Privacy-Aware in the IoT Applications: A Systematic Literature Review

Faiza Loukil, Chirine Ghedira, Benharkat Aïcha-Nabila, Khouloud Boukadi, Zakaria Maamar

Privacy-aware in the IoT applications: A systematic literature review

Faiza Loukil[1], Chirine Ghedira-Guegan[1], Aïcha Nabila Benharkat[2], Khouloud Boukadi[3], and Zakaria Maamar[4]

[1] University of Lyon, CNRS, IAE - University of Lyon 3, LIRIS, UMR5205, France
[2] University of Lyon, CNRS, INSA Lyon, LIRIS, UMR5205, France
firstName.lastName@liris.cnrs.fr
[3] Mir@cl Laboratory, Sfax University, Tunisia; khouloud.boukadi@fsegs.usf.tn
[4] Zayed University, Dubai, U.A.E; zakaria.maamar@zu.ac.ae

Abstract. The Internet of Things (IoT) emerged as a paradigm in which smart things collaborate among them and with other physical and virtual objects using the Internet in order to perform high level tasks. These things appear in a variety of application domains, including smart grid, health care and smart spaces where several parties share data in order to tackle specific tasks. Data in such domains are rich in sensitive data and data owner-specific habits. Thus, IoT raises concerns about privacy and data protection. This paper reports on a systematic literature review of privacy preserving solutions used in Cooperative Information Systems (CIS) in the IoT field. To do so, and after retrieving scientific productions on the subject, we classify the results according to several facets. In this paper, we consider a subset of them: (i) data life cycle, (ii) privacy preserving techniques and (iii) ISO privacy principles. We combine the facets then express and analyze the results as bubble charts. We analyze the proposed solutions in terms of the techniques they deployed and the privacy principles they covered according to the ISO standard and the data privacy laws and regulations of the European Commission on the Protection of Personal Data. Finally, we identifies recommendations to involve privacy principle coverage and security requirement fulfillment in the IoT applications.

1 Introduction

The rapid growth of the Internet of Things (IoT) technology has resulted into different advances in the IoT field that are affecting both businesses and persons. In general, IoT applications, such as smart grid and smart cities require the collaboration of several parties in order to achieve their goals. These parties can be data owners or requesters, including an individual, a group of individuals or an organization. For instance, the different parties can share their energy consumption in order to help energy provider to predict its energy production.

However, despite the bright side of IoT, several concerns continue to undermine its adoption. In fact, collecting data in IoT applications increases the

data owner's worries about the potential uses of these data. In fact, some of the collected data can be sensitive and the data owner wish not share them with other competitor organizations without retaining some level of control. Thus, this work focuses on one non-functional requirement of IoT applications, which is privacy protection for the collaborating parties.

As part of our research agenda on privacy in the IoT era, we deem necessary conducting a comprehensive analysis on this topic. To this end, we provide an overview of existing IoT privacy preserving solutions in order to identify gaps and come up with solutions and recommendations. This overview is the result of a systematic literature review.

According to [20], a systematic literature review consists of five interdependent steps including (i) choose a research scope by defining research questions, (ii) retrieve candidate papers by querying different scientific databases, (iii) select relevant papers that can be used for answering the research questions by defining inclusion and exclusion criteria, (iv) define a classification scheme by analyzing the abstracts of the selected papers to identify the terms that will be used as categories for classifying the papers, and (v) produce a systematic literature review by sorting papers according to the classification scheme.

Our objective is to identify open issues and trends regarding privacy preserving in the IoT applications. Therefore, our classification scheme consists of six facets that stress out application domains, IoT architectures, security properties and requirements, data life cycle, and privacy preserving techniques. We also define an additional facet to identify the ISO privacy principle that the IoT solutions consider.

This paper is organized as follows. Section 2 gives a general idea about IoT, security and privacy concepts. Section 3 describes our systematic review study. Section 4 identifies the recommendations in order to involve privacy principle coverage and security requirement fulfillment in the IoT applications. Section 5 discusses existing reviews that study privacy issue in the IoT applications. Section 6 concludes the paper and presents some future endeavors.

2 Internet of things, security, and privacy

In this section, we shed light on: IoT, security, and privacy by discussing their definitions, the existing IoT application domains and architectures as well as the security properties and requirements according to the ISO standard. Afterwards, we present the existing privacy legislation and the privacy preserving techniques.

2.1 Some definitions

Commonly agreed definitions of IoT, security and privacy do not exist. Thus, we summarize those that are deemed relevant for our work.

The Internet of Things (IoT) is a network of physical objects that contain embedded technology to communicate with the external environment [12]. According to Guillemin et al. [13], IoT connects people and things at anytime,

anyplace, with anything and anyone, ideally using any path or network and any service.

Security involves the application and management of appropriate measures that involve consideration of a wide range of threats. In this context, ISO standard [14] defines a set of security properties and requirements detailed in Section 2.3.

Privacy is "*the claim of individuals, groups, or institutions to decide for themselves when, how and to what extent information about them is communicated to others*" [25]. With the IoT applications, it is important to consider the context when dealing with privacy issue. Data privacy is about data security and while taking into account requirements from legal regulations and individual preferences [6].

