Peer to Peer for Privacy and Decentralization in the Internet of Things

Marco Conoscenti
Nexa Center for Internet & Society
DAUIN-Politecnico di Torino
ITALY
Email: marco.conoscenti@polito.it

Antonio Vetrò
Nexa Center for Internet & Society
DAUIN-Politecnico di Torino
ITALY
Email: antonio.vetro@polito.it

Juan Carlos De Martin
Nexa Center for Internet & Society
DAUIN-Politecnico di Torino
ITALY
Email: demartin@polito.it

***Abstract*—In the Internet of Things (IoT) era new connected devices will spread highly sensitive personal data. Sending this type of data to centralized companies represents a serious risk for people's privacy, since economical or political interests could lead to an illegitimate use of personal information (as shown by Snowden's revelations). With the purpose of overcoming such status-quo, our research goal is to develop software systems according to the notion of *decentralized private-by-design* IoT. The basic idea is that data produced by personal IoT devices are safely stored in a distributed system whose design guarantees privacy, leaving to the people -the real data owners- the decision of which of them to share and with whom. To achieve this goal, a possible solution is to leverage the use of Peer-to-Peer storage networks in combination with the blockchain. However, such architecture, despite promising, embeds still limitations, especially in terms of scalability. In this paper we discuss our research motivation, we describe our research idea applied in a possible scenario and we present the scalability problem.**

***Index Terms*—privacy; internet of things; peer to peer; blockchain.**

I. Motivation and research problem

As defined by ITU [1], the Internet of Things (IoT) refers to the network of numerous physical objects connected to the Internet. Such devices will acquire information about the surrounding environment and will exchange data with other devices or platforms, thus enabling several new services.

Despite the indisputable benefits provided by these services, the IoT can also entail serious privacy issues, as long as data spread by IoT devices reveal information about health, behaviors and private life aspects of their owners.

Nowadays, data that we produce through our devices are processed and stored by centralized Internet companies, i.e., companies whose control is in the hands of one or few people. Even if data are distributed among different servers of the company, there is a single point of control and access to them, that is the company which owns those servers.

This approach of entrusting people's data to centralized Internet companies has already proved to constitute a threat to people's privacy. As a matter of fact, Edward Snowden revealed numerous mass surveillance programs and -among them- the PRISM program [2], which allowed NSA to direct access servers of the main Internet companies (like Google, Facebook, etc.) in order to obtain stored sensitive data. Also according to Benkler, when storing people's data in a single centralized point, the risk is that whoever can access that point can exercise power over people [3]. Such risk is even more concrete in the context of the IoT, since billions of devices will continuously collect fine-grained personal information.

Starting from these considerations on the risks of the current data management model for our society, in this paper we present our idea for a *decentralized private-by-design* IoT. It consists of three main components. The first is a Peer-to-Peer (P2P) network where IoT devices data are privately stored (instead of entrusting them to centralized companies). The second is the blockchain: a P2P ledger firstly used in Bitcoin for economic transactions [4], which in our solution is needed for certifying IoT devices data and for incentivizing peers to store data. The third is access rules, that allow people to define which of their data to share and with whom. This scenario preserves the main benefit of the IoT - i.e., enabling useful services by processing data of new connected devices - and at the same time protects people's privacy.

However, before adopting P2P and blockchain technologies in the IoT, some important technical issues have still to be addressed, as we observed in previous work [5]. Among them, from the software engineering point of view, the most critical is the limited adaptability of the blockchain. The *adaptability* is the software quality defined in ISO-IEC 25010 [6]. As explained in Section IV, in our specific case it is intended as the scalability of the blockchain with the number of transactions.

II. Context: the blockchain [1]

As said in Section I, the blockchain was first used as a P2P ledger for registering Bitcoin economic transactions [4]. A transaction represents a transfer of the Bitcoin cryptocurrency. New transactions are relayed to all peers of the blockchain, which check their validity. Valid transactions are grouped into a block and stored in the blockchain in a way that tampering with them is nearly impossible, as it would require large computational power. Many peers store the entire history of the blockchain, therefore they are called full nodes. Each attempt of registering non valid transaction in the blockchain or changing its history is detected by those nodes and can be avoided.

[1]Here we give just an high-level description. For the technical details of the Bitcoin blockchain, we refer to [7].

The result is a system enabling secure economic transactions which, by virtue of its P2P nature, is not controlled by any centralized entity (there is no bank issuing the currency) and does not feature single points of failure.

The tamper-resistance property, the validation of data inserted and the P2P nature of the blockchain enlarge the spectrum of applications of this technology, going far beyond cryptocurrencies and financial uses. Indeed, the blockchain can be seen as an immutable, fully-decentralized log of events, since information is validated, stored in a precise order, timestamped and -once registered- is difficult to be tampered.

III. Research scenario

To foster the applicability of our research idea to everyday life, our reference scenario is a community of people owning a certain amount of interoperable IoT devices. An example of community can be the one constituted by people living in the same apartment building. In each apartment there are IoT devices, and some external entities - like the municipality, insurance companies, etc. - are interested in the data produced by those devices.

According to our vision, instead of providing the interested entities with the raw data produced by the IoT devices in the building, a privacy-preserving solution is a local P2P data storage combined to access policies, so that a person can decide the different levels of sharing data: anonymous sharing, sharing of aggregate data, sharing of obfuscated data or sharing of the entire (or a specific portion) raw data are some examples of possible policies.

