# Systematic Literature Review on Data Provenance in Internet of Things

Emrullah Gultekin[1,2(✉)] and Mehmet S. Aktas[1]

[1] Computer Engineering Department, Yildiz Technical University, Istanbul, Turkey
aktas@yildiz.edu.tr
[2] Ziraat Teknoloji, Istanbul, Turkey
egultekin@ziraatteknoloji.com

**Abstract.** Internet of Things(IoT) is a concept that develops day by day and is now an indispensable part of our lives. Although it has been developed a lot, it still has many problematic areas and has many aspects that need improvement. On the other hand, data provenance is a concept that covers the origin of the data, the changes made, and the processes it has gone through. Data provenance is an essential option for improvements in the IoT field. Although there are many research studies in the literature related to the use of data provenance in the IoT, we could not identify a comprehensive systematic literature review that focuses explicitly on this topic. The study aims to examine the current data provenance on the Internet of Things studies, identify potential shortcomings or improvements, determine the working areas that can be done to make the processes more efficient and solve the IoT and data provenance problems. We also aim to serve this study as a knowledge base for researchers for conducting further research and development in this area. The review is done by following a systematic literature review process. We presented an SLR-based method on data provenance in IoT. We investigated the challenges encountered in IoT applications and worked on using data provenance as a solution. We conducted a literature search covering the years 2012–2022. We found 140 published papers. After applying exclusion criteria for these papers, we have focused on the final 16 papers. In the IoT area, some challenges require work, such as constrained resources, heterogeneity, scalability, mobility, security and privacy. Our study shows that data provenance is primarily used in security, privacy, reliability, and trust. We can say that the studies in these fields have increased gradually over the last years. Blockchain is frequently used to provide data provenance.

**Keywords:** Data provenance · Internet of Things · IoT · Systematic literature review

## 1 Introduction

The Internet of Things (IoT) refers to a rapidly growing technological concept in which devices can be connected and can communicate even if they are in different

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
O. Gervasi et al. (Eds.): ICCSA 2022 Workshops, LNCS 13380, pp. 31–46, 2022.
https://doi.org/10.1007/978-3-031-10542-5_3

structures and platforms. Depending on the advancement of technology, new sensors and devices have emerged, and sometimes changes have been made to existing devices, and new features have been added. Although such sensors and devices have already existed, a completely new approach has emerged with the development of the internet and the connection of these devices to the internet. This approach implies that any object can be accessed by anyone by any device in any place and condition. The Internet of Things has now entered every aspect of our lives, from our homes to our cars, outside to inside, and even our bodies. It is estimated that by 2030, organizations can generate value using data generated by 8 billion people and 25 billion connected devices [13]. This shows the importance and impact of IoT in our current lives and future. Its rapid growth causes some problems.

Provenance can find answers to what, where, when, how, and what for a data object. It enables us to obtain information about the state of an object and the processes that are made for it to come to this state, the stages it passes through, when and how these processes and stages occur [1]. In general, pure data is insufficient to understand, evaluate and reuse data. In addition to the data, meta data information such as the method in which the data is provided, the purpose of receiving the data, and the processes applied to the data are also required. IoT devices have resource limitations. These limitations may result in a device failure or service failure. Provenance capacities for all system elements can help resolve these issues [2]. Provenance makes many contributions, such as data clarity, reliability, and data reuse [12].

Using provenance data provides many benefits but can also cause some disadvantages. For example, a particular increase in network traffic is expected by adding provenance data to existing data. In addition, with this new data, the IoT, which already has limited resources, will result in memory, processor, and energy consumption in the devices. It is crucial to study how much these consumptions will be and whether they will create disadvantageous situations, and possible solutions that can be applied to eliminate or minimize disadvantageous situations.

Both the Internet of Things and the use of provenance are among the popular topics studied recently. This study investigated academic studies on data provenance in the Internet of Things.

## 2 Review Approach

### 2.1 Research Questions

Although there are studies on data provenance on the Internet of Things, there is a lack of detailed studies. There are review studies related to data provenance and IoT; for example, Bai et al. [24] conducted a systematic literature review, but they focused on security and the Internet of Health Things (IoHT), an extension of the IoT. We could not identify a comprehensive systematic literature review that focuses explicitly on the use of data provenance on the Internet of Things. This literature review aims to fill this gap to some extent, as there is no

comprehensive review of the use of data provenance in the field of the Internet of Things. For this purpose, the main difficulties encountered in the current Internet of Things applications and data provenance to solve these problems have been studied and it has been tried to find answers to the following research questions.

RQ1: How much research activity has there been in IoT and Data provenance since 2012? How often and in which journals are articles published? The frequency of publication will show a trend analysis on the usage of provenance in the Internet of Things issues.

