Blockchain-based Identity Management and Data Usage Control (Extended Abstract)

Ricardo Neisse, Gary Steri, Igor Nai Fovino

Blockchain-based Identity Management and Data Usage Control (Extended Abstract)

Ricardo Neisse, Gary Steri, and Igor Nai Fovino

European Commission Joint Research Centre (JRC)
Via E. Fermi 2749 Ispra (VA), Italy, I-21027
{ricardo.neisse, gary.steri, igor.nai-fovino}@ec.europa.eu

The General Data Protection Regulation (GDPR) [1], which will be enforceable from May 2018, introduces significant changes on the obligations of data controllers and processors in the context of the data protection legistlation of the European Union (EU). These obligations are defined by a single set of rules that should be adopted by all EU Member States including, among others, the need for explicit consent with the possibility of withdrawal and the right to erasure. The GDPR applies to data controllers (organizations) that access data of a data subject (persons) and data processors (organizations) that process data on behalf of the controller.

The focus of our work is on a blockchain-based solution using smart contracts, in the scope of the GDPR, to support data accountability and provenance tracking when subject's data is accessed by controllers and possibly forwarded to data processors. The main goal is to empower subjects with a trusted and transparent solution allowing the tracking of who has accessed their data or identity attributes, to verify if the access and usage of the data did not violate their consent encoded in privacy preferences, and to give the possibility of withdrawing or modify their preferences in case they change their mind. Furthermore, such a solution also benefits controllers and processors with a way to prove they have rightfully obtained consent and are processing data without violating the data protection obligations. The main advantage of using blockchain technologies is the transparency, auditability, and immutability features that potentially enable trust and trasparency on the proposed solution.

In our analysis [2] we identified three possible models for the solution, which are depitect in Figure 1. In the first model data subjects express their privacy preferences by means of usage control policies that are embedded in specific smart contracts deployed in the blockchain for each controller or processor receiving their data. In the second model, subjects create smart contracts for each data item that is possibly shared with multiple data controllers. In the third model, each controller expresses their privacy conditions in a smart contract with an interface allowing users to join or leave the contract, meaning they are giving or withdrawing their consent for each data controller or processor. These policies, which can be selected before hand or on request from a library of policy templates, express the conditions for data access, usage, and transfer to data processors. Our contribution is the analysis of design choices, implementation, and performance/scalability analysis of these blockchain-based data accountability and provenance tracking solutions.

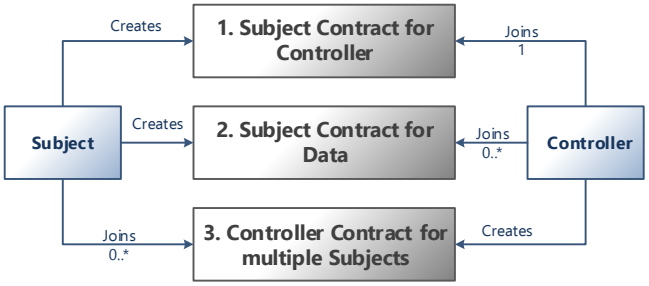

Fig. 1. Provenance and accountability tracking models using blockchain.

With respect to user privacy, data accountability, and data tracking granularity each model provides different properties. In the first model there is one contract per pair Subject/Controller, the contract tracks data provenance, events, and encodes specific policies for each controller. Since subjects can use a different pseudonym for each controller, contracts are unlinkable among controllers. In the second model there is one contract per pair Subject/DataInstance, the contract tracks data provenance, events, and a shared policy for all controllers acessing the respective data. Controllers may be able to uniquely identify a subject in case a unique identifier is shared (e.g. name, e-mail, etc.). In the third model there is one contract per controller that is shared for multiple subjects, the contract includes only the general privacy conditions of each controller without the possibility of customization for each data subject. Thee evaluation/tracking of events is done off-blockchain and subjects are also able to benefit from the use of pseudonyms for each controller.

From the three analyzed models we provided two concrete implementations for the first and third model described above, with an extensive analysis with respect to data accountability features, provenance tracking granularity, privacy, anonymity, performance, and scalability. The second model was excluded since it allows linkability of subjects across different controllers. For the first and third model contracts were implemented using a shared secret nonce to prevent linkability across multiple smart contracts of a subject, and to obfuscate the privacy preferences, data, and identity provenance information using a one-way hash function. We show that for more sensitive data with less frequent exchanges, such as medical data, a more fine-grained solution where subjects create contracts with each controller and processors is more adequate (first model). On the other hand, for more dynamic data with more frequent exchanges and strict scalability and performance requirements, controllers or processors should manage a contract that registers all subjects accepting all or part of the data usage conditions (third model).

A possible solution for scalability issues we are currently investigating is the use of sharding, where the blockchain is divided into separate chains that are responsible for contracts of a subset of all controllers and processors. These separate private chains then synchronize with the public chain on regular intervals,

for example every N blocks, in order to allow for public verifiability [5]. In case the separated chains are managed privately, data protection supervisory authorities can then join all chains just as observers in order to prevent censorship and guarantee that transactions of data subjects are not indiscriminately refused. As future work we also plan to investigate the possibility of using business blockchain approaches such as the Hyperledger solution, which uses a different algorithm for reaching consensus and also has a more ambitious scalability and performance goal with thousands of transactions per second [4, 3].

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Official Journal of the European Union L119/59 (May 2016), http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC
2. Neisse, R., Steri, G., Fovino, I.N.: A blockchain-based approach for data accountability and provenance tracking. In: Proceedings of the 12th International Conference on Availability, Reliability and Security, Reggio Calabria, Italy, August 29 - September 01, 2017. pp. 14:1–14:10. ACM (2017), http://doi.acm.org/10.1145/3098954.3098958
3. Vukolić, M.: The Quest for Scalable Blockchain Fabric: Proof-of-Work vs. BFT Replication, pp. 112–125. Springer International Publishing, Cham (2016)
4. Vukolić, M.: Rethinking permissioned blockchains. In: ACM Workshop on Blockchain, Cryptocurrencies and Contracts (BCC'17) (4 2017), available at: http://vukolic.com/rethinking-permissioned-blockchains-BCC2017.pdf
5. Wiki, E.: Sharding faq - on sharding blockchains. https://github.com/ethereum/wiki/wiki/Sharding-FAQ (2017), online; accessed April 06th 2017