Towards Enforcement of the EU GDPR: Enabling Data Erasure

Subhadeep Sarkar, Jean-Pierre Banâtre, Louis Rilling, Christine Morin

Towards Enforcement of the EU GDPR: Enabling Data Erasure

Subhadeep Sarkar[†]
Univ Rennes, Inria, CNRS, IRISA, France

Jean-Pierre Banatre[⋆]
Univ Rennes, Inria, CNRS, IRISA, France

Louis Rilling[◇]
DGA, France

Christine Morin[‡]
Univ Rennes, Inria, CNRS, IRISA, France

Email: {[†]subhadeep.sarkar, [⋆]jean-pierre.banatre, [‡]christine.morin}@inria.fr, [◇]louis.rilling@irisa.fr

***Abstract*—With the emergence of the Internet of Things (IoT), an increasing need for preserving the privacy of personal data has been realized. In this context, the EU has recently published the general data protection regulation (GDPR), which ensures strengthening of the privacy rights of the data subjects concerning their personal data. In this paper, we present the importance of having a holistic solution aimed towards the enforcement of the GDPR. As a first step towards the enforcement of the GDPR, we present the research challenges in facilitating the erasure of data as per the *right to erasure*. We also propose the envisaged technical solutions to work through the challenges.**

***Index Terms*—Personal data privacy, EU GDPR, Internet of Things (IoT), Data erasure**

1. Introduction

Over the past four months Alice has developed this habit of weighing herself every morning using the smart connected-scale she bought. Since last month, Alice has gained some weight and this morning while browsing through the Internet, she has seen a number of advertisements on 'weight losing products' and 'weight losing diets'. Given that Alice is always aware of her personal data privacy, specially those concerning sensitive personal information, this incident is recorded as a *breach of privacy*. Clearly, Alice's weight records have fallen in the hands of some third party organizations either through active data sharing or by undesired data leakage or security breaches.

In today's smart world, many individuals like Alice, own their personal set of Internet-connected devices, comprising of smartphones, tablets, wearables for well- being and ubiquitous health monitoring, and other connected objects. These devices often collect the individual's private information and personal data for the sake of service provisioning. Aggregation, analysis, and storage of these data are governed by service providers, beyond the visible range of the data subjects. The term *data subject* corresponds to the natural person, who is the subject of personal data (in oue example, Alice) [1]. Selling and buying of personal data without due consent of the concerned data subjects is now more of a trend than an exception [2], [3]. Also, the end users often end up allowing the service providers to use their personal data for business purposes when they agree to the *terms and conditions* put forward by the providers, in most cases, without reading them [4], [5]. It is often the case that these service providers use these data for organizational profit (for example, directly selling an individual's data to another party in exchange of money [6], [7]) and also for manipulation of the concerned individuals (like for advertisement purposes [8], [9]). Moreover, with the 'capture data at the source' policy, service providers cut the rights and privileges of the data subject on their own data at the very root.

With the advancement in embedded technology and communication systems, we stand today at the brink of a new era of ubiquitous computing, where *data* stand to be the new *currency* [10], [11]. The Internet-of-things (IoT) is not a technology of the future anymore; rather a reality which awaits the correct season to bloom. A report published by Gartner, Inc. states that an estimated 20.4 billion devices will be connected to the Internet, constituting the IoT ecosystem in 2020 [12]. The unprecedented amount of data generated every day by billions of smart things, form the backbone of a massive data-driven business model. In this context, data privacy and data security [13]–[15] are the key terms which demands to be revisited.

1.1. Motivation

Today's standard service model is typically dictated by the service provider with little or no room of customization. The end-users may not express their preferences over the data collection mechanisms and service specifications. The major limitations of the existing service model below are as follows.

(i) *Data subjects have no control over their personal data*: In the current service model the business agreements are designed solely by the authorities of the service providers. The service policies and contracts are usually compiled and put forward by the service provider, to which the clients of this service, i.e., the data subjecs must agree in order to be able to obtain the service. The users do not have means

[⋆]This work was done while Jean-Pierre Banatre was Professor Emeritus at Univ Rennes, Inria, CNRS, IRISA.

to express their preferences over the purpose of use, location of processing, sharing policy, or retention policy concerning their personal data. Disagreeing to the terms essentially implies opting out from the service. Hence, more often than not, individuals do not read the terms and conditions of these online contracts [16] and agree to receive the service.

