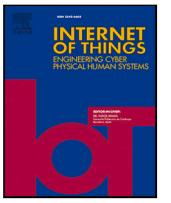
Internet of Things 22 (2023) 100722

Contents lists available at ScienceDirect

Internet of Things

journal homepage: www.elsevier.com/locate/iot



Research article

On the use of Blockchain to enable a highly scalable Internet of Things Data Marketplace

Víctor González *, Luis Sánchez *, Jorge Lanza, Juan Ramón Santana, Pablo Sotres, Alberto E. García

Network Planning and Mobile Communications Laboratory, Universidad de Cantabria, Plaza de la Ciencia s/n, Santander, 39005, Spain

ARTICLE INFO

Keywords:
Internet of Things
Data Marketplace
Blockchain
Smart Contracts
Data subscription
Performance evaluation

ABSTRACT

As data becomes the new fuel of the economy and a key asset to address our societal challenges, we cannot afford to have the data of businesses, public sector and individuals stored and kept unexploited or, what is worse, exploited by others that actually have the resources and capacity to do it. This is affecting not only our economic performance but also our security, safety and sovereignty. Among the plethora of data sources that exist nowadays, the Internet of Things (IoT) is recognised as a game-changer technology that expands its applicability to a huge variety of domains. Its main asset is, precisely, the data that the myriad of sensors embedded in the environment are constantly generating. The diffusion of platforms for IoT data sharing and monetisation is one of the key success factors which may help to drive the data economy and industrial transformation. In this paper, we are presenting a data sharing platform based on Blockchain, so-called Blockchain-based IoT Data Marketplace (BIDM) over which data producers and data consumers are able to share data in a decentralised and trustworthy manner. The BIDM enables a data marketplace where owners of IoT infrastructures can expose the observations that their devices generate while retaining control over who accesses each observation and directly getting revenues according to the price they have set. The evaluation that we have carried out of the BIDM's behaviour and performance in terms of operational execution times and scalability has been the basis for the discussion that we are presenting on the shortcomings that are typically associated with the use of Blockchain technologies as enablers for data marketplaces. This discussion also includes the evaluation of the challenges that must be considered for the creation of secure and interoperable Data Spaces based on Blockchain.

1. Introduction

The Economy of Data is growing very rapidly and, for example, in the EU area, it is estimated to be 800 billion Euros by 2025 [1]. Thus, significant focus should be given on data management techniques that can boost accessibility and discoverability of high quality, secure and privacy-preserved data. Among the plethora of data sources that can be identified, the Internet of Things (IoT) is particularly increasing its importance due to the large number of different application domains on which it is currently being employed. IoT-generated data has the potential to offer high value for richer smart services, but is not released for exploitation or not readily available.

* Corresponding authors.
E-mail addresses: vgonzalez@tlmat.unican.es (V. González), lsanchez@tlmat.unican.es (L. Sánchez), jlanza@tlmat.unican.es (J. Lanza), jrsantana@tlmat.unican.es (J.R. Santana), psotres@tlmat.unican.es (P. Sotres), aegarcia@tlmat.unican.es (A.E. García).

https://doi.org/10.1016/j.iot.2023.100722
Received 25 August 2022; Received in revised form 9 January 2023; Accepted 11 February 2023
Available online 18 February 2023

Firstly, the IoT solutions' landscape is still largely fragmented, with vertical platforms dominating and horizontal standards more mirroring experimental petri dishes rather than solutions to real problems [2,3]. In this sense, the IoT is still suffering from a major challenge that has not been sufficiently addressed already, which is the siloed approach followed in most of the scenarios. Many times, the owner of the IoT infrastructure generating the data is, at the same time, the service consumer. It is not easy to find IoT platforms that allow the exchange of data generated by heterogeneous providers and demanded by heterogeneous consumers. This situation leads to below critical-mass efforts in standardisation and commodity solutions. The landscape is deeply affected by the concerns raised by potential vendor lock-in, resulting in lowering users' confidence that smartisation strategies may achieve a major change [4]. Conversely, the fragmentation of emerging IoT-enabled platforms makes it difficult for entrepreneurs and Small-Medium Enterprises (SMEs), which fear platform lock-in, to achieve economies of scale by replicating innovative solutions from one scenario to another, even in the same application domain [5].

In addition to the interoperability issues, from a trustworthiness standpoint, data is, most of the time, only shared if not considered sensitive. As a result, success and actual value of services that make use of available data is often only modest [6]. A lot of data is still not shared because its owners, while conscious that it would be appreciated by potential consumers, do not have the skills to become themselves data providers. The alternative that they have is to use IoT platforms from third parties. However, they are not confident on how their data might be used, exploited or even monetised by the owners of those IoT platforms. There is a lack of incentives, market confidence and trust for organisations and individuals to share and economically exploit new datasets as the key ecosystem foundation for such a marketplace is still missing. It is necessary to develop a secure-by-design approach increasing the transparency of all the IoT assets governance flow shifting from the current paradigm of discrete centralised trusted authorities to a paradigm of liquid and decentralised trust.

Thus, there is a need for overcoming these major hurdles. On the one hand, we must avoid lock-in situations in which any of the ecosystem's stakeholders could gain a force position and discourage the rest. On the other hand, we must establish the mechanisms to promote an ecosystem in which as many relevant actors as possible are attracted. To enable this scenario, it is necessary to set up a platform that allows the design and implementation of novel services and, why not, so-called killer applications. A crucial aspect that has been highlighted is the need to include enablers associated to the incentives and rewards of participating, not only during the trial phases but also considering the long-term sustainability of the whole platform and IoT ecosystem [7].

In this shift, the creation of Blockchain platforms that can be employed following the "as a Service" paradigm will enable fully decentralised solutions delivering certified mechanisms to support the management of IoT devices, during their whole lifecycle. The blockchain-based distributed ledger concept can support provenance and quality-of-data guarantee, as well as reputation mechanisms for qualification of shared assets or traceability of data. The shifting from the current paradigm of discrete centralised trusted authorities to a paradigm of decentralised trust of the web as a whole, responds to the consumers' needs and to the providers' demands. The former because they require reassurance of the quality of the data. The latter because they are reluctant to share some datasets due to uncertainty of how and for what purposes the data is used.

In this overall context the work described in this paper aimed at answering the following Research Questions (RQ):

- RQ1: How to support flexible and trustworthy exchange of data within IoT data marketplaces leveraging Blockchain technologies?
  - RQ1_1: What have other proposed data marketplaces overlooked?
  - RQ1_2: What features of Blockchain technologies can be leveraged to overcome the limitations of existing proposals?
  - RQ1_3: What procedures have to be designed and developed to support the goals of a decentralised and trustworthy IoT data marketplace?
- RQ2: How to assess the potential performance degradation associated to the use of Blockchain technologies for the support of trustworthy IoT data exchange?
  - RQ2_1: What is the impact, in the time domain, of employing Blockchain technologies to support of trustworthy IoT data exchange?
  - RQ2_2: What is the evolution of the system performance with time and usage?

This paper describes in detail and assesses the performance of a Blockchain-based IoT Data Marketplace (BIDM). The BIDM creates an ecosystem on which transactions of IoT data can be supported, initially, establishing a price for each observation, but also with the possibility for extending the requirements to be fulfilled by the buyer (reputation, purpose, etc.) in order to have access to the observations.

The key contributions presented in the paper are: (i) the analysis of the current research in the field of data marketplaces in order to put the BIDM in context and pinpoint the key aspects and functionalities from the BIDM that have not received sufficient attention in previous solutions; (ii) the design and specification of the BIDM to support the secure and transparent exchange of data both in the form of datasets (i.e. synchronous access to specific observations) and data streams (i.e. asynchronous access to observations as soon as they are generated). In this respect, the paper presents the key design considerations, the functional architecture of the platform behind the BIDM and the main procedures that will take place in it; and (iii) the evaluation of BIDM's performance when handling the high loads that it has to support when integrated with large-scale IoT infrastructures so that it can demonstrate its highly scalable behaviour when faced against highly-demanding conditions.

The paper is structured as follows. Section 2 presents a review of some related works in the field of Blockchain-based data marketplaces. The focus is put on describing how the key aspects of the BIDM are positioned compared with previously proposed

solutions. This review helps in identifying the novelty and relevance of the proposed work. In Section 3 the BIDM architecture is described and its main functional components presented, with a description of the step-by-step data registration and purchase procedures. In particular, Section 3 describes the smart contracts, which are at the core of the BIDM, and support the main parts of the aforementioned procedures. Moreover, a qualitative analysis of the BIDM is presented. Section 4 presents the results of the performance evaluation that has been carried out over the BIDM. This evaluation has focused on assessing the behaviour of the BIDM in terms of execution time and scalability of the solution under heavy workloads. Finally, Section 5 concludes the paper highlighting the main contributions and the key results derived from the analysis performed.

2. Preliminaries

We first present the running case about a smart city Digital Twin (DT) use case in Section 2.1. Next, Section 2.2 reviews existing literature in order to settle the context in which the work described in this paper has been developed, thus providing a reference for the key contributions of this work.

2.1. Running case

The Smart Cities are complex systems of systems where different, but intertwined, stakeholders participate. Digital Twins are digital representations of real-world entities or systems that mirror physical objects, processes, organisations or persons. For this representation to be created, it is necessary to capture information from the physical world and make it available for processing.

Fig. 1 depicts the case of a DT for a Smart City scenario. In this scenario two main kinds of users are included. Firstly, producers can be citizens wearing some fitness bands or using their smartphones, local utilities managing public services or vehicles circulating on the city equipped with novel electronic devices (sensors and actuators). All of them are continuously creating pieces of data about its status, position and/or behaviour. Secondly, consumers are service providers that build their services on top of the available pieces of information that producers share. These service providers need to have access to as much information as possible from the city so that they can process it to smarten the way in which current city services are provided or to provide new smart services.

The DT is a specific instance of a consumer that gathers data from several producers so it can create the corresponding representation of the city as an aggregation of data streams coming from virtually everywhere. However, producers will only be willing to share their data if some kind of incentive and/or revenue is granted to them. Otherwise, they will keep their data locked. The most common situation to date is that the DT provider settles a dedicated platform and establishes agreements with the producers to have access to their data, typically, in an unrestricted and closed manner. This way, producers lose control of their data as soon as they inject it into the consumer's system.

The data marketplaces are meant to operate on a different business model where producers can offer their data to different consumers in parallel. However, among the different technical challenges that must be addressed in order to realise the data marketplace, the establishment of trustworthiness between producers and consumers is critical. In this sense, it is of utmost importance to fulfil the following design considerations: (1) guarantee that data is provided by reliable IoT producers; (2) guarantee that customers pay for the data at a specific instant in time; (3) guarantee that customers receive the pieces of data (i.e. measurements generated by the sensors embedded within the physical world) that they bought; and (4) guarantee secure and authorised-only access to the data. By achieving these features, a trustworthy system for both consumers (i.e. the marketplace customers) and producers (i.e. the sellers at the marketplace) is created.

This paper reports on the design and performance evaluation of the BIDM, which resolves the trustworthiness-related challenges for realising Fig. 1 running case.

2.2. Related work

IoT data marketplaces have already been a research topic for the past years. In fact, it is an evolution of the well-studied Sensing as a Service (SaaS) model [8] in which datasets and data-streams generated by IoT deployments are exchanged so that context information is served to the applications that need it using a service-oriented architecture. However, in most of the cases the proposed data marketplaces follow a centralised architecture mimicking the existing model of large cloud platforms that gather everything on a single central point. For instance, [9,10] make use of this type of architecture. In these scenarios, the central entity is the one that decides how products are sort in a client's search, the distribution of benefits between the data sellers and the platform operator, or the algorithm behind the review system. Both [9,11] are recent works that follow this approach. The former presents a conceptual design while the latter introduces a fairness parameter that applies to all data on sale and presents a mathematical model of its operation. The state of the art of this type of solutions has already been reviewed in works like [12,13]. They have made systematic reviews of articles and literature published on the topic, performing methodical analyses of different parameters such as kinds of data producers, types of monetisation or most common data sources.

However, data marketplaces that follow this approach have several important shortcomings that, due to their centralised nature and governance structure, they cannot overcome. Some of these limitations and the challenges that they imply are:

- Data consumers acquiring data from the marketplace will typically be paying extra charges for every piece of data in order to compensate platform expenses.

