Crime and Punishment in the Cloud
Accountability, Transparency, and Privacy

Stefan Berthold, Simone Fischer-Hübner,
Leonardo A. Martucci, and Tobias Pulls*

Karlstad University
651 88 Karlstad, Sweden
[firstname.lastname]@kau.se

Abstract. The goal of this work is to reason on the complexity of the relationship between three non-functional requirements in cloud computing; privacy, accountability, and transparency. We provide insights on the complexity of this relationship from the perspectives of end-users, cloud service providers, and third parties, such as auditors. We shed light on the real and perceived conflicts between privacy, transparency, and accountability, using a formal definition of transparency and an analysis on how well a privacy-preserving transparency-enhancing tool may assist in achieving accountability. Furthermore, we highlight the importance of the privacy impact assessment process for the realisation of both transparency and accountability.

1 Introduction

The complexity of the relationship between the non-functional requirements privacy, accountability, and transparency in cloud computing is high. They are subjective or social constructs, in the case of privacy, and are regulated mostly by legislation and regulation. Social constructs, legislation, and regulation are aspects that are linked to the cultural background of a country or region. Hence, cloud computing services that are delivered online to a global audience need to consider the local flavours and understanding of the privacy, accountability, and transparency.

In this paper, we address the relation between privacy, accountability, and transparency. We provide insights on the complexity of the relationship between the requirements from the perspectives of end-users, cloud service providers (CSPs), and third parties, such as auditors. All requirements are part of a system of checks and balances based on legislation, regulation, economical factors, and competition between CSPs. We do not consider every possible legislation and regulation, but abstract these local parameters as a set of policies that are defined by the CSPs and are communicated to the end-users and auditors. The complexity of the relationship between the requirements is not reduced, but the

* The authors have received funding from the Seventh Framework Programme for Research of the European Community under grant agreement no. 317550.

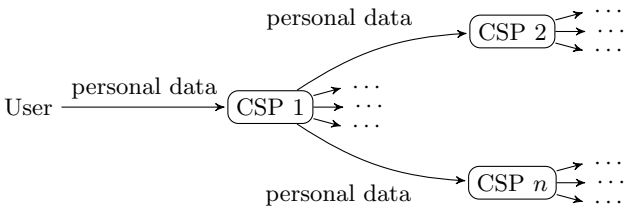

Fig. 1. The flow of personal data into the cloud. The first cloud service provider (CSP 1) offers cloud services to end-users. The personal data of the end-user is forwarded by CSP 1 to other CSPs as parts of the services are outsourced to these CSPs.

representation of local customs and practices is simplified as a set of defined rules. Our conclusions shed light on the real and perceived conflicts between privacy, transparency, and accountability, using a formal definition of transparency and an analysis on how well a privacy-preserving transparency-enhancing tool may assist in achieving accountability. Furthermore, we highlight the importance of privacy impact assessment (PIA) [2] for the realisation of both transparency and accountability.

This paper is organised as follows. Section 2 presents the background. Section 3 provides a formal definition for transparency and opacity. Section 4 presents the relationship between privacy, transparency, and accountability. Section 5 discusses privacy, transparency, and accountability in cloud computing. Section 6 outlines the relationship between all three non-functional requirements from the perspective of a distributed privacy-preserving log trails system for cloud computing. Finally, Section 7 presents the conclusions.

2 Background

A cloud service may be based on other cloud services, platforms or infrastructures. Hence, a user's personal data that is sent to a cloud service provider may be forwarded to other cloud service providers. For instance, a cloud-based application can run on top of a cloud platform that is hosted on top of a cloud-based infrastructure, and the application is a mash-up of other cloud services running on different platforms and infrastructures. We illustrate the complexity of such relationships from the perspective of personal data in Fig. 1.

As cloud computing services become commoditised, a cloud service can be easily replaced or offered by multiple providers simultaneously, e. g., a cloud computing service may store user data in different cloud-based infrastructures. The commoditisation of functional requirements does not result in the commoditisation of the non-functional requirements. CSPs that collect data from users are thus required to negotiate and compose policies in the service chains and are thereby contributing to the ex ante transparency [3] of the user. The commoditisation allows CSPs to offer differentiated services regarding the desired