2.2 Application domains and architectures of IoT

Different application domains in IoT exist. We categorize these applications into two domains:

1. **Personal and home**: including (i) location sharing, which the aim of providing services based on the collected location information (i.e. geographical position) of IoT terminals, (ii) health care, which consists of offering care monitoring services without necessary visiting hospitals and (iii) smart home, which automates the ability to control smart devices around the house.
2. **Government and industry**: including the smart city, which monitors all its critical infrastructures and smart grid, which allows grid monitoring in order to reduce energy consumption. Such IoT applications need collaboration between several parties in order to fulfill their ends. For instance, house, office and industry consumers should be aware of the collaboration benefits in a smart grid to reduce energy consumption. Besides, the smart grid should provide the adequate privacy protection for the collaborating members to reassure them.

Moreover, we distinguish four types of architecture that are: centralized, decentralized, third party and hybrid architecture.

1. **Centralized Architecture**: it is where all the entities in the network are passive: their only task is to provide data. The collected data will be stored, processed by a central server which is the only server that provides IoT services to the other entities [21]. The main challenge with this architecture is resilience. In fact, all the computation tasks are managed by a single server. Thus, in case of server failure, the IoT services will be unavailable.
2. **Decentralized Architecture**: each entity can process data and provide IoT services to other entities in the network. Moreover, the decentralized architecture overcomes the single point of failure issue of the centralized architecture. In fact, a failure in one entity in the network will not affect the whole system. However, malicious entities intrusion arises because any entity can connect with any other entity at any time.

3. **Third party Architecture**: it is where a public institution or a private corporation is responsible for data collection, transfer, storage, and/or processing. Example of ready to use platform is the Smart-Meter-Analytics (SAP) [2]. The main challenge with such architecture is that it gives full trust to the third party for the whole data management.
4. **Hybrid Architecture**: it consists of combining several architectures in order to take advantage of the existing architecture structures and overcome their disadvantages. For instance, Birman et al. [7] addressed the privacy issue in smart grid data collection phase by combining peer-to-peer communications with some elements of centralized control in order to help utilities to effectively use the collected data while preserving the consumers' privacy.

2.3 Security properties and requirements

According to ISO standard [14], the purpose of information security is to protect and preserve three essential properties, namely:

- **Confidentiality**: referring to data protection from unauthorized accesses, disclosures and processes.
- **Integrity**: referring to data accuracy and completeness protection from unauthorized modifications.
- **Availability**: referring to assure data accessibility and usability upon demand by an authorized entity.

Moreover, information security may also involve protecting the authenticity, the authorization and ensuring that entities can be held accountable.

- **Authentication**: referring to ensure that a claimed characteristic of an entity is correct.
- **Authorization**: referring to provide permissions towards information.
- **Accountability**: referring to the entity responsibility for its actions.

In the next subsection, we present the requirements defined by the ISO standard in order to preserve data privacy.

2.4 Privacy legislation

In 1980, the Organization for Economic Co-operation and Development (OECD) issued Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. These guidelines consist of eight principles known as Fair Information Practices (FIP) that enable individuals to express their privacy requirements and place obligations on organizations to follow those requirements.

Besides US privacy legislation, the European Union's application of a comprehensive legislation resulted in the Directive 95/46/EC [5] on the protection of individuals with regard to the processing of personal data and on the free movement of such data. The directive embeds the FIPs.

The ISO standard [14] also defines eleven privacy safeguarding requirements to protect sensitive information, namely: consent and choice; purpose legitimacy and specification; collection limitation; data minimization; use, retention and disclosure limitation; accuracy and quality; openness, transparency and notice; individual participation and access; accountability; information security and privacy compliance. We refer the readers to [14] for more information about these privacy principles.

In the next section, we propose a taxonomy of the existing privacy preserving techniques that are used to satisfy the requirements discussed above.

2.5 Privacy preserving techniques in IoT

Existing privacy preserving techniques are classified into: data perturbation and data restriction.

Data Perturbation Techniques. These techniques are a series of operations that modify or hide some sensitive parts on the original data to preserve privacy [10]. To this end, noise addition and anonymization techniques are adopted.

Noise addition techniques. These techniques transform confidential attributes by adding noise to the original data to prevent the identification of a particular individual [17]. They can be categorized into four groups: (1) data sampling techniques, which aim at releasing a new table that includes only the data of a sample for the whole population, (2) random-noise techniques, which consist of adding or multiplying the value of the sensitive attribute with a randomized number, (3) data swapping techniques, which modify a subset of the data by introducing uncertainty about the true data value [22], and (4) differential privacy techniques, which consist of adding Laplace noise to a database query result [17].

Anonymization protection techniques. These techniques hide a data owner's identity by removing any explicit identifier and makes the data less precise. There are three well-known privacy preserving methods: k-anonymity [23], l-diversity [16] and t-closeness [15]. The k-anonymity is a formal method that is proposed to counter the re-identification problem caused by the quasi identifier attributes. However, k-anonymity can be susceptible to background knowledge attacks. Therefore, researchers designed other versions, such as l-diversity [16], the main idea of which is that there must be at least l distinct values for the sensitive attribute in each quasi identifier group as well as t-closeness method [15], which requires the distribution of a sensitive attribute in any quasi identifier group to be close to the distribution of the attribute in the overall table.

Data Restriction Techniques. These techniques aim at limiting data use by blocking access or encrypting inputs. Data restriction methods include access control and cryptography-based techniques.

