Risk Analysis of a GDPR-Compliant Deletion Technique for Consortium Blockchains Based on Pseudonymization

Lelio Campanile[1], Pasquale Cantiello[2], Mauro Iacono[1], Fiammetta Marulli[1], and Michele Mastroianni[1](✉)

[1] Dipartimento di Matematica e Fisica, Università degli Studi della Campania, Caserta, Italy
{lelio.campanile,mauro.iacono,fiammetta.marulli, michele.mastroianni}@unicampania.it
[2] Osservatorio Vesuviano, Istituto Nazionale di Geofisica e Vulcanologia, Napoli, Italy
pasquale.cantiello@ingv.it
http://www.unicampania.it, http://www.ingv.it

Abstract. Blockchains provide a valid and profitable support for the implementation of trustable and secure distributed ledgers, in support to groups of subjects that are potentially competitors in conflict of interest but need to share progressive information recording processes. Blockchains prevent data stored in blocks from being altered or deleted, but there are situations in which stored information must be deleted or made inaccessible on request or periodically, such as the ones in which GDPR is applicable. In this paper we present literature solutions and design an implementation in the context of a traffic management system for the Internet of Vehicles based on the Pseudonymization/Cryptography solution, evaluating its viability, its GDPR compliance and its level of risk.

Keywords: Blockchain · Privacy · GDPR · IoV · Risk analysis · Pseudonymization

1 Introduction

Blockchain technologies are now well established and appreciated for their intrinsic characteristics of security, reliability, transparency and inalterability of the information stored. By virtue of these characteristics, in addition to the original applications relating to cryptocurrencies, their scope of use has expanded to many other fields. Different areas of use are constantly being studied and researched and in the literature there are now many scientific works that demonstrate their applicability, validity and the concrete advantages that can derive from their adoptions. In particular, they can certainly be considered an excellent alternative, also in terms of costs, to databases and registries managed centrally

O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12956, pp. 3–14, 2021.
https://doi.org/10.1007/978-3-030-87010-2_1

by recognized and regulated authorities. Even greater are the benefits that can be derived from using blockchains technologies in complex, federated and distributed ledgers that are participated by subjects that have potential conflicts of interests [4]. The design of GDPR-compliant Blockchain Systems is an emerging topic for researcher and practitioners, as documented in [17] and [10].

One of the fundamental properties of these technologies, the guarantee of the immutability of the stored information, can however, at first sight, place a constraint on particular areas of adoption where instead the possibility of modifying or deleting some information must be provided. In particular, in contexts in which data of natural persons must be processed, the European General Data Protection Regulation (GDPR) [6] guarantees a series of fundamental rights for the citizen. In detail, each natural person has the right, if there are no higher legal obligations that prohibit it, to request and obtain from the entity who processes the information (Data Processor) that his data be rectified (Article 16) or even erased (Article 17). In this case, the need to modify/delete personal data is in contrast with the immutability of the blocks making up the chain.

This problem has been addressed, among the others, by the French supervisory authority CNIL[1] [5] which has recognized, as a method to guarantee the right to erasure, the deletion, not of the data itself, but that of the means used to full decode it. In this way, although the information is still present, it will no longer be accessible.

In this paper one of these techniques, namely *Pseudonymization/Cryptography* [16] is analyzed and a possible implementation is proposed. This is contextualized in an Internet of Vehicle scenario with the necessity of management of all related information (traffic, safety, accounting, property) as described in [2]. Along with the implementation, a risk analysis is conducted in order to assess the good practice to pursue the GDPR compliance.

After this introduction, the paper continues in Sect. 2 with a brief introduction of the security advantages of blockchain systems in IoV contest. In Sect. 3 several approaches to data deletion in blockchain found in literature are presented, and one of them is detailed in Sect. 3.1. Two use cases with data requests and data update/deletion are shown in Sect. 4 with a risk analysis in Sect. 5. The paper ends in Sect. 6 with conclusions and future work directions.

