Paavolainen, Santeri; Nikander, Pekka

Security and Privacy Challenges and Potential Solutions for DLT based IoT Systems

Published in:
2018 Global Internet of Things Summit (GIoTS)

DOI:
10.1109/GIOTS.2018.8534527

Published: 15/11/2018

Please cite the original version:
Security and Privacy Challenges and Potential Solutions for DLT based IoT Systems

Santeri Paavolainen and Pekka Nikander
Department of Communications and Networking
Aalto University
santeri.paavolainen@aalto.fi, pekka.nikander@aalto.fi

Abstract—The use of distributed ledger technologies introduces new security and privacy challenges. These challenges are dependent on properties of the ledgers, such as transaction latency and throughput. Some use cases may be outright impossible to implement securely, or in a privacy-retaining manner. Consequently, it is important that these concerns are taken into account when distributed ledger technologies are evaluated and selected as building blocks for higher-level systems. In this paper, we illustrate these concerns through use case examples. We discuss the implications these concerns on the use of distributed ledgers within higher-level systems, such as in SOFIE, a DLT-based approach to securely and openly federate IoT systems.

Index Terms—Internet of Things; distributed ledgers; blockchain; security; privacy.

I. INTRODUCTION

Research and innovation on blockchains and other distributed ledger technologies (DLT) has proliferated after the success of Bitcoin. This initial popularity led Gartner to place blockchains at the top of the hype cycle in 2016 [1]. As with most highly hyped technologies, many mundane concerns, such as security and privacy, play a catch-up game. In this paper, we outline some major security and privacy challenges related to DLT technologies and discuss how they related to the Internet of Things (IoT) systems. We do this in the light of three simplified use cases, thereby illustrating the challenges and potential solutions involved.

In practical terms, a distributed ledger is a massively replicated append-only data structure. Data can be added to it, typically by anyone. Once data has been added to the ledger, it cannot be removed. This inability to remove data is perhaps the most important essential feature of distributed ledgers. If data could be removed, it is questionable if the system can any more be called a ledger at all.

The other essential feature of a distributed ledger is that the data is massively replicated. In present systems, such as Bitcoin and Ethereum, all maintainers keep an identical copy of all the data. Open ledgers allow anyone to join the network and download a copy of the data, at any time. Hence, there are thousands of copies of the data, stored all over the world. While the ledgers may become more efficient in the future in the sense that not all maintainers keep all data, we surmise that in order to retain the massive replication, even the future systems will store hundreds if not thousands of copies for each datum.

We focus on the security and privacy challenges related to the use of distributed ledgers in the context of IoT devices. More specifically, we leave beyond the scope of this paper any security problems in the ledgers themselves.1 Furthermore, security and privacy risks that are merely related to the payment aspect of blockchains are beyond the scope of this paper, unless they are directly related to IoT applications.

To our knowledge, this paper is among the first systematic reviews of the security and privacy risks related to combining DLT and IoT.

The rest of this paper is organised as follows. First, in Section II we outline three distinct use cases that we use to illustrate some of our observations. Then, in Section III, we discuss the main security and privacy challenges we have observed. In Section IV we briefly outline some tentative solutions. In Section V, which is very brief, we discuss the related work. Finally, Section VI summarizes this paper and discusses potential future work.

II. SAMPLE USE CASES

We focus on three illustrative use cases: a door lock, a transportation container as a part of a larger IoT system, and a smart building with multiple sensors and actuators. Starting with the lock, it should provide the following essential features:

- The lock shall open when an authorized “key” is present, and otherwise not.
- All attempts to open the lock, whether successful or not, must be duly recorded.
- The lock shall work also when there is no Internet connectivity, potentially with reduced functionality.

A trivial approach would store into the DLT an up-to-date list of the identities2 of the authorized “keys”. In a similar manner, all accesses may be recorded to the DLT as separate transactions. Limited offline functionality could be implemented by caching the latest known valid list of authorized keys in the device memory.

