J Supercomput (2017) 73:2763–2800
DOI 10.1007/s11227-016-1953-y

Privacy in cloud computing environments: a survey and research challenges

Amal Ghorbel[1] · **Mahmoud Ghorbel**[1] · **Mohamed Jmaiel**[2]

Published online: 23 January 2017

Abstract Definitely, cloud computing represents a real evolution in the IT world that provides many advantages for both providers and users. This new paradigm includes several services that allow data storage and processing. However, outsourcing data to the cloud raises many issues related to privacy concerns. In fact, for some organizations and individuals, data privacy present a crucial aspect of their business. Indeed, their sensitive data (health, finance, personal information, etc.) have a very important value, and any infringement of privacy can cause great loss in terms of money and reputation. Therefore, without considering privacy issues, the adoption of cloud computing can be discarded by large spectra of users. In this paper, we provide a survey on privacy risks and challenges for public cloud computing. We present and evaluate the main existing solutions that have made great progress in this area. To better address privacy concerns, we point out considerations and guidelines while giving the remained open issues that require additional investigation efforts to fulfill preserving and enhancing privacy in public cloud.

Keywords Survey · Privacy · Cloud computing · Privacy issues · Data confidentiality

✉ Amal Ghorbel
amal.ghorbel@redcad.org; ghorbelamal1@gmail.com

Mahmoud Ghorbel
mahmoud.ghorbel@redcad.org

Mohamed Jmaiel
mohamed.jmaiel@redcad.org

[1] ReDCAD Laboratory, ENIS, University of Sfax, B.P. 1173, 3038 Sfax, Tunisia

[2] Digital Research Center of Sfax, B.P. 275, Sakiet Ezzit, 3021 Sfax, Tunisia

Abbreviations

Acc	Accountability
CSA	Cloud Security Alliance
CSB	Cloud Service Broker
DLP	Data Leakage Prevention
FIP	Fair Information Practices
IaaS	Infrastructure as a Service
IDD	Illegitimate Data Dissemination
IDH	Illegitimate Data Handling
PaaS	Platform as a Service
PC	Privacy Compliance
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PII	Personally Identifiable Information
Re	Retention
SaaS	Software as a Service
SCI	System Call Interception
TCG	Trusted Computing Group
TPM	Trusted Platform Module
TTPM	Trusted Third Party Mediator
USU	Unauthorized Secondary Usage
VM	Virtual Machine
XACML	eXtensible Access Control Markup Language

1 Introduction

cloud computing is defined by the United States National Institute of Standards and Technologies (NIST) as
"cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [1].

The same NIST document describes briefly the five main cloud characteristics (on-demand self-service, broad network access, multi-tenancy, rapid elasticity and measured service), its three service model (SaaS: Software as a Service, PaaS: Platform as a Service, IaaS: Infrastructure as a Service) and some cloud deployment models (public, private, hybrid and community).

From the perspective of a common reader of this cloud computing definition, someone can extract several interesting keywords that highlight the great advantages of cloud computing like "convenient on-demand access," "configurable computing resources," "minimal management effort" and "elasticity." However, from the perspective of privacy, there is only one simple keyword in this definition that reveals one of the biggest concerns within the cloud computing concept, which is "shared pool." In fact, the cloud computing paradigm offers several services for processing user data

on machines that he does not control and, even more, does not operate. So when we confront our definition of the privacy: "the right of a person to not have his private data brought to the public knowledge" with the aforementioned cloud definition, a big challenge is raised: how can we be sure that private data put in cloud shared pool will not be brought to public knowledge?

This challenge comes out in several levels and aspects of the cloud. For example, one of the most famous cloud services is outsourcing the storage and the management of data. This service is so used in such a way that people use it unconsciously (through emails, remote storage services, social networks, backup services, etc.). People have no idea where (geographically) their data are, who can or cannot access to it, who keeps it, what really happens when they ask for data deletion and so on.

Another aspect embodying privacy risks resides in the multi-tenancy feature of the cloud. It means that multiple customers can be served from the same instance of software, by "securely" separating the resources on logical level [2]. This feature can achieve higher profit margin by maximizing resources usage and reducing cost through economies of scale. However, this gain cannot crop up without risking the leakage of confidential or sensitive data. There is no mean to verify whether a tenant data were accessed, copied or logged by another tenant [3].

Regarding cloud deployment models, of course, the privacy concerns does not affect all models with the same degree. The private cloud model is affected by this challenge less than the public or hybrid cloud models. In fact, the latter are available for open (more or less) use by the general public (particular, organization, etc.), whereas in private cloud, resources are customized solely for a single organization (e.g., an enterprise). The management and the control of these resources fall on the responsibility of this organization. Private cloud is none other than a (proprietary) information system designed respecting the cloud concept including its features, characteristics and service layers. Community clouds are supplied for several organizations with shared concerns, so the tolerance for the privacy issues is more critical (than a private cloud).

In general, privacy issues are not recent matters. Around the world, this has driven to the release of a large amount of laws and legislation to ensure the protection of individual's data. Examples include the Fair Information Practices [4], the European Directive 95/46/EC [5], the USA Health Insurance Portability and Accountability Act (HIPAA) [6], the USA Gramm–Leach–Bliley Act [7], etc. Privacy issues become more and more hazardous in cloud computing environment. So, these laws may not be applicable in such dynamic and public environment and need to be customized to cover all privacy problems. On the other hand, there is no guarantee for the enforcement of these laws in the cloud.

To overcome privacy issues in cloud computing environments, several research actors invested efforts taking into account privacy laws and user's preferences for data protection. The invested works engendered various techniques and approaches. Moreover, some researchers have conducted surveys to understand and target privacy issues in the cloud. However, in general, these surveys lack some privacy issues and current solutions since most of them are not exclusively dedicated to privacy in the cloud. Besides, they lack a clear classification either for privacy issues or for current data protection trends. Here, we provide a complete survey dedicated to the privacy

in the cloud to offer a better understanding of the privacy challenges and to identify problems that are unsolved till now. Thus, we exhibit the privacy issues according to four aspects. We classify the existing solutions into techniques and approaches, and we classify the approaches into data-centric, user-centric, CSP-centric and hybrid approaches. We assess the presented studies and tabulate advantages and disadvantages. Afterward, we discuss open research issues and give a guideline for privacy preservation to identify the most relevant criteria that must be further concerned in the future research directions. The main goal of this study is to offer a better understanding of the privacy challenges in the cloud environment and to focus on current imperfections to fulfill the privacy issues.

The remainder of the paper is organized as follows: Sect. 2 presents some related works. Section 3 gives some basic backgrounds on privacy and enumerates some privacy legislation. Section 4 highlights several risks and challenges of privacy for cloud computing. Section 5 provides an overview of current data protection trends (techniques and approaches) in the cloud environment. In Sect. 6, we describe a guideline for ensuring privacy protection in the cloud. Next, in Sect. 7, we discuss general privacy open issues that still exist in the cloud. The paper ends with a conclusion.

2 Related works

Several researchers have conducted researches and surveys to understand and target privacy issues in the cloud. In this section, we will position our contribution among the surveys already made.

Siani [8] gives a clear definition of privacy and enumerates types of data to be protected. She then presents an overview of challenges, issues and risks to privacy in the cloud. She also gives a set of privacy protection keys based on the FIP and a guideline for designing cloud services. This guideline is intended for software engineers to take account of privacy when designing cloud services in contrast to our guideline which provides practical recommendations that implicate most of the involved actors to achieve privacy protection when using cloud services. This may be more relevant since most current cloud services are not based on the privacy by design concept. Probably, Siani's guideline can be useful for designing the future intended cloud services. Besides, in our survey, we cover most of the proposed techniques used for privacy preservation and we present current approaches that use these techniques which is not the case of this survey. We also present some open issues to highlight weaknesses of the current solutions which require greater research efforts.

Miranda et al. [9] give an overview of privacy and cloud computing and highlight privacy issues when moving personal information in the cloud. The authors have also presented a set of promising techniques and methods that may be used to address these issues. They give an evaluation in which they describe drawbacks of these solutions and discuss what more is needed to be more efficient. Finally, they present some related issues for key management, design for privacy and accountability. This work presents a relevant survey that covers most of the privacy aspects in the cloud. Nevertheless, the exhibited solutions look like a strategic consulting or advice. It lacks a study and a classification of technical approaches that treated the privacy issues in the cloud. Con-

trary to this survey, in our survey, we distinguish between a technique and an approach to data protection. Firstly, we enumerated most of the technical research solutions and then we present some proposed approaches which use these techniques. We classify these approaches into four main categories depending on actors involvement in the privacy protection process: (i) data-centric approaches, (ii) user-centric approaches, (iii) CSP-centric approaches and (iv) hybrid approaches. Afterward, we assess the presented studies and present its advantages and disadvantages.

In other surveys presented in [10,11], the authors highlight some security and privacy issues in the cloud and enumerate the existing solutions to deal with these problems. The solutions are presented and briefly discussed in order to highlight the advantage and inconvenient of each one. They are also classified chronologically from the oldest one to the newest one. Nevertheless, there is a commingling between security and privacy in the presentation of the issues; no clear separation between the two aspects has been done. In our study, we have explained the difference between privacy and security and the relation between them. Besides, some techniques or approaches are not covered by the survey like sticky policy or TPM-based solutions.

A survey provided in [12] studies the Data Leakage Prevention (DLP) solutions dedicated to detect or prevent the leakage confidential data when it is in use, in transit, and at rest. In this survey, the authors discuss the DLP systems paradigm and describe the challenges facing DLP. They also categorize the current DLP methods and discuss the advantages and disadvantages of each method. The survey focuses only on data leakage issue. Instead, in our study we consider several privacy issues as well as data dissemination, lack of user control, data retention, etc.

The surveys presented in [13–18] enumerate some security and privacy issues in the cloud environment. The vast majority of these surveys was focused on security. The privacy issues presented covered only two main issues: legal aspect and data location aspect. In contrast to our study in which we present privacy issues according to four aspects: (i) issues related to the lack of user control such as lack of transparency, data loss and leakage, etc., (ii) issues related to the dynamic nature of cloud such as transborder data flow issues, retention and replication, (iii) compliance with laws and user's preferences and (iv) accountability. We also enumerate the main cloud actors that are involved in the private data life cycle in order to better understand the responsibility of each one.

In general, the studied surveys lack some privacy issues and current solutions. Besides, they do not provide a clear classification either for privacy issues or for current privacy preservation approaches. In this survey, we present privacy issues in the cloud computing and enumerate current research solutions done to confront them. We exhibit the privacy issues according to four aspects (lack of user control, dynamic nature of the cloud, compliance with laws and user's preferences, and accountability) and enumerate the main cloud actors that are involved in the private data life cycle. We classify the current solutions into techniques and approaches, and we classify the approaches into four main categories: data-centric approaches, (ii) user-centric approaches, (iii) CSP-centric approaches (CSP: Cloud Service Provider) and (iv) hybrid approaches. We assess the presented solutions according to many relevant criteria. Afterward, we discuss open research issues and give a guideline for privacy preservation.

The next section is dedicated to the definition of the privacy in general and the enumeration of different types of data and legal policy concerned by this aspect.

3 Privacy

Privacy is a fundamental human right that was included in the United Nations Universal Declaration of Human Rights since 1948. Privacy is a complex notion for which there is no standard accepted definition. Thus, privacy has various definitions as: "the right to be let alone" [19], "The state of being free from public attention or unsanctioned intrusion" [20], "the right to be free from secret surveillance and to determine whether, when, how, and to whom, one's personal or organizational information is to be revealed" [21].

By combining the meaning of the listed definition, we can define privacy in a much more simple way as "the right of a person to not have his private data brought to the public knowledge." We mean by private data any personal information, habits, lifestyle, health data, etc. In the computing context, privacy is defined as "the rights and obligations of individuals and organizations with respect to the collection, use, retention, disclosure, and disposal of personal information" [22].