RQ2: What research topics are being addressed? What are the main categories to which the issues are related? What are the main subjects, and what keywords do they use? The keywords used will show their relationship with other topics.

RQ3: What are the main problems being tried to find a solution for the Internet of Things using data provenance?

RQ4: What are the open issues and the future research directions of data provenance and the Internet of Things?

### 2.2 Identification of Relevant Articles

The study begins with searching in known digital libraries to find articles. For searches, popular digital libraries such as IEEE Xplore (https://ieeexplore.ieee.org/Xplore/home.jsp), ScienceDirect (https://www.sciencedirect.com), ACM (https://dl.acm.org/dl.cfm) and Scopus (https://www.scopus.com) were scanned. These libraries were searched with "(data provenance OR provenance) AND (Internet of Things OR IoT)" search criteria. The abstract section of the results was read by eliminating those not relevant. The articles related to data provenance in the Internet of Things were identified, and those that were more distant from the subject or those that were not suitable for the study were eliminated. The articles that conveyed the subject most efficiently were included in the study.

### 2.3 Inclusion and Exclusion Criteria

The criteria for the inclusion of articles in the research are as follows: Within the scope of the article, we tried to identify the articles directly related to the internet of things and data provenance. "Internet of Things AND Data Provenance", "Internet of Things AND Provenance", "IoT AND Data Provenance" and "IoT AND Provenance" search criteria were searched. In order to narrow the search results, the articles with document title or abstract or index terms containing these search criteria and published in the last 10 years (2012–2022) were selected.

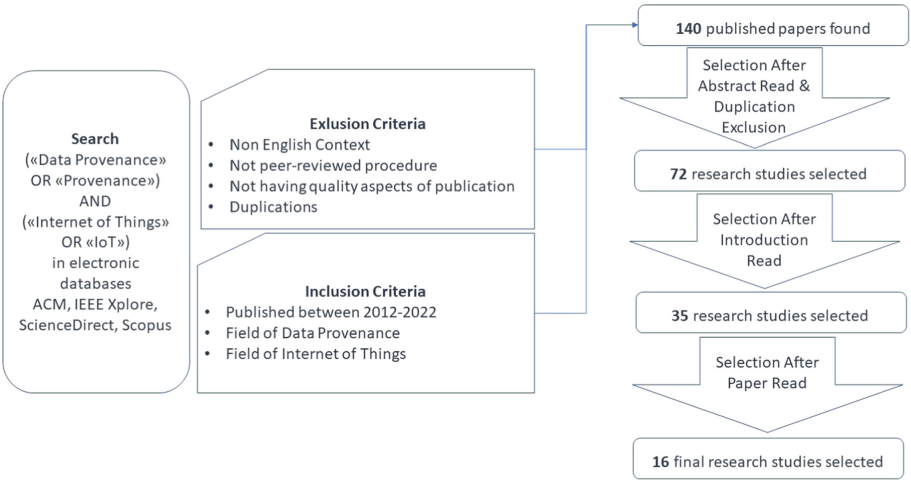


**Fig. 1.** Search progress

There are 27 articles searched by these search criteria from the IEEE Xplore database, 13 articles from the ScienceDirect database, 82 articles from the Scopus database, and 18 from the ACM database. 140 articles were found in total.

- 14 articles were filtered due to duplication.
- 54 articles were excluded after abstract reading because they were irrelevant.
- In total, 68 articles were filtered after abstract reading due to irrelevant to the data provenance in Internet of Things.
- 72 articles were selected to be read.
- 37 of 72 articles were eliminated after abstract and introduction reading because they were irrelevant.
- 35 articles were downloaded.
- After reading these 35 articles, 16 articles were selected to be studied.
- Only the manuscripts in English were included in the study.

### 2.4 Quality Assessment of Studies

When selecting the articles, it was put as a quality criterion that they were published in a well-known prestigious digital library and published after a particular examination. It was considered sufficient to be a primary research article evaluated and accepted by the referees and published in the last 10 years. No other criteria have been searched for.

### 2.5 Extraction and Synthesis of Data

The following data were extracted from studies that met the inclusion criteria, and the results were synthesized.

- Title
- Author
- Year of publication
- Publication database
- Data provenance and Internet of Things approaches discussed.
- Category of study
- Evaluation methods
- Implementation and simulation status
- Keywords used

## 3 Conducting the Review

### 3.1 Initial Search

Table 1 summarizes the search texts used in each digital library database. Inquiries were not made in the full text but on metadata fields such as article title, abstract and keywords (index terms).

**Table 1.** Search Strings

| Search String ID | Search String |
|---|---|
| SS1 | Internet of things and data provenance |
| SS2 | Internet of things and provenance |
| SS3 | IoT and data provenance |
| SS4 | IoT and provenance |

The database search took place on October 17th, 2021. The search was done for the last 10 years. The total number of papers from all databases was too much for a study. Then a preliminary screening of the search results was performed.