1See the literature review by Conoscenti et al. [2] for a summary of various security threats specific to distributed ledgers.

2Not really identities in the strict sense, but e.g. public cryptographic keys or their fingerprints.
The case of a transportation container is more complex. We focus on a container during transit. A number of IoT devices are relevant: the container itself, the vessels or vehicles used to transport it, any lifts or cranes used to handle it, and potentially also any storage spaces where the container may need to wait. All of these devices belong to potentially different parties, with partially conflicting interests, especially in view of liabilities, if a container gets lost, damaged, or compromised.

The essential features for the IoT system appear to be the following:

- The system shall always know the whereabouts of and the currently responsible party for all containers.
- When a container arrives at a transit terminal, the responsibility for the container shall be transferred from the arriving vessel or vehicle to the appropriate terminal operator.
- When a container leaves a transit terminal, the responsibility for the container shall be transferred from the terminal operator to the departing vessel or vehicle.
- All transfers of responsibility shall be stored in non-revocable and non-repudiable record.

Again, there appears to be a trivial solution: Simply use a DLT to record all events on all containers. And, again, there are a number of emerging challenges, discussed below.

Our final example is a smart building, with a number of sensors. In this case, we assume that the building and all the sensors and actuators are owned by a single party.

Here the essential features appear to be quite similar to the cases above:

- The lights, ventilation, etc, shall be adjusted based on human presence and action.
- The sensor data and actuator adjustments shall be recorded.
- The adjustment system must work (at some level) even if there is no Internet connectivity.

As in the two other use cases, there are a couple of simplistic ways to apply DLTs, both with their problems. Firstly, of course, a DLT can be used to record sensor data and actuation events. Secondly, it may appear clever to use the so-called smart contracts to process sensor data and generate actuation commands. However, in this case we have to question even the generic applicability of DLTs; their benefits seem meagre compared to the problems related with them.

III. CHALLENGES

Given our three example cases — lock, container, and building — we now consider the security, privacy, and some other challenges emerging from the proposed simplistic approaches. We cover the various aspects one at a time, and briefly note current problems in the light of the examples.

A. Security challenges

The usually considered computer security aspects include integrity, confidentiality, and availability. To achieve those, authentication, authorization, key management and timely revocation of access rights are needed. Furthermore, in the case of IoT we have to consider also physical security and safety as well as the storage and backup of private keys.

Integrity. One of the main benefits of DLT systems is the (near) impossibility of changing the data in the ledger. However, in today's DLT systems this comes with a high cost: the so-called full nodes must store the whole transaction history, which is easily gigabytes or terabytes, typically preventing IoT devices from acting as full nodes.\(^3\) Even in situations where memory saving techniques would enable some larger IoT devices to participate as essentially full nodes, care must be taken to ensure they will also meet future storage needs.

In quite practical terms, any individual IoT device must either have access to a trusted full node, or be one, in order to achieve the full security benefits. Furthermore, to cover situations with expected intermittent connectivity, the full node must be available locally, e.g. in the same local network with our lock, container, or building. From the devices' point of view, this is similar to trusting a centralized server. Some IoT systems may be able to avoid the use of fully trusted full nodes by either accepting increased latencies during transaction verification, or by accepting decreased security guarantees.

The container use case appears to have most to gain from the integrity guarantees of DLTs. In the container case, there are multiple parties that must record the movements of the container and refer to the recordings. For these parties, it is their interest to accept transfers of responsibility only when the integrity of the ledger can be confirmed. Without going into the details, the ability of the individual parties to record their view each independently and then reviewing the views of the other parties appears beneficial. Here the DLT facilitates the situation through providing an integrity protected storage without requiring any direct trust relationships. In the other two cases, the integrity of the historical record may not be as critical as in the container case, especially if the current state is secure and valid.

