A Systematic Review of Secure IoT Data Sharing

Thanh Thao Thi Tran[1], Phu H. Nguyen[2][a] and Gencer Erdogan[2][b]
[1]*University of Oslo, Oslo, Norway*
[2]*SINTEF, Oslo, Norway*

Keywords: IoT, Data Sharing, Security, Privacy, Systematic Review.

Abstract: The Internet of Things (IoT) is more and more omnipresent. The greater values of the IoT can be realized by enabling data sharing between different stakeholders. However, one of the biggest challenges is ensuring security and enabling trust for IoT data sharing. In this paper, we identify state-of-the-art (SotA) approaches and techniques for secure IoT data sharing. We present high-level results emphasizing the SotA trend and revealing the most addressed domains, as well as more in-depth details such as procedures and methods used to preserve security in the data sharing environment. The blockchain technology, smart contracts, and InterPlanetary File System (IPFS) are among the most widely used approaches. As today's solutions explore a more decentralized approach to data sharing, there are several aspects to consider. Based on the findings, we have identified potential research directions for future work, including the differences between public and private blockchains, the combination of sharing and analytic, the value of data quality, and the importance of data management and governance.

1 INTRODUCTION

In recent years, the Internet of Things (IoT) has become omnipresent. The reasons are due to the possibilities of connecting everyday objects, such as cars, kitchen appliances, and baby monitors, to the Internet, allowing for seamless communication between processes, people, and things. With billions of IoT devices currently connected to the Internet, Cisco expects that 30 billion connected IoT devices will be available in the IoT market by 2023 (Grossetete, 2018). The IoT enables intelligent and inter-operable digital and physical things, individuals, and services, generating a variety of ecosystems capable of enabling secure cross-domain interactions.

Recent global important factors suggest further increase and expansion of the use of IoT. These factors are, amongst others, related to the environmental aspects and global pandemic diseases, such as the global sustainability goals and the COVID-19 pandemic. Amazon has taken the lead in initiating an ask of net-zero carbon emissions by 2040, whereas companies like Mercedes Benz and Microsoft have committed to the Climate Pledge[1]. To achieve such an ambitious goal, companies will need to measure carbon emissions in coming years. The IoT can, e.g., provide data-driven analytic, while revolutionizing heavy industries, and life sustainable industrial activities (Fantana et al., 2013). The advantages of the IoT were, for example, also demonstrated during the COVID-19 global pandemic (Peerzade, 2021). Besides empowering the control and statistics of the high case rate, IoT has played a role in monitoring virus-infected patients. With numerous countries have experienced lock-downs, the fear of new lock-downs will only increase and further fuel the importance and expansion of the IoT and its use cases.

Indeed, the IoT is more than just a collection of interconnected devices that provide significant benefits. Connecting enterprises, consumers, suppliers, or any other stakeholders in the IoT ecosystems can bring much higher value. Such greater values come from sharing data (Jernigan et al., 2016; Nguyen et al., 2022a) between stakeholders in the IoT ecosystems. One of the biggest challenges is ensuring security (Rajmohan et al., 2022; Rajmohan et al., 2020) and enabling trust for IoT data sharing (Atluri et al., 2020; Ullah et al., 2021; Siegel et al., 2018) as well as data quality (Nguyen et al., 2022b; Sen et al., 2022). While the topic of secure IoT data sharing is becoming more

[a] https://orcid.org/0000-0003-1773-8581
[b] https://orcid.org/0000-0001-9407-5748
[1] https://www.aboutamazon.com/news/sustainability/mercedes-benz-joins-the-climate-pledge

Tran, T., Nguyen, P. and Erdogan, G.
A Systematic Review of Secure IoT Data Sharing.
DOI: 10.5220/0011674200003405
In *Proceedings of the 9th International Conference on Information Systems Security and Privacy (ICISSP 2023)*, pages 95-105
ISBN: 978-989-758-624-8; ISSN: 2184-4356

important, it may still be only the end of the beginning. Few existing studies like (Lo et al., 2019; Al-Ruithe et al., 2019; De Prieëlle et al., 2020) have examined related topics of secure IoT data sharing such as data governance and blockchain solutions, but none has done a systematic literature review (SLR) of secure IoT data sharing.

Because the IoT is a continuously evolving topic, it is important to assess the current state-of-the-art addressing the advantages of IoT data sharing as well as the security considerations that must be taken into account. Following the most commonly used guidelines (Kitchenham and Charters, 2007), we have conducted an SLR on the state-of-the-art of secure IoT data sharing, and identified open issues and gaps that need to be addressed in the future. The contributions of our SLR are our answers to the following research questions: ***RQ1:*** *What are the solutions for secure IoT data sharing?* ***RQ2:*** *What are the security and trust aspects of these IoT data sharing approaches?* ***RQ3:*** *What are the current limitations of the IoT data sharing, and what are the open issues to be further investigated?* In addition to revealing the most addressed domains, our high-level results and statistical numbers emphasize the publication increase and trends. We did, however, obtain more in-depth information on the procedures and methods used to preserve security in the data sharing environment. Using blockchain technology and smart contracts, as well as the InterPlanetary File System (IPFS), a decentralized peer-to-peer storage system, are among them. These and other important discoveries do all contribute to an increase in secure IoT data sharing.

The remainder of this paper is structured as follows. Section 2 gives the details of our SLR approach. Then, we present the results of our SLR in Section 3, and compare our study with related work in Section 4. In Section 5, we discuss some possible threats to validity and give our conclusions in Section 6.

2 REVIEW PROCESS

Using online inquiry features of popular publication databases is the most notable approach to scan for primary studies when directing supplemental studies (Kitchenham and Charters, 2007). We used five popular publication databases, i.e., IEEE Xplore[2], ACM Digital Library[3], ScienceDirect[4], Scopus[5], and Web of Knowledge[6] to search for potential primary studies. These databases contain peer-reviewed articles and provide advanced search capacities.