Security and privacy are two related but different terms. In fact, security represents a set of practices that are designed to ensure the confidentiality, availability and integrity of data, whereas privacy is defined as the appropriate use of personal data under certain circumstances [23] (e.g., customer and/or law policies). Security techniques must be employed to ensure the appropriate use of data according to the required circumstances. Therefore, security is not privacy but it is a basic foundation, among others, that helps to preserve privacy.

3.1 Types of private data

The privacy concerns cover several types of private information as depicted in Fig. 1. We detail each of these types as following.

3.1.1 Personally identifiable information (PII)

Personally Identifiable Information (PII) can be defined as information that can identify an individual with certainty. PII can be divided into two subgroups:

- *Key attributes* each of these attributes can uniquely identify an individual such as name, phone number, social security or national identity number, email address and passwords. These attributes must be removed when adopting anonymization techniques [24].
- *Quasi-identifier* a set of these attributes can uniquely identify an individual such as {ZIP code, date of birth, address}. These attributes can be used for linking anonymized dataset with other datasets and then identifying individuals [24].

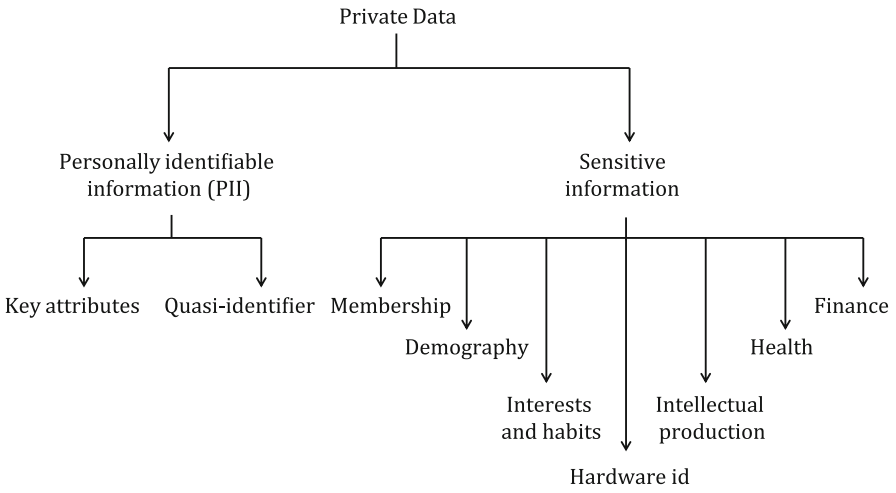

Fig. 1 Types of private data

3.1.2 Sensitive information

Sensitive private data can be categorized into different types as follows:

- *Membership* represents data regarding the subject affiliation in groups such as political, religious, community, etc.
- *Demography* represents the demographic characteristics of a data subject such as nationality, gender, education level, job position, criminal record, etc.
- *Interests and habits* represents the data subject's activities and preferences such as traceability, history of data use, web browsing activity, shopping activity, etc.
- *Finance* represents the data subject's financial information such as credit card number, account balance, financial transaction traces, etc.
- *Health* represents health information of the data subject such as medical record, diseases, diagnostics, medical images, prescriptions, etc.
- *Hardware id* represents the data subject's hardware identifiers such as computer IP address, radio frequency identity (RFID) tags, MAC address, hostname, etc.
- *Intellectual production* represents information related to the data subject's ideas, inventions before publication or validation.

In general, private data represents every information that is considered as personal. At the professional scale, private data concerns confidential enterprise information, enterprise employee information and enterprise customer information.

3.2 Privacy legislation

Privacy is extremely important to every business and individual concerned about protecting confidential and personal information. Around the world, this has driven to laying down of a large amount of laws and legislations. Indeed, most of the national governments have imposed their local privacy legislation.

The Fair Information Practices (FIP) was developed in the USA in the 1970s [4]. It represented the basis for most data protection and privacy laws around the world. These principles can be broadly described as follows:

Data collection limitation data collection should be performed with the consent of the data subject or owner and should be limited to the requested data.
Purpose specification the purpose of the collection should be stated at the time of data collection.
Purpose use limitation personal data should not be used for other purposes unless with the consent of the individual.
Individual participation an individual should be able to control his information and to obtain details about them. He should also be given the choice of whether he want this information to be collected, disclosed or not.
Visibility and transparency an individual should be able to find out how and by whom his data are handled.
Security personal data should be protected by a reasonable degree of security.
Compliance the data collector must be compliant with the specified circumstances.
Accountability the data collector should be accountable for complying (or not) with the specified circumstances.

The FIP has been proposed and emerged in the USA. Despite this, USA does not have a comprehensive regime of privacy. Indeed, data protection are disseminated among different sector-specific laws such as the Health Insurance Portability and Accountability Act (HIPAA) [6] which targeted health industry and the Gramm-Leach-Bliley Act [7] which is specifically designed for the financial services and applied for financial institutions. Another act which is relevant in this context is the USA-PATRIOT Act [25]. This act was emerged in 2001 to intercept and obstruct terrorism [26]. However, USA-PATRIOT Act does not conform to any of the FIP practices and presents a limitation for the data privacy [27]. Particularly, in Sects. 215 and 505, this act allows data collection and disclosure without the consent of the owner. Further, the purpose of the data collection lacks of clarity and accountability related to the data collector and processor is not considered [28]. In contrast to the USA approach for data protection, the European Union fixed privacy regulations through the Data Protection Directive 95/46/EC which was implemented in October 1995 [5]. The main purpose of the directive was to consolidate the privacy laws that existed in the states member of the European Union and to provide a basic standard on the privacy safeguard [29]. The European Data Protection Directive implements the USA Fair Information Principles (FIP), along with some additional preferences including transborder data flow restrictions. The European Data Protection Directive has been revised and reformed as a "General Data Protection Regulation" on January 25, 2012 [30]. This new version concerns data processing for the purposes of prevention, investigation, detection or prosecution of criminal offenses, and the execution of criminal penalties.

To regulate industry and organizations privacy, many legal arguments and standards were developed. The Safe Harbor agreement [31] was introduced in 2000 to regulate data transfers between the USA and the European Union. This argument states that a US-based organization which has business operations within the European Union has to rely on the Safe Harbor agreement to adhere to the Transborder Transfer principle

of the Directive 95/46/EC. However, because of the important number of unauthorized disclosures made in 2013 by the U.S. National Security Agency (NSA) surveillance programs and other U.S. intelligence collection operations in Europe, the Court of Justice of the European Union (CJEU) invalidated the Safe Harbor Agreement on October 6, 2015. Thus, the USA–EU revised the invalid agreement and provided the EU-US Privacy Shield as a replacement of the Safe Harbor agreement on February 29, 2016 [32]. Its main principles entail notice, choice, accountability, security, data integrity, purpose limitation, access, and recourse, enforcement, and liability. ISO/IEC 27018 is the first international standard intended to cover privacy aspects for the cloud industry [33]. It was emerged in 2014. It provides guidelines to protect Personally Identifiable Information (PII) and to implement the necessary controls to privacy risks. Its key principles include consent, transparency, communication, portability/data retention, compliance and confidentiality. Currently, some of CSPs such as Microsoft and Amazon announce their compliance with the ISO/IEC 27018 standard.

In general, private data have a very important value either for individuals or for professionals. It must acquire much more attention because of the misuse of such sensitive information can lead to many corruptions. Nevertheless, recent reports [34] indicate an increase of the data leakage in the business sector (about 50% of recorded data are leaked), in the government sector (about 30% recorded data are leaked) and in the health and education sectors (about 20% recorded data are leaked). The adoption of cloud computing technology makes privacy even worse. Indeed, moving private data in such dynamic environment will raise many problems and concerns. In the next section, we will highlight several issues related to the privacy in cloud environments.

4 Privacy issues in the cloud

To benefit from advantages offered by cloud computing, the user is led to move his private data to the Cloud Service Provider (CSP). Outsourcing this data to an external party provokes several problems and concerns such as: Who has access to the data? Where is the data stored? How many copies of data exist in the cloud? How to be sure that all copies are deleted when requested? Are the data actors compliant with laws and user privacy policies in the cloud? All these issues are mainly raised because of (i) the lack of user control, (ii) the dynamic nature of the cloud, (iii) the lack of technologies to ensure the compliance with laws and user's preferences and (iv) the difficulty to achieve accountability in the cloud environments. In order to better understand these privacy issues, let us have a look at the involved actors; then, we will detail each of the aforementioned issues.

4.1 Involved actors

The usage of the cloud covers several domains and levels; thus, several kinds of actors could have a relation with the life cycle of a personal data in the cloud. Figure 2 presents some possible actors that are involved in the data usage process in the cloud.

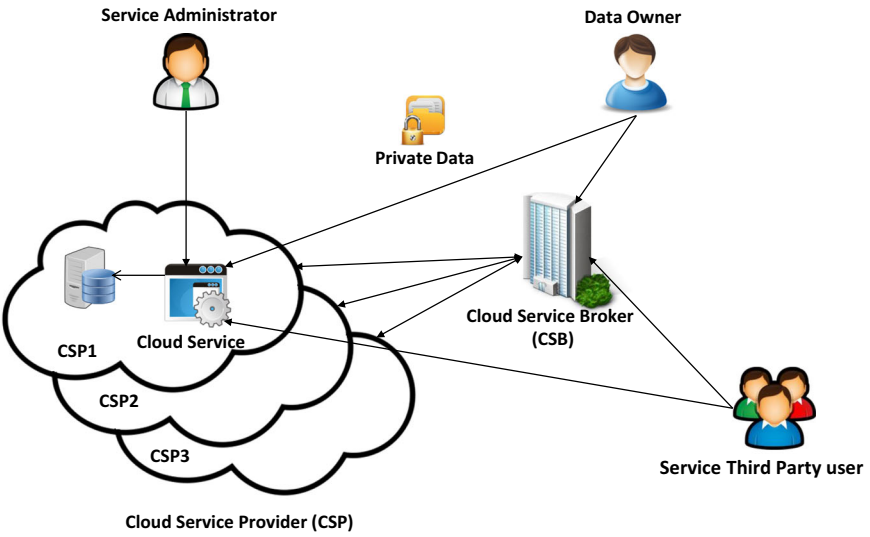

Fig. 2 cloud system model and involved actors

4.1.1 Data owner

This is the main actor. He is the owner of the private data and has chosen or not to host his data in the cloud or to use a service that is hosted in the cloud. It should be noted that in some cases, the data owners are several individuals, especially when the data are co-produced by several individuals or when it's about data collection regarding several individuals (like usage databases, statistical data, etc.).

4.1.2 Cloud service provider CSP

The CSP is a key actor since it is the data hosting entity. It could be embodied by both individuals (internal employees) or/and software entities (web services interfaces, distribution, backup and load-balancing algorithms, etc.). The CSP can perform cloud-specific processing on private data such as duplication, transfer, storage, research, indexation, statistic, data mining, segmentation, etc. CSP can also perform undeclared actions that can threaten the privacy of data. For example, it can disclose private data by reading them or peek in the customer's VM when running and make copies. In fact, the CSP could seek to gain information about its customers and their behaviors because that information can be a source of a great financial profit (it can sell information for example to the advertisement). From another side, a CSP can use information to make statistics, to improve his services and to do many other businesses.

4.1.3 Cloud service broker CSB

This actor is a company that adds values to existing cloud services (especially in public cloud) on behalf of customers of these services. The CSB uses three main roles

including aggregation, integration and customization brokerage. In some limited cases, the role of the CSB is limited to a consulting role. In this case, it is not considered for us as a real actor of the private data since it is only responsible for making the first linking between CSP and customers. Otherwise, it is a sustainable intermediate between them.

4.1.4 Cloud-based service

This actor represents applications and programs deployed in the cloud and which are used to accomplish a service such as CRM (customer relationship management), document management, Cloud storage service, accounting, medical applications, collaboration tools, etc. In some cases, these applications need to access to private data to perform some treatments. However, how to ensure that the service is not malicious and do not menace privacy of data (make copies of data, modify data without authorization, send data to a third party, etc.).