Availability. Another major purported benefit of DLTs is availability. With the thousands of replicated nodes, the DLTs are assumed to provide unprecedented availability. Unfortunately, this benefit is difficult to achieve with resource-constrained IoT devices, such as those used in the lock or building use cases. Their limited storage capacity prevent from keeping a full copy of the ledger, requiring the devices to rely on either remote full nodes, or a local trusted node. Dealing with intermittent connectivity can also affect availability due to the time required for DLT synchronization. The building case is probably the easiest to engineer for having high availability of DLT access, with the lock being the hardest and the container somewhere in between.

Hence, in the light of our example cases, the two main benefits of DLTs — integrity and availability — do not appear to provide much benefits to many IoT systems, at least if the

\(^3\) A typical IoT device today has at most a few megabytes of memory, often less, e.g. 64–512 kb.
DLTs are applied in a simplistic and straightforward manner. We surmise that this is a general property of so-called siloed IoT systems.

**Authentication and authorization.** For authentication and authorization, IoT devices could use the DLT as a repository of trust-related information and the IoT device would rely on the timeliness and security properties of the ledger to ensure that most recent and correct configuration was used. Alternatively, a smart contract in a DLT could be used to actively verify access and authorization by sending a transaction with suitably protected parameters to the smart contract, and then reading the response of the transaction from the ledger. In either case, the IoT device must be configured with cryptographic keys, smart contract addresses, etc. to provide security and integrity.

The first method can offer higher availability of up-to-date authentication and authorization information than centralized systems, although this only applies to IoT devices with reliable and timely DLT access either directly or via trusted nodes. For devices with intermittent or easily disrupted connectivity, the first method carries a possibility of using stale data, especially if timely operation is required (e.g. the lock case). The second method may allow for higher flexibility, but it suffers even more from intermittent connectivity, and unless some secondary authentication mechanism is used, it suffers greatly from DLT transaction latencies.

**Revocation.** In a situation where access or other rights are revoked from a party, it is often crucial that the revocation event is distributed in a timely and predictable manner. However, the large majority of today’s DLTs are relatively slow. In Bitcoin it may take several tens of minutes before a new transaction gets validated and recorded. While Ethereum is faster, writing new information may still take in the order of a minute. Here Iota[3] and Corda [4] appear to be substantially better, with the average recording time being in the order of seconds. However, it can be conceived that a resourceful adversary could arbitrarily delay a revocation from being accepted to the ledger by incentivising the individual nodes into not accepting the transaction.

Hence, while DLTs appear as a great mechanism for storing and revoking authorisation data, the long confirmation latencies may make the present day DLTs for IoT unusable in practice for use cases with short to moderate timeliness requirements.

**Confidentiality.** All the information in a DLT is replicated[4] and therefore public by definition. Of course, some of the data in the DLT may be encrypted. However, given the permanent nature of the data and the continuous development of cryptanalysis, there is a non-negligible probability that any encrypted public data will become decryptable at some point in the future. Therefore it is highly advisable to store confidential data into a DLT even in an encrypted format.

This applies especially to private or symmetric cryptographic keys, which should never be stored into a DLT or any other publicly available storage. In other words, the management of such keys must take place outside of the DLT. This also means that DLTs shall not be used to backup private keys.

Thus, if confidential information needs to be transferred to or from an IoT device, this requires alternate information paths to exist, which in turn may reduce the overall availability of the IoT system. Alternatively, a hybrid or multi-ledger system may be employed, with the confidential information stored in a private DLT, accessible only to the participating IoT devices, and only the public portion of operations (user identification, payments etc.) performed on the public DLT.

**Using public keys as identifiers.** In the IoT world, storing and backing up private keys may present a major problem, if the keys are associated with value or other key-specific semantic meaning. In general, while the IoT devices are small, they may still contain a handful of private keys that are specific to the device. In most cases, these keys cannot be stored or backed up anywhere else, or storing them elsewhere is cumbersome and adds additional security vulnerabilities into the system. Furthermore, many IoT devices operate in uncontrolled environments, and may be physically accessible by adversaries. Hence, a common practice is to keep the device specific keys as such, associating them just with the specific device and nothing else.

