022.2797040

[300] O. Gheibi, D. Weyns, and F. Quin, "Applying machine learning in self-adaptive systems: A systematic literature review," ACM Trans. Auton. Adapt. Syst., vol. 15, no. 3, aug 2021. [Online]. Available: https://doi.org/10.1145/3469440

[301] J. O. Kephart and D. M. Chess, "The vision of autonomic computing," Computer, vol. 36, no. 1, pp. 41–50, 2003.

[302] D. Weyns, B. Schmerl, M. Kishida, A. Leva, M. Litoiu, N. Ozay, C. Paterson, and K. Tei, "Towards better adaptive systems by combining mape, control theory, and machine learning," in 2021 International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS). IEEE, 2021, pp. 217–223.

[303] A. Asquith and G. Horsman, "Let the robots do it!–taking a look at robotic process automation and its potential application in digital forensics," Forensic Science International: Reports, vol. 1, p. 100007, 2019.

[304] D. Hayes and M. Kyobe, "The adoption of automation in cyber forensics," in 2020 Conference on Information Communications Technology and Society (ICTAS). IEEE, 2020, pp. 1–6.

[305] K. C. Moffitt, A. M. Rozario, and M. A. Vasarhelyi, "Robotic process automation for auditing," Journal of emerging technologies in accounting, vol. 15, no. 1, pp. 1–10, 2018.

[306] R. Verma, J. Govindaraj, and G. Gupta, "DF 2.0: designing an automated, privacy preserving, and efficient digital forensic framework," Journal of Digital Forensics, Security and Law (JDFSL), 2018.

[307] A. Patrascu and V.-V. Patriciu, "Cyber protection of critical infrastructures using supervised learning," in 2015 20th International Conference on Control Systems and Computer Science. IEEE, 2015, pp. 461–468.

[308] M. Litoiu, I. Watts, and J. Wigglesworth, "The 13th cascon workshop on cloud computing: Engineering AIOps," in Proceedings of the 31st Annual International Conference on Computer Science and Software Engineering, ser. CASCON '21. USA: IBM Corp., 2021, p. 280–281.

[309] IBM, "IBM Pak for AIOps," 2022, visited on 2022-09-01. [Online]. Available: https://www.ibm.com/cloud/cloud-pak-for-watson-aiop

[310] P. Notaro, J. Cardoso, and M. Gerndt, "A systematic mapping study in AIOps," in Service-Oriented Computing – ICSOC 2020 Workshops, H. Hacid, F. Outay, H.-y. Paik, A. Alloum, M. Petrocchi, M. R. Bouadjenek, A. Beheshti, X. Liu, and A. Maaradji, Eds. Cham: Springer International Publishing, 2021, pp. 110–123.

[311] Z. Chen and Y. F. Li, "Anomaly detection based on enhanced DBScan algorithm," Procedia Engineering, vol. 15, pp. 178–182, 2011.

[312] H. Asif-Iqbal, N. I. Udzir, R. Mahmod, and A. A. A. Ghani, "Filtering events using clustering in heterogeneous security logs," Information Technology Journal, vol. 10, no. 4, pp. 798–806, 2011.

[313] I. Syarif, A. Prugel-Bennett, and G. Wills, "Unsupervised clustering approach for network anomaly detection," in International Conference on Networked Digital Technologies. Springer, 2012, pp. 135–145.

[314] A. J. Hoglund, K. Hatonen, and A. S. Sorvari, "A computer host-based user anomaly detection system using the self-organizing map," in Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on, vol. 5. IEEE, 2000, pp. 411–416.

[315] A. I. Hajamydeen, N. I. Udzir, R. Mahmod, and A. A. A. Ghani, "An unsupervised heterogeneous log-based framework for anomaly detection," Turkish Journal of Electrical Engineering & Computer Sciences, vol. 24, no. 3, pp. 1117–1134, 2016.

[316] G. Münz, S. Li, and G. Carle, "Traffic anomaly detection using K-means clustering," in GI/ITG Workshop MMBnet, 2007, pp. 13–14.

[317] L. Tian and W. Jianwen, "Research on network intrusion detection system based on improved

K-means clustering algorithm," in Computer Science-Technology and Applications, 2009. IFCSTA'09. International Forum on, vol. 1. IEEE, 2009, pp. 76–79.

[318] M. Eslamnezhad and A. Y. Varjani, "Intrusion detection based on MinMax K-means clustering," in Telecommunications (IST), 2014 7th International Symposium on. IEEE, 2014, pp. 804–808.