4.1.5 Cloud-based service administrator

It is the owner of the service. It can be the CSP or another party. Its role is to control and to improve the service. The administrator of the service can have access to the service database and then can disclose private customer data. If the administrator is different from the CSP, the administrator of the service has the same motivations of the CSP to disclose user's information.

4.1.6 Cloud-based service third party

They are actors implicated in the usage of the service deployed in the cloud, for example, an employee of an enterprise that uses collaboration tool, a doctor and a clinic employee that use a medical application to manage patient medical information. Generally, these actors are trustworthy, but it can happen in worse cases that they can be unreliable and can threaten the privacy of the data (for example a clinic employee that disclose the medical information of patients to third parties, a malicious employee who sell confidential information of the enterprise to concurrent).

4.2 Lack of user control

In the public cloud, the private data are stored in remote machines controlled by the CSP. There is a lack of transparency of data processing, the storage location, the number of copies, etc. It is even difficult, and sometimes impossible, to know whether there were violations of privacy and who is the responsible. Let us consider the example of a user who enjoys using cloud storage services like Dropbox, Mega [35] or Tresorit [36]. The user does not know that some of these service providers can store his private data in the USA whose PATRIOT Act presents a limitation of privacy. Moreover, the uncontrolled private data can be susceptible to unauthorized usages. Illegitimate Data Handling (IDH) issue is when authorized actors perform

unauthorized data treatment such as keeping copies of data, performing unauthorized modifications, data publication, etc. Illegitimate Data Dissemination (IDD) is the case when authorized data actors disseminate or send plain text data to unauthorized third parties. Another example of data misuse is the Unauthorized Secondary Usage (USU) of data made by some CSP. Indeed, a CSP can draw incomes from the secondary usage of the data, mostly targeting advertisements or for direct profits. However, some usage may be undesirable for the data owner. We can consider the case of an enterprise which stores its confidential data in the cloud. These data have a clear business value, and so the enterprise can risk a large amount of money if these data are disclosed to a third party (for example, detailed sales data). However, to gain incomes, CSP can share or disclose these data to a third party who may be the concurrent of the enterprise.

Another concern which is caused by the lack of user control is the data loss and leakage. This presents one of the seven threats described in the CSA (Cloud Security Alliance). Data loss and leakage represents a strong barrier to adopt cloud services by enterprises and users. This can be justified as a number of incidents have occurred in cloud computing systems like:

- In March 2009, Google revealed documents saved by users of Google Docs service to third parties who do not have the permission to explore these documents [37].
- In 2010, several Hotmail accounts were hacked due to technical flaws in Microsoft software [14].
- In October 2007, a Salesforce.com employee fell victim to a phishing attack and leaked a customer list, which generated further targeted phishing attacks [37].
- In March 2009, Epic.com lodged a formal complaint to the FTC (Federal Trade Commission) against Google for its privacy practices. EPIC was successful in an action against Microsoft Passport [37].
- In 2011, Amazon customer services were unavailable for multiple days, and data were lost due to a logical flaw in the cloud storage design [14].
- In July 2007, Steven Warshak stops the government's repeated secret searches and seizures of his stored email using the federal Stored Communications Act (SCA) [37].

4.3 Dynamic nature of the cloud environment

The dynamic nature of cloud causes major problems for data privacy. Transborder data flow is one of these issues. In fact, dynamic algorithms are responsible for transfer and storage in the cloud, and thus, the user does not have any information about his data location. This issue is extremely relevant especially when data are handled in plain text form. In fact, data can be stored or processed in countries whose laws bring more privacy risks (e.g., USA-Patriot Act). Then, data became more susceptible to data disclosure issue. On the other hand, storing or processing data in some countries can present violation of legislation and user's preferences. Firstly, there are many laws which prohibited storage and transfer in foreign countries. For instance, some legislations of the European governments (e.g., France) impose that medical information should not be stored or transferred in strangers countries. Secondly, sometimes and for data criticality, the user can impose that his data must not be stored in foreign

regions. In some cases, the user can even impose that his data must be stored in the city where he resides. Particularly, when the city of Los Angelos decided to use the public cloud for mailing service and electronic agenda, it has imposed that sensitive data such as arrest records, criminal information and police officers' emails must be stored on servers inside the city [38].

Data replication is another problem that is caused by dynamic nature of the cloud. Indeed, dynamic algorithms perform duplication of data in many servers to ensure availability of services. Hence, it joins the transborder data flow problem as it is difficult to guarantee that a copy of data or their backups will not be stored or processed in some countries. Further, this problem boosts concerns about how to keep track of all the copies of data across the cloud servers. At the same time, it raises Retention (Re) issues since it is difficult to ensure that all instances of data will be deleted when it is requested by the data owner.

4.4 Compliance with laws and user's preferences

Another privacy key concern in the public cloud is the Privacy Compliance (PC). This concern can cover the aforementioned privacy issues as they can be regulated by expressing well-defined privacy policies (data owner preferences and/or regulations) and enforcing them. The effectiveness of privacy compliance almost depends on two aspects: (i) the precise definition of the privacy policies and (ii) the capability of the used enforcement mechanisms. However, most of the cloud customers does not have a clear idea about practices that data actors can perform with their private data. For example, a customer who is unaware about data collection and sharing in cloud may not specify policies to regulate these practices. Further, privacy policies are various and complex; thus, it is not trivial to express a policy that exactly meets the data owner needs for privacy. Second, the enforcement mechanism introduced till now does not consider all the cloud aspects in the enforcement process (dynamic data management, sharing pool, data transferring, lack of user control, etc.).

4.5 Accountability

Accountability (Acc) is a multi-dimension term that has multiple definitions. One common definition is given by the Galway Project in the context of corporate data governance state that [39]: "Accountability is the obligation to act as a responsible steward of the personal information of others, to take responsibility for the protection and appropriate use of that information beyond mere legal requirements, and to be accountable for any misuse of that information." Hence, the way to achieve accountability in the computing context is to introduce a strong emphasis on auditing mechanisms. These mechanisms must be able to keep track of data and to give a clear idea about all actions performed on data and who are the data processors. In the cloud context, this raises many concerns since data are randomly duplicated and transferred across the system. Thus, auditing and monitoring must be applied in an intelligent way taking into consideration all of the cloud aspects for data management.

4.6 Synthesis

As a summary, we have seen that cloud computing can magnify the existing privacy issues especially because of the lack of transparency concerning data handling and storage, the dynamic nature of the cloud, the lack of means for enforcement of privacy policy and the difficulty to achieve accountability in such environment. Figure 3 illustrates actual privacy issues in the cloud and the problems that may be engendered. Hence, it is logical that cloud customers concerns about their sensitive outsourced data are more and more increasing. In fact, a survey realized by Fujitsu Research Institute [40] conducted among more than 3000 cloud consumers found that 84% of the consumers are concerned about their data storage location and that 88% of the customers believe their data are not well protected and require more privacy. Privacy issues present a strong barrier to the adoption of cloud services and can lead to fear from data outsourcing. Thus, it is recommended to give this aspect more and more attention as it can be one of the major hindrances for cloud adoption. Thus, many potential consumers are not sure whether they can trust the cloud providers in offering dependable services [41]. In this context, several research works have been proposed to cope with privacy issues in the cloud. In the next section, we study the existing works.

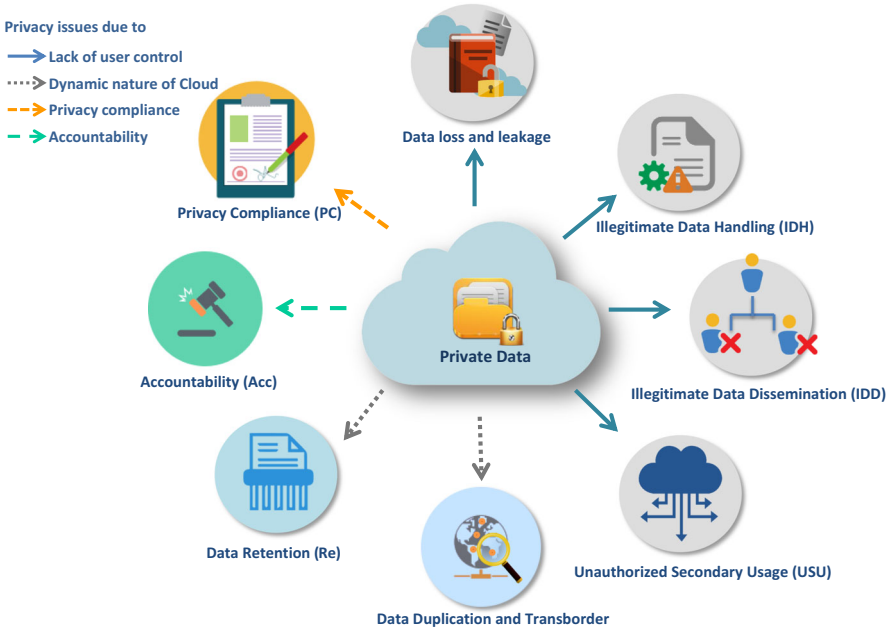

Fig. 3 Privacy issues in the cloud environment

5 Private data protection trends in the cloud: current techniques and approaches

Since the emergence of cloud computing in the IT world, several industrials and especially research actors invested efforts in tackling the privacy issues described above. This was a result of not adopting the privacy by design concept while developing the cloud computing paradigm [42]. The invested works engendered till now present various techniques and approaches that we will describe in this section.

5.1 Techniques

In this survey, we distinguish between a technique and an approach for the data privacy protection. A technique is a technical or practical method or skill for completing a specific task while an approach is theoretical ideas/actions/mechanisms intended to deal with a problem or situation related to the privacy in the cloud. In this subsection, we enumerate different techniques used to preserve privacy in cloud computing.

5.1.1 Encryption

The encryption is the most used and recognized security technique to ensure data confidentiality. It is about using a cryptographic solution to encrypt data stored in the cloud. This process could be done either by the data owner, by the CSP or by both actors. The cryptography in the cloud still represents an open issue since it requires the management of the cryptographic keys. The main question to ask is should we use symmetric key cryptography requiring the same key for encrypting and decrypting (it can be considered as a complex problem especially when different actors are involved), or should use asymmetric cryptography based on a couple of private and public keys (the main difficulty here is how many keys we have to use). Many encryption schemes that include symmetric and asymmetric encryption methods are emerged in the cloud [43]. Further questions are about the responsibility of data decryption and encryption and the key management. Some new solutions take a step further in this area. They propose to emerge an end-to-end encryption between cloud applications to reinforce user and provider trust [44]. In fact, encryption and decryption keys are located at the client side rather than the server side, and the key management is kept to an independently trusted third party data. In general, the encryption technique could be seen as a good solution for preserving privacy in the cloud; however, it can represent some limitation. In fact, in 2010, there was a Twitter incident exposing data theft attack from the cloud service provider which is Google [15]. Of course, here users do not have the possibility to encrypt their data. This task was supposed to be done by the provider, so when the attack has been done, all private data were exposed publicly. Another limitation of encrypting data in the cloud environment is that it prevents some applications from accomplishing some services such as indexing and searching since they cannot process encrypted data [45]. So if the encrypted data will be processed in the cloud, it must be decrypted first which compromise data privacy.

5.1.2 Processing encrypted data

In order to overcome the limitations of data encryption, the encrypted data processing technique was proposed to allow performing computation on encrypted data. Therefore, the cloud actor does not need to decrypt data for query execution and can execute queries directly on encrypted data. Multiparty computation such as Yao's secure two-party protocol enables a data user and a data owner to cooperate to calculate a function without data disclosure [46]. Homomorphic encryption such as Gentry's fully homomorphic encryption scheme [47] provides a general way of calculating a function of encrypted data in the cloud. It generates an encrypted result which, when decrypted, represents the result of the operations if performed on the plain text.