The keywords in the query texts return a large number of results. However, repetitive or irrelevant studies were eliminated from the query results. At the end of this phase, a total of 140 works remain. Figure-1 shows the search progress of the study.

# 4 Summary of Primary Studies

## 4.1 List of Primary Studies

**Table 2.** Selected papers

| Paper name | Paper Id |
|---|---|
| Using Provenance and CoAP to track Requests/Responses in IoT | P1 |
| Secure data provenance in Internet of Things based networks by outsourcing attribute based signatures and using Bloom filters | P2 |
| Data provenance to audit compliance with privacy policy in the Internet of Things | P3 |
| Secure data provenance in home energy monitoring networks | P4 |
| Light-Weight Security and Data Provenance for Multi-Hop Internet of Things | P5 |
| Towards a provenance collection framework for Internet of Things devices | P6 |
| Enhancing privacy in wearable IoT through a provenance architecture | P7 |
| Blockchain-based Data Provenance for the Internet of Things | P8 |
| Secure and efficient distributed network provenance for IoT: A blockchain- based approach | P9 |
| An index-based provenance compression scheme for identifying malicious nodes in multihop IoT network | P10 |
| BlockTrack-L: A lightweight blockchain-based provenance message tracking in IoT | P11 |
| Provenance-enabled packet path tracing in the RPL-based internet of things | P12 |
| A lightweight protocol for secure data provenance in the internet of things using wireless fingerprints | P13 |
| IoT big data provenance scheme using blockchain on hadoop ecosystem | P14 |
| Secure data provenance in internet of things using hybrid attribute based crypt technique | P15 |
| Provenance-based data flow control mechanism for internet of things | P16 |

The IoT vision, which benefits from the consistent advances in microelectronics, communications, and information technology, seems achievable. IoT-related applications and services are making more and more contributions to facilitating human life day by day. Therefore, this creates a diverse and large number of entities and services that require interaction with each other in various conditions and environments. The Internet of Things is expected to enable a world of intelligence and automation powered by seamless interactions between humans and their intelligent environments. The IoT offers many exciting solutions to problems encountered in various application areas. The potentially large number of

things, their diversity, their constraints, and the heterogeneous nature of communications encountered in the IoT create many significant challenges in terms of management, security, privacy, and interoperability, amongst many others [4]. Data provenance can be used for diverse purposes, such as evaluating the quality and reliability of data, trustworthiness, reproducibility of experiments, audit processes, data versioning, establishing property data, and discovering new data. However, scientists need to save provenance metadata during the data lifecycle stages to obtain the benefits of data provenance [11]. Provenance is not limited to data only. It is a crucial concept that provides features such as integrity, authenticity, and repeatability when it comes to digital information. These features help in detecting errors, analyzing errors, preventing unwanted behaviors and failures, and understanding the process of converting inputs into outputs, thus understanding the system [5].

This section summarizes studies on the use of data provenance in IoT applications. The studies are given by grouping according to their main categories. Sometimes a study may be close to more than one category. In this case, the closest category is selected. What we mean by the category is not the subject to be solved, but the main subject that the study focuses on. For example, in a study conducted in the field of security using blockchain for data provenance, security is the area where the solution is tried to be found, while blockchain is the method used. In this case, work can sometimes be categorized as blockchain and sometimes as security. The main focus of the study was chosen as the category.

**Blockchain:** Sigwart et al. [15] proposed a generic blockchain-based data provenance framework for the Internet of Things which can be applied to various use cases. They evaluated the framework concerning the defined requirements using an implementation with Ethereum smart contracts. An example of using provenance data in vaccine procurement is discussed. This example aims to determine whether the values, such as the temperature of the vaccine, go out of the desired value and thus whether the vaccine is spoiled or not with the provenance data. In another usage scenario, an example of provenance is given in a system where patient values are monitored, and assistants or health personnel are warned in case of an emergency. They proposed a generic data provenance architecture.

Siddiqui et al. [16] have proposed a permissioned blockchain solution for maintaining secure data provenance, utilizing outsourced attribute-based encryption. They use partial signatures to reduce the overhead of authentication and blockchain mechanisms from the IoT node to the Edge node.

Porkodi et al. [23] proposed a framework for securing data provenance in IoT systems, privacy and access control policies, and cryptographic techniques in the blockchain. They used Hybrid Attribute-Based Encryption (HABE) as a blockchain cryptographic technique. The proposed hybrid attribute-based encryption assures data security.

Liu et al. [21] inspected security challenges for operating distributed network provenance for the Internet of Things. They utilized the blockchain as the essential architecture for storing and querying cross-domain provenance data. They

also presented a unified model for efficiently instantiating provenance query modules in the verifiable computation framework.

