
XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

ISBN

A protocol for data exchange with free samples using smart contracts

Rafael Genés-Durán, Juan Hernández-Serrano, Oscar Esparza, Miquel Soriano, José Luis Muñoz-Tapia
Departamento de Ingeniería Telemática,
Universitat Politècnica de Catalunya (UPC)
(rafael.genes,j.hernandez,oscar.esparza,miquel.soriano,jose.luis.munoz)@upc.edu

Marta Bellés-Muñoz
Universitat Pompeu Fabra (UPF)
belles.mm@gmail.com

Distrust between data providers and data consumers is one of the main obstacles hampering digital-data commerce to take off. Data providers want to get paid for what they offer, while data consumers want to know exactly what are they paying for before actually paying for it. In this paper, we summarize a protocol that overcomes this obstacle by building trust based on two main ideas. First, a probabilistic verification protocol, where some random samples of the real dataset are shown to buyers in order to allow them to make an assessment before committing any payment; and second a guaranteed, protected payment process, enforced with smart contracts on a public blockchain, that guarantees the payment of the data if and only if the data provided meets the agreed terms, and that refunds honest players otherwise.

***Palabras Clave*—data exchange, smart contract, blockchain, DLT, payments**

I. INTRODUCCIÓN

The use of data has increasingly become a crucial factor in the success of businesses. Businesses not only collect and analyse the data they generate, but increasingly rely on third party data to enhance its business value. In general, making proper data agreements is not easy, specially the task of valuing data and convincing customers of their value without giving them away [1]. The creation of marketplaces addresses many of these problems. Allowing providers and consumers to deal with common interests in a platform where both parties can meet each other and trade information solves the integration problem of connecting consumers and providers.

This article focuses on the problem of convincing consumers of data value, which can be seen as a form of lack of trust towards data providers. Traditionally, this problem cannot be solved without previously establishing confidence between parties. This represents an entry barrier to new providers in the market, hurting competence and thus, reducing utility for consumers. To exchange value safely, it is essential to ensure that consumers get the product they are paying for and that providers get paid. These two things are often carried out without any strict protocols and guaranteed just by existing trust. Typically, counterparties that know each other from previous experience or that are aligned with future interests, are confident that no intent to scam will be made by the other party, since confidence is often more beneficial than gains from fraud.

But, when stronger assurance than that is needed, it is a common practice to use a trusted third-party (TTP) to whom all parties trust to guarantee that the process is carried out correctly by all individuals involved. TTPs entail an extra cost for all parties, and generate a single point of failure that could produce critic delays and denial of services. Distributed Ledger Technologies (DLTs) can be seen as a paradigm shift when it comes to the need of TTPs. Using DLTs, all participants in the network can maintain a set of synchronized data (who owns what) without the need for a central authority (TTP) guaranteeing integrity, fairness and data availability.

In this paper we summarize DEFS (Data Exchange with Free Sample Protocol), a protocol that addresses the lack-of-trust between providers and consumers in a data trade. DEFS preserves the security, privacy and fairness standards that marketplaces should guarantee, and it includes the capability of checking some sample portions of the dataset before committing to purchase.

II. BACKGROUND

A. Merkle Hash Trees

A Merkle hash tree (MHT) is an authenticated data structure where every leaf node of the tree contains the

cryptographic hash of a data block and every non leaf node contains the concatenated hashes of its child nodes [2]. MHTs allow to link a set of data to a unique hash value, the Merkle hash tree root (MHR), allowing efficient and secure verification of consistency and content of large sets of data.

Figure 1 contains an example of a MHT with 8 leaves. To show that a certain value is stored in a leaf of the MHT, one can create a Merkle proof (MP), which consists of a list of the additional nodes required to compute the root of the tree. For instance, a Merkle proof showing that h_3 is stored in the MHT from Figure 2 would consist of the nodes h_2, h_{01} h_{4567}, $h_{01234567}$. Note that with h_3 and the first three nodes of this list anyone can compute the root of the tree. If the root matches $h_{01234567}$, then the proof is valid proof of membership for h_3 in the tree.

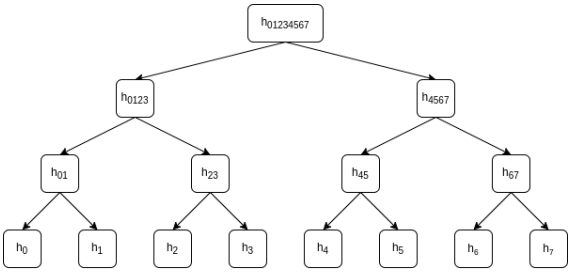

Fig. 1. MHT of 8 leafs.

