Accepted to IEEE DAPPS 2020 (2nd IEEE International Conference on Decentralized Applications and Infrastructures); to be published by the IEEE in 2020; detailed citation information to be added.

A Data-Driven Analysis of Blockchain Systems' Public Online Communications on GDPR

Rahime Belen Sağlam*¶, Çağrı B. Aslan†*, Shujun Li‡¶, Lisa Dickson‡ and Ganna Pogrebna§
*Ankara Yildirim Beyazit University, Turkey
†STM Defense Technologies Engineering and Trade Inc., Turkey
‡University of Kent, UK
§University of Birmingham, UK
¶Corresponding co-authors: Rahime Belen Sağlam (rahimebelen@gmail.com), Shujun Li (http://www.hooklee.com/)

***Abstract*—After the European Union's new General Data Protection Regulation (GDPR) became applicable in May 2018, concerns about the legal compliance of public blockchain systems with rights guaranteed by GDPR have emerged, e.g., on the "right to be forgotten". In order to better understand how the blockchain sector sees the challenges raised by GDPR and how such their communications could influence their users, this paper reports our data-driven analysis of GDPR-related public online communications of blockchain developers and service providers. Our analysis covers 314 public blockchain systems, and two different online communication channels: legal documents including privacy policies, T&C (Terms and Conditions) documents and other similar legal documents published on systems' official websites and public tweets of their official Twitter accounts.**

Our analysis revealed that only a minority (86/314 $\approx$ 27.5%) of the investigated blockchain systems had covered GDPR at least once using one or both communication channels. Among the 86 systems, only 27 systems (8.6%) had at least one legal document that actually talks about GDPR for the corresponding blockchain system. We noticed a systematic lack of detail about why and how the GDPR compliance issue was addressed, and most systems made questionable statements about GDPR compliance. The results are surprising considering that the GDPR was enacted in 2016 and has been in effect since May 2018.

***Index Terms*—GDPR, blockchain, distributed ledger, data protection, law, privacy, communication, transparency**

I. Introduction

Technically, a blockchain is a distributed ledger that contains all the transactions that are shared among participating parties, basically computers, that are maintained by a distributed consensus algorithm [1]. It exists on a P2P network where there is not a single central body to manage the ledger, but every participating node keeps a copy of the ledger. This concept is called distributed or decentralized trust, and transactions are validated and authorized by the consensus of participating nodes in a distributed network. In addition to being a distributed secure data storage (enforced by cryptographic hashes, digital signatures, and verification of multiple agents), new-generation blockchain system now also provide computer programs a mechanism to self-execute a set of simple instructions among different parties, via a new technology called smart contracts that are intended to facilitate, verify or enforce a contract [2]. Depending on what participating nodes can access data on chain, blockchain systems can be split into three types: public (permissionless) – anyone can access, consortium (permissioned) – only authorized parties can access, and private – only a centralized party can access. Some people do not consider private blockchains as real blockchains, and some only see public blockchains as real blockchains. Most permissioned blockchains are developed in the context of business-to-business applications, based on public versions with added access control policies.

The European Union's General Data Protection Regulation (more commonly referred to by its acronym GDPR) of 2016 is a new and far reaching regulation concerning data protection, which became enforceable across the whole European Union (EU) in May 2018 [3]. As the EU's most recent attempt to addressing data protection issues, the GDPR aims to protect the privacy of any data subject in the EU (not just EU citizens), regardless of the location of their data, and any personal data that are collected or processed in the EU (Article 3). In order to achieve its set goals, the GDPR provides an enforceable legal framework of rights for data subjects, whose data are collected and processed, and corresponding enforceable obligations being placed on data controllers and processors, to ensure that data is processed only within a set of stated principles (Article 5). Data controllers and data processors are defined within the Regulation (Article 4) and extend to natural or legal persons, public authorities, agencies or other bodies. The new legal framework introduced by the GDPR represents significant steps forward in protection from the EU's Data Protection Directive (DPD) of 1995 that it replaces.

The GDPR has created new challenges to the development and operation of blockchain systems because of some potential conflicts between a number of key principles in the GDPR and the technical nature of the blockchain technology, particularly between the data subject's "Right to Erasure ('right to be forgotten')" defined in Article 17 of the GDPR and the data immutability feature of blockchain systems. Such potential conflicts are echoed in a recent report from the EU Blockchain Observatory & Forum [4], which says "Public, permissionless blockchains represent the greatest challenges in terms of GDPR compliance". Considering the fact that a number of other countries (e.g., USA [5] and China [6]) are creating their

The paper's supplementary material is available at http://www.hooklee.com/Papers/DAPPS2020_suppl.pdf.

new GDPR-like data protection regulations and more countries may follow up in future, the problem will soon become a more global issue beyond the current territorial scope of the GDPR.

Although the GDPR compliance issue has been clearly identified in the EU Blockchain Observatory & Forum report [4] for public blockchain systems, we observed a general lack of direct communications from public blockchain developers and service providers to their users. This motivated us to conduct a comprehensive data-driven analysis on GDPR-related public online communications made by developers and service providers of 314 public blockchain systems with a cryptocurrency with a market size greater then $10 million at the time of our study. We looked at two different online communication channels for each system: legal documents including privacy policies, T&C documents and other legal documents on official websites, and public tweets from official Twitter accounts. Our analysis revealed some surprising results: only 86/314 $\approx$ 27.5% of the investigated blockchain systems explicitly talked about the GDPR, and many made questionable statements about GDPR compliance. Among the systems that communicated about the GDPR, there was a systematic lack of detail about why and how GDPR compliance has been or will be achieved. On a positive side, a very smaller number of (6) systems clearly admitted that some data subjects' rights (notably the right to erasure/be forgotten) cannot be respected due to the technical nature of the blockchain technology.

The results are surprising particularly considering the fact that the GDPR was enacted nearly three years ago and has been in effect for around a year. Given the wide discussion around the GDPR and the fact that all public blockchain systems will unavoidably fall into the rather wide territorial scope of the GDPR (Article 3), it is hard to believe that most blockchain developers and service providers were unaware of the relevance of the GDPR compliance issue to their systems. Our work therefore calls for more urgent research into topics such as legal aspects of the public blockchain technology, development of more legally sound technical solutions, human users' and organizations' perception and behaviors, as well as new data protection laws that are more "future-proof".

