Research Article **Open Access**

Md Mehedi Hassan Onik, Chul-Soo Kim, Nam-Yong Lee, and Jinhong Yang*

Privacy-aware blockchain for personal data sharing and tracking

https://doi.org/10.1515/comp-2019-0005
Received November 30, 2018; accepted March 4, 2019

Abstract: Secure data distribution is critical for data accountability. Surveillance caused privacy breaching incidents have already questioned existing personal data collection techniques. Organizations assemble a huge amount of personally identifiable information (PII) for data-driven market analysis and prediction. However, the limitation of data tracking tools restricts the detection of exact data breaching points. Blockchain technology, an 'immutable' distributed ledger, can be leveraged to establish a transparent data auditing platform. However, Art. 42 and Art. 25 of general data protection regulation (GDPR) demands 'right to forget' and 'right to erase' of personal information, which goes against the immutability of blockchain technology. This paper proposes a GDPR complied decentralized and trusted PII sharing and tracking scheme. Proposed blockchain based personally identifiable information management system (BcPIIMS) demonstrates data movement among GDPR entities (user, controller and processor). Considering GDPR limitations, BcPIIMS used off-the-chain data storing architecture. A prototype was created to validate the proposed architecture using multichain. The use of off-the-chain storage reduces individual block size. Additionally, private blockchain also limits personal data leaking by collecting fast approval from restricted peers. This study presents personal data sharing, deleting, modifying and tracking features to verify the privacy of proposed blockchain based personally identifiable information management system.

Keywords: general data protection regulation, GDPR, personally identifiable information, PII, privacy policy, distributed ledger, off-chain data, identity management, personal information tracking

1 Introduction

For providing user-centric services, websites gather a noticeable amount of personally identifiable information (PII) (e.g. age, race, social security numbers, house location, driving license etc.). Out of 4.021 billion internet user, over 2.77 billion people use several social networking sites (SNS) and have made available a vast amount of personal information [1]. These SNS sites and mobile applications offer sign-in or registration option for premium services. Moreover, PII is often used by organizations to authenticate a customer's identity. As most of these SNS sites and applications are for free, they earn money from personal data trading. Actually, these organizations store, distribute, analyse sensitive PII to generate a business model through user profiling. Tech giants also use third-party service providing enterprises to mine customer data. Those subsidiary organizations also collect, analyze and distribute data from other organizations too. Eventually, users have no clue where their data is going. We all are reaping advantages of the data-driven industry, but the dark side is illicit use of those personal data. Guardian [2] revealed in April 2018 that, Cambridge analytica breached 87 million personal information from the largest SNS site Facebook. Zou [3] listed top data breaching organizations of the 21st century are Yahoo (3 billion), eBay (145 million), Adobe (38 million), JP Morgan (76 million), US office of personnel management (22 million). Gemalto's breach level index (BLI) [4] reported, out of 10. 4 million yearly PII leak, 74% were identity stealing. Kumar [5] stated data as the new currency in trade marketing. Kumar [5] reported that around 200 billion USD is being invested every year for PII exchange. Business to business communication will be even extensive in the era of industry 4.0 which will lead us to think personal data management and tracking. Indirect identifiers (quasi-identifiers) also known as potential personally identifiable information (PPII) are immensely

Md Mehedi Hassan Onik: Department of Computer Engineering, Inje University, Gimhae 50834, Korea; E-mail: hassan@oasis.inje.ac.kr
Chul-Soo Kim: Department of Computer Engineering, Inje University, Gimhae 50834, Korea; E-mail: charles@inje.ac.kr
Nam-Yong Lee: Department of Applied Mathematics, Inje University, Gimhae 50834, Korea; E-mail: nylee@inje.ac.kr

***Corresponding Author: Jinhong Yang:** Department of Healthcare and IT, Inje University, Gimhae 50834, Korea;
E-mail: jinhong@inje.ac.kr

Open Access. © 2019 Md Mehedi Hassan Onik et al., published by De Gruyter. This work is licensed under the Creative Commons Attribution alone 4.0 License

used to generate PII. Sweeney [6] showed that combination of gender, birth dates or postal codes can reveal 87% identities of the USA citizen. False data injection attack in healthcare system has been a widely discussed topic in data security domain [7].

Blockchain technology has gained much attention from researchers to use beyond cryptocurrencies [8–11]. The blockchain is constructed as a sequence of blocks, which can hold any data in its block like a conventional public ledger. These blocks are linked and secured together using cryptography. Although few researchers have already used blockchain in PII management [12–16], conflict with data privacy terms put a question mark on them. Similarly, off-the-chain data storing style for blockchain adaptation to personal data managing is already in practice. However, this study shows the circulation of personal data through key GDPR components. So far, our understanding, no other study elaborated how the controller and processor accumulate sensitive user data. However, the motivation of recently executed GDPR [17] is to protect individual information therefore, institutions must pay special attention in public data sharing. Consent must obtain before any private data is being analyzed, there is also accountability to confirm that those data can be withdrawn or deleted independently (also known as 'the right to be forgotten'). Blockchains PII storing architecture is based on 'immutability' of the data. On the contrary, GDPR [17] demands, user request personal information modification and deletion.

An overview of existing data sharing architecture versus improved architecture proposed by this study is shown in Figure 1. It shows how a user can only track the first level data collector in legacy data tracking system. Oppositely, this study proposes a multi-layer data sharing and tracking architecture to track data destination. Proposed study stores the identification of the controller and processor with the help of blockchain technology. This study proposes a blockchain based personally identifiable information management system (BcPIIMS) designed for PII management and tracking throughout organizations. A separate storing of personal and non-personal data is proposed to generate a transparent, immutable scheme that complies with the GDPR.

In summary, this study makes the following significant contributions:

1. This study stores personal data (PII and PPII) off-the-chain and non-personally identifiable information (NPII) information in blockchain to satisfy Art. 42 and Art. 25 of GDPR [17].
2. The proposed system successfully tracks personal information movement among stakeholders. Easy tracking of data leaking sources improves personal data privacy. This study shows, only privacy improvement can add additional security to personal data.

