

Digital Privacy and Security


Meng Shen
Xiangyun Tang
Wei Wang
Liehuang Zhu

Security and Privacy in Web 3.0

 Springer


Digital Privacy and Security

Series Editor

Shui Yu , School of Computer Science, University of Technology Sydney, Ultimo, NSW, Australia

Editorial Board Members

Nirwan Ansari, University Heights, New Jersey Institute of Technology, Newark, NJ, USA

Muhammad Khurram Khan , College of Computer & Information Sci, King Saud University, Riyadh, Saudi Arabia

Rongxing Lu , University of New Brunswick, Fredericton, NB, Canada

Peter Mueller, IBM ZURICH RESEARCH LABORATORY, scanning tunneling microscopy, RUESCHLIKON, Switzerland

Tianqing Zhu, School of Software, University of Technology Sydney, Ultimo, Australia

Digitalization is a big trend at the moment in the whole history of human beings. However, the movement introduces unprecedented challenges in security and privacy.

The book series on Digital Privacy and Security aims to develop and disseminate understandings of innovations, paradigms, techniques, and technologies in the contexts of digital world related research and studies. It covers security, privacy, availability, and dependability issues for digital systems and applications. It welcomes emerging technologies, such as security and privacy in artificial intelligence, digital twin, blockchain, metaverse, semantic communications, and so on.

The series serves as an essential reference source for security and privacy in the digital space. It publishes thorough and cohesive overviews on state-of-the-art topics in cyber security and privacy, as well as sophisticated techniques, original research presentations and in-depth case studies in the domain. The series also provides a single point of coverage of advanced and timely emerging topics and a forum for core concepts that may not have reached a level of maturity to warrant a comprehensive textbook. The intended audience includes students, researchers, professionals, and industrial practitioners.

The quality is assured through rigorous peer review, based on the editors' review and selection and adequate refereeing by independent experts.

Meng Shen • Xiangyun Tang • Wei Wang •
Liehuang Zhu

Security and Privacy in Web 3.0

 Springer

Meng Shen 
School of Cyberspace Science and
Technology
Beijing Institute of Technology
Beijing, Beijing, China

Wei Wang
Ministry of Education Key Lab for
Intelligent Networks and Network Security
Xi'an Jiaotong University
Xi'an, China

Xiangyun Tang
School of Information Engineering
Minzu University of China
Beijing, China

Liehuang Zhu
School of Cyberspace Science and
Technology
Beijing Institute of Technology
Beijing, China

ISSN 2731-992X

Digital Privacy and Security

ISBN 978-981-97-5751-0

<https://doi.org/10.1007/978-981-97-5752-7>

ISSN 2731-9938 (electronic)

ISBN 978-981-97-5752-7 (eBook)

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

If disposing of this product, please recycle the paper.

Preface

Web 3.0 is the next generation of the Internet built on decentralized technologies such as blockchain and cryptography. It is designed to address issues encountered in the previous generation of the Internet such as imbalanced distribution of interests, monopoly of platform resources, and leakage of personal privacy. This book explores the challenges and solutions related to ensuring security and privacy in the context of the evolving Web 3.0 architecture.

Web 3.0 is characterized by decentralized networks, blockchain technology, and enhanced user control over data. As Web 3.0 evolves, the focus on addressing security and privacy concerns becomes increasingly crucial. This book provides a comprehensive understanding of the security and privacy issues specific to Web 3.0 and offers practical defense frameworks and methods to mitigate these challenges.

This book is dedicated to the specific aspects of security and privacy in Web 3.0, from introducing the architecture and addressing the inherent issues to presenting innovative methods for privacy-preserving computing, user behavior identification, and abnormal transaction detection. It is of particular interest to researchers in the field of Web 3.0, blockchain, and network security, as it summarizes the security and privacy concerns in Web 3.0, and brings a number of innovative technologies and practical solutions to protect sensitive data and maintain user privacy in Web 3.0 environments, equipping researchers with the knowledge necessary to design secure and privacy-aware applications and systems in Web 3.0.

Beijing, China
April 20, 2024

Meng Shen
Xiangyun Tang
Wei Wang
Liehuang Zhu

Acknowledgments

The authors would like to acknowledge our editors for their help and guidance, acknowledge the graduate students at Beijing Institute of Technology, Hanbiao Du, Hao Lu, Jin Meng, Yukai Liu, and Zhehui Tan, and acknowledge the graduate students at Beijing Jiaotong University, Bin Wang, for their contributions to the accomplishment of this book.

The authors would also like to acknowledge the great support from the National Key R&D Program of China with No. 2023YFB2703800, NSFC Projects with No. 62222201, U23A20304, and 62302539, Beijing Nova Program with No. 20220484174, and Beijing Natural Science Foundation with No. M23020.

The authors dedicate this book to their families in appreciation of their continuous love and support.