The rest of the paper is organized as follows. The next section discusses useful background information about blockchain and related work regarding the relationships between blockchain and GDPR, which will help the readers understand the research problem and the results we will discuss later. Section III explains how we collected and processed the data we used, with basic statistics of the data. Detailed observations from our data-driven analysis are reported in Section IV, and further discussions are given in Section V. Limitations of the study and future work are discussed in Section VI, and the related works are given in Section VII. The paper is concluded by the last section.

II. Background: Blockchain vs GDPR

In order to assess compatibility of blockchain systems with the GDPR, let us have a look at different types of personal data that may be stored on blockchains. The first class of personal data on blockchain systems are transaction data. Transactions are not limited to transfer of cryptocurrencies between pseudonymous individuals. Depending on the underlying application, a transaction can cover personal data such as financial or medical information relating directly or indirectly to individuals depending on the use case. Independent of the application, nothing prevents a malicious user to upload personal data of other people to a blockchain. In a blockchain system, transaction data can appear in three forms: plain, encrypted, or hashed, the latter two being considered "pseudoanonymised" but still requiring a reduced level of protection. The second set of data stored in blockchains that may qualify as personal data is metadata, which is necessary to coordinate individuals without centralized intermediaries. Particularly, public keys (i.e., addresses on blockchain) are essential metadata used for validating transactions, which have been recognized by the European Union Blockchain Observatory & Forum as valid personal data [4]. Another examples are IP addresses that are recorded in some blockchain systems.

To understand the implications of blockchain-based systems from a privacy perspective, it is important to understand some key technical characteristics of data on blockchains. The first one is data immutability, i.e., data cannot be deleted or changed once added. To be more precise, a blockchain is a series of data blocks, which are sequentially linked (chained) together through a cryptographic hashing process, where every block contains its own hash as well as the hash of the previous block for verification and sequencing.[1] For a hash function, any differences in input data will produce different output data. This means that, in blockchain systems, an attempt to change existing data will cause the hash of the corresponding data block to no longer match the hash value included in the next block, thereby breaking the chain. This characteristic of immutability has been discussed as one of the main concerns about the GDPR compliance of blockchain systems [9]–[12]. It would seem that the necessary characteristic of immutability is incompatible with the stipulated right of data subjects to require rectification and/or erasure of data under the "Right to Erasure ('right to be forgotten')" in Article 17 of the GDPR. In other words, the technical nature of blockchain seems to contradict the "right to be forgotten" as the blockchain technology implies that changing or omitting existing data (such as deleting records) will compromise the underlying trust principles of blockchain. One recommended solution to this problem is to put personal data off chain and only their hashes on chain [13], and another is to put encrypted data on chain and the key off chain (which correspond to the weaker protection called "pseudonymisation" in the GDPR) [14].

Another key feature of the public blockchain technology

[1]There are other more advanced distributed ledger technologies based on a graph-based model (e.g., DAG [7] and hashgraph [8]) rather than a linear blockchain, but the data immutability feature remains largely unchanged. This will not significantly influence our discussion in this paper, so we will ignore such level of technical details and use the term "blockchain" as an umbrella term for all distributed ledger systems.

that gives rise to questions of GDPR compliance is that it is public and permissionless, meaning that anyone may, without authorization, participate in the network as a node. Such public distributed ledgers, together with a distributed consensus process, do not require any centralized authority as a manager, thus leading to a large number of parties who do not trust and may not even be able to identify each other[2]. This potential lack of attribution as well as inherent anonymity of public distributed ledgers challenge the GDPR compliance of blockchain systems. Millard et al. discussed this feature, considering the data protection by design and default principle elucidated in the GDPR (Article 25), and raised some questions that need to be addressed [15]. One of the questions is that this openness leads to a confusion in identifying data controllers and data processors. Article 25 of the GDPR places legal obligations on the data controllers to ensure that "appropriate technical and organizational measures, such as pseudonymisation" are implemented "both at the time of the determination of the means for processing and at the time of the processing itself". In this context, we can ask whether each participating node of a public blockchain system holding a copy of the distributed ledger should be seen a data controller. In addition, the necessary replication of data on each node also gives rise to a contradiction to the principle of "data minimisation" stated in the GDPR [16], [17]. This is manifested in Article 5(1)(c) of the Regulation, which requires that the amount of personal data collected must be "limited to what is necessary" to achieve purposes for which the data processed. It is further reflected in the purpose limitation that requires that personal data be "collected for specified, explicit and legitimate purposes and not further processed" in Article 5(1)(b) of the GDPR.

In the GDPR, the data controller is defined as the individual or entity who decides the purposes and means of the processing (Article 4). For any systems that have the potential to retain personal data, the identification of the data controller is important in order to determine where these responsibilities and obligations reside. On public blockchain systems, even if the system does not store any data that can be definitely labeled as personal data, the public-key addresses used by such systems as pseudonyms of participating nodes can still be considered as personal data, as stated explicitly in the EU Blockchain Observatory & Forum report (on Pages 19-20) [4]. This implies that it may be impossible for any public blockchain system to argue that it does not store any personal data so the GDPR does not apply. Note that in some situations the data controller is obliged to conduct a mandatory DPIA (data protection impact assessment) that requires records of all processing activities be made and retained (Article 35), which would require clarifying the whole data storage and processing chain including responsibilities and obligations of other data processors involved.

[2]Several countries, e.g., China, are currently making attempts to regulate this space to require blockchain users register with state-issued IDs (see, e.g., https://www.theverge.com/2018/10/22/18008640/china-blockchain-registration-government-id for more information).

A further question raised by Millard et al. in [15] concerns the relationship between controllers and processors; more specifically, how controllers can instruct processors on the processing of personal data given the anonymous nature of public blockchains. In the literature, some researchers proposed the use of smart contracts to control data access, usage, and transfer to data processors. For instance, Neisse et al. proposed a solution based on smart contracts, which are deployed by data subjects for each data controller or data processor [18]. This solution allows data subjects (users) to track how their data are processed by data controllers and data processors and whether the processing of their data are compliant with their *ex ante* consent. In [19] Sousa proposed a conceptual model where each data subject's consent is stored in the data controller's back-end component, which enables the regulator to traverse the blockchain whenever a consent requires verification. In their two recent studies [20], [21], Loukil et al. proposed a similar approach that converts privacy policies describing the data subject's privacy preferences into custom smart contracts.

