Debunking blockchain myths

Roman Vitenberg*

Department of Informatics, University of Oslo, Norway
romanvi@ifi.uio.no

Abstract

The explosive popularity of the blockchain paradigm has brought upon an avalanche of startups, new strategies and departments at major industrial players, societal and scientific conferences, as well as government-level initiatives in Europe. While computer science has always been trendy and any emerging technology that bursts into popularity has always been accompanied with uncertainty, the situation with blockchain is unique. The uncertainty pertaining to Java, P2P, Web Services, cloud, IoT, and other technologies has been about standardization, competition between alternative solutions, functional extensions, and understanding of technological limitations. In the case of blockchain, however, it starts with the terminology, basic assumptions, and models.

The goal of this short paper is to reflect on the most common misconceptions found in non-scientific (and sometimes, also scientific) articles, and hopefully, contribute towards better understanding of the technology.

1 The rapidly evolving blockchain field

Since Bitcoin [12] inception and deployment in 2009, it has been creating waves of discussions and debates about its success. In the wake of Bitcoin came Ethereum [3] that aimed to facilitate programmatic mediation between cooperating parties by defining a feature-rich language for writing mediation software. When the blockchain wave hit the world, many dozens of competing industrial initiatives and proposals for blockchain software were put forward. A number of these proposals resulted in open source implementations, such as the Hyperledger [7] and Corda [2] projects promoted by IBM and the R3 consortium, respectively.

2 The architecture of Bitcoin

Bitcoin layered architecture is illustrated in Figure 1. The communication layer implements a specialized broadcast primitive in the Bitcoin P2P network. The data storage layer records all transactions in a distributed ledger implemented as a chain of blocks. The consensus layer realizes a consensus protocol through proof-of-work mining. The business logic layer implements a simple scripting language used for verifying transactions and redeeming their output. The discussion of blockchain myths in Section 3 mentions the requirements and design elements of the data storage, consensus, and business logic layers in Bitcoin and other blockchain systems.

3 Blockchain myths

Myth 1: A blockchain is a replacement for a database. A blockchain differs from a database in two fundamental ways. First, a database is an organized collection of data representing the current system state. The main functionality of a database is to allow efficient data

*The author presented this paper at the NISK 2018 conference.

Proceedings of the 11th Norwegian Information Security Conference, Sep. 19-20, 2018.
NISK Journal, e-ISSN 1894-7735. https://ojs.bibsys.no/index.php/NISK

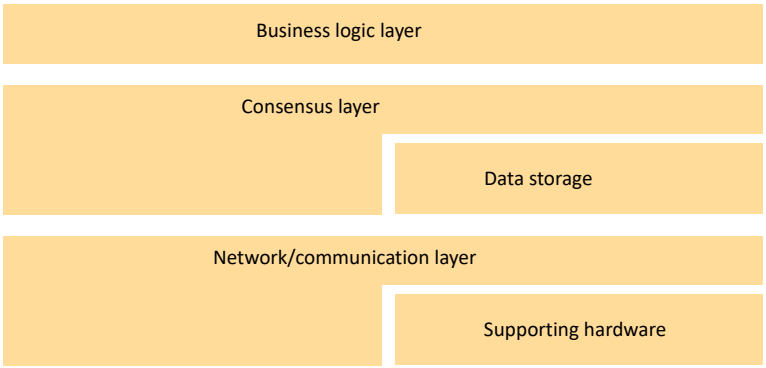

Figure 1: Layered architecture of Bitcoin

retrieval, fusion, and aggregation triggered by user queries. In contrast, most blockchain implementations represent a ledger in which a history of transactions (or, more generally, of changes to the system state) is recorded. For example, there is simply no concept of a user balance in Bitcoin! While Ethereum keeps track of a contract state, it only provides limited means to retrieve and process state data, as explicitly defined by the contract. It does not support abstractions of a flexible query language, data view, schema, join, etc. Besides, Ethereum records a history of all changes to the state, which results in blockchain space being more expensive and the storage less efficient compared to a database. As a result, only specific data elements (such as short transactions or indices) are stored on a blockchain. Most blockchain systems in application domains other than financial combine blockchain with offchain storage (databases or dedicated file systems). There are many implementations that support such a combination, e.g., StorJ [9], Filecoin [4], BigchainDB [1], etc.

Secondly, the trust model is radically different as observed in [8]. The database servers typically trust each other, even in federated databases, in the sense that they do not expect attacks from within the system. The main security focus is on making it difficult to compromise a server in the first place. To this end, database systems defend against malicious clients by using firewalls, strict access control, and many other methods. The situation is fundamentally different in the blockchain environment: the interests of participating nodes are inherently misaligned so that they need to verify information received from each other and run a consensus to agree on changes to the data. While being able to agree on changes and progress in absence of a trusted administrator is a powerful abstraction, it bears a cost tag in terms of performance. If there are no misaligned interests between the participants and attacks from within the system are unlikely, there is little point in using blockchain technologies.

Myth 2: Blockchain has a definition that is universally agreed upon. Distributed Ledger Technologies (DLTs, in short) is a well-defined term: it refers to a system that records a ledger of transactions or a history of changes to the system state. The ledger is usually hard to temper with, which is a boon for security yet it also makes it hard to prune the history or compact the ledger.

In many contexts and for many purposes, people equate blockchain with DLTs. Note, however, that both narrower and broader meanings of "blockchain" are in use. Literally, blockchain means "a chain of blocks", which implies a specific data structure for the ledger implementation. A chain of blocks precludes any parallelism between the transactions, however, which has a negative impact on the performance. Some ledger implementations use more a complex

data structure such as braids [11], which allow some degree of parallelism by retaining concurrently proposed competing blocks and merging them. Since the term of DLT does not imply any specific data structure, it covers such a generalization. On the other hand, the term of "blockchain" becomes a misnomer in that case, a distinction that many people are unaware of. In absence of more refined terminology today, "blockchain" is used in the literature to refer to a chain of blocks or generalized DLTs.