Pacooh et al. [22] proposed a secure framework that is blockchain-enabled. The framework is for large-scale IoT data storage. They considered Edge computing to be merged to facilitate the management of the authentications of the small IoT devices and perform data storage. They also deployed a lightweight mutual authentication scheme that enables performing authorization and authentication of IoT devices. They presented the implementation of the proposed security scheme for IoT data to ensure data provenance, verifiability, integrity, and traceability.

**Security:** Liu et al. [18] proposed an index-based provenance compression algorithm. It adopts the idea of typical substring matching, combined with path identifier and path index to represent the path information in the data provenance, thereby reducing the size of data provenance. Similarly, they amplify the data provenance scheme to attack detection and propose a based malicious node identification.

Aman et al. [14] used an analytical model to develop a mechanism to establish data provenance in IoT systems. They also proposed a lightweight security protocol for data provenance. They use physical unclonable functions (PUFs) and fingerprints extracted from the wireless channel, mutual authentication, and anonymity to achieve data provenance.

Kaku et al. [6] presented studies on provenance in the IoT field, it is emphasized that most of the studies aim to provide authenticity, trust, and quality of data by following the source of the data. Furthermore, it is stated that provenance is still an important issue when IoT considers it. It is necessary to look for answers such as who, timestamps, periods of processes, and why, where, and when from the data provenance IoT framework. In the IoT CoAP area, they discussed the use of provenance for request/response calls from a machine perspective. In addition, they presented the results by performing response time and throughput values and performance values.

In the study conducted by Chia et al. [3], A method for home energy monitoring networks that provides cross-check validation to prevent fraud in energy consumption is presented. With this method, it is aimed to prevent the manipulation of smart meters and to ensure that the correct measurement is transmitted. The proposed solution tried to validate the data sent via an independent sensor placed with a smart plug. Thus, it has been tried to provide secure data provenance. Their work investigated provenance models to reduce data integrity attacks in IoT systems. To identify the origin of the data, they used PUFs and wireless link fingerprints. They ensure the integrity of an information's origin since PUFs are difficult to clone. No provenance model and framework were used in the study.

Nwafor et al. [8] emphasize the need for integrating provenance into the Internet of Things system. They propose a provenance collection framework that provides provenance collection capabilities for devices in an Internet of

Things system. They record IoT devices' I/O data transformations and generate provenance data. In the confidential procurement of origin data, they presented a framework for collecting provenances traced data can be converted into provenance logs. With a sensor track with different configurations, their built-in solution achieves data lineage for IoT devices. The integrity of lineage logs is neglected.

Kamal and Tariq [7] introduced a lightweight solution for security and data provenance in a multi-hop IoT environment. They created link fingerprints using the Received Signal Strength Indicator (RSSI) value of the IoT devices. They used the correlated coefficient based on matching the link fingerprints to identify the path to the source node. To ensure security without an authorization mechanism, they relied on the physical parameters of RSSI. In this case, RSSI values can be easily manipulated. The system is vulnerable to external attacks.

In the study of Siddiqui et al. [10], It is stated that the use of data provenance is necessary for IoT devices. However, it is emphasized that these data can be manipulated from outside, and therefore it is crucial to use secure data provenance mechanisms. Since IoT devices have resource constraints, they focus on the appropriate solution. A secure data provenance mechanism based on attribute-based encryption has been proposed. In this mechanism, ciphertext policy attribute-based encryption (CP-ABE) is used instead of public-key encryption. A bloom filter stores provenance information to reduce storage costs due to resource constraints on IoT devices. In order to reduce the computation load of IoT devices, operations that require more computation were performed on edge nodes. The system was tested in a simulation environment, and the results were evaluated.

**Data Flow:** Rong-na et al. [19] proposed a provenance-based data flow control mechanism that uses provenance data to implement direct and indirect data flow control. Thus, they aimed to reduce the risk of indirect leakage. Provenance data is becoming more and more prominent as there is data flow between different devices; therefore, indirect control on the provenance data takes an increasing amount of time, critically affecting control performance. They simplified provenance data according to the type of operation generating the data and the association degree of data, and their provenance data; thus, they wanted to increase the control efficiency of the data flow. In order to reduce the storage space and improve the efficiency of querying the provenance data, they also proposed a provenance tree that records the relationship between data and their provenance data. They demonstrated the implementation and optimization method of provenance-based data flow control.

**Data Trustworthiness:** To ensure data trustworthiness during packet traversal, Suhail et al. [20] proposed a Provenance-enabled Packet Path Tracing scheme for RPL-connected IoT devices. They introduced a node-level provenance by embedding sequence numbers against the routing entry at the routing table of the respective forwarding node. They also added a system-level provenance.