2 Security Advantages of Blockchain Systems

A blockchain [7] can be seen as a distributed database of digital events and transactions shared between participating entities. The characteristics of immutability of data contained in blockchains [19] are guaranteed by the model, which is in fact an open and distributed ledger running on a peer-to-peer (P2P) network.

Transactions are intrinsically verifiable and traceable without involving parties external to the chains. In fact, each transaction is added to the public register and

[1] A recognition of a procedure by an EU based supervisory authority is legally valid and recognized by all Countries that adhere to the GDPR.

is verified by consensus by the majority of the parties. At all times, a blockchain contains a certain and verifiable record of every single transaction ever made. Precisely for this reason every single transaction, once inserted in the chain, cannot be altered or canceled in any way. The distributed consent thus allows to have the assurance that an event has occurred guaranteeing the irrefutable certainty of the associated information in what can be seen in all respects as a secure distributed public ledger. The partners involved in a chain obtain benefits both in terms of management costs and the reduction of associated risks to ensure data security. The improvement of cybersecurity and privacy protection using the blockchain was analyzed in [12], demonstrating how this technology can guarantee better performances than the cloud in terms of security and privacy, both as low susceptibility to manipulation and falsification by malicious entities and in terms of data breach containment.

The distributed nature of blockchain systems and the consensus mechanism acts as a protection against hacking, since it is necessary to hack more than 50% of the nodes in order to determine a real Data Breach. Moreover, these systems are obviously much less prone to DDoS attacks due to their distributed nature. Improvements on PKI can be also obtained, as publishing keys on a blockchain may eliminate the risk of false key propagation.

3 Methods for Deletion and Updating

The intrinsic nature of the blockchain (Read/Append only model) provides that no type of alteration to the information present is possible. Any attempt to perform modifications would invalidate the entire chain. Obviously, the impossibility of modifying data also prevents them from being deleted.

In order to allow the use of blockchains in the areas where this operations must be allowed, for example to comply with the guarantees provided for by the GDPR, some research lines have therefore headed towards the identification of alternative techniques to allow such cancellation being implemented, not in a direct form, but as an indirect effect.

The first approach is to avoid saving any personal data in blockchain blocks, rather storing them in a separate repository and/or periodically performing data pruning to erase older data. This can be adopted when only a small amount of the whole managed data consists of personal data and when not all of them are subject to the Regulation.

A remarkable example of this approach is the Delegated Content Erasure in IPFS [14], in order to address the off-chain erasure over the Interplanetary File System, upon which many chains are based. The authors propose an anonymous protocol for delegated content erasure requests to be integrated in the IPFS to distribute an erasure request among all the IPFS nodes and, ultimately, to fulfill the requirements foreseen in the Right to be Forgotten. In order to prevent censoring, erasure is only allowed to the original content provider or her delegates.

The *Hash Function Modification* approach has been presented in [1]. It is based on the so-called chameleon hash functions [11], that provide an undeniable commitment of the signer to the contents of the signed document (as regular digital signatures do) but, at the same time, do not allow the recipient of the signature to disclose the contents of the signed information to any third party without the signer's consent. In this way there is no possibility that personal data will be exposed to external entities, and this reduces the associated risk factor, but at the same time the technique does not fully guarantee the respect of the envisaged rights.

Another approach is the *Modification of Consensus Mechanism* presented in [15]. The authors introduce the concept of alternative versions of events and data (transaction set), and a shared consensus to determine the current (valid) version. In a transaction set, only one of the transactions is specified as active (typically the last one), while all the others are inactive alternatives. An update can be obtained by adding a new a transaction and specifying it the active one. Furthermore, every mutable transaction set includes the so-called nope transaction which is equivalent to "no operation" action. If selected as active, a nope transaction effectively hides the others and this removes a mutable transaction from the history.

Another way to address the problem is presented in [16] as *Pseudonymization/Cryptography Approach*. Personal data over the blockchains are subject to pseudonymization and so their status of "personal data" is valid only for those who are in possession of the additional information needed to associate those data to the natural person they belong. This method is used in our work, and it is described in detail below.