Based on our research questions, we identified an initial set of keywords to construct our search strings. This initial set of keywords were then used to identify the top relevant papers in the search engines. Finally, by reviewing the top relevant papers, we identified a more fine-grained set of keywords that are related to the broad topics of data sharing, application domain, security, and governance. This was an iterative process where we also tested the keywords on the search engines to see if the top relevant papers appeared using our fine-grained set of keywords. The final set of search keywords are listed below, and we structured the keywords by following the guidelines from (Kitchenham and Charters, 2007), e.g., having groups of keywords to construct our search string using Boolean ANDs and ORs. The search query was adapted to fit the search engine of each publication database.

(*”data sharing”* **OR** *”sharing”* **OR** *”data exchange”* **OR** *”context sharing”* **OR** *”context aware data sharing”* **OR** *”context-sensitive information sharing”* **OR** *”sharing of data”* **OR** *”sharing data”* **OR** *”ecosystem”* **OR** *”marketplace”* **OR** *”data marketplace”*)

AND (*”Internet of Things”* **OR** *”IoT”* **OR** *”Industry 4.0”* **OR** *”smart cities”* **OR** *”smart city”* **OR** *”smart contract”* **OR** *”manufacturing”* **OR** *”energy”* **OR** *”supply chain”*)

AND (*”access control”* **OR** *”secure”* **OR** *”security”* **OR** *”trust”* **OR** *”trustworthy”* **OR** *”encryption”* **OR** *”data security”* **OR** *”secure communication”* **OR** *”secure data sharing”* **OR** *”context-aware security”* **OR** *”management”* **OR** *”governance”* **OR** *”protocols”* **OR** *”standards”*)

Our selection process is based on the predefined selection criteria (Section 2.2). We first filtered the candidate papers based on their titles and abstracts. When the titles and abstracts were not enough to decide whether to discard or keep the papers, we continued to skim and scan through the contents of these papers. When a candidate paper appeared in more than one database, we kept it, at first, in multiple search results. Then, we consolidated the outcomes in group discussions among the authors to cross-check the selection decisions and acquire the first set of primary studies with no duplicates. The review process is illustrated in Figure 1.

[2] https://ieeexplore.ieee.org

[3] https://dlnext.acm.org

[4] https://sciencedirect.com/

[5] https://scopus.com

[6] https://www.webofscience.com/

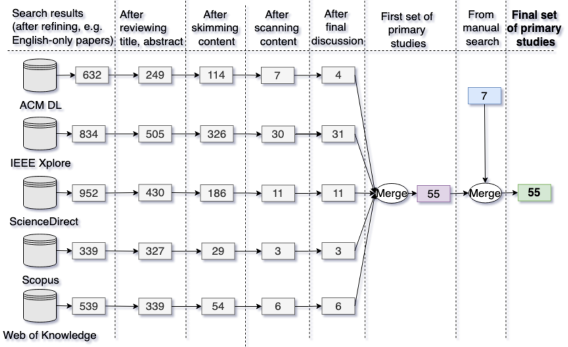

Figure 1: Overview of the search and selection steps.

2.1 Research Questions

This SLR aims to answer three RQs as explained in Section 1. RQ1 includes four sub-RQs. **RQ1.1 -** *What is the current trend of publications on secure IoT data sharing?* Answering this question allows us to see when the importance of secure data sharing became a topic of interest. **RQ1.2 -** *What are the reported application domains of IoT data sharing?* Answering this question allows us to identify the domains of interest in the IoT data sharing ecosystem. **RQ1.3 -** *What are the purpose and benefits of data sharing considered in the primary studies?* As we are interested in identifying the domains, we are also interested in the purpose of why different domains would like to share data and the advantages that could come from it. **RQ1.4 -** *What are the architectures for IoT data sharing in the primary studies?* Answering this question will give a high-level understanding of the architectures for IoT data sharing before we introduce in RQ2 the more in-depth findings on how different mechanisms and approaches have been utilized.

RQ2 has three sub-questions. **RQ2.1 -** *What are the most common threats and vulnerabilities to IoT data sharing today?* Answering this question would give an overview of current threats and vulnerabilities that come with IoT data sharing, which need to be considered. **RQ2.2 -** *What and how are the techniques and approaches used to preserve secure IoT data sharing?* After identifying the most common threats and vulnerabilities related to our scope, we extract how different contributions address these concerns. **RQ2.3 -** *What is the role of data management and governance, and how do their standards, policies, and guidelines support trusted and secure data sharing?* As data management and governance could contribute to secure IoT data sharing, the extraction of what policies that have been supported and how they preserve the security and privacy aspects is analyzed.

RQ3 does not consist of any sub-RQs. However, this RQ helps to express the current issues and suggest possible directions for future research.

2.2 Inclusion and Exclusion Criteria

Because the search strategy produced a wide range of primary studies with diverse content and outcomes, it was necessary to establish a set of inclusion and exclusion criteria that all primary papers had to meet. Our selection procedure was conducted in the most transparent and unbiased way possible, with all the primary studies having to meet all the following Inclusion Criteria (IC):

- (IC1) Must address IoT data sharing.
- (IC2) Must address IoT data sharing architecture models.
- (IC3) Must consider the security (& trust) aspects of IoT data sharing.

We excluded papers matching any of the Exclusion Criteria (EC):

- (EC1) Non-peer review publications are excluded.
- (EC2) Papers written in any other language than English will be filtered out and excluded in the search process.
- (EC3) Papers published before year 2010 will be excluded. According to (Wikipedia, nown), it is claimed that the IoT was ”born” around year 2008-2009, but did not gain popularity before year 2010, which is a good starting point to review.
- (EC4) Papers less than or equal to six pages (single column) and four pages (double column) are excluded. In general, short papers have shown to not include all necessary details. We therefore excluded these papers.

2.3 Search and Selection Strategy

The database search discovered 55 primary studies related to our topic (see Figure 1). Seven of these studies were discovered through the manual search and used as the test set for the database search. With the combination of both the database and manual search, we were left with a final set of 55 primary studies.

