Privacy & Cloud Services: Are We There Yet?

Davit Marikyan[1], Jose Llanos[2], Masoud Barati[3,5], Gagangeet Aujla[4], Yinhao Li[1,3], Kwabena Adu-Duodu[1],
Sabeen Tahir[3], Omer Rana[3], Savvas Papagiannidis[1], Rajiv Ranjan[1], and Madeline Carr[2]

[1]*Newcastle University, UK*; [2]*University College London, UK*; [3]*Cardiff University, UK*;
[4]*Durham University, UK*; [5]*Edinburgh Napier University, UK*

2021 IEEE International Conference on Service-Oriented System Engineering (SOSE) | 978-1-6654-3477-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/SOSE52839.2021.00006

Abstract— **The General Data Protection Regulation (GDPR) remains an important requirement for many electronic services which utilise user data. GDPR compliance verification for a cloud provider is aimed to confirm that personal data provided by a user is shared in-line with the requirements of this legislation, so that any subsequent audit carried out on the provider does not lead to a financial penalty. This verification involves two aspects: (i) ensuring that user consent has been obtained; (ii) sharing of data with external cloud providers is undertaken in a transparent way, so that the user is aware of which providers the information was shared with and for what purpose. Using a survey we describe why users are still ambivalent about the use of GDPR – and how its adoption can be improved using a Blockchain-based architecture that can provide greater transparency on how GDPR compliance is supported by cloud providers.**

I. Introduction & Motivation

Understanding the role of behavioural factors, such as perceived performance and ease of use [1], [2] influences how users adopt and utilise new software infrastructures. However, *embeddedness* of internet-enabled devices in our daily lives, and their integration in private and organisational routine, have led to unprecedented changes in user perceptions and behavioral patterns [3]. This is also particularly relevant in the context of privacy-enhancing technologies that attempt to limit exposure of personal user data.

The ubiquitous connectivity of people to the Internet have emphasised concerns over the degree to which devices ensure information privacy and security [4]. Perceived privacy is an individual's belief about how their personal information is acquired, controlled, stored and used [4]. While several researchers have empirically confirmed the direct effect of this factor on user behaviour, others found that the effect is non-significant [5]. This forms the key focus of this contribution – i.e. to what extent do users consider the utility and benefit of privacy-preserving technologies, including support provided by legislation such as the General Data Protection Regulation (GDPR). We motivate this work by two questions:

- Q1: how do users perceive benefit in using privacy technologies to support GDPR legislation, particularly in the context of cloud hosted services?
- Q2: is GDPR seen as a barrier to making more effective use of cloud services, i.e. do users consider GDPR as a barrier to more effective use of services from a cloud provider, or as an important requirement that needs to be fulfilled before initiating any interaction with a cloud provider?

In the context of Q1, we also inquire if providing user consent for cloud providers to use their personal data (Art. 6 of GDPR), a key tenet of many articles within the GDPR legislation, is fully understood by users.

In order to secure broad access to personal data under the semblance of GDPR compliance, website owners are increasingly relying on 'dark patterns' – i.e. interface designs which seek to nudge users into desired privacy-intrusive choices through *deceitful* interaction flows. Examples of GDPR-non-compliant dark patterns are intrusive default settings, the concealment of privacy-friendly choices, requiring extra effort from users to choose them, and take-it-or-leave-it options which bundle many data processing operations[1]. Empirical research has shown that many dark patterns on the Web are common, even widespread, such that the provision of the service on the basis of implicit consent (i.e. no consent asked) makes the rejection of all tracking technologies substantially more difficult than accepting them, with buried pre-ticked boxes for optional vendors (e.g. third-party trackers) or purposes/ categories of data processing [6]. Similarly, some websites relying on advertising as their main revenue source, coupled with the complexity and multiplicity of actors involved in the advertising actor chain, results in numerous instances where data processing by specific entities is not duly informed, and consequently users cannot possibly be aware of them.

In the ongoing pandemic, which has forced us to increase our reliance on digital technologies to conduct our lives and endure restrictions, the above mentioned practices are exacerbated, making the notion of individual control all the more illusory. For example, imagine you want to have a meal in your local restaurant – upon entering the premises you realise there is no 'traditional' customer service; rather, you need to download a booking and payment app., entering personal details to register. However, registration cannot be completed — and therefore you cannot be served -– until you tick a box signalling consent to terms and conditions that allow for extensive processing of personal data based on several legal basis, for multiple purposes unrelated to the transaction you had in mind (i.e. having a simple meal). Any consent given as a result is invalid under the GDPR, as the consent request involved no real choice. Yet, without the competent Data

[1] Norwegian Consumer Council (Forbrukerrådet), "Deceived by design" (2018) – available at: `https://www.forbrukerradet.no/undersokelse/no-undersokelsekategori/deceived-by-design/`

2642-6587/21/$31.00 ©2021 IEEE
DOI 10.1109/SOSE52839.2021.00006

Protection Authority (DPA) actually finding and penalising the breach, consent obtained in this way is 'fair game'[2].

The remainder of the paper is structured as follows: in Section II we present GDPR requirements for cloud services, especially where a number of interconnected cloud providers need to work together. In Section II-A we describe how user uptake of on-line services is impacted by concerns about data privacy, an aspect that has been amplified during the recent pandemic. Due to increasing requirements to work from home (for a distributed work force), employees are often required to make use of on-line service (and platforms) to work from home. Privacy considerations are therefore overlooked to ensure that a minimal level of work can continue from home. Under this general context, GDPR considerations become even more significant, as these provide protection for users when sharing their personal data to access on-line services. In Section II-B we describe our survey used to assess interest in GDPR compliance verification, and to what extent users are fully aware of how cloud providers manage GDPR compliance verification for user provided data. We conclude our work in Section V and offer our views on future work on how data privacy legislation can be used more widely.