[319] R. Ranjan and G. Sahoo, "A new clustering approach for anomaly intrusion detection," arXiv preprint arXiv:1404.2772, 2014.

[320] Y. Liao and V. R. Vemuri, "Use of K-nearest neighbor classifier for intrusion detection," Computers & security, vol. 21, no. 5, pp. 439–448, 2002.

[321] H. Mohammed, N. Clarke, and F. Li, "Evidence identification in heterogeneous data using clustering," in Proceedings of the 13th International Conference on Availability, Reliability and Security, ser. ARES 2018. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: https://doi.org/10.1145/3230833.3233271

[322] A. Makanju, A. N. Zincir-Heywood, and E. E. Milios, "Investigating event log analysis with minimum apriori information," in Integrated Network Management (IM 2013), 2013 IFIP/IEEE International Symposium on. IEEE, 2013, pp. 962–968.

[323] S. Varuna and P. Natesan, "An integration of K-means clustering and naïve bayes classifier for intrusion detection," in Signal Processing, Communication and Networking (ICSCN), 2015 3rd International Conference on. IEEE, 2015, pp. 1–5.

[324] Z. Muda, W. Yassin, M. Sulaiman, and N. Udzir, "K-means clustering and naive bayes classification for intrusion detection," Journal of IT in Asia, vol. 4, no. 1, pp. 13–25, 2016.

[325] R. M. Elbasiony, E. A. Sallam, T. E. Eltobely, and M. M. Fahmy, "A hybrid network intrusion detection framework based on random forests and weighted K-means," Ain Shams Engineering Journal, vol. 4, no. 4, pp. 753–762, 2013.

[326] M. Du, F. Li, G. Zheng, and V. Srikumar, "Deeplog: Anomaly detection and diagnosis from system logs through deep learning," in Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1285–1298.

[327] R. C. Nickerson, U. Varshney, and J. Muntermann, "A method for taxonomy development and its application in information systems," European Journal of Information Systems, vol. 22, no. 3, pp. 336–359, 2013.

[328] P. Simões, T. Cruz, J. Gomes, and E. Monteiro, "On the use of Honeypots for detecting cyber attacks on industrial control networks," in Proc. 12th Eur. Conf. Inform. Warfare Secur. ECIW 2013, 2013.

[329] IRM, "Amateyrs attack technology. professional hackers target people," Website article, 2015.

[330] Y. Yang, H.-Q. Xu, L. Gao, Y.-B. Yuan, K. McLaughlin, and S. Sezer, "Multidimensional Intrusion Detection System for IEC 61850-Based SCADA Networks," IEEE Transactions on Power Delivery, vol. 32, no. 2, pp. 1068–1078, 2017.

[331] L. Rosa, T. Cruz, P. Simões, E. Monteiro, and L. Lev, "Attacking SCADA systems: a practical perspective," in IFIP/IEEE International Symposium on Integrated Network Management, May 2017.

[332] C. Curt and J.-M. Tacnet, "Resilience of critical infrastructures: Review and analysis of current approaches," Risk analysis, vol. 38, no. 11, pp. 2441–2458, 2018.

[333] N. Chowdhury and V. Gkioulos, "Cyber security training for critical infrastructure protection: A literature review," Computer Science Review, vol. 40, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574013721000010

[334] P. Gromek, "Strategic training and exercises for critical infrastructure protection and resilience: A transition from lessons learned to effective curricula," International Journal of Disaster Risk Reduction, vol. 65, p. 102647, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212420921006087

[335] L. Galbusera, M. Cardarilli, M. Gómez Lara, and G. Giannopoulos, "Game-based training in critical infrastructure protection and resilience," International Journal of Disaster Risk Reduction, vol. 78, p. 103109, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212420922003284

[336] The White House, "NATIONAL SECURITY PRESIDENTIAL DIRECTIVE INSPD-54," https://irp.fas.org/offdocs/nspd/nspd-54.pdf, 2008, visited on 2021-10-19.

[337] Council of the European Union, "Council Directive 2008/114/EC," https://eur-lex.europa.eu/eli/dir/2008/114/oj, 2008, visited on 2023-05-09.

[338] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-physical systems security—a survey," IEEE Internet of Things Journal, vol. 4, no. 6, pp. 1802–1831, 2017.

[339] ENISA, "Communication network dependencies for ICS/SCADA Systems," https://www.enisa.europa.eu/publications/ics-scada-dependencies/at_download/fullReport, December 2016, visited on 2017-03-01.