3 RESULTS

Table 1 presents the *fifty-five* (55) selected primary studies. We conducted in-depth evaluations and analyses of these studies in order to answer the research questions as follows.

Table 1: The primary studies*.

#	Year	PV	Title (click to open the corresponding publication)
#1	2017	C	Towards Blockchain-based Auditable Storage and sharing of IoT data
#2	2017	J	Secure and Efficient Data Sharing with Atribute-based Proxy Re-encryption Scheme
#3	2017	J	IoT data privacy via blockchains and IPFS
#4	2017	C	Big Data Model of Security Sharing Based on Blockchain
#5	2018	J	A Peer-to-Peer Architecture for Distributed Data Monetization in Fog Computing Scenarios
#6	2018	J	Continuous Patient Monitoring with a Patient Centric Agent: A Blockchain Architecture
#7	2018	C	Towards a Decentralized Data Marketplace for Smart Cities
#8	2018	C	Providing Context Aware Security for IoT Environments Through Context Sharing Feature
#9	2018	C	A Blockchain-Based Framework for Data Sharing With Fine-Grained Access Control in Decentralized Storage Systems
#10	2018	J	Smart-toy-edge-computing-oriented data exchange based on blockchain
#11	2019	C	Security and Privacy of Electronic Health Records: Decentralized and Hierarchical Data Sharing using Smart Contracts
#12	2019	J	Accelerating Health Data Sharing: A Solution Based on the Internet of Things and Distributed Ledger Technologies
#13	2019	J	MedChain: Efficient Healthcare Data Sharing via Blockchain
#14	2019	C	Toward a Decentralized, Trust-Less Marketplace for Brokered IoT Data Trading Using Blockchain
#15	2019	C	Blockchain for Secure and Efficient Data Sharing in Vehicular Edge Computing and Networks
#16	2019	C	Towards Multi-party Policy-based Access Control in Federations of Cloud and Edge Microservices
#17	2019	C	BlendMAS: A Blockchain-Enabled Decentralized Microservices Architecture for Smart Public Safety
#18	2019	C	BPIIoT: A Light-Weighted Blockchain-Based Platform for Industrial IoT
#19	2019	C	Enabling Industrial Data Space Architecture for Seaport Scenario
#20	2019	C	Blockchain based Proxy Re-Encryption Scheme for secure IoT Data Sharing
#21	2019	J	IoT Passport: A Blockchain-Based Trust Framework for Collaborative Internet-of-Things
#22	2020	C	BEAF: A Blockchain and Edge Assistant Framework with Data Sharing for IoT Networks
#23	2020	C	A Blockchain-based Medical Data Marketplace with Trustless Fair Exchange and Access Control
#24	2020	C	Blockchain Smart Contract for Scalable Data Sharing in IoT: A Case Study of Smart Agriculture
#25	2020	J	Fully Decentralized Multi-Party Consent Management for Secure Sharing of Patient Health Records
#26	2020	J	Secure data exchange between IoT endpoints for energy balancing using distributed ledger
#27	2020	J	BDSS-FA: A Blockchain-Based Data Security Sharing Platform With Fine-Grained Access Control
#28	2020	J	EdgeMediChain: A Hybrid Edge Blockchain-Based Framework for Health Data Exchange
#29	2020	C	Decentralized patient-centric data management for sharing IoT data streams
#30	2020	C	Blockchain-Based Multi-Role Healthcare Data Sharing System
#31	2020	J	Data Sharing System Integrating Access Control Mechanism using Blockchain-Based Smart Contracts for IoT Devices
#32	2020	J	Subscription-Based Data-Sharing Model Using Blockchain and Data as a Service
#33	2020	J	Towards a remote monitoring of patient vital signs based on iot-based blockchain integrity management platforms in smart hospitals
#34	2020	C	TrustedChain: A Blockchain-based Data Sharing Scheme for Supply Chain
#35	2020	J	A multi-layered blockchain framework for smart mobility data-markets
#36	2021	J	Blockchain-Driven Trusted Data Sharing With Privacy Protection in IoT Sensor Network
#37	2021	J	MedShare: A Privacy-Preserving Medical Data Sharing System by Using Blockchain
#38	2021	J	Fortified-Chain: A Blockchain-Based Framework for Security and Privacy-Assured Internet of Medical Things With Effective Access Control
#39	2021	C	A Cooperative Architecture of Data Offloading and Sharing for Smart Healthcare with Blockchain
#40	2021	C	ITrade: A Blockchain-based, Self-Sovereign, and Scalable Marketplace for IoT Data Streams
#41	2021	J	Proxy re-encryption enabled secure and anonymous IoT data sharing platform based on blockchain
#42	2021	J	Medi-Block record: Secure data sharing using block chain technology
#43	2021	J	PrivySharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities
#44	2021	J	AgriOnBlock: Secured data harvesting for agriculture sector using blockchain technology
#45	2021	J	BlockHealth: Blockchain-based secure and peer-to-peer health information sharing with data protection and right to be forgotten
#46	2021	J	BCHealth: A Novel Blockchain-based Privacy-Preserving Architecture for IoT Healthcare Applications
#47	2021	J	A conceptual IoT-based early-warning architecture for remote monitoring of COVID-19 patients in wards and at home
#48	2021	J	A blockchain-based trading system for big data
#49	2021	J	MedHypChain: A patient-centered interoperability hyperledger-based medical healthcare system: Regulation in COVID-19 pandemic
#50	2021	J	SmartMedChain: A Blockchain-Based Privacy-Preserving Smart Healthcare Framework
#51	2021	J	eHealthChain—a blockchain-based personal health information management system
#52	2021	J	A Threshold Proxy Re-Encryption Scheme for Secure IoT Data Sharing Based on Blockchain
#53	2021	J	A Blockchain-Based Medical Data Sharing Mechanism with Attribute-Based Access Control and Privacy Protection
#54	2021	J	FAST DATA: A Fair, Secure and Trusted Decentralized IIoT Data Marketplace enabled by Blockchain
#55	2021	J	Blockchain Assisted Secure Data Sharing Model for Internet of Things Based Smart Industries

*PV: Publication venue; J: Journal; C: Conference; * Because of space restriction for this paper, we can not cite the primary studies in the references.*

3.1 High-Level Details (RQ1)

Answering RQ1.1: Figure 2 illustrates the 55 primary studies that have been published since 2017, which according to the histogram, have been the most relevant to our specific topic of secure IoT data sharing (2020: 14; and 2021: 20 papers).