II. GDPR use for Cloud Services

Even though online cloud service providers which host services, such as Amazon AWS and Microsoft Azure, try to convince end-users (consumers) about their security and privacy strategies, there still exist a number of *grey* areas around privacy legislation and adoption by cloud service providers. Specifically, it is hard to convince users when a multi-cloud service chain is involved where different cloud service providers act as data controllers and data processors. In such an environment, the liability of data leakage or unauthorized access to personal data of a user can be conflicted. To enforce liabilities and responsibilities for privacy, the GDPR legislation ensures that organizations must observe specific rules based on the following principles: lawfulness, fairness and transparency, purpose limitation, data minimisation, accuracy/ consistency, storage limitation, integrity and confidentiality (security), and accountability. Any non-compliance of the GDPR legislation associated with these aspects, e.g. leakage or non-reporting of an unauthorized use of personal data, may lead to financial consequences for cloud providers. For instance, the 2018 Cambridge Analytica scandal ended up in a financial penalty of US$5B for Facebook[3] for leakage of user profiles without user consent.

In [7] an extensible layered privacy language (LPL) was presented to formally express privacy policy, user consent and retention management in GDPR. However, the implementation of LPL in a real environment (e.g., cloud and IoT) was not examined. Moreover, the data movement across different data processors/ controllers was not taken into account in LPL. A privacy preference language, called YaPPL, was introduced in order to realize user consent requirements of GDPR within an IoT ecosystem [8]. This work also aligns with the focus of our efforts, however the verification of user consent based on the purposes of data processing was not undertaken in a transparent and automatic way in YaPPL. Another recent extension to YaPPL focuses on the development of a transparency information language and toolkit (TILT) [9] to improve the transparency of information enforced by GDPR, enabling a more automated use of such information in modern information system engineering. The scalability of the toolkit was not evaluated in cloud-based systems and its comparison with other technologies such as a Blockchain was not discussed.

Overall, addressing issues of privacy specifically in cloud-hosted services raises a serious question about how cloud providers need to handle personal (or sensitive) data that users entrust upon them while accessing cloud services. Due to the complexity of the cloud hosting process, cloud providers may host data and services at different global locations. Additionally, the user base can also be scattered across the globe, eventually leading to loopholes due to different data privacy regulations (at some locations, no regulations). Some of the critical challenges and research questions related to the use of GDPR legislation for cloud services include:

- How does a cloud provider understand what constitutes“personal data”?
- How do we design a compliance-aware platform to host cloud services? Compliance-aware implies that GDPR legislation is automatically enforced across such a platform, providing greater *trust* to a user that service access and sharing of personal data will automatically preserve their privacy.
- How can we identify and map data privacy regulations to monitoring granularity (i.e. what should a cloud provider monitor and at what frequency to support privacy audits) while provisioning cloud services?
- How do we verify compliance in an automatic manner and ensure the 'right to be informed' obligation?
- How do we equip existing cloud platforms with a monitoring strategy for logging information required for verifying GDPR compliance? This monitoring should not impede the performance of the service hosted by the platform but still ensure compliance with privacy legislation.
- How do we confirm GDPR compliance and provide a trusted solution to securely log what personal data is processed by which provider – especially where multiple providers are involved in offering a particular service to a user?
- What approach(es) ensures the translation of GDPR obligations (e.g. data protection and data transfer) into smart contracts and supports an automated verification of GDPR obligations over the activities of providers? Although, legal texts are often written in an open manner, as aspect that is considered to be a highly desirable

[2] Rana, Llanos, Carr, "Lessons from the GDPR in the COVID-19 era", available at: `https://www.academia.edu/45666233/Lessons_from_the_GDPR_in_the_COVID_19_era`

[3] `shorturl.at/hlmS2`

feature, as it leaves room for interpretation on a case by case basis, such ambiguity poses challenges for automatic compliance checking. Therefore, understanding how legal concepts can be translated into a form that can be automatically verified remains a challenge.

- The “right to be forgotten” requirement in GDPR can be difficult to realise, as user data may be fragmented across multiple services. How can cloud hosted services, which may involve invocation and interaction across a number of distributed platforms, ensure that this requirement can be achieved and verified?
- How can we consider the preference of users for verifying GDPR obligations (an essential requirement to ensure scalability of the approach)? This approach assumes that not all users care about privacy – or some users may have greater preference of privacy across a subset of their data.
- Increasing use of mobile devices and their integration with cloud platforms also poses scalability challenges for automated GDPR compliance checking. The transaction rate from devices can increase in frequency and complexity. If a blockchain based approach is to employed, the transaction rate of such a system needs to be scaled also.

The questions identified above describe key research challenges to assess GDPR compliance verification for cloud services. Based on this context we have proposed a compliance-aware multi-layered service stack – referred to as *compliance-aware cloud application engineering* (COM-PACE) for cloud services [10]. Fig. 1 shows the layered architecture comprising of: cloud providers, virtualization platform, compliance checking services & application layers. The compliance layer can be used to enforce whether privacy requirements are being supported for operations carried out within a conventional cloud architecture.

- *Compliance provisioning:* In this step, user privacy requirements to fulfill GDPR compliance requirements are identified. Thereafter, any hardware resources that can be used to instantiate trustworthy services (e.g. through the use of Trusted Platform Module (TPM) or other hardware-based components) are configured with software resources in a multi-stack web application environment.
- *Compliance monitoring:* A monitoring agent is activated to track and extract events that have GDPR compliance properties – and events are then forwarded to a blockchain network. These logs can be used to evaluate and audit GDPR regulations, and any possible violations that may have occurred. A monitoring manager coordinates between agents and the subsequent submission of log records to the blockchain network.
- *Compliance verification and enforcement:* GDPR relevant event logs submitted to the blockchain are verified for compliance based on the use of smart contracts. This verification helps to identify and disclose any possible violation(s) concerning unlawful disclosure, processing or transfer of personal data for purposes not agreed with the user. A violation alert is triggered and can be visualized providing transparent and trustworthy online services to end-users.

