technologies

MDPI

Article

Identity Management and Protection Motivated by the General Data Protection Regulation of the European Union—A Conceptual Framework Based on State-of-the-Art Software Technologies †

Pascal Birnstill *, Erik Krempel, Paul Georg Wagner and Jürgen Beyerer

Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB, Fraunhoferstr.1, 76131 Karlsruhe, Germany; erik.krempel@iosb.fraunhofer.de (E.K.); paul-georg.wagner@iosb.fraunhofer.de (P.G.W.); juergen.beyerer@iosb.fraunhofer.de (J.B.)

* Correspondence: pascal.birnstill@iosb.fraunhofer.de

† This paper is an extended version of two other papers by the authors that have been published in Proceedings of the 11th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA 2018), Island of Rhodes, Greece, 26–29 June 2017.

Received: 31 October 2018; Accepted: 30 November 2018; Published: 4 December 2018

Abstract: In times of strongly (personal) data-driven economy, the inception of the European General Data Protection Regulation (GDPR) recently reinforced the call for transparency and informational self-determination—not only due to the penalties for data protection violations becoming significantly more severe. This paper recaps the GDPR articles that should be noticed by software designers and developers and explains how, from the perspective of computer scientists, the summarized requirements can be implemented based on state-of-the-art technologies, such as data provenance tracking, distributed usage control, and remote attestation protocols. For this, the challenges for data controllers, i.e., the service providers, as well as for the data subjects, i.e., the users whose personal data are being processed by the services, are worked out. As a result, this paper proposes the ideal functionality of a next-generation privacy dashboard interacting with data provenance and usage control infrastructure implemented at the service providers to operationalize the legal rights of the data subject granted by the GDPR. Finally, it briefly outlines the options for establishing trust in data provenance tracking and usage control infrastructures operated by the service providers themselves.

Keywords: data protection; privacy; GDPR; identity management; data provenance tracking; usage control; remote attestation; trusted computing

1. Introduction

In an increasingly interconnected world, (personal) data are the asset on which innovative and successful business models are built. For example, the management of globally distributed supply chains is not imaginable without exchanging data across the boundaries of otherwise independent organizations. Obviously, once we share our key asset, data transfers must happen in a manner such that we remain in control over our data to a certain extent. This introduces challenging requirements for access control and data protection technologies.

On the other hand, more and more assistance technologies are entering our homes and working environments. Smart homes, driver's assistance as well as assistance systems at our workplaces aim to ease our everyday life, but at the same time strongly rely on personal data to offer services that are not only adequate to the current situation, but also to the needs of the individuals involved. While this may not seem to be exceptionally critical as long as personal data do not leave one's smart home,

circumstances change dramatically once smart home data are processed by cloud services, and, all the more, as soon as an assistance system is operated by our employer who may have interest in exploiting the data beyond their original purpose, e.g., for secretly evaluating the work performance of his personnel. Once we are talking about the personal data of users or customers from the European Union being processed, according requirements are laid down in the General Data Protection Regulation of the European Union (GPDR), whose inception recently reinforced the call for transparency and informational self-determination—not only due to the penalties for data protection violations becoming significantly more severe.

In this paper, we particularly describe and suggest technologies that support the rights of the data subject, i.e., the right of access (Article 15 GDPR), the right to rectification (Article 16 GDPR), the right to erasure ("right to be forgotten", Article 17 GDPR), and the right to restriction of processing (Article 18 GDPR). Please note that this paper is an extended version of two other papers published in Proceedings of the 11th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA 2018), Island of Rhodes, Greece, 26–29 June 2017: Erik Krempel analyzed the technical requirements and implications for data processing services according to the GDPR [1]. Pascal Birnstill addressed these requirements by introducing a conceptual framework, which integrates the interaction with state-of-the-art information security and data protection mechanisms into a next-generation privacy-dashboard [2].

As shown in [3], when trying to exercise the right of access today, one is mostly confronted with a contact form on the data controller's website. Usually, the reply is a download link delivering a hardly human-readable extensible markup language (XML)-style listing of the data records stored about us. Even if one digs through this pile of information, further contacting is necessary to get data rectified, deleted, or locked. However, next to this poor usability of exercising the rights of the data subject, such a reply leaves more questions than it answers, namely whether one can at all trust in the completeness of the answer, or even in the execution of a request to rectification, erasure or restriction of processing. Being confronted with customers' requests for information, the situation of most data controllers is by no means more convenient. Due to the low degree of automatization, processing requests for information generates significant additional workload for organizations. The same is of course true for requests for rectification, deletion, and restriction of processing.

Being able to automatize the processing of requests concerning the data subject's right to access requires that the provenance of personal data representations is being tracked. Only the knowledge about where personal data reside in an organization's IT infrastructure allows implementing mechanisms fulfilling the rights of rectification, erasure, and restriction of processing.

This paper is structured as follows. In Section 2, we explain the legal framework of the GDPR, in particular the aspects of lawful data processing, documentation and compliance, as well as transparency and control. Subsequently, we introduce several concepts and technologies, which we afterwards integrate into a concept for operationalizing and automatizing the legal rights of the data subject with particular focus on the aspects of transparency and control (cf. Section 3). Readers who are familiar with data provenance tracking, distributed usage control, and remote attestation protocols may skip these paragraphs. Based on this legal framework and these technical concepts, our first step in Section 4 is to discuss the requirements and challenges for data controllers and how they can be fulfilled based on data provenance tracking and distributed usage control technology [4]. When considering an average Internet user as a prototype data subject, then the data subject uses dozens of services processing his personal data. Therefore, in our second step, we also consider how user-friendly access to the rights of the data subject can be operationalized in a next-generation privacy-dashboard, the ideal functionality of which we outline in Section 5. In this step, we also explain the options regarding how trust in the implementations of data provenance tracking and usage control operated by the data controller can be established based on hardware trust anchors and remote attestation protocols. In Section 6, we describe how the considered security mechanisms interact with a new service requesting a user profile. We explain the agenda and the status of our evaluation in Section 7 and conclude in Section 8.