3.1 Pseudonymization/Cryptography

As already mentioned above, the characteristic of immutability of the information contained in the blockchains clashes with the right to modification and cancellation guaranteed by the regulation. The natural person, if there are no higher legal impediments that prohibit it, has the right to delete (or correct) his personal data. If those data are contained in a chain, that right cannot be guaranteed directly; but, if data in the chains are stored in such a way that they are not directly attributable to a specific natural person (pseudonymized), then they can no longer be considered as personal data. The only thing to ensure is that the information useful to permit the association between a natural person and his data is kept off-chains.

This pseudonymization can be achieved by encrypting the data with cryptographic hash functions applied over them, or by using pseudonymous identifiers. In this way only those participants who possess the additional information (encryption keys o person-pseudonym associations) required for attribution can act as controllers. In case of joint controllers, these key information must be shared among them with a specific agreement in order to establish clear responsibilities for compliance to the Regulation.

The right of erasure can be guaranteed by eliminating this additional key information. In this way the processor (or jointly with the other controllers as stated by agreements) will no longer have the ability to attribute that data to the person they belong to. This technical measure is reliable when based on a solution that ensures that the additional information required for the association can be shared securely and at the same time reliably deleted.

4 IoV Use Cases

The Internet of Vehicles represents a fairly new application context for blockchains. If we extend this scope to that of Smart Roads, we will find ourselves faced with an application scenario in which different actors are involved and exchange information mutually. In [2] a model is depicted in which a whole series of information relating to the world of both personal and commercial mobility is managed. In this sense, management includes data relating to movements, traffic, security, safety, accounting and the aspects relating to ownership of vehicles, most of them with privacy concerns [3].

As described in [4], in the IoV world some use cases with privacy implications can be identified. For this work we will describe more use cases: a vehicle data access request made by the owner or by a public authority, and a request to update vehicle data to fulfill the right to update guaranteed by the GDPR. A diagram is shown in Fig. 1. It can be easily seen that a request for deletion is a sub-case of update.

We know that, when a person buys (or rents) a new vehicle, his personal data begin to be associated with that vehicle. In case of buying, the transaction is registered in a public registry within a blockchain to assert the property, while in renting process the transaction is registered upon the renting company. In both cases the association is imposed by law, mainly for safety purposes, but also for taxes (and for commercial reasons in case of renting). In case of accident, injuries or violations of traffic rules the need arises for an authority to obtain, with no constraints, the full content of information records about the vehicle and who was in its possession at the time of the accident. The owner of the vehicle has the right to access all the personal data belonging to him too.

The public registry should be able to register pseudo identities, being the only entity that keeps the association between pseudo identities and user/vehicles.

In Fig. 2 the process is shown of accessing data that can be originated by a User or by an Authority. A request is made to the Public Registry and contains user and vehicle identification. The Public Registry performs a lookup on its private repository to find the corresponding pseudo-identity (if any). If found, a further lookup is executed to extract the ID that points to the block in the Blockchain containing data. After response, the block of data is decrypted using the key associated with the pseudo-identity to recover the original data that can be answered to the originating requester.

The sequence diagram of the other use case, related to the the updating of user data, is shown in Fig. 3. A request of update made by the User with updated