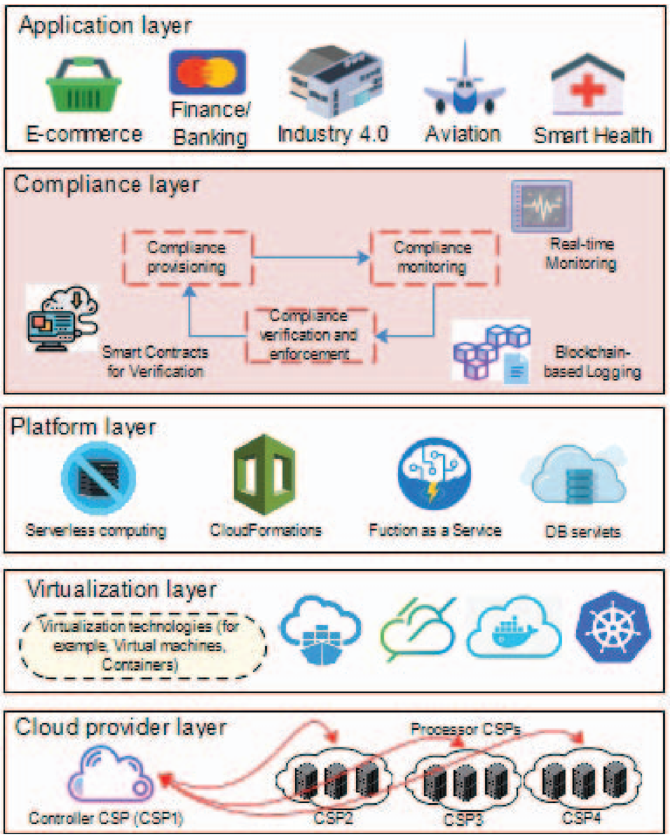


Fig. 1: Compliance-aware multi-layered service stack for Cloud services

A. Data Privacy Concerns under GDPR

Survey data consistently shows that people are concerned with how companies use their personal data. For example, in a survey published in August 2018 by the UK Information Commissioner’s Office, 53% of British adults said they were concerned about their ‘online activity being tracked’. Also, the European consumer protection organisation (BEUC – `https://www.beuc.eu/`) has reported that 70% of EU consumers are worried about how their data is being collected and processed. Similarly, in a study commissioned by IAB Europe, 11,000 people across the EU were asked about their attitudes towards online media and advertising, it was reported that only ‘20% would be happy for their data to be shared with third parties for advertising purposes.’ In the same vein, the 2019 Eurobarometer survey found that 30% of respondents who provide personal data online feel they have no control over it, and 51% stated feeling they have only partial control; of these respondents 62% claimed that they are concerned about this situation.

Concerns about data privacy are one of the main obstacles to greater use of on-line services. Acknowledging this reality, the GDPR was conceived and enacted to improve an individual's ability to allow users to control their data and give people 'efficient and operational means to make sure they are fully informed about what happens to their personal data'. Beyond Europe, a number of other legislations also describes similar requirements – such as the Personal Information Protection and Electronic Documents Act (PIPEDA), the California Consumer Privacy Act (CCPA), Australian Privacy Principles (APPs) and New Zealand's Privacy Act 1993 [11]. Consequently, the requirement for consent to be a valid ground for data processing were strengthened. In particular, consent must be given by 'a clear affirmative act establishing a freely given, specific, informed and unambiguous indication' of an individual's agreement to the processing of their personal data' – as described below:

Freely given and unambiguous: For consent to be freely given and informed, it must be a separate action from the activity the user is pursuing. Implicit or 'opt-out' consent — continuing to use a website without active objection to a notice — is not a clear positive action and consequently does not meet the requisite legal standard to legitimise the processing of personal data. Thus, pre-ticked boxes, which require a positive action to opt-out from, are an explicit example of invalid form of consent in the GDPR . Further, "[a] consent mechanism that emphasises 'agree' or 'allow' over 'reject' or 'block' represents a non-compliant approach, as the online service is influencing users towards the 'accept' option." Similarly, cookie boxes without a 'reject' option, or where it is located in a 'more information' section or on a third-party web page, are also non-compliant . To enable freedom of choice, both the accept and reject buttons must 'be presented on an equal footing'.

Specific and informed: The consent of the data subject must be given in relation to "one or more specific" purposes, and the data subject must have a real choice in relation to each of them. If the controller has conflated or bundled several purposes for processing and has not attempted to seek separate consent for each purpose, consent cannot be specific. Thus, when data processing is performed in pursuit of several purposes, the 'specific' criterion can be met on the basis of granularity, that is, the separation of these purposes, obtaining specific consent for each purpose. Accordingly, an 'accept all' button is only compliant if it is additional to the possibility of specifically consenting to each purpose.

Moreover, to fulfil the 'informed' criterion, information on the intended processing operations must be provided to data subjects in advance. The provision of this information enables data subjects to make informed decisions, understand what they are agreeing to and exercise their right to withdraw their consent. If the controller does not provide accessible information, user control becomes illusory and consent becomes an invalid basis for processing. The GDPR legislation sets out the information data controllers must provide to data subjects when processing their personal data . In particular, all 'recipients or categories of recipients' of personal data must be identified. If incomplete lists of recipients are provided – e.g. a list of potential vendors in the context of real-time bidding for advertising – this information will be insufficient to elicit informed consent . Since both options to 'accept' and 'reject' consent must be at the same level, approaches that force the user to navigate further to third party websites to reject tracking by specific vendors is non-compliant.

In accordance with the accountability principle, data controllers must be able to demonstrate that they comply with their data protection obligations, including that they have valid consent for each individual.

