NORTHWESTERN UNIVERSITY

Unchaining the Blockchain Network Layer

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Computer Science

By

Uri Klarman

EVANSTON, ILLINOIS

March 2019

ABSTRACT

Unchaining the Blockchain Network Layer

Uri Klarman

Blockchains are an exciting new type of Peer-to-Peer (P2P) distributed systems, which enable parties to transact directly, and maintain the record of said interactions in a distributed manner. A unique feature of blockchains is their ability to maintain a consensus without requiring knowledge on the number of participants, nor their identities, opening the door for cross-border, self-organized, decentralized, open ecosystems. The original blockchain, Bitcoin, aims to be a Global P2P Electronic Cash System, and to replace fiat money, banks, payment-processors, and other financial middlemen. Other blockchains, *e.g.*, Ethereum, aim to remove middlemen from other types of interactions, such as replacing custodians with Smart Contracts, and to securely store credentials, identities, health records, and private information.

While blockchains have the potential to be transformative in many fields, their real-world usage is held back by practical limitations. First and foremost, for a blockchain to be used at a global scale, it must be capable of handling a high volume of transactions; for Bitcoin to replace Visa, MasterCard, PayPal and other payment processors, it must be

capable of processing roughly $5,000$ transactions per second (TPS), the average number of TPS these companies process today [**26, 74**]. To support online shopping, it must support peak demand, which for Alibaba stands at $325,000$ TPS [**109**]. In contrast to these significant requirements, Bitcoin can only process 3–4 TPS. Other significant limitations include the centralization and wastefulness of blockchain mining, the procedure which records transactions in the blockchain.

Research in the blockchain field had thus far focused on new cryptographic primitives and alternative blockchain protocols to address these real-world challenges, while the networking aspects were largely ignored. In this thesis, I propose a definition for the blockchain Network Layer, and provide evidence that the Network Layer is the bottleneck and root-cause for some of the most pressing challenges blockchains face today. I further propose new networking primitives and novel network utilization methods, and explore how they can be used to overcome said challenges, including scalability, mining centralization, and mining wastefulness, as well as to utilize blockchains to decentralize existing knowledge silos. First, I provide the necessary background to understand the operation of blockchains, and a definition for the blockchain Network Layer. Then, I present an analysis which outlines how the Network Layer is the bottleneck for blockchain scalability, and suggest a new networking primitive, the Blockchain Distribution Network (BDN) to overcome both blockchain scalability and mining centralization. Lastly, I present novel networking methods which enable blockchains to be utilized in new fields, focusing on the decentralization of the search engines market, and mitigate mining wastefulness.

Thesis Committee

Aleksandar Kuzmanovic, Northwestern University, Committee Chair

Fabián Bustamante, Northwestern University, Committee Member

Peter A. Dinda, Northwestern University, Committee Member

Cristina Nita-Rotaru, Northeastern University, Committee Member

Acknowledgements

First, I would like to thank my advisor, mentor, and partner, Aleksandar Kuzmanovic, for his continuous support of my research, and for his wise guidance. His focus on exciting ideas, his confidence in our ability to overcome any obstacle we encounter, and his genuine enthusiasm about our joint work, are the cornerstones for my approach towards science, research, and academia.

I would further like to express my gratitude to my co-advisor, Fabián Bustamante, for supporting my research, for sharing his insights on conducting research across multiple domains and its implications, and for his valuable feedback and advice. My sincere thanks also go to the other members of my committee, Peter Dinda and Cristina Nita-Rotaru, for their valuable insight and candid feedback. Their insights and comments during my proposal had a significant effect on the shaping of my thesis, for which I am grateful.

I thank my lab-mates and fellow students at the Northwestern Networking Group: Marcel Flores, Ning Xia, Marc Warrior, Qurat-Ul-Ann Akbar, Andrew Kahn, and Alexander Wenzel, for their camaraderie, technical assistance, and wise advice. It had been my pleasure to work alongside each and every one of you. Additional thanks go to the entire bloXroute Labs team, and to Soumya Basu in particular, for investing countless hours in the development and deployment of bloXroute, bridging the gap between clever ideas and solid reality.

I would like to thank my parents, brothers, in-laws, nephews and nieces, for continuously supporting me in this endeavor, and throughout my life. Finally, I would like to thank my wife Yaara, without whom I would have never set out on this exciting academic journey, nor would it had been half as exciting.