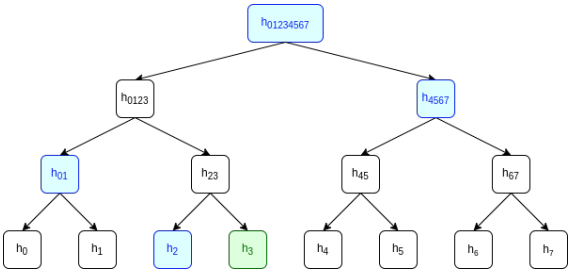

Fig. 2. Merkle proof for h_3. MP(h_3)=$h_{01234567}$,h_{4567},h_{01},h_2

III. State of the Art

Decentralized marketplaces have arisen as a solution to enhance security, sovereignty and trust in data exchanges [3], [4], [5].

One interesting initiative is GAIA-X [6], which is an European project to develop the foundations for a federated open data infrastructure connecting both classical architectures with decentralized infrastructures in order to build a transparent ecosystem for the end users taking advantage of the decentralized benefits.

One of the main technologies that is fostering data marketplaces is the Internet of Things (IoT) with huge amounts of data being generated from sensors and devices. The increasing necessity of monetizing these data is also pushing research. In the literature, we can find several works that propose decentralized marketplaces for IoT using distributed ledger technologies to enhance the data exchanges with transparency, trust and integrity [7], [8], [9]. Among others, decentralized marketplaces are being implemented in new disruptive scenarios such as artificial intelligence [10], smart cities [11], [12], and the connected car [13]. In fact, the value of the data is becoming more and more important to the business interactions which is reflected in the new technologies and their necessity to generate this new era of decentralized marketplaces.

An example of a decentralized data trading solution is presented in [14]. As in our protocol, the data on sale are not stored on the blockchain but in some external (and possibly distributed) storage platform. Similar to our protocol, the proposed solution symmetrically encrypts data on sale and uses a Merkle tree of cryptograms to register the associated trades on the blockchain. However, the solution proposed not only requires to generate symmetric cryptograms but also each of these cryptograms needs to be asymmetrically signed. Additionally, authors propose to use Plaintext Checkable Encryption (PCE) [15] to check on-chain that the cryptograms have been correctly encrypted. In DEFS, we avoid using asymmetric encryption, which is much slower than symmetric encryption.

Another remarkable implementation of a decentralized data trading solution is presented in [16], where authors present SDTE, a secure blockchain-based data trading ecosystem. As our protocol, SDTE tries to mitigate the existence of dishonest parties in data exchanges. However, SDTE focuses on an scenario in which the buyer does not need to have access to a complete dataset but it only needs the findings from the data analysis. For this case, SDTE proposes a data processing-as-a-service, where the buyer is paying for the analysis of the seller's dataset. SDTE is build using an Intel's SGX-based secure execution environment to protect the data processing, the source data and the analysis results. As we will show in the following section, DEFS is not designed as a data processing-as-a-service but as a data exchange-as-a-service. In the latter, the seller wants to buy the complete dataset not computed data. For this scenario, DEFS provides a probabilistic verification protocol and a conflict resolution protocol that is guaranteed and supported by a smart contract.

IV. Data Exchange Protocol

In this section we summarize DEFS, a protocol that addresses the problem of data trading between provider and consumers using a smart contract deployed in the blockchain as a broker. As we explained before, the use of DLTs can replace the role of TTPs in payment processes. When using DLTs, participants in the network can maintain synchronized data and share payment information without the need of a central authority, guaranteeing this way the integrity, fairness and availability of the data. In this manner, DEFS makes use of a smart contract to preserve the security and privacy standards that marketplaces should guarantee.

Another gap to cover in this data trading scenario is generating trust between data consumers and data providers. Here it comes the novelty of DEFS: our proposed data exchange protocol is designed with the capability of

checking random samples from the dataset, so that consumers are able to infer if the complete dataset is worth to be paid for, enhancing the trust of the consumer's side. On the other side, the smart contract acts as a broker during the payment procedure, ensuring providers that they will receive the payment for the data they exchanged.

A. Protocol Explanation

We assume that before starting the protocol, a data provider advertises her data to the public using off-blockchain means, such as a data marketplace. Then, a consumer interested in a particular dataset contacts the provider, who starts the DEFS protocol to perform the data exchange and payment. To prevent potential extensive leaks of the data, it is important that there is one DEFS protocol per each individual consumer. DEFS consists of three different phases:

1) **Protocol preparation**: in this initial phase, the provider prepares not only the data to be exchanged, but also all the parameters and cryptographic material necessary to demonstrate that the data exchange is secure and private. More specifically, the provider:
   - Divides the complete dataset in portions. These portions are chosen randomly from the dataset (not consecutively).
   - Generates a seed to generate symmetric cryptographic keys.
   - Uses these keys to create a MHT, whose root can be used to check the correctness of this cryptographic material.
   - Encrypts a random permutation of the data portions with the keys, obtaining an encrypted and randomized version of the whole dataset.
   - Creates another MHT using the hashes of these cryptograms as leaves, whose root can be used to verify the correctness of the cryptograms generated.
   - Deploys a smart contract in the blockchain containing certain public parameters and that smart contract acts as a broker during the rest of the protocol.