2. Legal Motivation

As of 25 May 2018, the European Union enforces the EU General Data Protection Regulation (GDPR) all over Europe. In the near future, this will affect data processing practices in multiple ways. The first aspect one should note about the GDPR is that it is a regulation and not a directive. Therefore, it applies directly in all member states without any translation into national law. While many of the underlying concepts are well known in the privacy community, e.g., the Fair Information Practices Principles (FIP) by the Organisation for Economic Co-operation and Development (OECD) are very similar [5], the regulation comes with two aspects that are fundamentally new to data protection.

First, data protection is obligatory and fines are huge. Infringements are fined up to 20 million EUR or up to 4% of the total worldwide annual turnover of the preceding financial year, whichever is higher. The second part is called **territorial scope**. This means that the regulation does not only apply to EU companies but to every company selling into the EU or marketing to EU citizens.

2.1. Legal Terms

For a better understanding of the legal situation, we shortly explain the most important legal terms:

- **Personal data:** Any information relating to an identified or identifiable natural person. This includes obvious cases, such as names, passport ID, or addresses, but also more opaque examples, such as a play list of favorite music titles, as long as one can establish a link to a natural person.
- **Data subject:** All personal data belong to an identified or identifiable natural person. This person is called the data subject.
- **Processing:** Processing of personal data including collection, storage, use, transmission, and erasure.
- **Controller:** The entity which determines the purposes and means of data processing.

As typical with new legislation, the current situation is somewhat under-defined. This is because we do not yet know how the articles of the GDPR will be interpreted and used in court. Especially in upcoming technologies, such as distributed data management and usage, designers will have to come up with long-term designs that will allow them to fulfill legal requirements which may alter.

When looking at the GDPR, one can identify four pillars that are essential for the compliance of personal data processing software systems:

1. Lawfulness of processing
2. Documentation and compliance
3. Data Protection by Design
4. Transparency and control

In the following sections, we give an overview on these four pillars and point out the aspects that are important for our further considerations.

(Please note that the authors are not legal scholars or lawyers, but engineers with a strong background in privacy and data protection. Whenever processing personal information, consult a legal expert for the legal requirements applicable to your specific case.)

2.2. Lawful Data Processing

One of the key aspects of data processing in general and the GDPR in particular is that there is a ban with permit reservation, which means that processing of personal data is generally prohibited if no legal permission applies. Article 6 GDPR states six conditions under which the processing of personal data is permitted:

- **Consent:** The data subject has given consent to the data processing.

- **Contract:** The data processing is necessary for the performance of a contract of which the data subject is party.
- **Legal obligation:** The data processing is necessary for compliance with legal obligations.
- **Vital interests:** The data processing is necessary to protect vital interests of the data subject or of another natural person.
- **Public interest:** The data processing is necessary for the public interest or in the exercise of an official authority vested in the controller.
- **Legitimate interest of the controller:** The data processing is necessary for a legitimate interest of the controller and not challenged by the interests or fundamental rights and freedoms of the data subject.

In this article, we focus on the first case, where a data subject declares explicit consent to the data processing. This is the most common case, e.g., when using Internet services as well as assistance systems in smart homes or at work. The consent has to be specific, free and informed.

Specific means that the user has to give consent to data processing for one or more specific purposes. A purpose has to be described as precise as possible at the moment the consent is requested. Processing data beyond these specific purposes is unlawful. Free means that the user is not forced into accepting data processing.Service providers are not allowed to demand consent for processing of information which is not required to fulfill the specified purposes. Thus, if a service only needs the name and the address, but requires that the user gives consent to additionally process his birthday, this circumstance can render data processing unlawful. Finally, informed is the most challenging of the three requirements of consent. The request for consent must be presented in a manner clearly distinguishable from other interactions with the given system, i.e., it must be separated from all other interactions with the system. It must also be presented in an intelligible and easily accessible form, using clear and plain language (Article 6). Currently, it is unclear how to provide this level of detail without confronting the user with extensive privacy policies that do not comply with the requirement of clear and plain language.

2.3. Documentation and Compliance

A significant part of the GDPR is concerned with documentation and compliance. The most important documents that must exist for lawful data processing are the following ones:

- **Privacy Notice:** Document for customers, business partners and website visitors that describes how their personal data are going to be processed (Article 12).
- **Records of processing activities:** A list of all processing activities including the purpose of the data processing, a description of the data, the recipients of the data and a general description of technical and organizational security measures to protect the data (Article 30).
- **Processor Agreements (when applicable):** When data are processed (this includes storage) by a third party on behalf of the controller, a controller processor agreement is needed.

While the scope and level of detail of these document vary, they show some common elements. To compile the document, an organization providing a data processing service needs to understand what data it processes, where the data are collected for what well-defined purposes, and when data are shared with third parties. Additionally, one needs to know where data enter the system, e.g., whether they are entered directly by the users or are received from a third party.

2.4. Transparency and Control

The final pillar of the GDPR that we discuss is transparency and control. The user, in the wording of GDPR the data subject, should always know what data are processed, understand the purpose of processing and have an adequate level of control over his data. This understanding is expressed in Articles 6 and 7 where consent and the conditions of consent are described (cf. Section 2.2) as well as in Articles 12–23.