Contents

1	Introduction of Web 3.0	1
1.1	Background	1
1.1.1	The Evolution of Web	2
1.1.2	The Categorization of AI in Web 3.0	3
1.2	Architecture of Web 3.0	4
1.2.1	Infrastructure Layer	5
1.2.2	Interface Layer	6
1.2.3	Management Layer	7
1.2.4	Application Layer	8
1.3	Security and Privacy Issues in Web 3.0	8
1.3.1	Challenge in Balancing Privacy Protection and Collaborative Services in Web 3.0	9
1.3.2	Challenge in Guaranteeing Authentication and Behavior Identification in Web 3.0	10
1.3.3	Challenge in Auditing Data Assets in Web 3.0	10
1.4	Summary	11
	References	12
2	Security and Privacy Defense Framework for Web 3.0	15
2.1	Overview of Defense Framework	15
2.2	Privacy Protection of Digital Assets	17
2.2.1	Verifiable Privacy-Preserving Federated Learning	17
2.2.2	Membership Privacy Protection for Federated Learning	18
2.3	Authentication and Behavior Identification	19
2.3.1	Secure Mobile Device Authentication	19
2.3.2	User Behavior Identification via Traffic Analysis	20
2.4	Audit of Data Asset Transactions	20
2.4.1	Malicious Transaction Detection of Digital Assets	20
2.4.2	Malicious Transaction Deanonymity	21
2.5	Summary	22

3	Verifiable Privacy-Preserving Federated Learning in Web 3.0	25
3.1	Background	25
3.2	Security Requirements	27
3.2.1	System Model	27
3.2.2	Threat Model	28
3.2.3	Security Goals	28
3.3	Privacy Computing Method	29
3.3.1	System Overview	29
3.3.2	Initialization Phase	30
3.3.3	Phase I: Model Perturbation	31
3.3.4	Phase II: Gradient Computation	32
3.3.5	Phase III: Gradient Verification	32
3.3.6	Phase IV: Model Update	41
3.4	Performance Evaluation	42
3.4.1	Environmental Settings	42
3.4.2	Performance Under Attacks	43
3.4.3	Time Overhead	45
3.4.4	Communication Overhead	47
3.5	Summary	47
	References	49
4	Membership Privacy Protection for Federated Learning in Web 3.0	51
4.1	Background	51
4.2	Privacy Attack	55
4.2.1	Notation	55
4.2.2	Membership Inference Attack	55
4.3	Privacy Attack Defense	56
4.3.1	System Model	56
4.3.2	Threat Model	57
4.3.3	Scheme Overview	58
4.3.4	Parameter Filter	60
4.3.5	Noise Generator	62
4.4	Performance Evaluation	63
4.4.1	Datasets and Model Architectures	63
4.4.2	Model Accuracy	64
4.4.3	Defense Effectiveness	65
4.4.4	Time Efficiency	65
4.4.5	The Impact of the Parameter Settings	66
4.5	Summary	68
	References	68
5	Secure Mobile Device Authentication in Web 3.0	71
5.1	Background	71
5.2	Development of Mobile Device Authentication	73
5.3	Security Requirements of Mobile Device Authentication	74

- 5.3.1 Threat Model 74
- 5.3.2 Security Requirements 75
- 5.4 Secure Mobile Device Authentication 76
 - 5.4.1 Secure Mobile Device Authentication Framework 76
 - 5.4.2 Details of Secure Mobile Device Authentication 77
 - 5.4.3 Security Analysis 83
- 5.5 Performance Evaluation 85
 - 5.5.1 Experimental Settings 85
 - 5.5.2 Evaluation on Time Efficiency 85
 - 5.5.3 Evaluation on Communication Overhead 89
- 5.6 Summary 91
- References 91
- 6 User Behavior Identification via Traffic Analysis in Web 3.0 95**
 - 6.1 Background 95
 - 6.2 Behavior Identification Requirements 97
 - 6.3 Development of Behavior Identification 98
 - 6.4 Behavior Identification Method 100
 - 6.4.1 Traffic Interaction Graph 100
 - 6.4.2 The Benefits of TIG 102
 - 6.4.3 GraphDApp Overview 103
 - 6.4.4 GNN Architecture 104
 - 6.4.5 Design of MLPs 106
 - 6.5 Performance Evaluation 107
 - 6.5.1 Preliminary 107
 - 6.5.2 Dataset Collection 108
 - 6.5.3 Parameter Tuning of GraphDApp 110
 - 6.5.4 Closed-World Evaluation 112
 - 6.5.5 Open-World Evaluation 113
 - 6.6 Summary 114
 - References 115
- 7 Malicious Transaction Detection in Web 3.0 117**
 - 7.1 Background 117
 - 7.2 Development of Malicious Transaction Detection 119
 - 7.3 Design of Malicious Transaction Detection Model 120
 - 7.3.1 Learning Phase 120
 - 7.3.2 Detection Phase 124
 - 7.4 Performance Evaluation 125
 - 7.4.1 Environmental Settings 125
 - 7.4.2 Experiments and Results 126
 - 7.5 Summary 129
 - References 129