   If the consumer has interest in obtaining the dataset, the protocol continues as follows:
   - The consumer receives the whole dataset encrypted but it cannot be decrypted at that very moment.
   - The consumer queries the smart contract to obtain the root of the tree of cryptograms and verifies that all the cryptograms belong to this tree.

   At this point, all entities (consumer, provider and smart contract) are ready to start the protocol execution phase, in which the consumer will have access to the complete dataset and perform the payment.
2) **Protocol execution**: in this phase, the consumer will be able to get some samples of the dataset (for free) to evaluate if it is worth to pay, and if so, it will obtain the dataset and the provider will be paid:
   - The consumer will choose at random some sample portions to be revealed.
   - The provider will disclose the keys for those samples, so the consumer can evaluate the quality of the dataset.
   - If the consumer is not convinced, the protocol ends here. However, if it decides that it is worth paying the dataset, it will commit the payment to the smart contract.
   - The provider is asked to publish the seed (that will disclose all the encryption keys) in the smart contract.
   - If the consumer is able to properly decrypt the dataset, after a timeout, the provider is paid and the protocol ends.
   - If the consumer is able to prove that there were problems with the previous procedure, it starts the conflict resolution phase to obtain a refund.

   The following phase will only be needed in case the consumer considers that is cheated on.

3) **Conflict resolution***: this phase is optional, it only takes place if the consumer detects a provider misbehaviour. The following are the cases that can end with a refund if he is able to demonstrate this misbehaviour:
   - Keys are not properly generated.
   - Cryptograms do not have the proper format.

B. Protocol Properties

The main properties provided by our protocol are the following:

1) **Data samples evaluation**. The consumer gets a free set of fair samples of the data being traded before paying. The protocol ensures that neither the consumer nor the provider are able to manipulate the chosen data or select specific samples.
2) **Payment guarantees**. The provider gets paid if and only if the consumer has access to the whole set of data. That is, the consumer can not get the data without paying for it and the provider does not get paid without disclosing the data.
3) **The solution is cost-efficient**. Due to high fees on public ledgers, DEFS minimizes the amount of data stored on the network, which is also independent of the quantity of data traded. This way, both the amount of data stored and the number of interactions with the distributed ledger is constant.
4) **Non-repudiation**. The DEFS protocol ensures that any party involved in the exchange is not able to cancel and/or deny the data exchange once an agreement is made. Since the hash function used to generate the MHT is assumed to be collision-resistant, the MRC and MRK logged in the smart contract creation will prevent other data to be faked as bought or sold this way. Moreover, the use of a public blockchain enhances the integrity of the actor actions.

5) **Liveness**. The different timeouts guarantee that the protocol reaches a final state, even when one of the parties quits in advance. The provider can cancel the smart contract if no consumer reaches him out and the timeouts set after payments ensure that any counterparty can finalize the execution of the protocol favourably for it if the other party does not act on time.

C. State Diagram

The protocol operation and the interactions between the different stakeholders and the smart contract are detailed in Figure 3.

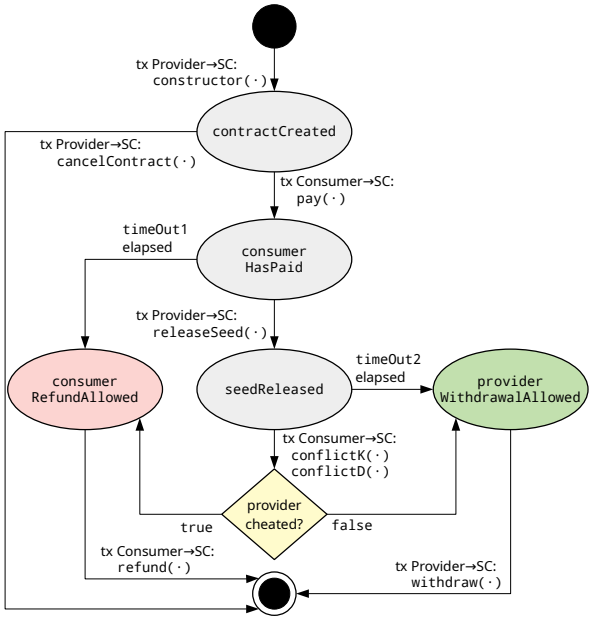

Fig. 3. State diagram of the smart contract.

