

**Roman Matzutt**

**Demystifying and Adjusting the Promises  
of Blockchain-based Data Management  
in the Permissionless Setting**

---

# Demystifying and Adjusting the Promises of Blockchain-based Data Management in the Permissionless Setting

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften  
der RWTH Aachen University zur Erlangung des akademischen Grades  
eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Master of Science

**Roman Matzutt**

aus Heinsberg

Berichter:

Prof. Dr.-Ing. Klaus Wehrle  
Prof. Dr. rer.nat. Frank Kargl

Tag der mündlichen Prüfung: 20.10.2023

---



# **Reports on Communications and Distributed Systems**

edited by  
Prof. Dr.-Ing. Klaus Wehrle  
Communication and Distributed Systems,  
RWTH Aachen University

Volume 24

**Roman Matzutt**

## **Demystifying and Adjusting the Promises of Blockchain-based Data Management in the Permissionless Setting**

Shaker Verlag  
Düren 2024

**Bibliographic information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Zugl.: D 82 (Diss. RWTH Aachen University, 2023)

Copyright Shaker Verlag 2024

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-9516-6

ISSN 2191-0863

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren

Phone: 0049/2421/99011-0 • Telefax: 0049/2421/99011-9

Internet: [www.shaker.de](http://www.shaker.de) • e-mail: [info@shaker.de](mailto:info@shaker.de)

## Abstract

---

The digital currency Bitcoin introduced the blockchain as a data structure that allows its users to establish consensus about who owns which coins in a decentralized manner. Since then, blockchain technology has evolved and now enables distrusting parties to engage in online interactions without the need for a trusted intermediary by immutably recording general events in transactions. This interaction model sparked a tremendous interest in blockchain technology, its potential, and applications.

However, the identification of multiple shortcomings has since dampened this initial spirit of optimism. These shortcomings are especially apparent for permissionless blockchains, such as Bitcoin, which openly encourage participation by anybody. For instance, Bitcoin has to secure its blockchain against malicious actors by relying on energy-intensive computations, which further leads to scalability issues as only few payments can be accepted at a time. While prior work has extensively studied such technical challenges, it neglected the influence of the data stored on the blockchain so far. Yet, this influence becomes undeniable: On the one hand, unknown actors can irrevocably append new data without a designated removal process. On the other hand, the operation of a blockchain system depends on a massive replication of its full and growing history. Hence, the impact of blockchain-recorded data requires thorough investigation to ensure the security and longevity of blockchain systems.

In this dissertation, we thus take a data-driven perspective to assess and improve the applicability of permissionless blockchains as building blocks for decentralized data management systems. We identify two core challenges of blockchain-based data management, i.e., the need for moderating what data is recorded and the need for alleviating the continually growing storage requirements stemming from the append-only nature of blockchains. Furthermore, we assess the potential of blockchains to enable additional applications by seizing their characteristic properties. We address these challenges on a technical level via the following contributions.

As our first contribution, we systematically analyze the phenomenon of blockchain content insertion on a conceptual, technical, and empirical level. Our analysis reveals that content insertion is a common practice and offers benefits for higher-level applications, but inserting illicit content can potentially create devastating consequences for the participants. As our second contribution, we explore means to mitigate these consequences, both before and after the fact, by proposing strategies to prevent the insertion of unwanted content as well as a redactable blockchain that enables a swift and transparent removal of content. Our third contribution addresses the challenge of growing blockchain sizes by defining a gradually deployable block-pruning scheme that is retrofittable to Bitcoin and enables users to retroactively forget obsolete data and thereby reduce their storage requirements. Finally, our fourth contribution shows that permissionless blockchains still hold untapped potential for fueling novel applications despite their limitations; namely, we demonstrate how Bitcoin can help securely bootstrapping decentralized anonymity services.

Overall, we shed new light on the potential impact of the data persisted on blockchains. Our analyses and technical contributions therefore widen the scope for resilient and durable blockchain designs for data management tasks.

## Kurzfassung

---

Die digitale Wahrung Bitcoin hat die Blockchain als eine Datenstruktur etabliert, mit der Nutzer dezentralisiert Konsens daruber erlangen konnen, wem welche Munzen gehoren. Seitdem hat sich die Blockchaintechnologie stetig weiterentwickelt, so dass nun sich misstrauende Parteien ohne Intermediar uber das Internet interagieren konnen, indem beliebige Ereignisse unwiderruflich aufgezeichnet werden. Dies hat ein enormes Interesse an der Technologie, ihr Potenzial und ihre Anwendungen entfacht.