Fig. 1. Conceptual running case for a Smart City Digital Twin.

- Data producers offering their data through the marketplace lack the certainty that their data is being treated fairly when it comes to searches and promotions (i.e. how the marketplace organises the "storefront", materialised in this case as the algorithm responding to data look-ups) and do not have control about the actual value of their data (i.e. how much benefit the marketplace gets from the data).
- Stakeholders need absolute trust in the proper behaviour of the platform. If consumers claim that they have not received a measurement that they have just bought, the platform is in charge of dealing with all possible scenarios, whoever the fraudulent user may be.
- The platform has access to all data, since it is in control of the payment system. If the platform itself decides to be dishonest, then the data marketplace immediately should stop working, and that is the best-case scenario; the data marketplace may keep operating fraudulently without users' knowledge.

As a result of these shortcomings, more recently, the focus has been put on decentralised solutions. In this paradigm shift, Blockchain is being employed as a crucial part of the proposed systems. Blockchain is a decentralised ledger where several nodes, connected in a peer-to-peer network, keep a record of the transactions made by users. This record is stored in blocks securely linked together. Thus, Blockchain-based data marketplaces have emerged following this decentralised philosophy, and are generating increased attention within the research community. In some works, only the essential ideas and concepts of decentralised Blockchain-based data marketplaces are proposed together with an initial design of their platform [14–17]. Other authors narrow down the scope of their solutions and specialise in IoT data marketplaces [18–20]. However, most of the existing works, as it can be seen in Table 1, have not taken the solutions that they propose into an actual implementation, and even less have accomplished a performance assessment of such implementations as we are doing in this article.

Nevertheless, it is still highly relevant to briefly discuss and analyse these and some other existing works that introduce novelties and aspects that are related to the BIDM that we are presenting and evaluating in this article. The work presented in [21] centres its design on the trading of anonymous private data focusing on the compliance with GDPR. Conversely in [22,23] the attention is put on the review system for data consumers to rate the data that they purchase, or data consumers and data providers to rate one another after a transaction. The data usage by the consumers is also a subject of interest that is worked upon in several articles. For instance, [24] proposes a data marketplace focusing on the distribution of massive amounts of data for machine learning applications, using trusted execution environments. In [25], the authors work towards the merchandising of medical data from

Table 1
Comparative of data marketplace proposed solutions.

Reference	Blockchain	Consensus	Storage	Implem.	Cost analysis	Perform.	General use	Monet.
I3 [9]	✗	N/A	Centralized	✔	✗	✗	✔	✗
Mišura et al. [10]	✗	N/A	Centralized	✔	✗	✗	✔	✗
Agarwal et al. [11]	✗	N/A	N/A	✗	✗	✔	✔	✔
Gupta et al. [14]	✔	N/A	Distributed	✗	✗	✗	✔	✗
Yoo et al. [15]	✔	N/A	Centralized	✗	✗	✗	✔	✗
Banerjee et al. [16]	✔	N/A	Centralized	✗	✗	✗	✔	✗
Lawrenz et al. [17]	✔	N/A	Centralized	✗	✗	✗	✗	✗
Kanhere et al. [19]	✔	Unknown(Besu)	Distributed	✗	✗	✗	✔	✗
Ramachandranet al. [20]	✔	IOTA(Tangle)	Distributed	✔	✗	✗	✔	✗
Ha et al. [21]	✔	PoA(Luniverse)	Distributed	✗	✗	✗	✗	✔
Park et al. [22]	✔	PoW	Distributed	✔	✔	✗	✔	✗
BlendSM-DDM [23]	✔	N/A	Distributed	✗	✗	✗	✔	✗
Sterling [24]	✔	Unknown	Unknown	✔	✗	✗	✗	✗
Alsharif et al. [25]	✔	PoA(Kovan)	Distributed	✔	✔	✗	✗	✗
Agora [26]	✔	PoA(Rinkeby)	Centralized	✔	✔	✔	✔	✗
Badreddineet al. [27]	✔	None(Ganache)	No Storage	✔	✗	✗	✔	✔
BIDM	✔	PoA(Clique)	Distributed	✔	✗	✔	✔	✗

anonymous sources. Finally, other works put their focus on the monetisation model that should be used to attract users to their platform [26,27].

Table 1 summarises some key aspects of all these works, which are not meant to be an exhaustive survey like others that have been already published [28–30], but are representative of the trends in this kind of solutions and the key challenges that are addressed by the BIDM. Thus, it qualitatively assesses the contributions of our work by comparing it with other solutions that have been proposed, thereby setting the context for the value propositions that make this work a significant step forward towards analysing the feasibility and performance of Blockchain-based distributed data marketplaces.

Specifically, the analysis has looked, firstly, at whether the proposed IoT data marketplace makes use of Blockchain technology. In the cases where the proposed solution is a Blockchain-based marketplace, the second parameter analysed, "Consensus", is the consensus algorithm used for the creation of the blocks. As it can be seen in Table 1, in many cases, this aspect is not applicable or not available because the existing work lacks an actual implementation or the specification published does not reach such level of detail. It is important to highlight this aspect since it clearly shows the additional value and progress of the state of the art of the work described in this article, and the relevance of such algorithm in aspects such as energy consumption. "Storage" refers to the approach followed to support data persistence. In this case, the analysis carried out has not delved into the specific technology used but rather whether the proposed marketplaces store data in a centralised or distributed manner (or if data is stored at all). Moreover, it is important to highlight that, for all the cases, data is stored in an off-chain manner, keeping in the Blockchain only the metadata necessary for retrieving the actual data. With the "Implem." feature we have examined whether the proposed data marketplace has been actually materialised in some form or if it has been simply described and specified up to a conceptual system.

Pertaining to the evaluation of the solutions reviewed, we have focused on two completely different aspects. On the one hand, "Cost analysis" indicates if the system has been evaluated from the point of view of the resources consumed in their transactions (e.g. gas in Ethereum), which translates to the cost associated with the use of a Blockchain, using this evaluation as the basis to support the data marketplace operation. In this regard, it is important to mention that this cost is heavily dependent on the consensus mechanism used. Thus, for some of the proposed solutions, this analysis is not applicable as the Blockchain technology that they use do not have such a cost (or it is not directly comparable with others that use public Blockchains, for instance). On the other hand, "Perform." refers to a more general evaluation, either theoretical or experimental, made to assess the computational behaviour and performance of the developed data marketplace.

Regarding the scope of the reviewed marketplaces, our analysis has also covered whether the data on sale through the marketplace can be of any nature or if the proposed system is restricted to a specific domain. This is indicated in the "General use" column. Finally, "Monet." shows if the proposed solution has described or implemented a business model for monetisation of their platform and/or the data traded through it.

Considering the review made, some conclusions can be derived. Firstly, while there have been several proposals that can be analogous to the BIDM, very few have gone beyond a detailed design of their platforms, and even less have provided an estimation of their performance. Besides, for those who have completed an actual implementation of their marketplaces, the evaluation has been usually centred in the transactions' cost in Blockchain technologies, meaning that the basic operational performance is neglected while the economic dimension is put on the main stage.

All in all, as it has been already stated, this review sets the context for the key contributions of the work presented in this paper, namely: (i) a complete implementation of the data marketplace, proving that the proposed design performs as intended when put into practice; (ii) a performance analysis of the BIDM focusing on the use of computing resources, specifically execution time and disk usage; (iii) a comparison among the two operation modes supported by the BIDM, both qualitatively and quantitatively; and (iv) an integration with real IoT data providers, regardless of their inner workings.

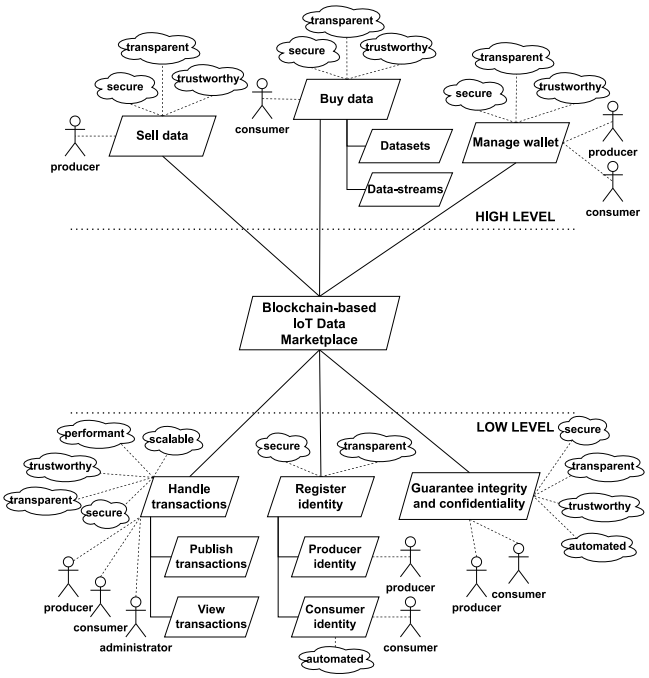


Fig. 2. AOM of the BIDM's requirements.

3. System model

3.1. BIDM key requirements

Before delving any deeper into the actual BIDM's architecture it is necessary to summarise the key functional requirements that were considered in the architecture's design. Fig. 2 shows the Agent-Oriented Modelling (AOM) elements used for deriving the requirements of the BIDM. The elements are based on standard AOM principles [31].

As it was introduced in the running case presented in Section 2.1, there are four key design considerations that must be taken into account in order to enable trustworthy data transactions between producers and consumers within an IoT data marketplace.

Firstly, the data marketplace has to be a neutral actor in the scenario in which both types of users, producers (i.e. sellers) and consumers (i.e. buyers), have equal rights. Indeed, the so called "prosumer" role (i.e. a combination of producer and consumer) should be possible. Thus, it should be implemented using decentralised technologies for shifting from the current paradigm of discrete centralised trusted authorities to a paradigm of liquid trust. It is important to highlight that the marketplace will always play a central role, which should not be mistakenly considered a centralised behaviour. On contrary, it keeps its decentralised operation as long as it employs a technology that avoids hierarchical trust relationships among the actors (i.e. marketplace, producers and consumers).

Next, the access to the marketplace has to be open but there should be mechanisms to identify the participants, both the consumers and the producers. This way not only the "buyers" (i.e. the DT in the running case we are using as an example of BIDM's operation) are identified but also the "sellers" (i.e. all the potential IoT data producers) are controlled.

Moreover, there has to be an immutable record for each data transaction. In this sense, not only the DT has to be able to demonstrate that the corresponding price for the obtained IoT measurement has been paid, but also the producers have to have the means for proving that the measurement was made available to the DT on time and form. Additionally, the BIDM has to protect the pieces of data that producers make available (and announce on the marketplace) so that only authorised consumers (i.e. those that have paid for them) can get them.

Last but not least, given the permanent flow of data registrations (i.e. the process by which producers make new IoT measurements available) and data consumptions (i.e. the process by which consumers get one of the already registered IoT data measurements), it is necessary that the BIDM does not imply a bottleneck from a non-functional point of view. This is, the technological solutions developed to address the aforementioned requirements must show a good performance and scalability.

3.2. BIDM functional architecture

In this section we outline the BIDM's core architecture and its main functionalities as they have been defined to satisfy the requirements from both the data providers and the data consumers previously outlined in Section 3.1.

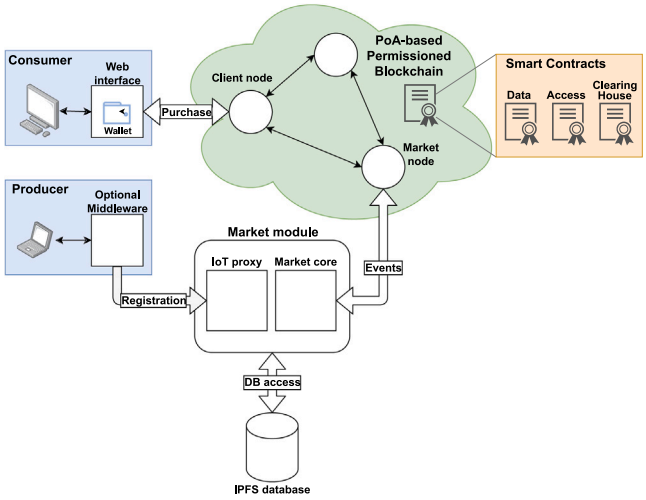


Fig. 3. Structure of the data marketplace.