Access Control. These techniques are effective for ensuring data sharing [10]. Data owners can express their individual preferences about who can access to what data and how others manipulate their shared data. Control mechanisms include Role Based Access Control (RBAC) and Attribute Based Access Control (ABAC). RBAC assigns access permissions based on the roles whereas ABAC defines permissions based on attributes, such as subject, resource and environment attributes [10].

Cryptographic protection. These techniques are heavily used when preserving privacy. They can be categorized into three major groups: (1) secure multiparty computation, which aggregates inputs of distributed entities to produce outputs, while preserving the privacy of inputs [22], (2) asymmetric/symmetric encryption, which uses keys to protect the data, and (3) public key infrastructure, which delivers the entity a certificate to make sure that the public key belongs to the identified entity.

In recent years, the blockchain technology emerged. In fact, this technology successfully overcomes the problem related to trusting a centralized party. The first system was Bitcoin [18], which allows users to transfer securely the currency (bitcoins) without a centralized regulator. Specific nodes in the network known as miners are responsible for collecting transactions, solving challenging computational puzzles (proof-of-work) in order to reach consensus and adding the transactions in form of blocks to a distributed public ledger known as the blockchain. Since then, other projects demonstrate how these blockchains can serve in other domains, such as the Storj project [3], which is a decentralized peer-to-peer cloud storage network, and the Onename project [1], which is a distributed and secured identity platform. Blockchain technology is also used in order to address the privacy issue in the IoT domain. However, the existing blockchain-based solutions [19] [27] concentrate at addressing the access control issue in the IoT applications. In fact, they adapt the blockchain by eliminating financial bitcoin and introducing new types of transactions in order to limit unauthorized access. Moreover, the examples cited above show that the existing approaches are only concerned with one phase, and generally not address the whole data life cycle.

Based on the above overview of IoT, security and privacy, we work on a systematic literature review detailed in the following section.

3 Systematic literature review

To analyse privacy in the IoT applications, a systematic literature review as defined in [20], has been carried out and reported in this section. Figure 1 depicts the systematic literature review process consists of five steps.

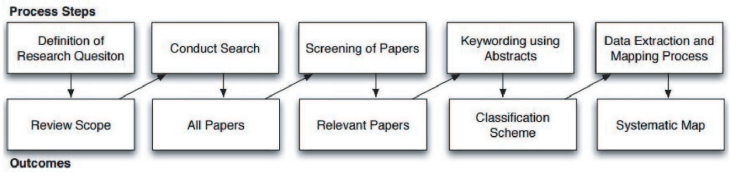

Fig. 1. Systematic literature review process([20])

3.1 Step 1: Definition of research scope

This step consists of defining research questions. The main goal of our study is to (i) categorize the contributions of the research carried out on privacy preserving in the IoT applications from an end to end view, (ii) discover the limitations of the existing works from a privacy principle coverage view, and (iii) verify if using new technology, such as the blockchain can overcome the existing problems of the privacy issue.

Table 1 lists our research questions.

Table 1. Research questions for our systematic literature review

Research Question	Aim
RQ1: How and what are the techniques used by published papers to preserve privacy during data life cycle in the IoT application domains?	This question aims at identifying the used techniques in order to preserve privacy from an end to end view. It will also help to understand in which phase of the data life cycle privacy should be more enhanced.
RQ2: What are the privacy principles that have been supported by the proposed solutions and in which architecture and life cycle phase?	This question will help to understand the actual privacy principle coverage state by the existing solutions. It will also help to identify the least considered principles that should be addressed in the future.
RQ3: What are the privacy preserving techniques that have involved the privacy principle coverage and in which architecture of the literature?	The objective of this question is to know how the chosen technique can involve the privacy principle coverage in the IoT area and its effect on the architecture choice.
RQ4: What are the privacy preserving techniques that have involved the security property respect and the security requirement fulfillment on the IoT applications?	This question will help to know how the chosen technique can involve the respect of security properties in the IoT area and its impact on the fulfillment of security requirements.

3.2 Step 2: Research conducting

This step consists of collecting papers from relevant electronic databases like ACM, IEEE Xplore, Science Direct, and Web of Science. We restrict our search

to papers published between 2010 and 2017. Moreover, a set of keywords is chosen and used to retrieve papers from databases. Thus, we used the following query:

Internet of Things AND privacy AND
(preserving OR principle OR blockchain)

3.3 Step 3: Paper Screening

This step consists of choosing the relevant papers that would help answer the research questions. To do so, a set of inclusion and exclusion criteria are defined. Our study is restricted to papers published in English and addressing privacy issue in the IoT applications. Publications that are table of contents, foreword or summary of conference are deleted. As a result of the filtering process, we exclude 1345 publications. Table 2 summarizes the number of papers included in and excluded from each scientific database. The outcome of this step is 90 papers to include in our study.

Table 2. Number of papers included in and excluded from each database

Database	Amount	Included	Excluded
ACM	113	25	88
IEEE Xplore	271	9	262
Science Direct	945	33	912
Web of Science	106	23	83
Total	1435	**90**	1345

3.4 Step 4: Keywording using abstracts

This step consists of defining a classification scheme composed of facets that group frequent relevant terms that are derived from papers' abstracts. After analyzing the abstracts of papers derived from the previous steps, we consider the frequent relevant terms as dimensions. Then, we cluster the final set of dimensions in order to form the categories (i.e., facets) for our map.