In the above discussion, we have mentioned user consent. The GDPR defines a set of obligations for data controllers and processors, which include obtaining explicit consent from the data subject for the processing of any personal data. Explicit consent means freely given, specific, informed and unambiguous indication of the data subject's preferences about the processing of personal data relating to him or her. In order to have a lawful basis for processing activity of personal data by a controller, the data subject must have given consent for the processing to occur for one or more specific purposes explicitly. Article 7 of the GDPR sets out a framework for consent, providing three fundamental principles or rules; controllers are responsible for demonstrating consent was given, a data subject has the right to withdraw consent at any time, and finally written requests for consent must be clear. The regulation also makes it clear that "it shall be as easy to withdraw as to give consent". This statement of the importance of consent is further strengthened by Article 22, which notes that the data subject has the right not be subjected to automated decision making unless this kind of processing is based on the data subjects' explicit consent. It is made clear that a lack of explicit consent requires the controller/processor to stop all automated processing of the data.

In a public blockchain, once a transaction has been made on the blockchain, the same set of data will be processed by the all nodes in the chain. Consequently, gaining explicit consent is essential at the beginning before the download or execution of the blockchain software. In the literature, there are studies emphasizing that each executed transaction needs to include a statement of consent to be acceptable by data subjects [22], which can be hard to manage for most public blockchain systems. A proposed solution to this problem is to use smart contracts to automatically handle consent management [18], [20], [21], but smart contracts are normally based on a public blockchain so the solution can be seen circular reasoning.

Yet another important aspect is the territorial scope. In a

public blockchain, anyone can run a node by downloading the transaction history of a blockchain disregarding the territorial scope of the laws that regulate the protection of personal data on chain. However, the GDPR defines a wide territorial scope (Article 3), which says that the GDPR applies for processing of personal data of any data subjects (not just EU citizens) by data controllers and processors in the EU, or personal data of any data subjects in the EU by any data controllers and processors for the purposes of providing goods or services or behavior monitoring [23], [24]. The (pseudo-)anonymous nature of public blockchains means that it is very hard to manage the territorial scope, and the only feasible approach is to assume that the GDPR always applies in all situations.

III. Data collection and processing

In this section we explain how we selected the public blockchain systems and how we collected public online communications made by developers and service providers behind those selected systems.

A. Selection of public blockchain systems

Since the first public blockchain system Bitcoin [25] appeared in 2008, many blockchain systems have emerged in domains such as supply chain, health care, Internet of Things, and governmental services [26]. Most public blockchain systems are associated with one or more cryptocurrencies, which are used as an effective mechanism for incentivizing people to participate in maintaining the blockchain [1].

In order to conduct our data-driven analysis of public online communications of blockchain developers and service providers, we first needed to decide what public blockchain systems to choose. Due to the rapid development of new blockchain systems, there was not a well maintained list of such systems with the needed indicators for us to consider. We therefore decided to use associated cryptocurrencies with a large market capitalization size as a proxy to "reverse engine" public blockchain systems that are popular among blockchain users, who are the people our study will benefit. This method should have missed some public blockchain systems that do not have an associated cryptocurrency on the market yet, but we noticed that such systems had been reasonably rare or less mature (e.g., under development) so missing them should not significantly skew our results. The use of cryptocurrencies also naturally allowed us to avoid most permissioned and private blockchain systems, which rarely use cryptocurrencies because they do not normally have the need to incentivize participants (who are normally organizations co-running the system rather than individuals attracted to an existing system).

To decide the market capitalization size of cryptocurrencies, we used CoinMarketCap (https://coinmarketcap.com/), a popular website maintaining a large list of cryptocurrencies with their market capitalization figures. We used a snapshot of the CoinMarketCap list captured on 17 April 2019 (08:42:19 UK time) to all cryptocurrencies with a market capitalization size greater than $10 million. This led to 320 active cryptocurrencies (see Section 1 of the paper's supplementary material for a full list), which correspond to 314 valid public blockchain systems. In this paper we use the term "system" to refer to the following three different cases: 1) a single cryptocurrency appearing on a single dedicated website (which is normally managed by a single company or a group of developers); 2) more than one cryptocurrency appearing on the same website and managed by the same company or a group of developers; 3) a single cryptocurrency co-developed/maintained by more than one company/group of developer (e.g., a company and a foundation, or a number of collaborating organizations). Some systems following the above definition may actually maintained by the same company or the same group of core developers[3], but we did not attempt to consider this due to the complexity of obtaining such information (which is not always in public domain).

B. Online communication channels

After investigating the selected public blockchain systems' websites and other online activities, we identified two main channels that blockchain developers and service providers normally make GDPR-related public online communications:

1) legal documents including privacy policies, T&C (Terms and Conditions) documents and other legal documents published on official websites;
2) public tweets posted by official Twitter accounts.

There are other online communications we also considered (e.g., white papers, posts on blogs, web forums and chat rooms) and did some preliminary analysis, but plan to cover them more thoroughly in future (see Section VI for more).

C. Data collection and statistics

All legal documents were collected on 23 May 2019, almost exactly one year after the GDPR became effective.

Among the 314 blockchain systems, 189 ones provided links to "Privacy Policies" or "Terms and Conditions" (T&C) or other legal documents on their official websites. The most common legal documents provided by the systems are privacy policies where 169 systems provided privacy policies, 128 ones provided T&C documents and 34 other legal documents. In order to obtain relevant Twitter data, we identified 310 official Twitter accounts and downloaded their most recent tweets (up to 3,200 tweets for each system due to the limitation of Twitter's API).

In order to eliminate documents that do not mention the GDPR at all, we followed a simple approach by searching for the keyword "GDPR" or "General Data Protection Regulation" in each document we collected. Although being simple, we consider this approach valid as any sensible discussions on the GDPR should include at least one mention of the word "GDPR" or its full title. Some basic statistics of results of the keyword searches are summarized in Table I. Note that one system did not have any relevant legal documents or a Twitter account, so the total number of relevant systems is 313.

[3]For instance, a company maintains two different blockchain systems associated with two different cryptocurrencies and sets up separate websites for them as well.

TABLE I
BASIC STATISTICS OF COLLECTED DOCUMENTS MENTIONING GDPR (BASED ON NUMBERS OF SYSTEMS, NOT DOCUMENTS)

Data Source	GDPR Mentioned (%)	Total
Privacy Policies	48 (28.4%)	169
T&C Documents	9 (7.0%)	128
Other	2 (5.9%)	34
Twitter Accounts	43 (13.7%)	313
Any Channel	86 (27.5%)	313

Among all the legal documents we used in our study, privacy policies proved to be the one covering the most relevant information about the GDPR. However, even for this channel the ratio of privacy policies that include our two query terms is still quite low. Figure 1 shows more detailed statistics regarding how the 314 systems are split into different groups for each communication channel.