The BIDM is organised around a permissioned Blockchain using a Proof-of-Authority (PoA) consensus protocol (e.g. an Ethereum Blockchain using Clique [32] consensus protocol has been used for the BIDM's implementation whose evaluation results are presented in Section 4). We opted for the use of a PoA consensus protocol because it avoids unnecessary computational costs by reducing the mining difficulty to a minimum. This allows eliminating the need for fees which add extra cost to the mining of blocks. In a PoA-based network, identity is used as a way to allow nodes to validate and sign blocks. Clique allows for current validator nodes to vote on the addition or removal of new validators. More than half of the current validator nodes are required to agree before any change is made. This way, we can control the validator nodes in the development phase, and easily move to a fully decentralised production deployment through Clique's proposals. Additionally, we think that the Blockchain ought to be permissioned (private or public depending on the needs of the owner of the platform) so that all the nodes have the same level of trustworthiness. Finally, by using PoA instead of other consensus mechanisms such as Proof-of-Work, it is possible to fix the amount of time between two mined blocks. This is important because given the mainly asynchronous (i.e. event-based) nature of the IoT, this feature helps minimising the number of empty blocks in the Blockchain. It is important to highlight that the use of a permissioned Blockchain using a PoA consensus protocol does not go against the decentralised nature of the BIDM. Even if PoA-based Blockchains establish restrictions on which nodes participate in the consensus (i.e. generate new blocks), all the information is available to all the nodes joining the Blockchain, which is the key feature to guarantee decentralisation of information and overall trustworthiness.

The main role of the Blockchain is to register the key information about the new measurements generated by data producers and about the purchases made by data consumers. Moreover, it logs other aspects such as the creation of new user accounts or the transfer of tokens (i.e. transference of credit from buyer to seller). Finally, it hosts the Smart Contracts (SCs) that rule every operation taking place within the BIDM.

Fig. 3 shows all the components that are organised around the central structure that are the nodes supporting the Blockchain. The rest of components and/or entities that are part of the BIDM's architecture are, on the one hand, its users (whether data consumers or data producers) and, on the other hand, the sub-system in charge of the off-chain data storage. In this regard, users will have to create a Blockchain account through any of these nodes. In Fig. 3, for the sake of clarity, we have called Client Node to the Blockchain nodes in which Data Consumers create their accounts and Market Node to those that are used by Data Producers and the data storage sub-system to offer their observations through the BIDM, but they can take both roles at the same time.

Elements pertaining to the Blockchain are represented in green, namely the nodes and the SCs. The three SCs that gather the main functionalities of the BIDM can be seen in orange, and they are named based on their functionality: Data SC, Access SC, and Clearing House SC. The specific content of these contracts will be explained later.

Moreover, the users of the Blockchain are represented in blue, both a Data Producer (acting as seller in the marketplace) and a Data Consumer (acting as buyer in the marketplace). Consumers have a web API at their disposal, which they can use to interact with the Blockchain and the SCs in order to purchase measurements, see their value when purchased, or check their account balance, to name a few examples. The web API interacts through a Blockchain node to manage the clients' accounts. On the other end, producers have the possibility to include a middleware whose function is to adapt their data model to the one used in the BIDM.

Finally, the Market module and the InterPlanetary File System (IPFS) Database are the components used by data producers to inject their measurements into the BIDM and to store them, respectively. The Market module is the only component with administrator credentials in the Blockchain, so it can perform special operations on Smart Contracts. It is responsible, among other things, for receiving requests from producers to register new measurements in the marketplace, completing the purchase process initiated by consumers, or adding new accounts to the BIDM. In addition, it has access to the IPFS database both to store new measurements and to retrieve them when requested. This module is divided into two smaller components: the IoT proxy and the Market core.

3.2.1. Smart contracts

Smart contracts, which perform transactions on the Blockchain at users' request, are instantiated during the BIDM's start-up, generating an Ethereum address that is made public for easy access. Both the web API and the Market module are aware of these addresses in advance and can interact with the SCs automatically. There are three SCs, each one in charge of a part of the BIDM's functionality.

The main purpose of the Data SC is the management of the observations from IoT devices that are registered within the Blockchain. It is responsible for registering new measurements in the Blockchain, as well as returning the details about registered measurements and their respective sellers. In addition, it allows an IoT data producer to register new Topics in the Blockchain. These Topics describe a basic characteristic of a measurement (e.g. generated by sensor X, temperature measurement, etc.) so that any observation tagged with one of these Topics is implicitly part of a stream uniquely identified by that Topic. As it will be described in Section 3.3.3, the Topics are key to support the subscription-based model of the BIDM so that IoT consumers are able to have access to data-streams through the BIDM.

In addition, Topics have a secret key associated to it, which is also registered during the Topic creation. This key is used to encrypt all the measurements in that Topic before they are stored in the database. This way, the access to those measurements is restricted to the consumers that have a valid subscription to that stream of measurements. Access to these keys is only allowed for the platform administrator and the producer that generated them. We will explain more about the Topic keys when discussing the IoT Proxy component.

The Access SC is responsible for managing the customers' Blockchain accounts. Its functions allow producers to register or remove their IoT devices from the white list of the Blockchain. Only measurements originated from addresses in this list will be registered and sold on the platform. These functions only respond to calls from the administrator. Moreover, this contract automatically registers and publishes the public key of all users who perform transactions on the Blockchain.

This design aims for a simple yet effective access control mechanism. However, more complex trust and data access techniques such as the one presented in [33] could be integrated within the BIDM's operation.

Everything related to purchases and transactions of tokens is handled in the Clearing House SC. Consumers willing to purchase a measurement can make use of its functionalities, which register the request on the Blockchain immutably. If a user, whether buyer or seller, wants to check a measurement's price, they can do so by means of this SC. Moreover, it includes several functions for transferring tokens between accounts as well as for providing access to basic public information about the Blockchain. Besides, each Topic has a subscription price associated with it, which can be obtained freely through this contract. To subscribe to a given Topic, consumers must also do so through the Clearing House SC as it will be explained later.

3.2.2. Market module components

The Market module, containing both the Market Core and the IoT Proxy, is the main component of the BIDM besides the Blockchain, but the users are not allowed to access the Core's functionalities directly. Rather than that, interaction with the consumers happens through the SCs in the Blockchain, whereas the producers communicate with the IoT Proxy. In this sense, it is important to highlight, in order to avoid false impressions, that the decentralised nature of the BIDM is not compromised. Even when in the architecture of the BIDM, shown in Fig. 3, the Market module might appear as a central component, this is not the case in reality as the Market core functionalities of the Market module are either automatic (and open code, so every user knows what is happening and how) or accessible through the SCs that are replicated at any node participating on the Blockchain. Thus, implicitly inheriting the distributed and decentralised nature of Blockchains.

Following, we will discuss the functionalities and the logic behind the division into the two subcomponents. Firstly, the Market Core has access to an administrator account on one of the Blockchain nodes. Therefore, this component has permissions to perform any type of operation on the Blockchain. It is in charge of the communication with the SCs, it listens to the purchase events published at the Blockchain in order to complete those that are valid, and it contains the logic to ensure seamless and efficient operation of the internal processes. Additionally, the Market Core is also in charge of the storage of measurements in the database. These IoT measurements are stored in JSON format, but their structure is intentionally left open and will depend on the IoT producer.

The IoT Proxy is the second functional block at the Market module. It is the interface towards the IoT data producers and is continuously listening to their publications. This interface is implemented as a REST API that receives requests from the IoT data producers. The measurement that the producer wants to register in the BIDM is sent within the body of the request. Besides the measurement itself, additional information, such as the timestamp in which it was generated, has to be included in the POST request. After receiving a measurement, a verification process is initiated at the IoT Proxy to check whether the IoT data producer performing the registration attempt has the rights to make such registration. In this verification process, the credentials of the IoT data producer are checked for authentication and authorisation. If this verification is successful, the measurement is signed with the administrator's key and the information to be published in the Blockchain is extracted. This information includes the measurement's hash, the Topics to which the producer wants to associate the measurement with, and the URL to visualise it after a valid purchase. If it is the first time that one of these Topics is defined, the IoT Proxy automatically registers it before continuing with the measurement registration process.

The abovementioned URL is generated by the IPFS database after storing the encrypted measurement. As it was previously mentioned, symmetric encryption is performed to measurements before they are stored. The key used results from the combination of the keys bound to each of the Topics the measurement belongs to. When a measurement belongs to several Topics, which is usually the case, it will be encrypted with a key formed as the bit-wise XOR operation of all the respective Topics' keys. Every Topic's key is generated from a cryptographically secure random number generator. This is done by the IoT Proxy component

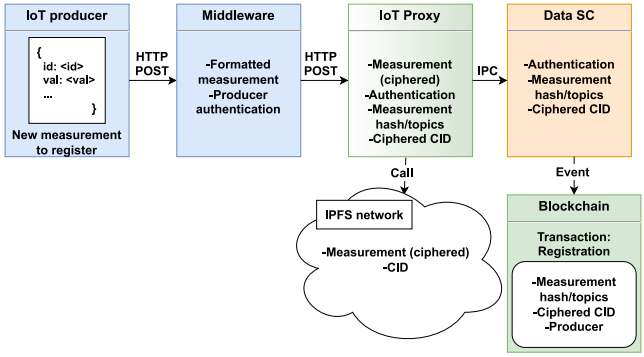


Fig. 4. Flow diagram of the registration of measurements.

whenever a new Topic is registered. It is important to note that IoT data producers can always get the keys associated to the Topics that they have created themselves from the Data SC in case they lose them from their local private keyring. The Data SC is the only place in the BIDM where the Topic keys are stored, and access is restricted to the producer that generated the Topic and the platform administrator.

Finally, the process is completed by calling two functions on the Data and Clearing House SCs respectively. The one made on the Data SC is meant to store the measurement information and the one made on the Clearing House SC aims at recording its selling price.

3.3. BIDM procedures

The functionality of the platform is, from the user's perspective, fundamentally based on the execution of operations that modify the state of the data marketplace. Considering that users can be either producers or consumers, the basic operations are the registration of new measurements by the former and their purchase by the latter, in addition to subscriptions to measurement flows or data streams. The main objective of subscriptions is to increase the flexibility of the BIDM, so that consumers are allowed to select the most convenient option for them. All three operations require the communication and cooperation between the elements covered previously in this section, and the execution of several functions in both the components and the SCs. We will now present a thorough analysis of these operations, detailing the functionalities performed by each component and the communications between them.

