Blockchain for the Internet of Things: a Systematic Literature Review

Original

Availability:
This version is available at: 11583/2650266 since: 2017-09-21T10:53:12Z

Publisher:
IEEE

Published
DOI:10.1109/AICCSA.2016.7945805

Terms of use:
openAccess
This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)
Abstract—In the Internet of Things (IoT) scenario, the blockchain and, in general, Peer-to-Peer approaches could play an important role in the development of decentralized and data-intensive applications running on billion of devices, preserving the privacy of the users. Our research goal is to understand whether the blockchain and Peer-to-Peer approaches can be employed to foster a decentralized and private-by-design IoT. As a first step in our research process, we conducted a Systematic Literature Review on the blockchain to gather knowledge on the current uses of this technology and to document its current degree of integrity, anonymity and adaptability. We found 18 use cases of blockchain in the literature. Four of these use cases are explicitly designed for IoT. We also found some use cases that are designed for a private-by-design data management. We also found several issues in the integrity, anonymity and adaptability. Regarding anonymity, we found that in the blockchain only pseudonymity is guaranteed. Regarding adaptability and integrity, we discovered that the integrity of the blockchain largely depends on the high difficulty of the Proof-of-Work and on the large number of honest miners, but at the same time a difficult Proof-of-Work limits the adaptability. We documented and categorized the current uses of the blockchain, and provided a few recommendations for future work to address the above-mentioned issues.

I. INTRODUCTION

As defined by ITU [1], the Internet of Things (IoT) refers to the network of numerous physical objects (20 billion by 2020, according to Gartner [2]) which are provided with Internet connection. Such devices acquire information about the surrounding environment, and they communicate with each other and with software systems through the Internet. As a consequence of such rich interaction, they also produce a large amount of data, in turn usable to enable dependent services.

Despite the benefits provided by these services, critical privacy issues may arise. That is because the connected devices (the things) spread sensitive personal data and reveal behaviors and preferences of their owners. People’s privacy is particularly at risk when such sensitive data are managed by centralized companies, which can make an illegitimate use of them: as a matter of fact, Edward Snowden’s revelations showed that people’s data stored by Internet and telecommunication companies have been exploited within a mass surveillance program, i.e., the PRISM program [3].

With the purpose of preventing this situation, the goal of our research is to encourage a decentralized and private-by-design IoT, where privacy is guaranteed by the technical design of the systems. We believe that this can be achieved by adopting Peer-to-Peer (P2P) systems. In particular, the blockchain could be very helpful in building such privacy-preserving IoT. The blockchain is a P2P ledger, firstly used in the Bitcoin cryptocurrency [4] for economic transactions. It is tamper-proof and contains only authentic information; in addition, since it is P2P, it is not controlled by any single centralized entity. For these reasons, cryptocurrencies are just one of the possible applications of this technology.

A private-by-design IoT could be fostered by the combination of the blockchain and a P2P storage system. Sensitive data produced and exchanged among IoT devices are stored in such storage system, whose P2P nature could ensure privacy, robustness and absence of single points of failure. Combined with this storage system, the blockchain has the fundamental role to register and authenticate all operations performed on IoT devices data. Each operation on data (creation, modification, deletion) is registered in the blockchain: this could ensure that any abuse on data can be detected. Moreover, access policies can be specified and enforced by the blockchain, preventing unauthorized operations on data. In this framework, people are not required to entrust IoT data produced by their devices to centralized companies: data could be safely stored in different peers, and the blockchain could guarantee their authenticity and prevent unauthorized access.

In the paper we performed a Systematic Literature Review (SLR) to verify whether documented use cases in the state of the art confirm this possibility, and, on the other side, to investigate which are the main factors that affect the levels of integrity, anonymity and adaptability of the blockchain.

The remainder of the paper is structured as follows. In Section II first we briefly delineate the main features of the blockchain, and then we describe goal and research questions of the SLR and the research process. In Section III we report the results obtained from the literature survey and in Section IV we discuss them. Finally, in Section V we provide conclusions and future work.