In practice, the above mentioned standards have failed to translate into substantial improvements to individual control. In order to secure broad access to personal data under the semblance of GDPR compliance, website owners are increasingly relying on 'dark patterns' – i.e. interface designs which seek to *nudge* users into desired privacy-intrusive choices through deceitful interaction flows. Examples of GDPR-non-compliant dark patterns are intrusive default settings, the concealment of privacy-friendly choices requiring extra effort from users to choose them, and take-it-or-leave-it options which bundle many data processing operations. Empirical research has shown that many dark patterns on the Web are common, and even widespread, such as the provision of a service on the basis of implicit consent (i.e. no consent asked), making the rejection of all tracking technologies substantially more difficult than accepting them, and buried pre-ticked boxes for optional vendors (e.g. third-party trackers) or purposes/ categories of data processing. Similarly, the fact that most websites rely on advertising as their main revenue source coupled with the complexity and multiplicity of actors involved in the advertising actor chain results in numerous instances where data processing by specific entities is not duly informed.

Importantly, consent is only one of the six legal grounds that legitimise the processing of personal data, and controllers typically rely on more than one, without specifying what basis justifies the processing of specific data for a specific purpose. This practice enables scenarios of unlawful processing and causes a level of ambiguity which undermines a data subjects' choices – e.g. a data subject may deny consent to the processing of their browsing behaviour for targeted advertising, but such data is nevertheless processed for that purpose on the basis of the controller's legitimate interests.

In the ongoing pandemic, which has forced us to increase our reliance on digital technologies to conduct our lives and endure restrictions, the above mentioned practices are exacerbated, making the notion of individual control all the more illusory. For example, imagine you want to have a meal in your local restaurant (when restrictions are relaxed). Upon entering the premises, you realise there is no 'traditional' customer service; rather, you need to download a booking and payment app., entering personal details to register. However, registration cannot be completed – and therefore you cannot be served – until you tick a box signalling consent to terms and conditions that allow for extensive processing of personal data based on several legal basis, for multiple purposes unrelated to the transaction you had in mind (i.e. having a simple meal). Any consent given as a result is invalid under the GDPR, as the consent request involved no real choice. Yet, without the competent Data Protection Authority (DPA) actually finding and penalising the breach, consent obtained in this way is *fair game*.

B. Apathy towards the GDPR: User Acceptance Survey

The practices above have translated into an overall apathy towards the GDPR and the mechanisms through which GDPR seeks to attain data usage agreement from an individual. For the study we utilised a cross-sectional research design to ensure the generalisability of the findings. We conducted a survey to measure the importance of privacy protection as afforded by the GDPR legislation. In total, 506 valid responses were collected using an independent research company (Prolific), which had distributed the survey online. To ensure compliance with research ethics, the responses were collected anonymously and following the consent of survey participants.

The survey contained two parts. The first part was aimed at collecting socio-demographic data about the cohort for the study. The respondents included 313 male (61.7%) and 195 female (38.3%) respondents, 82.6% were between 18 and 44 years old. 501 of the survey participants, accounting for 98.8% of the total sample, had 10 or more years of experience of using the Internet. More details of the demographic makeup of the respondents can be found in III in the Appendix.

The second part was focused at measuring an individual's familiarity with the GDPR, and the extent to which users perceived the GDPR legislation to be an important consideration for privacy protection. The study adopted a validated scale from prior literature [12] and tested the reliability of the adopted scale (table 1). Given that the awareness is a latent construct and cannot be measured directly, we used a multi-item scale to assess it. The awareness of GDPR was measured using a 7-point Likert style scale. The reliability of the scales and the descriptive analysis of responses were assessed using SPSS v27. The reliability (described in table 2) of GDPR awareness scale was satisfactory, the factor loadings are above 0.4 and Cronbach's alpha above 0.8 [13].

The results are summarised in tables 1-2 and figure 2. In terms of GDPR awareness, 320 respondents (representing 63.2% of the sample) are aware of GDPR (Table 2 and figure 1). While the awareness is high this does not indicate that they follow the rules and pay attention when companies breach them. It can be seen that 52.8% (267) of respondents are familiar with the GDPR, while 47.2% (239) of respondents score low on this scale. In terms of importance, the majority of individuals (286 respondents, 56.50% of the sample) believe that GDPR is not important.

The finding that more than 50% of respondents did not perceive GDPR to be relevant can be interpreted as a lack of objective knowledge of GDPR benefits – and also aligns with other work on user perception of the Computer Misuse Act [14]. Another, assumption can be that individuals do not trust the effectiveness of GDPR. This indicated that either users feel that the GDPR is not likely to be fully complied with by service providers or that the it is unlikely that the Regulation will lead to prosecution.

The apathy towards the privacy protection provided by the GDPR is problematic. If left unattended, it can culminate in a 'dysfunctional equilibrium', in which controllers realise that the level of individual control they offer is inconsequential for driving demand, as users expect that they have no control over their personal data to begin with. Thus, there is an urgent need for new approaches aimed at making consent an effective mechanism to signify and uphold our privacy choices, and more generally at facilitating GDPR compliance by online operators and law enforcement by Data Protection regulators.

III. GDPR COMPLIANCE CHECKING

In this section we describe a systems architecture that can be used to support GDPR compliance checking. This includes a container framework that can support automated GDPR compliance verification using smart contracts. The realization of our architecture contains a ratification phase that provides an agreement between a user (data subject) and a provider (data controller) before service delivery and any data usage. A sequence diagram representing the protocol of this phase is illustrated in Fig. 2. Through a smart contract, the purposes of data processing which include "*actor ID*", "*operation*", "*personal data*" and "*usage aim*" are sent into the blockchain by the data controller. The data subject is then provided with the deployment address of the contract, who is able to retrieve and observe the specified purpose of data usage from the blockchain and provide positive/ negative consent (the outcome of this decision is also stored in the blockchain).