3.3.1. Measurement registration

The registration of measurements takes place whenever an IoT data producer wants to push IoT measurements into the BIDM. A diagram of the execution flow of this operation is represented in Fig. 4. The colour coding is kept to match the architecture description in Section 3.2. Blue boxes represent steps happening at the user side with no direct access to the Blockchain. White-coloured steps take place at the Market module. Steps that are handled at the Blockchain or at the components that have access to it through one of its nodes are coloured green. Finally, those involving the SCs are shown in orange. Elements that have a colour gradient indicate that the step is initiated in one component but another, from a different class, is also involved. Following, the measurement registration process is described in detail:

1. When a new measurement is generated, it is sent to the marketplace, directly or through the intermediate middleware if needed. This communication is performed with an HTTP POST request where the measurement is included as the HTTP message body.
2. The optional middleware is in charge of receiving the measurement and formatting it as the required JSON object containing the necessary attributes so that it can be accepted by the platform. It also extracts the producer's authentication credentials and sends it all to the IoT proxy through an HTTP POST request.
3. The IoT proxy receives the measurement and the producer's authentication credentials. This component extracts, from the received JSON object, the information about the Topics to which the measurement is related, and generates a hash of the measurement. Subsequently, it makes a call to the IPFS network to store the measurement after encrypting it to restrict access, and in response it receives its Content Identifier (CID), which is the necessary index to access the measurement in the future.
4. Next, the IoT proxy connects to the Market Node through its Inter-Process Communication (IPC) interface and calls the Data SC to register the measurement information in the Blockchain. The CID has also been previously encrypted to prevent it from becoming public information.
5. The Data SC generates an event in the Blockchain to complete the registration transaction, and the basic information necessary to securely purchase this measurement is stored in the Blockchain: the producer's Blockchain address, the measurement's hash, which is used as a unique identifier for that particular measurement, the CID to access its value in the database and the Topics

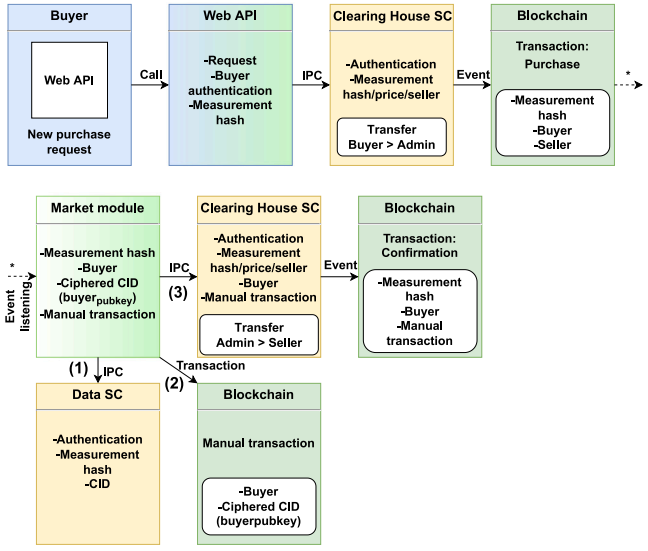


Fig. 5. Flow diagram of the purchase of measurements.

to which that measurement is associated with, so that interested consumers can be duly notified in case they had subscribed to one of the Topics previously. Since the CID is encrypted, users cannot access the measurement even if the transaction information is public, and either a purchase or a subscription is required to acquire it.

3.3.2. Measurement purchase

When data consumers are interested in a particular measurement or set of measurements, they have to initiate purchase procedure. The execution flow diagram of this operation is shown in Fig. 5. Colour code shows the component or components (in case of colour gradient) that are involved in each step of the procedure. Following, is the description of each of these steps and the transactions that they trigger:

1. Data consumers can request the measurements that they are interested in through the web API.
2. At the web API back-end, an IPC-based communication is established with the Blockchain through the Client Node. After extracting the authentication information of the client, who has to be logged in order to request the purchase, the Clearing House SC at the Blockchain is called indicating the consumer's credentials and the hash of the measurement to be bought.
3. Upon this call, the Clearing House SC generates an intermediate transfer of tokens from the consumer's account to a BIDM administrator-controlled account. This transfer generates an event in the Blockchain.
4. A purchase transaction is therefore recorded in the Blockchain, including the measurement's hash and the addresses of both seller and buyer. Subsequently, given the public nature of the Blockchain, when this new event is published, other components with access to the Blockchain can listen to it and trigger the next actions.
5. The Market Core module is the component that is listening for these purchase events. When one of them is detected, it extracts the hash of the purchased measurement and the identity of the buyer. With this information, it makes a call through the Market Node to the IPC interface of the Data SC, more specifically it looks at the fields that were registered within the measurement. With proper authentication, since it is an administrator node, the Market can obtain the CID from where the measurement can be retrieved.
6. Next, the Market Core sends a second transaction to the Blockchain, in which the buyer's identity and the CID, encrypted now with the buyer's public key so that he/she is the only one who can use it for retrieving the measurement from the database, are recorded. The generation of this transaction returns a transaction hash.
7. The Market Core then accesses the Clearing House SC through the IPC interface, where the corresponding transfer of tokens is performed from the account of the BIDM's administrator to the seller's one. If something goes wrong during the process, the purchase is revoked and the tokens are returned to the buyer.
8. Finally, one last transaction is generated on the Blockchain through an event, so-called the purchase confirmation. The information recorded in this transaction includes the measurement's hash, the buyer's address and the hash of the transaction performed to complete the payment. Once this transaction is registered in the Blockchain, the data consumer can claim access to the just acquired measurement.

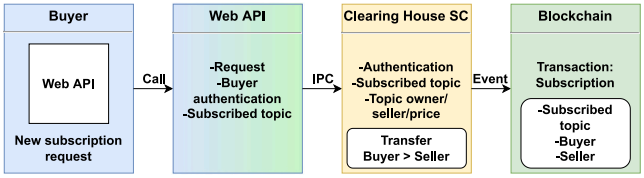


Fig. 6. Flow diagram of subscriptions.

3.3.3. Datastream subscription

While the previously described synchronous process of purchasing measurements is the basic way for data consumers to acquire IoT data, as it has been described, the purchase process is complex and requires the intervention of many of the BIDM's components. Additionally, each purchase generates three transactions on the Blockchain, which can become a significant overhead as the number of measurements and buyers grow.

Moreover, for many of the applications that will actually take the role of data consumer, the interest, rather than on specific measurements, is on the continuous streams of data that IoT infrastructures generate. This is, instead of constantly looking for specific measurements and pulling them from the BIDM, they will declare interest on one kind of data and ask for getting it as soon as such data is available at the BIDM.

Thus, as an alternative, a model of Topics and subscriptions has been implemented in the BIDM. Each registered measurement can be part of different data-streams, which are identified by a Topic. As it has been already described, when a measurement is registered, it is associated with as many Topics as data-streams the producer wants it to be part of. For example, a measurement is generated by sensor X (Topic 1: URI of the sensor or area in which the sensor is located), captures ambient temperature data (Topic 2: physical phenomenon) and was generated on February 29 (Topic 3: date). This way, consumers with, for example, interests on a particular area, phenomenon and/or date, could declare their interest on any or many of these Topics. Topics can be dynamically created so the level of granularity for the creation of different data-streams is unbounded.

The outline of the execution flow of this operation is shown in Fig. 6. The phases of this flow are explained below:

1. Data consumers, by using the web API, scroll through the available Topics and decide to subscribe to any of them. To this effect, they make a request specifying the Topic in question. This request is then received by the web API back-end.
2. The web server extracts the credentials from the buyer and the Topics of interest included in the request, and makes a subsequent call to the Clearing House SC via the IPC interface of the Blockchain Client Node.
3. The function that is run on the Clearing House SC, besides authenticating the buyer, performs a transfer of tokens from the buyer's (i.e. data-stream consumer) account to the seller's (i.e. data producer) account. As it has been already mentioned, the data producer is the one responsible for registering the Topics and for assigning them their corresponding price. Upon the transaction of tokens is made, a new event is triggered in the Blockchain.
4. Finally, a subscription transaction is recorded in the Blockchain, containing the addresses of the buyer and seller, as well as the Topic to which the former is now subscribed.

This latter transaction is used as a timestamp to find out which measurements had been generated before or after a specific buyer subscribed to a Topic. Also, it is worth noting that subscriptions only require one transaction and, from that point on, the measurements are acquired implicitly, which means that there is no need for additional transactions. Any new registered measurement, associated to a Topic that has active subscriptions, will be directly forwarded to the corresponding subscribers through their respective wallets. As it will be discussed in Section 4, this improves BIDM's scalability considerably.

In terms of fairness, as it happens with any subscription-based service, the subscriber has to pay in advance for getting real-time measurements that might or might not arrive, as it depends on whether the criteria defined in the subscription are met by the incoming measurements. However, the provider cannot block the notifications as the last transaction recorded in the BIDM acts as an immutable proof of when the subscriber got the rights to be notified. Moreover, intentionally blocking registration of new measurements on the BIDM would be incoherent for any provider which would be perceived by buyers and subscribers as a pointless investment. In fact, providers would gain credit and recognition the more measurements and satisfied buyers/subscribers they have.

3.4. Qualitative analysis of operations

Having detailed the internal behaviour of the operations, we will perform a qualitative analysis of both operating models. Both the individual purchase model and the subscription model have intrinsic advantages at a conceptual level, and the objective of the following discussion is to highlight them.

It is worth mentioning that we will be using typical use cases and situations to showcase how it may be advantageous to use one of the two models, but BIDM users are free to make use of the two alternatives in the way that they consider best fit in their applications. The rationale behind having two different models in operation is to provide clients with as much flexibility as possible. Therefore, one model does not replace the other, even if, from a performance point of view, the subscription-based one outperforms the individual purchase model.

In this sense, regardless of the results that are obtained at the performance analysis that will be presented in Section 4, at a functional level both models are of interest to potential consumers looking for a specific way of using the BIDM. Thus, this comparison is not meant to be a confrontation between the two models to find out which one is better, but rather as showing which are the benefits that the BIDM offers to its users by putting at their disposal two different alternatives for acquiring IoT data. In this sense, the two models seen here work in cooperation to provide the best possible service. This comparison is therefore a compilation of situations in which the use of one model or the other may be beneficial to the consumer.

3.4.1. Individual purchase model

Following, we describe the key features that the specific measurement acquisition model provides to the BIDM as well as the motivation for supporting it within the BIDM operation.

- *Interest in specific measurements*: The purchase of individual measurements is essential when the consumer needs to have information about a specific phenomenon, at a specific place and time. This case can occur in a plethora of situations, for instance, to have information about a particular physical phenomenon at the current moment measured by a sensor located in a specific place.
- *Reduced price*: IoT producers are the ones who decide the price of their measurements and subscriptions, but most likely individual measurements will always cost significantly less than a subscription to the data-stream to which they are associated, since the latter enables access to multiple measurements. Consumers can adapt their consumption pattern so that their purchases in the platform are made in the most capital-efficient manner.
- *Instant purchase*: Another possible situation a consumer may find themselves in is the need to access measurements quickly. In a subscription, the buyer has to wait for the producer to generate the measurements corresponding to that data-stream. In contrast, when purchasing individual measurements, the data is available the instant the purchase is made, although it will be a measurement generated prior to the purchase.
- *Historical measurements*: As already mentioned, the subscription-based model allows access to all the measurements of a given data-stream that are generated after the subscription. However, with this mechanism it is not possible to acquire past measurements. If a consumer is in need of historical data, the individual purchase model is the only possible option.

3.4.2. Subscription model

Following, we have identified the key features that the subscription-based model provides to the BIDM together with the motivation for supporting it within the BIDM operation.

- *Interest in a specific kind of measurements*: As in the previous model, there is a use case that is the main motivation for the subscription-based model supported by the BIDM. IoT data consumers interested in a stream of measurements, such as real-time temperature at a given location, can subscribe to that data-stream to receive the measurements as soon as they are generated.
- *Convenience*: One of the advantages of the subscription model at a functional level is that, once the subscription is made, the buyer does not have to do anything other than being prepared for handling the subsequent notifications that are triggered when a new measurement is registered and tagged with the Topic to which the data consumer has just subscribed. However, in the individual purchase model, the consumer has to go through the purchase process each time they want to acquire a new one.
- *Efficiency*: Just as in the previous case it was reasonable to assume that individual measurements will be cheaper than subscriptions, it is also foreseeable that producers will establish prices for subscriptions so that they are profitable for the consumer in the long run. Otherwise, any informed customer would simply avoid using subscriptions. Therefore, buyers interested in making efficient use of economic resources in the long run will find the subscription model to be the best option.
- *Scheduling and planning*: Typically, buyers will use measurements within the logic of their own applications. The best way to ensure consistent data flow without the need for user interaction is a subscription. In this way, buyers can schedule and automate processes that use the measurements they are subscribed to as an input.

4. Performance evaluation

Once presented the architecture and key functionalities of the BIDM, in this section we present the results of the evaluation that we have carried out over the actual implementation of the marketplace. Through both synthetic and real IoT data sources, we are emulating the flow of IoT data from one or many producers which are registering new IoT measurements continuously. Analogously, we also emulate the purchase process that the DT would have to perform in order to keep an up to date digital representation of the real world.

As it is presented in Table 1, while previous works have already proposed the use of Blockchain technologies to support data marketplaces, there has not been, to the best of our knowledge, an evaluation of the potential bottleneck that using this kind of technology could imply in the marketplace performance. Moreover, as it has been indicated in Section 3.1, besides the functional requirements, it is equally important to satisfy some non-functional needs, from which execution time and scalability are two major ones. Thus, this evaluation has been focused on assessing BIDM's performance under heavy load conditions that could, for example, mimic the running case scenario depicted in Section 2.1.