II. STUDY DESIGN

A. Context

The blockchain is a P2P ledger for transactions. To issue transactions, public key cryptography is employed. A user is
TABLE I: Use Cases of the Blockchain beyond Cryptocurrencies (RQ1 and RQ2)

<table>
<thead>
<tr>
<th>Category</th>
<th>Paper</th>
<th>Usage of the blockchain</th>
<th>IoT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data storage management</td>
<td>6</td>
<td>Management of access policies and references to users’ data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Management of data storage contracts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Management of document storage contracts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Tamper-proof log of events and management of access control to data</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Management of metadata of data kept in a storage system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>Automatic compensation of clients of a storage server in case stored data are lost</td>
<td></td>
</tr>
<tr>
<td>Trade of goods and data</td>
<td>12</td>
<td>Purchase by devices or human beings of assets such as sensors data or goods</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>Purchase of sensors data in IoT</td>
<td>✓</td>
</tr>
<tr>
<td>Identity management</td>
<td>14</td>
<td>Management of identity verification and certificate revocation of PGP certificates</td>
<td></td>
</tr>
<tr>
<td>Rating system</td>
<td>15</td>
<td>Tracking of users and contents points in a social voting system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>Rating system where customers can give feedbacks about a purchase</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>17</td>
<td>Management of software license validation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>Timestamping service, in order to prove a content has been produced before a specific date</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>Implementation of a lottery</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>Banking applications such as automated and distributed bank ledgers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>Implementation of a social cryptocurrency, to quantify social influence</td>
<td></td>
</tr>
</tbody>
</table>

provided with a public and a secret key: the secret key is used for signing transactions, while the public one is used as address in the system. So, no real-world identity is needed for transactions: this is a form of pseudonymity. A transaction can have multiple inputs and outputs. For instance, in economic transactions - i.e. transactions representing transfers of coins - the inputs are the addresses where coins come from and the outputs are the addresses of the recipients of the coins. Each input must be signed with the secret key corresponding to the address it represents.

Transactions are relayed in the P2P network and some peers, called miners, collect them together into a data structure, called block. Once a new block is assembled, it is relayed in the P2P network and, if valid, is chained to the current last block of the blockchain. Each block contains a reference to the previous block (that’s why it is called blockchain). After some time a block is stored the blockchain, the transactions of the block are considered confirmed.

A block is valid if it contains valid transactions and if miners have performed a computationally-hard puzzle, which consists in finding an hash of the block lower than a predefined target. The miner which adds the next block to the blockchain is the first which has assembled a valid block and has found a valid solution to the puzzle. This specific mining technique is called Proof-of-Work (PoW). The PoW allows to achieve distributed consensus, which means that all nodes agree on the same version of the blockchain and this blockchain contains valid transactions. Forks could happen in this chain of blocks, that is, there could be two contrasting branches of the chain. However, thanks to PoW, eventually one of the branches should be discarded and all nodes should agree on the same blockchain. In case of forks, the rule is that miners extend the longest branch or the one that has the most recent PoW. Moreover, thanks to PoW, the blockchain is hard to be tampered.

### B. Goal and Research Questions

The goal of our research is to understand whether the blockchain and, in general, P2P approaches can foster a private-by-design IoT, where IoT devices data are not entrusted to centralized companies, instead are property of the devices owner, who can decide which data share and with whom. As a first step of this research, we conducted a SLR to collect use cases of the blockchain and to collect evidence from the literature about the level of adaptability, integrity and anonymity of the blockchain.

To achieve the goal of our SLR, we formulated the following research questions (RQ):

- **RQ1** What are the use cases of the blockchain beyond cryptocurrencies?
- **RQ2** Are there any use cases applicable to the IoT?
- **RQ3** What are the implementation differences with respect to the Bitcoin blockchain?
  - **RQ3.1** Which data are stored in the blockchain?
  - **RQ3.2** Which mining techniques are used?
  - What is the degree of integrity (RQ4), anonymity (RQ5) and adaptability (RQ6) of the blockchain?