Figure 2 shows our proposed facets and dimensions for our study of the privacy issue in the IoT era.

We define seven facets for our study. These facets cover the eleven privacy principles defined by the ISO standard (see Section 2.4). Each IoT application should follow those principles according to the privacy legislation in order to preserve privacy.

- **Application domains**: such as smart city, smart home, smart grid, health care, location sharing and smart space.
- **Architectures**: such as centralized, decentralized, third party and hybrid.
- **Security Properties**: such as confidentiality, integrity and availability.

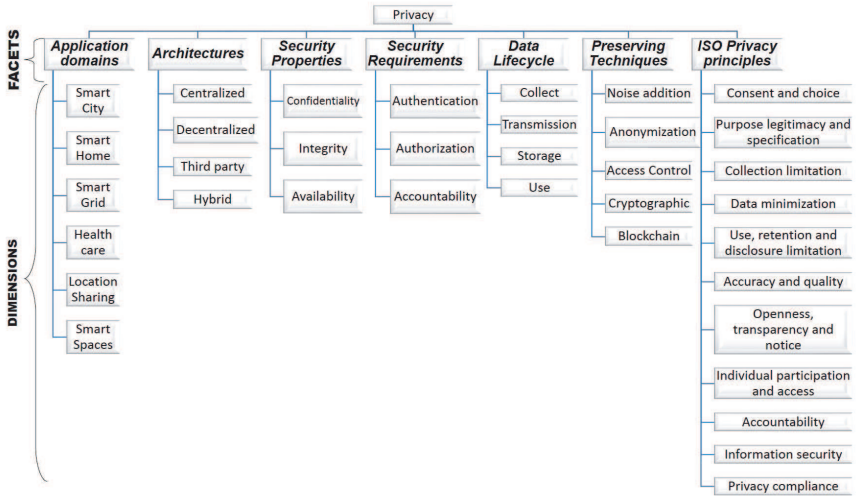

Fig. 2. Classification scheme of facets and dimensions

- **Security Requirements**: such as authentication, authorization and accountability.
- **Data lifecycle**: such as collecting, transmission, storage and use phases.
- **Privacy preserving techniques**: such as noise addition, anonymization, access control, cryptography and blockchain.
- **ISO privacy principles**: such as consent and choice; purpose legitimacy; collection limitation; data minimization; use, retention and disclosure; accuracy and quality; openness, transparency and notice; individual participation and access; accountability; information security and privacy compliance.

3.5 Step 5: Data extraction and mapping process

The facets are combined and the results presented in bubble charts to provide answers to our research questions. It should be noted that for the horizontal axis, the plotted values are related to the total number of publications (i.e. 90 publications) whereas the plotted values on the vertical axis are related to the horizontal axis values.

RQ1: ***How and what are the techniques used by published papers to preserve privacy during data life cycle in the IoT application domains?***
To answer this question, we combine the privacy preserving techniques, the application domains, and the data life cycle facets. Figure 3 shows that cryptography

(67 publications - 74.44%) is the most used privacy preserving technique in the current proposed solutions in each application domain. According to our study, few solutions (2.22%) are found to cover the whole data life cycle. The most addressed data phase by the studied publications is the use phase (47 publications - 52.22%) followed by the transmission phase (38 publications - 42.22%). Both of these phases are based on cryptography technique to provide data protection. The storage phase is the least addressed by publications (9 publications - 10%). It seems that with the emergence of cloud computing, solutions count on cloud data security guarantees and consider it as a trust party. Moreover, anonymization (47 publications - 52.22%) and access control techniques (32 publications - 35.56%) have also emerged in the whole data life cycle. Noise addition is the least used privacy preserving technique in the current proposed solutions for many reasons. First, noise addition leads to a significant utility loss of data [24]. Second, the obfuscated data produced by the classical obfuscation techniques are easy to be detected by an adversary [9]. Finally, noise addition requires smart devices with high storage and computation capabilities to support data storage, aggregation and communication [7].

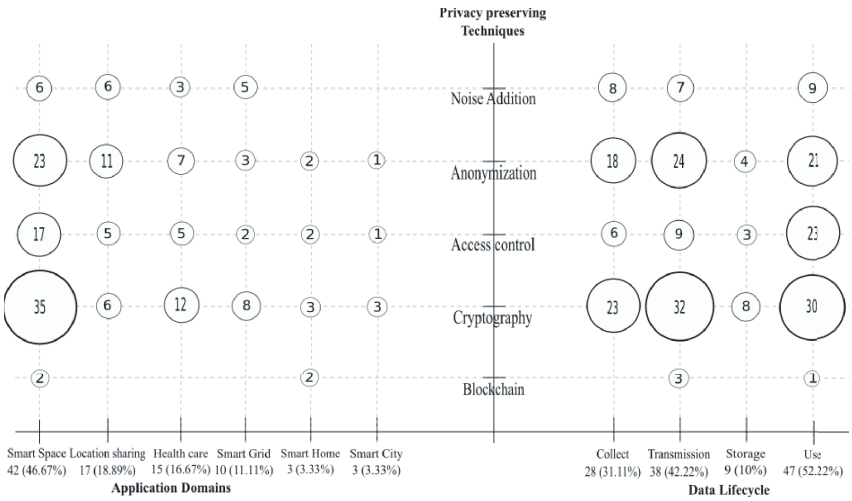

Fig. 3. Privacy preserving techniques, application domains and data lifecycle facets