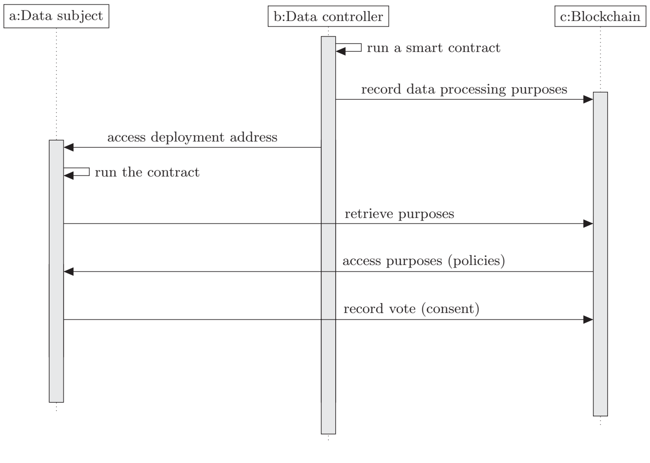


Fig. 2: A protocol for the ratification phase

Figure 3 shows the data flow between components supporting automated GDPR compliance verification. It presents all the actions that are taken in the back end of our proposed framework for tracking actors. First, users provide the cloud container with their personal data. A cloud provider then requests access to personal user data hosted within the container. Based on these data requests, the activity of providers on user

TABLE I: Awareness scale reliability

Items	loadings	Cronbach's alpha
I follow news and developments about GDPR rules	0.899	0.902
I discuss GDPR rules with friends and people around me	0.859	
I am interested in GDPR rules	0.865	
I read about GDPR rules on web sites and magazines	0.893	

TABLE II: GDPR Awareness, familiarity and importance

Category	Frequency-Low	Frequency-High
GDPR Awareness	186 (36.8%)	320 (63.2%)
Importance	286 (56.5%)	220 (43.5%)
Familiarity	239 (47.2%)	267 (52.8%)

data will be monitored and recorded by the container, and recorded operations are sent to a blockchain network for the purpose of verification. This step facilitates compliance with the accountability principle described in GDPR. Each record includes: an anonymised version of *provider IP*, the *operation* (e.g., read, write, etc.) executed by provider on personal data, and the *processed personal data items* by the provider. We note that the actual value associated with personal data fields is not submitted to the blockchain. After the execution of all operations on user data during service execution, a trusted third party, called a *verifier*, is able to run a transaction to retrieve the block contents and flag any observed GDPR violations in an automatic way. In particular, a smart contract is deployed to identify the providers who carried out operations on a user's data without getting her positive consent, or that executed a data processing operation in violation of the GDPR. An example of such violation would be the collection of personal data without the implementation of appropriate technical and organisational measures to ensure the security of such data, such as anonymisation and encryption of the data (Art.32(1) GDPR).

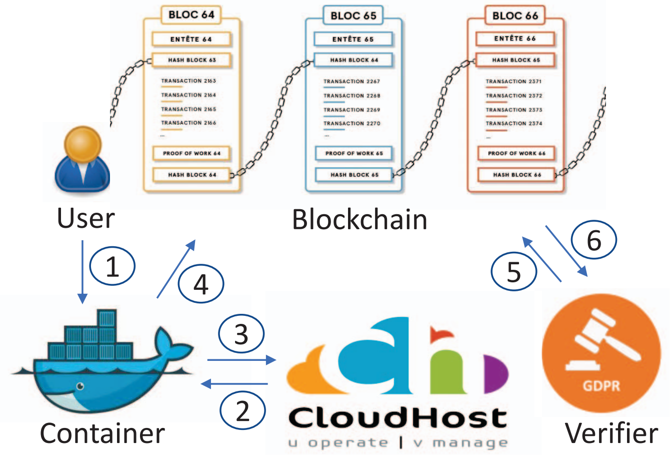


Fig. 3: Data flow in GDPR-compliance architecture

Based on the types of GDPR-relevant event being captured, we consider a number of agents that support this capture [10]. These agents are deployed alongside container-hosted services at each cloud service provider. The agents in charge of monitoring consider four key metrics related to GDPR compliance verification: read, write, transfer, and profile operations. A key challenge involves the selection of suitable compliance-related events to be monitored across the entire topology based on a trade-off between monitoring overhead, response time, scalability and compliance coverage. Figure 4 shows the data flow between different components that are used to realise our system. A client agent *GDPR agent* is used to submit an event record (read, write, transfer, profile) to a recording/ monitoring environment. Details of the interaction are extracted and added to the blockchain. Using encryption, we ensure that no personal data is visible to non-trusted users when the audit process is carried out.

A. *Prior Work*

In [10], [15] we have converted a number of GDPR operation requests into smart contract to verify their compliance by cloud providers in an automatic manner. However, contrary to machine-readable instructions that are concise, typically involving binary 'if/then' type of language and therefore rigid, legal rules tend to be 'open-textured', flexible and subject to interpretation. This is particularly the case of provisions in the GDPR, which feature terms like 'appropriate', 'reasonable', 'necessary', 'incompatible' or 'fairness', to name a few, that require highly-contextual interpretation and consequently human intervention. Thus, we have been able to translate into code only those rules that are strongly specified, whose violation can be directly detected through logging, and administrable – i.e. rules having low representational complexity and thus well-suited to be accurately represented in code. Given the significant volume of transactions that a cloud provider often needs to deal with, we believe this will significantly strengthen privacy in the operations of cloud providers.

Our focus in this work is therefore on those aspects of GDPR that apply to cloud service providers and which can be measured based on operations carried out on personal data of a user. Our approach also offers transparency and lawfulness, as the data recorded in the blockchain can be used as a basis for this. However, additional aspects of transparency are limited by the type of logging supported by the cloud provider. Data minimisation and the duration of storage, two additional requirements of GDPR, can also be measured through the blockchain submissions. Data integrity (i.e. the data is not modified by the provider) can be indirectly (to an extent) interpreted by the write operations log recorded in the blockchain. Similarly, the data protection smart contract verifies whether or not personal data records are encrypted, a technique that in and of itself ensures 'integrity and confidentiality'.