(ii) *Lack of appropriate data privacy protection measures for IoT*: Although there have been disconnected work done on various aspects, such as data confidentiality, access control, and accountability, a holistic solution dedicated to provided end-to-end protection for data privacy is long overdue. Also, the aspect of '*control over data*' has always been overlooked and data subjects today have no control over their personal data once the data leaves their personal space. Especially, in the context of the IoT, where data streams generated by the IoT devices flow across the domains of multiple stakeholders (such as the fog and cloud service providers), introduction of a cross-domain and coherent solution for privacy protection is important.

Herein comes the importance of the two primary objectives of our research endeavor: (a) giving back the end-users the control over their personal data and (b) implementing the privacy solutions by the principle of design. Although there have been plenty of discussions on issues such as 'privacy by design' [17] and 'privacy as operating system service' [18], implementation of privacy preserving measures at the system level is long overdue.

1.2. The Legal Front: The EU GDPR

Cyber-security and data protection measures have failed to evolve apace with the technological shift and evolution of the data-centric business models. However, in recent years, one of the much discussed issues in the cyber-security community has been protection of personal data privacy and rights of citizens over their personal data. As far as Europe is concerned, in 2012, the European Union (EU) drafted a new set of regulation aimed towards protection of data. The General Data Protection Regulation (GDPR) was finally accepted in April 2016 [1], and comes as a replacement the old Data Protection Directive (95/46/EC). The goal of this new regulation is designed to harmonize data privacy laws across Europe with a vision to protect personal data, empower all EU citizens with data privacy and protection rights, and to reshape the way organizations approach data privacy [1], [19]. Enforced on May 25, 2018, the key changes introduced through the GDPR are as follows.

- *Rights of the data subject*: One of the main changes proposed in the GDPR is empowerment of the data subjects.
  - *Right to be forgotten*: One of the fundamental changes introduced through the GDPR is the data subject's right to be forgotten, also known as the right to erasure. This specifically concerns the retention period of the personal data collected and on-demand erasure of all records of the information. The service provider must comply with such erasure requests unless it has compelling reasons for acting otherwise (as listed in Art. 17(3)).
  - *Right to access*: This grants the data subjects to right to obtain information from the service providers concerning the purpose and location of the processing of the personal data obtained from them. The service provider is also responsible to provide the data subject with an electronic copy of all her personal data in its possession, if demanded.
- *Consent of the data subject*: The GDPR also specifies that instead of long and hard-to-understand terms and conditions, service providers must provide precise and intelligible terms and conditions to the data subjects. Revocation of consent also should not be unnecessarily complicated.

Apart from these, the GDPR also regulates the issues of data portability (from one service provider to another), breach notification, legal assistance through the data protection officers, and penalty for service providers for breach of contract. Thus, the GDPR strengthens the legal front of data privacy protection.

This work focuses particularly on empowering the data subjects by allowing them to express their preferences over the usage of their personal data and towards enforcement of the *right to erasure*. In our example, this will, allow Alice to dictate the purpose of use, the location of the processing, the retention period, and sharing policies concerning her personal data. Alice may also for erasure of all records of her. On the legal side, Alice can lodge a complain or issue an inquiry after she suspects a breach in contract.

1.3. Contributions

In the context of the IoT, where billions of Internet-connected devices continuously transmit heterogeneous data streams at a high velocity demanding for real-time services, existing solutions fail to provide a holistic solution for preserving data privacy. The main contribution of this work this to identify the research challenges in providing an end-to-end solution for data privacy protect in a multi-stakeholder, distributed service architecture, and present the envisaged potential solutions to those challenges.

We present the fundamental research challenges with respect to implementation of data erasure as per the data subjects' preferences, in this work. At first, we discuss the significance of having a policy language which is customizable by the data subjects based on their preferences, and also highlight the importance and the challenges associated with policy enforcement in a multi-stakeholder service architecture. We also propose the principal mechanisms and design