Processing encrypted data is a promising technique that has the potential to be used in the cloud [48]. Nevertheless, Processing encrypted data has its own disadvantages such as the high computational and bandwidth overheads as generated ciphertexts are more large and complicated [9]. Another weakness is that this technique is specific for a limited kind of processing and it is not always applicable in practice since most existing applications cannot process encrypted data and must be re-writing.

5.1.3 Obfuscation

Data obfuscation or data masking refers to the process that enables concealing sensitive information by replacing it with realist-looking values based on secret masking rules. These values look similar but are significantly different (unrelated) from the real data. Data obfuscation offers three main techniques that fall under the category of privacy preserving techniques. These techniques are detailed below.

Data randomization the data are blurred by adding either a random variable or by data discretization, for example, multiplying a column of data with a secret factor, replacing customer identities with pseudonyms or random values [49].

Data Swapping swaps entries within a single field in a record set so that the individual record entries are unmatched or intelligently swaps the pixels of an image [49]. Swapping is based on secret rules that enable the reversibility of the original data. The effectiveness of this technique depends on the amount of data (columns, pixels...) to be overlapped. This technique enables many applications to perform sufficient calculation accuracy on the degraded sensitive data.

Anonymization falls within obfuscation techniques that consists in removing personally identifiable information (PII) from a data record. Then, the data actor can use the real data without compromising data owners privacy. However, Sweeney demonstrated that this technique can be bypassed using a linking attack to identify an individual by exploring data from another known database [24]. To overcome this drawback, Sweeney proposed the K-anonymization technique. In fact, she proposed to remove the PII record from the private data and then classify the rest of data records to quasi-identifier (e.g., ZIP code, race) and sensitive attributes (e.g., disease). Quasi-identifying

attributes will be generalized and/or removed. Generalization means replacement of attributes with less specific, but semantically consistent values. The goal of k-Anonymity is that each record in the database cannot be distinguished from at least $k-1$ persons present in the same table. k-Anonymity does not provide privacy if there is a lack of diversity in a class of sensitive attributes and the attacker has a background knowledge.

The main advantage of the obfuscation techniques is that it produces usable and protected datasets that enable many applications to perform sufficient calculation accuracy [50]. However, in general, obfuscation techniques give a weaker secure protection than encryption [9]. Besides, as it is mentioned, not all applications deployed in the cloud can process the obfuscated data. Furthermore, it is clear that the anonymization concept (simple or advanced anonymization, generalization, l-diversity, etc.) target especially databases containing PII information. Other types of data are not covered by this technique (image, document, etc.). Even more, many de-anonymization techniques can recover PII from some types of anonymized data [51].

5.1.4 Sticky policy

The sticky policy is an advanced technique coming from the need to gather data with the user preferences regarding privacy. This technique allows sticking individual privacy circumstances and preferences (formatted as policies) directly to personal data while moving across the cloud [52]. Hence, data processing is permitted unless the attached policies are respected. Policy enforcement is ensured through policy management components. The Policy Decision Point (PDP) evaluates and compares processing requests against sticky policies and decides whether to authorize the access or not. The Policy Enforcement Point (PEP) is responsible for the policy enforcement.

The effectiveness of sticky policies technique strongly relies on the effectiveness of the enforcement in the cloud. Therefore, it is crucial to introduce powerful means for policy enforcement in the cloud. Further, according to Bezzi et al., the processing of sticky policies adds an important computational overhead and its applicability to realistic scenarios has to be investigated [53].

5.1.5 Trusted platform module

Trusted Platform Module (TPM) is a tamper-resistant hardware component developed by the Trusted Computing Group (TCG) [54]. This hardware component acts as the root of trust. In fact, TPM provides the ability to perform secure actions as well as identification and authentication, encryption and decryption, key generation, signature creation, hashes generation, integrity checking of software, nonvolatile storage of secrets, etc. The overall TPM hardware architecture is depicted in Fig. 4. TPM provides a shielded location to protect user's data secrets, but it is not intended to perform secure data processing [55]. The main limitation of this technique is that it is a hardware-based solution, that means that the CSP should integrate it physically when deploying their data centers.

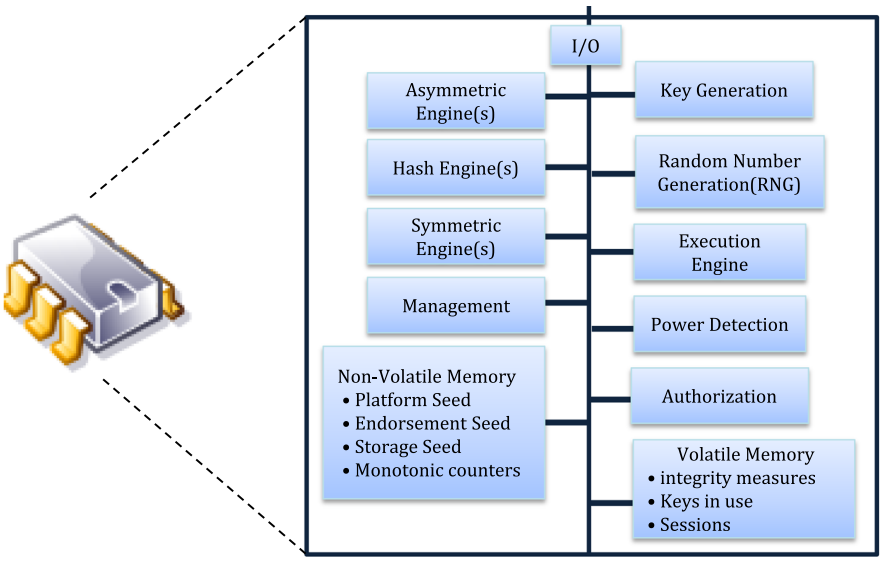

Fig. 4 TPM architecture [56]

5.1.6 Data segmentation

Data segmentation is an additional solution to ensure data confidentiality when relying on external cloud providers for storing and processing data. This technique is based on the fact that private data are not the only sensitive information; the associations between data are also sensitive. Thus, privacy can be guaranteed by storing different chunks of the data in separate non-linkable fragments [57]. For example, we consider patients private data in a hospital. The name of patients may be considered as not sensitive and so the diseases treated by the hospital, although the association between patients' names and their diseases is sensitive and should be maintained confidentially.

This technique is generally used by default in the cloud in an arbitrary manner to store data. To ensure the protection of data privacy, this technique can be optimized such a way that data segmentation is performed according to sensitive associations. The main drawback of this technique is that the loss of a chunk of data leads to the loss of the data in all. Furthermore, if data must be handled in the cloud then it must be reconstructed and returned to the initial plain text from. Hence, data segmentation is like encryption since it can be an effective solution only for storing private data in the cloud.

5.1.7 Trusted third party mediator (TTPM)

Trusted third party would act as a mediator (TTPM) between the customer and the cloud actor to ensure the policy enforcement and carry out the audit. This technique is not new, it was adopted to build online customer trust when using e-commerce applications. For instance, trusted third party can act as an anonymizer which hides

the identifying information of the customer. Moreover, TTPM was used to maintain QoS (Quality of Services) in a certain domain such as mobile vehicular cloud based on QoE (Quality of Experience) metrics [58,59]. It was also used to check cloud actor for compliance with customer's preferences when handling private data in the cloud and carry out audit and control [60]. Accordingly, trusted third party must evaluate each request for data performing which can considerably affect system performance.

5.1.8 Synthesis

A review of the existing research works shows that the presented techniques are applicable in the cloud and look promising to cope with privacy issues either used alone or combined with each other. Nevertheless, all the privacy techniques have advantages and disadvantages. We summarize strengths and weaknesses in the Table 1. Encryption, for example, can partially address the challenges associated with malicious insiders by preventing them from obtaining private data in their plain text format. However, encryption cannot be an eventual solution when data must be decrypted to be processed. Therefore, encryption could be an effective solution to the data at-rest in the cloud [61]. In this case, indexing and searching data issue can be solved by enabling searching over the encrypted data. Many researchers have realized advances in this area [62,63]. Moreover, it is required to continuously check the confidentiality and integrity of the encrypted data stored in the cloud due to its dynamic nature and the data transborder issue [64]. For data processing, encryption may be combined with many others techniques to increase efficiency, for instance, we can use encryption with sticky policies (e.g., to encrypt data and to stuck it to policies) [65] or encryption with segmentation [57]. Even more, encryption can be combined with many others security techniques. For instance, the work in [66] leverage the encryption capability with the identity management and firewall to maintain integrity and confidentiality of Business clouds.

Encryption generates several problems due to the additional complexity introduced in cloud computing models. As it is aforementioned, many questions may be raised about who is responsible for doing the encryption? Who has the right to decrypt the data? How and where can we store secrets? The TPM, for example, can be used to provide a shielded location to protect user's secrets. Thus, the cloud infrastructure could be set up as a trusted system where private data (encryption and decryption keys) are stored by trusted location registries. However, as it is aforementioned, the TPM cannot perform secure data processing. In fact, it is used only to store secret, encryption and decryption, signature creation, hashes generation and integrity checking of software. TPM can be either integrated into the cloud infrastructure or provided by a Trusted Third Party Mediator (TTPM). Further, a TTPM can serve as a privacy policy manager. Indeed, added to the fact that it can integrate a TPM to store secret, the TTPM can perform identification, authentication and compliance checking before delivering the encryption keys [60]. Moreover, it can carry out control and audit [67]. Hence, the TTPM presents useful method that can be combined with each of the presented techniques. However, the adoption of a TTPM solution adds a significant overhead caused by the communication traffic requesting the TTPM services.

Another solution that can replace the encryption is the obfuscation technique. One main strength of this technique over the encryption is its ability to provide multiple

Table 1 Privacy techniques strengths and weakness

Techniques	Strengths	Weakness
Encryption	Enables strong protection for data at rest	Data Must be decrypted to be processed
	Supports any type of data	Prevent indexing and searching
Processing encrypted data	Enables processing the encrypted data	Adds a high computation overhead
	Ensures data protection throughout its lifetime	Not always feasible in the cloud
	Supports any type of data	
Obfuscation	Enables to perform sufficient calculation accuracy masked data	Week protection than encryption
	Provides multiple obfuscation techniques	Not always feasible in the cloud
Sticky policy	Stucks data to policy	Adds a relatively high overhead
	Processing is permitted unless policies are respected	Its efficiency depend on the PEP
TPM	Provides a shielded location to protect user's data secrets	Cannot perform secure processing
		Presents a hardware solution
Segmentation	Enables protection for data at rest	The loss of a chunk of data leads to the loss of data in all
TTPM	Enables a trusted party to check for privacy compliance	Adds high-level computation due to the additional communication traffics

TPM Trusted Platform Module, *TTPM* Trusted Third Party Mediator

degrees of data protection depending on the end user needs [68]. In addition, contrary to the encryption, obfuscation enables applications to perform sufficient calculation over obfuscated data without the need to de-obfuscate it. Currently, obfuscation is not always a feasible technique since many existing applications deployed in the cloud cannot process the obfuscated data and they must be re-written. Similarly, processing encrypted data suffer from the same problem. This can present probably the main barrier to adopting those techniques. The integration of a middleware that enables traditional applications to process encrypted or obfuscated data seems likely to be a useful solution to this problem.

5.2 Approaches

As aforementioned, the approach differs from the technique in the sense that it is based on assumptions and theoretical ideas to solve a problem. The literature presents many approaches that tackle the problem of private data protection in cloud environments. In this section, we will provide an overview of the current data protection approaches. These approaches employ one or more of the techniques presented previously. We classify these approaches into four categories: (i) data-centric approaches that focus on how to allow data to protect itself, (ii) user-centric approaches that target users involvement in the data protection process either by policy specification, data encryption or data obfuscation, etc., (iii) CSP-centric approaches that target mechanisms and frameworks integrated in the cloud infrastructures to ensure privacy and (iv) hybrid approaches which combine two or more types of approaches. In the following, we enumerate some relevant research works according to this categorization.