In particular, we have analysed the time taken to accomplish the main procedures in the BIDM (i.e. the registration of measurements coming from an IoT infrastructure, and the purchase of one of these measurements by a data consumer). Additionally,

Table 2
Number of samples.

Inter-arrival time	SmartSantander	Length	Scalability
5,000	10,000	5,000	>500,000

we have also analysed the BIDM footprint in terms of disk space. These two key parameters have been assessed considering the BIDM's behaviour in the long term in order to evaluate how scalable the proposed solution would be. The Blockchain setup implemented for the evaluation has three nodes, of which one acts as a validator and signer. The other two act as a Client node and a Market node as described in the architecture. Most of the configuration depends on the specific analysis scenario and will be detailed separately. The source code of the implemented BIDM is publicly available in GitHub.[1]

The main objective of the evaluation carried out is to be able to conclude if, besides the functional qualitative benefits that have already been described in Section 3, the BIDM could actually be employed in real-world scenarios and support the exchange of large amounts of data. For this, it has to, firstly, avoid introducing excessive execution times to the data registration and purchase procedures; secondly, keep the storage size footprint as controlled as possible; and, thirdly, maintain these parameters constant as the amount of data exposed through the marketplace increases.

For the evaluation, we have used mostly synthetic data sources and consumers on which we could control the measurement generation and purchase patterns (mainly in terms of the size of the measurements and the amount of time between consecutive events). However, we have also carried out the analysis using the SmartSantander smart city IoT infrastructure [34].

For this purpose, we have developed a system that connects the BIDM with SmartSantander and registers the IoT platform that manages the IoT deployment in Santander as one of the BIDM's data producers. We have used the subscription API offered by SmartSantander [35] in order to implement the integration to the BIDM. The role of the middleware is to receive the measurements generated by the SmartSantander infrastructure, map them to the data model used by the BIDM, and register them in the platform through the Market module, or more specifically, through the IoT proxy component.

4.1. Time domain behaviour: Procedure execution time

The main objectives of the evaluation that we carried out were to assess the performance of the BIDM in terms of the execution time of each registration and purchase operation, as well as to evaluate if this performance is stable as the amount of operations carried out over the BIDM increases with time.

In this section we are presenting the results of the evaluation of the measurement registration and measurement purchase execution time, as well as the analysis in terms of time of the measurements' subscription-notification procedure. These results were obtained after executing a large number of the aforementioned procedures. The number of consecutive executions of each experiment is summarised in Table 2. In this regard, we have analysed how the time between two consecutive procedures (whether they are two registrations or two purchases), so-called interarrival time, might affect the BIDM's performance. Moreover, we have also considered the impact that the amount of information contained in the measurements, noted as Length in Table 2, might have in the execution times. Finally, while the previous two analyses were carried out using synthetic data sources and data consumers, we have also analysed the behaviour of the measurement registration process when a real-world smart city IoT infrastructure with several thousands of sensors (i.e. SmartSantander) was injecting the measurements that it generates into the BIDM.

For the synthetic data sources, we used a Python script emulating an IoT infrastructure where we were able to set the probability distribution function and corresponding mean of the time between two consecutive measurement registrations, as well as the size of the generated measurements. Similarly, for the synthetic data consumers, a simulated client located in the Market module requested, on a periodic basis, one measurement. The script was also configurable to fix the probability distribution and mean value for the time between two of these requests.

4.1.1. Measurement registration execution time

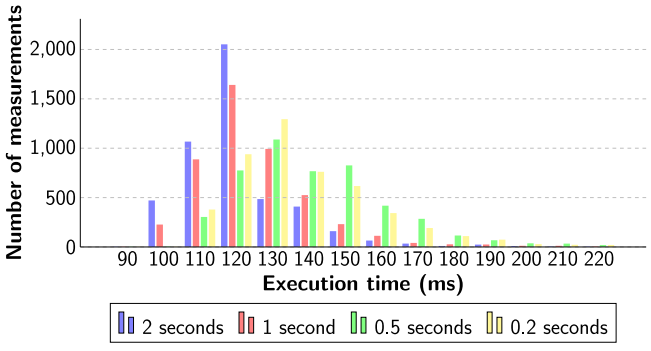
First, we present the analyses performed on the measurement registration procedure. In each of them, we show the results obtained in a graphical manner and we discuss the conclusions derived from them.

– Execution time under variable inter-arrival time

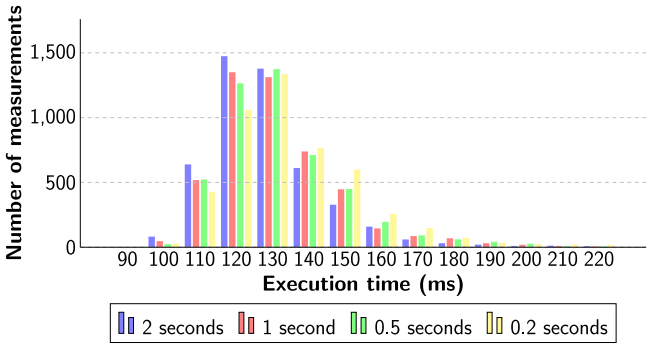
For this analysis scenario, we have used the synthetic measurement generator to produce and register measurements both at regular intervals and with an inter-arrival time following a Poisson distribution. The objective of this experiment is to observe the influence of the inter-arrival time parameter on the BIDM's performance. This is, assessing if the BIDM has a saturation point if overloaded with too frequent registrations.

There are already a number of works that have validated the Poisson process as a good approximation to model aggregated IoT traffic [36,37]. Thus, we have used it as a more realistic approximation compared to the fixed inter-arrival time cases. As a result, when using this other way of emulating the IoT data producer, there can be, so-called, "bursts" of IoT measurements arriving at the BIDM, where new measurements will arrive while the previous one has not yet been processed, thus producing concurrency.

[1] GitHub repository - BIDM(https://github.com/vgonzalez7/BIDM).



(a) Registration of measurements with fixed inter-arrival times.



(b) Registration of measurements with Poisson-modelled inter-arrival times.

Fig. 7. Execution time of measurement registrations with different inter-arrival times.

As indicated in Table 2, 5,000 samples have been taken for all the configured cases. The execution time results obtained when separating the registrations at regular intervals of different duration (2, 1, 0.5 and 0.2 s between measurements, respectively) can be seen in Fig. 7(a). In the case of the sources generating measurements with an inter-arrival time modelled as a Poisson process, the same parameters (i.e. 2, 1, 0.5 and 0.2 s) were used but, in this case, as the average value of the Poisson distribution function. It is worth noting that a Poissonian source producing measurements with mean N seconds is equivalent to M parallel sources producing measurements with mean $M * N$ seconds. This implies that the synthetic generator used can be interpreted either as one single fast IoT producer or as several IoT producers working simultaneously at proportionally slower rates. Results of the execution time can be seen in Fig. 7(b). In all cases the size of the measurement that was registered was 229 bytes, which matches the average measurement size in the SmartSantander IoT infrastructure [35].

As it can be seen in Fig. 7, the registrations take an average of 130 ms, with a variance of $5.0 \cdot 10^4$ ms^2 for regular interval generation and $4.3 \cdot 10^4$ ms^2 in the case of the source modelled as a Poisson process. It can be noted that the statistical distribution is reminiscent of a Gaussian, with a clearly identifiable mean and “tails" on both sides. In these cases, the slope of the tail on the lower side is more pronounced since it is impossible for the processing to be much faster than the mean. Another important detail is the negligible difference between the four cases of the same type. In this sense, it might seem reasonable to think that if the measurements arrive at a higher rate, the processing would start to stress the BIDM and produce it to slow down. However, the behaviour is not dependant on the inter-arrival time. This allows us to conclude that the BIDM is capable of handling registration operations on the Blockchain at a high rate without having its performance hindered.

– *Execution time under variable measurement length*

In the previous scenario, the size of the measurements used was constant. In contrast, the objective in this case is to determine whether a variation in the length of the measurements affects the execution time. We have generated three groups of 5,000 measurements synthetically, separating the registration operations in time at fixed intervals of 1 s. We wanted to avoid two consecutive registration processes to affect each other so we purposely made the inter-arrival time long enough to guarantee that when a new measurement arrived, the BIDM was idle.

The three groups of measurements are divided according to their length. The short measurements are 55 bytes long, the standard measurements occupy 229 bytes, and the body of the long measurements is 997 bytes long. In principle, the processing does not depend on the length of the measurements, but it is necessary to consider the possibility of variations due to the inner workings of the Blockchain. Testing whether this possibility is a reality is the motivation for this experiment. The execution time results obtained are shown in Fig. 8.

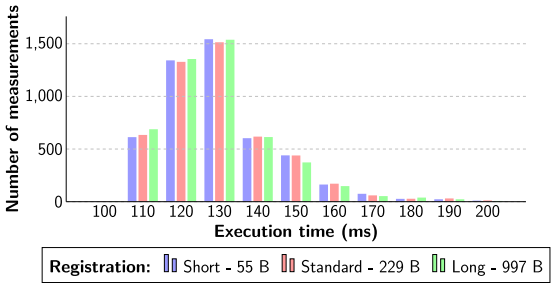


Fig. 8. Execution time of measurement registrations with variable length.

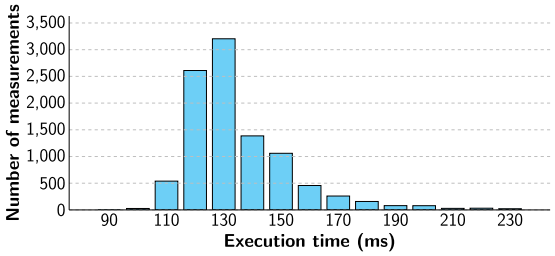


Fig. 9. Execution time of the registration of SmartSantander-generated measurements.

Table 3
Comparison of measurement registrations' execution time for different numbers of validator nodes.

Number of validator nodes	0.5s fixed inter-arrival times	Poisson-modelled inter-arrival times (0.5s avg.)	Generate by Smart-Santander
3	+9%	+7%	+5%
6	−14%	−13%	+14%

The influence of the size of the measurements on the performance of the BIDM is null. Except for minimal variations due to the statistical nature of the analysis, the registration procedures show virtually identical behaviour despite one set of measurements being more than 10 times the length of another.

– *Real IoT deployment data source: SmartSantander*

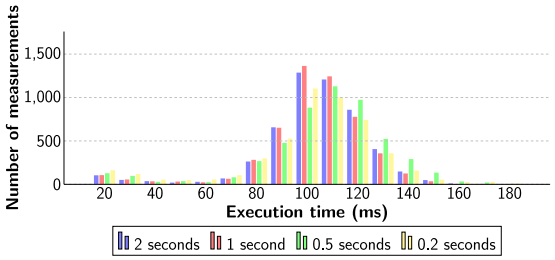
For this scenario, we will make use of the integration with SmartSantander, which will act as the IoT data producer used, instead of using the synthetic measurements generator for the performance analysis. As indicated in Table 2, 10,000 registrations have been generated. Fig. 9 shows the results obtained.

It can be seen that the execution time probability density function and the average execution time are very similar to those seen in the previous cases. This shows that the performance of the BIDM shows no apparent change when used under real measurement generation conditions (considering that SmartSantander counts with thousands of sensors, these are, indeed, high-load real-world conditions). In view of these results, it can be concluded that the platform is able to support its integration with real IoT producers.

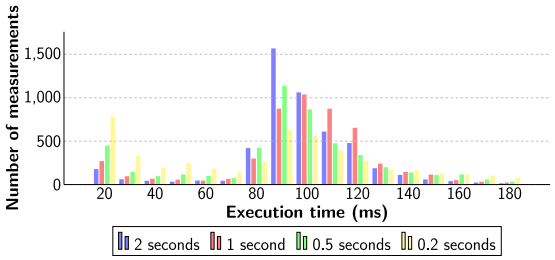
– *Different number of validator nodes*

In this scenario, we repeated all the previous scenarios presented in Section 4.1.1 using a different number of validator nodes in the BIDM's Blockchain network. The previous scenarios were carried out with one validator node. In Table 3, we compare these results with the ones obtained by replicating the exact same scenarios with 3 and 6 validator nodes, respectively. It is worth noting that we used a 4-core CPU in all the cases.