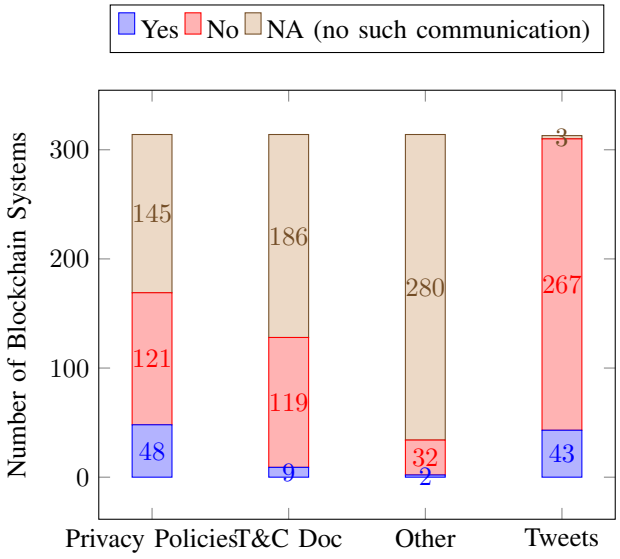

Fig. 1. Statistics of GDPR mentions in different types of online public communications by blockchain systems

After the data collection and cleaning process, we manually inspected all legal documents that mention the GDPR. We noticed that some legal documents were written on personal data collected and processed by the websites rather than on data on blockchains. We eliminated legal documents clearly written for website visitors only since the aim of the study is to assess the GDPR compliance of the blockchain software systems. This led to legal documents of 27 blockchain systems for further analysis. Within the 27 systems, only 6 ones[4] have legal documents explicitly stating what data are for website and what for the blockchain software systems. For other systems, we made an assumption that the legal documents do cover data on blockchain as well, and reported our findings in Sections IV and V based on this assumption. If this assumption is wrong for some systems, then we will only over-estimate how well public blockchain systems communicated about the GDPR compliance issue, which should be better than under-estimate it since our findings are mostly negative (i.e., public blockchain systems did not communicate enough about the GDPR compliance issue). We did a similar process on tweets collected. We eliminated retweets and general expressions about the GDPR, which do not give any useful information about the GDPR compliance issue of the system concerned. We identified 18 tweets that are directly written to announce GDPR compliance of the systems mainly posted in May and June 2018 which are around the date that the GDPR came into effect. However, there are also tweets that underline the challenges and question the personal data on the chain.

[4]Selected texts from privacy policies of the six "good" systems can be found in Section 3 of the paper's supplementary material.

IV. OBSERVATIONS

A. Legal documents

In our study, we focused on legal documents published by the systems including privacy policies, T&C documents and other legal documents that are all written in English. It has been observed that rights of data subjects provided by the GDPR are mainly covered in privacy policies where only a few T&C documents mention some rights and none of the legal documents provide any related information. For those T&C documents mentioning some rights, the statements are also covered in privacy policies, so we decided to focus on privacy policies for the further analysis.

Privacy policies are pervasive feature of websites and applications required in many countries with the aim of informing users about how information about them is gathered and processed. This allows users to make a more informed decision on accepting or abstaining from using the website/application. However, it is a known fact that, in practice, privacy policies, which make up the largest portion of our dataset, are often not read by users, hard to understand, and do not support rational decision making [27]. Consequently, previous studies on analyzing privacy policies mainly focused on manually assessing their usability in means of their accessibility, writing quality, content and evolution over time [28]. There are also manual assessments that focus on compliance with self-regulatory requirements, e.g., Cranor et al. evaluated privacy policies of online tracking companies against self-regulatory guidelines with the aim of understanding collection and use of sensitive information and linkage of tracking data with personally-identifiable information [29]. Similarly, in our study, we aimed to evaluate the privacy policies of public blockchain systems against the GDPR and assessing whether they contain sufficient and clear information relevant for users to make privacy decisions considering theirs rights provided by the GDPR.

There are past studies to interpret policies for users through natural language processing (NLP) tools [30], [31] and crowd sourcing [32], [33] as well. However, these efforts will succeed only if privacy policies contain relevant information. In this context, we aimed to evaluate privacy policies of blockchain systems and to check existence of relevant information about

data subjects' rights provided in GDPR, that can provide valuable insights into future studies.

We identified several factors that make blockchain systems' privacy policies very questionable or incomplete in terms of their coverage for data subject's rights. First of all, except for a few ones privacy policies do not make it clear if they cover the website itself and/or the related blockchain applications, which leads confusion about the scope of privacy policies. Second and perhaps surprisingly, only six of the 27 systems provided information about their applications disclosing the possible problem to exercise some rights due to the underlying blockchain technology. However, even among those six systems, each right is not covered in an equally transparent way, e.g., the immutability issue is discussed more than other rights. Only one policy covers all rights in a transparent way.

In the following, we give our observations on how the investigated blockchain systems' legal documents communicated in a number of key areas of the GDPR. The observations are based on a categorical encoding scheme that indicates how each document addresses each area, which is explained in detail in Section 4 of the paper's supplementary material.

1) Explicit consent: Ten of the 27 privacy policies do not provide clear information about how the corresponding blockchain systems obtain consent from their users. There are vague statements such as "Our processing of your Personal Information is based on the consent where you have consented to our use of your Personal Information". However, there is no further explanations about how they obtain it. Four of the policies state that users will consent processing of their personal data when they use their Website or the applications. There are two policies that state that consent is gathered via use of applications. Eight policies provide better explanations and state that consent is obtained from the user as they register to their platforms, submit information to their systems or use their services. Those policies ask people to actively opt in and try to meet the standard of an unambiguous indication by clear affirmative action in Article 4. However, better policies are given in two other systems where it is required to fulfil a contract with the user to obtain consent. In those privacy policies, blockchain systems and the technologies under these systems are briefly introduced to the users and the policies for the website services and application are written separately to avoid possible confusion. The consent is said to be obtained by the active submission of the wallet address or after the fulfillment of the contract between user and the system. One policy does not provide any explanation about this right.