Allerdings hat die Entdeckung einiger Nachteile diese Aufbruchstimmung derweil gedampft. Diese Nachteile kommen insbesondere bei den frei zuganglichen Permissionless Blockchains, wie Bitcoin, zum Tragen. Beispielsweise muss Bitcoin die Blockchain uber energieintensive Berechnungen gegen Angreifer absichern, was zu Skalierbarkeitsproblemen fuhrt, da so nur wenige Zahlungen auf einmal akzeptiert werden konnen. Obwohl solche technischen Herausforderungen bereits intensiv studiert wurden, wurde der Einfluss der Daten in der Blockchain bisher vernachlassigt. Dabei ist dieser Einfluss unbestreitbar: Einerseits konnen Unbekannte Daten unwiderruflich und ohne vorgesehenen Loschprozess anhangen. Andererseits bedingt der Betrieb einer Blockchain eine massive Replizierung ihrer gesamten und stetig wachsenden Historie. Daher bedurfen die Auswirkungen der Blockchain-Daten einer sorgfaltigen Analyse, um die Sicherheit und Langlebigkeit dieser Systeme sicherzustellen.

In dieser Dissertation fokussieren wir uns auf ebendiese Daten, um die Eignung von Permissionless Blockchains als Bausteine fur dezentralisierte Datenmanagement-Systeme zu beurteilen und zu verbessern. Wir identifizieren zwei Kernherausforderungen des Blockchain-basierten Datenmanagements, namlich die Moderierbarkeit der Daten und den Bedarf, die wachsenden Speicheranforderungen aufgrund stetig angehangter Daten abzumildern. Zudem bemessen wir das Potenzial der Blockchain, mittels ihrer spezifischen Eigenschaften weitere Anwendungen zu ermoglichen. Diese Herausforderungen adressieren wir auf technischer Ebene mittels folgender Beitrage.

Als ersten Beitrag analysieren wir das Phanomen des Einfugens von Blockchaininhalten auf konzeptioneller, technischer und empirischer Ebene. Unsere Analyse zeigt, dass dieses Vorgehen gebrauchlich ist und Vorteile fur Anwendungen bietet, jedoch hat das Speichern illegaler Inhalte potenziell verheerende Konsequenzen. Als zweiten Beitrag untersuchen wir Mittel, diese Konsequenzen sowohl im Vorfeld als auch im Nachgang einzudammen, indem wir Strategien, die das Einfugen unerwunschter Inhalte verhindern, und eine editierbare Blockchain, die eine rasche und transparente Loschung erlaubt, vorschlagen. Unser dritter Beitrag adressiert das Problem wachsender Blockchaingroen, indem ein graduell ausrollbares Block-Pruning-Verfahren definiert wird, das Bitcoin-Nutzern nachtraglich ermoglicht, obsoletere Daten zu vergessen und so ihren Speicherbedarf zu reduzieren. Zuletzt zeigt unser vierter Beitrag, dass Permissionless Blockchains noch unerschlossenes Potenzial haben, trotz ihrer Limitationen neue Anwendungen zu realisieren, indem wir demonstrieren, wie Bitcoin das sichere Bootstrapping dezentraler Anonymisierungsdienste unterstutzen kann.

Insgesamt werfen wir ein neues Licht auf die moglichen Auswirkungen der auf Blockchains gespeicherten Daten. Unsere Analysen und technischen Beitrage erweitern so den Raum fur resiliente und bestandige datenzentrierte Blockchainedesigns.

## Acknowledgments

In a way, working on a PhD could retrospectively largely be described as sitting in the office most days, being occupied with activities, sometimes exciting, sometimes mundane, and occasionally going to a conference to celebrate the prior acceptance of a paper of hard work. What I mean to imply by this: On paper, this type of work does not sound like the blueprint for a thrilling or moving story. Yet, despite having felt incapable of believing otherwise, I have to come to the conclusion that my personal journey was all but uneventful. I started this journey on the brink of losing a fight against demons I was painfully aware of, yet could not fully grasp them. Now, I am able to write these words and feel genuinely happy.

A major part of this turn is owed to the great persons I met along the way. I will only highlight some of them and definitely forget someone in the following. But I count on you to know your contributions and how much I value them. Thank you.

First off, I want to thank my advisor Klaus Wehrle for giving me the chance to work on my PhD at COMSYS and believing in my capabilities when I could not. The freedom of research you provide has had a significant impact on my personal and professional development and helped me overcome challenges I would not have felt suited for otherwise. Further, I want to thank Frank Kargl for taking on the role of the second opponent during my PhD defense.