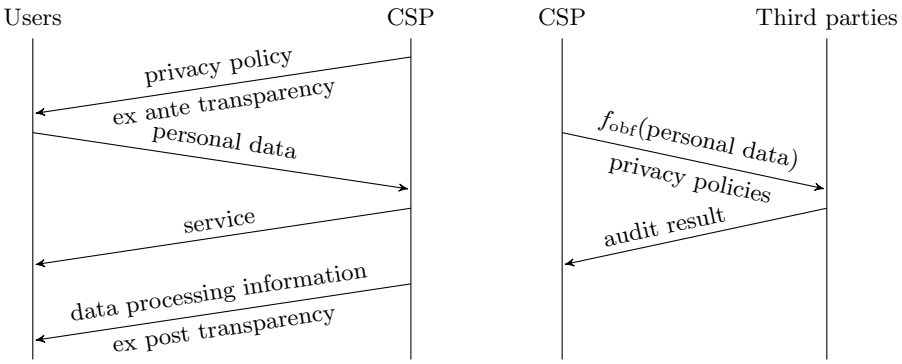

Fig. 2. The relationship between users, cloud service providers (CSP), and third parties, such as auditors or law enforcement. On the diagram to the left, the CSP sends to the users the privacy policy (or agrees upon one) related to the requested cloud service. This step provides users with ex ante transparency. We assume that service provisioning incurs in the transfer of personal identifiable information to the CSP. Ex post transparency [3] is provided by the CSP to the users by offering means for them to verify all the processing information on their personal data. On the diagram to the right, third parties, e. g., auditors or data protection agencies, verify if privacy policies are been executed accordingly by accessing logs that may contain personal data, which is required to be obfuscated to protect the users' privacy.

level of privacy and transparency. However, the flexibility of CSPs to change their subordinate CSPs may diminish, since the privacy policies have to be renegotiated with all actors involved.

The data flows between end-users, CSPs, and third parties are presented as a sequence diagram in Fig. 2. Only data flows that are relevant for accountability, transparency, and privacy are illustrated. Nevertheless, there are additional accountability aspects that are omitted in Fig. 2 for the sake of simplicity. Third parties should be held accountable for their actions regarding the collected data towards the CSP and users, and the CSP is accountable to the third parties, which may represent civil society organisations or governmental agencies. Furthermore, Fig. 2 simplifies some tensions regarding data access by aggregating all users under a single designation. Naturally, users should have access only to the personal data that they own.

3 Transparency and privacy

This section defines the terms transparency and privacy, and discusses their relation. The intuitive meaning of transparency is that nothing is hidden from anyone. In the context of information processing, the scope of the term transparency can be limited to no information is hidden from anyone or, equivalently, all

information is available to everyone. We put this idea in the words of Shannon's information theory [8] and define the term transparency in Definition 1.

Definition 1 (Transparency). *Transparency is the state when every party in the target group possesses perfect knowledge about the observable of interest. In other words, no party in the target group could learn any information (in Shannon's [8] sense) about the observable of interest.*

The observable is an object, a subject, or a process (with inputs and outputs) that can be measured. Its transparency state is determined by the knowledge of all parties in the target group at the same time. Their knowledge has to be perfect. The target group is a group of individuals or information processors that has to be defined for observables of interest before their transparency state can be determined. A party has perfect knowledge, if no fact can be presented to the party that would add to its knowledge. A fact like that would be information in terms of Shannon's information theory.

The observable is opaque, if there is a fact which is information, i. e., a fact that would add to the knowledge of one party. The existence of the fact is sufficient, it does not have to be available for the party. The more information could be learned by one party the greater is the opacity of the observable. We define opacity as a dual to transparency.

Definition 2 (Opacity). *The opacity of the observable of interest is the maximum amount of information one party in the target group could learn about the observable.*

Zero opacity means that every party in the target group possesses perfect knowledge, thus, the observable is transparent. Non-zero opacity implies that at least one party in the target group could learn more about the observable, i. e., add information to its knowledge, and thus the observable is not transparent. Transparency and opacity may vary over time when new information about the observable is created which may not be available to all parties immediately. Thus, transparency and opacity depend on the time of measurement.

In April 2013, the French government made it mandatory for members of the national cabinet to declare their wealth [10]. This can be understood as transparency where the observable is the wealth and the target group is the French society. In this context, transparency for the public and privacy for the state ministers are conflicting objectives.

In data protection (EU Data Protection Directive 95/46/EC), the obligation of the data controller to inform the data subject about the data processing can be understood as transparency where the data processing is the observable and the data subject is the only member of the target group. This does not conflict with the privacy of the data subject as long as the data processing is opaque for everyone except the data subject and the data controller. The term privacy is informally defined in Definition 3.

Definition 3 (Privacy). *Privacy is the right of individuals to control the flow and use of their personal data.*

Privacy as in Definition 3 is also known as the right to informational self-determination. The terms 'control', 'flow', and 'use' mean that individuals are in the position to make informed decisions about data disclosure, storage, and processing, and can impose their decisions on the data controller. This includes the right to minimise the data disclosure (data minimisation), binding the processing of personal data to specific purposes, and deleting the data after specific time periods. Privacy also implies the right of individuals to be informed about the storage and the processing of their personal data. This right is required for making informed decisions.