RQ1 and RQ2 aim at discovering in the literature the uses of the blockchain beyond Bitcoin and cryptocurrencies and which of them are applicable to the IoT (according to the authors of the papers).

By means of RQ3 we want to know the implementation choices described in found papers which differs from the ones of the Bitcoin blockchain: in fact, some of them could prove to be useful when applying the blockchain in the IoT.

Regarding RQ4, we take as reference the definition of integrity from the ISO 25010 [24], and we intend to characterize the attacks to which the blockchain is vulnerable and which could mine its integrity.

RQ5 concerns the need to further protect users’ privacy, by avoiding that IoT devices can be linked to their owner. Anonymity in blockchain systems is intended as pseudonymity plus unlinkability, where the latter is the impossibility to link an address of the blockchain system with a real identity or an IP address, and also the impossibility to understand that different addresses of the system belong to the same user. We
answer to this question by describing the techniques found in the literature which undermines anonymity.

Finally, RQ6 aims at verifying whether the blockchain is adaptable to the number of transactions. This is fundamental if we want to employ the blockchain in the IoT, where the number of transactions produced by IoT devices could be very large. Also for this question we employ the generic definition of adaptability from \cite{24}, and we narrow it by intending the adaptability of the blockchain as its ability to scale with the number of transactions.

C. Search process

To conduct the study, we followed the guidelines on SLR provided by Kitchenham \cite{25}. We used the string blockchain to search in the following digital libraries: IEEE Xplore; ACM Digital Library; SpringerLink; ScienceDirect; Google Scholar. We gathered 1511 papers. In order to decide which of them deeply analyze, we performed two exclusion stages - one based on titles and the other on abstracts - and we excluded papers regarding non-engineering aspects (e.g., papers addressing ethical issues of the blockchain or purely economic aspects of cryptocurrencies). Finally, we left 35 papers, from which we extracted the information necessary to answer our questions.

III. RESULTS

In this section we report the results extracted from the analyzed papers, organized by research question. Discussion on the results will follow in Section IV.

RQ1 and RQ2: Use cases and IoT

The results answering RQ1 and RQ2 are shown in Table I. Specifically, for each paper we report: the category \[^1\] in which we classified the paper, the reference to the paper in the bibliography, the usage of the blockchain, and whether the authors of that paper believe that it can be applied in the IoT.

RQ3: Implementation differences with Bitcoin

We report the answer to RQ3 in two tables: Table II refers to RQ3.1, while Table III to RQ3.2.

RQ4: Integrity

We report the attacks found in the papers analyzed to which the blockchain is vulnerable.

In \cite{28}, it is shown that most of the peers known by a peer of the Bitcoin network reside in its same autonomous system. This means that the P2P network is not well connected and there could be difficulties in the relay of new blocks added to the blockchain. This makes the achievement of distributed consensus hard.

The authors of \cite{29} shows that an attacker which controls a large number of nodes, even if with not high computational capabilities, could achieve an high fraction of the total computational power in small blockchain systems where there are few miners. This could threaten the integrity of the system, because the attacker would be able to cause intentional forks.

In \cite{30}, it is introduced the selfish mining attack. In this attack, a malicious mining pool decides not to publish the blocks it finds, thus creating a fork in the blockchain, where there are the public branch of the honest branch of the malicious pool. It keeps mining on its private branch until the public one approaches the private one in length. At this point, it publishes its own private branch; denial of service, because, if the attacker controls several nodes, it can prevent dissemination of information.