Table of Contents

List of Figures

and indexing, (5) block and index propagation completion, (6) Web crawling completion, (7) block mining begins. 101

CHAPTER 1

Introduction

The Bitcoin white paper [**75**], published in October 2008, describes a new type of Peer-to-Peer (P2P) distributed system – the blockchain. In a blockchain ecosystem, participants may transact among themselves, and their transactions are periodically batched into *blocks* which are recorded by all peers. This results in chain of blocks, the blockchain, which defines a consensus over which transactions took place. To add a new block of transactions to the blockchain, the Bitcoin protocol requires a Proof-of-Work (PoW), *i.e.*, a proof of performing a difficult computational-intensive task. This requirement prevents sybil attacks from influencing the consensus, and alleviate the need for peers to have knowledge of the number of participants or their identities. In the decade since Bitcoin's inception, many blockchains were launched in an attempt to improve upon, repurpose, substitute, or complement Bitcoin. While some blockchains are almost identical to Bitcoin, others had experimented with wide range of changes to the protocol, including altering its PoW, replacing its PoW with Proof-of-Stake (PoS) and other alternatives, altering the time interval between blocks and the number of transactions each can contain, running complex scripts, obfuscating transactions, and others.

Despite Bitcoin becoming a household name, and despite a market cap of hundreds of billions of dollars for blockchain-based cryptocurrencies, blockchain usage remains negligible when compared to major payment processors such as MasterCard, Visa, and PayPal,

or to online shopping giants such as Alibaba and Amazon. While the credit card companies and PayPal combined handle roughly $5,000$ transactions per second (TPS) on average, and Alibaba had processed over $325,000$ TPS at peak demand, Bitcoin can only process 3–4 TPS. This limitation, which is common to all blockchains, is often referred to as the "scalability problem", and is considered amongst the most pressing and difficult challenges blockchains face. Other significant challenges include mining centralization, i.e., the coalescence of miners into mining pools which undermine blockchains' security, blockchain utilization, and the wasteful nature of PoW mining.

1.1. The Blockchain Network Layer

Possible solutions for the scalability problem and other pressing challenges had been the source of heated discussions among Bitcoin and blockchain enthusiasts, researchers, and developers. However, due to to cyber-security and cryptography origins of this community, these discussions had mostly focused on novel cryptographic primitives, new protocols, and alternative consensus schemes. The networking aspects of blockchains had largely been ignored, and often considered to be a blackbox infrastructure which is agnostic to the operation of blockchains. Indeed, it is not uncommon to see all networking aspects abstracted away in these discussions, marked only as edges in graph representations of blockchain systems.

In this dissertation I propose a definition for the blockchain Network Layer, and the functionalities which fall under it. I further show that the Network Layer is the root-cause of the scalability problem, and in contrast to common belief, it is the Network

Layer which holds the key for some of the most pressing challenges blockchains face, including scalability and mining centralization.

1.2. Blockchain Networking Resources

Most of the networking resources consumed by Blockchain systems are utilized at the blockchain Network Layer. However, it is possible for blockchains to utilize their networking resources in other fashions to achieve additional goals. In this dissertation, I present how novel network utilization techniques enable blockchains to consume networking resources, rather than computational resources, which can be leveraged to decentralize critical networking tasks and to mitigate mining wastefulness, focusing on Web search engines.

1.3. Thesis Organization

The remainder of this dissertation is structured as follows: First, in Chapter 2 I present the thesis statement, which captures the overarching scheme of the individual chapters incorporated in this dissertation. Then, in Chapter 3, I provide the necessary background to understand the operation of blockchains and some of the the major challenges they face, and provide a definition for the blockchain Network Layer. In Chapter 4, I present an analysis of the scalability problem, and demonstrate that its root-cause is a bottleneck at the blockchain Network Layer. I further outline a new networking primitive, the Blockchain Distribution Network (BDN), which enables blockchains to overcome both the scalability problem and blockchain mining centralization, and present empirical results from the deployment of bloXroute, the first Blockchain Distribution Network. Next, in Chapter 5 I present a novel networking utilization method to enable the application of

blockchains in new fields, focusing on the decentralization of the search engine industry, and to mitigate mining wastefulness. Finally, in Chapter 6 I summarize my contributions and conclude.