4 Accountability

This section defines accountability and discusses the relation of it to transparency and privacy. In simple words, accountability is complementing the privacy of individuals with transparency and liability provisions for data controllers. ISO/IEC 29100 puts this in more specific wording.

> Accountability: document policies, procedures and practices, assign the duty to implement privacy policies to specified individuals in the organization, provide suitable training, inform about privacy breaches, give access to effective sanctions and procedures for compensations in case of privacy breaches. [4]

ISO/IEC 29100 provides data controllers with guidelines on how to achieve accountability. We aim to define what accountability is.

Definition 4 (Accountability). *A data controller is accountable, if privacy breaches are transparent to the respective data subjects and the data controller is sanctioned and/or the data subject is compensated in case of privacy breaches.*

Accountability imposes transparency and liability on data controllers. Liability assigs resposibility that may lead to sanctions or compensations. Data controllers that breach privacy store or process personal data of data subjects beyond their the control.

Definition 4 requires the detection of privacy breaches, but does not determine how they are detected. Conceivable are independent third parties, auditors, who check the data processing logs. The auditor would announce the result of the check, i. e., the presence or absence of privacy breaches. However, the public announcement of privacy breaches that are assignable to specific data subjects pose new privacy breaches, since personal data of the first privacy breach is then published in the announcement beyond the control of the data subjects. As a consequence, auditors need to be accountable as well.

An alternative to sending the the full data processing log to the auditor is to send anonymised logs which cannot be used to identify data subjects. An even more fundamental alternative is to send the data processing mechanism as a black box instead of data processing logs. The mechanism can be checked by auditors with random input data. In both cases, the auditors would avoid to commit new privacy breaches.

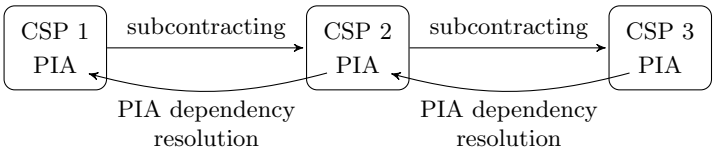

Fig. 3. CSPs are subcontracting the services of other CSPs. Even contracting loops are conceivable. The privacy impact assessment (PIA) of a contractor depends on the outcome of the PIA of the subcontractor.

5 In the cloud

This section discusses transparency, privacy, and accountability in the context of cloud services. In the cloud, cloud service providers (CSPs) may become data controllers and cloud users may become the data subjects. However, cloud services introduce two dimensions that are uncommon in the classic data controller model. On the vertical dimension, each CSP may offer more than one service. On the horizontal dimension, CSPs may be subcontracting services of other CSPs.

The implications of the vertical dimension have been discussed in prior work [5]. Privacy needs to be preserved and transparency to be established for each service independently. The implementation and, if necessary, the trade-off between privacy and transparency must be the result of careful planning, e. g., by applying privacy by design rules [1] or carrying out a PIA. Part of the implementation can be the involvement of third parties who need to be accountable as well, at least if they receive personal data.

On the horizontal dimension, even called the chain of accountability [6], the implementation of privacy and transparency measures are not independent. Again, careful planning has to precede the implementation, in contrast to the vertical dimension, however, the contractor's planning depends on the subcontractors' planning and has to be adjusted whenever the subcontractor is changing its implementation. Fig. 3 illustrates the case for PIAs among subcontracting CSPs.

6 Distributed privacy-preserving log trails

In this section, we look at an example system that was designed with the privacy of end-users in mind. The system is applicable in a cloud setting for making data processing of personal data transparent to end-users.

The goal of the *distributed privacy-preserving log trails* system presented in [7] is to make data processing of users' personal data transparent towards the users whose data is being processed. The system facilitates the transfer of data processing information from CSPs to users while protecting the privacy of the users. Only the information logged for a user is transparent to that user, to all other parties the information is opaque. This is accomplished by cryptographic means, in terms of encryption but also by ensuring that users are only identified

by transaction pseudonyms where both log entries and identifiers are unlinkable. When a user's personal data is spread within the CSP's service chain, as illustrated in Fig. 1, a new transaction pseudonym for the user is generated for each system that performs data processing. The service chain can be both distributed and dynamic, i. e., there is no need to know before data disclosure which or how many data processing systems the CSPs use. Users can anonymously reconstruct and verify the integrity of all descriptions of data processing logged for them across all of the systems that performed data processing as a consequence of the user's disclosure of personal data. These privacy protections minimise the amount of personal data generated by the transparency-enhancing tool, which in turn ensures that using the tool preserves the users' privacy.