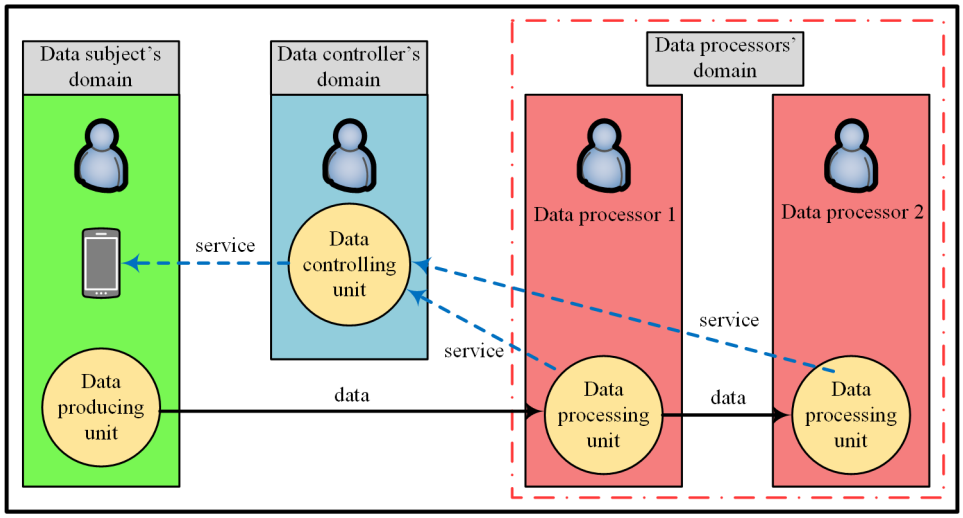

Figure 1: Service model highlighting the domains of the different stakeholders

principles of the different distributed algorithms, which are envisaged as potential solutions to the challenges identified.

The EU GDPR is the one of the primary catalysts behind our research endeavor. The broader objective of this endeavor is to analyze the technical challenges envisioned in course of enforcement of the GDPR and to propose a set of solutions for protecting data privacy *by design*.

1.4. Paper organization

The rest of the manuscript of is organized as follows. In Section 2, we present the technical challenges associated with this research undertaking. In Section 3, we present the solutions envisaged to facilitate policy customization and efficient policy management, while the potential solutions towards enabling data erasure are presented in Section 4. Finally, the manuscript is concluded in Section 5.

2. Technical Challenges

Before we discuss the technical challenges concerning our research endeavor, we present the reference service model. As shown in Figure 1, the service model has three primary categories of stakeholders, i.e., data subject, data controller, and data processor. As per the GDPR, data controller refers to the stakeholder, which dictates the purpose and means of processing of the personal data, and data processor corresponds to the stakeholder which processes the data on behalf of the data controller [1]. In the model, at one end are the data subjects, who own a set of wearable and non-wearable smart devices. Each such device, provides a set of services to the data subject. The data controller is responsible for provisioning of the services concerned. The data subject and the data controller must be bound by a set of business agreements for the service model to be valid. The data controller's ability to obtain the data produced by the IoT devices is based on these business agreements. Note that, any data produced by the device, in this case, will be considered as personal to the data subject. The data controller may outsource processing of the acquired data to one or more data processors, who are responsible for processing the data on its behalf and sending back the results. These data processors can be referenced as typical fog or cloud service providers.

Next we illustrate, with examples, the challenges in implementing our solution. The right to erasure concerns the data subjects right to ask the data controller to erase all her personal information after a pre-agreed retention period or on-demand. Considering the in an IoT environment, data are transmitted in form of a stream, a data segment is defined as a finite set of data tuples, which arrive successively in a data stream, and data tuple refers to the smallest data unit transmitted by a data producing unit. Data erasure, in this context, is defined as follows.

Definition 1. Data erasure is defined as the process invocation of which, either by the data subject or as per the agreement between the data subject and the data controller, ensures that
(i) replication and dissemination of the data segment requested to be erased, must be stopped,
(ii) all replicas of the data segment, which are in possession of the data controller and the data processors, must be erased,
without undue delay, unless there is a compelling reason as per Art. 17(3) of the EU GDPR for not doing so.

We now present the fundamental research challenges concerning implementation of the right to erasure.

2.1. Challenges in Policy Design

First we present the challenges identified in design and management of the privacy policies, in the context of enabling data erasure.

Determining the granularity of erasure: Our first goal is to allow the data subjects to express their preferences over the retention period of their personal data during the contract negotiation phase. The challenge, here therefore, is to optimally choose the granularity of the erasure operation. Now to erase the data, we rely on the cryptographic erasure techniques [20], [21], as they provide the most secured yet fast mode of data erasure and also preserves the integrity of the devices. Erasing data using the cryptographic means essentially corresponds to secured deletion of the encryption keys corresponding to the data. Therefore, smaller erasure granularity (such as hourly or daily) will correspond to managing a higher number of encryption keys (also higher cost). Therefore, based on the sampling rate of the data, the data controller must optimize and ascertain the minimum allowable size of a data segment for the erasure operation, and accordingly construct the erasure-related policies. Considering the heterogeneity of the data and the varied data-rates of different IoT devices (e.g., a wearable ECG sensor transmits a data frame every 0.2 seconds, whereas a smart scale reports only when someone stands on it, i.e., typically 1-5 times in a day), it is a challenge to determine the erasure granularity for different data streams.