### B. Privacy-related challenges

From the privacy point of view, both of the main DLT properties — immutability and availability — may endure privacy challenges. Furthermore, there are challenges related to privacy laws, including the European GDPR and the right to be forgotten.

**Immutability** of the data stored in blockchains can easily cause problems with privacy. The increasing pool of data, available in the ledger, can be mined for insights, and dedicated techniques, such as correlation attacks, can reveal even obfuscated information. Therefore, careful analysis is needed to determine what information should and should not be stored in any DLTs and what methods should be used to protect the information. In most situations only hashes of the actual data (e.g., the root of a Merkle tree) will be stored to blockchains. Extremely sensitive data, such as cryptographic keys, must only be stored privately.

Transactions are always traceable in DLT, by the very definition a distributed ledger. While transactions cannot be directly tied to individuals — unless they contain unencrypted personally identifiable information — any leakage of identifiable information will allow the tracing of all past and future transactions made by the entity tied to a transaction. While information-hiding techniques such as the use of tumblers makes it possible to obfuscate in some cases the transaction parties, the success of obfuscation depends on the properties (and security) of the tumbler service used and the type of transaction attempted. It should also be noted that future developments in identifying transaction patterns may in the

---

[4] We surmise that even in the future DLT systems where the nodes do not need to store the full data, the replicas of each datum must still be stored at essentially random nodes. Doing otherwise is likely to unnecessarily complicate the system and may easily lead to new security problems, e.g. open venues for new types of denial of service attacks.
future lead to previously obfuscated transactions becoming traceable.

C. Internet of Things point of view

In addition to the traditional security and privacy concerns, IoT devices pose a number of challenges that are specific to the very nature of the IoT devices. That is, contrary to most other ICT systems used widely, IoT systems tend to be used long after their installation, from several years to even half a century. Furthermore, most of today’s IoT devices do not have any practical means of upgrading them, other than physical replacement, which may be prohibitively expensive. In this section, we have a brief look at some of these aspects.

To start with, IoT devices often have limited reconfigurability or none at all. They may become “stuck in the past” regarding newer technological developments. Changes in the DLT infrastructure and protocols may cause IoT devices to either become isolated from the DLT, or to only have limited functionality available. While it may be feasible to run a deprecated, or backward-compatible version of IoT backend systems, it would seem unlikely that it is possible to run an alternate DLT network for the purpose of supporting old IoT devices. In this manner an IoT system trying to gain reliability and security advantages of a public DLT system is also at the mercy of that DLT system’s later developments. Long-term changes in a DLT’s development may also be difficult to predict, as even an open ledger has an implicit governing body subject to potentially diverging interests and incentives [5]. Considering the use cases, locks and buildings are relatively accessible for upgrades, while containers would be likely to be upgradeable only during select time windows during their travels.

Power requirements are often critical for IoT devices, and a large portion of IoT protocol concerns are related to the power requirements of transmission of data over the network. As noted before, devices with constrained CPU, storage, and/or power capacities cannot participate in DLT networks as full nodes, and are unable to store or process the full ledger. Furthermore, integrating DLTs is likely to increase in network traffic which in turn impacts the power usage of the devices. Thus, it becomes important that any use of DLTs takes the limited power budget of IoT devices into account, for example, through the use of protocols that allow tradeoffs between DLT security, and latency guarantees and power requirements. Power requirements for the lock case is especially problematic, as locks are needed to operate on battery power for extended periods of time. The same logic applies to unpowered containers, but is not so relevant for powered containers.

Another difference between many IoT systems and typical ICT systems is that the IoT systems directly control real life utilities or other functions whose failure may have severe or even fatal consequences to humans. Hence, their resilience and robustness requirements may be decades more stringent than even e.g. for financial systems.