RQ2: ***What are the privacy principles that have been supported by the proposed solutions and in which architecture and life cycle phase?*** By combining the privacy principles, the architectures and the data life cycle facets (see Figure 4), we can observe the privacy principle coverage by the different publications. In general, the existing solutions cover only six out of eleven principles. The most covered principle is the 'information security' (76 publications - 84.44%) followed by the 'use, retention and disclosure limitation' principle (57 publications - 63.33%). This can be explained by the relationship between

these principles and the most addressed data life cycle phases, such as transmission and use, respectively. However, 'privacy compliance' (2 publications - 2.22%) and 'accountability' (4 publications - 4.44%) are the least considered principles. The rest of privacy principles are covered by less than 50% of the publications. Moreover, both centralized and decentralized solutions involve covering all the privacy principles.

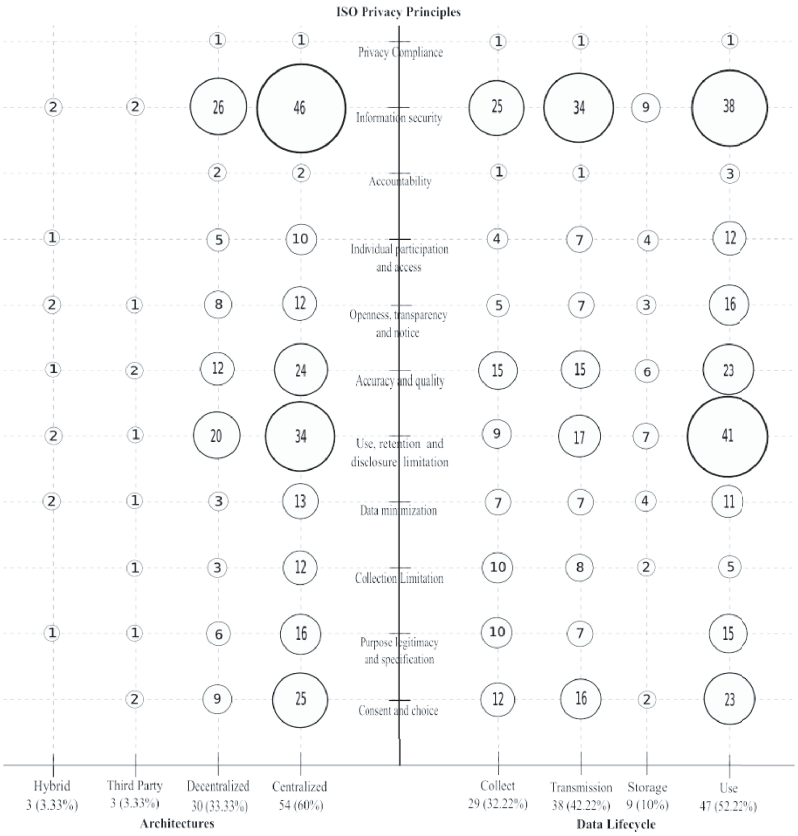

Fig. 4. ISO privacy principles, architectures and data lifecycle facets

RQ3: ***What are the privacy preserving techniques that have involved the privacy principle coverage and in which architecture of the literature?***

The result of combining the privacy preserving techniques, the privacy principles and the architectures facets is shown in Figure 5. Cryptography is the dominant technique in most of the current proposed solutions that helps to cover all privacy principles. This technique is used to cover the 'information security' principle with 70% of the total publications, while the blockchain, as a decentralized

solution, covers only the average of privacy principles (i.e. 6 principles). This can be explained by the life cycle coverage by this technique. In fact, blockchain is used only in transmission and use phases. It can be said that addressing the life cycle coverage by the blockchain use can involve privacy principle coverage. Moreover, centralized is the most used architecture (54 publications - 60%) followed by the decentralized one (30 publications - 33.33%). Independently of the used techniques, these architectures suffer from several technical and legal limits, such as vulnerabilities to attacks, performance and scalability issues as well as need to purpose and data storage duration specifications.

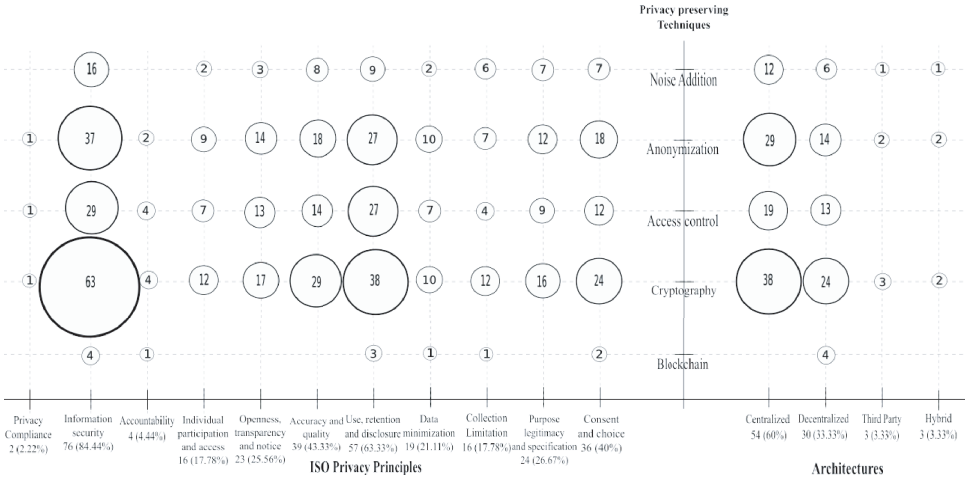

Fig. 5. Privacy preserving techniques, ISO privacy principles and architectures facets