Policy management: The business policies, which serve the same purpose are usually compiled together to constitute a business agreement. Management of these business agreements, however, is no less important than design and formalization of the customizable policies. Ensuring the coherence and consistency of the business agreements between the different parties is highly challenging, particularly in a multi-stakeholder and distributed service architecture.

Legal aspects of the policies designed: While our goal is to give back the data subjects the right over their personal data by facilitating policy customization, conformation with the legal regulations is advocated. Also, after the policy negotiation phase is over, and both the stakeholders come to an agreement, it is important to have a verification mechanism that neither party tamper with the policies agreed, and any dispute must be settled through by invoking the mechanism. Design of such mechanism involves challenges both in the technological and legal fronts.

2.2. Challenges in Enabling Data Erasure

In order to enforce the right to erasure, invocation of the erasure operation must ensure that the replicas of a given data segment are erased (or will eventually be erased). This includes all replicas of the data located within any of the storage units of the multi-domain, distributed framework. Below, we present the fundamental technical challenges associated with enforcement of the data erasure facility.

Identification of data replication: Given that the EU GDPR is applicable to not only the original data, but also any replicas of the same, it is important for us to identify all replicas of a given data segment. The next research challenge, therefore, is to detect efficiently the process of data replication within a machine. Data, in an IoT ecosystem, are generated and transmitted in the form of streams, and are considered to be stored segment-wise in some persistent storage unit (such as files and databases). Considering that a process either replicates an entire storage unit, or simply a subset of it, identification of the replication operation is a highly challenging task.

Designing the distributed data erasure algorithms: Design and implementation of the data erasure algorithms for a distributed, multi-domain service model is also not straightforward as it involves a number of technical challenges. We present the three main challenges involved with the process of data erasure below.

- Concerning the erasure technique of data, the question of 'how sensitive personal data can be erased efficiently and securely' itself looms as the biggest research challenge. While cryptographic erasure of data is considered to be fast and secured [20], [21], 'losing' the key is never simple, as digitally erased data can be retrieved through manipulation of the physical storage disk. Secure management of the keys is also an open research issue.
- Enforcement of the right to erasure fundamentally implies design and implementation of (a) an algorithm that would facilitate automatic erasure of a data segment after expiration of its retention period and (b) another algorithm that would facilitate erasure of data segments on demand of the corresponding data subjects. Now, considering that these algorithms are executed on a distributed, multi-domain service architecture guaranteeing atomicity, mutual exclusion, and consensus is crucial and challenging.
- By definition, data erasure corresponds to erasure of all replicas of a given data segment. However, some data storage sites may be offline or inaccessible when the erasure operation is invoked. Therefore, our final challenge is to ensure 'eventual' erasure of all replicas of the data.

3. Enabling Policy Customization

While there are multiple languages specifically designed for expression of privacy policies, such as APPEL [22], EPAL [23], XPref [24], and XACML [23], none of the existing languages support customization of the policies based on the data subjects' preferences. To enable customization of the personal data privacy policy as per the preferences of the data subject, it is important to identify the different types of business agreements between the different stakeholders. It is also crucial to provide system solutions to ensure the consistency of the different stakeholders in terms of the version of the contract they agreed to, in order to facilitate policy enforcement.

3.1. Business Agreements

First, we define and illustrate the different types of business agreement which may be in place between different pairs of stakeholders.

Definition 2. *Service level agreement* (*SLA*) is a formal contract between a service provider and its customer, which details about the nature, quality, scope, span, and responsibilities regarding the service to be provided.

An SLA, in general, includes details of the type of services to be provided, the reliability, responsiveness, expected performance of the services, fallibility and means of recovery, penalty for breach of the agreement, and other constraints. As shown in Figure 2, for the SLA between the data subject and the data controller, the data controller is the service provider offering its services against the data obtained from the data producing unit owned by the subject. For the SLAs between the data controller and the data processors, however, the data controller acts as the client and the data processors provide the services after processing the data obtained from the data controller.

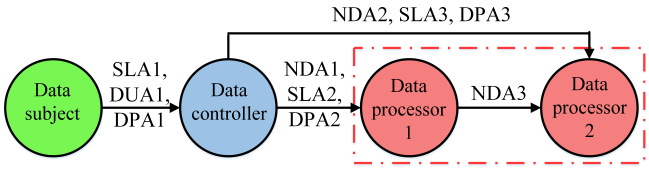

Figure 2: Data flow outline

Definition 3. *Data use agreement* (*DUA*) is a legal contract which dictates the terms and conditions for certain non-public data which are subjected to restricted use, and it takes place between two entities (individuals and/or organizations), one of which is the rightful owner of the data.