This provenance type included destination and source node IDs and was used for capturing the complete packet trace. They aimed to verify the provenance with node-level and system-level sources.

**Privacy:** Lomotey et al. [17] proposed a broadcast-subscriber IoT model where users' data only is shared with healthcare facilities or user-authorized devices and nodes. For this purpose, they proposed two main approaches. In the first approach, uniquely identifiable components in communication devices encrypted with meta-data level encryption techniques and keys are known only to devices in the subscription pool. This hardware and software level encryption technique means that privacy breaches are minimal, as demonstrated in various experiments. It is tested with various encryption-decryption algorithms, including AES, DES, 3DES, MD5, and SHA. The second approach proposes the provenance technique in broadcast subscriber IoT architecture. In this way, transparency and a full digital audit trail are targeted. Provenance information from the data source to its current state facilitates the audit of the IoT data. The interconnectedness of Wearable IoT Devices is also investigated to determine who receives what data in the IoT architecture. The proposed system was tested with several evaluations, and its contributions were presented.

Pasquier et al. [9] examine provenance to provide privacy policy and control mechanisms in the IoT area. They explain the idea of the study as follows: Provenance can be analyzed to determine whether the implementation mechanisms and the objectives of the regulations are met. Expected behaviors and

**Table 3.** Category of Papers

| Paper Id | Category | Implementation | Simulation |
|---|---|---|---|
| P1 | Network | ✓ | ✓ |
| P2 | Security | ✓ | ✓ |
| P3 | Privacy Policy | ✗ | ✗ |
| P4 | Security, Privacy | ✓ | ✓ |
| P5 | Security | ✓ | ✓ |
| P6 | Data collection | ✓ | ✗ |
| P7 | Privacy | ✓ | ✓ |
| P8 | Blockchain | ✓ | ✗ |
| P9 | Trust, Blockchain | ✓ | ✓ |
| P10 | Security | ✓ | ✓ |
| P11 | Blockchain, Security | ✓ | ✓ |
| P12 | Data trustworthiness, Routing | ✓ | ✓ |
| P13 | Authentication | ✓ | ✓ |
| P14 | Big data, Blockchain | ✓ | ✓ |
| P15 | Security, Blockchain | ✓ | ✓ |
| P16 | Data flow | ✓ | ✓ |

actual behaviors can be reported to improve the implementation mechanism. The difficulties encountered in obtaining provenance data were also discussed.

## 5 Results and Discussion

In this section, the study results are presented and discussed regarding the aim. Answers to the research questions determined for the study were presented.

RQ1: How much research activity has there been in IoT and Data provenance since 2012? How often and in which journals are articles published? The frequency of publication will show a trend analysis on the usage of provenance in the Internet of Things issues.

When the frequency of studies and the databases from which the studies are located are analyzed, it can be seen that data provenance and IoT issues have been studied more in recent years.

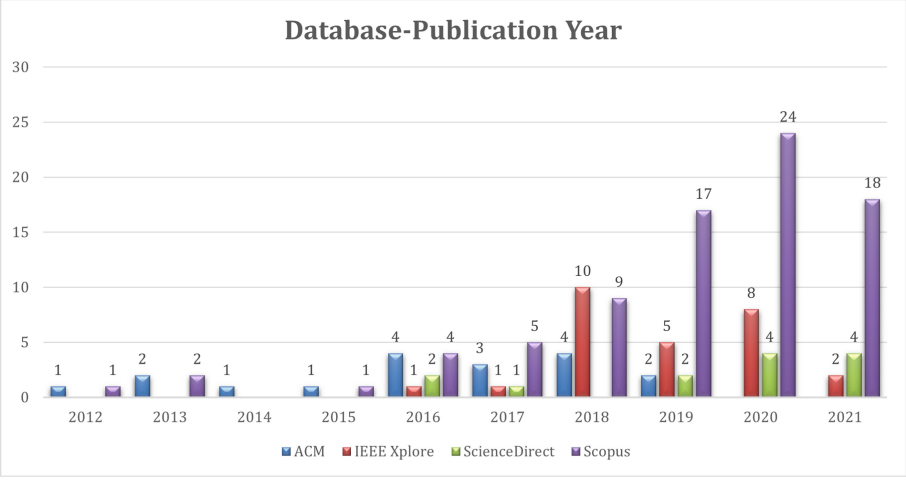


**Fig. 2.** Publication years of the papers

Figure-2 shows a graph of working years and journals. In 2012, there were 1 IEEE Xplore and 1 ACM work in total 2, 3 in 2013, 2 in 2015,11 in 2016, 10 in 2017, 23 in 2018, 26 in 2019, 36 in 2020 and 24 in 2021 140 articles in total. When evaluating these data, it should be considered that 2021 has not yet been completed. Therefore, there may be articles in progress and are in the process of publication. These data show an increasing trend, especially in the number of studies in recent years. There is a steady increase here until 2021. It should not be overlooked that the rates here are the rates of the papers we work with. Considering the studies that we did not include, different rates may occur. However, in general, it is possible to observe that data provenance studies in the field of IoT have increased recently.