5.2.1 Data-centric approaches

Data-centric approaches emphasize mechanisms and techniques to automate sensitive data protection anywhere in the distributed systems. One of the relevant data-centric approaches is the self-protecting data solutions. These solutions enable data to defend itself even though it is located in an untrusted environment. Squicciarini.AC et al. introduce the Self-Controlling Objects (SCO). This object encapsulates sensitive data along with their policies and assures their protection by means of object-oriented programming techniques [69]. The SCO policy concerns data owner policy and legal policy. Each time the access to the SCO protected content is attempted, its policy is evaluated according to the requester's credential and location. The policy enforced by the SCO comply with the legal regulations of the SCO storage location. For example, the SCO is stored on an EU server, and therefore, the data disclosure laws are applied. The content is then rendered according to the granted privilege. The SCO manages copies of data synchronizes and updates it if there was a change. The authors discuss possible attacks to the SCO such as reverse engineering of JAR, policy modification, unauthorized copy of protected content and bypassing authentication. The authors prove that SCO can address these security issues. However, even it can manage copies of data, SCO cannot retrieve plain text once delivered to the data actor.

Chen et al. present in [70] an open and flexible software solution called SelfProtect Object (SPO). The user provides the sensitive data and the corresponding policies (specified in XACML) to the SPO generator. This latter bundles the data content and the policy files in an object (SPO) that can protect its content by itself anywhere and anytime. Additionally, the SPO incorporates policy management components (Policy Enforcement Point, Policy Information Point, Policy Decision Point) to maintain the policy enforcement. The SPO concept is built on .NET platform and is presented as a .dll extension.

In a later work, the authors enhance the SPO model to prevent unauthorized uses by authorized parties [71]. They introduce a generic scheme called SafeProtect that leverage the SPO capability by the use of a hardware-based TPM module called the Trust Extension Device (TED) to enable secure data sharing. The device attests with a certifying cloud-based authority service that it is legal and valid, verifies data owner and data actor identification and decrypts the data if so. This hardware must be owned by all data owners and data users to enable secure data sharing and access and to prevent dishonest authorized users from illegally redistributing sensitive data to unauthorized parties. The solution introduces a monitoring service as a cloud-based storage service that stores application-based actions carried out by data consumers. To demonstrate the proposed ideas, the authors implement a plugin for Microsoft Word to enables policy interpretation and enforcement for data stored in the TED device. However, the solution is based on the fact that the cloud applications that process private data are trusted and are able to carry out the policy enforcement. Nevertheless, this assumption is not logical since, in general, we cannot trust in applications we not own or operate.

5.2.2 User-centric approaches

This kind of approach focuses on how to protect data from the data originator side which is involved either in the expression of his preferences or in the consideration of legal texts. Some of this category of work try to define advanced tools to enable the defining policy for protection data when processed in the cloud. Researchers in [72] come up with a toolkit for automating compliance of cloud computing services. This toolkit allows the semantic annotation and natural-language processing of policy texts (regulation text and/or user's preferences text) to generate machine-readable rules. The implementation of the proposed tool is incorporated in the EnCoRe Policy Enforcement Framework [73].

In [74], the author proposes an approach to achieve compliance for the sharing and the disclosure of patient data between the Member States across Europe in a health grid. This approach is based on the semantic modeling of privacy obligations of legal, ethical or cultural nature. Indeed, the author aims to formalize privacy policies intercepted from the EU directive through the use of ontology modeling (OWL) and semantic web rule language (SWRL). These policies are then mapped to the XACML language in order to be enforced in the cloud.

Privacy Manager presented in [68] is another user-centric framework that offers different ways to protect sensitive data through several features. Obfuscation feature allows obfuscation of data before being sent to the cloud. The result of processing

done on the obfuscated is de-obfuscated by the privacy manager to generate the correct result. Preferences feature allows the users to define their preferences regarding the handling of data. These preferences can then be bound to data sent to the cloud to form sticky policies. Personae feature allows the use of the appropriate personae when interacting with cloud services. Privacy Manager can be deployed under different possible architectures in the user side, in a private cloud, and in a hybrid cloud. To prove their solution, the authors look at several different privacy manager operations for Online share portfolio, Magcloud and Printcloud, and cloud photo application. They also provide an analysis of the scalability and efficiency of this approach within such scenarios.

In general, this category of work emphasizes especially the user implication for policy specification. These works remain incomplete since they lack efficient frameworks or mechanisms to ensure enforcement of the generated policies in the cloud environment.

The previous user-centric research works tackle the specification of privacy policy in general; other user-centric solutions consider the data protection when it is stored in cloud servers (data-at-rest). The work in [57] defines a set of confidentiality requirements based on encryption (with indexing) and fragmentation to protect data confidentiality. These requirements are based on the fact that a given attribute is not the only sensitive information. In fact, the associations among some attributes are also sensitive. Hence, data segmentation will be performed according to the defined sensitive attributes and associations. The authors apply the defined requirements on different protection paradigms. These paradigms are to be used by the user to make chunks of data and then send them to the cloud. The keys of encryption and fragmentation are owned by the user and are not released to the data actors.

Li et al. [61] come up with an intelligent cryptography scheme entitled Security-Aware Efficient Distributed Storage(SA-EDS) to securely store data in the cloud. This model consists of two components: (i) the Deterministic Process (DP) intelligently retrieves sensitive data from overall data packets and (ii) the Data Distributed Storage Process (D2SP) that includes two steps: (1) divides the private data and separately stores it in the distributed cloud servers in a first step, (2) retrieves the data from the cloud and merges it to obtain the original data. The Deterministic Process (DP) and the Data Distributed Storage Process (D2SP) component are to be introduced in the user host to maintain secure data storage. The authors propose three main algorithms to implement their proposed scheme, which includes Alternative Data Distribution (AD2), Secure Efficient Data Distributions (SED2) and Efficient Data Conflation (ED-Con) algorithms. They evaluate their approach and prove that the proposed scheme consumed less computation time than AES for data encryption and data retrieval processing.

Bahrami et al. [75] propose the cloudPDB: a light-weight data privacy scheme for cloud-based databases that scrambles data on each selected bucket (multiple records, or fields) of a database. The proposed scheme includes several components and algorithm, to securely outsource data to an untrusted cloud. The evaluation of the implemented scheme on a TPC-H database with different query size shows that it provides a better performance over the AES encryption scheme for data protection at-rest.

5.2.3 CSP-centric approaches

The CSP-centric approaches focus on mechanisms and framework to be deployed and integrated into the CSP infrastructure. In this kind of approaches, the user's involvement is negligible compared to the CSP one. The work presented by Yau.S et al. [76] aims to protect data from the CSP. They propose three conditions under which the CSP may reveal the private data: (i) CSP knows the storage location data, (ii) CSP has the privilege to access and collect data and (iii) CSP may understand data. The authors believe that if they can prevent the CSP to fulfill the three conditions simultaneously, they can protect data. For that reason, they propose to separate software service providers and infrastructure service providers (data storage) in the cloud, anonymize data and finally, integrate a middleware in the cloud infrastructure to allow the use of the system resources. Further, the middleware sets up an encryption key with the user, encrypts any data being stored in the physical storage of the cloud or being transmitted through the network, and obfuscates sensitive data being processed in the cloud service. The authors illustrate their solution with an example of the online video conference to demonstrate how the approach can protect data in cloud computing system. However, the authors do not present how to implement the proposed cloud architecture. Furthermore, it is not clear when and how the proposed middleware encrypts and obfuscates data.

Another category of work relies on tamper-proof facilities to solve the problem of securely processing sensitive data in the cloud computing infrastructures. For example, the work presented in [77] defines a novel service that offers a set of security protocols to improve the privacy of data in the cloud. This service allows the data storage and processing with full security taking the advantages of inviolable capabilities of a cryptographic coprocessor. The proposed protocols define data and software transfer to the cloud, define steps performed by coprocessor to execute software and enforce privacy of data and feedback of user's data usage. The authors present a prototype implementation of the proposed protocols which tested on a banking application. The authors present an economic study of the solution that show the high cost of the usage of coprocessors for privacy preservation. In fact, if the cost of the used coprocessors is $120,000, CSP incur $60,000 (50%) and the rest is divided among security service customers $60,000. The authors explain that this cost is very reasonable in return of the privacy service provided.

5.2.4 Hybrid approaches

A hybrid approach is a solution that combines two or more of the categories already presented. Moreover, data, user and CSP, all or some of them can be involved in the data protection process. Some of the hybrid approaches rely on sticky policy technique and political management components to ensure the policy enforcement in the cloud environment. For instance, authors in [65] propose to introduce a data protection module called CDPM (Client Data Protection Module) deployed on the user side to allow the policy specification. This module generates the PDE (Privacy Data Envelope) [78] which present an encrypted envelope that embodies sensitive data raw coupled with the associated properties and policies. A specific virtual machine

called PaVM (Privacy-aware VM) is introduced to enable the policy enforcement in the cloud based on the political management components. The authors implement a java prototype of the proposed solution. They consider an e-commerce enterprise that operates in different countries as an illustrative example. Performance tests done show that overhead ranges from a factor 2.6 w.r.t native case in a local system. For a remote user, only the overhead percentage of each component is presented, the global overhead is not mentioned. In the same direction, Trabelsi et al. present in [79] a security service called SPACE based on the sticky policy technique. This service incorporates an enforcement engine whose main functionality is to perform access and usage control of the private data and enforce associated obligations. This engine is composed of political management components (PEP, PDP) to enforce policy and an obligation handler to manage associated obligations. The authors propose the implementation of several maps (access map, history map and geolocalization map) to allow real-time data control.

The approach presented in [60] introduces a trusted third party to ensure the policy enforcement rather than relying on the policy management components that can be mistrustful. Using this approach, the user must generate policies which will be stuck to his sensitive data. Once the data consumer wants to use the data, it must send an authorization message to a trusted authority. The latter evaluates this request with respect to the specified policies and send the decryption key in the case of policies compliance. The authors present different implementation of the proposed sticky policy management approach using PKI-based solution, IBE-based solution, leveraging secret solution, multiple trust authorities and partially trusted trust authorities.

Brown et al. [80] address the problem of controlling information sharing in heterogeneous distributed environments where hosts might have very different hardware and software architectures. The proposed solution aims to allow different combinations of trusted components to meet the requirements for managing sensitive information. For this, the data owner defines two types of sticky policy: (i) hosting policies are used to define entities that are able to provide local enforcement and only further transferring data under the same policies, and (ii) usage policies define the users action capabilities based on their attributes. This approach is enabled by certified attributes that the systems can present when requesting sensitive information. The authors implement a prototype based on Microsoft Excel add-in to demonstrate application level enforcement of policies. The remote trust is established using the TPM 2.0 simulator provided with the TSS.NET library for Windows. The performance tests done show that a full decision on 1000 resources need over twenty seconds.

Brandic et al. present a novel approach for compliance management in cloud environments that rely on fragmentation to achieve data privacy [81]. The authors propose a novel language for specifying compliance requirements based on a model-driven technique using Unified Modeling Language (UML). They further propose to introduce the C3 middleware that is responsible for the deployment of certifiable and auditable applications in the cloud infrastructure. This middleware is based on the usage of a certification mechanism for authentication and compliance management to help the cloud users to select the cloud provider that is susceptible to be compliant. The authors underpin their approach with an illustrative use case discussing how the C3 framework consider compliance management of security, privacy and trust in cloud environment.

The presented solutions offer high-level software to enforce a high-level privacy policy. The compliance with these policies is enforced in a preventive way, where the violations of policies are prohibited from accessing to data. Nevertheless, this type of enforcement cannot prevent unauthorized uses of sensitive information after gaining access to the data. Current solutions in this area consider that the System Call Interception (SCI)technique presents till now the best solution to face this issue. Indeed, this technique can detect and prevent a non-desired behavior at the "intention" stage. UC4Win [82] is a data loss prevention solution for Microsoft Windows operating systems that is based on the SCI technique. The solution intercepts application calls to the Windows API, evaluates their policy compliance, and blocks or modifies them upon detected policy violations. A dynamic data flow model is incorporated to track flows of sensitive data through the system. The authors introduce a detouring library that implements the monitoring functionality. They study the ability of UC4Win to face attacks on the policy enforcement, on the policy evaluation and on the availability. The executed performance tests show that for opening a file, adding some characters and saving it there is 17.32% computation time overhead without applying policies and 22.12% with policies. Yet, one major weakness of this solution is that the proposed data flow model is designed to the local system. Hence, it cannot track data when it is sent via the network. This means that the proposed model cannot satisfy the constraint of distributed environments as well as the cloud computing.