My time at COMSYS could have never been as great without my inspirational and great colleagues. This holds especially, but not exclusively, for my partners in crime of the Security & Privacy group. Martin, Henrik, Jens, and René; you helped me kick off and develop my researcher career back when the group was comparably small. I am grateful to have learned from all of you on a professional and personal level. Markus, Ina, Konrad, Eric, Christian, and Johannes; you make up the not-so-small next generation, and I am confident that you will excel at your endeavors. No, I did not forget you, Jan. I will never forget our paper-writing marathon in 2019, it was a great demonstration of how well our different approaches complement each other. And, “besides,” you became a great friend along the way and I could not have wished for anyone else to share an office with. Keep up the great work! Further, I have to thank those who endured my rants and kept up the pragmatism, which sometimes also works wonders in research, who would have known. I am looking at you, Torsten, Jan, Ike, Constantin, and also Claudia! Of course, also everyone whom I did not mention explicitly here contributes their fair share to keeping COMSYS a great place. Besides all other colleagues, this also holds for all the students I had the honor of helping take their first steps, of whom there are too many to name here.

Finally, I want to thank my family for all their support and believing in me. To my mother Lydia, my father Karli, and my brother Nico: Thank you for everything. The same holds for all my supportive friends, you are the best! The best gift of all this time, however, was having the indescribable luck of randomly meeting my wife. Roxanna, I do not know how my life would have turned out without you. But, frankly, I do not care in the slightest. Getting to know you was the single best and most impactful event of my life. Not only am I genuinely happy, but impatiently anticipating what our future will hold for us. I love you.





# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problem Analysis . . . . .	3
1.1.1	Promises of Blockchain Technology . . . . .	3
1.1.2	Problems of Blockchain-based Data Management . . . . .	5
1.2	Research Questions . . . . .	8
1.3	Contributions . . . . .	9
1.3.1	Attribution of Contributions . . . . .	14
1.4	Outline . . . . .	16
<b>2</b>	<b>Bitcoin and Blockchain Technology</b>	<b>17</b>
2.1	Bitcoin Overview . . . . .	18
2.2	Blockchain Structure . . . . .	23
2.2.1	Overview . . . . .	23
2.2.2	Block Structure . . . . .	24
2.2.3	Merkle Trees . . . . .	27
2.3	Transactions . . . . .	29
2.3.1	Overview . . . . .	29
2.3.2	Coin Ownership . . . . .	31
2.3.3	Scripting System . . . . .	35
2.3.4	Segregated Witnesses . . . . .	42
2.3.5	Managing Unspent Transaction Outputs . . . . .	43
2.4	Consensus . . . . .	44
2.4.1	Mining Process . . . . .	45
2.4.2	Distribution and Validation Process . . . . .	47

2.4.3	Forks . . . . .	49
2.5	Peer-to-Peer Network . . . . .	52
2.5.1	Network Overview . . . . .	52
2.5.2	Node Connections . . . . .	55
2.5.3	Data Dissemination . . . . .	56
2.5.4	Initial Synchronization Process . . . . .	58
2.6	Summary . . . . .	59
<b>3</b>	<b>Systematic Analysis of Non-Financial Blockchain Content</b>	<b>61</b>
3.1	Motivation . . . . .	61
3.1.1	Contributions . . . . .	62
3.2	Problem Analysis . . . . .	64
3.2.1	History of Bitcoin Content Insertion . . . . .	64
3.2.2	Model for Blockchain Content Insertion and Distribution . . . . .	66
3.3	Analysis of Content Insertion Methods . . . . .	67
3.3.1	Low-Level Content Insertion Methods . . . . .	68
3.3.2	Content Insertion Services . . . . .	72
3.4	Benefits and Risks of Blockchain Content . . . . .	74
3.4.1	Benefits of Blockchain Content Insertion . . . . .	74
3.4.2	Negative Consequences of Blockchain Content Insertion . . . . .	75
3.5	Analysis of Non-Financial Content . . . . .	81
3.5.1	Measurement Framework and Methodology . . . . .	81
3.5.2	Quantitative Analysis of Content Insertion . . . . .	86
3.5.3	Assessment of Blockchain Files . . . . .	95
3.6	Related Work . . . . .	99
3.7	Conclusion and Future Work . . . . .	101
<b>4</b>	<b>Mitigation of Unwanted Blockchain Content</b>	<b>103</b>
4.1	Motivation . . . . .	103
4.1.1	Constraints for Mitigation Schemes . . . . .	105
4.1.2	Contributions . . . . .	107
4.2	Prevention Strategies Against Unwanted Content . . . . .	107

