Enabling Multi-Party Consensual Data Exchange Through Blockchain

Jack Hickey[1] and Ruairí O'Reilly[2]

[1] Jaguar Land Rover,
jhickey3@jaguarlandrover.com
[2] Cork Institute of Technology, Ireland,
ruairi.oreilly@cit.ie,
www.cs.cit.ie

Abstract. The quantity and quality of data available to an organisation plays an increasingly important role in its operation. This data can relate to a variety of subjects, from the internal logistics to consumer sentiment towards a product in a specific market. This data provides increasingly optimal behaviour derived from its analysis, e.g. improved decision making. As more advanced, application-specific, machine learning models are developed, the organisation with the largest share of data will gain an advantage over its competitors.
It is postulated that smaller entities with minority shares in data within a domain possess only a fragmented view of a market. This fragmented view puts the smaller entity at a disadvantage and enables larger entities to reap unfair, competitive advantages. This unequal dynamic should be rectified. To that end, in this work, a model is proposed that enables the consensual sharing of data between multiple parties using blockchain.

Keywords: Blockchain, Data Analytics, Distributed Systems, Federated Learning

1 Introduction

Growing quantities of data enable the creation of increasingly accurate and reliable models of consumer behaviour, sentiment, and risk-analysis [4]. Notably, companies with the largest share of market data have an advantage over their competitors within their domain. The insights the smaller entities have present a fragmented view of the market. This fragmented view damages their ability to monitor trends and consumer sentiment effectively, ultimately resulting in a less competitive market space.

When a dominant player emerges, they will utilise their data advantage for analytics to garner insights into their business revealing information, such as consumer behaviour and price points, as well as trends within the industry itself. This domination could potentially hurt the organic growth of the industry. Developing a fuller picture of the market would give smaller entities the ability to compete against dominant players, fostering a more competitive market

place. Sharing fragmented data amongst entities within the same domain could generate a more complete view of the market.

Sharing data between competing entities is challenging for numerous reasons. Forming an agreement between multiple organisations is difficult, as organisations are reluctant to share their data as they fear it provides a competitive advantage [9]. The difficulty of reaching an agreement is further complicated due to privacy concerns [3] and legal restrictions [5] surrounding the security of the data involved.

The benefits of multi-party data sharing have the potential to increase the value that can be derived from data analytics for these companies. Fully executing an agreement that addresses data security, privacy concerns, and legal considerations is an addressable challenge.

Blockchain is a technology built around consensus. Consensus in blockchain allows a decentralised network of members who do not have to trust each other to form an agreement. Consensus is achieved through immutable smart contracts, which is code that acts as the form of trust for these members.

This paper proposes the use of blockchain to allow competing entities to reach a consensus of sharing fragmented data to contend against the dominant players. Ultimately, it seeks to present a model that both facilitates competition and enables the collaborative sharing of data such that optimal market insights can be garnered. The challenge in providing such a solution can be summed up in three points:

1. Consensus, creating a consensus model where all parties involved can agree that the data is shared equitably.
2. Data privacy, data is appropriately handled such that it is not personally identifiable.
3. Legal, ensuring the system is lawful with regards to data sharing and GDPR.

This paper focuses on the technical challenges in provisioning these three characteristics through the use of digital agreements via blockchain, ultimately providing a means for competing parties, who do not trust each other, to exchange data. Furthermore, the potentials of such a system to improve the ability of smaller entities to compete against dominant players within a data market are explored.

This paper postulates that the consensus of blockchain can be utilised to provide a means for multiple competing parties who do not trust each other to exchange data. This hypothesis is examined in the context of allowing smaller entities to share their fragmented market view in order to garner a complete perspective and compete against the dominant players within their respective domains by exchanging data.

Future developments of the model propose its usefulness within the areas of data analytics and machine learning. A use case examined is the application of the model to federated learning in which the design holds a particular affinity.

The remainder of this paper is organised as follows: Section 2 reviews work previously carried out that addresses similar. Section 3 presents an overview

of the model and introduces the approach taken to achieve consensus between parties. Section 4 details the proof of concept implemented and explores how a fully working model could be realised. Section 5 presents the conclusion and findings of this paper and discusses some planned future work.

2 Related Work

The value of data is increasingly accepted and is what enables companies to compete and stand out within their respective markets. In markets that are heavily data-dependent, a few key players have dominated. The hypothesis presented is that smaller entities often only have a fragmented view of the market place, making it difficult to compete against a dominant player.

This section looks at how blockchain can be utilised to bring about consensus to share data between these smaller entities who are competing with each other. Application of blockchain in this context would allow smaller entities to garner a more complete view of the market and increase their competitiveness against a dominant player.