To deal with this shortcoming, work in [83] come up with a distributed data flow tracking model. This model extends a generic model for intra-system data flow tracking to the cross-system case. The model makes it possible to transfer usage control policies along with the data moving in the distributed environment and to be aware of the existence of copies of the data across the distributed system. As a proof of concept, the authors concretize "transfer of data" usage to the Transmission Control Protocol (TCP). They propose to use Systrace for system call interception and policy enforcement on the OpenBSD OS. They tested the proposed solution against non-TCP Communication, Portable media, Fool Infrastructure, a man-in-the-middle and Denial-of-Service. The calculated performance overhead ranges from a factor of 0.14 to 11.66 in the best case and between a factor of 0.65 to 13.15 in the worse case.

In a later work [84], the authors enhance the proposed model to tackle the problem of enforcing global data usage control policies when it refers to events happening within several distributed systems. For this, they propose a fully decentralized infrastructure for the preventive enforcement of data usage policies. The enhanced model not only tracks the flow of usage controlled data within and across systems but also coordinates the decision process of multiple distributed and independent decision points. The authors provide an evaluation of communication overhead when enforcing policy and the PDP decision overhead.

Several approaches believe that preserving data privacy on software layer may not be sufficient. Chen et al. [85] present DataSafe, a software-hardware solution for protecting the confidentiality of data when processed by unvetted applications. The main purpose of this architecture is to prevent illegitimate secondary dissemination of protected plain text data by authorized recipients. For this, DataSafe enables translation of high-level software policies to efficient hardware tags at runtime. It propagates these hardware tags whenever data flow related hardware instructions are executed. DataSafe

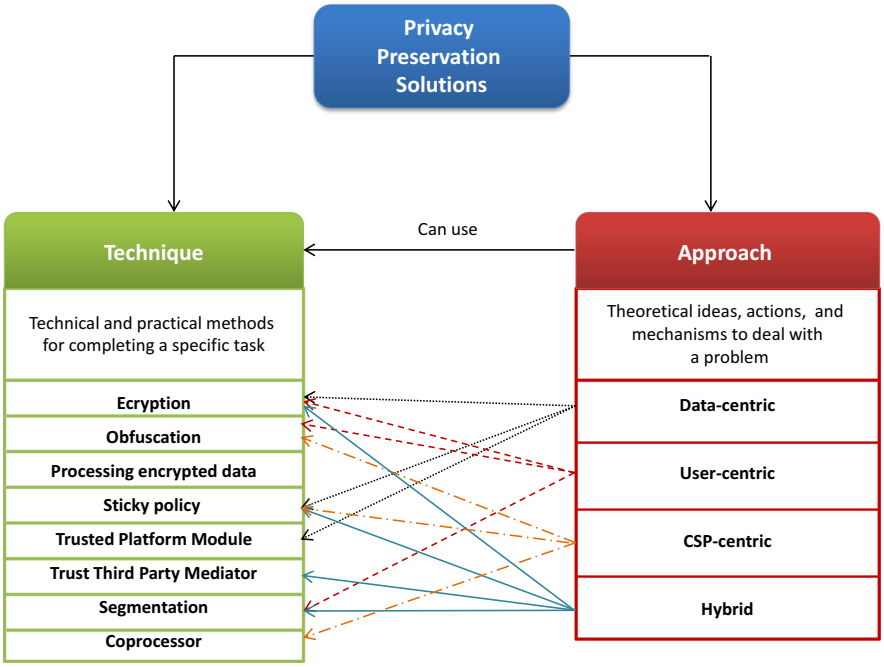

Fig. 5 Classification of privacy preservation solutions in the cloud

is built upon additional hardware as well as an output memory map (new hardware structure inside the processor), a trusted hypervisor, which are expected to enforce and to propagate the hardware tags. These devices are to be deployed on all devices that are in prospect to use the protected data. As a matter of fact, the implementation of this approach in the context of cloud must be investigated as the overall cloud infrastructure must be changed to include the proposed hardware. Yet, DataSafe is not able to enforce high-level obligations such as the deletion of private data after the retention duration. Moreover, a prototype implemented in the Legion simulator of the OpenSPARC platform shows that there is an increase of 50% in the software complexity cost (hypervisor code base) and 15.6% increases in the storage overhead for a 10-bit tag per 64-bit word. This presents important complexity and economic overheads.

In this section, we have given an insight of different methods and approaches to tackle privacy issues using different techniques as illustrated in Fig. 5. In the next section, we will discuss the advantages and limitations of each of the presented methods to try to completely solve privacy issues.

6 Evaluation

In this paper, we have enumerated the state of the art about techniques and approaches proposed to overcome the privacy issues in the cloud. These works have shown that

most of these problems were recognized and that there was a great progress in this area. In this section, we assess these works and discuss what more is needed to enhance privacy in a cloud environment. We evaluate each research work according to relevant technical criteria that ensure data safeguard in the cloud (Table 2). We also review the proposed approaches according to the addressed privacy issues (Table 3).

6.1 Technical feasibility

We have studied the techniques adopted within the current approaches to cope with the privacy issues in the cloud computing. We can note that all presented works endorse one or more of the techniques presented previously. As it is presented in Table 2, the most used techniques for privacy preservation in the cloud are probably: the encryption, the sticky policy and the privacy policy alone without being bound to data. This is due to the fact that these techniques look promising to address the privacy problems in the cloud with reasonable complexity, performance and economic overheads. From another side, some approaches come up with concepts and assumptions that are difficult to apply in a real context. For example, the solutions [76,77,85] introduce radical low-level modifications like change the cloud architecture or integrate new hardware devices (memory, hypervisor, coprocessor, specific processor). These solutions represent a very important overhead that makes them hard to be realized.

6.2 Policies specification

To ensure a better protection of his private data in the cloud, the data owner must make a precise specification of his policies and also must consider eventually legal policies. From the Table 2, we can notice that not all approaches consider the specification of the two types of policies [65,68,69,72,81]. Some other works do not consider the privacy policy specification. This is because they rely on other techniques like segmentation, obfuscation and encryption to protect data only when it is at rest [57, 61,75] and/or in use [76]. The rest of the approaches consider only the data owner policies. Some of these works suppose that the data owner policies are already specified [60,69–71,80,83–85]. Other works introduce new tools and languages to facilitate the policy specification [65,72,79,81]. According to our study, we have perceived that the proposed tools suppose that the data owner has the ability to find the appropriate policies for his context of privacy. However, most of the cloud customers are unaware of what the cloud actors can do with their data. In fact, cloud actors may, implicitly, perform undesirable uses of data (e.g., secondary usage of data, data dissemination, and data collection and share). For this, the proposal of a tool that guides the data owner while specifying his policies has become a high priority.

6.3 Enforcement level

The privacy enforcement can be done at different levels including the high-level software, low-level software and hardware. The high-level software solutions introduce

Table 2 Data protection approaches in the cloud: Technical synthesis

	Approach	Applied techniques	Policy specification		Enforcement level			Data Flow tracking		Data lifetime protection		
			Legal policy	Data owner policy	High level software	Low level software	Hardware	Local	Distribute	In-transit	In-use	At-rest
D-c	[69]	Ecry-SP	✓	✓	✓			✓	✓	✓		✓
	[70]	Ecry-SP		✓	✓					✓		✓
	[71]	Ecry-SP-TPM		✓	✓		✓			✓		✓
U-c	[72]	SP	✓	✓								
	[74]	n-a	✓									
	[68]	Obf-An	✓	✓						✓	✓	✓
	[57]	Ecry-Seg								✓		✓
	[61]	Ecry-Seg								✓		✓
	[75]	Obf								✓		✓
C-c	[76]	Obf-An			✓		✓			✓	✓	✓
	[77]	Co-PP		✓			✓			✓	✓	✓
Hybride	[65]	Ecry-SP	✓	✓	✓					✓		
	[79]	SP		✓	✓							
	[60]	Ecry-SP-TA		✓	✓					✓		
	[80]	PP		✓	✓					✓		
	[81]	Seg-PP	✓	✓	✓					✓	✓	✓
	[82]	PP		✓		✓		✓		✓	✓	✓
	[83]	PP		✓		✓		✓	✓	✓	✓	✓
	[84]	PP		✓		✓		✓	✓	✓	✓	✓
	[85]	SP		✓		✓	✓	✓	✓	✓	✓	✓

D-c Data-centric, *U-c* User-centric, *C-c* CSP-centric, *SP* Sticky Policy, *Seg* Segmentation, *Co* Coprocessor, *Encry* Enryption, *An* Anonymization, *Ob* Obfuscation, *TA* Trusted Authority, *PP* Privacy Policy, *TPM* Trust Platform Module

Table 3 Review of the privacy issues concerned by the presented approaches

	LUC			PC	Acc	Re
	IDH	IDD	USU			
[69]	−	−	−	+	+	−
[70]	−	−	−	+	−	−
[71]	−	−	−	+	+	−
[68]	−	−	−	−	+	−
[57]	+	+	+	−	−	−
[61]	+	+	+	−	−	−
[75]	+	+	+	−	−	−
[76]	+	+	+	+	−	−
[77]	+	+	+	+++	+	−
[65]	−	−	−	+	+	−
[79]	−	−	−	+	+	+
[60]	−	−	−	+	+	−
[80]	−	−	−	+	−	−
[81]	+	+	+	+	−	−
[82]	+	+	+	++	−	−
[83]	+	+	+	++	++	+
[84]	+	+	+	++	++	+
[85]	+	+	+	+++	++	−

LCU Lack of User Control, *IDH* Illegitimate Data Handling, *IDD* Illegitimate Data Dissemination, *USU* Unauthorized Secondary Usage, *PC* Privacy Compliance, *Acc* Accountability, *Re* Retention
− not supported, + supported, ++ sufficient, +++ robust

software that enable enforcement of high-level policies at application level [60,65,69–71,79,80]. One shortcoming of such solutions is the loss of the control over data once the usage request is checked and the data are released (e.g., the data actor can keep copies of data, send data to unauthorized third parties). As it is shown in the Table 3, the privacy compliance (PC) of these approaches is done at the application level, and thus, it is not strong enough (PC +: supported) to prevent Illegitimate Data Handling (IDH), Illegitimate Data Dissemination (IDD) by authorized user and Unauthorized Data Usage (UDU) issues (−: not supported). To enhance the high-level enforcement resilience, the solution presented in [81] introduces a middleware that is responsible for the deployment of certifiable and auditable applications for data processing in the cloud. This enhancement allows a stronger PC (+: supported) that enables IDH, IDD and USU issues prevention (+: supported). The approach [76] relies on the obfuscation technique and on a middleware integrated into the cloud infrastructure to ensure PC (+: supported) and prevent IDH, IDD and USU issues (+: supported). Nevertheless, the efficiency of the PC of these two solutions mainly relies on the trustworthiness and the resilience of the introduced middlewares.

The low-level software solutions rely on low-level policies that are enforced at the operating system level [82–84]. The System Call Interception (SCI) technique enables such solutions by intercepting system calls for data usage and checking their compliance regarding the low-level policy. Hence, these solutions can detect the intention of data misuses and prevent them after the data release. This explains the ability of the SCI-based solutions to ensure the (PC) (++: sufficient) and to prevent IDH, IDD and USU issues (+: supported) (Table 3).