The results are expressed in a relative manner, taking as the reference the average execution time for the 1-validator configuration, since we think this fits the purpose of the results' analysis better. None of the results deviate more than 15% from the 1-validator case. The experimental nature of the analysis is behind the fluctuating results, which sometimes show increased and others reduced performance of the BIDM. However, there is no clear pattern on it. It is worth mentioning that the higher the number of validator nodes, the higher the variance of the results. Thus, it is sensible to conclude that the size of the Blockchain network does not actually have a clear direct systematic impact on the BIDM behaviour. This would mean that the size and, correspondingly, the decentralisation of the BIDM, can grow in practical terms without degrading its performance. Indeed, from the results obtained with the 6-validators configuration, it can be the case that enlarging the BIDM's underlying Blockchain network, actually has a beneficial effect.



(a) Purchase of measurements with fixed inter-request times.



(b) Purchase of measurements with Poisson-modelled inter-request times.

Fig. 10. Execution time of measurement purchases with different inter-request times.

4.1.2. Measurement purchase execution time

In this section, we present the analyses performed on the measurements purchase procedure. As in the case of the registration, the results obtained for every scenario are presented and we will discuss the conclusions derived from them.

– *Variable time between consecutive requests*

Similarly to the synthetic IoT data generator used in the registration scenarios, in this case we have used an automated function of the Market core component that purchases measurements from a given customer account. The objective of this experiment, as in the case of the registration, is to observe if the time between two consecutive purchase events has any influence on the BIDM's performance. For this purpose, 5,000 samples have, once again, been taken in several cases. We have made the analyses using both a fixed time between events (purchases in this case) and modelling the time between purchases as a Poisson process. The experiment parameters are identical to the registration case. The measurement purchase execution time results can be seen in Fig. 10(a) for the cases of 2 s, 1 s, 0.5 s and 0.2 s between measurements, respectively. In the case of the Poisson modelling, these values (i.e. 2, 1, 0.5, and 0.2 s, respectively) were used as the mean, and the results are shown in Fig. 10(b). In all cases we have used a measurement size of 229 bytes.

The purchase procedure has a mean execution time slightly over 100 ms, and the variance is $6.5 \cdot 10^2$ ms^2 when purchases are made at regular intervals, and $1.4 \cdot 10^3$ ms^2 in the case that purchases are modelled as a Poisson process. The conclusions that can be drawn do not differ from those found for the measurement registration process, with only one exception. In the experiment where the time between purchases was defined with a Poisson distribution, there is a larger number of purchases that are processed in a very short time. In fact, the smaller the average time between consecutive events, the more probable it is that this fast processing occurs.

The rationale behind this behaviour can be found in the internal functioning of Blockchain and the PoA mechanism. In general, when a transaction occurs on the Blockchain, the mining nodes work on it until they succeed in mining a block. When this happens, all pending transactions are included in that block, which is then added to the Blockchain. In the PoA mechanism used in this platform, mining is practically instantaneous since the cryptographic difficulty of the mining process has been reduced to a minimum. In all the previous cases, every time a transaction takes place, part of the execution time is due to the validation and mining of the block, although it is a practically negligible effect. However, when using a more realistic model, such as the Poisson process used in the experiments, for emulating the load of the BIDM, we expose the BIDM to bursts of purchase requests. Thus, it is more probable that several purchase processes start to overlap. This overlapping causes the blocks to include several transactions automatically. The beneficial side effect of this situation is that there are transactions that do not need to wait for the validation and mining of a new block, since they already find one waiting. This is what can be observed in the purchase graphs, where there are more purchases with shorter execution times the faster the arrival rate is. Therefore, it can be concluded that the system performs even better under heavier load.

– *Variable length of purchased measurements*

As it was the case for the measurement registration, the scenarios above used fixed-length measurements. Similarly, the objective of this second analysis is to determine whether a variation in the length of the purchased measurements affects the execution time of this operation. Three groups of 5,000 measurement purchase processes have been generated synthetically, separating two

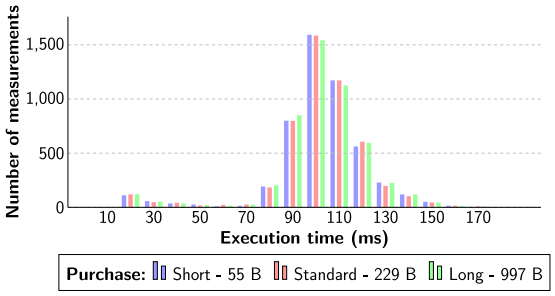


Fig. 11. Execution time of measurement purchases with variable length.

Table 4
Comparison of measurement purchases' execution time for different numbers of validator nodes.

Number of validator nodes	0.5s fixed inter-arrival times	Poisson-modelled inter-arrival times (0.5s avg.)
3	−11%	+6%
6	−18%	+7%

consecutive purchase operations by a fixed 1-second interval. The sizes of the purchased measurements are, again, 55, 229 and 997 bytes for short, average and large measurements, respectively. The execution time results obtained for each case are shown in Fig. 11.

As expected based on the previous experiments, the histograms show that the purchase execution time is not affected by the size of the purchased measurements.

It is worth mentioning that the purchase process shows no difference when acquiring measurements registered by SmartSantander. Therefore, we have not included a graph depicting this specific case since it does not provide new information.

– *Different number of validator nodes*

Similarly to the case of the data registrations' execution time analysis, we have also repeated the two previous scenarios presented in Section 4.1.2 using a different number of validator nodes. In Table 4, we compare the results obtained with the 1-validator configuration with new ones obtained by replicating them with 3 and 6 validator nodes, respectively.

Results are expressed in relative manner, taking a reference analogous to the case of the measurement registrations (i.e. the average execution time of measurement purchases made with a fixed 0.5 s period between purchases and 1-validator configuration). The execution times in this case are slightly faster when purchasing measurements with fixed inter-arrival times when compared to the 1-validator scenarios. Conversely, they were slightly slower when purchasing measurements with Poisson-modelled inter-arrival times. We concluded that this fluctuating behaviour is due to the experimental nature of the analysis. In any case, the fact that the relative differences between the three configurations employed is low allows us to answer one of our RQs, and to neglect the impact that enlarging (thus, making it even more decentralised in practice) the BIDM might have on its behaviour.

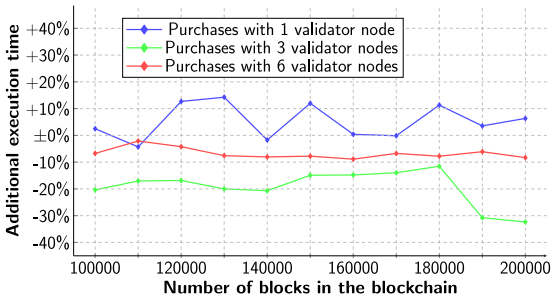
4.1.3. Temporal analysis of the subscription-notification process

Once we have analysed the synchronous operations of registering and consuming IoT measurements from the BIDM, we now focus on the analysis of the subscription model. We will present the evaluation of its performance as well as how it compares with the performance of the individual measurement purchase model.

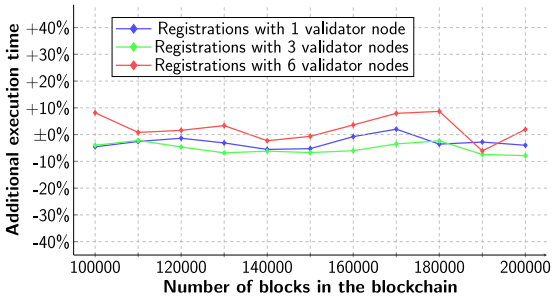
In terms of execution time, this model has a conceptual advantage. Although subscription may be a costlier operation, it only has to be performed once to acquire multiple measurements. Therefore, the execution time of the subscription procedure is a value with much less impact on the final performance, since it is only going to be executed on limited occasions. In other words, regardless of how long it takes to create a subscription, dividing this initial-time among the tens, hundreds or thousands of measurements that are obtained through the subsequent notifications, makes the per-measurement execution time negligible. However, we have performed some basic analyses on this procedure anyway.

First of all, we have analysed the subscription procedure execution time. We have logged over 100 timestamps at the beginning and at the end of the process, in order to obtain a reasonably reliable result. While its statistical behaviour is not as interesting as that of repeating operations, the execution time of a subscription has been observed to be in the range of 100 to 200 ms, with a mean of approximately 140 ms (and a variance of $2.0 \cdot 10^4$ ms^2). Therefore, despite being a rarely performed operation, its computational cost and resultant execution time is similar to that of the registration and purchase procedures, if only slightly higher.

Another interesting result that has not been considered so far is the user experience when using the web API. Much of it depends on the efficiency of the client (i.e. browser or native HTTP client) and the specifications of the device used to access the site, and this is the main reason why it has not been given a lot of attention throughout this section. However, it is interesting to compare which one of the two different models performs better. To do this, we have measured the loading time experienced by a standard user using a well-known browser running on an Intel® Core™i5-6600K processor. The samples used for this experiment have been



(a) Purchases with different number of validator nodes.



(b) Registrations with different number of validator nodes.

Fig. 12. Execution time for different Blockchain loads.

taken at the time of displaying all user's purchased measurements on screen. In the case of the standard measurement purchase model, the web server calls the Clearing House SC and filters the confirmed purchase events from the logged-in user's Blockchain account. This involves going through the entire list of the contract's events. In the subscription model, the server calls the Clearing House SC and filters the subscription events from the user's account, which are fewer in number. Once the program knows the exact moment when the user subscribed to a given topic, it calls the Data SC and filters the registration events where at least one of the topics associated with the measurement matches one of those to which the user is subscribed.

At first glance it may appear that the second model implies a higher computational load, but the lists traversed are smaller and the filters are more efficient. We have taken 100 samples of the loading time for both visualisations: individually purchased measurements, and those implicitly purchased through a subscription. We noted that the loading time in the first case has an average of 1,310 ms, while in the subscription model it drops to 920 ms. The second process is therefore about 40% faster, although both are relatively normal loading times compared to standard commercial web pages. In any case, the most remarkable result is that the subscription model presents a lower delay when searching, filtering and visualising measurements when compared to the individual purchase model.

4.2. Long term scalability

One of the most important characteristics of a system of this kind is its ability to maintain its performance as its load increases and it has to handle large amounts of information; in other words, its scalability. The following experimental analysis consists of registering and purchasing massive amounts of measurements, and then testing whether after a long period of time, with a Blockchain containing several hundred thousand blocks and a database with approximately the same number of measurements, the BIDM behaves as it did at the beginning.

4.2.1. Execution time for different blockchain sizes

First of all, we are going to analyse the execution time of the BIDM procedures for different sizes of the Blockchain (i.e. with a growing number of blocks chained). This could potentially deteriorate the behaviour of the BIDM. However, we can see in Fig. 12 that there is no performance degradation and that both procedures (i.e. registration and consumption of IoT measurements) have a reasonably constant execution time, regardless of the system's load.

The automatic generator with fixed 0.5 s interval between measurements' registrations was used, and the data purchases followed a similar process with its corresponding automatic system within the Market core. We started the analysis from a Blockchain that already contained 100,000 blocks, and another 100,000 operations (both registrations and purchases) have been performed. The same experiment has been carried out with the three configurations of the Blockchain network (i.e. with 1, 3 and 6 validators, respectively).

The ordinate axis in Fig. 12 represents the relative difference in the execution time of each operation. The results shown consider that the average execution time of the first 30,000 operations (i.e. the ones that created the initial 30,000 blocks) has been used as a reference. As shown in Fig. 12, for the three Blockchain configurations, the execution time remains quite stable, with fluctuations up and down a more or less constant average with the evolution of the number of blocks in the Blockchain. This is, as time goes by and more measurements are continuously being registered and consumed, more and more blocks containing the different transactions and events that have been specified in Section 3 are chained. The points shown in Fig. 12 have been calculated every 10,000 blocks and are computed taking the respective previous and following 5,000 operations from that point. Despite the small variations, it is possible to conclude that the size of the Blockchain has no direct correlation with the execution time of the BIDM operations.

Moreover, looking at the other variable introduced in the experiments (i.e. the amount of validator nodes deployed for the BIDM's Blockchain), it is also evident that there is no interrelation between the BIDM's performance and the size of the Blockchain network. In the case of the registration procedures, the execution times for the three configurations are almost the same, while in the case of the purchase procedures, the difference in performance does not seem to be dependant on the number of nodes in the network as the best performance has been achieved for 3 validators. In any case, from the results obtained, it can be concluded that increasing the size of the BIDM's Blockchain, if any, has a positive impact on the BIDM's performance.