As long as personal data are processed by a controller, the data subject has a right to access this data (Article 15). This is independent from the legal purpose for data processing (cf. Section 2.2). A service provider must therefore even inform the data subject about data processing, if some kind of legal obligation prescribes the data processing. The controller must provide an overview of the purpose of data processing to the data subject, including the categories of personal data and potential recipients, especially if data are transmitted to third parties outside the EU. Additionally, the controller has to inform the data subject about his rights to rectification, restriction of access, or erasure and the right to lodge a complaint at a data protection supervisory authority.

Additionally, the data subject has a right to rectification of personal data (Article 16). Therefore, each system should provide means to allow the data subject to do so, either directly or by contacting somebody at the data controller.

A data subject has the right to demand erasure of personal data from a controller for a wide range of reasons (Article 17). Data processing could have been unlawful, the subject withdraws his consent or the personal data are no longer necessary in relation to the purpose for which they were collected. If a data subject has a valid claim, the controller has to delete the data without undue delay. If the data were shared with a third party, the controller needs to contact these third parties and inform them about the request to erasure. This can be achieved by organizational or technical means and should take into account available technology and the cost of implementation.

In similar cases, the data subject can utilize the right to restriction of processing (Article 18). Once the processing of data is restricted, it can only be processed for very restricted matters, e.g., in the context of a legal claim. Unlike with deleted data, a controller still has to fulfill the data subject's right to access of restricted data, while at the same time enforcing restricted access from within the organization.

The rights of rectification, erasure, and restriction of processing come with additional obligations for the controller (Article 19). If data were shared with third parties, the controller has to forward the request to enable the enforcement the data subject's rights at these parties. In addition, the data controller has to inform the data subject about the recipients if the subject wishes so.

These rights and obligations come with certain requirements and balances, which consider the rights of the controller and the subject. A subject cannot demand deletion of a bill and thus escape payment, or change data in a way that prevents the controller from fulfilling his legal obligations. However, on the basis of the GDPR, the users have stronger rights, and systems have to be designed in a way to provide a significantly higher level of transparency and control.

2.5. Conclusion on Legal Motivation

As shown in the previous sections, the GDPR sets a high standard for transparency, control, and consent of the user. While the regulation offers neither aid nor restrictions, it clearly recommends to look into the state of the art of technology to fulfill the demands.

In this paper, we show how provenance tracking and usage control can be instantiated to automatize large parts of the controllers obligations. While the implementation of such a system goes along with a certain overhead, we emphasize on the benefits:

1. **Users' trust:** A hight level of transparency increases the users' trust in the system and might offer a competitive advantage.
2. **Low running costs:** Many of the data subjects' rights can either be implemented by organizational or technical matters. While a technical implementation generates higher initial costs, it reduces the workload of the staff and will therefore reduce costs in the long run.
3. **High level of flexibility:** The presented technical solutions offer a much higher level of flexibility compared to organizational measures, thus increasing the potential competitive advantage.

3. Technology Overview

In the following paragraphs, we introduce several concepts and technologies, which we afterwards integrate into a concept for operationalizing and automatizing the legal rights of the data subject. Readers who are familiar with data provenance tracking, distributed usage control, and remote attestation protocols may feel free to skip this section.

3.1. Data Provenance Tracking

The idea of data provenance tracking originates from the domain of scientific computing, where it is traditionally used to prove the lineage of data [6,7]. Data provenance tracking provides information on the origins of data and the derivations of data from other data [8]. It makes traceable who is accountable for the modifications of data, how and where this modification happened and which other data influenced the process of creating new data items. This kind of traceability is clearly similar to the requirements a data controller is confronted with so as to be able to fulfill its data subjects' right to access. Bier et el. [4] therefore transferred data provenance tracking to data protection and showed how it can serve as an approach towards automatized processing of information requests by data subjects. For this, data provenance tracking components have to be integrated into the data processing infrastructure of a data controller.

When employed for data protection purposes, data provenance tracking technology needs to keep track of personal data usage and transfers within the IT infrastructure of a data controller. In particular, it aims to track personal data flows between the organizations' IT systems in terms of sources and sinks, as well as information concerning data access since the initial collection of personal data. Figure 1 shows an exemplary data provenance tracking architecture. In the example, a webshop transmits customer data from its frontend to a database. Both systems have to be extended with a component, which observes outbound and inbound data flows. Speaking in terms of the well-established conceptional architecture of eXtensible Access Control Markup Language (XACML) [9], this is similar to a policy enforcement point (PEP). This PEP signals data flow-related events, i.e., data collection, access, and deletion, to a second component, which is responsible for keeping track of representations of personal data items on a given physical or virtual host system and to store this knowledge persistently. These two conceptional components (usually there will be several PEPs on a host system) enable provenance tracking of personal data flows for a single host system.For inter-system provenance tracking across the entire IT (information technology) infrastructure of a data controller, a centralized provenance collection component is required to compile provenance reports over multiple host systems. For this, data provenance for a specific personal data item, which can be represented as a tree, is generated by collecting and connecting the subtrees from the particular systems' provenance storage components.