Returning to Fig. 2, the system in [7] is suited for providing ex post transparency of data processing towards users. Ultimately, this plays a role in making the CSP accountable towards their users. In [7], the CSP shares the complete set of data with the users, however, the system does not facilitate support for sharing to third parties anything but the complete set of data provided by the CSP to users. The lack of obfuscation makes the system far from ideal, since ideally the personal data shared with third parties should be fully obfuscated, i. e., opaque. There has been some work on obfuscating audit logs, such as [9] in the context of intrusion detection. However, there is a need for future work on minimising the amount of personal data disclosed while still retaining the ability for an auditor to keep a CSP accountable. Furthermore, we note that while the logging system in [7] facilitates ex post transparency of data processing by CSPs towards users, there is still a need for further work on supporting redress once a user learns of data processing which violates the previously agreed to privacy policy. Without redress, e. g., in the form of financial compensation to the user or sanctions towards the misbehaving CSP, the CSP may not be considered accountable.

7 Conclusions

Cloud computing is becoming increasingly complex as services are turning into commodities, easily swapped out and replaced, into dynamic service chains. Making CSPs accountable while respecting the privacy of users is a truly daunting task. This is because in general, accountability, transparency, and privacy are perceived as conflicting goals. In this paper, we have shown that when it comes to *end-user* privacy there is conceptually no conflict with providing transparency and accountability while at the same time respecting the privacy of end users. The CSP should be accountable for privacy breaches.

Towards end users, the CSP can make all processing on a user's personal data transparent. The CSP needs to make sure that the processing is only transparent to the user to whom the personal data belongs while the processing remains opaque to everyone else. This does not constitute a privacy breach, because each user only learns of the processing on their own data. We discussed an example system for realising this kind of transparency in Section 6.

Towards third parties, the CSP only needs to make the privacy breaches transparent, not the personal data of users. This requires that the third parties are able to inspect, at the very least, each processing mechanism the CSP uses to be able to detect breaches. Again, this does not breach the privacy of end-users.

The privacy impact assessment (PIA) as a transparency tool plays a central role, especially in complex and dynamic settings such as cloud computing. PIA is used to uncover privacy issues and possible privacy breaches for the CSP and to convey these issues to other CSPs and end-users. This makes accountability as defined in Definition 4 possible.

If we expand the concept of privacy to include the "privacy" of CSPs, there is a real conflict between privacy, transparency, and accountability. Detecting privacy breaches, in the sense of illegitimate data processing, requires CSPs to make their data processing transparent. Such transparency may reveal intellectual property or trade secrets of the CSPs. This is recognised, e. g., in recital 41 of the EU Data Protection Directive 95/46/EC.

Unlike the moral dilemma presented in Dostoyevsky's *Crime and Punishment*, our work offers a happier ending. Accountability, transparency, and privacy are conceptually realisable without negatively effecting the privacy of end-users. This however comes at the cost of the "privacy" of CSPs, whose data processing needs to become more transparent towards both end-users and auditing third-parties. The question is: Is this morally justifiable?

References

1. Cavoukian, A.: Privacy by design. White paper, Information and Privacy Commissioner of Ontario (2009)
2. Clarke, R.: Privacy impact assessment: Its origins and development. Computer Law & Security Review 25(2), 123–135 (2009)
3. Hildebrandt, M.: Behavioural biometric profiling and transparency enhancing tools. Project Deliverable 7.12, Future of Identity in the Information Science (FIDIS, Network of Excellence within the European Community's 6th Framework Program, No. 507512) (2009)
4. ISO/IEC: Privacy framework. ISO/IEC 29100, ISO/IEC (2011)
5. Pearson, S.: Taking account of privacy when designing cloud computing services. In: Software Engineering Challenges of Cloud Computing, 2009. CLOUD '09. ICSE Workshop on. pp. 44–52 (2009)
6. Pearson, S., Tountopoulos, V., Catteddu, D., Südholt, M., Molva, R., Reich, C., Fischer-Hübner, S., Millard, C., Lotz, V., Jaatun, M., Leenes, R., Rong, C., Lopez, J.: Accountability for cloud and other future internet services. In: 4^{th} IEEE Int. Conf. on Cloud Computing Technology and Science (CloudCom). pp. 629–632 (2012)
7. Pulls, T.: Privacy-preserving transparency-enhancing tools. Licentiate thesis 2012:57, Karlstad University, Department of Computer Science (2012)
8. Shannon, C.E.: A mathematical theory of communications. Bell System Technical Journal 27, 379–423, 623–656 (1948)
9. Sobirey, M., Fischer-Hübner, S., Rannenberg, K.: Pseudonymous audit for privacy enhanced intrusion detection. In: SEC. pp. 151–163 (1997)
10. The Economist: Transparency days. Printed Edition **407** (8832) (20 Apr 2013)