A total of 140 articles were found from the 4 researched databases. 27 of 140 articles were found in IEEE Xplore, 13 in ScienceDirect, 82 in Scopus, and 18 in ACM. Some of these articles (14) were found in more than one database. Detailed information was described in the Search Strategy section.

As a result, there is an increasing trend in the field of data provenance in IoT with the increasing popularity of these fields.

RQ2: What research topics are being addressed? What are the main categories to which the issues are related? What are the main subjects, and what keywords do they use? The keywords used will show their relationship with other topics.

Table-3 shows the category of selected 16 articles and the work done. The article states whether there is Implementation and whether the results are evaluated by simulation.

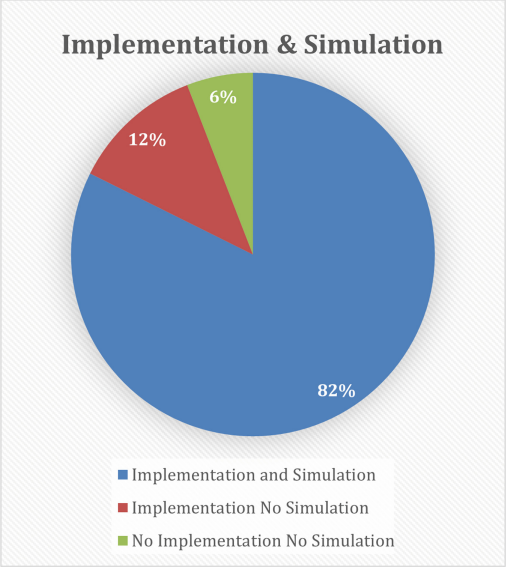


**Fig. 3.** Implementation/Simulation

Figure-3 shows the implementation and simulation status of selected papers. Studies have been conducted in security, privacy, network, request/response, verification, data trustworthiness, privacy policy, blockchain, trust, data collection, and authentication. These are related studies. Data provenance and IoT are the main fields of study. IoT and data provenance issues such as security, privacy, reliability, confidentiality, and trust have been frequently addressed.

The keywords used will show their relationship with other topics. The keywords used in conjunction with data provenance and IoT can give information about the areas in which these studies are carried out. Figure-4 shows the distribution of most keywords used in the articles studied. From the perspective

of keywords, we can see that Internet of Things (IoT) and Data Provenance (Provenance) keywords are mostly used. This is natural because these keywords are our search criteria. By looking at other commonly used keywords, we can understand the topics covered in this field. Blockchain, Security and Cloud Computing are used frequently. We can say that the studies in these areas are more intense. The Other keywords used are:Smart Contract, Authentication, Link Quality Indicator, Wireless Channel Characteristics, Physical Unclonable Functions, Attack Detection, Compression, Cloud-based IoT, Privacy, Bloom Filter, Sensors, Lightweight Authentication, Light-weight Signature Generation, Mobile Devices, Middleware, Wearables, Hyperledger Fabric, Attribute Based Encryption, Hadoop, Big Data, Traceability, Link-fingerprints, Light-weight, Multihop, Data Trustworthiness, Access Control, LLN, URI, REST, RPL, Cryptography, Sensor Networks, 6LowPAN, Home Area Networks, Location Authenticity, Smart Grid, IPv6, Distributed Network Provenance, Trust, Ciphertext Policy Attribute Based Encryption, Hybrid attribute based crypt technique, History, Lossless provenance, Data Models, Temperature sensors, Computational Modeling, Meta Data, COAP

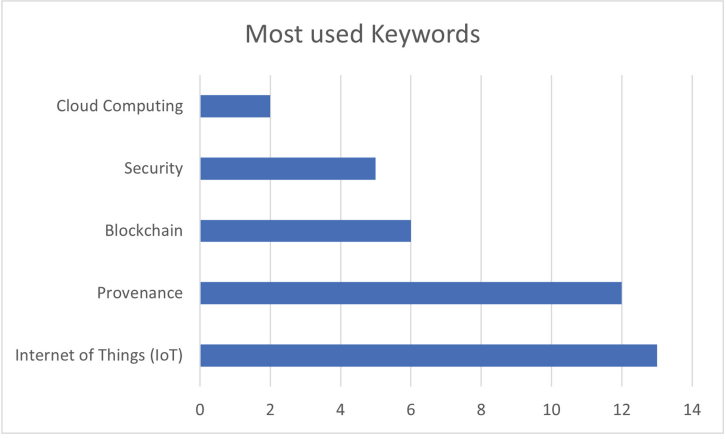


**Fig. 4.** Keywords most used

RQ3: What are the main problems being tried to find a solution for the Internet of Things using data provenance?

