COM SYS
Chair of
Communication and Distributed Systems

RWTH AACHEN UNIVERSITY

Roman Matzutt

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

Reports on Communications and Distributed Systems
Editor: Prof. Dr.-Ing. Klaus Wehrle

VOL 24

# 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 • e-mail: 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 Währung Bitcoin hat die Blockchain als eine Datenstruktur etabliert, mit der Nutzer dezentralisiert Konsens darüber erlangen können, wem welche Münzen gehören. Seitdem hat sich die Blockchaintechnologie stetig weiterentwickelt, sodass nun sich misstrauende Parteien ohne Intermediär über das Internet interagieren können, 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 gedämpft. Diese Nachteile kommen insbesondere bei den frei zugänglichen Permissionless Blockchains, wie Bitcoin, zum Tragen. Beispielsweise muss Bitcoin die Blockchain über energieintensive Berechnungen gegen Angreifer absichern, was zu Skalierbarkeitsproblemen führt, da so nur wenige Zahlungen auf einmal akzeptiert werden können. Obwohl solche technischen Herausforderungen bereits intensiv studiert wurden, wurde der Einfluss der Daten in der Blockchain bisher vernachlässigt. Dabei ist dieser Einfluss unbestreitbar: Einerseits können Unbekannte Daten unwiderruflich und ohne vorgesehenen Löschprozess anhängen. Andererseits bedingt der Betrieb einer Blockchain eine massive Replizierung ihrer gesamten und stetig wachsenden Historie. Daher bedürfen die Auswirkungen der Blockchain-Daten einer sorgfältigen 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 für dezentralisierte Datenmanagement-Systeme zu beurteilen und zu verbessern. Wir identifizieren zwei Kernherausforderungen des Blockchain-basierten Datenmanagements, nämlich die Moderierbarkeit der Daten und den Bedarf, die wachsenden Speicheranforderungen aufgrund stetig angehängter Daten abzumildern. Zudem bemessen wir das Potenzial der Blockchain, mittels ihrer spezifischen Eigenschaften weitere Anwendungen zu ermöglichen. Diese Herausforderungen adressieren wir auf technischer Ebene mittels folgender Beiträge.

Als ersten Beitrag analysieren wir das Phänomen des Einfügens von Blockchaininhalten auf konzeptioneller, technischer und empirischer Ebene. Unsere Analyse zeigt, dass dieses Vorgehen gebräuchlich ist und Vorteile für 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 einzudämmen, indem wir Strategien, die das Einfügen unerwünschter Inhalte verhindern, und eine editierbare Blockchain, die eine rasche und transparente Löschung erlaubt, vorschlagen. Unser dritter Beitrag adressiert das Problem wachsender Blockchaingrößen, indem ein graduell ausrollbares Block-Pruning-Verfahren definiert wird, das Bitcoin-Nutzern nachträglich ermöglicht, obsolete 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 unterstützen kann.

Insgesamt werfen wir ein neues Licht auf die möglichen Auswirkungen der auf Blockchains gespeicherten Daten. Unsere Analysen und technischen Beiträge erweitern so den Raum für resiliente und beständige datenzentrierte Blockchaindesigns.

# 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