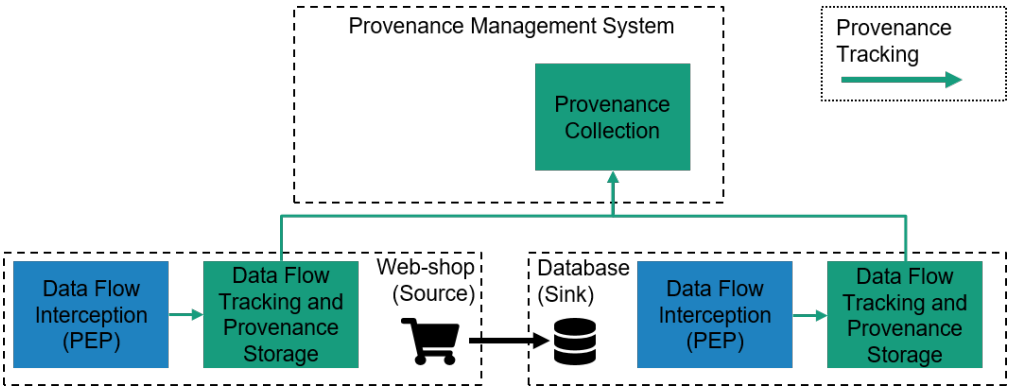

Figure 1. An exemplary data provenance tracking architecture.

3.2. Distributed Usage Control

Distributed usage control (DUC) can be conceived as an extension of access control, which aims to control the usage of data after access has been granted—also across system boundaries (under a "closed world" assumption, i.e., the systems exchanging data are equipped with inter-operable usage control technology). (Many of the architectural challenges of a distributed usage control systems are omitted or simplified here. For a deeper understanding please refer to Pretschner et al. [10]. The authors explain for example how to enforce policies, which only allow a limited number of copies of a piece of data.)

Typical usage control requirements are, for example, "Delete data item after 30 days", "Anonymize data item before sharing", "Do not forward data item", or "Do not forward data item to a system where usage control enforcement cannot be ensured".

If, for instance, a user was granted access to a data item at some point in time, he could have created copies of that data item (e.g., by forwarding it via email), which are beyond the control of access control, i.e., even if we revoke the access permission of the original representation of the data item, we cannot enforce any restrictions on the copies. This is the goal of usage control approaches, which—similar to data provenance tracking—rely on observing and keeping track of flows of protected data items.

Figure 2 extends the webshop example from the previous section with an additional usage control component, the so-called policy decision point (PDP). The PDP evaluates events intercepted by PEPs against usage control requirements specified in policies. As already stated for data provenance tracking, all software systems that are able to collect, alter or store data have to be extended with PEP functionality.

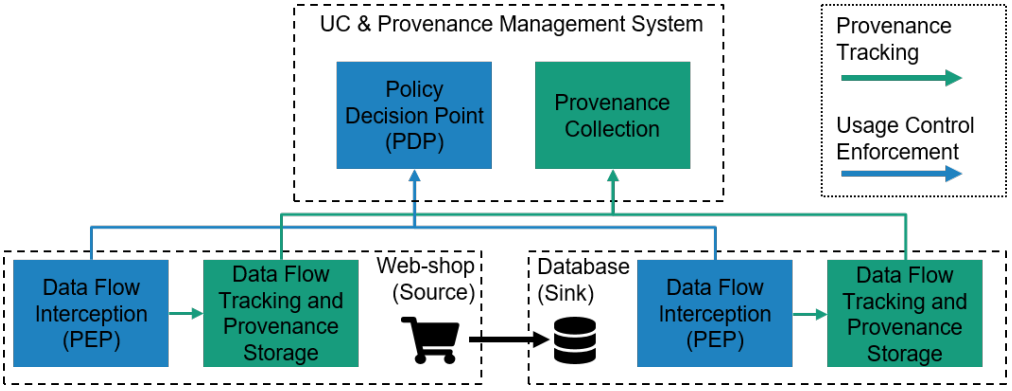

Figure 2. An exemplary data provenance and usage control architecture.

The simplified workflow of data provenance tracking and usage control enforcement proceeds as follows. Whenever a PEP observes an event indicating a data usage or data flow, the event is forwarded to the PDP. The PDP evaluates the event against deployed usage control policies. If, for example, the event indicates that the system tries to transmit protected data to a destination where usage control enforcement cannot be ensured, the PDP will advise the PEP to inhibit the event. In case the PDP decides that the event is to be allowed, the event is also forwarded to the host system's data flow tracking and provenance storage component so that the respective data provenance information is kept up-to-date.

3.3. Remote Attestation

Distributed usage control relies on a client-side reference monitor (CRM), which continuously enforces policies on data under protection. This CRM has to be trustworthy, i.e., it has to be protected in a hostile environment, where the legitimate user of the data is a potential attacker with physical access to the host system. Approaches for protecting a CRM against undiscovered manipulation are usually based on remote attestation. Remote attestation denotes a family of technical measures to

establish trust in the state of the computer system used by a communication partner. The intuition behind remote attestation is to measure the integrity of hard- and software of a system in a fingerprint, store it at a tamper-proof location, and make it available on-demand for a validation by remote systems requesting a proof of integrity. By this means, a so-called Trusted Computing Base (TCB) is established, which comprises the identity of the system's hardware as well as a verified software stack, on which further trustworthy applications can be built. Executing remote attestation of a system of course requires that the requester can compare a received fingerprint with a certified value provided by a trusted third party. The value attested by the trusted third party may for example be based on measurements of a certified system configuration.

Measuring the integrity of a computer system can be achieved based on a Trusted Platform Module (TPM), which acts as a so-called root of trust for measurement. A TPM is a dedicated hardware chip that extends a computer with basic security-related features [11]. The TPM holds several cryptographic keys that can be used to encrypt data, identify the computer system and attest to its current configuration. Furthermore, the TPM contains volatile platform configuration registers (PCRs), which save a summary of the current hardware and software configuration as an unforgeable hash (denoted as measurement). During the boot process, the TPM expects each boot stage to hash the software at the next stage and to extend the PCRs with this measurement. As a result, the PCR values reflect all the software measurements up to that point and hence attest to a certain set of software that is running on the system. Remote attestation protocols can be based on these PCR values, i.e., a third party can remotely verify that the target system is in a certain state by attesting to certain PCR values.