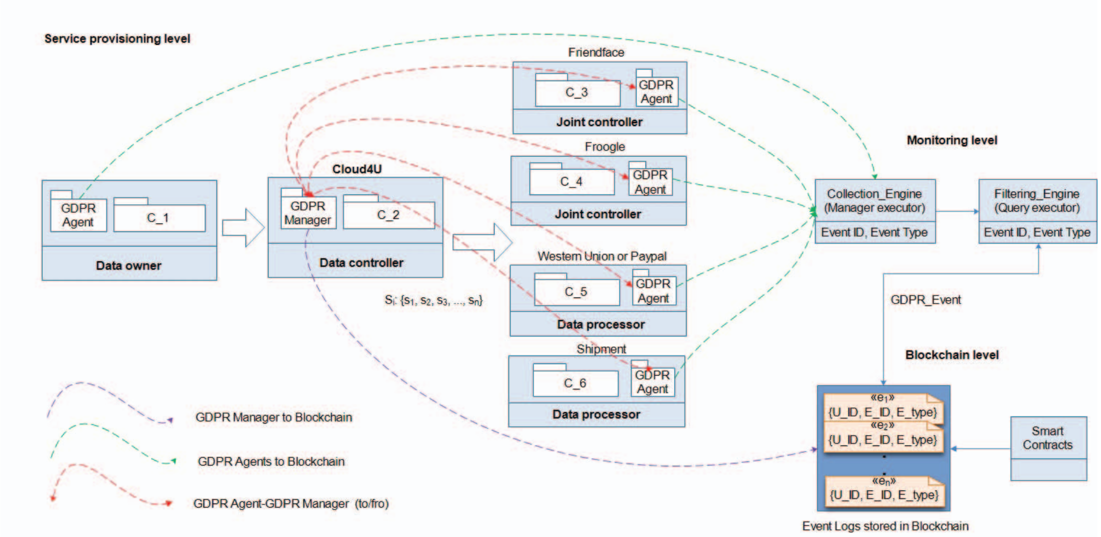

Service provisioning level
Friendface
C_3
GDPR Agent
Joint controller
Froogle
C_4
GDPR Agent
Joint controller
Cloud4U
GDPR Agent
C_1
Data owner
GDPR Manager
C_2
Data controller
S_i: {s_1, s_2, s_3, ..., s_n}
Western Union or Paypal
C_5
GDPR Agent
Data processor
Shipment
C_6
GDPR Agent
Data processor
Monitoring level
Collection_Engine (Manager executor)
Event ID, Event Type
Filtering_Engine (Query executor)
Event ID, Event Type
GDPR_Event
Blockchain level
«e_1» {U_ID, E_ID, E_type}
«e_2» {U_ID, E_ID, E_type}
«e_n» {U_ID, E_ID, E_type}
Smart Contracts
Event Logs stored in Blockchain
GDPR Manager to Blockchain
GDPR Agents to Blockchain
GDPR Agent-GDPR Manager (to/fro)

Fig. 4: Component interaction – showing how a GDPR agent can be deployed for each cloud service provider

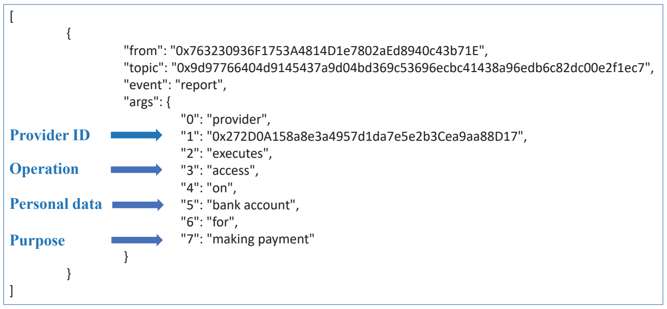
Existing efforts have focused on developing a semantic model (encodedb b with OWL/OWL2) for representing GDPR rules, using a policy language [16]. This can be used to express consent, business policies, and regulatory obligations, primarily as a step towards the automated compliance verification of GDPR obligations. The proposed policy language is also contextualised with reference to other formal representation of legal knowledge and reasoning, e.g. Legal RuleML [17]. Understanding how legal clauses can be mapped into automated rules remains a challenge however, and many of the approaches that currently exist are often conceptual in nature, leaving the more important consideration of automated mapping up to the user. Conversely, where GDPR compliance is attempted, the focus in on a questionnaire that needs to be completed by a human expert [18] – with limited support available to automate this process.

B. GDPR-Compliant Privacy Policy

As a GDPR requirement, a privacy policy should explicitly describe the purpose for personal data usage by actor(s). For example, in a cloud-based payment service, a policy can be expressed as “provider accesses bank account details for making an online payment”.The following definition determines a privacy policy:

Let $\mathcal{P}_S = \langle P, Pr, \mathcal{A}, D, \mathcal{D}_h \rangle$ be a composite cloud service, where P is a set of providers involving in the composite service, Pr is a set of purposes of data processing determined by P, A is a set of operations (e.g., access, store etc.) executed by P on personal data, D is a set of personal data that will be under processing by providers, and $\mathcal{D}_h \subseteq P \times \mathcal{A} \times D \times Pr$ be a data handling relation set that shows what operations will be executed by which providers (actors) on what personal data and for what purposes.