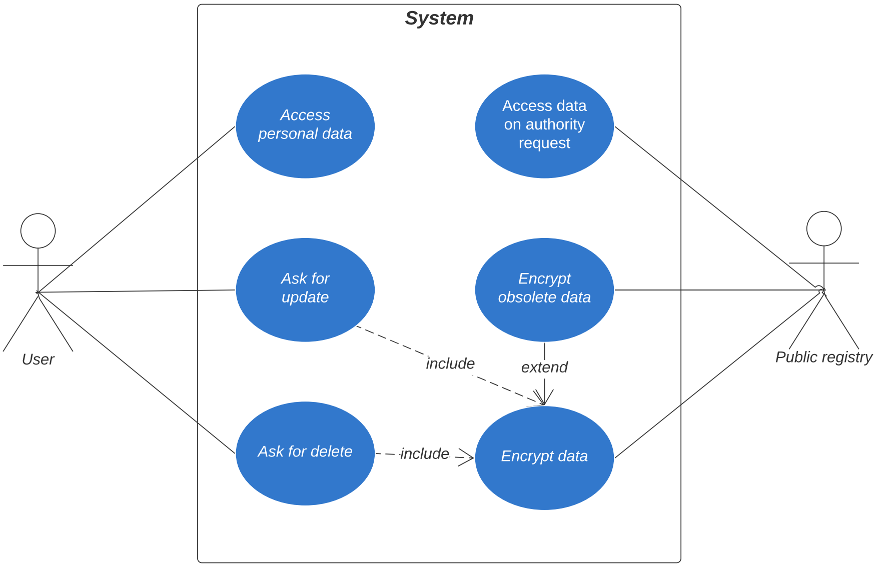


Fig. 1. Use case diagram

data (Art. 16 of GDPR - *Right to rectification*) causes the following operations to be performed: (i) a lookup of the pseudo-identity is performed on Public Registry private repository, (ii) a delete on that pseudo-identity is executed, in order to kill the associated private key, (iii) a new pseudo-identity is generated and stored for the User, (iv) data is encrypted with the new identity, (v) encrypted data is saved as a block in the Blockchain, (vi) the related block ID is stored in the private repository of the Public Registry and associated to the pseudo-identity. (vii) status code is sent to the User.

It is easy to notice that the previous block of data continues to be present in the chain but, since the key to decrypt it no longer exists, there is no way to access contained plain text data, therefore it is possible to consider them logically deleted.

Request to delete data (Art. 17 of GDPR - *Right to be forgotten*) starts as the previous one, but simply ends after deleting the pseudo-identity, so causing the impossibility to decryption the original data.

5 Risk Analysis

In the field of information security risk assessment, the most commonly used approach is the qualitative assessment approach. In the qualitative approach, classes and relative values are used to show the impact and probability of a particular scenario. This approach is widely used, due to the ease of understanding