As shown by Wagner et al. [12], certain problems are inherent to TPM-based remote attestation protocols. This is because the attacker model of TPMs does not consider the legitimate user of a host system as a potential attacker. The authors therefore propose a hybrid trusted computing approach based on a TPM as well as Intel Software Guard Extensions (SGX)-based enclaves ((Intel Software Guard Extensions are built into current generations of Intel's x86 CPU architecture) and enclave attestation [13] to implement a trustworthy runtime environment for distributed usage control.

4. Step I: Tracking and Enforcement—Data Provenance Tracking and Distributed Usage Control

As shown by Bier et al. [4], data provenance tracking (cf. Section 3.1) implemented within the infrastructure of a data controller can deliver the kinds of information required to fulfill the data subject's right of access. Furthermore, operationalization of the rights to rectification, to erasure, and to restriction of processing requires the ability to enforce obligations concerning personal data items after they have been collected or after access to them has been granted. From an IT security perspective, enforcing obligations on remote IT systems is a requirement addressed by concepts from the field of usage control languages and technologies (cf. Section 3.2). Usage control technology relies on the kind of information that provenance tracking provides. If provenance tracking helps us to keep track of the locations of representations of a specific personal data item within the IT infrastructure of an organization, usage control allows us to specify and to enforce obligations for data items such as deletion, modification, notification, etc. Based on data provenance tracking, an organization can visualize the information the data subject is entitled to obtain in provenance trees, which show the particular representations and transactions for each personal data item stored within the organization's IT infrastructure accompanied with the possibility to exercise further rights concerning these data items. While these technologies improve the usability of exercising the rights of the data subject and also the degree of automatization of the operationalization of these rights from the perspective of the data controller, several issues remain.

First, why should I, as a data subject: (I.i) trust in the data controller to show me the complete provenance of data items referring to me; and (I.ii) trust in the data controller to fulfill my requests for rectification, erasure, or restriction of processing? Somehow, we have to establish trust in the systems of the data processing chain of the data controller's organization to not be tampered with as well as in the provenance and usage control components within this infrastructure on which our

operationalization of the rights of the data subject is based. Moreover, from the perspective of the data subject, the usability is still improvable, e.g., by providing a central point of contact (I.iii), from where he can access the data protection services of all the organizations he shares data with. This central point of contact could also ease up the administration of the data subject's digital identities, i.e., managing the user profiles he uses for different services as well as reusing particular attributes between different user profiles. Now, we are obviously talking about the next-generation privacy dashboard. In the subsequent section, we discuss what we believe is the ideal functionality of the next-generation privacy dashboard fulfilling (I.iii), i.e., a "privacy dashboard-as-a-service", and how it interacts with the technologies described before.

5. Step II: User-centric Design—Ideal Functionality of a Next-generation Privacy Dashboard

The ideal functionality of a next generation privacy dashboard comprises the following tasks on which we elaborate in the following:

- Creating and managing digital identities (user profiles) for digital services
- Authorizing services to access profiles and their attributes and providing information about the communication endpoints where these resources can be obtained
- Providing access to data provenance interfaces and visualizing data provenance of any used service and shared data item
- Managing usage restrictions concerning the user profiles and their attributes for each used service
- Supporting the management of different levels of trustworthiness of used services
- Supporting established authentication standards—for authenticating the privacy dashboard's user as well as for authenticating services previous to authorization for data access

Creating and managing digital identities (user profiles) for digital services means that the user should be supported to generate user profiles—typically a vector of data attributes and data sources—from service offers in which service providers describe the data requirements of their services. This comprises permissions, i.e., "What data do I share with a specific service?" and consent, i.e., "What data about me do I allow a service to collect/process?" In case a user gives consent to the collection of certain personal attributes, he may also want to decide whether he wants to reuse this attribute for other services and whether he wants to allow the service provider to persistently store the attribute in his data center. Permissions should therefore include granting write access for an attribute on a resource server under the user's control, and also allowing a service provider to persistently store certain attributes in other cases. (Of course, one could also imagine that a service provider collecting a certain personal attribute, e.g., by exploiting some data source he has access to, acts himself as a resource server for which the data subject can manage permissions from within the privacy dashboard. However, this is out scope of this paper.).

A next generation privacy dashboard also needs to be able to authorize services to access profiles and their attributes, which are accessible via arbitrary resource servers. This means that resource servers have to be managed, permissions have to be set on these resource servers, and information about the communication endpoints of these resource servers have to be communicated to the services with whom respective attributes should be shared. As mentioned above, permissions may also include write permissions for specific resources on specific resource servers.

Providing access to data provenance interfaces of service providers means being able to read a standardized data provenance format in order to visualize internal transactions and processing of the service provider as well as data transactions to further data controllers, e.g., a payment provider used by the data controller. Bier et al. [3] showed that, in terms of transparency as well as usability, a suitable implementation of the right of access can be obtained by visualizing the provenance of personal data in provenance graphs. As discussed above, the knowledge about the provenance of personal data items is a necessary requirement for improving the degree of automatization of the rights of the data subject to rectification, erasure, and restriction of processing. Bier et al. also proposed

to implement facilities to exercise these rights from within the provenance graph visualization, e.g., simply by providing according request forms after the user has clicked on a personal data item. Ideally, the service provider also runs a usage control infrastructure, which automatically enforces these requests in case no other regulation contradicts. Providing access to data provenance interfaces of arbitrary service providers can also bring additional benefit for transparency in terms of being able to match an outgoing transaction of a personal data item reported by a data controller A to an incoming transaction reported by data controller B, who is named as the sink of the given transaction by A. By this means, data provenance can be made traceable across the boundaries of data controllers, i.e., organizations and companies.

