SoK: Blockchain Technology and Its Applications

Anonymous

***Abstract*—Abstract goes here.**

***Keywords*-keyword;**

Possible reader goals. Jeremy's suggestions:

1) ???
2) ???
3) I no longer have to read the 100+ industry whitepapers, as they are summarized here.

Scott's suggestions:

1) What is Blockchain and what is it not? What are its key properties? What are not properties of Blockchain?
2) What are possible use case areas for Blockchain research. How do I know if my research area might benefit from Blockchain?

I. Introduction (Rob/Arkady/Jeremy)

Short history of Bitcoin (1–2 paragraphs) Summary of the research ideas that are part of Blockchain (Jeremy's article) Bitcoin itself

Blockchain technology Evolved from bitcoin What can we do besides cryptocurrency?

Why industry? Academia was slow to appreciate Bitcoin and later Blockchain technology Industry saw potential and began running with it Most of the interesting work is happening outside of academia Academia is more focused on important, but more ticky-tacky details Industry has built up knowledge around Blockchain, its properties, and its uses

Open question: how do we integrate what we've learned from academia into the intro and elsewhere in the paper.

II. Methodology (Scott)

Motivation Separate out the data from the hype Collective knowledge of the masses Justify grounded theory.

A. Source Selection

B. Data Analysis

After collecting our initial set of **UPDATE # XXX** papers, we analyzed them using a four-stage grounded theory approach (open coding, axial coding, selective coding, and theory generation). Throughout the analysis of the documents we kept detailed research notes that outlined our thoughts as we reviewed and analyzed the literature. Additionally, we conducted intensive discussion between the various researchers to ensure that we were correctly understanding and evaluating the source material. As is often the case in grounded theory, these notes and discussion were every bit as important, if not more so, than the concepts, categories, and theories we generated.

1) Step 1—Open Coding: In this first step, documents were assigned to one of four reviewers. Each reviewer would read the document, assign codes to words and sentences in the document. These codes were generated using a mixture of open coding (assigning a code that summarizes the document's statement) and in situ coding (using the document's own words as the code). To ensure that we were assigning the correct codes, we paid careful attention to the context of each statement.

In particular, reviewers made sure to code the following four concepts found in documents:

1) **Properties.** What are the building blocks for Blockchain technology? What capabilities does it provide?
2) **Challenges.** What challenges must be addressed when building systems using Blockchain technology?
3) **Limitations.** What inherent limitations are their when using Blockchain technology?
4) **Use cases.** What use cases are suitable for Blockchain technology?

At this stage of the grounded theory process, reviewers were instructed to avoid evaluating the validity of the coded concepts. Instead, every attempt was to made to include all possible codes, helping to ensure that our results were grounded in the data and not reviewers' biases.

The reviewers continued reviewing documents until each felt that the last 3–5 documents they had read had no concepts that had not already been brought up by previous documents. This is a commonly accepted stopping criteria in grounded theory and is indicative that all core (i.e., not truly one-off) ideas have been discovered. In total, **UPDATE # XXX** documents were coded in this stage.

2) Step 2—Axial Coding: In the second stage, our research team used the constant comparative method to group codes into concepts. Specifically, we collapsed distinct codes referring to the same topic (e.g., one was an open code, the other in situ) into a single code, reducing the original set of **UPDATE # XXX** codes to a more manageable **UPDATE # XXX** codes. As needed, we referred back to the original documents to ensure that our understanding of the code was fresh, and that we were assigning it to the appropriate concept. Also, at this stage we continued to avoid evaluating the validity of concepts, ensuring that the ideas of the reviewed documents were fully reflected in the codes.

C. Interlude—Additional Open Coding

After completing axial coding, one reviewer coded (i.e., open coding) another **UPDATE # XXX** documents. These documents were all blog posts, representing the most up-to-date thinking on Blockchain technology. In this process, no new codes were discovered, indicating that our process had produced concepts that thoroughly describe Blockchain technology.

1) Step 3—Selective Coding: In the third stage, two researchers transfered all of the concepts onto sticky-notes. They then drew connecting lines between the concepts, describing how the concepts related to one another. Based on these interconnections, concepts were divided into five different categories:

1) **Primitives.** These are the primitives that are used to enabled Blockchain technology. Unlike properties, they have no useful by themselves, but only when combined with other primitives to achieve specific properties. Examples include on-chain tokens, authenticated data-structures, and zero-knowledge proofs.
2) **Technological properties.** Technological properties are the core building blocks and features of Blockchain technology. Examples include decentralized governance, append-only ledgers, and data replication.
3) **Normative properties.** Normative properties differ from technological properties in that they are not technical, but rather represent properties that people hope to achieve through the use of Blockchain technology. Critically, these properties cannot be achieved through reliance on the technical properties alone, but require the system built on top of Blockchain technology be designed to accomplish these goals. Examples include censorship resistance, ease-of-entry for miners, low cost to participate.