In \cite{33}, an expansion of selfish mining called stubborn selfish mining.

\[^{1}\]The following categories have been defined: data storage management, trades of goods and data, identity management, rating system, other

\begin{table}[h]
\centering
\caption{Data Inserted into the Blockchain (RQ3.1)}
\begin{tabular}{|l|l|}
\hline
Paper & Data in the blockchain \\
\hline
5 & Access policy and reference to data \\
12 & Key to access sensors data and multisigned transactions to exchange bitcoins with commodities \\
14 & Revocation or verification address of PGP certificates \\
15 & Triples (ID, PK, action), where action refers to registration, update or verification of the public key PK \\
11 & Messages metadata of the decentralized application \\
19 & Specifications useful for license validation \\
13 & Data purchased from sensors \\
17 & Reference to published content \\
18 & Rating information \\
20 & Hash of contents to be timestamped \\
9 & Spend conditions, file contract, storage proofs and arbitrary data \\
7 & Payment contracts \\
21 & A lottery contract \\
16 & Information for registration, revocation and update of public keys \\
8 & Access policy, reference to data and other information to recover data \\
3 & Reference to data and other metadata \\
10 & Transactions to safely deposit bitcoins \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{Mining Techniques (RQ3.2)}
\begin{tabular}{|l|l|}
\hline
Paper & Mining technique \\
\hline
5 & New measure of trust to give more weight to trusted nodes in mining \\
26 & The miner producing the minimum block hash is selected for mining \\
23 & Proof of Stake Velocity \\
27 & Proof of Space \\
\hline
\end{tabular}
\end{table}
mining is described. Results show that in some situations it could be more advantageous than selfish mining.

<table>
<thead>
<tr>
<th>De-anonymization technique</th>
<th>Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple inputs</td>
<td>[34]</td>
</tr>
<tr>
<td>Change address</td>
<td>[34]</td>
</tr>
<tr>
<td>Associations with IP</td>
<td>[28], [37]</td>
</tr>
<tr>
<td>Usage of centralized services</td>
<td>[26], [27]</td>
</tr>
</tbody>
</table>

**RQ5: Anonymity.** In Table IV we classified the papers according to the de-anonymization techniques they mention. We identified four categories of de-anonymization techniques: multiple inputs, change address, associations with IP and usage of centralized services.

When an user issues a transaction with multiple addresses as inputs, she reveals to own all those addresses. For this reason, in [34]–[36] the authors can safely state that all the input addresses of the same transaction belong to the user that issued that multiple-inputs transaction.

In systems like Bitcoin, in some transactions users send coins to a particular address that belongs to themselves, called change address. In [34]–[36], the authors are able to link this change address to other addresses of the same user.

In [28] and [37], starting from some hypotheses and analyzing network traffic, the authors are able to associate Bitcoin addresses with IP addresses.

In [36]–[38], usage of centralized services that keep track of associations between more addresses of the same user or real identity of the user and her address is considered a risk for the anonymity of the user.

**RQ6: Adaptability.** We found only three papers reporting information on adaptability, with a coarse detail level. In [13], where the blockchain is used to purchase sensors data via bitcoins, the authors state there are scalability issues due to the exploding number of transactions and sensors data permanently stored on every Bitcoin node. In [8], according to the authors the blockchain cannot scale to deal with many complex transactions. For this reason, they propose that computations and data storage should not be done by each node of the network, instead by a small subset of them working on different parts of data. The authors of [31] point out that scalability is a problem because every node of the blockchain should verify each block and transaction issued.

**IV. Discussion**

**Use cases and IoT.** As one may notice in Table I cryptocurrencies like Bitcoin are just one of the possible use cases of the blockchain. In some cases, the blockchain is employed for decentralizing services that so far have been provided by centralized trusted entities (e.g., PKI or timestamping). Moreover, we observed that only 4 of the 18 found use cases are considered applicable to the context of the Internet of Things. Two of them, [12], [13], use the blockchain for trading data collected by sensors of IoT devices and other goods. In the third [15] the IoT is mentioned as a possible field in which each device is identified by a public key to interact with other devices through the blockchain. Finally, the mechanism described in the fourth [8] can be employed to store and manage data collected by IoT devices, in a decentralized and private-by-design fashion. This last use case and all the others classified as “Data storage management” (even if not explicitly thought for the IoT) are in line with the goal of our research: encouraging a private-by-design IoT where devices data are not entrusted to centralized companies. Just to mention some use cases of this category, [9] and [6] are both decentralized storage platforms, where the blockchain is employed for implementing storage audits, useful for detecting any non-authorized deletion or modification of data. These audits are performed by storing the hash of the data in the blockchain. Then the data owner periodically sends a challenge to the host of the data and checks the correctness of the response using the hash in the blockchain. Any non-authorized deletion or modification of the data entails a wrong response, so any abuse can be detected. In [5] the blockchain enforces access policies that define which data of a user share and with whom. It leverages public key cryptography: each entity is represented by a public key and the policy specifies restricted accesses for the public keys of the interested entities. Only the data owner has full access to her data. Policies are stored in the blockchain and the nodes of the blockchain verify whether they are respected. From observing such applications of the blockchain in the literature, we can conclude that:

*It has been documented that the blockchain can be used for detecting abuses on data and defining access policies, without the need of entrusting people’s data to centralized companies.*

**Implementation differences.** Regarding RQ3.1, we observed that in some papers data are inserted in the Bitcoin blockchain, employing the 80 bytes of Bitcoin transactions reserved for arbitrary data; in other papers, a customized blockchain is used to store the data. Regarding RQ3.2, the mining techniques reported are all less computationally-expensive alternatives to PoW: in [5] the PoW is facilitated to trusted nodes; in [26] the selection of the miner which adds the new block depends on luck and not on the computations performed by the miner; in [23] no computations are required, the miner is chosen according to the age of coin she owns; in [27] the miner is chosen according to her amount of space, and not her computational capabilities. In the IoT scenario, it could be useful to take into consideration one of the less computationally-expensive alternatives to PoW showed in Table III. In fact, PoW requires very high computational power, and so IoT devices with limited capabilities would not be able to add blocks in the blockchain. However, before designing a blockchain with an alternative mining technique which allows all IoT devices to fairly participate in the system, we should further analyze what are the security properties provided by the PoW, which up to now is one of the key factors allowing to achieve distributed consensus. We refer to the following discussion on integrity for more on that. Therefore, from our analysis we conclude that:

*Arbitrary data can be inserted in the blockchain, so in theory any applications (not only cryptocurrencies) can be*
developed on top of it. Some less expensive alternatives to PoW are documented in the literature.

**Integrity.** Several countermeasures have been proposed for some of the attacks described in Section II. For example, for selfish mining attacks, [30] and [39] propose modifications in the way miners decide which block to extend, in order to decrease $\gamma$, i.e. the portion of honest miners which extend the blockchain proposed by selfish miners.

However, what we evince from the results is that the greatest risk for the blockchain integrity is represented by the presence of misbehaving miners which own an high proportion of the computational power of the system. They could cause forks to the blockchain, bringing to a situation where distributed consensus is difficult to achieve and some past data could be lost. In addition, they could pollute the blockchain with invalid data or transactions. Such risk is avoided in already large and stable blockchains like the Bitcoin one, because obtaining an high proportion of the computational power is hard thanks to the difficulty of the PoW and to the great number of miners, which in addition are incentivized to act honestly. For this reason, starting a completely new blockchain, which does not have a critical mass in the initial phase, is risky. Even if understanding in depth the security of the Bitcoin blockchain is not trivial, because it depends also on socioeconomics factor, in this moment Bitcoin is the most stable and secure blockchain system. So, instead of designing a new blockchain from scratch, our suggestion is to develop distributed applications for the IoT on top of the Bitcoin or another secure and stable blockchain. This can be done by leveraging a layered architecture, like the one proposed in Blockstack [40]. In this solution, the additional functionalities of the application are defined in another layer on top of the blockchain. Moreover, the blockchain is hidden at the application level, so low-performance IoT devices are not required to compute the PoW. To conclude:

> **We believe that the most secure approach is to develop IoT applications on top of an already existing stable blockchain, where PoW and the great number of honest miners ensure integrity, and avoid that misbehaving miners can obtain a large portion of computational power.**