As already mentioned, distributed usage control technology is a beneficial complement to data provenance tracking. However, in addition to enforcing the rights to rectification, erasure, and restriction of processing, an integration of usage control into next-generation privacy dashboards and services can further improve control over personal data from the perspective of the data subject, and the ability to prove accountability from the perspective of the data controller. In this respect, obvious usage control requirements would be enabling the data subject to already set deletion deadlines prior to releasing data, prohibiting persistent storage of specific personal data items, and also expressing conditions under which the data subject explicitly prohibits or allows the transfer or disclosure of specific personal data items to (specific) third parties for specific purposes. In case the data subjects expresses such requirements, according polices are transferred to the services and deployed on their provider's usage control infrastructure, which is responsible for their enforcement. In the following, we discuss what needs to be done so that we can trust in the enforcement of our usage control policies (cf. Section 4, I.ii).

Whether a data subject would demand the enforcement of usage control requirements as described before strongly depends on the degree of trust the data subjects has into the data controller. For example, in some cases or in certain cultures, people will more likely tend to trust in public administration and authorities to treat their data in a responsible manner and in legal compliance than in private sector companies. In some other cultures, it will be the other way round. Therefore, another desired ingredient of a next-generation privacy dashboard is attributing different trust levels to service providers. In case of high confidence in a service provider, i.e., given a high level of trust, the data subject may not even have usage control requirements. In case of a low level of trust, the data subject would not even trust the service provider to not tamper with the provenance and/or usage control infrastructure it uses and offers. In the latter case, the data subject needs to make sure that the service provider's software including usage control and/or data provenance components is trustworthy. For this, the software components must be certified (or verified using formal methods), and their integrity must be preserved. Concerning the latter question of ensuring the integrity of usage control infrastructures and their runtime environments, we refer to Wagner et al. [12]. They described how remote attestation protocols based on hardware trust anchors can be used to protect usage control infrastructures (cf. Section 3.3). Data provenance tracking components are not mentioned in this paper, however they can be protected in the same manner as usage control components. We believe that at least two trust levels should be represented in a next-generation privacy dashboard indicating whether the data subject demands remote attestation of a service's trustworthy integrity state before deploying usage control policies or granting access to personal data.

Having discussed mostly authorization related aspects so far, finally we need to briefly mention authentication. We believe that next-generation privacy dashboards will be the password managers of next-generation data-driven services. This means that well-established authentication standards have to be integrated in privacy dashboards for authentication of the data subject with a service to be used and also for checking the authenticity of service endpoints before sharing data. For authentication of the data subject with services, we think of managing authentication features like keys, biometric features, tokens, etc., but also redirections to identity provider services like for example OpenID [14]. The latter is also required for verifying the authenticity of services, since it is very likely that services

requesting user profiles will themselves use tokens from trusted third parties, i.e., identity provider services, to prove their identity.

In the next paragraph, we outline how the mechanisms and protocols described above could interact with each other to build up to the functionality of a next-generation privacy dashboard that we claimed above.

6. Interaction of Security Mechanisms

In Figure 3, we outline how the mechanisms described in Section 3 interact and contribute to obtain an identity management infrastructure, as described above (cf. Section 5). In the first step, we assume that a service transmits a service offer to the privacy dashboard, which describes the kinds of data the service provider wants to access or to collect before/while performing its service. The service offer also includes information concerning whether it supports data provenance tracking, usage control enforcement, or remote attestation. In case the service is provided via a website, we assume that this website will allow the user to direct it to the identity management infrastructure, e.g., by entering an uniform resource identifier (URI) of a communication endpoint of the identity management infrastructure. The user will then open the privacy dashboard to create or update his user profile for the given service. He can set permissions concerning live data acquisition to be performed by the service and also concerning resources he wants to provide to the system by granting access to an external resource server. The former is particularly relevant when we think of interactive assistance systems, which learn certain characteristics of the user while he is already using it, i.e., by evaluating and exploiting data collected by various sensors. In case some service requires an additional identity attribute, which is not yet stored on any attached resource server, the user can add this attribute via the privacy dashboard and directly dispatch it to an existing resource server. Any permission or newly added attribute is automatically synchronized with the resource server responsible for the respective resource. In case the user wants to specify usage restrictions concerning his data, he also needs to choose according usage control policies, e.g., he wants his heart rate related data to be deleted by the requesting service after a trial period of seven days. Finally, a reply containing the user's service selection is sent to the service.

In the second step, the service demands authorization to access the required resource(s) via the identity management server. However, before the identity management server will grant access to any resources, it demands remote attestation of the service's integrity state. For this, it asks the service for a fingerprint reflecting its integrity state, a so-called quote. At the same time, it queries the remote attestation server for a quote of the last state known as trustworthy (the remote attestation server is a trusted third party, which maintains the quotes capturing the states of somehow verified or certified systems). If the quotes are authentic and matching, we proceed with the deployment of usage control policies. Otherwise, the process is aborted.

In the third step, usage control policies are deployed at the service provider's usage control infrastructure in case the user specified such policies in the profile for the requested attributes. In case the service provider supports data provenance tracking, this mechanism is also initiated via deploying a provenance tracking policy for the given user profile [4]. In other words, this step is only performed if the profile for the given service demands data provenance tracking and/or contains usage control policies. However, in case it is performed, policy deployment must be successful, otherwise the process is aborted.