A further interesting issue arising in this area concerns a data subject's right to withdraw consent. Nine policies acknowledge this right explicitly and require the data subject to send an email to request withdrawal of consent. This compares with four policies that do not specifically mention the data subject's ability to withdraw consent. Further 13 policies claim to support the right without giving any information as to how a user can do that. Considering the nature of blockchain systems and the obvious challenges to allowing a subject the ability to exercise the right to withdraw consent to personal data, it is very surprising that only one policy explicitly states that the immutability of blockchain systems may affect the users' ability to exercise some rights. This included the ability to object to, or restrict, the processing of their personal data. Another important finding is that even though there are several conceptual models based on smart contracts proposed in the literature, with the aim of access control, none of the blockchain systems offer solutions for controlling data access including usage, and transfer to data processors.

2) Right to erasure (right to be forgotten): Considering the immutable nature of blockchain, the right to erasure (right to be forgotten) is the most challenging data subject's right for the blockchain community. However, perhaps surprisingly, the majority of the privacy policies (21 out of 27 ones) do recognize the issues and discussions about the right to erasure but they make claims that personal data can be deleted upon request by the data subject. Two policies include discussions on immutable nature of the blockchain and states the impossibility of erasure whereas two propose anonymization as an alternative solution. One policy claims to cover no personal data on the blockchain except Wallet IDs on the blockchain, which are argued as non-personal data.

3) Transparency and portability: Among 27 blockchain service providers, for 13 systems transparency is limited to personal data collected solely by the website services, which leads to some ambiguity about the scope of the policies. Nine of the policies covered personal data collected by their applications and only four of the systems are transparent about keeping the personal data on blockchain. One of those systems argued that all the data on the blockchain is pseudo-anonymized so that the data on chain does not reveal any personal data. One of the systems did not provide any information about personal data collected by their systems.

The right to data portability assures the possibility of transferring data "from one electronic processing system to and into another, without being prevented from doing so by the controller" upon the request of a data subject. Data subjects have the right to request their personal data in a common and easily readable computer format, which is a relatively easier task compared to other rights such as the right to erasure. Among the 27 systems, only three of them explicitly claimed to transfer personal information directly to another "controller", where technically feasible. Nonetheless, 10 of such systems included just the keyword "data portability" in their policies, claiming that they provide this right to their users, however they did not elaborate it. The remaining 14 policies do not provide any information about this right.

4) Data retention: The GDPR states that principles of "data minimisation" require that data is retained only as long as it serves a necessary purpose, which requires service providers to give information about how long they are going to store personal information and the conditions regarding this procedure. Majority (16) of the blockchain systems' privacy policies cover vague statements about this right, stating that the personal information will be erased if further storage is not required or permitted by applicable laws. Seven systems are

transparent about the exact conditions or the periods of their personal data storage practices. Four of them do not specify any information about this right.

5) Transfer of personal data to third countries or international organizations: The GDPR requires the data controllers to provide information to data subjects about the transfer of their personal data to third parties, third countries or international organizations. Considering its distributed nature, for a blockchain system it is inevitable to transfer personal data to third parties. During this step of our analysis, we focused on how service providers had explained the transfer of personal data related to distributed ledger technology. Surprisingly, only one system's privacy policy explicitly states that interacting with their blockchain systems can lead any personal data written on the blockchain to be transfered and stored across the globe due to its global decentralized public network. Except one, 22 systems' policies state that that they may share personal data with third-party services including law enforcement, government officials and regulators as well as other service providers that they use to support their businesses. Four systems' policies stated that they will or do not share personal information with third parties beyond the scopes such as the European Economic Area.

6) Data minimization: "Data minimisation" (British English spelling in the GDPR) is the principle of storing and processing personal data that is relevant, adequate and limited to what is necessary. Two main features of the blockchain clearly conflict with this principle; storing a copy of all data on every node and immutability. However none of the privacy policy we examined, including those from the systems that are transparent about the immutability nature of the blockchain systems, has statements underlining this conflict. 22 systems' policies do not explicitly cover this principle and five of them provide only very brief explanations such as "Processing of your data is carried out by the principle of data minimization, accuracy and limited data storage."

7) Right of access: The right of access, which has been reported to be entirely compatible with the blockchain technology [34], gives data subjects the right to access their personal data held by any service provider subject to compliance with the GDPR. Among privacy policies of the 27 systems, four of them give detailed information informing the users about their rights not only to access but also to know whether they process their personal data and certain information how they use it and who they share it with. Two systems' policies provide relatively limited information and recognize the right to access personal data limited to the information about how and why they use it. 13 systems' policies provide brief information stating that the users have the right to access their data and/or a copy of their data in a machine-treatable format. Six of the systems' policies provide very brief information and just cover the name of this right in the list of rights that are recognized. Finally, two systems' policies do not provide any related information about this right.

8) Data protection by design and by default: The GDPR explicitly mentions the "Data protection by design and by default" principle in Article 25, which is a data protection law version of the more widely known "privacy by design and by default" principle. In order to analyze discussions around this principle, we focused on explicit mentions of the related terms "privacy by design", "data protection by design", "data protection by default" and other forms of those words (e.g., "privacy-by-design"), which led to only one legal document: the privacy policy at https://casinocoin.org/privacy-policy/. This policy only states that "We will need to update this Privacy Policy from time to time in order to make sure it stays current with the latest legal requirements and changes to our privacy by design practices." This does not give any concrete information about their practices on privacy by design.

B. Twitter accounts

When we evaluated the tweets posted by the 43 blockchain systems' official accounts that mention GDPR in at least one of their tweets, we identified 16 accounts that had announced GDPR compliance of their systems. Their tweets announcing the compliance were mainly posted in May, June or July 2018 shortly after the the date GDPR became enforceable. However, it was surprising to notice that 11 of those systems did not have a legal document, which made it impossible to validate their claims. This finding is valuable for the potential social media analysis on the GDPR compliance issue. Other than the tweets that announced GDPR compliance, we identified six tweets belonging to six different accounts that had claimed their blockchain systems to be GDPR compliant without giving a link to explain why. There are six accounts that had shared negative opinions about the GDPR or the GDPR compatibility of blockchain systems underlining regulatory challenges.

V. More discussions

Our analysis drew a largely negative picture of how public blockchain systems were communicating about the GDPR compliance issue. Our study does not give answers to why the public blockchain community behaved the way we observed, which will be our future work. However, we can reasonably speculate that either there was an insufficient level of awareness on GDPR or that the sector was aware of this issue but simply unwilling to make the sensitive matter transparent to their users and the general public (maybe out of fear of the resulting legal punishment or loss of business opportunities).