RQ4: ***What are the privacy preserving techniques that have involved the security property respect and the security requirement fulfillment on the IoT applications?***
The result of combining the privacy preserving techniques, the security properties and the security requirements facets is shown in Figure 6. Cryptography (67 publications - 74.44%) is the most used technique in the current proposed solutions that helps to respect all security properties and fulfill all security requirements. Moreover, access control technique is used with 20% of the total publications in order to fulfill both authentication and authorization security requirements. It seems that few of the existing solutions address all the security properties. Especially availability (1 publication - 1.11%) that is the less considered property compared with confidentiality (16 publications - 17.78%) and integrity (17 publications - 18.89%).

It is worth noting that the systematic review results may have been influenced by multiple factors such as researchers' opinions, selection of databases, the used query string for search, and time constraints.

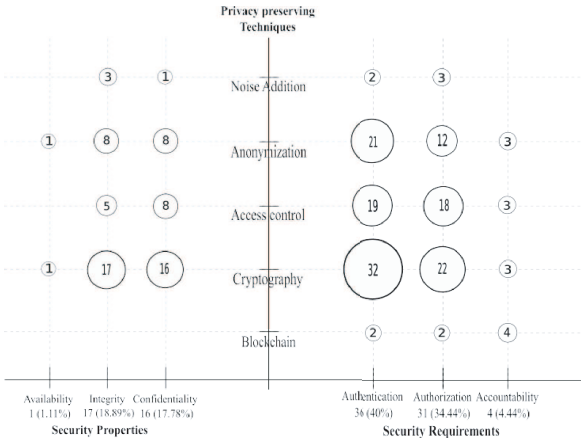

Fig. 6. Privacy preserving techniques, security properties and security requirements facets

4 Privacy and security concerns in the IoT applications

According to our study, many issues in the IoT data protection and privacy preserving area are still to be dealt with. In fact, privacy should be protected in each data phase to preserve sensitive data of the data owner, who can be an individual, a group of individuals or an organization. Considering all the privacy principles defined by the ISO standard [14] is the best way to respect data privacy laws [4][5]. Nevertheless, assuring security is essential to ensure privacy.

For this purpose, we identify and suggest for IoT application consumers and designers some recommendations in order to aware them about key points for protecting data and involving privacy principle coverage from an end to end view. In our work, we distinguish two types of privacy principles: (i) general principles, such as 'accountability', 'information security' and 'privacy compliance', which need to be covered during the whole data life cycle and (ii) specific principles, including the remaining nine principles, which are bound to a particular phase.

According to our study results discussed above, addressing privacy principle coverage can be achieved by programming laws and principles into the blockchain. In fact, law enforcement can be automatically ensured by the use of smart contracts. In practice, Ethereum allows for an easy implementation of such smart contracts [8].

The next subsections identify the privacy principles for each data phase.

4.1 Privacy at collection time

Smart devices collect periodical data from the environment and human bodies. The sensitive information may be leaked out in case of unauthorized manipulation in these devices. For example, an attacker can reprogram a surveillance camera to gather data like the legitimate server. Thus, Privacy by Design as well

as defining an appropriate authentication are important for devices that gather sensitive data in order to prevent illegal device access.

Data perturbation techniques, such as data aggregation, noise mechanism and differential privacy are the used solutions to preserve privacy in this phase.

Regarding the privacy principles, both 'consent and choice' and 'purpose legitimacy and specification' principles should be considered before beginning the collection phase. Each data owner has the right to know the reasons behind collecting each data by his smart device. Thus, the respect of these two principles can help data owner to choose his preferences about the collect frequency that can also influence privacy and data granularity he wants to disclose to third party. For instance, unlike the traditional electricity architecture in which metering data are read monthly, in the smart grid, more detailed energy data are collected. Thus, these data can expose a great amount of valuable and intimate information about the customers. Besides, specifying the reasons of collecting particular data can lead to consider the privacy principle of the 'data collection limitation'.

Preserving privacy in the collection phase is essential and can affect the whole data life cycle. Thus, privacy should be preserved before the transmission phase instead of trying to preserve it when the data are already stored in the utility data center. Thus, privacy should be preserved at the smart devices. However, smart devices generally do not support the privacy preserving techniques. For that, the use of an access point between smart devices with low memory and storage capabilities and the main data system can enable to locally store the data for pre-processing before the transmission phase. Such access point should provide a portal to the data owner to manage his smart devices and choose his preferences about how others access and manipulate his data. Thus, data owners in the IoT applications will able to keep control on their shared data and preserve their privacy. Besides, this access point should have the capability to interact with a blockchain.

4.2 Privacy at transmission time

After being temporally stored in the smart devices or in the broker, the collected data will be sent to external servers. Many techniques are used in order to protect the data from attacks during the transmission phase.