If all steps have been passed through successfully, the identity management server will hand out an authorization ticket to the service. For this ticket, the resource server will finally grant access to the requested resources/identity attributes. While the data are processed at the service provider, the usage control infrastructure will monitor the enforcement of the according usage control policies and forward information related to personal data flows to the data provenance tracking infrastructure.

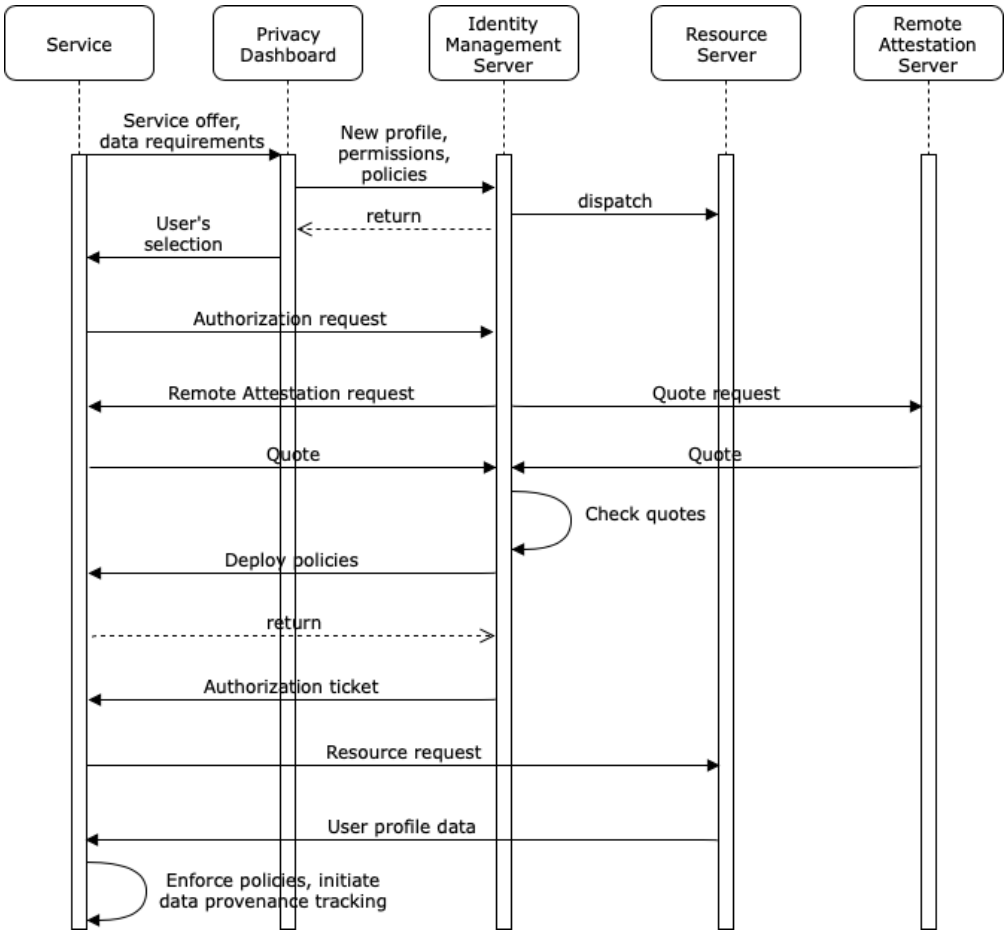

Figure 3. Interaction sequence of the described security mechanisms before using a service.

7. Evaluation

In the beginning of Section 5, we provided a list of the ideal functions of a next generation privacy dashboard. Our prototype PrivacyCockpit [15] integrates this ideal functionality except for the data provenance tracking part, i.e., it includes functions for user profile editing, managing access rights and usage policies, as well as handling trust levels defining whether or not the integrity of a service platform is to be attested. PrivacyCockpit is based on many earlier works (e.g., [16,17]) on data flow tracking for distributed usage control and [12] on integrity attestation of distributed usage control infrastructures. Preliminary results of a user study with 15 participants with a technical background show that the functions of PrivacyCockpit are well understood and manageable [15]. In this user study, participants had to interact with PrivacyCockpit in order to configure user profiles for IOT (the internet of things) devices and web services requesting access to personal data. The tasks the participants had to solve included creating user profiles, creating and maintaining personal data attributes, configuring access rights and usage policies for the user profiles as well as trust levels for the services requesting access to these data. The next refinement of PrivacyCockpit according to the feedback given by the participants is going to be evaluated with a more heterogeneous target group.

Another prototype we developed earlier [3] implements a data provenance tracking infrastructure, as outlined in Section 3.1, including another privacy dashboard dedicated for accessing and visualizing provenance information. This application called PrivacyInsight provides a web-interface for exercising one's right of access (cf. Section 2.4). Its usability and comprehensibility have been successfully evaluated in a detailed user study [3], where data provenance tracking has been integrated into a demo webshop. Participants had to solve certain tasks on their provenance data in PrivacyInsight after

buying a product in the webshop, such as finding out to which third party their payment data had been transmitted after they inserted them into a form in the webshop.

The works described above show that prototypical implementations exist for all the mechanisms and components outlined in this paper. Earlier works have also shown that each particular technology we integrated in our concept is suitable for the purposes we intend. However, we are still working on an integration of PrivacyCockpit and PrivacyInsight as well as on an evaluation of the overall approach. This evaluation will not only include user studies concerning usability and comprehensibility, but also analyses on the expense and feasibility when integrating our mechanisms in legacy products, e.g., web and IOT-based services.