A DUA is an agreement between the data subject and the data controller. It includes the privacy rights associated with the data, data confidentiality details, access control, and accountability concerning the violation of the agreement.

Definition 4. *Data protection agreement* (*DPA*) is a legal contract which describes the obligations to safeguard the data, details about the security measures adopted to protect the data from from both physical harm and cyber-attacks, and about the accountability and compensation in case of a breach, and is offered by an entity (as an individual or an organization) to another from which it acquires the data.

A DPA, therefore, may be in place either between the data subject and the data controller or between the data controller and the data processor, with the latter entity acquiring the data from the former, in both cases.

Definition 5. *Non-disclosure agreement* (*NDA*) is a legal contract between two or more organizations who agree not to disclose confidential information which is shared among them for certain purposes, but wish to restrict from generalized use.

As shown in Figure 2, there exists an NDA between *data processor 1* and *data processor 2*, i.e., the organizations who exchange data or information in between them for business purposes. Note that, there may be NDAs between the different data processors which are connected in terms of data exchange, as well.

3.2. Policy Management

In a multi-stakeholder service model, where different domains are owned and managed by different stakeholders, attaining coherence is challenging. Our objective, thus, is to ensure by design, that all the stakeholders concerning a business agreement are consistent and coherent in terms of the version of the agreement agreed upon, at any point of time. This prevents tampering of the business agreement either by a stakeholder or by an unauthorized third party. The fact that any modification of a local copy of an agreement, in possession of a particular stakeholder, must be approved by every other stakeholder concerned before it is committed, essentially disregards any unapproved modification to an agreement. This, therefore, prevents any of the stakeholders from modifying their local copy of the agreement for their business benefits. Also, cases in which the security of a particular stakeholder is compromised and certain business agreements are maliciously tampered, the principle of unanimous consensus before commit, localizes the breach and helps in recovery. To put it into perspective, in our example with Alice, after successful negotiation of the business contracts between Alice and her service provider, it is important to ensure that both parties have with them the exact same version of the contracts. Therefore, when Alice speculates that her personal data may have been misused and initiates an investigation against the same, she is at least certain of the fact the service provider and she have the same version of the business agreements. In other words, in case of a possible breach in a contract, the service can not come up with a modified (tampered) version of the contract so as to negate the breach. Architecturally, there are two possible ways of solving this issue, which we discuss on below.

Centralized approach: The centralized approach involves a central trusted verification authority (TVA), which acts as the management authority for the business contracts between different groups of stakeholders. The TVA can either operate at a global scale or in a region-specific manner, based on the location of the stakeholders. The operating principle is based on simple timestamp-based verification. After successful negotiation of contracts between a group of stakeholders, each stakeholder independently sends a digitally signed copy of the contract (the one, which is at their possession) to the TVA. After receiving the duly signed copies from all the stakeholders concerning the agreement, the TVA performs a coherence check on the coherence of the agreement timestamp (i.e., the time when the negotiation was successful), followed by a consistency check on

the different copies of the agreements. One verified to be coherent and consistent, the TVA retains a copy of the signed contracts and sends a message bearing a unique validation reference to the corresponding stakeholders. Otherwise, it simply sends to all the stakeholders, informing about the disparities observed. In cases of modification of an existing contract, the same process is followed and a modification is only considered to be in place, only after the validation reference is issued by the TVA against the modified version of the contract. However, there are a couple of disadvantages of this centralized approach.

- *Trust issues*: A major point of concern for this approach is the trustworthiness of the centralized authority, and therefore, its composition. The most plausible solution to this, would be to entrust the cyber-police to take up this role. However, as it stands, this needs more careful investigation before we entrust someone with the responsibility.
- *Single point of failure*: Like most centralized solutions, this approach also suffers from single point of failure, i.e., in this case, if the TVA crashes or is compromised, all records are lost. One potential solution to this issue is to add redundancy to it.

The other way of managing the business contracts is a distributed one with no central authority involved.

Distributed approach: The distributed policy management approach basically thrives on the principle of distributed ledger technology (DLT) [25], [26]. The idea is that once agreed upon, each stakeholder stores the contract as a part of a continuous ledger stored at their end. Any modification of this ledger may take place only after a consensus is reached among all stakeholders of the agreement, and in form of appending to the list. A distributed ledger is essentially a database which is managed and updated independently by each participant node of the network. Every node processes every transactions taking place within the network independently and derives its own conclusions. A node then casts its vote against its derived conclusions. Once a consensus is reached, all nodes update their own version of the ledger, and thus, a single identical copy of the ledger is maintained throughout the network. Every entry in the ledger has a timestamp associated with it and bears a unique cryptographic signature, which ensures the integrity and consistency of the ledger.