To locally store the raw data, we can exploit the P2P network consisting of the IoT devices in the building. Each data can be split in more pieces and each piece stored in a different peer. [2] In this way, there is no single point where people's data can be accessed. Therefore, by ensuring that only the data owner can recompose her data spread among the devices of the community, privacy is guaranteed by design. In addition, P2P storage techniques are characterized by high levels of robustness: since some redundancy is added, even if a peer crashes and some pieces of data get lost, it is still possible to recover the original data. Examples of P2P solutions that can fit in our scenario are Tahoe-LAFS [8] or IPFS [9].

The second P2P technology to be helpful in this scenario is the blockchain, described in Section II. In our scenario the blockchain should accomplish two functions: certification of data and incentivization for the peers storing data. Regarding the certification, by registering in the blockchain the hash of data produced by IoT devices, we could be able to detect any abuse on those data. In fact, since the blockchain is tamper-resistant, the data hashes in the blockchain cannot be tampered, and any non-authorized modification of the data implies a mismatch of the hash. As proposed in Storj [10], it is possible to implement periodical audits based on hash challenges. A correct answer to those challenges is a proof the data are actually stored and have not been tampered. Such certification is not only useful for data owners - that in this way are sure their data are still stored in the system: it is also a guarantee for the external entities interested to data, because it assures that data are authentic and have not been tampered. In addition, in the blockchain it is possible to code a storage contract which automatically rewards with some amount of cryptocurrency the peer which correctly answers the hash challenges. This serves as an incentive for peers of the P2P storage network to store data, needed in case peers of the storage network are not known and may cheat the data owners.

Regarding the access policies, we want to give the data owner the possibility to easily specify the rules of access, defining which entities can access which kind of data (aggregate data, anonymous data, etc.). This could be done at the application layer. Public key cryptography can be employed: each entity is identified by a public key and data owners specify which public keys can access data. Then an entity willing to obtain data authenticates itself by its private key and an application checks whether it has the right to access.

IV. Blockchain scalability

As mentioned in Section I, one of the main barriers to enable a decentralized private-by-design IoT supported by the blockchain is scarce scalability of the present blockchain system (see our literature review [5] for more detailed information).

By scarce scalability we mean that the Bitcoin blockchain cannot support high transaction throughput. A parameter coded in the Bitcoin code, that is the maximum block size [11], limits to 7 the number of transactions per second that can be written in the blockchain. The reason of setting a maximum block size is that it limits the cost of running a full node. In fact, with a higher block size limit the throughput would be higher, but at the same time the blockchain would more rapidly increase in size, requiring more disk space. For that reason, there would be less full nodes storing the entire blockchain. A lower number of full nodes, however, implies a more centralized system, since less nodes would have the power to decide which transactions are valid. If these nodes collude together, they could influence the system in their favor.

To avoid the centralization of the blockchain, recent research is focusing on how to scale Bitcoin without changing the block size. One promising approach is the Bitcoin Lightning Network [12], which will be implemented in the next release of Bitcoin [13].

V. Contributions

To be profitably employed in the IoT, the blockchain should support the high throughput of data production which characterizes IoT devices. For this reason, in our research work we will focus on the scalability issue of the blockchain, by performing simulations that provides empirical measures on its scalability degree. We will also consider the possibility of employing blockchains whose scalability is higher than the Bitcoin blockchain.

[2] Here we are assuming that devices are equipped with enough data storage capability. However, we retain that this precondition is not hard to be satisfied, since nowadays the cost and dimension of memory drives are low and will decrease in the following years according to Moore's law.

REFERENCES

[1] International Telecommunication Union, "Measuring the Information Society Report," International Telecommunication Union (ITU), Report, 2015.

[2] G. Greenwald and E. MacAskill, "NSA Prism program taps in to user data of Apple, Google and others." [Online]. Available: http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data

[3] Y. Benkler, "Degrees of Freedom, Dimensions of Power," *Daedalus Winter 2016*, vol. 145, pp. 18 – 32, 2015. [Online]. Available: http://www.mitpressjournals.org/doi/pdfplus/10.1162/DAED_a_00362

[4] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2009. [Online]. Available: http://www.bitcoin.org/bitcoin.pdf

[5] M. Conoscenti, A. Vetrò, and J. C. De Martin, "Blockchain for the Internet of Things: a Systematic Literature Review," in *The Third International Symposium on Internet of Things: Systems, Management and Security (IOTSMS-2016)*, 2016.

[6] ISO/IEC, "ISO/IEC 25010 System and software quality models," Tech. Rep., 2010.

[7] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J. A. Kroll, and E. W. Felten, "SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies." in *IEEE Symposium on Security and Privacy*. IEEE Computer Society, 2015, pp. 104–121.

[8] "Tahoe-LAFS." [Online]. Available: https://tahoe-lafs.org/trac/tahoe-lafs

[9] J. Benet, "IPFS - Content Addressed, Versioned, P2P File System." *CoRR*, vol. abs/1407.3561, 2014. [Online]. Available: https://github.com/ipfs/papers/raw/master/ipfs-cap2pfs/ipfs-p2p-file-system.pdf

[10] S. Wilkinson, "Storj A Peer-to-Peer Cloud Storage Network." [Online]. Available: http://storj.io/storj.pdf

[11] "Bitcoin source code." [Online]. Available: https://github.com/bitcoin/bitcoin/blob/master/src/consensus/consensus.h

[12] J. Poon and T. Dryja, "The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments," 2016. [Online]. Available: https://lightning.network/lightning-network-paper.pdf

[13] "Segregated Witness benefits." [Online]. Available: https://bitcoincore.org/en/2016/01/26/segwit-benefits/