While most privacy policies are very vague in how and why the GDPR compliance is (or not) achieved, we noticed a number of systems whose privacy policies are particularly transparent. For instance, GNOSIS's privacy policy (https://gnosis.io/privacy-policy) warned its users as follows:

> "Accordingly, by design, a blockchains records cannot be changed or deleted and is said to be 'immutable'. This may affect your ability to exercise your rights such as your right to erasure ('right to be forgotten'), or your rights to object or restrict processing, of your personal data. Data on the blockchain cannot be erased and cannot be changed. Although smart contracts may be used to revoke certain access rights, and some content may be made invisible to others, it is not deleted.
> ...

IF YOU WANT TO ENSURE YOUR PRIVACY RIGHTS ARE NOT AFFECTED IN ANY WAY, YOU SHOULD NOT TRANSACT ON BLOCKCHAINS AS CERTAIN RIGHTS MAY NOT BE FULLY AVAILABLE OR EXERCISABLE BY YOU OR US DUE TO THE TECHNOLOGICAL INFRASTRUCTURE OF THE BLOCKCHAIN.IN PARTICULAR THE BLOCKCHAIN IS AVAILABLE TO THE PUBLIC AND ANY PERSONAL DATA SHARED ON THE BLOCKCHAIN WILL BECOME PUBLICLY AVAILABLE"

The same policy covers the following statements for "Right to erasure (right to be 'forgotten')":

"HOWEVER, WHEN INTERACTING WITH THE BLOCKCHAIN WE MAY NOT BE ABLE TO ENSURE THAT YOUR PERSONAL DATA IS DELETED. THIS IS BECAUSE THE BLOCKCHAIN IS A PUBLIC DECENTRALIZED NETWORK AND BLOCKCHAIN TECHNOLOGY DOES NOT GENERALLY ALLOW FOR DATA TO BE DELETED AND YOUR RIGHT TO ERASURE MAY NOT BE ABLE TO BE FULLY ENFORCED. IN THESE CIRCUMSTANCES WE WILL ONLY BE ABLE TO ENSURE THAT ALL PERSONAL DATA THAT IS HELD BY US IS PERMANENTLY DELETED."

Five more examples of such transparent privacy policies can be found in Section 3 of the paper's supplementary material.

The explicit statements about the challenges in GDPR compliance due to the nature of blockchain technologies cast more doubts on other privacy policies' claims about the GDPR compliance. It is questionable to assure that those potentially false claims are due to misunderstanding of GDPR or only written for ordinary websites' services without any tendency to hide details or misdirection. On the other hand, it is also possible to argue that those systems are capable of handling GDPR by storing personal data on central databases other than blockchains by-passing any possible challenge introduced by the nature of the blockchain technology. Since a few number of public communications we have covered in our study provides details about data storage practices within the systems, those possibilities remain open for discussion.

The lack of transparency about GDPR compliance issues of blockchain systems could be related to the ongoing development phase of such systems. Since some systems have not produced a working software system yet, it is possible for the developers to procrastinate those issues. However, due to the data protection by design principle it is necessary to consider privacy and data protection principles right from the start.

VI. Limitations and Future Work

Although we worked on a comprehensive dataset, there are limitations of the work that require further studies.

First of all, although our analysis work was assisted by automated tools, it involved a lot of human efforts and expert knowledge of people involved (the first three co-authors, and a number of helpers – see the acknowledgement section), so human errors and biased judgments are inevitable. We adopted a protocol to have information of each cryptocurrency checked by at least two independent human encoders, so we believe the error rate is low so any errors that may still remain should not significantly influence the results reported in this paper.

Second, the results we present offer just a snapshot of public online communications of the studied blockchain systems at a particular time. Even though we have tried to obtain a representative data set, it is not an easy task to make this claim considering the variety of blockchain systems. We have also limited our in-depth analysis to the ones that we retrieve by a naive keyword search and we eliminated the ones that do not cover our keyword 'GDPR'. It is reasonable to assume the existence of this keyword in any GDPR related discussions, however, enriching the query terms with some important terms in GDPR such as the names of different data protection principles have a potential to enrich the data set.

In the data-driven analysis reported in this paper, we considered legal documents and public tweets of official accounts. There are other important public online communications we did not cover due to the complexity of collecting such information including end user licence agreements (EULAs), official blogs and emails sent from blockchain systems to their users that were sent out mostly around 25 May 2018, the date when GDPR became effective. For Twitter, there are likely more accounts we should consider in addition to the official one, e.g., accounts of core developers and senior managers of the underlying company.

As a result, the results discussed in this paper give only a partial picture of how blockchain developers and service providers talked about the GDPR compliance issue publicly online. In our future work, we plan to look into some other types of public communications listed below and conduct a longitudinal analysis of public communications of blockchain systems. We expect that more blockchain systems will be captured by this source of information.

The first type of public communications we will consider are end user licence agreements (EULAs) of blockchain software, which may mention privacy and GDPR. EULAs are often shipped as part of software and it is not straightforward to automate the collection of such information. Considering most blockchain systems have their software code released on popular repositories such as GitHub, we plan to develop a dedicated crawler to automatically search, download, analyze and extract EULAs. We also envisage that some software systems may have to be manually inspected if their EULAs do not follow a standard pattern. We also noticed some blockchain systems published their EULAs on their websites, so we will also try to check their websites for collecting EULAs.

Another important source of public communications on GDPR are emails sent from blockchain systems' developers and service providers to their users. Such emails are normally not publicly available and were sent out mostly around 25 May 2018, the date when GDPR became effective. We plan to run a crowdsourcing based campaign to solicit such emails from as many blockchain systems as possible, in order to gain insights about how blockchain developers and service providers communicated to their end users regarding GDPR. We also plan to conduct a survey to gather end users' perception on

GDPR and the GDPR email(s) they received.

For public communications on official websites, we looked at privacy policies and white papers only. Some blockchain systems may have explained how the GDPR compliance issue is addressed on other web pages such as news and press releases. We therefore also plan to run a more in-depth analysis of all web pages on selected blockchain systems' websites to see if such additional communications can be found.

Yet another potentially important online communication channel is official blogs. Most systems we studied have official blogs, but we decided not to include blog articles because of the human efforts required to analyze such articles.