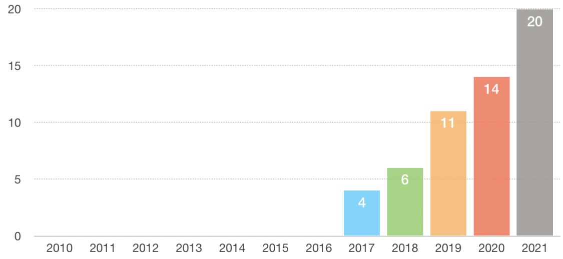

Figure 2: The publication years of the primary studies.

Answering RQ1.2: As an adaptable technology, IoT is utilized across multiple areas. We have divided the application domain according to IDS Data Space Radar[7] as follows: smart cities, manufacturing, energy, supply chain, automotive, and cross-domain/other. The application domain labeled ”Cross-domain/other” is the most dominant, represented by 44 out of 55 primary studies. The papers addressing cross-domain/other are divided into four subcategories: healthcare, surveillance, smart toys, and generic domains. Healthcare and the generic topic have a shared first place, being represented by 21 out of 55 primary studies. Unlike the topics of smart toys and surveillance, which are each addressed in a single study. For the other domains, energy is represented by one paper, while as for supply chain, manufacturing, and automotive, they are presented in two papers each, and smart city in four.

Answering RQ1.3: All the primary papers share a common purpose and goal in their studies and work: to develop a reliable and efficient data sharing solution that allows data owners and users to securely exchange their data while making data sources more accessible to authorized actors. There are some studies that only offer a broad overview of the purpose and benefits, while others delve further into the purpose and benefits of data sharing applicable specifically to the domain they address, such as papers #18 and #28 (see Table 1).

Our findings show that the majority of primary papers with publications addressing the healthcare domain deal with data sharing within the same system. The work by paper #28, on the other hand, may be categorized as addressing data exchange in cross-healthcare systems. Doctors are one of the stakeholders, as they give healthcare to patients, which implies data sharing within the same system. However, there is another stakeholder referred to as the ”requestor” of data. As the division of stakeholders is represented by doctor, patient, and requestor, it indicates that the ”requestor” may be from a different system.

Data sharing also has a significant impact in the manufacturing industry, which involves many different stakeholders, e.g., customers, employees, and supply-chain organizations. Paper #18 explains how the traditional manufacturing environment is complex, with various manufacturing data, e.g., equipment data, which is often stored in separate systems. Because these systems may belong to multiple service providers, and manufacturing organizations may not have direct control over this type of data and may be unable to comprehend the true and full value of the massive amount of data generated. As a result, the aim and benefits of sharing equipment data are explored in depth in this paper. The data on the equipment comprises not just their capacity, but also their status data. Data sharing can empower R&D, making manufacturing and distribution audits more effective, which assists production companies in reducing operating and manufacturing costs.

Answering RQ1.4: The extractions were categorized and filtered based on the data sharing models that were utilized. To emphasize, the differences between these three data sharing models are in terms of data management and governance perspectives. In domain-specific sharing, only stakeholders and interests from inside the domain are allowed to participate in the process. Peer-to-peer refers to the lack of a middleman; in other words, the absence of a central system in the solution. A marketplace, in contrast to peer-to-peer, implies the presence of a central system, at least to initiate the sharing process. Our findings reveal that peer-to-peer is the most common model utilized. To provide a more exact overview, the peer-to-peer model is addressed in 23 out of 55 primary studies, domain-specific sharing in 21 out of 55 primary studies, and the marketplace is represented in 11 out of the 55 primary studies we have analyzed.

In summary, it has been challenging to fully answer RQ1 in terms of today’s solution in the aspects of analytics. This is due to a lack of information as many studies do not cover or address the topic of data sharing in combination with analytic aspects, at least not in-depth. However, it is important to note that data analysis comes with great benefits, such as reducing the amount of analysis that other stakeholders must perform. With this in mind, one can debate whether data analysis before sharing data to a stake-

[7] https://internationaldataspaces.org/adopt/data-space-radar/

holder is restricting, in the sense that the stakeholder could have an interest in raw data for their own analysis tools and goals.

3.2 Low-Level Details (RQ2)

Answering RQ2.1 We evaluated the contributions against the OWASP top ten for the IoT domain[8]. The results show that the issues of *insecure network services*, *insufficient privacy protection*, and *insecure data storage and transfer* were the most addressed OWASP issues in regard to our scope of secure IoT data sharing. Our findings, however, reveal that the ten OWASP security concerns are not equally addressed. The majority of studies are focused on the three security threats outlined, whereas two studies deal with *insecure ecosystem interfaces*, one with *lack of secure update mechanism*, and six with *use of insecure or outdated components*. Furthermore, *weak, guessable, or hard-coded passwords*, *lack of device management*, *insecure default settings*, and *lack of physical hardening* have not been considered in any way. We may conclude that the issues of *insecure network services*, *insufficient privacy protection*, and *insecure data storage and transfer* have been the current most important factors to consider while using IoT data sharing. Nevertheless, the evidence gives us an overview that several OWASP threats and vulnerabilities have been neglected and overlooked.