CHAPTER 2

Thesis

The blockchain Network Layer and the networking resources utilized by blockchains hold the key to some of the most pressing challenges blockchains face, including scalability, mining centralization, utility, and wastefulness.

CHAPTER 3

Blockchain Fundamentals and the Blockchain Network Layer

In this chapter I provide an overview of the operation of Bitcoin to demonstrate the fundamental principals of blockchains, and the necessary background to understand the importance of the blockchain Network layer. I then continue to outline the blockchain stack, and provide a definition for the blockchain Network Layer and the functionalities it includes.

3.1. The Operation of Bitcoin

3.1.1. Bitcoin Users

Bitcoin [**75, 10**] is the first blockchain system, and the first cryptocurrency to gain significant popularity and to become a household name. While some cryptocurrencies predate Bitcoin [**9, 29**], none had gained any significant traction outside the cypherpunk community [**50**]. At its core, Bitcoin is a Peer-to-Peer (P2P) Electronic Cash System which allows its users to hold a balance of Bitcoins, and to transact Bitcoins among themselves. Said transactions are batched into *blocks*, which are recorded by all peers, and the resulting chain of blocks, *i.e.*, the *blockchain*, defines a consensus over which transactions took place. While the Bitcoin white paper itself describes the blockchain as a distributed timestamp server, it is oftentimes referred to as a distributed ledger, as its main function is to record financial transactions and balances.

To understand how Bitcoin transactions are created and the blockchain maintained, assume a user, Alice, wishes to buy an item from a different user, Bob, and to pay in Bitcoin. Each Bitcoin user controls a *wallet*, which is a simple private and public key pair. The public key (or *addresses* derived from the public key) are used to maintain user balances, while the private key is used to prove ownership over said balances. To pay Bob, Alice locally creates a new transaction $t_{A \to B}$, which specifies an amount of Bitcoins which is to be passed from her address to Bob's address, and then signs it using her private key. Alice then sends $t_{A \to B}$ to the Bitcoin nodes she is connected to, to be propagated to the entire Bitcoin P2P network.

It is worth noting that Alice and Bob are Bitcoin *users*, and therefore in control of their Bitcoin wallets, and that any Bitcoin user may control any number of wallets, each of which controlling any number of public keys and addresses. However, Alice and Bob may or may not be running a Bitcoin *node*.

3.1.2. Nodes, Miners and Validators

The Bitcoin white paper refers to participants in the P2P network which maintain a local copy of the blockchain and attempt to produce new blocks as "nodes". However, later work had led to a more fine-grained, yet somewhat ambiguous, terminology to describe the different kinds of blockchain participants. First, a comparison made in the Bitcoin white paper between the creation of a new block and the mining of gold had led to the use of the term *mining* to describe the creation of new blocks using Proof-of-Work (PoW), and the term *miner* to describe a node engaged in mining. Second, some [**4**] consider the term "node" to also describe participants which maintain a local copy of the blockchain *without*

attempting to produce new blocks. These participants are often referred to as *non-mining nodes*. Lastly, some blockchains had experimented with the use of Proof-of-Stake (PoS) and other mechanisms to replace Bitcoin's use of PoW, where the producers of new blocks are referred to as *validators* [**114, 113**] or block-producers (*BPs*) [**24**]. Throughout this thesis I use the term "miner" to describe participants which attempt to produce new blocks by any means, and the term "non-mining node" to describe participants which do not attempt to produce new blocks. Lastly, I use the general term "nodes" to describe participants of all kinds, including both miners and non-mining nodes.

Every Bitcoin node to receive $t_{A \to B}$ would validate that:

(1) The balance of Alice's addresses specified in $t_{A \to B}$ is greater or equal to the amount spent.

(2) $t_{A \to B}$ contains a signature which can only be created with knowledge of Alice's private key.

If these two conditions are met, $t_{A \to B}$ is deemed valid, and any node to receive it will propagate it to its peers. While all Bitcoin nodes would receive and propagate transactions, only miners attempt to aggregate the transactions they receive into a new block, to be appended to the blockchain. It is only after transactions are included in the blockchain that they are considered to have taken place, while transactions which still await to be included are not.

There are two monetary incentives for miners to engage in mining. First, each block contains a unique transaction, called the *coinbase transaction*, which passes some amount of Bitcoins to its miner's address. Thus, miners do not attempt to produce a new block from the kindness of their hearts. Rather, they compete, as each miner attempts to