One possible implementation of this DLT is blockchains [27], [28]. DLT offers multiple advantages over the centralized approach. First of all, dis-intermediation by a centralized authority makes the process of validation simpler and issue of single point of failure is resolved. Moreover, DLT does not require trust to be placed on any authority and can operate in a ecosystem comprising of untrusted parties as well. It also offers transparency among all stakeholders of the system and guarantees timely validation of the requests. However, it may require additional redundant computations to be performed, which may lead to higher energy consumption and computational cost.

Given the multi-fold advantages of the distributed approach over its centralized counterpart, we are inclined towards adopting it for the sake of efficient and secured policy management. Also, each agreement must have an exclusive reference, known as the unique agreement identifier (UAI), through which any agreement can be uniquely identified. We will have to design a means to uniquely identify the business agreements, as it will be used as a reference to identify the contracts associated with a data stream.

4. Implementation of Data Erasure

Towards enabling data erasure in a multi-stakeholder, distributed service model, we present the primary objectives.

A. **Tracking all replicas of data:** Tracking all replicas of the data segments in a multi-domain architecture by correctly identifying the data storage location.
B. **Enabling data erasure:** Design of a set of distributed services, which will ensure 'eventual' erasure of all replicas of the data segment.

4.1. Tracking of Data

In order to facilitate erasure of all replicas a given data segment, the first milestone is to identify the (persistent) storage location of the data segments. Considering the multi-domain service model, as described in Figure 1, the flow of data spans multiple processing devices located at different domains. In each of these devices the data may be replicated multiple times during the course of its processing. We, therefore, have two primary challenges associated with data tracking are as follows.

(i) Identifying the event of data replication
(ii) Locating the storage location of all replicas of the data segment

We present the solutions envisaged to address each of these challenges below.

Identification of the event of data replication

We investigate into how to track all replicas of a given data segment, and for this we first need to distinctively identify the system processes, which are responsible for replication of data. Note that, by replication, here we specifically refer to duplication of data segments where for the original and replica data segments reside in the same computing device. Any cross-device duplication of data are categorized under data flow management.

To efficiently detect the process of data replication we envision to exploit the operating system capabilities. By restricting the purpose (read or write) for which a process may access a persistent memory location, we first of all regulate the privilege under which the data may be accessed. The next phase is to detect whether or not a process with sue access privilege is making a replica of the data. Considering that a process either replicates an entire storage unit, or simply a subset of it, identification of the

replication operation is a highly challenging task. Herbster *et al.* [29] proposed the concept of privacy capsules (PCs), which is an Android platform execution model for mobile applications that prevents unauthorized flow of information. Although, the scope of the work is restricted to monitoring data flow within a single device, the principles of PCs can be useful for process monitoring, in our context. We will develop a novel mechanism by linking the principles of PC with the operating system capabilities, to put restrictions on processes to access certain memory addresses containing the sensitive personal data.

Locating all data replicas

In order to locate the location of the data replicas, we first need to track the flow of the data. Data slow may be broadly categorized into two types – (a) flow of data from one device to another within the same domain, and (b) flow of data from one device to another, where the devices are located in different domains. For this, we propose a novel tainting method which will help in uniquely tainting [30] the data stream generated by each data producing unit. Every data stream is tainted at the source and the same taint is carried forward even when the data transits across domains. The taint contains, as a part of it, the relating UAI, which assists in quick referencing to the business agreements. The taint-tables store the necessary information about which tainted data-stream has been re-directed to which physical machine. The taint-tables can only be accessed or modified by processes (or threads) with due capabilities, i.e., appropriate access rights. Another challenge, in this regard, which will be interesting to examine, is ensuring cross-domain integrity of the information on data flow and storage.

It is, therefore, evident that operating system capabilities are to play an important part in design of our proposed solution. For this, the services provided by the operating system, and thus, the choice of operating system would be crucial.

4.2. Enabling Data Erasure

The final piece of the puzzle is to design a distributed mechanism to facilitate the erasure of data. Given that we now have the required information about the storage location of all replicas of the data segment to be erased, we have to design data erasure mechanism. Towards this, we have identified the following challenges.

(i) Design of a secure data erasure technique
(ii) Design of the data erasure algorithms
(iii) Enabling *eventual* erasure of inaccessible data

We discuss about each of the challenges separately below.