To add to the confusion, some systems in this domain do not maintain a distributed ledger at all. For example, Corda [2] allows participating nodes to agree upon and maintain shared knowledge in a non-trusted environment typical for blockchain. However, each piece of information is only shared within a subset of nodes to which the information pertains. Yet, the term of "blockchain" is sometimes used to collectively refer to all systems in the domain including Corda.

Myth 3: All blockchain technologies are similar to Bitcoin. While most blockchain initiatives use a layered architecture similar to that of Bitcoin, with each layer serving a similar purpose, there are also significant differences in application requirements on the one hand and operational conditions and attack models on the other. In some cases, the functionality is significantly extended or even replaced. For example, the business layer implementation in Bitcoin is rather rudimentary compared to Ethereum and other later systems. The implementation of layers such as consensus in Hyperledger is radically different compared to Bitcoin. The hypothesis is that this diversity will continue to increase over time. Instead of a single homogeneous blockchain area, multiple technological subareas will be identified as the technology matures.

Perhaps the most basic distinction is that while Bitcoin is open and self-organizing (there are no validated identities, anyone can read the ledger and propose new transactions to include), other Blockchain implementation such as Hyperledger are managed, with authenticated identities, membership, and access control. Such implementations are commonly referred to as private blockchains.

Myth 4: All blockchain technologies use mining and consume a lot of energy. All blockchain technologies need to solve a consensus problem. The proof-of-work mining mechanism is Bitcoin solution for the consensus problem. It is notoriously known for consuming exorbitant amounts of energy [6]. Significantly more efficient mechanisms such as proof-of-capacity or proof-of-stake have been proposed. For example, proof-of-stake does not require any significant computation; the mining process is commonly referred to as "virtual" in this situation. However, consensus solutions in smaller-scale systems may not require mining at all. In fact, first consensus solutions appeared in the eighties, long before their applications in the area of cryptocurrency. The PBFT protocol [10] designed in 1999 has been adapted for blockchain needs and adopted as one of the principal consensus solutions in Hyperledger.

Myth 5: All blockchain systems use a P2P network It is true that most blockchain systems in existence use a P2P network. For self-organizing systems with many thousands of participating nodes, such as Bitcoin, that might be the most practical way to build a communication between the nodes and scale. For managed, private, member-only systems like Hyperledger or Corda, however, there might be little need to utilize a P2P network. They use it by-product rather than by-design: their architects took the Bitcoin design as a basis and then focused on the innovation related to the business logic, consensus, and storage solution design. Since the communication is not perceived as a bottleneck in blockchain, it does not receive a lot of attention in these rapidly developing systems. In a sense, it is possible to always use a P2P network inside any data center or across a small number of data centers, even though it

might be an unnecessary element of the design that leads to a small negative impact on the performance. It is envisioned, however, that in the future, some of the emerging blockchain systems will do away with P2P networks.

Myth 6: Private blockchains are necessarily lacking transparency. Perhaps most of the misunderstandings in non-technical literature are related to the concept of private blockchains. A lot of banks and other companies deploy blockchain solutions without disclosing any technical details, giving the word "private" a negative connotation. However, as explained above, private blockchain refers to a completely different concept; it implies managed rather than self-organizing deployment. It is called private because there is a members-only club. Perhaps "managed" blockchain would be a better term because the code can be transparent and even open source, as it is the case for Hyperledger. In fact, the ledger may also be publicly available for querying; only the right to propose modifications may be restricted to validated identities or members. This may be a suitable model for some blockchain applications such as a blockchain-based voting system.

While it can be argued that protected data query access diminishes transparency compared to Bitcoin, this is actually a desired property in some application domains. Imagine a blockchain system storing sensitive healthcare data which exposes the data to all and allows everyone to store a copy. This would be unacceptable even if the data are encrypted, not to mention incompliant with GDPR.

Myth 7: Private blockchain are intended for the private market. Another common misconception is that private blockchain solutions are only deployed by private companies and enterprises. Arguably, most blockchain deployments will be managed and will require validated identities. The managing entities might well belong to the public sector: universities, public healthcare clinics, and even governmental organizations as indicated by the publicized X-Road deployment in Estonia [5].

References

[1] BigchainDB – The blockchain database. https://www.bigchaindb.com/.

[2] The corda project. https://www.corda.net/.

[3] The ethereum project. https://www.ethereum.org/.

[4] Filecoin – A blockchain-based storage network and cryptocurrency. https://filecoin.io/.

[5] Forbes: The tiny european country that became a global leader in digital government. https://e-estonia.com/tag/blockchain/.

[6] How much power does the bitcoin network use? https://www.thebalance.com/how-much-power-does-the-bitcoin-network-use-391280.

[7] The hyperledger project. https://www.hyperledger.org/.

[8] On distributed databases and distributed ledgers. https://www.corda.net/2016/11/distributed-databases-distributed-ledgers/.

[9] StorJ – Decentralized Cloud Storage. https://storj.io/.

[10] M. Castro and B. Liskov. Practical byzantine fault tolerance and proactive recovery. *ACM Trans. Comput. Syst.*, 20(4):398–461, Nov. 2002.

[11] B. McElrath. Braiding the blockchain. https://scalingbitcoin.org/hongkong2015/presentations/DAY2/2_breaking_the_chain_1_mcelrath.pdf.

[12] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system, Oct. 2008. http://www.bitcoin.org/bitcoin.pdf.