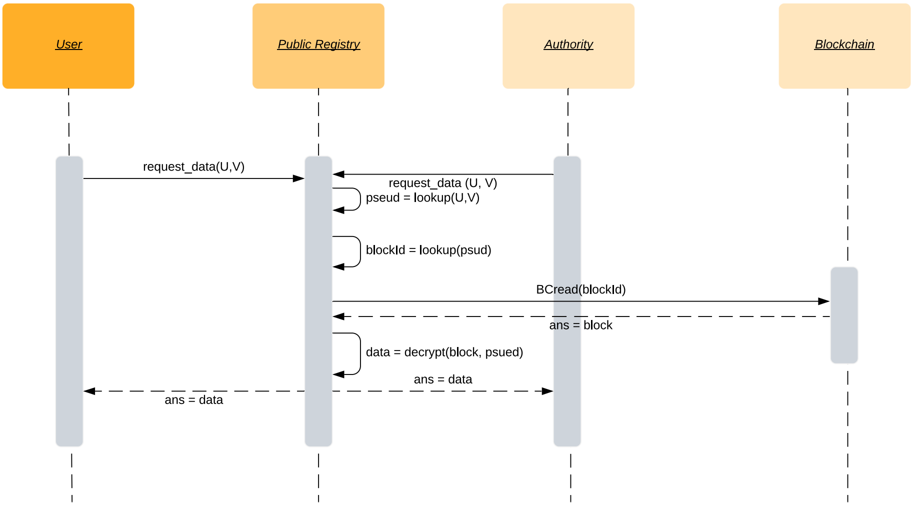


Fig. 2. Sequence diagram of user data access request

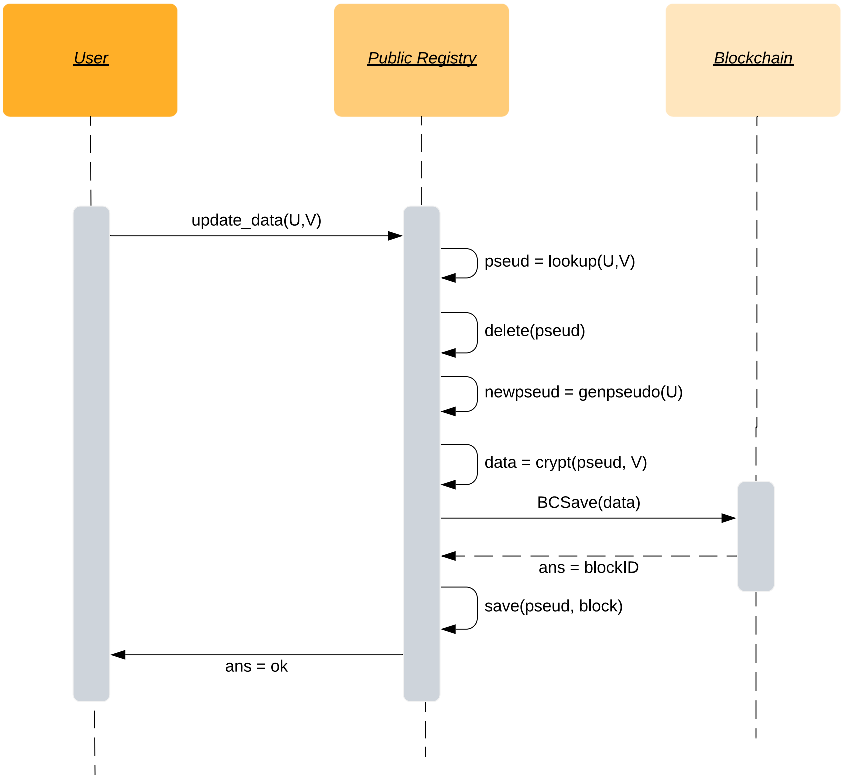


Fig. 3. Sequence diagram of user data update request

and implementation; the calculations involved are simple and the valuation of assets, threats, and vulnerabilities is easier [13] with respect to alternatives.

In this work, we carry out the assessment using CNIL-PIA [9] methodology, which uses a qualitative approach. The use of this methodology is suggested by many other EU data protection authorities (including the Italian *Garante per la Protezione dei Dati Personali*). We also used the software developed by CNIL, namely PIA, to conduct analysis, in order to avoid any possible difformity in the evaluation process with respect to the recommended procedures.

The assessment is carried out estimating Severity and Likelihood using qualitative criteria. Severity represents the magnitude of a risk: it is primarily estimated in terms of the extent of potential impacts on data subjects, taking into account existing, planned or additional controls (which should be mentioned as justification). The estimation is done regarding the possible damage that would occur to the user/data subject in case of problems. Likelihood represents the feasibility of a risk to occur: it is estimated in terms of the level of vulnerabilities of the supporting assets concerned and the level of capabilities of the risk sources suitable to exploit them, taking into account existing, planned or additional controls. The scale for classifying likelihood is related to the feasibility of the occurrence of a risk for the selected risk sources by exploiting the properties of supporting assets.

In the CNIL methodology, feared events are classified as follows:

- Illegitimate access to personal data **(I)**
- Unwanted modification of personal data **(U)**
- Disappearance of personal data **(D)**

Both Severity and Likelihood are classified in a 1–4 scale, while 1 is the lower level of risk and 4 is the higher level:

1. Negligible;
2. Limited;
3. Significant;
4. Maximum.

The Severity level may be lowered by including additional factors to oppose identification of personal data, such as encryption, pseudonymization, anonymization, and so on. On the other hand, the Likelihood level may be lowered by including additional factors, such as firewalls, logging, monitoring, and similar solutions [9].