It is worth mentioning that it is not possible to assure that, in the future, when the platform is loaded with a much higher number of measurements, there will not be a point when performance begins to degrade. However, given the large amount of blocks that the Blockchain had at the end of the evaluation and the lack of significant impact, it is reasonable to conclude that such a situation will only happen after a much larger period of usage of the BIDM and that it could be compensated by using more powerful computing infrastructure, if it happens at all.

Given the results obtained in this analysis, we can now establish a mathematical model to obtain the maximal throughput of the BIDM. In this regard, we firstly derive the value of the mean execution time of all procedures, which requires a mean value for every kind of operation and the probability of that given operation as well as considering the probability of registration and purchase of measurements. Eq. (1) shows the analytical expression derived to get this result. $E.T.$ stands for execution time, and therefore $\overline{E.T.}$ is the average execution time. REG refers to registration operations, whereas PUR refers to purchase operations. It is important to note that $Pr\{REG\} + Pr\{PUR\} = 1$.

$$\overline{E.T.} = \overline{E.T._{REG}} \cdot Pr\{REG\} + \overline{E.T._{PUR}} \cdot Pr\{PUR\} \tag{1}$$

Eq. (2) shows the result obtained when using the average execution time values obtained in the analysis carried out on top of our evaluation testbed and considering the situation in which the BIDM is equally used for registering and for purchasing measurements (i.e. $Pr\{REG\} = Pr\{PUR\} = 0.5$).

$$\overline{E.T.} = 160 \text{ ms} \cdot 0.5 + 100 \text{ ms} \cdot 0.5 = 130 \text{ ms} \tag{2}$$

Knowing the average execution time of procedures, the maximum throughput can be calculated as an inverse of it, resulting in the number of operations that the BIDM can perform in a given time span. Eq. (3) shows this calculation applied to the value obtained previously. The term *op* refers to an operation.

$$\text{Throughput} = \frac{1000 \text{ ms/s}}{\overline{E.T.} \text{ (ms/op)}} = \frac{1000 \text{ ms/s}}{130 \text{ ms/op}} \approx 7.7 \text{ ops/s} \tag{3}$$

As it has been previously mentioned, this result has been obtained considering that the BIDM is deployed at the small-footprint computing infrastructure used for the evaluation testbed (i.e. one signer node and one IPFS database instance). However, multiple instances of these elements would further improve the throughput of the BIDM. However, the evaluation of the impact of using a larger infrastructure for the deployment of the BIDM on its performance is not within the scope of this paper.

4.2.2. Disk usage of the BIDM

So far, the analyses have been focused on the time it takes for certain operations or processes to execute, but this is not the only interesting metric in the evaluation of the computational performance of the BIDM. Another parameter of great importance is disk usage. That is, how many additional bytes are generated in the storage system each time an operation takes place. This parameter is necessary to assess how bearable in terms of storage resources is the continuous operation of the BIDM.

The process used to evaluate this parameter is as follows. First, we have noted the size of the virtual machines with an empty Blockchain. Afterwards, we have registered 20,000 measurements, then stopped all the processes on the platform, and noted the size again. On the same Blockchain, we have purchased all 20,000 measurements, stopped all the processes once more, and noted the size one last time. This way, we have been able to calculate the difference in size between each of the three instants, and obtain an estimate of the increase in disk usage caused by each of the operations. Having performed a large number of procedures, we can calculate a sufficiently accurate average. Eqs. (4) and (5) show these calculations and their result.

$$\text{Registration: } \frac{427 \cdot 10^6 \text{ bytes}}{20{,}000 \text{ registrations}} = 21{,}350 \ \frac{\text{bytes}}{\text{registration}} \tag{4}$$

$$\text{Purchase: } \frac{305 \cdot 10^6 \text{ bytes}}{20{,}000 \text{ purchases}} = 15{,}250 \ \frac{\text{bytes}}{\text{purchase}} \tag{5}$$

As shown above, the measurement registration leads to an increase in disk usage of approximately 21 KBytes, and is significantly higher than the purchase process, at about 15 KBytes. This difference is due to the fact that registration entails inserting a new

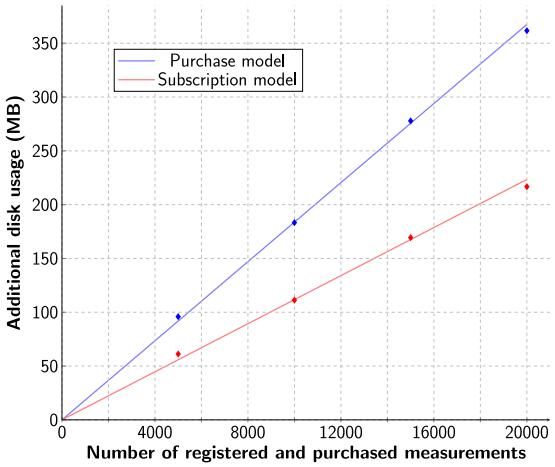


Fig. 13. BIDM evolution in terms of disk usage.

measurement into the database and several of its parameters into the Blockchain, all of which implies an overhead when compared to a purchase, which does not insert new data.

With this information, we have compared the individual measurement purchase model and the subscription model in terms of disk usage. It is important to remember that subscriptions have the advantage that measurement acquisitions are implicit (i.e. consumers only have to detect the new measurements registration events and they will be able to directly get them without adding further information to the Blockchain). This avoids the multiple transactions present in the purchase process. That said, the scalability comparison of the two models, making use of the disk usage results obtained above, is shown in Fig. 13.

The evolution in terms of disk usage, taken directly in MB, can be observed. This is a linear regression that makes use of the previous results, and shows that the slope in the case of the purchase model is almost twice as steep as in the subscription model. This happens because the purchases in the latter are implicit and do not imply additional data or metadata to be stored. Furthermore, the experimentally measured points, respectively obtained after the registration and purchase of 5,000, 10,000, 15,000 and 20,000 measurements, are also plotted. This way, we have experimentally confirmed the analytical results derived from Eq. (4) and Eq. (5), which model the BIDM's scalability with a linear behaviour.

With the results at hand, we can conclude that the subscription model entails a great advantage over the individual purchase model. Its scalability, or in other words, the disk usage of the platform, grows at a much slower rate and therefore allows operation on more limited computing infrastructures.

4.3. Comparative discussion

In this section, we have analysed the performance of a blockchain-based marketplace when carrying out the basic operations of purchase, registration and subscription. Looking back at Section 2, we identified two works [11,26] that also conducted some kind of performance analysis over a blockchain-based marketplace. Now, we will compare the results in order to pinpoint possible limitations or open issues in our work.

In [11], the performance analyses are purely theoretical and therefore, not easy to compare with ours, which are more practice-oriented. In this sense, we can say that even if we feel that our practical results are very complete, and we have found clear trends in the results, we are lacking a more complex theoretical model that analytically describes the BIDM's behaviour.

In the case of [26], the analyses are more focused on the gas cost of procedures rather than their performance. However, they do present an evaluation of the computation time during data brokerage. It shows a linear increase with the number of purchases. Nevertheless, this result alone is not sufficient to make a fair comparison. We can see that, as we discussed in Section 2, none of the works really focus on answering our RQ2 through a detailed performance analysis as we have done. On the other hand, we have not focused on the gas cost in our analyses, given that it was already a fairly discussed topic and we decided to employ PoA-based Blockchain, which do not imply such costs.

In conclusion, the performance evaluation that we have carried out is, actually, one of the key and novel contributions of our work as, to the best of our knowledge, there are no similar studies in the literature. We believe it is not possible to make a direct comparison between the results of our work and others that have proposed analogous solutions for addressing IoT data exchange requirements. We have shifted the usual focus from a gas cost perspective to a more performance-oriented one, resulting in a comprehensive set of performance analyses while giving up some of the more common metrics.

5. Conclusions

This paper presents a platform that enables a decentralised marketplace, so-called Blockchain-based IoT Data Marketplace (BIDM), where information can be exchanged between IoT infrastructures generating it and applications exploiting it, in a fast, self-sovereign and totally transparent way. Furthermore, the use of this technology enables the traceability of the information, thus assuring at every moment who purchased a measurement and when the transaction was carried out. In addition, the authentication mechanism implemented in the platform guarantees the provenance of the data. Thus, providing reliability and quality to the platform.

Moreover, the behaviour of the BIDM has been assessed through several performance evaluation tests. This evaluation has demonstrated that the impact of the BIDM in the measurement registration and purchase procedures is negligible and, what is more important, such small footprint is kept stable even for continuous and long-term operation of the BIDM. Moreover, the BIDM has proved that it is able to support heavy loads with frequent data registration and/or consumption maintaining the same streamlined operation.

In this sense, the paper has answered all the research questions introduced in Section 1. Firstly, the design of the BIDM's architecture and key functionalities, described in Section 3 addresses RQ1 by specifying, in detail, the main procedures that take place within the marketplace, namely, the production and the consumption of IoT-generated measurements. In this latter aspect, the BIDM does not only support synchronous purchases of specific measurements, but also implements the necessary mechanisms to allow subscriptions to streams of data, a functionality that has demonstrated, through the performance evaluation carried out, an excellent behaviour and the capability to optimise the BIDM's scalability.

Pertaining to RQ2, the results obtained from the performance evaluation carried out have shown that the execution time of the registration of new IoT data and the consumption of available IoT measurements in the BIDM is negligible, even under heavy load conditions. Moreover, the performance studies that were performed also considered the evolution of the BIDM's behaviour with time in order to analyse the scalability of the solution. Precisely, to avoid the scalability problems associated with on-chain data storage, the BIDM keeps storage off-chain. This way, the process of synchronisation is significantly streamlined. Otherwise, it would be necessary for any new actor (i.e. producer or consumer) arriving at the marketplace to synchronise all the historical information which, for large IoT scenarios, could be a cumbersome process.

All in all, this paper has described how Blockchain technologies can be leveraged to support flexible and trustworthy exchange of data within IoT data marketplaces. The conclusions derived from the analysis of prior existing works have helped us in defining the key functional and non-functional requirements. Moreover, as from the results of the evaluation presented in the paper, it is also possible to conclude that Blockchain technologies in general do not imply a performance bottleneck, and they can be used as a baseline for the support of IoT data marketplaces, if they are properly introduced in the system architecture, as it is the case of the BIDM in particular.

Finally, in the domain of data sharing, the current trend is on the establishment of Data Spaces. In particular, International Data Spaces (IDS) [38] is at the forefront of the alternatives with a higher critical mass and community support. Within the IDS model, the Clearing House component is responsible for the recording and attestation of data transactions. We are working on integrating the functionality of the BIDM as an actual implementation of the IDS Clearing House. Moreover, from the performance evaluation viewpoint, while the study carried out has allowed us to derive long-term conclusions, we want to extend the evaluation to other performance metrics and aspects of such a socio-technical dApp besides the pure performance. As it has been indicated in the comparative discussion in Section 4, no similar behaviour and performance evaluation has been done so direct comparison has not been possible. By extending the behaviour analysis to other metrics it will be possible to make a thorough comparison with related works. In this regard, we will extend the review of related work that has been presented in Section 2 into a Systematic Literature Review that, besides surveying current state of the art in this field, would allow us to identify other aspects and metrics that might be matter of analysis on the BIDM.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the Spanish State Research Agency (AEI) by means of the project FIERCE "Future Internet Enabled Resilient CitiEs" under Grant Agreement No. RTI2018-093475-A-I00 and the project SITED "Semantically-enabled Interoperable Trustworthy Enriched Data-spaces" under Grant Agreement No. PID2021-125725OB-I00.