V. Conclusions

Distrust is one of the main obstacles to implement exchanges between data providers and data consumers in a decentralized way. In this article, we summarize a protocol that allows a consumer to probabilistically obtain and check a subset of a dataset on sale from a provider before committing the payment. The protocol is executed using a smart contract deployed in a public distributed ledger. Once the consumer accepts to buy the dataset, the payment process, the agreed terms, and the possible refunds are managed and enforced by the smart contract. To expose the dataset, our protocol splits the data in portions and encrypts and stores each portion off-chain. Then, we create a MHT for the cryptograms and another MHT for the encryption keys. The encryption keys are related to each other using a cryptographic hash function in a way that allows us to implement a cost-efficient conflict resolution mechanism. The security analysis of our protocol shows that consumers and providers are economically protected and that the provider can reduce the risks of identity-replication attacks by adjusting the amount of free samples disclosed to the consumer.

Agradecimientos

This research has been funded by i3Market (H2020-ICT-2019-2 grant number 871754). This work is also supported by the TCO-RISEBLOCK (PID2019-110224RB-I00), ARPASAT (TEC2015-70197-R), Project RTI2018-102112-B-I00 (AEI/FEDER,UE) and by the Generalitat de Catalunya grant 2014-SGR-1504.

Referencias

[1] L. D. W. Thomas and A. Leiponen, "Big data commercialization," *IEEE Engineering Management Review*, vol. 44, no. 2, pp. 74–90, Second 2016.

[2] F. Haider, "Compact sparse merkle trees," Cryptology ePrint Archive, Report 2018/955, 2018, https://eprint.iacr.org/2018/955.

[3] H. Yoo and N. Ko, "Blockchain based data marketplace system," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, 2020, pp. 1255–1257.

[4] L. Mikkelsen, K. Mortensen, H. Rasmussen, H.-P. Schwefel, and T. Madsen, "Realization and evaluation of marketplace functionalities using ethereum blockchain," in *2018 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC)*, 2018, pp. 47–52.

[5] V. P. Ranganthan, R. Dantu, A. Paul, P. Mears, and K. Morozov, "A decentralized marketplace application on the ethereum blockchain," in *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, 2018, pp. 90–97.

[6] A. Braud, G. Fromentoux, B. Radier, and O. Le Grand, "The road to european digital sovereignty with gaia-x and idsa," *IEEE Network*, vol. 35, no. 2, pp. 4–5, 2021.

[7] K. R. Ăzyilmaz, M. DoÄan, and A. Yurdakul, "Idmob: Iot data marketplace on blockchain," in *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, 2018, pp. 11–19.

[8] D.-D. Nguyen and M. I. Ali, "Enabling on-demand decentralized iot collectability marketplace using blockchain and crowdsensing," in *2019 Global IoT Summit (GIoTS)*, 2019, pp. 1–6.

[9] P. Tzianos, G. Pipelidis, and N. Tsiamitros, "Hermes: An open and transparent marketplace for iot sensor data over distributed ledgers," in *2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 2019, pp. 167–170.

[10] V. Arya, S. Sen, and P. Kodeswaran, "Blockchain enabled trustless api marketplace," in *2020 International Conference on COMmunication Systems NETworkS (COMSNETS)*, 2020, pp. 731–735.

[11] S. Musso, G. Perboli, M. Rosano, and A. Manfredi, "A decentralized marketplace for m2m economy for smart cities," in *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 2019, pp. 27–30.

[12] G. S. Ramachandran, R. Radhakrishnan, and B. Krishnamachari, "Towards a decentralized data marketplace for smart cities," in *2018 IEEE International Smart Cities Conference (ISC2)*, 2018, pp. 1–8.

[13] B.-G. Jeong, T.-Y. Youn, N.-S. Jho, and S. U. Shin, "Blockchain-based data sharing and trading model for the connected car," *Sensors*, vol. 20, no. 11, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/11/3141

[14] Y.-N. Li, X. Feng, J. Xie, H. Feng, Z. Guan, and Q. Wu, "A decentralized and secure blockchain platform for open fair data trading," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 7, p. e5578, 2020, e5578 cpe.5578. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5578

[15] S. Ma, Y. Mu, and W. Susilo, "A generic scheme of plaintext-checkable database encryption," *Information Sciences*, vol. 429, pp. 88–101, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025517301640

[16] W. Dai, C. Dai, K.-K. R. Choo, C. Cui, D. Zou, and H. Jin, "Sdte: A secure blockchain-based data trading ecosystem," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 725–737, 2020.