Data provenance is primarily used to solve security, privacy, reliability, and trust-related problems. We can say that the studies in these fields have increased gradually towards the last years. Blockchain is frequently used to provide data provenance.

RQ4: What are the open issues and the future research directions of data provenance and the Internet of Things?

In the IoT and data provenance studies, security and privacy are often the natural subjects of the study, even if it is not the main topic of the study. IoT

devices have some security difficulties due to resource constraints. Therefore, it may be necessary to take additional precautions to ensure security in IoT devices. Privacy is an essential issue in IoT applications because it often senses information in these environments. Data provenance can provide security, reliability, and trust in IoT devices from a data provenance perspective. However, data provenance can provide these plus features while at the same time creating a disadvantageous situation because data provenance provides features such as security, reliability, trust, not only the current state of the data but also the origin, the stages, and processes it undergoes, the information such as when and by whom these processes are performed. Therefore, this information may also create privacy issues. Security mechanisms such as encryption and authentication are required to restore privacy. The issues of security, reliability, trust, and privacy in IoT devices and data provenance are open issues that will continue to be worked on in the future.

## 6 Conclusions and Future Work

This review presented an SLR-based method on data provenance in IoT. We investigated the challenges encountered in IoT applications and worked on the use of data provenance as a solution. We conducted a literature search covering the years 2012–2022. We found 140 published papers between these years. After applying exclusion criteria for these papers, we have focused on the final 16 papers. In the IoT area, some challenges require work such as constrained resources, heterogeneity, scalability, identity, mobility, searching and discovering, security and privacy. In our studies, we have seen that data provenance is mainly used in security, privacy, transparency, reliability, confidentiality, and verification. We can say that the studies in these fields have increased gradually towards the last years. Security, reliability, trust, and privacy are currently being studied and needed in the future. Regarding the SLR-based method, we may not have analyzed all available studies. Non-English, non-peers were reviewed, and editorial articles, book chapters, and survey articles were excluded. In this review, we conducted a comprehensive survey of approaches to using data provenance in IoT through the findings of more than 100 authors and different studies. However, given the increasing number of studies in this area, it is impossible to ensure that all studies are considered. Our research ended in March 2022.

**Acknowledgement.** The authors would like to express to thank to Ziraat Teknoloji for their support. This study was produced from the first author's doctoral thesis, prepared under the supervision of the second author.

## References

1. Aktas, M.S., Astekin, M.: Provenance aware run-time verification of things for self-healing Internet of Things applications. Concurrency Computat. Pract. Exper. **31**, e4263 (2019). https://doi.org/10.1002/cpe.4263