We notice that the most used common technique is cryptography. In our study, we distinguish three transmission types: data that are periodically transmitted, data that are transmitted as a replay to a request and the third type is the use of Publish/Subscribe system. In order to ensure data confidentiality during transmission, data encryption is absolutely required. Digital signature and certificate are also needful in order to ensure integrity and prove the identity.

In this phase, the privacy principle that needs to be covered is 'accuracy and quality'. Data should not be modified during the transmission phase. Moreover, detecting malicious entities that try to inject data in order to congest the network or influenced the analysis results is still an issue to be solved. Note that it is difficult to separate privacy and security preserving solutions in this phase. Therefore, assuring security is essential in order to ensure privacy.

The blockchain can involve privacy preserving in the IoT applications and ensure the 'accuracy and quality' principle coverage by enhancing collaboration between all entities in the network in order to verify data accuracy, integrity and reject unauthorized data access. Moreover, the blockchain offers non-repudiation principle compliance, which consists of preventing an entity from denying actions that are performed by itself since blockchain can ensure auditing functions. The main purpose of using the blockchain technology is to prevent any privacy violation attempts in the IoT applications. In fact, thanks to the immutable characteristic of the blockchain, the malicious nodes cannot modify the blockchain.

4.3 Privacy at storage time

After being periodically collected by devices and temporally sent through the network, data should be stored to be available for analyzing. A high storage capability is required to support the huge amount of data generated by IoT devices.

To conceal the real identity linked to the stored data, cryptography and anonymization techniques are used. Secure multiparty computation and asymmetric/symmetric are the solutions the most used by the existing approaches to preserve privacy during the storage phase. We distinguish two solutions types. The first one consists of storing encrypted data and the second solution type aims at decrypting data before the data storage. Each solution had its advantages and disadvantages. Although storing encrypted data can overcome the trust issue, data querying process will be more complicated. Contrary to the first solution, storing unencrypted data simplifies data querying by end-users (i.e., data consumers), but it necessitates giving trust to cloud computing that will not disclosure sensitive data.

Regarding the privacy principles, French and European data privacy laws state that personal data collected and stored within a European Union country territory should be stored for a reasonable time duration [4][5]. Thus, the 'use, retention and disclosure limitation' principle is to be considered in this phase. This principle aims at limiting the retention of personal information to fulfill the specified purpose as long as necessary and thereafter securely destroying data.

IoT generates a large amount of data that cannot be supported by traditional data storage solutions. For this purpose, the cloud computing seems to be the best scalable solution for storing data for many reasons. First, cloud computing offers a reduced cost. Second, it allows benefiting from scalability and high performance computing. Finally, it guarantees data security and recovery. Moreover, data cannot be altered. In fact, when storing the data hash in the blockchain, the data owner can detect any change in his stored data by comparing the hash of his data in the data center with the stored hash in his gateway. Thus, data owners can store their data without relying on a trusted Third Party Authority.

4.4 Privacy at processing time

Data processing is an important phase in analyzing and using the collected and stored data by end-users. The generated data in the IoT applications can be

shared between multiple parties for two purposes. The first purpose is when the data owner should be known, such as billing purpose or patient's treatment. The second purpose is when the data are used for governmental programs and research. In this case, data must be anonymized.
In order to preserve privacy in query output, secure multiparty computation, anonymization, and differential privacy are used in the use phase.

According to European data privacy laws [4], data should have multiple levels of disclosure (i.e., no data sharing, restricted access, open access). Thus, without explicit acceptance of the data owner, personal data should not be disclosed to third parties. To this end, the privacy principle 'use, retention and disclosure limitation' should be considered in this phase. Furthermore, 'data minimization' principle, which aims at minimizing the processing of personal information to just fulfill the specified purpose should be more studied. Moreover, it is time that data owners be aware of their rights to have clear, complete and accessible information to correct inaccuracies. These rights are presented by 'openness, transparency and notice' and 'individual participation and access' privacy principles.

Access control techniques are necessary in order to fulfill authorization security requirement and help limit unauthorized access. However, the traditional access control models, such as RBAC and ABAC cannot support the data distribution in the IoT applications because of the lack of flexibility, scalability and usability [10]. Therefore, inspiring from the blockchain technology and proposing a permission token rather than a bitcoin can ensure a new access control solution with auditing functions and non-repudiation compliance. Thus, the blockchain-based access control can help to determine who accesses to what according to the data sensibility level (i.e., production data are less sensitive than decision support data) and under what circumstances in CIS and consequently enable data owners to own and control their shared data.

5 Related Work

Compared to security, privacy in the IoT applications has only received more attention since last year.

Fernández-Alemán et al. [11] conducted a systematic literature review concerning the security and privacy of electronic health record (EHR) systems. The authors defined and analyzed their selected articles based on several security areas. The study conclusion shows that most of the solutions defined EHR system security controls, but these are not fully deployed in actual tools.

A review of privacy threats related to the IoT applications was conducted by Ziegeldorf et al. [26]. The authors classified the evolving technologies used in the IoT applications and highlighted the most important features considered in the context of privacy. Afterwards, the authors studied and analyzed seven threat categories. Their study identified privacy-preserving approaches from related work to determine whether they could mitigate in an IoT context.

To sum up, it can be said that the existing systematic review works concerning privacy preserving issue in IoT focus on analyzing the challenges and threats

of IoT in the context of entities and information flows. Our work extends the existing works by examining the IoT-specific solutions considering the security property and requirement fulfillment and studying privacy principle coverage.