1.1 Roadmap of the study

Section 2 elaborates related studies to introduce personal information privacy, blockchain technology and GDPR. Section 3 discusses the proposed BcPIIMS. This section elaborates data sharing, management and tracking mechanisms too. In section 4, prototype testing using multichain 2.0 (alpha 3) is mentioned. Section 5 evaluates proposed BcPIIMS model from a qualitative perspective. Future goals and limitations are discussed in conclusions followed by necessary references.

2 Related works

This section discusses personal information types and associated privacy risks. We also discuss blockchain technology and its components. GDPR terms and conflicts with blockchain are revealed too.

2.1 Personal data privacy

Solutions and regulations are already in use to secure personal information. To protect data privacy, leading countries and companies have also started their individual regulations. Software level proprietary verification that works on OAuth protocol was proposed by [18]. To reduce personal data leaking issues, [19] suggested, either to reduce PII share or to track PII flow. Weingärtner [19] practiced a data mining technique to identify the risk factors of PII leaking. Kshetri [20] identified three main factors of end-user privacy leaking (a) higher use of cloud storage (b) gathering of more information than actually needed and, (c) excessive data analysis and distribution. Several regulations like GDPR [17], privacy protection amendment by Australian government [21], Canada's personal information protection act [22] and ISO27001 by 'International Organization for Standardization' (ISO) [23] are currently practised for data protection and management. In privacy, several studies have identified digital identities as a vibrant source of user identification. Pfitzmann [24] defined PII as "*any subset of attribute values of an individual person which sufficiently identifies this individual person within any set of persons*". National institute of standards and technology (NIST) [25] defined PII as "*any in-*

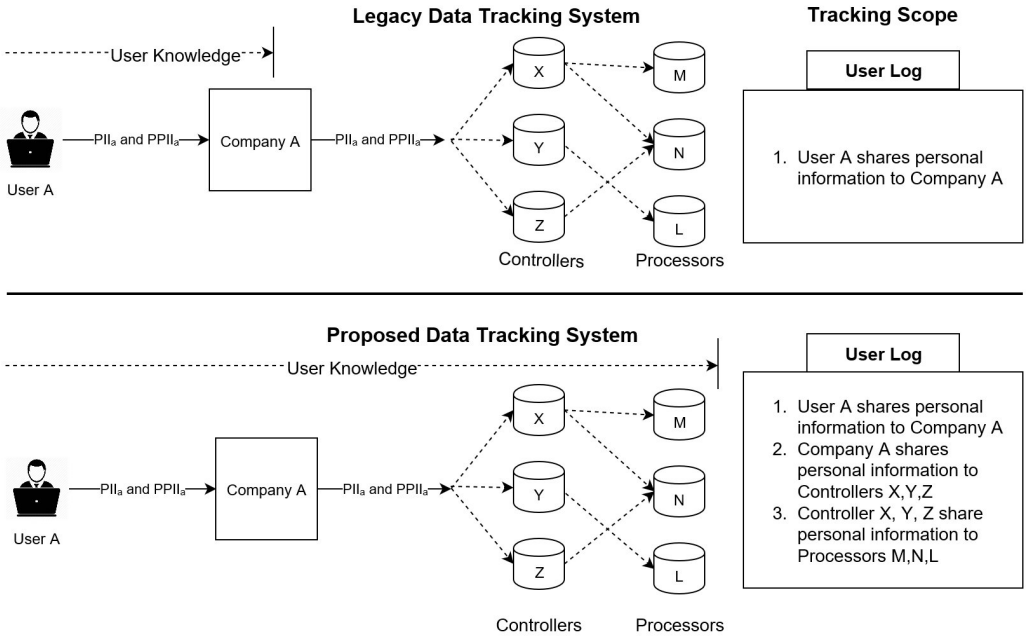

Figure 1: User data sharing and tracking architecture (legacy vs proposed).

formation about an individual maintained by an agency, including (PII) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (linked PII or PPII) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information". Few personally identifiable information (PII) are listed in Table 1.

Similar to PII, indirect identifiers are known as quasi-identifiers or potential personally identifiable information (PPII). Pfitzmann [24] defines PPII as "*a subset of attribute values of a complete identity, where a complete identity is the union of all attribute values of all identities of this person*". Few examples of potential personally identifiable information are listed in Table **??**.

Personal information leaking is risky to both organizations and individuals [26]. To diminish this risk, privacy by design was proposed by [27]. Posey [28] worked on PII breach classification from text data mining and identifies 8 major PII leaking. On the contrary, information that is unable to detect any identity individually or in a group is non personally identifiable information (NPII). NPII are shareable and risk-free information that carries no risk to user privacy. Australian privacy law and practice [29] defined NPII as "*non-identifiable data, which have never been labelled with individual identifiers or from which identifiers have been permanently removed, and by means of which no specific individual can be identified. A subset of non-identifiable data are those that can be linked with other data so it can be known that they are about the same data subject, although the person's identity remains unknown.*"

Different monitoring applications were also proposed by researchers [30–32]. Although most of them ended up increasing security instead of privacy. We will now discuss a few privacy-enhancing tools. Farzaneh [33] introduced "Data Track", a tracking tool that allows data visualization before sharing. The study allowed users to edit the data scope before sharing to service providers. Although the study provides a tool for tracking data movement, the processing and storing of information are still vulnerable. Similarly, studies [30, 34, 35] also focused on usability oriented privacy enhancing tool. They proposed a mobile health data tracking tool (mhealth) to increase the transparency in fitness data sharing. Another study by [30] mentioned the privacy of things (PoT) privacy awareness tool. It allows data privacy tracking of the internet of things (IoT) apparatus. However, the study did not consider energy issue of electric (IoT) equipment [36].

2.2 Blockchain technology

The first venture of blockchain technology was bitcoin (a cryptocurrency) by Nakamoto [37] to initiate a fast, cheap and transparent peer-to-peer money transaction. Progressively, the future industrial revolution also demands blockchain to improve the privacy of data-driven

Table 1: Personally identifiable information (PII).

Personally Identifiable Information (PII)
full name, home or office address, email, national identification number, passport, vehicle id, driving license, fingerprints, handwriting, credit card numbers, digital identity, birth date, birthplace, biological information, phone number, login name, social security number (SSN).

Table 2: Potential personally identifiable information (PPII).