The application of such a system is also examined for the potential for federated learning to improve machine learning models. This is considered a potential use case of the consensus model. A working assumption of the paper is that those with a minority share of the data domain have a fragmented view of their market.

2.1 Blockchain

Blockchain in its basic form is an immutable ledger which allows transactions to take place in a decentralised manner [10]. Blockchains use a consensus algorithm which dictates what constitutes a legitimate write to the network and is agreed upon by all parties involved. Three main types of blockchain have emerged: public blockchain, consortium blockchain, and private blockchain [2].

Public blockchain, as implied by the name, allows anyone to view records and take part in the consensus process. **Private blockchain** only grants write permissions to one centralised organisation. **Consortium blockchain** consists of only a pre-selected group of nodes able to participate in the consensus algorithm, where the right to read may be public or private.

The application proposed is open to organisations to run models on data; there must be a mechanism in place to ensure data privacy by restricting access to the data itself as well as to the entire network. Public blockchains are automatically exempt for this reason. Similarly, the network does not want the control to be in a single, centralised authority; thus, private blockchains also do not meet the requirements. A consortium blockchain would allow for each organisation within the network to have an equal say within the consensus.

Blockchain is a technology that will disrupt many fields, and its uses for sharing data has already been explored [7]. The use of blockchain's consensus, as opposed to a centralised authority, enables decentralised organisations to work together, using the rules and consensus algorithm of the blockchain network

to dictate an agreement and remove any reliance on a centralised authority. Applying blockchain consensus to our model would allow multiple parties to collaborate in sharing their data while using the transparent rules to regulate its sharing.

Blockchain has also been utilised for preserving data privacy [11] by using blockchain as a transparent data access control manager that relinquishes the need for human intervention when providing data access, relying on the power of the consensus algorithm. The same idea of a data access control manager by abstracting data access through the blockchain is applied in this work. The blockchain network will have direct access to all data, while the parties will only be able to access the sanitised data that the blockchain permits.

Blockchains often already use a certain kind of consensus algorithm with little flexibility in how consensus is achieved. The model requires a consensus algorithm that can accommodate the various needs of different markets. Flexibility is required so that rules can be put in place that allows each party to reach consensus.

This is why we propose using Hyperledger Fabric (HF). HF is an open-source and permissioned blockchain with a modular consensus algorithm [1].

2.2 Consensus

Blockchain technologies are typically associated with a fixed consensus protocol such as proof of work (POW) or proof of stake (POS) [10]. Consensus is agreeing about the legitimacy and order of transactions that occur on the blockchain network. A minority of blockchain technologies utilise a modular consensus protocol that gives the developer a say in what constitutes an accepted block in the network.

HF is unique amongst blockchain technologies, in that HF's consensus covers the entire process from proposal and endorsement to ordering, validation and commitment [1]. The consensus protocol provided with HF is exceptionally flexible, leaving the architect to decide what constitutes an agreement on the network.

Consensus for this project is an agreement between the entities exchanging data on the network. The consensus protocol outlines the rules that the entities involved will agree upon as to what is allowed on the network in terms of writes, queries, and data access as well as anything specific to each industry request. to The blockchain network enforces this. Achieving consensus between the competing entities to allow them to share data is automated through the design of the blockchain consensus algorithm. The algorithm is simply code that enforces the contract between these entities.

Using blockchain allows these parties to never directly interact with each other and not explicitly trust each other. The transparency of the consensus protocol and the ability to audit the chain and its transactions allows these parties to trust the model, as opposed to each other.

2.3 Federated learning

Federated learning is a potential future use-case that this model could be applied to and so is worth discussing. Federated learning is an approach that enables collaborative training of machine learning models without sharing raw data [6].

Federated learning trains the model on edge devices using the local data and summarises the changes as an update to the centralised model that the devices contribute to. Edge devices, by their definition, could be considered minority data stakeholders. This data privacy-oriented approach to machine learning has been by adopted by organisations in much the same way, known as multi-party federated learning.

Federated learning comes with its own set of challenges, most notably the high communication cost of a massively distributed system and security [8]. As the number of participants involved grows, the more challenging management of the system becomes. The different hardware and network communication variants make fault tolerance and stragglers more prevalent than in typical data centre environments. In a system where competitors intend to share data, data ownership is a requirement, as is the capacity to verify and validate that ownership.

The blockchain model removes the need for a centralised model, while also providing the basis for training the models without the need for exchanging data with the other participants. The blockchain model also provides easily verified data ownership and the coordination of results and trained models through the ordering of blocks.

3 Methodology & Design

This paper explores how to form an agreement to share data between competing entities. Consensus enables parties to agree to how they will share data-this is the true value of using blockchain.