6 Conclusion

Actually, IoT is considered as a promising technology that may improve collaborative working at anytime, anyplace, with anything and anyone. However, IoT opens the collaborators up to a possible loss of privacy. For this reason, both privacy and security should be carefully considered in this technology. The paper's aim is to present a detailed study about the privacy preserving solutions in the IoT applications. To achieve that, we have conducted a systematic literature review. Our analysis of the existing works helped us to identify recommendations in order to involve privacy principle coverage and security requirement fulfillment in the IoT context. We expect that our elaborated study will be an interesting contribution. In fact, researchers who want to target privacy in the IoT field can be based on our exhaustive analysis for proposing future scientific contributions that overcome existing solution limits.

In our ongoing work, we intend to propose a blockchain-based solution that takes into account our recommendations in order to preserve privacy in the IoT applications. In fact, the promising technology blockchain that successfully overcomes the problem related to trusting a centralized party in several domains, can be adapted by the IoT application designers in order to improve collaborative working in CIS and overcome privacy issue in the IoT applications.

References

1. Onename. https://www.onename.com/
2. Sap. https://www.sap.com/product/analytics/smart-meter-analytics.html
3. Storj. www.storj.io
4. Loi 78-17 du 6 janvier 1978 modifiee. https://www.cnil.fr/fr/loi-78-17-du-6-janvier-1978-modifiee (1978)
5. Regulation (eu) 2016/679 of the european parliament and of the council. http://eur-lex.europa.eu/eli/reg/2016/679/oj (2016)
6. Bertino, E.: Data security and privacy: Concepts, approaches, and research directions. In: Computer Software and Applications Conference (COMPSAC), 2016 IEEE 40th Annual. vol. 1, pp. 400–407. IEEE (2016)
7. Birman, K., Jelasity, M., Kleinberg, R., Tremel, E.: Building a secure and privacy-preserving smart grid. ACM SIGOPS Operating Systems Review 49(1), 131–136 (2015)
8. Delmolino, K., Arnett, M., Kosba, A., Miller, A., Shi, E.: Step by step towards creating a safe smart contract: Lessons and insights from a cryptocurrency lab. In: International Conference on Financial Cryptography and Data Security. pp. 79–94. Springer (2016)
9. Elkhodr, M., Shahrestani, S., Cheung, H.: A semantic obfuscation technique for the internet of things. In: 2014 IEEE International Conference on Communications Workshops (ICC). pp. 448–453. IEEE (2014)

10. Fang, W., Wen, X.Z., Zheng, Y., Zhou, M.: A survey of big data security and privacy preserving. IETE Technical Review pp. 1–17 (2016)
11. Fernández-Alemán, J.L., Señor, I.C., Lozoya, P.Á.O., Toval, A.: Security and privacy in electronic health records: A systematic literature review. Journal of biomedical informatics 46(3), 541–562 (2013)
12. Gartner: Internet of things. http://www.gartner.com/it-glossary/internet-of-things (2017)
13. Guillemin, P., Friess, P., et al.: Internet of things strategic research roadmap. The Cluster of European Research Projects, Tech. Rep (2009)
14. International Organization for Standardization: Information technology security techniques privacy framework, ISO/IEC 29100 (2011)
15. Li, N., Li, T., Venkatasubramanian, S.: t-closeness: Privacy beyond k-anonymity and l-diversity. In: 2007 IEEE 23rd International Conference on Data Engineering. pp. 106–115. IEEE (2007)
16. Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M.: l-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data (TKDD) 1(1), 3 (2007)
17. Mivule, K.: Utilizing noise addition for data privacy, an overview. arXiv preprint arXiv:1309.3958 (2013)
18. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (2008)
19. Ouaddah, A., Abou Elkalam, A., Ait Ouahman, A.: Fairaccess: a new blockchain-based access control framework for the internet of things. Security and Communication Networks 9(18), 5943–5964 (2016)
20. Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M.: Systematic mapping studies in software engineering. In: EASE. vol. 8, pp. 68–77 (2008)
21. Roman, R., Zhou, J., Lopez, J.: On the features and challenges of security and privacy in distributed internet of things. Computer Networks 57(10), 2266 – 2279 (2013), towards a Science of Cyber SecuritySecurity and Identity Architecture for the Future Internet
22. Sharma, M., Chaudhary, A., Mathuria, M., Chaudhary, S.: A review study on the privacy preserving data mining techniques and approaches. International Journal of Computer Science and Telecommunications 4(9), 42–46 (2013)
23. Sweeney, L.: k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10(05), 557–570 (2002)
24. Ukil, A., Bandyopadhyay, S., Pal, A.: Privacy for iot: Involuntary privacy enablement for smart energy systems. In: 2015 IEEE International Conference on Communications (ICC). pp. 536–541. IEEE (2015)
25. Westin, A.F.: Privacy and freedom. Washington and Lee Law Review 25(1), 166 (1968)
26. Ziegeldorf, J.H., Morchon, O.G., Wehrle, K.: Privacy in the internet of things: threats and challenges. Security and Communication Networks 7(12), 2728–2742 (2014)
27. Zyskind, G., Nathan, O., et al.: Decentralizing privacy: Using blockchain to protect personal data. In: Security and Privacy Workshops (SPW), 2015 IEEE. pp. 180–184. IEEE (2015)