   TODO: Does this belong somewhere else? Seems to be too much commentary for here." In general, normative properties roughly correlate to the hype attached to Blockchain technology. Based on our analysis, we believe that much of the confusion surrounding Blockchain technology arises from a failure to separate technological and normative properties. Conflating this two properties makes it difficult to see the technology benefits of Blockchain technology, as they are overshadowed by normative properties that are difficult to achieve in real-world applications.
4) **Capabilities.** Capabilities are high-level features provided by Blockchain technology. Examples include on-chain asset provenance, anonymity, and resilience.
5) **Use cases.** Use cases are the areas where the properties and capabilities of Blockchain technology would be useful in building systems. Examples include cryptocurrencies, supply chain management, and identity management.

While we divide the concepts into five categories, we note that there were inter-category connections, indicating that concepts frequently relied on or were relied on by other concepts.

At this stage of the research, we allowed researcher expertise to begin influencing the results. First, by its very nature drawing connections between concepts is subjective. As much as possible, we attempted to identify text in the underlying documents that supported our connections, but in several cases we created connections that were not explicitly mentioned in the text. Second, a handful of primitives were added that we determined were necessary to build some of the technological properties, but had not been discussed in the documents. Third, we identified several misconceptions that either shared no connections with the rest of the concepts or were obviously false (e.g., cryptographic signatures do not provide confidentiality). In each of these three cases, our research notes kept track of what was explicitly supported by the analyzed data and what was the result of researcher interpretation.

2) Step 4—Theory Generation: In the fourth and final stage, we used the categories, their connections, and our results to derive several theories (i.e., research results from our analysis) regarding Blockchain. First, we derived a concrete set of properties and capabilities describing precisely what Blockchain technology is, and what it isn't. Second, we were able to produce a clean split between Blokchain technology's technological primitives and its normative properties (i.e., hype). Third, we identified criteria that help determine whether a given problem can benefit from the use of Blockchain technology.

D. Limitations

Due to the nature of grounded theory, our analysis of the data represents one view on that data. Different researchers coding the same data may have focused on different aspects leading to differences in categories, connections, and the theories they focused on. To address this limitation, we will make the documents we reviewed and our coding of those documents public.

III. Results (Ben)

Types of items Primitives Properties Technological Normative Capabilities Provenance Off-chain physical asset provenance Off-chain digital asset provenance On-chain asset provenance Access control for tokens Data discoverability Transaction rules/smart contracts (Most applications rely on this) Auditability Internal auditability Global auditability

Governance is central Anonymity is orthogonal Game theory is core to decentralized governance

Resilience Broadly supported Easy to overlook parts of this property when not taken as a whole

Normative and technical properties are cleanly separable Ledger=/data store

Blockchain is not a global platform by definition
Challenges Limitations

IV. Lessons Learned (Scott)

What was missing in the graph? Off-chain stapling Anonymity MPC Functional encryption Authenticated data structure

Terminology

Normative vs. technical properties public participation is a huge normative property. Not gauntleted to be in a Blockchain. There are risks of thinking they are the same Using blockchain assuming you get normative properties, not just technical properties Get saturated on normative properties, ignore the technical properties Diffused trust vs. trustfulness Decentralization Governance and communication are required to be decentralized Disintermediation is normative, and may or may not be part of a Blockchain There are always intermediaries. Decentralization limits their power.

How do we know which use cases are good fits for Blockchain? Criteria 1) Decentralized governance (*) 2) Auditability 3) Resiliency Is there a question tree we can create?

V. Use Case Evaluation (Jeramey)

Open question: should we also discuss applications that are discussed in the literature, but turn out to not be great fits? If so, where does this fit. This section and the last are really more centered around us imposing knowledge.

List application areas Why are they good Interesting research questions

VI. Discussion (All)

Fast/cheap is because of deregulation Do things that are legally questionable

Degrees of decentralization Single - not really a blockchain, or is it? Oligarchy Embarrassingly decentralized

There are risks of thinking they are the same Using blockchain assuming you get normative properties, not just technical properties Get saturated on normative properties, ignore the technical properties

Blockchain is not only for global system

Lighter options when all of blockchain not needed Distributed data sources with auditability/replication

VII. Conclusion

Blockchain is good when used right. Let's use it correctly!

[1]

References

[1] I. Freely. A small paper. *The journal of small papers*, -1, 1997. to appear.