The hardware-level enforcement approaches are solutions that make changes in the hardware system by integrating new hardware components [77,85] or by modifying the cloud architecture [76]. According to Table 3, the solution based on the integration of new hardware devices provide, in general, a better resilience compared to the software-based solutions. This except the work proposed in [71] because the integrated hardware devices (TPM) is only used for authentication purposes. Yet, the integration of new hardware devices must be investigated as it adds a high complexity and economic overheads. The modification of the cloud architecture did not improve significantly the PC efficiency [76].

6.4 Data flow tracking

The data usage generates multiple copies and derived data in different locations across a computing system. The usage control and the policy enforcement have to be enforced for all the copies of data and the derived data through the overall system. Hence, it is crucial to track data through the system to ensure privacy preservation. For a single system, a data flow tracking model tracks the flow of data within and across the local system layers [82]. Yet, this model cannot track data which are transmitted via the network. In distributed context, the data flow tracking is even more necessary because data can be scattered across different connected hosts. Many approaches introduce dynamic models to take care of distributed data tracking to enable the local usage enforcement in each host [83–85]. As depicted from Table 3, the data flow tracking-based solutions can efficiently achieve accountability (++: sufficient) since they are able to keep tracks of where data has been outsourced, who processed it, for what purpose, etc. Still more, some of these solutions enable the retention obligation enforcement [83,84]. Some other approaches introduce simple monitoring and auditing techniques for the data tracking to ensure accountability [65,68,69,71,77,79,80].

6.5 Private data lifetime protection

The private data must be protected throughout its lifetime, whenever it is in-transit, in-use or at-rest. Data in transit are the data being transmitted internally or externally from one host to another over different networks. Large amount of mechanisms and approaches are provided to ensure a secure traffic of private data based on encryption [60,65,69–71], obfuscation [68,75,76], and segmentation [57,61,81].

Data in use are the data being processed. This state is the most vulnerable one as data are, in general, processed in plain text form. We can perceive from Tables 2 and 3 that some of the solutions based only on high-level software enforcement cannot permit an effective policy enforcement when data are being processed [60,65,69–71,79,80]. In fact, as it is aforementioned, the data usage control is lost once the data requester retrieve the sensitive information. Then, they cannot prevent IDH, IDD and USU issues (−: not supported). To deal with these issues when processing data, some approaches introduce new frameworks for applying obfuscation on sensitive information [68, 76]. The solution presented in [77] relies on cryptographic coprocessor capability. Nevertheless, one major drawback of this approach is that coprocessor is expensive

(range in price from several hundreds to several thousands U.S. Dollars) and it must be optimized to handle a large number of operations; otherwise, it can slow down transactions.

Data at rest are when the data are stored either in local or in remote servers. Several solutions focus on how to efficiently protect data stored on untrusted servers. These approaches rely basically on obfuscation [68,75,76], encryption and/or segmentation [57,61]. From the Table 3, we can notice that these approaches prevent IDH, IDD and USU issues since data are only stored and not handled in the cloud servers.

In this section, we have presented several research works that aim to address privacy issues and we have assessed each of these them according to technical criteria. We have concluded that such a criteria have an influence on their ability to face some privacy issues. In the next section, we will provide guidelines for cloud privacy technology enablers to establish mechanisms taking into consideration all needed aspects for the privacy safeguard in the cloud.

7 Guidelines for assuring data privacy in the cloud

Tackling privacy aspects in cloud environments require the consideration of some good practices and advices. These practices are highlighted when studying the privacy issues, techniques and approaches. From our point of view, achieving the privacy protection in the cloud environment is based on four main aspects: (i) how to define the data owner's preferences and to consider legal policies, (ii) how to enforce the defined policies and to track data flow among the system and (iii) how to ensure accountability processing in the cloud.

7.1 Policies specification

The user may not be aware of risks that can threaten his privacy when transferring his data to the cloud. These risks come either from outside of the cloud or from the cloud provider itself since he can use data for his personal benefit. For example, the user may have no idea that a CSP can share, disclose and use his data to gain revenues from targeted ads. Also, users may have no idea that his data may be stored in some foreign countries where local laws can affect data privacy. For this reason, it is necessary to guide the user to define all necessary policies that will cope with forbidden practices that may menace his privacy. Particularly, these policies may include:

- requiring the declaration of the usage purpose before any access to data.
- identifying authorized and forbidden data processing (collect, share, copy...).
- defining geographical authorized regions for storing data (including replication, backups, and others) or for application and process accessing personal data.
- identifying the duration of data retention and removal.
- requiring notification before any changes in any situation.

It is also required to consider legal policies and to integrate them with the data owner's preferences. The user can also define some sanctions in case of policies ignorance or violation. It is recommended to merge all the specified policies into a contrac-

tual agreement to establish accountability in the form of an enforceable commitment overlooked data handling (for instance, it is possible to integrate these policies in the Service Level Agreement-SLA). After ensuring a complete specification of policies, it is necessary to provide a machine-readable policy for bridging the level gap between the high-level policies and the technically enforceable policies.

7.2 Policies enforcement

One other crucial aspect for achieving privacy compliance is to ensure policies enforcement in the cloud infrastructures. Hence, it is required to accommodate intelligent software and mechanisms within the cloud to automatically intercept requests for private data on a low level in the system, evaluate them and make a decision regarding the defined policies. Besides, these mechanisms must continuously track copies of data and derived data and ensure their protection like the originate data. An existing law (policy) can prevent and dissuade someone from the violation; however, a non-enforced one is only worth the paper it is written on. That's why the policies enforcement is a very important step to tackle privacy issues in the cloud. Besides, the policy enforcement opens the door to the control and audit process.

7.3 Control and audit

The concepts described above are necessary to decrease the privacy risks in the cloud but they are not sufficient. Indeed, the risk of privacy breaches is always present regardless of the strength of the used techniques. The only way to cope with this is to introduce elements of audit and control to keep track of how data are processed, for what purpose, where the data have been outsourced, etc. Moreover, by introducing audit in the cloud services, we can address the accountability issue. This must be achieved by the user or by a third party that he trusts in to look after his privacy interests. It can be seen as an added value service that can be integrated into a global business model. Anyway, the user should be able to view and be consent of what is really performed on his personal information stored in cloud area. From another side, by means of audit and control, we can identify and prosecute the accountable entity in case of policies tampering. Moreover, audit and control offer to the CSP providing services an additional credibility and valuable marketing arguments.

7.4 Hybrid approaches

Data protection in the cloud is a complex process that requires the intervention of all involved actors ranging from the data owner, data consumer, CSP and all possible involved actors. For this reason, we recommend the usage of hybrid approaches. In fact, these approaches may combine many involved actors in the data protection process. Particularly, the data owner must carry out all required actions to protect his data, (specify preferences, encrypt data, make sticky policy, etc.). From another part, the

CSP must integrate software and frameworks intended to enforce the privacy policy in the cloud. He must also adopt mechanisms for control and audit.

8 Open issues

Although the guideline presented previously can theoretically fulfill many privacy challenges in the cloud, there are still many open issues in the practices that need to be resolved.

Firstly, the lack of user control in cloud paradigm still need great efforts to be resolved. In fact, the presented works try to solve this problem by enabling the data owner's policy enforcement and auditing to ensure that data are well protected and controlled. Yet, as long as there is a lack of visibility and transparency in the processing of data, there is still degrees of vulnerability that can threaten data privacy. Moreover, the user consent about data handling presents a legal requirement. Hence, many research works are required to retain control for data owner when his data are stored and processed in the cloud environment.

The second challenge is related to the policy enforcement within the dynamic cloud infrastructure. In fact, the data usage in the cloud engenders many copies of data and derived data that must have the same degree of protection as the original data. Absolutely, the leak of one copy of data means the leak of the data at all. However, it may be difficult to track all generated data and maintain the same enforcement level. This represents a very challenging issue since for many organizations or enterprises; the data leakage can be catastrophic as it can engender important loss of money and reputation.

Furthermore, we discuss the challenges of the lack of awareness about the impact of privacy problems and the CSP incentive for adopting privacy enhancing technologies in his system. The lack of awareness about the impact of privacy problems from the part of users makes the data protection in the cloud harder and harder. In fact, the majority of users do not know the real value of their private data and the risks that can threaten them if these data are misused. This category of users tends to easily disclose their private data to benefit from the remote cloud services. Actually, the cloud data actors may perform implicit usage of data (e.g., secondary usage of data, data dissemination and data collection) that sometimes would be very unwelcome by the data owner. From another side, sometimes the CSP warns his customer about some risks when using his services in the general conditions of sale (GCS). At the same time, the CSP can express his discharging from consequences. However, a few of the customers read the GCS before accepting it. From another side, as it is known, the CSP always seeks for realizing a maximum of profit with the minimum of expenditures. So, the main questions here are: how should we convince a CSP or a data processor, in general, to invest for introducing such technologies in his infrastructure? Are the data owners ready to pay for achieving privacy of their personal data? To solve this problem, it is recommended to optimize existing technical solutions to avoid the important cost and computing overheads. For professionals, the enterprises and the organizations can monetize the value of their confidential information as they can risk a very important amount of money if such data are lost or disclosed. Indeed, the enterprises can risk

their reputation in case of inappropriate usage of their customer data. For example, in the case of hospitals or banks, the reputation is the basis of their business. For these reasons, enterprises and organizations may pay for better protection of their data.

We can conclude that the privacy protection in the cloud requires further research efforts. More approaches and solutions are needed in order to address all privacy issues in the cloud and to fellow its evolution. These solutions must consider all cloud aspects in the protection process to enables stronger privacy enforcement. Nevertheless, this remains difficult since privacy by design is not adopted during the design of the cloud environment.

9 Conclusion

In this paper, we studied the different risks and issues that threaten privacy in cloud computing environments. An investigation effort conducts us to fly over the existing solutions that we classified into techniques and approaches. The study of these research works has shown that the privacy in the cloud still has some open issues. These issues are either technical like policy enforcement effectiveness and data flow tracking or strategic like the awareness and sensitivity about privacy importance and the involvement of the CSP in the privacy preserving process. In fact, the categories of privacy preserving approaches we have proposed have been defined mainly according to the degree of involvement of the data, the CSP and the data owner. That's why we believe that the hybrid approach is the best approach that can handle the privacy issues in the cloud. Thereby, we are currently working on a new hybrid approach allowing a policy based end-to-end solution for preserving private data in the cloud. This approach will learn from the limitations of existing ones in order to overcome the main obstacles. However, there is a claim that someone should avoid when proposing such solution which is: there could exist one generic solution that covers all privacy issues and threats in the cloud. In fact, the variety of private data we can find in the cloud (photos, documents, collected databases, etc.) and the variety of processing and execution context and services in the cloud make a generic solution hard to find. Some assumptions have to be done in order to come up with a feasible solution.