IV. Tentative approaches

Given the considerations above, we conjecture that in most cases individual IoT devices should not be directly connected to any DLT. Instead, typical well engineered approaches will be hybrid systems where the individual, resource constrained, IoT devices talk only to a handful of local, trusted “gateways.” These gateway nodes will then have more resources, be better protected and upgradeable, and — perhaps most importantly — any mission critical functionality will not depend on any DLTs being continuously available.

Hence, in the rest of this section, we focus on the consequences of this conjecture, discussing how the security and privacy challenges might be addressed within the IoT platforms, i.e. infrastructure nodes (including above mentioned local “gateways”) that process IoT data and take part on the coarse grained control of the missions the IoT devices are implementing through actuation. This approach may be considered as an example of the so-called hybrid DLT systems, where a part of the system is an “open” blockchain while other parts of the system are “closed” or permissioned.

From the integrity, confidentiality, authentication and authorisation point of view, the baseline approaches are well progressing in some of the ongoing work, e.g. in the Sovrin Foundation “identity” blockchain [6]. One basic idea is to separate all identity information into individual attributes, such as birth date, first name, and present only the attributes necessary and nothing else. From the DLT point of view, this means that the DLT itself works in a role somewhat similar to a traditional certificate authority (CA) or certificate revocation list (CRL). In other words, the DLT stores data about trust anchors and their relationships, while the actual data relating to privacy sensitive identities is stored by the parties themselves. Hence, the DLT is used to maintain the integrity of the trust anchors while session and data confidentiality and any decisions requiring authentication and/or authorisation take place outside of the open DLT.

Considering availability and blockchain latency, one approach is to combine several blockchains, c.f. e.g. Polkadot [7] and SOFIE [8]. Polkadot outlines a scalable, heterogeneous multi-chain protocol aimed to be backwards compatible with existing blockchain networks. The goal is an extensible system that has a lower cost structure than a standard blockchain design. The SOFIE project attempts to take the approach one step further in the IoT space, by federating IoT systems by with an inter-ledger transaction layer.

From the privacy point of view, an attribute-oriented approach, promoted e.g by the Sovrin blockchain, may completely dismiss the use of permanent identifiers, replacing them with secure but ephemeral peer-to-peer connections that are associated with security-related attributes. Especially when the attributes are combined with zero knowledge protocols, a party may prove that it has certain rights or posses certain attributes without revealing anything about its identity.

Another approach, promoted by e.g. Sovrin, MyData [9], SOFIE, and many others, is storing all privacy critical data
off-chain and only referring to the data from the chain, if so desired. In general, strictly confidential data must not be directly stored in a DLT, not even in an encrypted form, due to the high probability that all encryption algorithms will become weak sooner or later. Hence, for example, the Sovrin approach is that the parties themselves store their privacy-critical attributes and may use zero knowledge proofs to show that they possess certain attributes without revealing any non-ephemeral identifiers or other knowledge that would allow their “identities” to be linked.

A variant of this would be storing only partial data in a DLT. In such an approach, the data would be cryptographically split (or “shared”) [10]. An almost opposite approach would be storing the whole state in a DLT [11]. W.r.t. our use cases, such hybrid approaches would most probably be very useful for the lock and building cases, while the container use case could possibly be based on a more direct DLT approach, depending on the latency requirements.

V. Related work

There appears to be very few peer reviewed papers in the domain of applying blockchains to IoT. Furthermore, those published seem to err more to the side of proposing how blockchains could be used with IoT rather than systematically analysing the potential problems. In this section, we briefly summarize the few papers and about a dozen of newsletters and blog posts covering the security and private issues relating to DLT and IoT integration.