Potential Personally Identifiable Information (PPII)
partial name, a portion of the address, some parts of an email, area code, eye colour, blood group, car colour, cache data, handwriting, bank name, phone company, race, religion, food habit, pet name

enterprise architecture [11]. Onik [11] stated opportunities of blockchain in industry 4.0. Zheng [38] mentioned four privacy improving blockchain features: decentralization, persistency, anonymity and auditability. Few frequently used jargons of blockchain technologies are discussed here (Figure 2) [11].

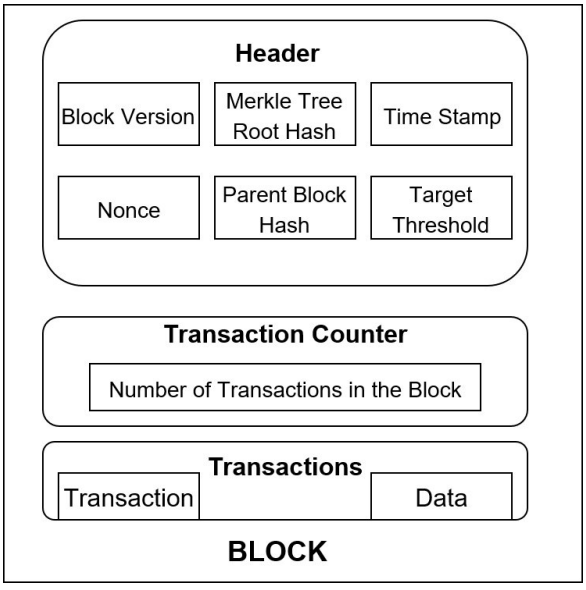

Figure 2: The architecture of a block [11].

- Blockchain types: Based on usability, blockchain is separated into three types. Among public, private and consortium blockchain, private and consortium are used at the organization level. Alternatively, security increases but privacy decreases in public blockchain.
- Node: Node generally represents a computer, owned by a participating organization or user. It is the core owner of any blockchain that verifies transactions with other nodes. A node serves as a connection point between blockchain technology and the user (Figure 2).
- Consensus algorithm: A consensus algorithm is used to approve decisions for nodes (Figure 2). Example: Proof of work (PoW), proof of stake (PoS), proof of burn (PoB) etc.
- Block: Transaction decision added to the current chain after effective consent is a block (Figure 2).
- Header: Block version represents currently accepted block version in blockchain network. Previous block hash is also stored in the header. Mining difficulty (block creation difficulty) is mentioned as a target threshold. A nonce is used only once to change the hash output of the block.
- Transaction counter: Transaction counter expresses the serial number of the current block.
- Transaction data: Depending on the usability, the purpose of this field fluctuates. It can be a bitcoin transaction, records, personal information, healthcare info, the hash of personal data etc.

2.3 Blockchain for personal data management

The blockchain has already been used for identity management by several other studies [13–15, 39, 40]. However, newly imposed EU-GDPR [17] has brought a set of new regulations that affect the way previous researchers dealt with personal information. Joshi [12] has combined data protection ontology and blockchain technology to propose (linkshare), to facilitate users to apply customized regulations on data. Other studies have also used blockchain technology for storing PII. Benhamouda [13] used multi-party computation (MPC) and hyper ledger to use 'chain code' that set secure communication among peers without revealing the identity. Zyskind [15] created a platform that provides personalized user-centric services by using blockchain with an access control consensus mechanism. Chen [16] managed cloud data privacy by

blockchain. It facilitates the traceability of data and allows access to healthcare resources after an approval procedure. Blockchain operated secured cyber-physical system was used for communication among cotton producers in [41].

2.4 General data protection regulation (GDPR) and blockchain technology

The GDPR by the European Union (EU) [17] is recently proposed as a new privacy regulation. According to this law, businesses organizations must comply with a set of rules, if they handle any personal information of the EU citizen [17]. A recent survey with 2000 IT professionals within the UK reported, only 47% of the respondents were fully aware of GDPR. Alternatively, 41% reported, they were aware but require more understanding of the subject. Finally, 9% of them were not aware of GDPR at all [42]. GDPR [17, 42, 43] classified data handling organizations as controller and processor. At this moment we will elaborate a few GDPR terminologies:

- Controller: According to the Art. 4 of GDPR [44], controllers are those legal persons or public authorities who process personal information of citizens from EU or member states.
- Processor: Art. 28 and Art. 4 defines, processers as the legal persons (third party) or public authorities who further process personal information on behalf of a controller [17].
- Right to erasure [45]: “The right to be forgotten” was mentioned in section 1 of Art. 17 of GDPR [17]. Since blockchain data are immutable, this term draws a line between blockchain and GDPR.

In addition to that, to solve the conflict of blockchain and GDPR, the adaptation procedure of blockchain is under study by GDPR observatory forum [46]. Filippi [47] mentioned the compatibility issues of blockchain with available laws. Similarly, another study [45] mentioned current blockchain architecture issues to comply with GDPR, both from the law and technical aspects. The study also noted several blueprints to adopt GDPR (Art. 42, Art. 4 and Art. 25):

- Privacy by design can be a way to modify current blockchain architecture to comply GDPR. "In code, we trust" the act of any protocol will be subject on how it codes a particular blockchain [48]. So, personal data removable blockchain technology development is not impossible, but it shall lessen the security features of blockchain technology (Art. 25 GDPR).
- GDPR only deals with data that are not anonymous. If created blockchain deals only with NPII, it can easily comply with blockchain. Oppositely, quasi-identifier or PPII can also lead to a user identity; which breaches GDPR indirectly. Therefore, a blockchain deals with the hash of personal information (PII and PPII) or NPII complies with GDPR (Art. 26 GDPR).

3 Proposed method

This section provides a detail description of the blockchain based personally identifiable information management system (BcPIIMS). This study uses a permissioned blockchain, where stakeholders (a user, U; a controller, C and a processor, P) form a private blockchain network.

3.1 Blockchain-based personally identifiable information (PII) management system (BcPIIMS):

To demonstrates the overall BcPIIMS architecture, a connection of stakeholders (U, C and P) and associated blockchain components are discussed here (Table 3).