- 8 Malicious Transaction Deonymity in Web 3.0** 133
 - 8.1 Background..... 133
 - 8.2 Development of Malicious Transaction Deonymity 135
 - 8.3 Data Preparation 136
 - 8.3.1 Data Description 136
 - 8.3.2 Data Collection 137
 - 8.3.3 Ground-Truth Construction 138
 - 8.4 Design of Mixing Interaction Graph 140
 - 8.5 Construction of MixBroker 142
 - 8.5.1 Overview 142
 - 8.5.2 Account Feature Extraction 143
 - 8.5.3 GNN Link Prediction Mechanism 144
 - 8.5.4 Account Correlation 146
 - 8.6 Performance Evaluation 147
 - 8.6.1 Environmental Settings 147
 - 8.6.2 Evaluation on Accuracy 148
 - 8.6.3 Evaluation on Time Efficiency 150
 - 8.7 Summary 151
 - References 151

- 9 Conclusion and Future Direction** 155
 - 9.1 Conclusion, Insights, and Lessons Learned..... 155
 - 9.2 Future Research Directions 156

Chapter 1

Introduction of Web 3.0



Abstract Web 3.0 represents the next evolutionary phase of the Internet, characterized by decentralized networks, enhanced user sovereignty over data, and the integration of technologies such as blockchain, artificial intelligence (AI), and the Internet of Things (IoT). This chapter provides a comprehensive overview of Web 3.0 from its concept to its architecture, highlighting the significant role of AI in enhancing its features across various layers, including infrastructure, interface, management, and application. Furthermore, it delves into the pressing challenges related to security and privacy within the Web 3.0, including ensuring effective privacy protection, guaranteeing robust user authentication, and behavior identification, and the auditing and traceability of data assets. Through in-depth analysis, this chapter explores existing research efforts and proposes potential directions within the context of Web 3.0 for addressing these challenges, emphasizing the importance of balancing privacy protection with usability, improving authentication methods, and ensuring the integrity and traceability of data assets.

1.1 Background

Web 3.0, the latest advancement in Internet technology, has attracted considerable attention from academic and commercial spheres due to its dynamic development trajectory. Positioned as the next evolutionary phase of the Internet, Web 3.0 is underpinned by decentralized technologies such as blockchain and cryptography. Contrary to its predecessors, Web 1.0 and 2.0, Web 3.0 is characterized by its intelligent design, decentralized architecture, and heightened security protocols. These distinctive features have broadened its application scope across diverse domains, encompassing blockchain, artificial intelligence, and the Internet of Things. Despite being at an incipient research stage with formidable technological challenges awaiting resolution, Web 3.0 holds immense promise for innovation and progress. Its transformative potential is poised to catalyze significant advancements in human society in the foreseeable future.

This section commences by providing a historical overview of the World Wide Web, delineating its evolution, i.e., three distinct generations: Web 1.0, Web 2.0, and Web 3.0. We provide an overview of these three generations' traits and development processes. AI technology plays a crucial role in Web 3.0, spanning detection, generation, optimization, and prediction tasks, contributing to enhanced security, content creation, asset pricing forecasts, and operational efficiency improvements. Subsequently, we discuss the classification and applications of AI technologies within the context of Web 3.0.

1.1.1 The Evolution of Web

Web 1.0 The World Wide Web was created in 1989 by Tim Berners-Lee at CERN in Switzerland, which launched the Internet's application era. The next year, he and his colleagues built the first website <http://info.cern.ch/> [4]. The key elements of a webpage, such as HTML [29], HTTP [13], browsers, and so forth, were created in the years that followed. Web 1.0 brings the Internet application to the era of personal computers, although most users are limited to using search and access features for information retrieval rather than content editing.

Web 2.0 The idea of the Web as a platform was initially presented at the inaugural Web 2.0 conference in 2004, which was organized by O'Reilly Media and MediaLive [39]. Facebook was created in the same year by Mark Zuckerberg. Following the establishment of several businesses like Google and Amazon, users were able to upload, like, comment on, and publish material on the network. Web 2.0 makes it possible for people to change material on the network, creating a massive social circle encircling the entire planet. But it has given rise to digital behemoths like Google and Facebook, which is seriously problematic for privacy.