8. Conclusions

Starting out from the rights of the data subject granted by the General Data Protection Regulation of the European Union, we discuss the technical challenges and requirements for data controllers to operationalize these rights in an automatized fashion. Data provenance tracking and distributed usage control provide technically suitable approaches for this automatization. Most importantly, data provenance tracking technology can be implemented to operationalize the right of access, while distributed usage control technology provides mechanisms for implementing the right to rectification, the right to erasure, and the right to restriction of processing. We outlined how a next-generation privacy dashboard-as-a-service could interact with these mechanisms on remote systems. We also need to make use of remote attestation protocols based on hardware trust anchors to ensure that these mechanisms have not been tampered with and that the according remote systems are in a trustworthy state. Based on TPMs only, this cannot yet be done with satisfying guarantees, yet more recent technologies for secure computation enclaves such as Intel SGX can complement TPM-based trusted computing for protecting a client-side reference monitor for distributed usage control in a hostile environment.

Our preliminary results show that our prototypical implementations are already usable and comprehensible for users with a technical background, but the concrete design of the workflows for user interaction require further refinements.

Implementing such mechanisms as we described within a data controllers' computing infrastructure do of course introduce non-negligible overheads in the beginning, which we are going to analyze in future work. Nevertheless, we believe that the advantages of a technical automatization of the rights of the data subject will in the midterm outweigh these initial costs by requiring less personnel for working on data subjects' information requests in time.

Author Contributions: P.B. contributed to the usage control and data provenance tracking aspects and came up with the concept for integrating the technologies described in the paper. E.K. contributed the legal analysis and contributed to the overall concept. P.G.W. did the work related to trusted computing and remote attestation. As the scientific advisor of the other authors, J.B. contributed to the vision of a next generation privacy dashboard as described in the paper.

Funding: This work was partially funded by the KASTEL project (16KIS0754), the GUIDE project (16SV7891), and the InDaSpacePlus project (FKZ 01IS17031) of the German Federal Ministry of Education and Research, BMBF.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Krempel, E.; Beyerer, J. The EU general data protection regulation and its effects on designing assistive environments. In Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference, Corfu, Greece, 26–29 June 2018; pp. 327–330.
2. Birnstill, P.; Beyerer, J. Building blocks for identity management and protection for smart environments and interactive assistance systems. In Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference, Corfu, Greece, 26–29 June 2018; pp. 292–296.

3. Bier, C.; Kühne, K.; Beyerer, J. PrivacyInsight: The next generation privacy dashboard. In Proceedings of the Annual Privacy Forum, Frankfurt am Main, Germany, 7–8 September 2016; Springer: Berlin, Germany, 2016; pp. 135–152.
4. Bier, C. How usage control and provenance tracking get together—A data protection perspective. In Proceedings of the 2013 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, 23–24 May 2013; pp. 13–17.
5. Organization for Economic Cooperation and Development. *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*; OECD: Paris, France, 1981.
6. Simmhan, Y.L.; Plale, B.; Gannon, D. A survey of data provenance in e-science. *SIGMOD Rec.* **2005**, *34*, 31–36. [CrossRef]
7. Freire, J.; Koop, D.; Santos, E.; Silva, C.T. Provenance for computational tasks: A survey. *Comput. Sci. Eng.* **2008**, *10*, 11–21. [CrossRef]
8. Moreau, L.; Groth, P.; Miles, S.; Vazquez-Salceda, J.; Ibbotson, J.; Jiang, S.; Munroe, S.; Rana, O.; Schreiber, A.; Tan, V.; et al. The provenance of electronic data. *Commun. ACM* **2008**, *51*, 52–58. [CrossRef]
9. Godik, S.; Moses, T. *Oasis Extensible Access Control Markup Language (XACML)*; OASIS Committee Secification cs-xacml-specification-1.0; OASIS: Burlington, MA, USA, 2002.
10. Pretschner, A.; Hilty, M.; Basin, D. Distributed Usage Control. *Commun. ACM* **2006**, *49*, 39–44. [CrossRef]
11. Group, T.C. *TCG Specification Architecture Overview*; Trusted Computing Group: Beaverton, OR, USA, 2007.
12. Wagner, P.G.; Birnstill, P.; Beyerer, J. Distributed usage control enforcement through trusted platform modules and SGX Enclaves. In Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies, Indianapolis, IN, USA, 13–15 June 2018; pp. 85–91.
13. Costan, V.; Devadas, S. Intel SGX explained. *IACR Cryptol. ePrint Arch.* **2016**, *2016*, 86.
14. Recordon, D.; Reed, D. OpenID 2.0: A platform for user-centric identity management. In Proceedings of the Second ACM Workshop on Digital iDentity Management, Alexandria, VA, USA, 3 November 2006; pp. 11–16.
15. Hornung, M. Ein Privacy-Dashboard für VerläSsliches Identitätsmanagement in Interaktiven Umgebungen. Bachelor Thesis, Karlsruhe Institut für Technology (KIT), Karlsruhe, Germany, 2018. (In German)
16. Harvan, M.; Pretschner, A. State-based usage control enforcement with data flow tracking using system call interposition. In Proceedings of the IEEE 2009 Third International Conference on Network and System Security, Gold Coast, QLD, Australia, 19–21 October 2009; pp. 373–380.
17. Birnstill, P.; Bier, C.; Wagner, P.; Beyerer, J. Generic semantics specification and processing for inter-system information flow tracking. In *Computer and Network Security Essentials*; Springer: Berlin, Germany, 2018; pp. 445–460.