4.2.1	Related Work . . . . .	109
4.2.2	Problem Statement . . . . .	110
4.2.3	Filtering Content-Holding Transactions . . . . .	112
4.2.4	Mandatory Minimum Transaction Fees . . . . .	114
4.2.5	Hardened On-Chain Addresses . . . . .	120
4.2.6	Summary and Future Work . . . . .	124
4.3	Moderation of Blockchain Content . . . . .	125
4.3.1	Related Work . . . . .	126
4.3.2	Missing Swift and Transparent Redactions . . . . .	129
4.3.3	Cryptographic Building Blocks . . . . .	131
4.3.4	RedactChain Overview . . . . .	135
4.3.5	Detecting Unwanted Blockchain Content . . . . .	137
4.3.6	Decentralized Redaction Process . . . . .	139
4.3.7	Coordinating Short-Lived Redaction Juries . . . . .	142
4.3.8	Evaluation . . . . .	145
4.3.9	Summary and Future Work . . . . .	150
4.4	Conclusion . . . . .	152
<b>5</b>	<b>Retrofitting Blockchains with Pruning Capabilities</b>	<b>153</b>
5.1	Motivation . . . . .	154
5.1.1	Problem Analysis . . . . .	155
5.1.2	Contributions . . . . .	158
5.2	Survey of Related Work . . . . .	159
5.2.1	Survey Criteria . . . . .	159
5.2.2	Retrospective Changes to Established Blockchain Systems . . . . .	159
5.2.3	Alternative Blockchain Designs with Pruning Capabilities . . . . .	162
5.3	Design Goals . . . . .	164
5.4	CoinPrune Overview . . . . .	165
5.5	Retrofittable Block Pruning . . . . .	167
5.5.1	Data Management . . . . .	167
5.5.2	Coordination of Established Nodes . . . . .	169
5.5.3	Bootstrapping of New Nodes . . . . .	169

5.6	Handling Application-Level Data . . . . .	170
5.6.1	Obfuscation of Illicit Data from the UTXO Set . . . . .	170
5.6.2	Preservation of Application-Level Data . . . . .	173
5.7	Seamless Integration into Bitcoin . . . . .	174
5.7.1	Additional On-Chain Data . . . . .	174
5.7.2	Adaptions to the Peer-to-Peer Protocol . . . . .	174
5.8	Security Discussion . . . . .	175
5.8.1	Snapshot Validity . . . . .	175
5.8.2	Reliability of Reaffirmations . . . . .	177
5.8.3	Peer-to-Peer Attacks . . . . .	178
5.9	Performance Evaluation . . . . .	179
5.9.1	Testbed Setup for Synchronization Measurements . . . . .	179
5.9.2	Storage Savings . . . . .	179
5.9.3	Evaluation of Synchronization Performance . . . . .	181
5.10	Conclusion and Future Work . . . . .	183
<b>6</b>	<b>Blockchain-based Bootstrapping of Anonymity Services</b>	<b>185</b>
6.1	Motivation . . . . .	186
6.1.1	Problem Analysis . . . . .	186
6.1.2	Contributions . . . . .	190
6.2	Design Goals . . . . .	191
6.3	AnonBoot Overview . . . . .	192
6.3.1	Design Intuition . . . . .	192
6.3.2	Overview of Bootstrapping Steps . . . . .	193
6.4	Sybil-Resistant Index of Peers and Services . . . . .	195
6.4.1	Message Types . . . . .	195
6.4.2	Pulse-based Message Release . . . . .	198
6.5	Bootstrapping Secure Anonymity Services . . . . .	199
6.5.1	Connecting Users and Privacy Peers . . . . .	200
6.5.2	Local Selection of Privacy Peers . . . . .	200
6.5.3	Service Requests for Peer Election . . . . .	201
6.6	Realization of Use Cases . . . . .	202

6.6.1	Decentralized Onion Routing via AnonBoot . . . . .	202
6.6.2	Shuffling Networks and Cryptocurrency Tumblers . . . . .	203
6.6.3	User-Centered Redaction Juries . . . . .	204
6.7	Security Discussion . . . . .	205
6.7.1	Proof of Work Against Sybil Attacks . . . . .	206
6.7.2	Security of Bootstrapped Services . . . . .	207
6.8	Performance Evaluation . . . . .	210
6.8.1	Synchronization with Host Blockchain . . . . .	210
6.8.2	Small Blockchain Footprint and Low Costs . . . . .	211
6.9	Related Work . . . . .	212
6.10	Conclusion and Future Work . . . . .	213
<b>7</b>	<b>Conclusion</b>	<b>215</b>
7.1	Contributions . . . . .	216
7.1.1	Systematic Analysis of Non-Financial Blockchain Content . .	216
7.1.2	Mitigation of Unwanted Blockchain Content . . . . .	216
7.1.3	Retrofitting Blockchains with Pruning Capabilities . . . . .	217
7.1.4	Blockchain-based Bootstrapping of Anonymity Services . . . .	218
7.2	Future Work . . . . .	219
7.3	Closing Remarks . . . . .	220
	<b>Abbreviations and Acronyms</b>	<b>221</b>
	<b>Bibliography</b>	<b>223</b>