- Node: U, C and P are represented as a blockchain node in BcPIIMS. BcPIIMS allows sharing of personal data (PII and PPII) among the stakeholders (U, C and P). If separation of NPII from personal data (PII and PPII) is not done within the EU boundary, C and P nodes locate within legal boundaries of GDPR. But, the proposed system improvises no constraint on the location of the user node, U.
- Block: A new block is formed and added to the existing blockchain after a successful information sharing happens among U, C and P.
- Block header: A Block header stores general information of a personal data sharing incident. It stores transaction time, data hash of the previous header, personal data encoding style etc.
- Transaction counter: Transaction counter stores the serial of a data sharing incident. Each time a U, C and P interact with each other to exchange data, this field increments by one.
- Transaction data: This field often conflicts with GDPR privacy terms for storing sensitive personal information [42]. Therefore, BcPIIMS stores hash of personal information (PII and PPII), data sharing terms and NPII.

– Consensus algorithm: BcPIIMS uses a round-robin scheduling system (consensus algorithm). Permitted nodes (U, C and P) validate a data sharing incident to generate a valid block. For robustness of the system, mining diversity is set to 0.75. To authenticate a transaction, at least 75 % of total nodes (U, C and P) must respond.

3.2 Working procedure of BcPIIMS

Firstly, the organization frequently collects personal information (PII and PPII) for market analysis and prediction. BcPIIMS delivers a set of personal information (PII and PPII) ρ, and NPII σ, from a user μ, to the controllers α_{1-n}. During this flow of information from μ to any α, a smart contract, η is created between α_{1-n} and μ. As, GDPR does not allow preserving of ρ on the blockchain, β, it only stores σ of μ. To separate σ and ρ, heuristic and contextual information classification approaches are followed in this study. Therefore, η, σ and other non-sensitive information are stored in the blockchain database, β. Data privacy terms, regulation, usability, distribution measures, breaching notification process and consensus between μ and α are stored in η. Local database, ψ stores the identities of data sharing entities (μ and α). Current hash of ψ, υ is also added to β. So, if the information is shared between μ and α, BcPIIMS stores associated information like this:

Local database, $\psi = \{\rho\}$

Blockchain, $\beta = \{\eta, \sigma, \upsilon, \mu, \alpha\}$

Secondly, controllers use several other processors for detailed analysis, prediction, profiling, business modelling, consumer estimation etc. Therefore, a controller α, provides the processor $€_{1-n}$, a set of personal information (PII and PII) ρ, of a user μ. At this time, available blockchain nodes (μ, α and €) again undergo a consensus algorithm to store all NPII, σ and hash of PII and PPII, υ. Oppositely, local databases, ψ of each node (μ, α and €) stores ρ and stores current hash, υ of local database as a new block in β. So, if the information is shared between α and €, BcPIIMS stores associated information like this:

Local database, $\psi = \{\rho\}$

Blockchain, $\beta = \{\eta, \sigma, \upsilon, \mu, \alpha, €\}$

Upon a successful consensus between users and controllers, a new block is added to the current blockchain (Figure 3). Similarly, another block is added after a successful consensus from the user, controllers and related processors (Figure 3). Personal information or user data is moving from left to right (from user to controller and from the controller to the processor) (Figure 3).

– Creation of block1: User shares PII, PPII and NPII to controller1. Block 1 is created after the consensus from the user and controller1 (Figure 3).
– Creation of block2: User shares PII, PPII and NPII to controller2. Block 2 is created after the consensus from the user and controller2 (Figure 3).
– Creation of block3: Controller1 shares PII, PPII and NPII to processor1 and processor2. Block 3 is created after consensus from the user, controller1, processor1 and processor2 (Figure 3).
– Creation of block4: Controller2 shares PII, PPII and NPII to processor1. Block 4 is created after consensus from the user, controller2 and processor3 (Figure 3).

3.3 Sharing of personal information from the user to the controller

– Processing of personal data: A user shares personal data (PII and PPII) with the controllers and processors. Controllers separate these data into three groups PII, PPII and NPII. Personal information separation mechanism from universal information is still in the infant stage. For now, heuristic and contextual approaches [24, 26] are applied to existing PII and PPII datasets in proposed BcPIIMS.
– Consent and smart contract: GDPR complied terms and conditions of personal information sharing are stored in a smart contract. This part fully depends on the law, and data collection policy.
– Adding a new block to blockchain: This step creates a block containing the smart contract, NPII and hashes of personal data (PII and PPII). This new block is then added to the current blockchain.

3.4 Sharing of personal information from the controller to the processor

– Sharing personal data: A controller shares a user's personal data to the connected processors. The controller also informs about this data sharing incident to the user. Processor again separates personal information (PII and PPII) and NPII.
– Consent and smart contract: All nodes create a new smart contract, holding terms and user consent for using personal information. After all the processors agree upon the conditions by controllers and user, processors are allowed to process personal information (PII and PPII).

Table 3: Example of a block in BcPIIMS.

Blockchain components	Stored data
Node	A user, Controller and Processor (data sharing entities)
Block version	0100000 (how the block is treated by this network)
Timestamp	Sunday, 28-Oct-10 03:56:10 UTC (time of this data sharing incidents)
Nonce	465323695 (authentication-difficulty deciding arbitrary number)
Transaction counter	126 (serial number of this data exchange)
Transaction data (PII and PPII)	0206CC8C3068416C9F24C2FC5D3A0D221EB97D8E416D78D46B3B7EC8F9AE2941 (Hash of PII and PPII)
Transaction data (NPII)	Red pen, 32 GB flash drive, New York, browser cache and cookies, "delete data after a month" (NPII and privacy terms)