[340] R. Morabito, J. Kjällman, and M. Komu, "Hypervisors vs. lightweight virtualization: A performance comparison," in 2015 IEEE International Conference on Cloud Engineering, March 2015, pp. 386–393.

[341] J. Proença, T. Cruz, E. Monteiro, and P. Simões, "How to use software–defined networking to improve security—A survey," in Proc. 14th Eur. Conf. Cyber Warfare Security (ECCWS), 2015, p. 220.

[342] N. M. Karie and H. S. Venter, "Taxonomy of

challenges for digital forensics," Journal of forensic sciences, vol. 60, no. 4, pp. 885–893, 2015.
[343] I. Ghafir, V. Prenosil, J. Svoboda, and M. Hammoudeh, "A survey on network security monitoring systems," in 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW). IEEE, 2016, pp. 77–82.
[344] H. Guo, B. Jin, and D. Huang, "Research and review on computer forensics," in International Conference on Forensics in Telecommunications, Information, and Multimedia. Springer, 2010, pp. 224–233.
[345] R. A. Hansen, K. Seigfried-Spellar, S. Lee, S. Chowdhury, N. Abraham, J. Springer, B. Yang, and M. Rogers, "File toolkit for selective analysis & reconstruction (FileTSAR) for large-scale networks," in 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018, pp. 3059–3065.
[346] B. Data, "for better or worse: 90% of world's data generated over last two years," https://www.sciencedaily.com/releases/2013/05/130522085217.htm, 2013, visited on 2022-12-01.
[347] Y. Xie, D. Feng, Z. Tan, and J. Zhou, "Unifying intrusion detection and forensic analysis via provenance awareness," Future Generation Computer Systems, p. 26–36, 2016.
[348] C. Ware, Information visualization: perception for design. Morgan Kaufmann, 2019.
[349] Datavizcatalogue, "Data visualization catalogue," https://datavizcatalogue.com/, 2022, visited on 2021-12-01.
[350] W. B. Glisson, T. Storer, and J. Buchanan-Wollaston, "An empirical comparison of data recovered from mobile forensic toolkits," Digital Investigation, vol. 10, no. 1, pp. 44–55, 2013.
[351] S. O. Sandy Ryza, Uri Laserson and J. Wills, Advanced Analytics with Spark, PATTERNS FOR LEARNING FROM DATA AT SCALE. O'REILY, 2015.
[352] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero trust architecture," National Institute of Standards and Technology, Tech. Rep., 2020.
[353] F. Menges, T. Latzo, M. Vielberth, S. Sobola, H. C. Pöhls, B. Taubmann, J. Köstler, A. Puchta, F. Freiling, H. P. Reiser et al., "Towards GDPR-compliant data processing in modern SIEM systems," Computers & Security, vol. 103, p. 102165, 2021.

JOÃO HENRIQUES is a PhD candidate in Informatics Engineering in the area of Architectures, Networks and Cybersecurity at the University of Coimbra (UC), being a researcher at the Centre for Informatics and Systems of the same institution. He is Professor at the Department of Informatics Engineering of the Polytechnic University of Viseu (PUV), also having experience for more than 20 years in software engineering and management of R&D projects. His research interests include forensic and audit compliance for critical infrastructures protection.

FILIPE CALDEIRA is an Adjunct Professor at the Informatics Department of the Polytechnic Institute of Viseu, Portugal. He acts as program director of the Informatics Engineering program since 2014. He is also a researcher at the Centre for Informatics and Systems of the University of Coimbra and at the CISeD research centre of the Polytechnic Institute of Viseu. He has been recently involved in several international and national research projects. His main research interests include ICT security, namely, policy-based management, trust and reputation systems, Security and Critical Infrastructure Protection.

TIAGO CRUZ (SM'19) is Associate Professor in the Department of Informatics Engineering, University of Coimbra. His research interests include areas such as management systems for communications infrastructures and services, critical infrastructure security, broadband access network device and service management, Internet of Things, software-defined networking, and network function virtualization (among others). He is the author of more than 100 publications, including chapters in books, journal articles, and conference papers.

PAULO SIMÕES is Associate Professor at the University of Coimbra. His main research interests are Security, Network Management and Critical Infrastructure Protection. He has over 180 journal and conference publications in these areas. He has been regularly involved in several European- and industry-funded research projects, with both technical and management activities.