**Anonymity.** From the results documented in papers of Table IV, it is possible to de-anonymize a user by analyzing network traffic or the blockchain itself, since it is public. So, pseudonymity is not enough to guarantee total anonymity. Countermeasures are proposed in [15], [31], [38], [41], [42]. In [38], [41], [42], mixing protocols are analyzed. The main idea behind mixing protocols is that a user sends some coins from an address and receives them back to another address in a way that it is difficult to discover the correspondence between input and output addresses of the same user. Also the fair exchange protocol described in [31] is based on the same principle of mixing protocols and allows two parties to securely exchange money. In the work regarding the blockchain used as PKI [15], the authors describe a method for the user to update her public key without linking it to her ID in the system. To conclude:

> **Pseudonimity is not enough to achieve total anonymity.**

Solutions that reduce the possibility of linking IoT devices to their owner should still be analyzed in future work.

**Adaptability.** As documented in Section III, the scalability issue of the blockchain is reported in three papers. Actually, there are two main scalability issues. The first is that, when the number of transactions grows, the blockchain increases in size, and it becomes expensive to store it, especially for IoT devices with limited resources. This issue can be addressed by the layered architecture described also for the integrity. In this architecture, where the blockchain is separated from the application layer, IoT devices with limited resources store only the portion of the blockchain they need for their own transactions (the so-called thin clients, already present also in Bitcoin).

The second issue is the low throughput of transactions - a typical issue of the Bitcoin blockchain, which we did not find in the papers but is largely discussed within the Bitcoin community. The low throughput is due to the difficulty of the PoW and to the maximum size of a block, which is set to 1 MB. This issue represents a tradeoff between scalability and security. In fact, regarding the PoW, if its difficulty is reduced, the throughput will be higher, but at the same time it will be easier for an attacker to cause forks in the blockchain. Regarding the block size, if its maximum is increased, the throughput will increase too, but it will be more difficult to validate transactions: this implies that only few nodes will be able to do it, and so the power of Bitcoin will be concentrated in few hands. Again, a solution could be a layered architecture, where not all operations performed at the application layer require a transaction in the underlying blockchain. However, this may not be enough for the IoT, where the blockchain should support billion of devices. For this reason, we retain that, even if in this moment the Bitcoin blockchain is the most secure, it could be prohibitive to leverage it in the IoT because of its scalability issues. Instead, it could be more convenient to employ another stable and secure blockchain which provides higher level of scalability with respect to Bitcoin. To conclude:

> **The scalability issues of the Bitcoin blockchain make it poorly suitable for the IoT, so we suggest to develop IoT applications on top of another secure but scalable blockchain.**

**In future work, we will test different blockchains to find a suitable one, in which the trade-off between scalability and security is acceptable.** Moreover, we suggest to adopt a layered architecture which supports thin clients to allow IoT devices with limited resources to store only a portion of the blockchain.

**V. Conclusions and Future Work**

We conducted a Systematic Literature Review to investigate which are the uses cases of the blockchain in the literature and which factors affect integrity, anonymity and adaptability of this technology. The ultimate goal of our research is to leverage the blockchain and P2P approaches for a private-by-design IoT where data produced by devices are not entrusted to centralized companies.

We reported several uses of the blockchain. Even if few of them are explicitly thought for the IoT, we found several
use cases for a private and decentralized data management, which are in line with the goal of our research. Regarding the integrity and the adaptability, we found that large blockchain systems like Bitcoin are the most secure, but at the same time Bitcoin scalability issues make it little suitable for the IoT. Regarding the anonymity, we found that in the blockchain only pseudonymity is guaranteed.

To address the integrity and the adaptability issues, our future work will consist in testing existing secure and scalable blockchains and in designing a layered architecture for IoT applications on top of the most suitable of them. Moreover, we will investigate mixing protocols and other solutions to achieve anonymity and further protect people’s privacy.

ACKNOWLEDGMENT
This work has been done with the Joint Open Lab SWARM and it is supported by a fellowship from TIM. We also thank Dr. Simone Basso for his valuable feedback.

REFERENCES