Designing the data erasure technique

Considering the fact that the work deals with erasure of personal and sensitive personal data, it is important to wisely choose an erasure mechanism, which would securely erase the data, leaving behind minimum contingent of retrieval of the erased data. For this, we rely on encryption-based erasure techniques, where instead of erasing the data segments in question, we 'lose' the encryption (or decryption) key corresponding to the segments. However, 'losing' the key is never very straightforward, as retrieval of the key, through any possible means, indicates access to all the erased data. For this, we will investigate secure key storage techniques, such as secure element (SE) [31] and hardware security module (HSM) [32], which offer hardware-based solutions for secure key storage and application execution. However, these solutions suffer from the problem of single-point of failure, i.e., if the device stops operating or the authentication unit within the device does not function properly, all data associated with the keys, which are stored inside the device, effectively becomes 'lost'. Also, hardware-based solutions are often costly and may induce performance overhead. Our research, therefore, aims towards designing of an erasure technique, which is secured, cost-effective, and resistant to failures.

Design of the data erasure algorithms

In this work, we focus on erasure of data driven by two factors – (a) erasure of data after the expiration of its retention period and (b) erasure of data on the demand of the data subject. We plan to design two separate algorithms for each type of erasure. Erasure of data after expiration of its retention period is directly associated with the contract associated with the data segment in question, and the erasure algorithm will be automatically invoked in due course of time. On the other hand, the erasure algorithm for on-demand erasure of data has to be manually invoked by the administrator for it to take effect. The main challenge in designing these algorithms is to ensure that integrity of the storage policy is preserved following the erasure mechanism. Considering that information derived from the raw data may also be categorized as personal data, lossless and integrity-preserving erasure of data in a multi-domain service model is a challenge. Note that, the erasure operation is invoked only if the erasure policy or erasure request complies with Art. 17(3) of the GDPR. Otherwise, for data which may be retained by the data controller for research purposes and in public interest, the data segments in question undergo the process of pseudonymization. Pseudonymization of data, however, is beyond the scope of this work, and is a part of our future work.

Enabling *eventual* erasure of data

The erasure algorithms, discussed till now, are designed to erase all accessible replicas of the data segment to be erased. However, some data storage sites may be offline or inaccessible when the erasure operation is invoked. Our goal, therefore, is to ensure *eventual* erasure of all replicas of the data. For this, we will design a persistent erasure technique, that holds the last state of the uncommitted or partially committed erasure requests, and effects opportunistic erasure of the same. This will guarantee that data will never be used anymore once an erasure request is issued.

5. Conclusion

This work is poised as a position paper aiming for the enforcement of the EU GDPR focusing on the right to erasure. We identified the research challenges in enforcement of the *right to erasure* – one of the crucial changes introduced through the GDPR. We also presented some technical solutions envisaged to overcome the challenges and facilitate erasure of data as per the GDPR norms. Our next objective is to implement the proposed solutions and validate their correctness. We also plan to extend this work by investigating on how to provide support at system level for the other regulations of the GDPR.

References

[1] "Regulation (EU) 2016/679 of the European Parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," *Official Journal of the European Union (Legislative Acts)*, pp. L 119/1 – L 119/88, May 2016.

[2] A. Acquisti, L. K. John, and G. Loewenstein, "What Is Privacy Worth?" *Chicago Journals*, vol. 42, no. 2, pp. 246 – 274, Jun 2013.

[3] P. Boutin. (2016, May) The Secretive World of Selling Data About You. Newsweek. [Online]. Available: http://goo.gl/2k34zz

[4] T. Brewster. (2014, May) Facebook, Google and personal data: What's your worth? BBC. [Online]. Available: http://goo.gl/bfL7ZR

[5] J. Morris and E. Lavandera. (2012, Aug) Why big companies buy, sell your data? CNN. [Online]. Available: http://goo.gl/t7bH9R

[6] L. Mirani and M. Nisen. (2014, May) The nine companies that know more about you than Google or Facebook. Quartz. [Online]. Available: https://goo.gl/CQD7Kr

[7] J. Sadowski. (2016, Aug) Companies are making money from our personal data but at what cost? The Guardian. [Online]. Available: https://goo.gl/JioyW5

[8] D. Gross. (2012, Jun) Facebook to show you ads based on your Web browsing. [Online]. Available: http://goo.gl/prqwrP

[9] S. Felix. (2012, Sep) This Is How Facebook Is Tracking Your Internet Activity. Business Insider. [Online]. Available: http://goo.gl/BPvUoR

[10] W. D. Eggers, R. Hamill, and A. Ali, "Data as the new currency: Governments role in facilitating the exchange," *Deloitte Review*, vol. 13, pp. 19 – 31, Jul 2013.