A privacy policy on $\mathcal{D}_h$ denoted by $Pol(\mathcal{D}_h)$ is a set of statements: “p_i executes α on d for pr”, for each $\langle p_i, \alpha, d, pr \rangle \in \mathcal{D}_h$, where $p_i \in P$, $\alpha \in \mathcal{A}$, $d \in D$, and $pr \in Pr$. It states that a privacy policy contains a number of statements on data processing operations, each of which must clearly expresses a data processing purpose. In case a provider processes personal data without determining a specific purpose, it is flagged as a *violator* of the purpose limitation principle (Art. 5(1)(b) GDPR) and potentially of Art. 6(1) when consent is the appropriate legal basis. Hence, if there is a handling relation $\gamma = \langle p_i, \alpha, d, pr \rangle \in \mathcal{D}_h$ such that $pol(\gamma) = \emptyset$, where $pol(\gamma) \in Pol(\mathcal{D}_h)$, there is a violation of the GDPR.



```
[
        {
                "from": "0x763230936F1753A4814D1e7802aEd8940c43b71E",
                "topic": "0x9d97766404d9145437a9d04bd369c53696ecbc41438a96edb6c82dc00e2f1ec7",
                "event": "report",
                "args": {
                        "0": "provider",
                        "1": "0x272D0A158a8e3a4957d1da7e5e2b3Cea9aa88D17",
                        "2": "executes",
                        "3": "access",
                        "4": "on",
                        "5": "bank account",
                        "6": "for",
                        "7": "making payment"
                }
        }
]
```

Fig. 5: A part of a block containing “purpose” of data processing

We have generated and recorded such privacy policy using smart contracts and blockchain [15], [19]. Our implementation improves the legibility of privacy policies in an electronic format. Figure 5 represents an instance of a created block that encompasses a blockchain ID for a provider executing access operation on bank account data to support a payment process. Based on such purposes retrieved from the blockchain, a user can give positive or negative consent (for each purpose).

IV. Improving User Engagement with GDPR

User disinterest in GDPR compliance that can be seen from the survey presented in Section 2.2 is directly related to practices by providers of digital services. Such service providers generally do not disclose particular operations that are to be carried out on user data, especially when it relates to data processing operations on personal user data.

User trust in a cloud provider can be improved by identifying these data processing operations and the actors involved in carrying out these operations. We believe this enhances transparency – as required by Art. 5(1)(a) of GDPR. This is achieved through the approach presented in Section 3.2. The informative value of privacy policies is enhanced by breaking down and explicitly identifying each purpose of data processing. In turn, if individuals can identify the purposes of the data processing operations, concerning specific items of personal data relating to them, they gain the ability to make informed decisions about their data privacy. Moreover, when the visibility of data processing purposes is combined with the ability to give or deny consent for each purpose, individual control - one of the objectives of GDPR as acknowledged in Recital 7 - is also attained.

Our architecture and implementation [20] enables messages exchanged between the client app. and the cloud-hosted service to also be explicitly identified. The use of a blockchain enables operations carried out on user data to be recorded and subsequently audited by an independent third party. Our previous efforts have focused on assessing scalability limitations in using a blockchain network [15], primarily by modelling the cost associated with carrying out such operations over a blockchain. However by limiting the type of operation being tracked (restricting this to read, write, transfer and profile) we can improve scalability of the associated implementation. Other approaches that use a multi-layered blockchain/ parachain, can also be used to improve scalability.

V. Conclusions & Future Work

GDPR remains an important requirement for many on-line services which utilise user data. GDPR compliance verification for a cloud provider is aimed to confirm that personal data provided by a user is shared based on the requirements of this legislation, so that any subsequent audit carried out on the provider does not lead to a financial penalty. This verification involves two aspects: (i) ensuring that user consent has been obtained – in line with Art. 6 of GDPR; (ii) sharing of data with external cloud providers is undertaken in a transparent way, so that the user is aware of which providers the information was shared with and for what purpose.

Our survey results show that users are ambivalent to GDPR benefits, and often are not fully conversant with the actions carried out by a cloud provider to achieve compliance with GDPR. With increasing take up of on-line services for a distributed workforce during Covid19, users often need to rely on cloud-hosted service to carry out their work. Data privacy needs have often been overlooked just to be able to access such on-line services. With increasing use of mobile devices, understanding how data privacy requirements can be extended on such devices also remains an important challenge.

The proposed work offers a transparent, blockchain-based auditing framework to ensure that GDPR compliance can be verified [15]. We believe this provides a better mechanism to improve take up and use of GDPR, and create better understanding of how cloud service providers handle and manage personal user data. The overhead of using a blockchain implementation to record transactions carried out on user data remains a challenge, addressed in this work by limiting the types operations that should be recorded for subsequent privacy audits.

Acknowledgement

This work has been funded by EPSRC project "PACE: Privacy-aware Cloud Ecosystems" under grants EP/R033439/1 and EP/R033293/1.

References

[1] S. A. Brown, V. Venkatesh, and H. Hoehle, "Technology adoption decisions in the household: A seven-model comparison," *Journal of the Association for Information Science and Technology*, vol. 66, no. 9, pp. 1933–1949, 2015.

[2] V. Venkatesh, J. Y. Thong, and X. Xu, "Unified theory of acceptance and use of technology: A synthesis and the road ahead," *Journal of the association for Information Systems*, vol. 17, no. 5, pp. 328–376, 2016.

[3] S. Papagiannidis and D. Marikyan, "Smart offices: A productivity and well-being perspective," *International Journal of Information Management*, vol. 51, p. 102027, 2020.

[4] P. A. Pavlou, "State of the information privacy literature: Where are we now and where should we go?" *MIS quarterly*, pp. 977–988, 2011.

[5] D. Marikyan, S. Papagiannidis, and E. Alamanos, """smart home sweet smart home": An examination of smart home acceptance," *International Journal of E-Business Research (IJEBR)*, vol. 17, no. 2, pp. 1–23, 2021.

[6] M. Nouwens, I. Liccardi, M. Veale, D. Karger, and L. Kagal, "Dark patterns after the GDPR: scraping consent pop-ups and demonstrating their influence," in *CHI '20: CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, April 25-30, 2020*, R. Bernhaupt, F. F. Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjøn, S. Zhao, B. P. Samson, and R. Kocielnik, Eds. ACM, 2020, pp. 1–13. [Online]. Available: https://doi.org/10.1145/3313831.3376321

[7] M. Barati, O. Rana, I. Petri, and G. Theodorakopoulos, "Gdpr compliance verification in internet of things," *IEEE Access*, vol. 8, pp. 119 697–119 709, 2020.