We evaluated each study against security and privacy terms as well, to underline what security and privacy aspects are handled in the solution (see Figure 3). There are 20 studies addressing confidentiality; 20 studies addressing integrity; and 13 studies addressing availability. Authentication, on the other hand, is addressed in 22 studies, whereas authorization is represented in 23 studies. Privacy is the most frequently mentioned quality, appearing in 42 of the 55 primary studies. According to the extraction, only six papers explicitly address the CIA triad (Confidentiality, Integrity, and Availability). Because the CIA triad (Gollmann, 2010; Nguyen et al., 2015) forms the core foundation for developing secure information systems, a lack of awareness could lead to threats and vulnerabilities when applying the various IoT (data sharing) solutions to everyday life.

Answering RQ2.2: The main contributions of the studies show that many of the recent solutions for trusted IoT data sharing leverage the possibilities of blockchain technology. The findings show that 51 out of 55 primary studies utilize blockchain technology; 32 studies out of 55 primary studies utilize smart contracts; 10 papers address access control; and one paper, namely #19, utilizes the IDSA architecture. Although some articles use blockchain alone, the majority of the papers use it in combination with smart contracts, access control, identity management, or a mix among these.

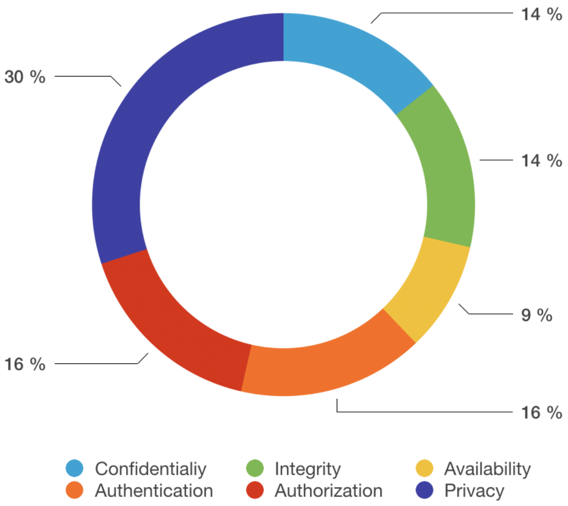

Figure 3: Security and its subgroups of specification.

The utilization of smart contracts varied from the studies. For example, paper #28 elaborates on how they use smart contracts to ensure automated regulation of rules and policies that govern access to the shared health data in a non-deniable way, whereas paper #27 uses the smart contract in two ways, the first one being for validation, while the other contract is for decryption purposes. Just like how smart contract solutions and use cases differ from study to study, so does access control management. Paper #18 utilizes a verification node, which is responsible for access management, with the access control policies written on the blockchain. In paper #28, they allow sellers of data to enforce their own access control policy on their encrypted records.

While 51 primary studies utilize blockchain technology, 17 of these papers address what kind of blockchain type they deploy. From the selection of primary studies, there are 9 studies that are taking advantage of the Hyperledger Fabric, while 8 studies are addressing Ethereum. Hyperledger Fabric is mentioned in papers #22 and #54, and Ethereum in papers #14, #20, #40, and #48 of the studies with solutions addressing the generic domain. In the healthcare domain, papers #33, #45, #49, #50, #51, and #53 leverage Hyperledger Fabric while papers #25, #28, and #37 use Ethereum. Paper #43 within the smart city domain use Hyperledger Fabric, whereas paper #26 in the domain of the smart grid use Ethereum. There are many possibilities with each one of the dif-

[8] https://owasp.org/www-pdf-archive/OWASP-IoT-Top-10-2018-final.pdf

ferent blockchain types. However, without any permission and the total transparency of Ethereum, the cost may affect the performance of scalability and privacy. Whereas for the Hyperledger, the consensus and access control have to be well-defined. The healthcare sector is a frequently mentioned domain in our primary studies, which tends to be associated with sensitivity and personal information. It is therefore worth looking into why some studies, specifically papers #25, #28, and #37 have been utilizing the public Ethereum platform in this domain, as the Ethereum platform consists of total transparency.

Paper #28 has implemented a prototype on the Ethereum blockchain to validate and evaluate the feasibility as well as the performance. They point out that the Ethereum blockchain consists of a large, global development community in addition to being completely open source and supporting a variety of use cases such as smart contracts and decentralized applications. In their solution, all functions executed in the contract are logged in the ledger. Therefore, they exude an energetic and positive view of Ethereum's transparency. As all transactions are recorded, the blockchain is considered hacker-resistant, which leads to it being more difficult to commit fraud within the system. Since everyone can join the Ethereum network, the solution assumes that the health data being shared is encrypted and anonymized to protect the privacy and real identity of patients.

As for the contributions by papers #25 and #37, they do combine the public Ethereum blockchain in combination with smart contracts as well. For the first one, a lot of the focus is related to the implementation of the prototype and, thereafter, the performance evaluation. As for the latter one, they do address that even though their contribution keeps all patient data secure using public-key infrastructure (PKI) cryptography, certain sensitive files cannot be shared on the public Ethereum network due to privacy issues. However, they do address one possible solution to this problem, which is to use forks of Ethereum that are permissioned and private, such as Quorum and Hyperledger Besu.

Not only do the trust aspects and mechanisms such as blockchain types have an effect on data sharing. It is also the layer that the data sharing is being done. We have been following the IoT World Forum Reference Model[9], which consists of seven layers. However, we will divide the seven layers into three layers, namely perception (physical devices and controllers), network (connectivity and edge computing), and application (data accumulation, data abstraction, application, collaboration and processes). There is a noticeable layer that stands out from the crowd, namely the network layer, which is the most leveraged layer to share data on from our findings across the selection of our primary studies. The network layer is presented 35 times, followed by the application layer being represented 17 times, and finally the perception layer being represented three times. The lack of coverage in the perception layer, especially the physical layer, raises the concern over the cross-domain physical-to-cyber attacks (Lanotte et al., 2020; Yampolskiy et al., 2012) that could make secure IoT data sharing meaningless.