At this point, determining the impact the users (Data Subjects) could face off in case of a data breach is necessary. The impacts taken into account, and the pertinent severity according to CNIL-PIA, are:

- Cost rise for Data Subjects (e.g. increased insurance prices) **(severity = 2)**;
- Targeted online advertising on a confidential aspect **(severity = 2)**;
- (Temporary) Denial of access to IoV services **(severity = 2)**;
- Fraud **(severity = 3)**.

The Risk Assessment procedure has been carried out by the authors. Each author is a computer security expert, and two of the authors are also Data Protection Officers (DPO). The working group is composed as follows: a IoT/IoV Expert (evaluation of threats and risks), a Software Expert (evaluation of threats and risks related to software modules), a Privacy Expert/DPO (review of the assessment), another Privacy Expert/DPO (assessment approval). What is here reported is the result of a panel discussion, after a separate analysis performed by each author in its own role, aiming at avoiding possible biases and ambiguity in the interpretation of partial reports and to synthesize the basis for the presented analysis.

Other impacts with higher severity, like "Loss of evidence in the context of litigation", have been excluded due to the distributed features of blockchain systems, which ensure virtually no permanent data loss.

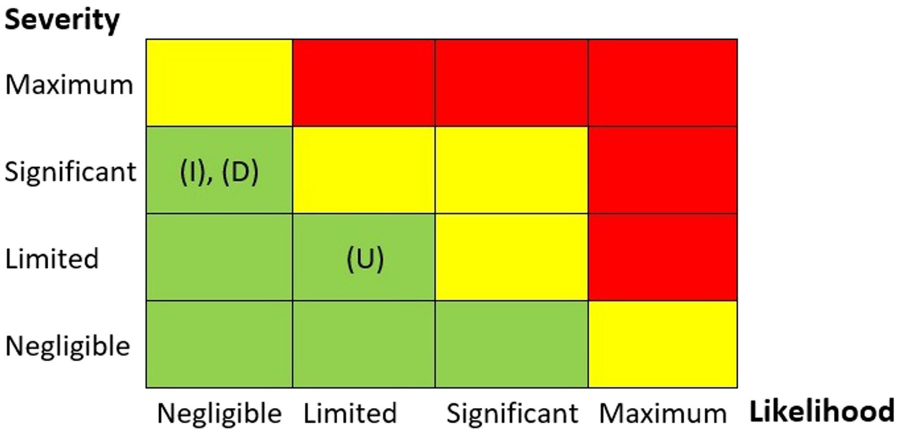
Thereafter, a choice is necessary of which threats must be taken into account, to define the scope of this analysis and state the actual extent of its results. As this study aims to be general and is not related to a single practical case arose by a specific situation from the real world, in order to keep generality and to ensure realism in the process, our choice is based on data retrieved on Verizon 2019 Data breach Investigation report [18] and EY Global Information security Survey 2018–19 [8] about the most common and relevant threats that may be directed against a system based on the architecture proposed in [2]. The threats taken into account are:

- *Hacking*: it is the most frequent threat, circa 54%; due to the intrinsic robustness to hacking attacks [2], the likelihood is limited **Likelihood = 2**;
- *Use of stolen credential*: almost 30% of threats; in this case, the likelihood is significant, due to the possibility for a subject to access to important users' data **Likelihood = 3)**;
- *Privilege abuse*: circa 10% of threats; for the same reason of the preceding point, the likelihood is significant **Likelihood = 3)**;
- *Natural disasters*: although considered infrequent events, (circa 2% of total breaches), they are taken into account because they can lead to a severe data loss; the likelihood is negligible **(Likelihood = 1)**;
- *DDOS Attacks*: usually this kind of attack does not lead to a data modification or illegitimate access, but leads to loss of availability (data disappearing), which is considered a breach event in GDPR; due to the fact that Blockchain systems are virtually immune to DDOS attacks, the likelihood is negligible **(Likelihood = 1)**.

It is noticeable that the low Likelihood values for the last two threat categories are a direct consequence of the design choices behind the proposed architecture, in compliance with the purposes of our research. Interested readers may find further details and results about the risk analysis for the overall system, complementary for the analysis presented in this paper, in the Appendix of [4], together with a detailed description of the general technique applied here with relation to what required by the GDPR.