To our knowledge, Fremantle and Scott [12] were the first authors that discussed IoT security and privacy, also considering blockchains. However, they merely remarked that blockchains may have potential in solving the cloud integrity and authentication problems for IoT, not considering the potential challenges. Conoscenti et al [2] gave a systematic literature review on blockchains and IoT, finding only four use cases explicitly designed for IoT. They briefly considered blockchain security and noted that user-related privacy issues may arise, without really going much deeper. Dorri et al [13], [14] have proposed a solution where each smart building has a separate local blockchain, though without proof-of-work mining and with a hierarchical structure. Their approach to privacy issues related to the use of blockchains is to store the private information primarily on the user-controlled private blockchain. While this approach is suitable for environments entirely under user’s control, it cannot be extended directly to situations where separated IoT systems need to communicate. Kshetri [15] discussed the applicability of blockchains to IoT security in the light of a number of IoT security incidents, most of the time suggesting straightforward solutions, and therefore probably suffering from many of the problems we have outlined above. Laszka et al. [16] considered a electricity trading use case, where public trading transactions in a DLT have a potential to expose personal information (e.g. electricity usage patterns) through automated trading by the IoT devices comprising of the smart grid. However, they consider a narrow concern and do not discuss more general problems related to the use of DLTs by IoT devices. Khan et al. [17] perform a systematic review of possible attacks specific to IoT systems, and discuss potential benefits of using blockchains regarding the discussed attack categories. Many of the attacks described by Khan et al. can also be used to disrupt IoT devices’ access to DLTs; however, to us their approach of using blockchains appears optimistic and glosses over a large portion of the practical problems discussed in this paper.

In an BBVA Open Mind blog post, Banafa [18] claimed that “Blockchain technology is the missing link to settle scalability, privacy, and reliability concerns in the Internet of Things.” As should be clear by now from above, we by-and-large disagree. His second article in the IEEE Internet of Things newsletter [19] appears to be somewhat more balanced, but still claimed that the “Blockchain technology is the missing link to settle privacy and reliability concerns in the Internet of Things.” However, in addition to heavily promoting blockchains as the “perhaps [being] the silver bullet needed by the IoT industry”, he acknowledges that the are challenges related to blockchain scalability, power and storage consumption, confirmation latencies, general lack of human skill, and legal and compliance issues.

The media and industry analysts are — more often than not — focusing on the apparent benefits of IoT and DLTs. Consider, for example, reports from Accenture [20] and Forbes [21] which are quite uncritical in their portrayal of IoT and DLTs. Even in situations where potential problems are highlighted, there seem to be focus on the technology, operational, legal, and compliance issues.

VI. Conclusions

Topping at Gartner hype cycle in 2016, blockchains and other DLT have been suggested as a security solution to numerous areas, some people even claiming it perhaps being “the silver bullet needed by the IoT industry” [19]. We have briefly but systematically discussed a number of security and privacy challenges related to using DLT in the context of IoT systems. Based on our admittedly early analysis, while admitting that DLTs may have a role in securing some IoT system use cases, to us it appears unwise to use (open) DLTs directly with IoT devices or for storing IoT related data as such. On the other hand, using more advanced solutions where the DLT role is diminished to that of a traditional trusted third party and/or for storing fingerprints of data, possibly with smart contract oracles, may well appear quite useful. In such solutions the security and privacy critical data is stored off-chain, in more traditional and separately protected systems, using open DLTs only to facilitate interoperability by providing distributed trust anchors.

Hence, to us it appears that more work is needed before we can integrate open DLTs into IoT systems in such a way that where the business benefits clearly outweigh the potential security and privacy problems. Firstly, we believe that a viable inter-ledger approach needs to be developed, allowing multiple ledgers to be used in the same time. Secondly, we need to identify the typical patterns of which data should be stored
into a public ledger, which is better left in a private ledger, and what should be left outside of ledgers altogether. In general, we expect various hybrid approaches to emerge, wherein the DLTs will typically have a relatively minor but important role.

Acknowledgments
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 779984.

References