To leverage the great benefits of data sharing, the technical aspects of how the data is stored need to be considered as well. The primary studies show that the current trend is to remove themselves from a traditional centralized solution and move to a more decentralized solution. With this in mind, the primary studies we have collected are introducing a distributed protocol for sharing and storing called the Interplanetary File System (IPFS) (IPFS, nown; Steichen et al., 2018). Compared with cloud storage, IPFS, as the peer-to-peer storage network it is (IPFS, nown), can contribute to preventing the problem of a single-point-of-failure.

More specifically, #3, #5, #7, #9, #24, #25, #27, #28, #30, #34, #39, #50 and #52 utilize the peer-to-peer IPFS data storage protocol, while #2, #20, #32, #38, #41 and #53 still utilize the traditional centralized cloud storage. The use of a third-party cloud service has been incorporated in some way or another, either as a primary storage of data, or just a back-up storage.

When storing the data, studies have pointed out the huge difference between storing raw data and encrypted data directly in their storage system. The findings show that a great number of selected primary studies have addressed this difference. There are 17 papers that utilize encryption. There are, however, only 11 papers addressing what type of encryption they utilize. From this number, study #2, #20, #25, #48, #52, and #54 utilize proxy re-encryption, while studies #9, #11, #27, #28, and #37 practice attribute-based encryption (ABE) in their solutions.

Answering RQ2.3: The data governance establish the policies and procedures, which the data management implements. A few studies look at the necessity for access control from the standpoint of the data owner, while none of them mention usage control (Lazouski et al., 2010). A few ideas, in particular, would allow data owners control over their own data. Papers #7 and #43 address GDPR, while paper #7 also ad-

[9] https://www.cisco.com/c/dam/global/en_ph/assets/ciscoconnect/pdf/bigdata/jim_green_cisco_connect.pdf

dresses the California Consumer Privacy Act. However, most of the primary studies do not go into as much detail or emphasis on the importance of data management and governance as we would like to see. In general, primary studies have revealed relatively little information about the role and impact of data management and governance.

Because the value of data sharing is derived from the data itself, it is important to assess the data's quality for the purposes of reuse and valid analysis. DNV are experts in assurance and risk management. They are using their expertise to improve safety and performance while also setting industry benchmarks, e.g., a framework for data quality assessment[10], which explains how to assess data quality. The framework emphasizes the importance of data quality assessment and the need for continual monitoring of the quality. ISO 8000[11] is a standard that provides approaches for managing, measuring, and improving the quality of data and information. However, none of the selected primary studies included inputs from DNV or ISO 8000, or other relevant data quality frameworks, demonstrating a lack of data quality management in the context of IoT data sharing.

To wrap up RQ2, there are many mechanisms and approaches utilized in the scope of secure IoT data sharing. However, when relating to the security aspects, there are only six contributions that cover the CIA principles. Even if other studies might cover the aspects of the CIA triad, contributions that lack elaboration and explanation have not been included. We can conclude that we have some shortcomings in the findings that contribute to our inability to fully answer RQ2. In particular, RQ2.3, in terms of the impact and role of data management and governance. This is because the majority of the primary studies we looked at did not include any specific standards, policies, or procedures that might have contributed to their solution for secure IoT data sharing. Even though GDPR was mentioned a few times, there was no in-depth explanation of how it was used as a contributing factor to secure the data sharing process.

3.3 Gaps and Limitations (RQ3)

This section discusses our findings to answer RQ3.

Limitations. A limitation that has been identified through a number of studies and findings is the lack of stress testing (*limitation1*). However, there are only four studies that explicitly address this limitation with their contribution, namely the contributing paper #3, #11, #19 and #33. Stress testing means testing against different use cases and seeing the results of utilizing a larger scale network. In addition, a variety of the studies operate on one blockchain platform, but do mention the wish to test their solution on other blockchain platforms to see if the results could have significant meaning. As the IoT technology is widely used and connects a variety of devices within different domains and areas, it would be useful to see results on the performance when putting the solution under stress.

Furthermore, another limitation of IoT data sharing is the concern about the heterogeneity nature of IoT (*limitation2*). As IoT technology makes human life more connected than ever over the Internet, it also implies the existence of a wide range of links from devices to multiple points such as endpoint devices, applications, and cloud platforms. More specifically, the studies show that the limitation occurs when combining and adding different technologies could lead to the growth of complexity.

Another limitation that is addressed is scalability (*limitation3*). The definition of scalability, when related to the domain of computer science, is the measure of a system's ability to increase or decrease in performance and cost in response to changes in application and system processing demands (Gartner, nown). When mentioning scalability as a limitation, it refers to the growth in the number of participants and transactions. As a side note, as the complexity of a system increases, it will have an effect on the scalability of the same system. A hand-full of solutions find it difficult to handle moving parts and components, in addition to the foundational scalability of a decentralized marketplace. To summarize, because the primary studies are concerned about the complexity of their solution, there will also be an indirect concern about scalability as well.

Last, we have the limitation related to the throughput and performance of the solutions (*limitation4*). One of the papers elaborates on their blockchain architecture, which is hardly supported by high throughput performance. More specifically, paper #10 revealed that the maximum throughput was estimated to be a maximum of 50,000 requests per second. Moreover, for (near) real-time secure IoT data sharing, especially involving dynamic contexts, heavy-weight blockchain-based solutions are less applicable. The advanced context-aware (de Matos et al., 2018) access control/usage control-based approaches should be more favorable. Such approaches can still employ blockchain technology for ensuring the trust-ability of IoT data sharing by logging only

[10] https://rules.dnv.com/docs/pdf/DNV/RP/2017-01/DNVGL-RP-0497.pdf

[11] https://www.iso.org/obp/ui/#iso:std:iso:8000:-61:ed-1:v1:en

the metadata of importance for the data sharing transactions like data quality metrics, the necessary headers of shared data. In general, as the IoT is often connected to critical infrastructure domains such as healthcare (Solanas et al., 2014), its performance would have significant meaning, where the outcome of the effects could be critical.