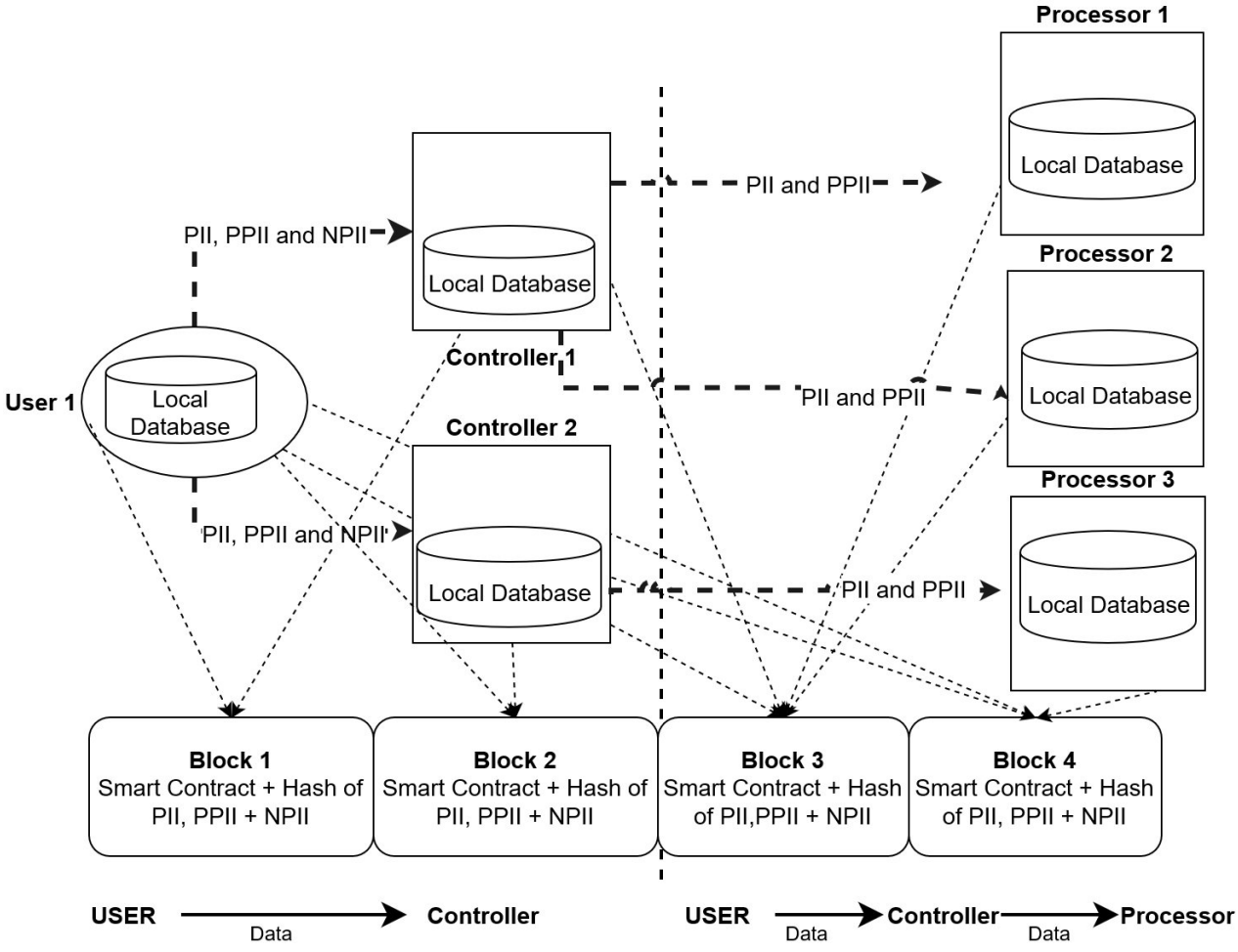

Figure 3: Overall architecture of the proposed system.

– Creating a new block: The final step creates a new block that holds the smart contract, NPII and hash of the processor's local database. This new block is then added to the existing blockchain.

4 Case study and testing

4.1 Case study: Personal information management scenarios

A use case scenario is described in detail to elaborate BcPIIMS. Four situations shown here are: user-controller scenario, user-controller-processor scenario, data delete, and data modify scenario.

– **User-controller data sharing scenario**

In the beginning, the user provides information to the controller (01). The controller then separates the data into PII, PPII and NPII (02) and generates the hash value of the shared personal information (PII and PPII). A list of the data (PII, PPII and NPII) along with the hash value will be published between nodes (03). The consensus of terms, conditions and user agreement are done between the users (04). The consensus must comply with GDPR regulation to be a smart contract. The final step creates a new block, BC1 with smart contract, NPII and hash of the local database. BC1 is then added as a new block to the current blockchain. As shown in Figure 4, to add a new block BC1, a consensus from user and controller is considered.

– **User-controller-processor data sharing scenario**

A controller shares personal information (PII and PPII) to a processor for analysis (06). After separation of personal information (PII and PPII) and NPII, processor gathers consensus from entities (user, controller and processor). Afterwards, a set of data (PII, PPII and NPII) and the hash value of the local database of the processor is published again among nodes. A consensus from all entities is gathered and stored in blockchain as a new smart contract (07). As shown in Figure 5, for adding the new block BC2, the consensus from all parties are taken into consideration.

– **Data modify scenario**

A user informs associated nodes (controller and processor) for modifying personal data (PII and PPII). Controllers and processors then check the previously stored smart contract. Afterwards, stakeholders (user, controller and processor) gather consensus on that modification. Eventually, controllers and processors modify user data in their local database and create a new block with the updated hash of local data. For double checking, modified data (PPII, PII) and updated hash are announced to connected nodes (user, controller and processor). The user can cross check requested update by comparing the hash stored in blockchain and local database (Figure 6).

– **Data delete scenario**

A user informs associated controllers and processors about the deletion of personal data. All nodes check terms of the smart contract and undergo a consensus for removal of PII, PPII or NPII. Afterwards, controllers and processors delete the user data from local databases. The deleted information (PII, PPII and NPII) and the hash value of local databases are published among other nodes. As blockchain hash is immutable, the hash of personal data remains in the block but, actual data is deleted from the local database. A user can now cross the previous hash and updated hash (hash of no data) to confirm successful delete of personal data from controller and processor. As the leftover hashes is of no use without the actual data, BcPIIMS complies GDPR (Figure 7).

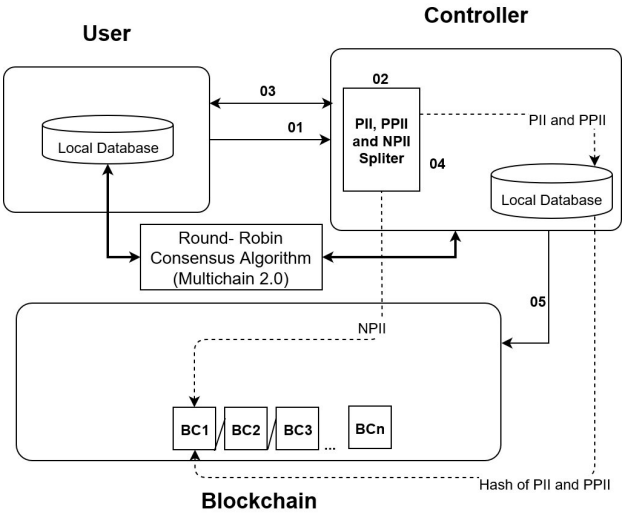

Figure 4: User and controller data sharing scenario.