References

[1] European Commission, The European data strategy: Shaping europe digital's future, 2020, http://dx.doi.org/10.2775/987881, https://ec.europa.eu/commission/presscorner/api/files/attachment/862109/European_data_strategy_en.pdf, (Accessed: 05-04-2022).
[2] K.J. Singh, D.S. Kapoor, Create your own internet of things: A survey of IoT platforms, IEEE Consum. Electron. Mag. 6 (2) (2017) 57–68, http://dx.doi.org/10.1109/MCE.2016.2640718.
[3] J. Kim, J. Yun, S.-C. Choi, D.N. Seed, G. Lu, M. Bauer, A. Al-Hezmi, K. Campowsky, J. Song, Standard-based IoT platforms interworking: implementation, experiences, and lessons learned, IEEE Commun. Mag. 54 (7) (2016) 48–54, http://dx.doi.org/10.1109/MCOM.2016.7514163.
[4] M. Noura, M. Atiquzzaman, M. Gaedke, Interoperability in internet of things: Taxonomies and open challenges, Mob. Netw. Appl. 24 (3) (2019) 796–809, http://dx.doi.org/10.1007/s11036-018-1089-9.
[5] P. Sotres, J. Lanza, L. Sánchez, J.R. Santana, C. López, L. Muñoz, Breaking vendors and city locks through a semantic-enabled global interoperable internet-of-things system: A smart parking case, Sensors 19 (2) (2019) 229, http://dx.doi.org/10.3390/s19020229.
[6] C. Harrold, Beyond the hype–smart today and tomorrow, in: Practical Smart Device Design and Construction, Springer, 2020, pp. 37–45, http://dx.doi.org/10.1007/978-1-4842-5614-5.
[7] L. Sánchez, J. Lanza, L. Muñoz, From the internet of things to the social innovation and the economy of data, Wirel. Pers. Commun. 113 (3) (2020) 1407–1421, http://dx.doi.org/10.1007/s11277-020-07321-2.
[8] X. Sheng, J. Tang, X. Xiao, G. Xue, Sensing as a service: Challenges, solutions and future directions, IEEE Sens. J. 13 (10) (2013) 3733–3741, http://dx.doi.org/10.1109/JSEN.2013.2262677.
[9] X. Zhao, K.K. Sajan, G.S. Ramachandran, B. Krishnamachari, Demo abstract: the intelligent IoT integrator data marketplace-version 1, in: 2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI), IEEE, 2020, pp. 270–271, http://dx.doi.org/10.1109/IoTDI49375.2020.00042.
[10] K. Mišura, M. Žagar, Data marketplace for Internet of Things, in: 2016 International Conference on Smart Systems and Technologies, SST, IEEE, 2016, pp. 255–260, http://dx.doi.org/10.1109/SST.2016.7765669.
[11] A. Agarwal, M. Dahleh, T. Sarkar, A marketplace for data: An algorithmic solution, in: Proceedings of the 2019 ACM Conference on Economics and Computation, 2019, pp. 701–726, http://dx.doi.org/10.1145/3328526.3329589.
[12] F. Stahl, F. Schomm, G. Vossen, The data marketplace survey revisited, Tech. rep., ERCIS Working Paper, 2014.
[13] F. Stahl, F. Schomm, G. Vossen, Data marketplaces: An emerging species, in: DB&IS, 2014, pp. 145–158, http://dx.doi.org/10.3233/978-1-61499-458-9-145.
[14] P. Gupta, S. Kanhere, R. Jurdak, A decentralized IoT data marketplace, 2019, http://dx.doi.org/10.48550/arXiv.1906.01799, arXiv preprint arXiv:1906.01799.
[15] H. Yoo, N. Ko, Blockchain based data marketplace system, in: 2020 International Conference on Information and Communication Technology Convergence, ICTC, IEEE, 2020, pp. 1255–1257, http://dx.doi.org/10.1109/ICTC49870.2020.9289087.
[16] P. Banerjee, S. Ruj, Blockchain enabled data marketplace–design and challenges, 2018, http://dx.doi.org/10.48550/arXiv.1811.11462, arXiv preprint arXiv:1811.11462.
[17] S. Lawrenz, P. Sharma, A. Rausch, Blockchain technology as an approach for data marketplaces, in: Proceedings of the 2019 International Conference on Blockchain Technology, 2019, pp. 55–59, http://dx.doi.org/10.1145/3320154.3320165.
[18] K. Christidis, M. Devetsikiotis, Blockchains and smart contracts for the internet of things, Ieee Access 4 (2016) 2292–2303, http://dx.doi.org/10.1109/ACCESS.2016.2566339.
[19] P. Gupta, V. Dedeoglu, S.S. Kanhere, R. Jurdak, Towards a blockchain powered IoT data marketplace, in: 2021 International Conference on COMmunication Systems & NETworkS, COMSNETS, IEEE, 2021, pp. 366–368, http://dx.doi.org/10.1109/COMSNETS51098.2021.9352865.
[20] G.S. Ramachandran, R. Radhakrishnan, B. Krishnamachari, Towards a decentralized data marketplace for smart cities, in: 2018 IEEE International Smart Cities Conference (ISC2), IEEE, 2018, pp. 1–8, http://dx.doi.org/10.1109/ISC2.2018.8656952.
[21] M. Ha, S. Kwon, Y.J. Lee, Y. Shim, J. Kim, Where WTS meets WTB: A blockchain-based marketplace for digital me to trade users' private data, Pervasive Mob. Comput. 59 (2019) 101078, http://dx.doi.org/10.1016/j.pmcj.2019.101078.
[22] J.-S. Park, T.-Y. Youn, H.-B. Kim, K.-H. Rhee, S.-U. Shin, Smart contract-based review system for an IoT data marketplace, Sensors 18 (10) (2018) 3577, http://dx.doi.org/10.3390/s18103577.
[23] R. Xu, G.S. Ramachandran, Y. Chen, B. Krishnamachari, Blendsm-ddm: Blockchain-enabled secure microservices for decentralized data marketplaces, in: 2019 IEEE International Smart Cities Conference (ISC2), IEEE, 2019, pp. 14–17, http://dx.doi.org/10.1109/ISC246665.2019.9071766.
[24] N. Hynes, D. Dao, D. Yan, R. Cheng, D. Song, A demonstration of sterling: a privacy-preserving data marketplace, Proc. VLDB Endow. 11 (12) (2018) 2086–2089, http://dx.doi.org/10.14778/3229863.3236266.
[25] A. Alsharif, M. Nabil, A blockchain-based medical data marketplace with trustless fair exchange and access control, in: GLOBECOM 2020-2020 IEEE Global Communications Conference, IEEE, 2020, pp. 1–6, http://dx.doi.org/10.1109/GLOBECOM42002.2020.9348192.
[26] V. Koutsos, D. Papadopoulos, D. Chatzopoulos, S. Tarkoma, P. Hui, Agora: a privacy-aware data marketplace, IEEE Trans. Dependable Secure Comput. (2021) http://dx.doi.org/10.1109/TDSC.2021.3105099.
[27] W. Badreddine, K. Zhang, C. Talhi, Monetization using blockchains for IoT data marketplace, in: 2020 IEEE International Conference on Blockchain and Cryptocurrency, ICBC, IEEE, 2020, pp. 1–9, http://dx.doi.org/10.1109/ICBC48266.2020.9169424.
[28] S.A. Azcoitia, N. Laoutaris, A survey of data marketplaces and their business models, 2022, http://dx.doi.org/10.48550/arXiv.2201.04561, arXiv preprint arXiv:2201.04561.
[29] F. Schomm, F. Stahl, G. Vossen, Marketplaces for data: an initial survey, ACM SIGMOD Record 42 (1) (2013) 15–26, http://dx.doi.org/10.1145/2481528.2481532.
[30] M. Spiekermann, Data marketplaces: Trends and monetisation of data goods, Intereconomics 54 (4) (2019) 208–216, http://dx.doi.org/10.1007/s10272-019-0826-z.
[31] L. Sterling, K. Taveter, The Art of Agent-Oriented Modeling, MIT Press, 2009.
[32] P. Szilágyi, Clique proof-of-authority consensus protocol, 2017, https://eips.ethereum.org/EIPS/eip-225/, (Accessed: 2022-04-05).
[33] A. Norta, A. Kormiltsyn, C. Udokwu, V. Dwivedi, S. Aroh, I. Nikolajev, A blockchain implementation for configurable multi-factor challenge-set self-sovereign identity authentication, in: 2022 IEEE International Conference on Blockchain (Blockchain), 2022, pp. 455–461, http://dx.doi.org/10.1109/Blockchain55522.2022.00070.
[34] L. Sanchez, L. Muñoz, J.A. Galache, P. Sotres, J.R. Santana, V. Gutierrez, R. Ramdhany, A. Gluhak, S. Krco, E. Theodoridis, et al., SmartSantander: IoT experimentation over a smart city testbed, Comput. Netw. 61 (2014) 217–238, http://dx.doi.org/10.1016/j.bjp.2013.12.020.
[35] J. Lanza, P. Sotres, L. Sánchez, J.A. Galache, J.R. Santana, V. Gutiérrez, L. Muñoz, Managing large amounts of data generated by a smart city internet of things deployment, Int. J. Semantic Web Inf. Syst. (IJSWIS) 12 (4) (2016) 22–42, http://dx.doi.org/10.4018/IJSWIS.2016100102.
[36] F. Metzger, T. Hoßfeld, A. Bauer, S. Kounev, P.E. Heegaard, Modeling of aggregated IoT traffic and its application to an IoT cloud, Proc. IEEE 107 (4) (2019) 679–694, http://dx.doi.org/10.1109/JPROC.2019.2901578.
[37] X. Jian, X. Zeng, J. Huang, Y. Jia, Y. Zhou, Statistical description and analysis of the concurrent data transmission from massive MTC devices, Int. J. Smart Home 8 (4) (2014) 139–150, http://dx.doi.org/10.14257/ijsh.2014.8.4.13.
[38] B. Otto, S. Steinbuß, A. Teuscher, S. Lohmann, Reference architecture model for the international data spaces, Tech. Rep., International Data Spaces Association, 2019.

Víctor González is a research fellow in the Network Planning and Mobile Communications Laboratory at Universidad de Cantabria, Spain. He received his Master's degree in Telecommunications Engineering from Universidad de Cantabria in 2021. His research interests are on the application of Blockchain technology for the distributed and secure sharing of IoT Data. Moreover, he is also active on applying semantic web principles to data sharing and, this way, developing a fully distributed secure data sharing ecosystem.

Dr. Luis Sánchez is Associate Professor at Universidad de Cantabria (Spain). He received M.Sc. (2002) the Ph.D. (2009) in Telecommunications Engineering. He is active on the IoT-enabled smart cities, and the application of AI for data enrichment. He has led and/or participated in more than 15 projects belonging to different EU Framework Programs. He has authored more than 60 papers at international journals and conferences. He often participates in panels discussing about innovation supported by IoT in Smart cities. He also acts as expert for several European countries national funding agencies.

Dr. Jorge Lanza is an Associate Professor at the Network Planning and Mobile Communications Laboratory at the University of Cantabria (UC), Spain. He received his Ph.D. in telecommunications engineering from University of Cantabria in 2014. He has participated in several research projects, national and international, with both private and public funding. Currently his research is focused on IoT infrastructures towards federating deployments in different locations using semantics technologies. In addition, his work has included combined mobility and security for the wireless Internet.

Dr. Juan Ramón Santana received the Ph.D. degree in Telecommunication Engineering from the University of Cantabria in 2021. He has participated in a number of international projects, mostly within the Internet of Things (IoT) paradigm, and its application in the Smart City domain. His research interests include the data management plane of IoT infrastructures, as well as their federation using semantic-enabled technologies. He is currently a Research Fellow with the Network Planning and Mobile Communications Laboratory, University of Cantabria, and has authored more than 30 publications, including conferences, journals, and book chapters.

Dr. Pablo Sotres is a senior research fellow in the Network Planning and Mobile Communications Laboratory, which belongs to the Communications Engineering department at the University of Cantabria, Spain. He received Telecommunications Engineering degree and Ph.D. from the University of Cantabria in 2008 and 2021 respectively. He has been involved in several different international projects framed under the smart city paradigm, such as SmartSantander; and related to inter-testbed federation, such as Fed4FIRE, Fed4FIRE+ and Wise-IoT.

Dr. Alberto E. García studied telecommunication engineering at Universidad de Cantabria where he received his Ph.D. in 2009. Since 2002 he is Assistant Professor in the Telematics Engineering Group of the Department of Telecommunication University of Cantabria (Santander, Spain). He is actually involved in several national and international projects. He has participated in specialised conferences and technical courses, including more than 50 publications.