Open Issues. Throughout our extraction, paper #2, #7, #9, #29, and #38 mention more or less the open issues regarding their contributions in their papers. The other papers did not explicitly mention open issues, but we still identified and extracted some open issues. The main findings related to the open issues have been divided into four topics, namely GDPR, access policy, cost, and storage.

Paper #29 discusses permission-less data management that empowers patients to securely and selectively share their own personal data. This paper discusses data management but mentions their lack of focus on how to address the compliance of GDPR; for instance, the erasing of personal data (*issue1*). The open issues were stated by the authors and are specific to their paper. On the other hand, this is an overall issue that applies to IoT data sharing or IoT security in general, as elaborated on in our results.

Paper #9 discuss the data storage and sharing scheme in their contribution. They elaborate that their solution is safe only if the Ethereum blockchain network and the attribute-based encryption scheme are safe. They do mention that their contribution does not include an access policy update (*issue2*).

Paper #7 on decentralized data marketplace for smart cities do address the concern of cost (*issue3*). The concern about cost is associated with the transactions and the contracts used, which must be minimized to reduce the cost involved in using a decentralized data marketplace. They do, however, propose a solution that is worth investigating: a special short data format for storing metadata in the blockchain.

A common topic that turns out to be a recurring point in many, if not all the studies, is the security issues regarding data storage (*issue4*). Some of today's solutions utilize the cloud server, including paper #2, #20, #32, #38, #41 and #53. Even though the storage phase is done in combination with encryption, the solution utilizing a third-party solution still suffers from a single-point-of-failure.

To summarize, there should be further research on the security and privacy aspects of today's data sharing solutions, including the methods and strategies used to ensure secure IoT data sharing. By looking into the relationship between various trust and security mechanisms and methodologies as well as the security and privacy evaluation of studies, one may be able to uncover underlying reasons why, for instance, the CIA triad is not being addressed as much as desired. Additional future research should also be the elaboration of the significance and influence of data management and governance. For example, full analyses of current standards, policies, and guidelines, as well as explanations of how each standard or policy helps to data sharing security, should be supplied.

4 RELATED WORK

While the topic of secure IoT data sharing is becoming more important, it may still be only the end of the beginning. Few existing studies like (Lo et al., 2019; Al-Ruithe et al., 2019; De Prieëlle et al., 2020) are relevant to secure IoT data sharing, but none having the scope of our study, nor answering our questions.

Kuang Lo et al. (Lo et al., 2019) elaborate on the issues that data management solutions face, as well as the key issue of single-point-of-failure caused by the use of centralized management servers. Our study and (Lo et al., 2019) complement each other by pointing out technologies as a solution to the single-point-of-failure, such as blockchain as smart contracts. However, our work further extracted on the architectural layers (Rajmohan et al., 2022) used for the execution of data sharing, who the stakeholders in different domains are, and why data sharing is of interest to adapt for these stakeholders and domains.

Al-Ruithe et al. (Al-Ruithe et al., 2019) presents an SLR of data governance and cloud governance in their use of data. They highlight the need for more advanced research in data governance, in addition to suggesting areas for further research within data governance, which can be taken into account when conducting our research. However, they do not go into detail on the implications for IoT data sharing because their main focus is on data and cloud governance.

The necessity of ecosystem data governance for data platforms is discussed by Prieëlle et al. (De Prieëlle et al., 2020). Future research directions are elaborated, such as the importance of data platform governance in access and usage as a primary concern. They also emphasize that there is a lack of research on the many types of benefits that data sharing generates, which is an important future research direction as well. However, we focus on data sharing as the primary topic, with data governance and its impact on data sharing as a subtopic. Furthermore, they do not address various standards, policies, and guidelines that have been considered.

5 THREATS TO VALIDITY

In this section, we address the various concerns and risks to the validity of our SLR. The validity analysis is carried out by combining our work experience with the knowledge gathered from Kitchenham *et al.* (Kitchenham and Charters, 2007)'s principles.

One of the pitfalls when performing an SLR can be related to how the search process has been conducted. More specifically, in regard to the keywords and queries. The query can be used as proof of the validity and limitations of the study, as there are an infinite number of other keywords that could be included in the search query. However, to defend the selection of our search query, we performed a test case to have a handful of quality test papers that we believe are relevant and should be included in our research. These test papers were used to evaluate the quality of our search query results. If the search results in the electronic database X include all the test set papers published by X, we know that the other papers in the results are relevant in some way. Otherwise, we may conclude that there is a lot of noise in our search.

In addition, there was a need for filtering the results from IEEE Xplore as the result gave a number greater than 2000. We filtered the results to only papers related to the ”Internet of Things”. This can be considered a limitation, as there may be papers in the original IEEE Xplore result that could be of interest.

Another threat to our study is with regard to the selection of primary studies. The studies that have been included in our selection of primary studies from our SLR, by passing our inclusion and exclusion criteria, in addition to being relevant based on our predefined taxonomy, can still be doubtful. There might be relevant studies we have overlooked or publications we have missed out on during our search phase.

Last, the snowballing procedure could be employed as part of the search process, after the automatic and manual searches have been completed. Often, this method is used to enrich the set of primary studies. For possible additional primary studies, the technique comprises examining the list of references of the existing primary studies, and new publications citing them. Even though we did not go through this process, our database search was extensive and backed up by the test set from the manual search.

6 CONCLUSIONS

After systematically reviewing 55 primary studies relevant to the scope of secure IoT data sharing, we have discovered that the topic is getting hotter in recent years. In the IoT data sharing domain, centralized cloud-based solutions have traditionally dominated. This means that the solutions are relying on a single trusted third-party. However, there is a growing interest in utilizing blockchain's decentralized qualities. From our findings, the objectives behind the use of blockchain are: the removal of the centralized third party, immutability, improved data sharing, enhanced security, and reduced overhead costs in distributed applications.