Our data-driven analysis focused on documents on the studied blockchain systems' websites, but such documents may not be a true reflection of the actual awareness and understanding of people who are running the systems, i.e., blockchain developers and managers of the organizations involved, on the GDPR compliance issue. For instance, the online documents may be prepared by lawyers and tweets written by marketing officers, without a proper discussion with technical people or senior managers. In our future work, we will explore the possibility of interviewing some blockchain developers and key management people of organizations behind selected blockchain systems to get their self-reported views on the GDPR compliance issue, how their systems actually store personal data, and on how they decided (not) to communicate about this issue to their users. This can be extended to empirical studies investigating how different blockchain developers and service providers manage privacy related risks in general. We expect that such empirical studies will be welcomed by some (if not all) blockchain developers and service providers since they will benefit from having a better understanding of the important legal issue.

In addition to the above planned future work, yet another interesting direction for research is to study how different demographic factors (country, culture background, gender, age, size of business, etc.) influence the attitude and decision of blockchain developers and service providers in terms of GDPR compliance. Most public blockchain systems have published their core team members on their official website or in the system's white paper, so such demographic information can be collected. Some statistical factorial analysis can be conducted to identify factors that have a main effect.

VII. Related work

The GDPR compliance issue has received increasing attention from both industry and academia, and there is growing focus on how applications can be evolved to eliminate the risk of being declared incompatible. However, due to its relative recency, studies that conduct data-driven analysis about legal compliance issues around GDPR are rare.

To the best of our knowledge, this is the first study that aims to collect and analyse public communications specific to blockchain developers and service providers. However, there are some related studies that focus on GDPR awareness for business. One of the first attempts was performed in 2014 in Finland, right before the GDPR lifespan, to understand awareness of companies (the controllers of the personal data) and their willingness to act towards compliance regarding GDPR [35]. The study showed that he general level of awareness was low and only 43% of data controllers were aware of GDPR. It was also reported that only 31% of controllers were planning to act towards compliance during the time of the study.

Another related study 'Cyber Security Breaches Survey' was published in 2018, which is a survey on business and charity actions regarding cyber security and the costs and impacts of cyber breaches and attacks in the UK [36]. It covers in-depth interviews undertaken in January and February 2018 to follow up with organizations that participated in the survey, as well as higher education institutions. It reported that 38% of businesses and 44% of charities were aware of GDPR at the time of the study. On top of this, 13% of businesses and 9% of charities had amended their cyber security policies or processes specifically in preparation for GDPR.

In [37], Sirur et al. reported the results of interviews made with several organizations with the aim of gaining in-depth understanding of real challenges faced by organizations in engaging with the GDPR. By conducting the interviews, they found that large organizations had felt compliance was reasonable and doable. However, the compliance attempts of general small-to-medium organizations (SMEs) were of a lower maturity than that of large companies and data protection focused SMEs. A similar conclusion was reported by Archibald and Renaud in [38].

VIII. Conclusion

We have conducted a study based on a large database of the public communications made by developers and service providers of 314 blockchain systems including legal documents (such as privacy policies and T&C documents) on their websites and public tweets made from official Twitter accounts. We found out that most systems had not communicated about GDPR. As a general trend, those legal documents lacked an explicit acknowledgment and warnings to users on the legal challenges introduced by the nature of the blockchain technology. In addition, those documents also mix terms for data on blockchains and data collected for other purposes (e.g., for websites). Only a very small number (6) of blockchain systems explicitly explained issues around data on blockchains. The status of such public communications and disclosure of legal issues and privacy risks on users is not satisfactory, and we call for urgent more research into the interfaces between data protection law and the blockchain technology.

Acknowledgment

The authors would like to thank Nikhil Patnaik, Yang Lu and Benjamin Nogué for helping collect some raw data about the cryptocurrencies we studied.

Shujun Li's work was partly supported by the research projects ACCEPT (https://accept.cyber.kent.ac.uk/) and PriVELT (https://privelt.ac.uk/), funded by the EPSRC (Engineer-

ing and Physical Sciences Research Council) in the UK, under grant numbers EP/R033749/1 and EP/P011896/2, respectively.

REFERENCES

[1] A. Narayanan, J. Bonneau, E. W. Felten, A. Miller, S. Goldfeder, and J. Clark, *Bitcoin and Cryptocurrency Technologies*. Princeton University Press, 2016.

[2] N. Szabo, "Smart contracts: building blocks for digital markets," *Extropy: Journal of Transhumanist Thought*, vol. 18, no. 16, 1996.

[3] European Parliament, "Regulation (EU) (2016) 2016/679 of the European Parliament and of the Council of 27 April 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 59(L 119), 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj

[4] European Union Blockchain Observatory & Forum, "Blockchain and the GDPR," thematic report, 2018. [Online]. Available: https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf

[5] J. De Groot, "What is the California Data Privacy Protection Act?" Digital Guardian's Blog, 2018. [Online]. Available: https://digitalguardian.com/blog/what-california-data-privacy-protection-act

[6] China Daily, "Personal information protection law," news report, 2019. [Online]. Available: http://www.chinadaily.com.cn/a/201903/06/WS5c7f06cba3106c65c34ecf67.html

[7] C. Bai, "State-of-the-art and future trends of blockchain based on DAG structure," in *Structured Object-Oriented Formal Language and Method: 8th International Workshop, SOFL+MSVL 2018, Gold Coast, QLD, Australia, November 16, 2018, Revised Selected Papers*. Springer, 2018, pp. 183–196.

[8] L. Baird, M. Harmon, and P. Madsen, "Hedera: A governing council & public Hashgraph network," White Paper 1.0, 2018. [Online]. Available: http://hedera-hashgraph.s3.amazonaws.com/hh-whitepaper-v1.0-180313-2.pdf

[9] R. Teigland, H. Holmberg, and A. Felländer, "The importance of trust in a digital europe: Reflections on the sharing economy and blockchains," in *Trust in the European Union in Challenging Times*. Springer, 2019, pp. 181–209.

[10] S. Castell, "The future decisions of RoboJudge HHJ Arthur Ian Blockchain: Dread, delight or derision?" *Computer Law & Security Review*, vol. 34, no. 4, pp. 739–753, 2018.

[11] D. Townend, "Conclusion: Harmonisation in genomic and health data sharing for research: An impossible dream?" *Human Genetics*, vol. 137, no. 8, pp. 1–8, 2018.

[12] C. Bartolini and L. Robaldo, "PrOnto: Privacy ontology for legal reasoning," in *Electronic Government and the Information Systems Perspective: 7th International Conference, EGOVIS 2018, Regensburg, Germany, September 3–5, 2018, Proceedings*. Springer, 2018, p. 139.