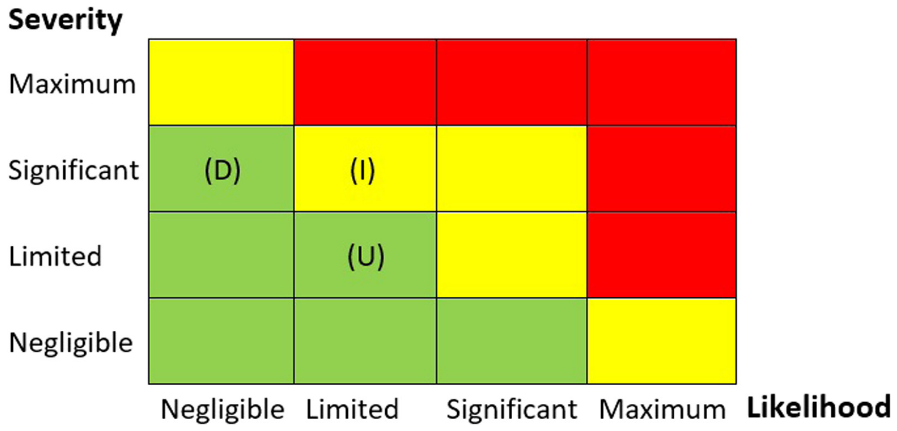
Figure 5 presents the results of the Risk Analysis conducted about the chosen deletion strategy. It may be noted that all three feared events, namely illegitimate access (I), unwanted modification (U), or disappearing (D) of personal data, are in the green zone (limited risk).

In order to evaluate the impact of the additional feature with respect to the overall architecture of the system, a comparison with the results of the overall risk analysis, available in [4], is needed. We report those results in Fig. 4, using the same format of Fig. 5 for an easier comparison. As evident, the new feature presents characteristics that do not lower the overall risk level of the system, as all evaluations for the new feature are in green zone and in a position on the grid that is equivalent or less severe with respect to the evaluations for the same events related to the system.



Severity				
Maximum	🟨	🟥	🟥	🟥
Significant	🟩 (I), (D)	🟨	🟨	🟥
Limited	🟩	🟩 (U)	🟨	🟥
Negligible	🟩	🟩	🟩	🟨
	Negligible	Limited	Significant	Maximum	Likelihood

Fig. 4. Risk analysis results for the deletion feature



Severity				
Maximum	🟨	🟥	🟥	🟥
Significant	🟩 (D)	🟨 (I)	🟨	🟥
Limited	🟩	🟩 (U)	🟨	🟥
Negligible	🟩	🟩	🟩	🟨
	Negligible	Limited	Significant	Maximum	Likelihood

Fig. 5. Risk analysis results for the system [4] (Color figure online)

6 Conclusions and Future Work

In this article we focused on understanding the effects of one between all techniques literature offers to make blockchain compliant with the GDPR, namely the Pseudonymization/Cryptography Approach, on the risk exposure of an example blockchain-based complex data management architecture that involves data covered by GPDR. In particular, our interest focused on how to permit operations legally provided for by the Regulation, but which cannot be carried out directly on a blockchain by its immutable nature, in a system that is devoted to IoV management. Our study shows that this approach does not introduce into the system additional risk factors that are worsening the overall risk analysis results of the system.

Future work includes the analysis of other techniques and a comparison between the results, in order to evaluate the best alternative with respect to risk minimization, and a general evaluation of other dimensions of exploration of the advantages and disadvantages of the alternatives, in order to provide a global choice criterion.