[11] B. Schmarzo. (2017, Jan) Data is a New Currency. Dell EMC Services. [Online]. Available: https://goo.gl/haf3uo

[12] (2017, Feb) Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016. Gartner, Inc. [Online]. Available: https://goo.gl/X7GHNj

[13] J. H. Ziegeldorf and O. G. M. andd Klaus Wehrle, "Privacy in the Internet of Things: Threats and Challenges," *Security and Communication Networks*, vol. 7, no. 12, p. 2728 2742, Dec 2014.

[14] M. Abomhara and G. M. Koien, "Security and privacy in the Internet of Things: Current status and open issues," in *2014 International Conference on Privacy and Security in Mobile Systems*, May 2014, pp. 1 – 8.

[15] Y. H. Hwang, "IoT Security & Privacy: Threats and Challenges," in *Proceedings of the* 1^{st} *ACM Workshop on IoT Privacy, Trust, and Security*. New York, NY, USA: ACM, 2015, p. 1.

[16] D. Berreby. (2017, Mar) Click to agree with what? No one reads terms of service, studies confirm. The Guardian. [Online]. Available: https://goo.gl/Nkjag3

[17] (2009, Jan) Privacy by Design. Information & Privacy Commissioner of Ontario. [Online]. Available: https://www.ipc.on.ca/wp-content/uploads/2013/09/pbd-primer.pdf

[18] S. Ioannidis, S. Sidiroglou, and A. D. Keromytis, "Privacy As an Operating System Service," in *Proceedings of the* 1^{st} *USENIX Workshop on Hot Topics in Security*, ser. HOTSEC'06. Berkeley, CA, USA: USENIX Association, 2006, pp. 8–8. [Online]. Available: http://dl.acm.org/citation.cfm?id=1268476.1268484

[19] "The Guide to Data Protection," Inofrmation Commissioner's Office (ICO), Tech. Rep., Jan 2017. [Online]. Available: https://ico.org.uk/media/for-organisations/guide-to-data-protection-2-7.pdf

[20] P. Gutmann, "Secure Deletion of Data from Magnetic and Solid-state Memory," in 6^{th} *Conference on USENIX Security Symposium, Focusing on Applications of Cryptography - Volume 6*, ser. SSYM'96. Berkeley, CA, USA: USENIX Association, 1996, pp. 8–8.

[21] A. R. Regenscheid, L. Feldman, and G. A. Witte, "NIST Special Publication 800-88, Revision 1: Guidelines for Media Sanitization," *NIST ITL Bulletin*, Feb 2015.

[22] L. Cranor, M. Langheinrich, and M. Marchiori, "A P3P Preference Exchange Language 1.0 (APPEL1.0)," *W3C Working Draft*, Apr 2002. [Online]. Available: https://goo.gl/DiKj1t

[23] A. Anderson, "A Comparison of Two Privacy Policy Languages: EPAL and XACML," Tech. Rep., 2005.

[24] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "XPref: a preference language for P3P," *Computer Networks*, vol. 48, no. 5, pp. 809 – 827, 2005.

[25] P. Evans-Greenwood, "Distributed Ledgers & Linked Data," in *Proceedings of the* 26^{th} *International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2017, pp. 1451–1451.

[26] C. Khan, A. Lewis, E. Rutland, C. Wan, K. Rutter, and C. Thompson, "A Distributed-Ledger Consortium Model for Collaborative Innovation," *IEEE Computer*, vol. 50, no. 9, pp. 29–37, 2017.

[27] K. Christidis and M. Devetsikiotis, "Blockchains and Smart Contracts for the Internet of Things," *IEEE Access*, vol. 4, pp. 2292–2303, 2016.

[28] L. Luu, D.-H. Chu, H. Olickel, P. Saxena, and A. Hobor, "Making Smart Contracts Smarter," in *ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2016, pp. 254–269.

[29] R. Herbster, S. DellaTorre, P. Druschel, and B. Bhattacharjee, "Privacy Capsules: Preventing Information Leaks by Mobile Apps," in 14^{th} *Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2016, pp. 399–411.

[30] W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, "TaintDroid: An Information-flow Tracking System for Realtime Privacy Monitoring on Smartphones," in 9^{th} *USENIX Conference on Operating Systems Design and Implementation*. USENIX Association, 2010, pp. 393–407.

[31] P. Urien, "Towards secure elements for the Internet of Things: The eLock use case," 2^{nd} *International Conference on Mobile and Secure Services*, pp. 1–5, Feb 2016.

[32] B. Koppel and S. Neuhaus, "Analysis of a hardware security module's high-availability setting," *IEEE Security Privacy*, vol. 11, no. 3, pp. 77–80, May 2013.