[8] M. Ulbricht and F. Pallas, "Yappl - A lightweight privacy preference language for legally sufficient and automated consent provision in iot scenarios," in *Data Privacy Management, Cryptocurrencies and Blockchain Technology - ESORICS 2018 International Workshops, DPM 2018 and CBT 2018, Barcelona, Spain, September 6-7, 2018, Proceedings*, ser. Lecture Notes in Computer Science, J. García-Alfaro, J. Herrera-Joancomartí, G. Livraga, and R. Rios, Eds., vol. 11025. Springer, 2018, pp. 329–344. [Online]. Available: https://doi.org/10.1007/978-3-030-00305-0_23

[9] E. Grünewald and F. Pallas, "TILT: A gdpr-aligned transparency information language and toolkit for practical privacy engineering," in *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021*, M. C. Elish, W. Isaac, and R. S. Zemel, Eds. ACM, 2021, pp. 636–646. [Online]. Available: https://doi.org/10.1145/3442188.3445925

[10] G. Singh Aujla, M. Barati, O. Rana, S. Dustdar, A. Noor, J. T. Llanos, M. Carr, D. Marikyan, S. Papagiannidis, and R. Ranjan, "Com-pace: Compliance-aware cloud application engineering using blockchain," *IEEE Internet Computing*, vol. 24, no. 5, pp. 45–53, 2020.

[11] A. Aljeraisy, M. Barati, O. Rana, and C. Perera, "Privacy laws and privacy by design schemes for the internet of things: A developer's perspective," *ACM Computing Surveys*, vol. 18, 2021.

[12] T. Dinev and Q. Hu, "The centrality of awareness in the formation of user behavioral intention toward protective information technologies," *Journal of the Association for Information Systems*, vol. 8, no. 7, p. 23, 2007.

[13] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, “Multivariate data analysis: Pearson new international edition,” *Essex: Pearson Education Limited*, vol. 1, p. 2, 2014.

[14] D. S. Wall, “Cybercrime and the culture of fear,” *Information, Communication & Society*, vol. 11, no. 6, pp. 861–884, 2008. [Online]. Available: https://doi.org/10.1080/13691180802007788

[15] M. Barati and O. Rana, “Tracking gdpr compliance in cloud-based service delivery,” *IEEE Transactions on Services Computing*, 2020.

[16] P. A. Bonatti, S. Kirrane, I. M. Petrova, and L. Sauro, “Machine understandable policies and GDPR compliance checking,” *Künstliche Intell.*, vol. 34, no. 3, pp. 303–315, 2020. [Online]. Available: https://doi.org/10.1007/s13218-020-00677-4

[17] T. Athan, G. Governatori, M. Palmirani, A. Paschke, and A. Z. Wyner, “Legalruleml: Design principles and foundations,” in *Reasoning Web. Web Logic Rules - 11th International Summer School 2015, Berlin, Germany, July 31 - August 4, 2015, Tutorial Lectures*, ser. Lecture Notes in Computer Science, W. Faber and A. Paschke, Eds., vol. 9203. Springer, 2015, pp. 151–188. [Online]. Available: https://doi.org/10.1007/978-3-319-21768-0_6

[18] S. Agarwal, S. Steyskal, F. Antunovic, and S. Kirrane, “Legislative compliance assessment: Framework, model and GDPR instantiation,” in *Privacy Technologies and Policy - 6th Annual Privacy Forum, APF 2018, Barcelona, Spain, June 13-14, 2018, Revised Selected Papers*, ser. Lecture Notes in Computer Science, M. Medina, A. Mitrakas, K. Rannenberg, E. Schweighofer, and N. Tsouroulas, Eds., vol. 11079. Springer, 2018, pp. 131–149. [Online]. Available: https://doi.org/10.1007/978-3-030-02547-2_8

[19] M. Barati, O. Rana, G. Theodorakopoulos, and P. Burnap, “Privacy-aware cloud ecosystems and gdpr compliance,” in *2019 7th International Conference on Future Internet of Things and Cloud (FiCloud)*. IEEE, 2019, pp. 117–124.

[20] M. Barati and O. Rana, “Privacy-Aware Cloud Ecosystems: Architecture and Performance,” *Concurrency and Computation: Practice and Experience*, 2020.

APPENDIX

TABLE III: Profile of respondents

Demographic Characteristics	**Type**	**Frequency (n = 506)**	**%**
Age	18 to 24 years	91	18
	25 to 34 years	164	32.4
	35 to 44 years	163	32.2
	45 to 54 years	49	9.7
	55 to 64 years	24	4.7
	65 or above	15	3
Gender	Male	313	61.7
	Female	195	38.3
Education	Completed some high school	122	24.1
	Completed some college (GSCE/AS/A-Level)	122	24.1
	Bachelor’s degree	183	36.1
	Master’s degree	64	12.6
	Ph.D.	11	2.2
	Other degree beyond a Master’s degree	4	0.8
Income	Less than £25,000	180	35.5
	£25,000 to £34,999	115	22.7
	£35,000 to £49,999	82	16.2
	£50,000 to £74,999	61	12
	£75,000 to £99,999	36	7.1
	£100,000 to £149,999	17	3.4
	£150,000 to £199,999	10	2
	£200,000 or more	5	1
Marital Status	Single (never married)	372	73.4
	Married or in civil partnership	128	25.2
	Separated	1	0.2
	Widowed	1	0.2
	Divorced	4	0.8
Internet Use – No. of Years	1-5 years	1	0.2
	5-10 years	4	0.8
	10-15 years	70	13.8
	15-20 years	214	42.2
	More than 20 years	217	42.8