2. Dundar, B., Astekin, M., Aktas, M.S.: A big data processing framework for self-healing internet of things applications. In: 2016 12th International Conference on Semantics, Knowledge and Grids (SKG), pp. 62–68 (2016). https://doi.org/10.1109/SKG.2016.017
3. Ming H.C., Sye L.K., Tang, Z.: Secure Data Provenance in Home Energy Monitoring Networks. In: Proceedings of the 3rd Annual Industrial Control System Security Workshop (ICSS 2017), pp. 7–14. ACM, New York (2017). https://doi.org/10.1145/3174776.3174778
4. Elkhodr, M., Alsinglawi, B., Alshehri, M.: Data provenance in the internet of things. In: 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), Krakow, pp. 727–731 (2018). https://doi.org/10.1109/WAINA.2018.00175
5. Kaku, E., Orji, R., Pry, J., Sofranko, K., Lomotey, R., Deters, R.: Privacy Improvement Architecture for IoT. In: 2018 IEEE International Congress on Internet of Things (ICIOT), San Francisco, CA, pp. 148–155 (2018). https://doi.org/10.1109/ICIOT.2018.00028
6. Kaku, E., Lomotey, R.K., Deters, R.: Using Provenance and CoAP to track Requests/Responses in IoT. Proc. Comput. Sci. **94**, 144–151 (2016). https://doi.org/10.1016/j.procs.2016.08.023, ISSN 1877–0509
7. Kamal, M., Tariq, S .: Light-Weight Security and Data Provenance for Multi-Hop Internet of Things. IEEE Access **6**, 34439–34448 (2018). https://doi.org/10.1109/ACCESS.2018.2850821
8. Nwafor, E., Campbell, A., Hill, D., Bloom, G.: Towards a provenance collection framework for Internet of Things devices. In: 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), San Francisco, CA, pp. 1–6 (2017). https://doi.org/10.1109/UIC-ATC.2017.8397531
9. Pasquier, T., Singh, J., Powles, J., Eyers, D., Seltzer, M., Bacon, J.: Data provenance to audit compliance with privacy policy in the Internet of Things. Pers. Ubiquit. Comput. **22**(2), 333–344 (2017). https://doi.org/10.1007/s00779-017-1067-4
10. Siddiqui, M.S., Rahman, A., Nadeem, A., Alzahrani, A.M.: Secure data provenance in internet of things based networks by outsourcing attribute based signatures and using bloom filters. Int. J. Adv. Comput. Sci. Appli. (IJACSA) **10**(5), 221–226 (2019). http://dx.doi.org/10.14569/IJACSA.2019.0100529
11. de Silva, D.L. Batista, A., Corrêa, P.L.P.: Data Provenance in Environmental Monitoring. In: 2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Brasilia, pp. 337–342 (2016). https://doi.org/10.1109/MASS.2016.050
12. Jensen, S., Plale, B., Aktas, M.S., Luo, Y., Chen, P., Conover, H.: Provenance capture and use in a satellite data processing pipeline. IEEE Trans. Geosci. Remote Sens. **51**(11), 5090–5097 (2013). https://doi.org/10.1109/TGRS.2013.2266929
13. Nitschke, P., Williams, S.P.: Conceptualizing the internet of things data supply. Proc. Comput. Sci. **181**, 642–649 (2021). https://doi.org/10.1016/j.procs.2021.01.213, ISSN 1877–0509
14. Aman, M.N., Basheer, M.H., Sikdar, B.: A lightweight protocol for secure data provenance in the internet of things using wireless fingerprints. IEEE Syst. J. **15**(2), 2948–2958 (2021). https://doi.org/10.1109/JSYST.2020.3000269
15. Sigwart, M., Borkowski, M., Peise, M., Schulte, S., Tai, S.: Blockchain-based Data Provenance for the Internet of Things. In: 9th International Conference on the

Internet of Things (IoT 2019), Bilbao, Spain, October 22–25, 8 pages. ACM, New York (2019). https://doi.org/10.1145/3365871.3365886
16. Siddiqui, M.S., Syed, T.A., Nadeem, A., Nawaz, W., Albouq, S.S.: BlockTrack-L: a lightweight blockchain-based provenance message tracking in IoT. Int. J. Adv. Comput. Sci. Appli. (IJACSA) **11**(4) (2020). https://doi.org/10.14569/IJACSA.2020.0110462
17. Lomotey, R.K., Sofranko, K., Orji, R.: Enhancing privacy in wearable iot through a provenance architecture. Multimodal Technol. Interact. **2**, 18 (2018). https://doi.org/10.3390/mti2020018
18. Liu, Z., Wu, Y.: An index-based provenance compression scheme for identifying malicious nodes in multihop IoT network. IEEE Internet Things J. **7**(5), 4061–4071 (2020). https://doi.org/10.1109/JIOT.2019.2961431
19. Rong-na, X., Hui, L., Guo-zhen, S., Yun-chuan, G., Ben, N., Mang, S.: Provenance-based data flow control mechanism for Internet of things. Trans. Emerging Tel. Tech. **32**, e3934 (2021). https://doi.org/10.1002/ett.3934
20. Suhail, S., Hussain, R., Abdellatif, M., Pandey, S.R., Khan, A., Hong, C.S.: Provenance-enabled packet path tracing in the RPL-based internet of things. Comput. Netw. **173**, 107189 (2020). https://doi.org/10.1016/j.comnet.2020.107189, ISSN 1389-1286
21. Liu, D., Ni, J., Huang, C., Lin, X., Shen, X.S.: Secure and efficient distributed network provenance for IoT: a blockchain-based approach. IEEE Internet Things J. **7**(8), 7564–7574 (2020). https://doi.org/10.1109/JIOT.2020.2988481
22. Honar Pajooh, H., Rashid, M.A., Alam, F., Demidenko, S.: IoT Big Data provenance scheme using blockchain on Hadoop ecosystem. J. f Big Data **8**(1), 1–26 (2021). https://doi.org/10.1186/s40537-021-00505-y
23. Porkodi, S., Kesavaraja, D.: Secure data provenance in Internet of Things using hybrid attribute based crypt technique. Wireless Pers. Commun. **118**(4), 2821–2842 (2021). https://doi.org/10.1007/s11277-021-08157-0
24. Baogang, B., Shah, N., Yuhe, B., Amir, A.: Security and provenance for Internet of Health Things: a systematic literature review. J. Soft. Evolution Process **33**, e2335 (2021). https://doi.org/10.1002/smr.2335