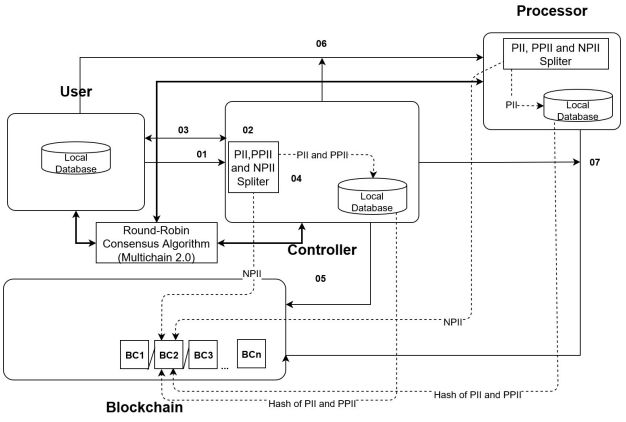

Figure 5: The user, controller and processor data sharing scenario.

4.2 System implementation

This section presents a pilot implementation scenario of BcPIIMS.

Node and network setup: Initially, this study builds a private network with 1 user, 3 controllers and 4 processors equipped with a blockchain implementation plat-

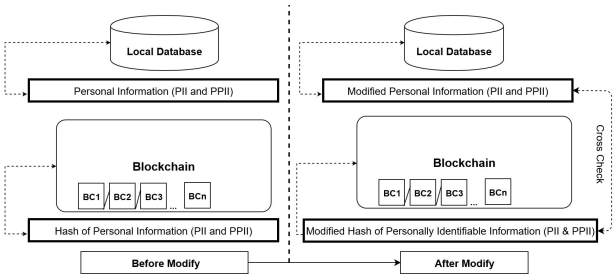

Figure 6: Data modify scenario.

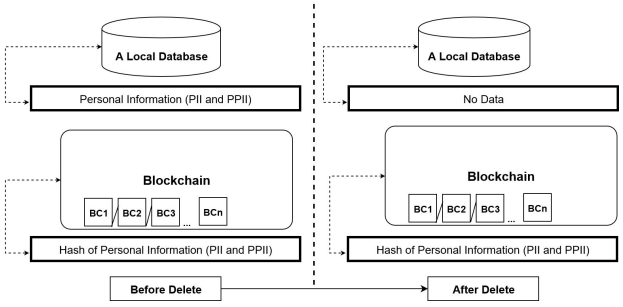

Figure 7: Data delete scenario.

form, multichain 2.0 (alpha) [49] (Figure 8). Each of them represents a computer set up, to create a virtual data sharing and tracking environment among themselves. We have implemented this prototype of BcPIIMS on 8 computers (Windows 10, i7-7700HQ CPU @ 2.8GHZ Samsung Inc. Republic of Korea) each representing a blockchain node. We have selected multichain 2.0 (alpha 3) that facilitates an on-chain and off-chain data storing mechanism [49]. As the proposed method is a private blockchain, consensus requires minimum difficulties to be achieved. Since nodes identities are known under a private blockchain, no nonce is used and the consensus is achieved with a round robin mechanism. However, mining diversity is set to 0.75, which requires a response from 75% of the total participants before finalizing a data exchange. An example of a block data storing (blockchain data vs actual data) is shown in Figure 9.

MultiChain 2.0 (alpha 3) [49] adds the following advantages to BcPIIMS to comply with GDPR [17].

- Writes large personal information in off-chain (local database). That is how the proposed study efficiently handles large-scale data.
- Built-in off-chain functionality allows easy management of PII, PPII and NPII among nodes.
- Provides the facility to add a hash of the off-chain data to the existing block.

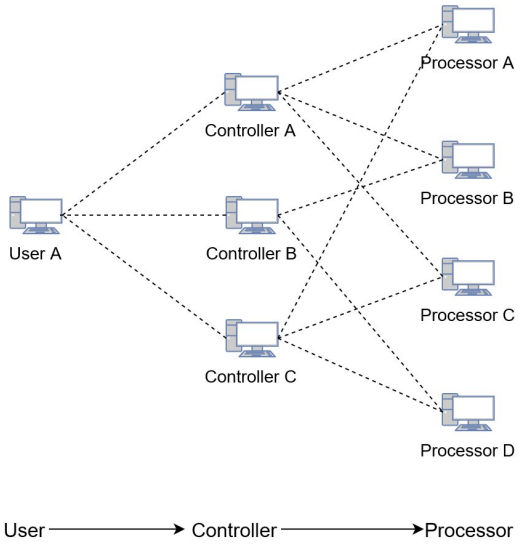

Figure 8: Pilot experimental network setup (tested with Multichain 2.0).

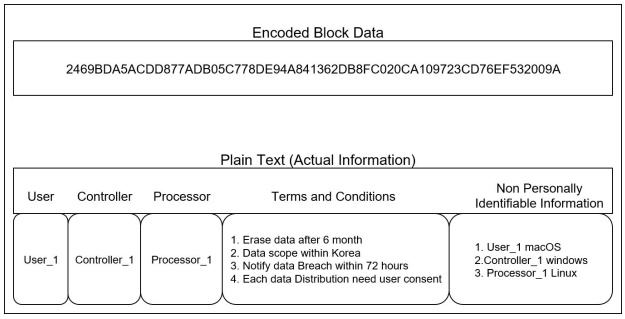

Figure 9: Blockchain data versus actual data.

5 Discussion and evaluation

We will now discuss a few architectural differences and subsequent advantages of BcPIIMS:

5.1 Functional advantages

- Efficient Storage Management: User stores identities (PII and PPII) in a local database and hashes with other NPII in the blockchain. As block storage is limited and off-chain (local storage) storage is expandable, therefore this study reduces the block size many times.
- Delete and modification capability: To comply with GDPR, proposed BcPIIMS architecture allows a user to modify and delete personal information. Since off-chain data is erasable and leftover hashes of personal information (PII and PPII) on blockchain are useless,

Table 4: Qualitative comparison of the proposed method (BcPIIMS) with other studies.