The backbone of blockchain is its consensus algorithm. Each blockchain offers a method for reaching consensus that a block is valid. This method is a combination of algorithm and smart contracts. Using smart contracts a flexible means of reaching consensus can be provided for the sharing of data, while also providing transparency into how this is achieved.

We assume that entities will only participate if there is value, security, and most importantly, consensus.

1. Value: The contribution of other entities matches the value of a data contribution. The output of this contribution must also be equally valuable to the other entities.
2. Security: The guarantee that their data is secure, as well as the security of upholding the contract between parties.
3. Consensus: All parties must agree to the consensus algorithm that the blockchain network enforces. This consensus upholds the previous two points and is the underpinning technology that supports this system.

The goal of the system is to achieve consensus to share data between competing parties. The blockchain network is the backbone of the system for providing this consensus.

Figure 1 outlines the system. Member refers to any participating entity of the network who shares data within their respective domains. They can perform several actions relating to the chain and management of their data.

Running queries is abstracted through a service, as depicted in Figure 1. The service acts to ensure the consensus is preserved and runs on the blockchain network. The entity runs its own set of data related queries through the service. The service queries the blockchain network by first checking the contract is being upheld. It then runs the queries which are returned to the service. The results of the queries are compiled within the service, and personally identifiable information (PII) is removed. The final result of all these queries is returned to the member. Finally, the result writes a log relating to the type of data accessed so that the other members of the system can audit this.

The blockchain network is the core of the system that enables the parties to interact with the shared data. Abstraction of the data helps to uphold data privacy and trust.

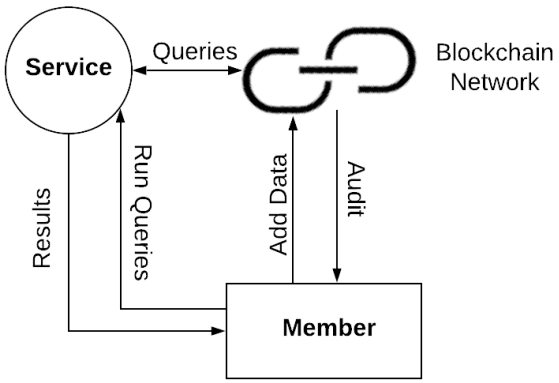

Fig. 1. An overview of member interaction with the blockchain model

3.1 The Consensus Protocol

In HF, consensus is the entire transactional flow, which serves to generate an agreement on the order and to confirm the correctness of the set of transactions constituting a block.

In the context of the proposed model, consensus describes the process by which the members of the network reach an agreement to share data. This consensus is backed up by the chaincode, which is the term used by HF to describe smart contracts and the permissions of the HF network.

There are several principles of the system that should be upheld by the network to ensure consensus: Data Integrity, Data Equality, Model Integrity

- **Data Integrity** is a means that ensures each organisations data cannot be tampered with or exposed.
- **Data Equality** addresses concerns that some organisations may monopolise data ownership. Providing mechanisms that discourage inequity and enable greater sharing of data amongst minority entities will assist in the prevention of data monopolisation.
- **Model integrity** enables each organisation to keep the models and their respective results private.

Smart contracts are simply code that enforces the rules of the network. In order for a business to perform transactions of data on the network, the contracts must be defined. These smart contracts in HF cover terminology, data, rules, concepts definitions, and processes. Smart contracts govern all interactions between the parties involved. This is the most critical part of the HF network for reaching consensus on how data is exchanged on the network. Smart contracts are how consensus is achieved.

HF has two methods for keeping data private within a transaction: channels and private data collections. Channels work as methods of communicating between organisations that enable the data of a transaction to be revealed only to those involved in the channel. Unlike channels, private data collections enable portions of the data to remain private between organisations.

HF has the concept of roles which includes peers, orderers, client applications, administrators and more. HF uses this to determine permissions. The permissions are incredibly flexible and can be related to the member's organisation, unit, role or even specific identity. The use of applying rules to an organisation can enable system equality to be upheld by ensuring that each organisation abides by the same general rules. The flexibility of the permissions then enables each organisation to define the permissions of roles further so that they can have control over the actions its members can take. This is all done through the membership service in HF.

Policies are another part of HF that will be utilised to reach consensus. Policies contain the list of organisations that have access to a given resource and also specifies how many organisations need to agree on a proposal to update a resource, such as a channel or smart contract. With policies, changes to the network between organisations are agreed upon and implemented without the need for third-party intervention.

4 Implementation

A proof of concept (POC) was developed to investigate the feasibility of the proposed solution. The POC used Hyperledger Composer, a tool for developing on Hyperledger Fabric that has now been designated for developing POCs.