Based on the synthesized data from analyzing the primary studies, we have been able to make the following remarks. (1) To make IoT data sharing more secure and valuable, based on our findings, we suggest looking into the gap between public versus private blockchain, and which is the most preferable blockchain technology type that should be utilized. It may also be worth analyzing the utilization of specific blockchain types in specific domains and the ripple effects that may occur. (2) As our findings show, there is a lack of research on the combination of sharing and analytics. However, there are numerous papers addressing the issue, which can occur when the data owner analyses the raw data and shares this analyzed data instead of the raw data. More specifically, there is further need for research to discover whether receiving already analyzed and processed data or unprocessed data makes a difference to stakeholders in the IoT ecosystem. (3) A clear standout is the topic of data management and governance and its lack of research. Future research should therefore highlight both the need and importance of data management and governance, in addition to addressing specific standards, policies, and guidelines that should be or are considered as well, and how they increase the security aspect in data sharing.

The great value of data sharing often derives from the data itself. Therefore, it is important that the data being sent and received is of high quality to ensure that analyses and the use of data are valid. Future research should emphasize data quality in the context of IoT data sharing, where consideration of ISO 8000 for data quality could be of interest. We further suggest collaboration across domains, as the combination of different data resources with data sharing across domains can enable IoT technology to reach even further and discover even more possibilities in the future.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Commission's H2020 Programme under the grant agreement numbers 958357 (InterQ), and 958363 (DAT4.Zero).

REFERENCES

Al-Ruithe, M., Benkhelifa, E., and Hameed, K. (2019). A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing*.

Atluri, V., Hong, Y., and Chun, S. A. (2020). Security, privacy and trust for responsible innovations and governance. In *The 21st Annual International Conference on Digital Government Research*, dg.o '20, page 365–366, New York, NY, USA. Association for Computing Machinery.

de Matos, E., Tiburski, R. T., Amaral, L. A., and Hessel, F. (2018). Providing context-aware security for iot environments through context sharing feature. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pages 1711–1715.

De Prieëlle, F., De Reuver, M., and Rezaei, J. (2020). The role of ecosystem data governance in adoption of data platforms by internet-of-things data providers: Case of dutch horticulture industry. *IEEE Transactions on Engineering Management*, pages 1–11.

Fantana, N. L., Riedel, T., Schlick, J., Ferber, S., Hupp, J., Miles, S., Michahelles, F., and Svensson, S. (2013). Iot applications—value creation for industry. *Internet of Things: Converging technologies for smart environments and integrated ecosystems*, page 153.

Gartner (Unknown). Scalability.

Gollmann, D. (2010). Computer security. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5):544–554.

Grossetete, P. (2018). IoT and the Network: What is the future?

IPFS (Unknown). What is IoT?

Jernigan, S., Kiron, D., and Ransbotham, S. (2016). Data sharing and analytics are driving success with iot. *MIT Sloan Management Review*, 58(1).

Kitchenham, B. A. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Software Engineering Group.

Lanotte, R., Merro, M., Munteanu, A., and Viganò, L. (2020). A formal approach to physics-based attacks in cyber-physical systems. *ACM Trans. Priv. Secur.*, 23(1).

Lazouski, A., Martinelli, F., and Mori, P. (2010). Usage control in computer security: A survey. *Computer Science Review*, 4(2):81–99.

Lo, S. K., Liu, Y., Chia, S. Y., Xu, X., Lu, Q., Zhu, L., and Ning, H. (2019). Analysis of blockchain solutions for iot: A systematic literature review. *IEEE Access*, pages 58822–58835.

Nguyen, H.-H., Phung, P. H., Nguyen, P. H., and Truong, H.-L. (2022a). Context-driven policies enforcement for edge-based iot data sharing-as-a-service. In *2022 IEEE International Conference on Services Computing (SCC)*, pages 221–230.

Nguyen, P. H., Kramer, M., Klein, J., and Traon, Y. L. (2015). An extensive systematic review on the model-driven development of secure systems. *Information and Software Technology*, 68:62–81.

Nguyen, P. H., Sen, S., Jourdan, N., Cassoli, B., Myrseth, P., Armendia, M., and Myklebust, O. (2022b). Software engineering and ai for data quality in cyber- physical systems - sea4dq'21 workshop report. *SIGSOFT Softw. Eng. Notes*, 47(1):26–29.

Peerzade, E. (2021). The role of IoT sensors in the COVID-19 fight.

Rajmohan, T., Nguyen, P. H., and Ferry, N. (2020). Research landscape of patterns and architectures for iot security: A systematic review. In *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pages 463–470.

Rajmohan, T., Nguyen, P. H., and Ferry, N. (2022). A decade of research on patterns and architectures for iot security. *Cybersecurity*, 5(1):1–29.

Sen, S., Husom, E. J., Goknil, A., Tverdal, S., Nguyen, P., and Mancisidor, I. (2022). Taming data quality in ai-enabled industrial internet of things. *IEEE Software*, 39(6):35–42.

Siegel, J. E., Kumar, S., and Sarma, S. E. (2018). The future internet of things: Secure, efficient, and model-based. *IEEE Internet of Things Journal*, 5(4):2386–2398.

Solanas, A., Patsakis, C., Conti, M., Vlachos, I. S., Ramos, V., Falcone, F., Postolache, O., Perez-martinez, P. A., Pietro, R. D., Perrea, D. N., and Martinez-Balleste, A. (2014). Smart health: A context-aware health paradigm within smart cities. *IEEE Communications Magazine*, 52(8):74–81.

Steichen, M., Fiz, B., Norvill, R., Shbair, W., and State, R. (2018). Blockchain-based, decentralized access control for ipfs. In *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pages 1499–1506.

Ullah, I., De Roode, G., Meratnia, N., and Havinga, P. (2021). Threat modeling—how to visualize attacks on iota? *Sensors*, 21(5):1834.

Wikipedia (Unknown). Internet of things.

Yampolskiy, M., Horvath, P., Koutsoukos, X. D., Xue, Y., and Sztipanovits, J. (2012). Systematic analysis of cyber-attacks on cps-evaluating applicability of dfd-based approach. In *2012 5th International Symposium on Resilient Control Systems*, pages 55–62.