Factors	Liang [39]	Zyskind [15]	Zhu [40]	Proposed BcPIIMS
Separate storing of personal information	×	×	×	√
GDPR compliance	×	×	×	√
Efficient block size	√	√	×	√
The off-chain capable consensus algorithm	×	×	×	√
Confidentiality	√	√	×	√
Scalability	√	√	√	√

thus BcPIIMS fully complies GDPR. Cross-checking the hashes generated after data modification can confirm the accuracy of the change. Similarly, personal data (PII and PPII) deletion is also verifiable with help.

- Security: The consensus among users, controllers and processors are stored blockchain. As this consensus cannot be altered, modification of term is impossible. If changes happen on local databases of controllers and processors, a user can cross check from local databases of controller, processor and blockchain. By doing this, the identification of corrupted controllers or processors is possible. Although this study uses the local database, core blockchain benefits are not sacrificed.
- Transparency and verification: As the proposed system stores data terms and identities of controllers and processors, everything related to personal data is transparent. A user can track a data life cycle from the beginning to the end. Similarly, consensus terms are also intelligible, easily accessible and supported by all parties (user, controller and processor).

5.2 Core contribution and advantages of BcPIIMS

Firstly, sharing of information from a user to a controller or even to a processor are fully tracked. As BcPIIMS architecture stores hash of personal information (PII and PPII) in an immutable blockchain after consensus from all nodes, a user can easily acknowledge any changes. However, unintended data modifications are fully recoverable from other trusted nodes.

Secondly, a user can easily predict a data breaching controller or processor. The user can also track the exact set of breached data (PII and PPII). Associated leaking nodes are easily detectable. Finally, the user can claim compensation according to the terms stored in a smart contract.

Finally, the proposed study explains a comprehensive data sharing and tracking mechanism among core GDPR entities. This study explains the correct way of dealing with personal data from user's, controller's and processor's point of view.

5.3 Overall comparison

Table 4 shows a comparison of BcPIIMS and similar works.

6 Conclusions

To reduce the risk of personally identifiable information (PII) and potential personally identifiable information (PPII) leaking, we have proposed a blockchain based personally identifiable information management system (BcPIIMS). Our proposed model tracked the life cycle of personal data throughout the controllers and processors. By storing the personal data off-the-chain (not in actual blockchain), the system complied to the GDPR rule (right to be forgotten). Our research reveals that the BcPIIMS model ensures higher security and complies to the existing privacy regulations. Misuse, mismanagement and lesser scope for PII tracking were identified as major causes of privacy breaching. Using an off-chain blockchain with data hash checking, the proposed system successfully addressed those pitfalls. Besides, this study recommends user and stakeholders to use data tracking tools. Special attention and monitoring facilities should be improvised in information (PII, PPII and NPII) classification before storing in the blockchain. Privacy by design should apply in blockchain development for efficient privacy preservation. The future research goal is to develop a fully-fledged PII tracking and managing system. A future goal of this study is to present an in-depth performance comparison with existing systems. As the work is in progress, this version presents the theoretically technical advantages only.

In future study, experimental results of the conducted research will be presented.

Acknowledgement: This work was supported by the 2018 Inje University research grant.