[13] C. Compert, M. Luinetti, and B. Portier, "Blockchain and GDPR: How blockchain could address five areas associated with GDPR compliance," IBM Security White Paper, 2018. [Online]. Available: https://www.ibm.com/downloads/cas/2EXR2XYP

[14] G. Maldoff, "Top 10 operational impacts of the GDPR: Part 8 - pseudonymization," IAPP (International Association of Privacy Professionals) Westin Research Center online resource, 2016. [Online]. Available: https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-8-pseudonymization/

[15] C. Millard, "Blockchain and law: Incompatible codes?" *Computer Law & Security Review*, vol. 34, no. 4, pp. 843–846, 2018.

[16] M. Berberich and M. Steiner, "Blockchain technology and the GDPR – how to reconcile privacy and distributed ledgers," *European Data Protection Law Review*, vol. 2, no. 3, pp. 422–426, 2016.

[17] N. Fabiano, "Internet of Things and the legal issues related to the data protection law according to the new European General Data Protection Regulation," *Athens Journal of Law*, vol. 3, no. 3, pp. 201–214, 2017.

[18] R. Neisse, G. Steri, and I. Nai-Fovino, "A blockchain-based approach for data accountability and provenance tracking," in *Proceedings of ARES 2017*. ACM, 2017.

[19] H. R. de Sousa and A. Pinto, "On the feasibility of blockchain for online surveys with reputation and informed consent support," in *Ambient Intelligence – Software and Applications –, 9th International Symposium on Ambient Intelligence*. Springer, 2018, pp. 314–322.

[20] F. Loukil, C. Ghedira-Guegan, K. Boukadi, and A. N. Benharkat, "Towards an end-to-end IoT data privacy-preserving framework using blockchain technology," in *Web Information Systems Engineering – WISE 2018: 19th International Conference, Dubai, United Arab Emirates, November 12-15, 2018, Proceedings, Part I*. Springer, 2018, pp. 68–78.

[21] ——, "Semantic IoT gateway: Towards automated generation of privacy-preserving smart contracts in the Internet of Things," in *On the Move to Meaningful Internet Systems. OTM 2018 Conferences: Confederated International Conferences: CoopIS, C&TC, and ODBASE 2018, Valletta, Malta, October 22-26, 2018, Proceedings, Part I*. Springer, 2018, pp. 207–225.

[22] M. Walther, "EU General Data Protection Regulation and distributed ledgers (blockchain)," 2018. [Online]. Available: https://www.researchgate.net/publication/325069696_The_EU_GDPR_and_Distributed_Ledgers_Blockchain_Solutions_to_a_Worst_Case_Scenario

[23] P. de Hert and M. Czerniawski, "Expanding the European data protection scope beyond territory: Article 3 of the General Data Protection Regulation in its wider context," *International Data Privacy Law*, vol. 6, no. 3, pp. 230–243, 2016.

[24] European Data Protection Board, "Guidelines 3/2018 on the territorial scope of the GDPR (Article 3) - version for public consultation," 2018. [Online]. Available: https://edpb.europa.eu/our-work-tools/public-consultations/2018/guidelines-32018-territorial-scope-gdpr-article-3_en

[25] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008. [Online]. Available: http://www.bitcoin.org/bitcoin.pdf

[26] B. Rodrigues, T. Bocek, and B. Stiller, "The use of blockchains: application-driven analysis of applicability," in *Advances in Computers*, 2018, vol. 111, pp. 163–198.

[27] A. M. McDonald and L. F. Cranor, "The cost of reading privacy policies," *I/S: A Journal of Law and Policy for the Information Society*, vol. 4, no. 3, pp. 543–568, 2008.

[28] C. Jensen and C. Potts, "Privacy policies as decision-making tools: An evaluation of online privacy notices," in *Proceedings of CHI 2004*. ACM, 2004, pp. 471–478.

[29] L. F. Cranor, C. Hoke, P. G. Leon, and A. Au, "Are they worth reading: An in-depth analysis of online trackers' privacy policies," *I/S: A Journal of Law and Policy for the Information Society*, vol. 11, no. 2, pp. 325–404, 2015.

[30] S. Zimmeck and S. M. Bellovin, "Privee: An architecture for automatically analyzing web privacy policies," in *Proceedings of 23rd USENIX Security Symposium*. USENIX Association, 2014, pp. 1–16.

[31] S. Zimmeck, Z. Wang, L. Zou, R. Iyengar, B. Liu, F. Schaub, S. Wilson, N. Sadeh, S. Bellovin, and J. Reidenberg, "Automated analysis of privacy requirements for mobile apps," in *Proceedings of NDSS 2017*. Internet Society, 2017.

[32] S. Wilson, F. Schaub, R. Ramanath, N. Sadeh, F. Liu, N. A. Smith, and F. Liu, "Crowdsourcing annotations for websites' privacy policies: Can it really work?" in *Proceedings of WWW 2016*. ACM, 2016, pp. 133–143.

[33] S. Wilson, F. Schaub, F. Liu, K. M. Sathyendra, D. Smullen, S. Zimmeck, R. Ramanath, P. Story, F. Liu, N. Sadeh, and N. A. Smith, "Analyzing privacy policies at scale: From crowdsourcing to automated annotations," *ACM Transactions on the Web*, vol. 13, 2018.

[34] F. Martin-Bariteau, "Blockchain and the European Union General Data Protection Regulation: The CNIL's perspective," Blckchn.ca Working Paper Series, 2018.

[35] T. Mikkonen, "Perceptions of controllers on EU data protection reform: A Finnish perspective," *Computer Law & Security Review*, vol. 30, no. 2, pp. 190–195, 2014.

[36] K. Finnerty, H. Motha, J. Shah, Y. White, M. Button, and V. Wang, "Cyber security breaches survey 2018," Governmental report from the Department for Digital, Culture, Media & Sport, UK, 2018. [Online]. Available: https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2018

[37] S. Sirur, J. R. Nurse, and H. Webb, "Are we there yet?: Understanding the challenges faced in complying with the General Data Protection Regulation (GDPR)," in *Proceedings of 2nd International Workshop on Multimedia Privacy and Security*. ACM, 2018, pp. 88–95.

[38] J. Archibald and K. Renaud, "POINTER: A GDPR-compliant framework for human pentesting (for SMEs)," in *Proceedings of HAISA 2018*. University of Plymouth, 2018, pp. 147–157.