References

1. Ateniese, G., Magri, B., Venturi, D., Andrade, E.: Redactable blockchain - or - rewriting history in Bitcoin and friends. In: 2017 IEEE European Symposium on Security and Privacy (EuroS&P), pp. 111–126 (2017). https://doi.org/10.1109/EuroSP.2017.37
2. Campanile, L., Iacono, M., Levis, A.H., Marulli, F., Mastroianni, M.: Privacy regulations, smart roads, blockchain, and liability insurance: putting technologies to work. IEEE Secur. Priv. **19**(1), 34–43 (2021). https://doi.org/10.1109/MSEC.2020.3012059
3. Campanile, L., Iacono, M., Marulli, F., Mastroianni, M.: Privacy regulations challenges on data-centric and IoT systems: a case study for smart vehicles. In: Proceedings of the 5th International Conference on Internet of Things, Big Data and Security, vol. 1, AI4EIoTs, pp. 507–518. INSTICC, SciTePress (2020). https://doi.org/10.5220/0009839305070518
4. Campanile, L., Iacono, M., Marulli, F., Mastroianni, M.: Designing a GDPR compliant blockchain-based IoV distributed information tracking system. Inf. Process. Manag. **58**(3), 102511 (2021). https://doi.org/10.1016/j.ipm.2021.102511
5. Commission Nationale de l'Informatique et des Libertés: Blockchain and the GDPR: solutions for a responsible use of the blockchain in the context of personal data. https://www.cnil.fr/en/blockchain-and-gdpr-solutions-responsible-use-blockchain-context-personal-data
6. Council of European Union: General Data Protection Regulation (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj
7. Crosby, M., Pattanayak, P., Verma, S., Kalyanaraman, V., et al.: Blockchain technology: beyond bitcoin. Appl. Innov. **2**(6–10), 71 (2016)
8. EY: EY Global Information security Survey 2018–19 (2019). https://assets.ey.com/content/dam/ey-sites/ey-com/en_ca/topics/advisory/ey-global-information-security-survey-2018-19.pdf

9. French Data Protection Authority (CNIL): Privacy Impact Assessment (PIA) - Knowledge Bases (2018). https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-3-en-knowledgebases.pdf
10. Haque, A.B., Islam, A.K.M.N., Hyrynsalmi, S., Naqvi, B., Smolander, K.: GDPR compliant blockchains-a systematic literature review. IEEE Access **9**, 50593–50606 (2021). https://doi.org/10.1109/ACCESS.2021.3069877
11. Krawczyk, H., Rabin, T.: Chameleon hashing and signatures. Cryptology ePrint Archive, Report 1998/010 (1998). https://eprint.iacr.org/1998/010
12. Kshetri, N.: Blockchain's roles in strengthening cybersecurity and protecting privacy. Telecommun. Policy **41**(10), 1027–1038 (2017)
13. Landoll, D.: The Security Risk Assessment Handbook: A Complete Guide for Performing Security Risk Assessments, 2nd edn. Second Edition. CRC Press Inc, Boca Raton (2011)
14. Politou, E., Alepis, E., Patsakis, C., Casino, F., Alazab, M.: Delegated content erasure in IPFs. Future Gener. Comput. Syst. **112**, 956–964 (2020)
15. Puddu, I., Dmitrienko, A., Capkun, S.: μchain: How to forget without hard forks. IACR Cryptology ePrint Archive (IACR), February 2017. https://eprint.iacr.org/2017/106
16. Rieger, A., Guggenmos, F., Lockl, J., Fridgen, G., Urbach, N.: Building a blockchain application that complies with the EU General Data Protection Regulation. MIS Q. Executive **18**(4), 263–279 (2019). https://doi.org/10.17705/2msqe.00020
17. Shi, S., He, D., Li, L., Kumar, N., Khan, M.K., Choo, K.K.R.: Applications of blockchain in ensuring the security and privacy of electronic health record systems: a survey. Comput. Secur. **97**, 101966 (2020)
18. Verizon Enterprise: 2019 data breach investigation report (2019). https://enterprise.verizon.com/resources/reports/2019-data-breach-investigations-report.pdf
19. Zheng, X., Zhu, Y., Si, X.: A survey on challenges and progresses in blockchain technologies: a performance and security perspective. Appl. Sci. **9**(22), 4731 (2019)