References

[1] Number of social media users worldwide 2010-2021, Statista, https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ [Accessed: 06-Mar-2019]
[2] Cadwalladr C., Graham-Harrison E., Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach, The Guardian, 17 Mar 2018
[3] Zou Y., Mhaidli A. H., McCall A., Schaub F., "I've got nothing to lose": Consumers' risk perceptions and protective actions after the equifax data breach, In: Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018), 2018, 197–216
[4] First Half 2017 Breach Level Index Report: Identity Theft and Poor Internal Security Practices Take a Toll, https://www.gemalto.com/press/pages/first-half-2017-breach-level-index-report-identity-theft-and-poor-internal-security-practices-take-a-toll.aspx [Accessed: 06-Mar-2019]
[5] Kumar V., Reinartz W., Customer privacy concerns and privacy protective responses, In: Customer Relationship Management: Concept, Strategy, and Tools, Third Edition, Springer Texts in Business and Economics, 2018, 285–309
[6] Sweeney L., k-anonymity: a model for protecting privacy, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5), 557–570
[7] Ahmed M., Ullah A. S. S. M. B., False data injection attacks in healthcare, In: Boo Y., Stirling D., Chi L., Liu L., Ong KL., Williams G. (Eds.), Data Mining, AusDM 2017, Communications in Computer and Information Science, Springer, Singapore, 2017, 845, 192–202
[8] Miraz M. H., Ali M., Applications of Blockchain Technology beyond Cryptocurrency, Annals of Emerging Technologies in Computing (AETiC), 2018, 2(1), 1–6
[9] Miraz M. H., Donald D. C., Application of Blockchain in Booking and Registration Systems of Securities Exchanges, In: International Conference on Computing, Electronics & Communications Engineering (iCCECE), IEEE, 2018, 35–40
[10] Onik M. M. H., Miraz M. H., Kim C.-S., A Recruitment and Human Resource Management Technique using Blockchain Technology for Industry 4.0, In: Proceedings of the Smart Cities Symposium (SCS-2018), 2018, 11–16
[11] Onik M. M. H., Ahmed M., Blockchain in the Era of Industry 4.0, In: Ahmed M., Pathan A. S. K. (Eds.), Data Analytics, CRC Press, 2018, 259–298
[12] Joshi K. P., Gupta A., Mittal S., Pearce C., Joshi A., Finin T., Semantic approach to automating management of big data privacy policies, In: 2016 IEEE International Conference on Big Data (Big Data), 2016, 482–491
[13] Benhamouda F., Halevi S., Halevi T., Supporting private data on Hyperledger Fabric with secure multiparty computation, In: 2018 IEEE International Conference on Cloud Engineering (IC2E), 2018, 357–363
[14] Bahri L., Carminati B., Ferrari E., Decentralized privacy preserving services for online social networks, Online Social Networks and Media, 2018, 6, 18–25
[15] Zyskind G., Nathan O., Decentralizing privacy: Using blockchain to protect personal data, In: Security and Privacy Workshops (SPW), IEEE, 2015, 180–184
[16] Chen L., Hoang D. B., Novel data protection model in healthcare cloud, In: 2011 IEEE 13th International Conference on High Performance Computing and Communications (HPCC), 2011, 550–555
[17] EUGDPR – Information Portal, https://eugdpr.org/ [Accessed: 06-Mar-2019]
[18] Sucasas V., Mantas G., Althunibat S., Oliveira L., Antonopoulos A., Otung I., Rodriguez J., A privacy-enhanced OAuth 2.0 based protocol for Smart City mobile applications, Computers & Security, 74, 258–274
[19] Weingärtner R., Westphall C. M., A design towards personally identifiable information control and awareness in openid connect identity providers, In: 2017 IEEE International Conference on Computer and Information Technology (CIT), 2017, 37–46
[20] Kshetri N., Big data's impact on privacy, security and consumer welfar, Telecommunications Policy, 2014, 38(11), 1134–1145
[21] Caron X., Bosua R., Maynard S. B., Ahmad A., The Internet of Things (IoT) and its impact on individual privacy: An Australian perspective, Computer Law & Security Review, 2016, 32(1), 4–15
[22] Simpson J. J., The Supreme court of Canada rules on when lenders may share personal information without violating federal privacy legislation, Banking & Finance Law Review, 2017, 32(2), 417
[23] Humphreys E., Implementing the ISO/IEC 27001 information security management system standard, Artech House, Inc., 2007
[24] Pfitzmann A., Hansen M., Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management-a consolidated proposal for terminology, Version v0.31, Feb. 15, 2008
[25] McCallister E., Guide to protecting the confidentiality of personally identifiable information, Diane Publishing, 2010
[26] Murphy R. S., Property rights in personal information: An economic defense of privacy, In: Barendt E. (Eds.), Privacy, Routledge, 2017, 43–79
[27] Schaar P., Privacy by design, Identity in the Information Society, 2010, 3(2), 267–274
[28] Pose C., Raja U., Crossler R. E., Burns A. J., Taking stock of organisations' protection of privacy: categorising and assessing threats to personally identifiable information in the USA, European Journal of Information Systems, 2017, 26(6), 585–604
[29] Butler D. A., Rodrick S., Australian media law, Thomson Reuters (Professional) Australia Limited, 2015
[30] Onik M. M. H., Al-Zaben N., Yang J., Kim C. S., Privacy of Things (PoT): personally identifiable information monitoring system for smart homes, In: Proceedings of Symposium of the Korean Institute of Communications and Information Sciences (KICS), 2018, 256–257
[31] Razaghpanah A., Nithyanand R., Vallina-Rodriguez N., Sundaresan S., Allman M., Kreibich C., Gill P., Apps, trackers, privacy, and regulators: a global study of the mobile tracking ecosystem, In: Procedings of the Network and Distributed Systems Security (NDSS) Symposium 2018, 18-21 February 2018, San Diego, CA, USA
[32] Sui P., Li X., Bai Y., A study of enhancing privacy for intelligent

transportation systems: k-correlation privacy model against moving preference attacks for location trajectory data, IEEE Access, 2017, 5, 24555–24567

[33] Karegar F., Lindegren D., Pettersson J. S., Fischer-Hübner S., User evaluations of an app interface for cloud-based identity management, In: Paspallis N., Raspopoulos M., Barry C., Lang M., Linger H., Schneider C. (Eds.), Advances in Information Systems Development, Lecture Notes in Information Systems and Organisation, Springer, 2018, 26, 205–223

[34] Murmann P. , Fischer-Hübner S., Tools for achieving usable ex post transparency: a survey, IEEE Access, 2017, 5, 22965–22991

[35] Murmann P., Usable transparency for enhancing privacy in mobile health apps, In: MobileHCI 2018, 2018, 440–442

[36] Onik M. M. H., Al-Zaben N., Phan Hoo H., Kim C. S., MUXER – A new equipment for energy saving in Ethernet, Technologies, 2017, 5(4), 74

[37] Nakamoto S., Bitcoin: A peer-to-peer electronic cash system, http://bitcoin.org/bitcoin.pdf. [Accessed: 06-Mar-2019]

[38] Zheng Z., Xie S., Dai H., Chen X., Wang H., An overview of blockchain technology: Architecture, consensus, and future trends, In: 2017 IEEE International Congress on Big Data (BigData Congress), 2017, 557–564

[39] Liang X., Shetty S., Tosh D., Kamhoua C., Kwiat K., Njilla L., Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability, In: Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2017, 468–477

[40] Zhu L., Wu Y., Gai K., Choo K. K. R., Controllable and trustworthy blockchain-based cloud data management, Future Generation Computer Systems, 2019, 91, 527–535

[41] Yin S., Bao J., Zhang Y., Huang X., M2M security technology of CPS based on blockchains, Symmetry, 2017, 9(9), 193

[42] Carey P., Data protection: a practical guide to UK and EU law, Oxford University Press, Inc., 2018

[43] Voss W. G., European Union data privacy law reform: General data protection regulation, privacy shield, and the right to delisting, Business Lawyer, 2017, 72(1), 221–233

[44] EUGDPR – Information Portal, https://gdpr-info.eu/art-4-gdpr/ [Accessed: 06-Mar-2019]

[45] Wirth C., Kolain M., Privacy by blockchain design: a blockchain-enabled GDPR-compliant approach for handling personal data, Reports of the European Society for Socially Embedded Technologies, 2018, 2(6), (https://hdl.handle.net/20.500.12015/3159)

[46] Ferrari V., EU blockchain observatory and forum workshop on GDPR, data policy and compliance, Institute for Information Law, Research Paper No. 2018-04, Available at SSRN: https://ssrn.com/abstract=3247494 or http://dx.doi.org/10.2139/ssrn.3247494

[47] De Filippi P. D. F., Blockchain and the Law: The Rule of Code, Harvard University Press, 2018

[48] Yermack D., Corporate governance and blockchains, Review of Finance, 2017, 21(1), 7–31

[49] Off-chain (Alpha 3) Multichain 2.0, https://www.multichain.com/blog/2018/06/scaling-blockchains-off-chain-data/ [Accessed: 06-Mar-2019]