The POC developed is a data exchange network with permissions in place for controlling data access. The POC enables data to be stored privately on the network, only accessible to those with permission. Functionality was developed that would query the data on the network only returning that which an entity had been granted permission to access. The queries, however, did not sanitise the data. Each entity can exchange data on this network in order to increase the data they can access. This simple data exchange shows the feasibility of using smart contracts and permissions on HF to provide a means of reaching consensus.

Continuing from this demonstration of data exchange, the intent is to show the added value of the increase in data. This value is realised by demonstrating that through the combination of fragmented data, one can increase the accuracy of machine learning techniques and data analytics.

Beyond this, we would then develop further the permissions to allow direct data access to the owners of the data. The service would be added to sanitise and control data access and uphold the full rules of the contract discussed in section 3.

5 Conclusions

Achieving consensus of data exchange between competing parties is challenging, though it is a challenge worth addressing, as the value of a more complete market view allows parties with minority stakes of data to contend against dominant players.

This work postulates that the permissioned blockchain technology Hyperledger Fabric will enable competing parties to achieve consensus for data exchange. Introducing blockchain allows for smart contracts to act as an immutable means of consensus for parties in determining data exchange. The privacy-preserving features of Hyperledger Fabric could enhance the value of a system, ensuring the security of each parties data, as well as potentially being more legally compliant.

Blockchain and its immutability can provide a viable means for competing parties to achieve consensus for data exchange. This will enable parties with minority stakes in market data to garner better insights into their respective markets and compete against dominant players.

The consumer has as much to gain as a business. Providing a more complete view of the market to smaller parties allows for greater competition and creativity within a market. The enhanced view may open up new avenues of revenue and provide creative solutions to opportunities that may have previously gone

unnoticed. It is envisaged that the greater the level of access smaller entities have to data, the greater a range of options that can be provided for the consumer.

Beyond business, there is also added values to education, healthcare, or any data-dependent institution. It will enable these institutions to share data in a consensual manner. This increase in data will enable the institutions to increase their understanding of the field they are operating within.

Two aspects of intended future work is outlined below:

Private Data The focus of this work is on achieving consensus between competing entities. Another element addressed was data privacy. Data privacy ensures the adoption of the model, as it presents one less concern for parties who participate. It also appeals to entities who would have benefited from multi-party data exchanges previously but were restricted by their inability to share this data, due to it containing sensitive information. Future developments of this model will require data privacy. Privacy was ultimately a deciding factor in why Hyperledger Fabric was chosen as the blockchain technology.

Federated Learning Machine learning is a rapidly growing field that is entirely data-oriented. Access to larger and more varied data could generate more accurate models. Achieving consensus between smaller entities could allow them to bring more to the field of machine learning. Federated learning is an application that could benefit from this design of data exchange model.

Institutions and business alike would benefit from this. The institutional sharing of data would allow researchers to solve issues that would have been previously restricted by access to data and the ability to share it.

References

1. Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., De Caro, A., Enyeart, D., Ferris, C., Laventman, G., Manevich, Y., et al.: Hyperledger fabric: a distributed operating system for permissioned blockchains. In: Proceedings of the Thirteenth EuroSys Conference. p. 30. ACM (2018)
2. Buterin, V.: On public and private blockchains (2015)
3. Clifton, C., Kantarciolu, M., Doan, A., Schadow, G., Vaidya, J., Elmagarmid, A., Suciu, D.: Privacy-preserving data integration and sharing. In: Proceedings of the 9th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery. pp. 19–26. ACM (2004)
4. Hofacker, C.F., Malthouse, E.C., Sultan, F.: Big data and consumer behavior: Imminent opportunities. Journal of Consumer Marketing 33(2), 89–97 (2016)
5. Mauro, F., Stella, D.: Brief overview of the legal instruments and restrictions for sharing data while complying with the eu data protection law. In: International Conference on Web Engineering. pp. 57–68. Springer (2016)
6. McMahan, B., Ramage, D.: Federated learning: Collabvest2010healthorative machine learning without centralized training data. Google Research Blog 3 (2017)
7. Peterson, K., Deeduvanu, R., Kanjamala, P., Boles, K.: A blockchain-based approach to health information exchange networks. In: Proc. NIST Workshop Blockchain Healthcare. vol. 1, pp. 1–10 (2016)

8. Smith, V., Chiang, C.K., Sanjabi, M., Talwalkar, A.S.: Federated multi-task learning. In: Advances in Neural Information Processing Systems. pp. 4424–4434 (2017)
9. Vest, J.R., Gamm, L.D.: Health information exchange: persistent challenges and new strategies. Journal of the American Medical Informatics Association 17(3), 288–294 (2010)
10. 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). pp. 557–564. IEEE (2017)
11. Zyskind, G., Nathan, O., et al.: Decentralizing privacy: Using blockchain to protect personal data. In: 2015 IEEE Security and Privacy Workshops. pp. 180–184. IEEE (2015)