References

1. Mell P, Grance T (2011) The NIST definition of cloud computing
2. Sellami W, Kacem HH, Kacem AH (2014, December) Elastic multi-tenant business process based service pattern in cloud computing. In: 2014 IEEE 6th International Conference on Cloud Computing Technology and Science (CloudCom), pp 154–161
3. Ali M, Khan SU, Vasilakos AV (2015) Security in cloud computing: opportunities and challenges. Inf Sci 305:357–383
4. US Privacy Protection Study Commission (1977) Personal Privacy in an Information Society-the Report of the Privacy Protection Study Commission
5. Directive EU (1995) 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Off J EC 23(6)
6. Act HIPAA (1996) Health insurance portability and accountability act of 1996. Public Law 104:191

7. Code US (1999) Gramm-Leach-Bliley Act. Gramm-Leach-Bliley Act/AHIMA, American Health Information Management Association
8. Pearson S (2009, May) Taking account of privacy when designing cloud computing services. In: Proceedings of the 2009 ICSE workshop on software engineering challenges of cloud computing. IEEE computer society, pp 44–52
9. Mowbray M, Pearson S (2012, September) Protecting personal information in cloud computing. OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Springer, Berlin, pp 475–491
10. Shankarwar MU, Pawar AV (2015) Security and privacy in cloud computing: a survey. In: Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014. Springer International Publishing, pp 1–11
11. Xiao Z, Xiao Y (2013) Security and privacy in cloud computing. IEEE Commun Surv Tutor 15(2):843–859
12. Alneyadi S, Sithirasenan E, Muthukkumarasamy V (2016) A survey on data leakage prevention systems. J Netw Comput Appl 62:137–152
13. Subashini S, Kavitha V (2011) A survey on security issues in service delivery models of cloud computing. J Netw Comput Appl 34(1):1–11
14. Zhou M, Zhang R, Xie W, Qian W, Zhou A (2010) Security and privacy in cloud computing: a survey. In: IEEE 2010 Sixth International Conference on Semantics Knowledge and Grid (SKG), pp 105–112
15. Jyothi P, Anuradha R, Vijayalata DY (2013) Minimizing internal data theft in cloud through disinformation attacks. Int J Adv Res Comput Commun Eng 2(9):
16. Gholami A, Laure E (2016) Security and privacy of sensitive data in cloud computing: a survey of recent developments. arXiv:1601.01498
17. Hussein NH, Khalid A (2016) A survey of cloud computing security challenges and solutions. Int J Comput Sci Inf Secur 14(1):52
18. Khan MA (2016) A survey of security issues for cloud computing. J Netw Comput Appl 71:11–29
19. Warren SD, Brandeis LD (1890) The right to privacy. Harvard Law Review 4:193–220
20. Farlex INC (2009) The free dictionary. Retrieved 28 June 2012
21. Dictionary B (2012) Business dictionary. Retrieved 17 April 2012
22. American Institute of Certified Public Accountants/Canadian Institute of Chartered Accountants (AICPA/CICA) (2009) Generally Accepted Privacy Principles
23. Swire PP, Bermann S (eds) (2007) Information Privacy: Official Reference for the Certified Information Privacy Professional (CIPP). International Association of Privacy Professionals
24. Sweeney L (2002) k-anonymity: a model for protecting privacy. Int J Uncertain Fuzziness Knowl Based Syst 10(05):557–570
25. McCarthy MT (2002) USA Patriot Act
26. Ruiter J, Warnier M (2010) Privacy regulations for cloud computing. TU Delft, Delft
27. Baase S (2008) A gift of fire: social, legal, and ethical issues for computing and the Internet. Prentice Hall, Upper Saddle River
28. Regan PM (2004) Old issues, new context: privacy, information collection, and homeland security. Gov Inf Q 21(4):481–497
29. Birnhack MD (2008) The EU data protection directive: an engine of a global regime. Comput Law Secur Rev 24(6):508–520
30. Hornung G (2012) A general data protection regulation for Europe. Light Shade Comm Draft 25:64–81
31. Bull G (2001) Data protection safe harbor: transferring personal data to the USA. Comput Law Secur Rev 17(4):239–243
32. Weiss MA, Archick K (2016) US-EU data privacy: from safe harbor to privacy shield. Congr Res Serv
33. De Hert P, Papakonstantinou VN, Kamara I (2014) The new cloud computing ISO/IEC 27018 standard through the lens of the EU legislation on data protection
34. datalossdb (2015) Datalossltatistics. Retrievedfrom http://datalossdb.org
35. Mega: secure cloud storage. https://mega.nz/
36. Tresorit: End-to-End Encrypted Cloud Storage for Businesses. https://tresorit.com/
37. Pearson S, Yee G (eds) (2012) Privacy and security for cloud computing. Springer, Berlin
38. Jansen W, Grance T (2011) Guidelines on security and privacy in public cloud computing. NIST Spec Publ 800:144
39. Pearson S (2011) Toward accountability in the cloud. IEEE Internet Comput 15(4):64

40. Sato M (2010) Personal data in the cloud: a global survey of consumer attitudes
41. Habib SM, Hauke S, Ries S, Mhlhuser M (2012) Trust as a facilitator in cloud computing: a survey. J Cloud Comput Adv Syst Appl 1(1):1
42. Cavoukian A (2010) The 7 foundational principles: implementation and mapping of fair information practices
43. Bessani A, Correia M, Quaresma B, Andr F, Sousa P (2013) DepSky: dependable and secure storage in a cloud-of-clouds. ACM Transactions on Storage 9(4):12
44. Song Y, Kim H, Mohaisen A (2014, September) A private walk in the clouds: Using end-to-end encryption between cloud applications in a personal domain. In: International Conference on Trust, Privacy and Security in Digital Business. Springer International Publishing, pp 72–82
45. Han F, Qin J, Hu J (2016) Secure searches in the cloud: a survey. Future Gener Comput Syst 62:66–75
46. Yao A (1986, October) How to generate and exchange secrets. In: IEEE 27th Annual Symposium on Foundations of Computer Science, pp 162–167
47. Gentry C (2009) May) Fully homomorphic encryption using ideal lattices. STOC 9:169–178
48. Atayero AA, Feyisetan O (2011) Security issues in cloud computing: The potentials of homomorphic encryption. J Emerg Trends Comput Inf Sci 2(10):546–552
49. Vishwakarma B, Gupta H, Manoria M (2016, March) A survey on privacy preserving mining implementing techniques. In: IEEE Symposium on Colossal Data Analysis and Networking (CDAN), pp. 1–5
50. Goroff DL (2015) Balancing privacy versus accuracy in research protocols. Science 347(6221):479–480
51. Narayanan A, Shmatikov V (2008, May) Robust de-anonymization of large sparse datasets. In: 2008 IEEE Symposium on Security and Privacy (sp 2008), pp 111–125
52. Mont MC, Pearson S, Bramhall P (2003, September) Towards accountable management of identity and privacy: Sticky policies and enforceable tracing services. In: Proceedings of the 14th International Workshop on Database and Expert Systems Applications, pp 377–382
53. Bezzi M, Trabelsi S (2011) Data usage control in the future internet cloud. Springer, Berlin
54. Chen L, Mitchell CJ, Martin A (eds) (2009) Trusted Computing: Second International Conference, Trust 2009 Oxford, UK, April 6–8, Proceedings, vol 5471. Springer
55. Sadeghi AR, Schneider T, Winandy M (2010) Token-based cloud computing. Trust and trustworthy computing. Springer, Berlin, pp 417–429
56. TCG Public Review. Trusted Platform Module Library. Part 1: Architecture. Family 2.0. March 13, 2014, Committee Draft, Level 00 Revision 01.07
57. di Vimercati SDC, Erbacher RF, Foresti S, Jajodia S, Livraga G, Samarati P (2014) Encryption and fragmentation for data confidentiality in the cloud. In: Foundations of security analysis and design VII. Springer International Publishing, pp 212–243
58. Aloqaily M, Kantarci B, Mouftah HT (2014, December) On the impact of quality of experience (QoE) in a vehicular cloud with various providers. In: 2014 11th Annual High Capacity Optical Networks and Emerging/Enabling Technologies (Photonics for Energy), pp 94–98
59. Aloqaily M, Kantarci B, Mouftah HT (2015, December) An auction-driven multi-objective provisioning framework in a vehicular cloud. In: 2015 IEEE Globecom Workshops (GC Wkshps), pp 1–6
60. Beiter M, Mont MC, Chen L, Pearson S (2014) End-to-end policy based encryption techniques for multi-party data management. Comput Stand Interfaces 36(4):689–703
61. Li Y, Gai K, Qiu L, Qiu M, Zhao H (2016) Intelligent cryptography approach for secure distributed big data storage in cloud computing. Inf Sci
62. Wang C, Cao N, Ren K, Lou W (2012) Enabling secure and efficient ranked keyword search over outsourced cloud data. IEEE Trans Parallel Distrib Syst 23(8):1467–1479
63. Song W, Wang B, Wang Q, Peng Z, Lou W, Cui Y (2016) A privacy-preserved full-text retrieval algorithm over encrypted data for cloud storage applications. J Parallel Distr Comput
64. Erway CC, Kp A, Papamanthou C, Tamassia R (2015) Dynamic provable data possession. ACM Trans Inf Syst Secur 17(4):15
65. Betge-Brezetz S, Kamga GB, Dupont MP, Guesmi A (2013, November) End-to-end privacy policy enforcement in cloud infrastructure. In: 2013 IEEE 2nd International Conference on Cloud Networking (CloudNet), pp 25–32
66. Chang V, Kuo YH, Ramachandran M (2016) Cloud computing adoption framework: a security framework for business clouds. Future Gener Comput Syst 57:24–41

67. Wang C, Chow SS, Wang Q, Ren K, Lou W (2013) Privacy-preserving public auditing for secure cloud storage. IEEE Trans Comput 62(2):362–375
68. Mowbray M, Pearson S, Shen Y (2012) Enhancing privacy in cloud computing via policy-based obfuscation. J Supercomput 61(2):267–291
69. Squicciarini AC, Petracca G, Bertino E (2013, February) Adaptive data protection in distributed systems. In: Proceedings of the third ACM conference on Data and application security and privacy. ACM, pp 365–376
70. Chen S, Thilakanathan D, Xu D, Nepal S, Calvo R (2015, May) Self protecting data sharing using generic policies. In: 2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), pp 1197–1200
71. Thilakanathan D, Chen S, Nepal S, Calvo R (2016) SafeProtect: controlled data sharing with user-defined policies in cloud-based collaborative environment
72. Papanikolaou N, Pearson S, Mont MC, Ko RK (2014) A toolkit for automating compliance in cloud computing services. Int J Cloud Comput 23(1):45–68
73. EnCoRe 2011. The EnCoRe project. http://www.encore-project.info/
74. Rahmouni HB (2011) Ontology based privacy compliance for health data disclosure in Europe. Doctoral dissertation, University of the West of England, Bristol
75. Bahrami M, Singhal M (2016, February) CloudPDB: A light-weight data privacy schema for cloud-based databases. In: 2016 International Conference on Computing, Networking and Communications (ICNC), pp 1–5
76. Yau SS, An HG (2010, November) Protection of users' data confidentiality in cloud computing. In: Proceedings of the second Asia-Pacific symposium on internetware. ACM, p 11
77. Itani W, Kayssi A, Chehab A (2009, December) Privacy as a service: Privacy-aware data storage and processing in cloud computing architectures. In: Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, DASC'09, pp 711–716
78. Ghorbel M, Aghasaryan A, Betg-Brezetz S, Dupont MP, Kamga GB, Piekarec S (2011, July) Privacy data envelope: concept and implementation. In: IEEE 2011 Ninth Annual International Conference on Privacy, Security and Trust (PST), pp 55–62
79. Trabelsi S, Sendor J (2012, July) Sticky policies for data control in the cloud. In: IEEE 2012 Tenth Annual International Conference on Privacy, Security and Trust (PST), pp 75–80
80. Brown J, Blough DM (2015, August) Distributed enforcement of sticky policies with flexible trust. In: 2015 IEEE 17th International Conference on High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conference on Embedded Software and Systems (ICESS), pp 1202–1209
81. Brandic I, Dustdar S, Anstett T, Schumm D, Leymann F, Konrad R (2010, July) Compliant cloud computing (c3): Architecture and language support for user-driven compliance management in clouds. In: 2010 IEEE 3rd International Conference on Cloud Computing (CLOUD), pp 244–251
82. Wchner T, Pretschner A (2012, November) Data loss prevention based on data-driven usage control. In: 2012 IEEE 23rd International Symposium on Software Reliability Engineering, pp 151–160
83. Kelbert F, Pretschner A (2013, February) Data usage control enforcement in distributed systems. In: Proceedings of the Third ACM Conference on Data and Application Security and Privacy. ACM, pp. 71–82
84. Kelbert F, Pretschner A (2015, June) A fully decentralized data usage control enforcement infrastructure. In: International Conference on Applied Cryptography and Network Security. Springer International Publishing, pp. 409–430
85. Chen YY, Jamkhedkar PA, Lee RB (2012, October) A software-hardware architecture for self-protecting data. In: Proceedings of the 2012 ACM Conference on Computer and Communications Security. ACM, pp 14–27