The Conception of Web 3.0 The idea behind Web 3.0 is to overcome the drawbacks of the centralized Web that exists now. Tim Berners-Lee recommends a system called Solid [5], which is currently regarded as the Web 3.0 prototype, to address the issues of privacy leakage and the monopoly of large enterprises. It is a decentralized social application platform that guarantees user data management separate from the application that accesses it. Users store their data in pods, and the application needs to comply with certain protocols for access. Data privacy is simultaneously protected by distributed authentication and access control systems. The data of users is no longer dispersed over several systems. They are free to switch between platforms and control which services get access to their data. User data is no longer scattered across multiple platforms. Users have the freedom to move between platforms as well as manage which services can access their personal information.

Subsequently, Web 3.0, a system that combines the World Wide Web with distributed technologies like blockchains and smart contracts, was formalized by Gavin Wood [44]. Decentralization is widely seen as Web 3.0's most significant

feature [8]. In summary, Web 3.0 is the next iteration of the Internet, rebuilt using distributed technologies and emphasizing value expression and data ownership. Additionally, it functions on the tenet that data and people, not large corporations, should be the owners and managers of digital assets. Web 3.0 is characterized by its decentralized, blockchain-based, privacy-protected, and artificial intelligence-powered features.

1.1.2 The Categorization of AI in Web 3.0

The rapid advancement of AI technology in recent years has introduced innovative solutions to the challenges encountered in the development of Web 3.0. Through our research and analysis, it has become evident that AI holds a unique and critical role within the framework of Web 3.0, which can be segmented into four key categories: detection, generation, optimization, and prediction. In terms of detection, AI demonstrates efficacy in identifying malicious addresses, detecting counterfeit content, and ensuring the security of smart contracts. For generation tasks, AI is adept at creating diverse and high-quality digital avatars and assets. Additionally, AI excels in predicting the future prices of digital assets by leveraging extensive datasets. Moreover, AI contributes to optimization efforts by enhancing network architecture and refining the incentive mechanisms of Web 3.0 to augment operational efficiency.

In the subsequent sections, we delve deeper into how AI facilitates various aspects of Web 3.0 across different layers. Machine learning techniques can broadly be classified into traditional machine learning methods and deep learning techniques.

- (1) *Traditional machine learning*: Earlier machine learning techniques and algorithms, such as Support Vector Machine (SVM), Decision Trees, K-Nearest Neighbors, and Naive Bayes, that are based on statistical and mathematical concepts are referred to as traditional machine learning.

Support Vector Machine [26] divides data into groups or predicts output values by identifying the optimal boundaries or hyperplanes. Sigmoid, Gaussian, polynomial, and linear kernels are common kernel functions found in SVMs. For instance, by recognizing dangerous users and monitoring user behavior, SVM can be used to strengthen the security of the Web 3.0 identity management system [48]. An algorithm for statistical learning is Naive Bayes [43]. Its foundation is the Bayes theorem, which offers a method of estimating an event's probability using past knowledge of the circumstances most likely to be related to it. Data pricing can be done with Naive Bayes [3]. The way decision tree [25] functions is by creating a tree model of choices and possible outcomes. The tree is made up of edges that show test outcomes and nodes that reflect feature tests. The price of digital assets can be predicted using decision trees [2]. Several decision trees are combined in Random Forest [31], an ensemble learning

technique, to produce predictions for machine learning problems involving regression and classification. In Web 3.0, the random forest method can be applied to perceive network activities [35].

- (2) *Deep learning techniques*: It is a branch of machine learning that draws inspiration from the neural networks that make up the structure of the brain. It entails using a sizable dataset to train artificial neural networks, which are made up of layers of connected nodes or artificial neurons. Until the final output is produced, each layer processes the input before passing it on to the one after that. The hidden layers are those that lie between the input and output layers.

One particular kind of deep learning is convolutional neural networks [19]. Convolutional layers are used to identify local characteristics, whereas pooling layers are used to decrease spatial resolution. CNN offers an abundance of Web 3.0 applications that can be used to identify harmful content [31, 41, 42], enhance the blockchain incentive mechanism [9], and adjust the picture style [15, 23]. Neural networks with a focus on processing consecutive inputs are known as recurrent neural networks [16]. By keeping a concealed state, the network is able to retain and learn from the context of earlier time steps for many types of data, including time series, text, and speech. They work successfully on projects that call for being aware of the input's context. RNN is utilized in Web 3.0 to generate content [27], estimate the price of encrypted assets [32, 54], and identify unwholesome content [12]. A deep learning architecture called Graph Convolutional Network [53] is intended for graph-structured data. The graph convolution procedure, which creates a new representation for the current node by combining data from nearby nodes, is the fundamental part of GCN. In Web 3.0, GCN is extensively utilized for security and privacy protection, including transaction identification of malicious transactions [11], entity recognition [34], and network perception conduct [35].

1.2 Architecture of Web 3.0

This section presents a brand-new Web 3.0 architecture from application standpoint environments and scenarios, as depicted in Fig. 1.1. The infrastructure layer, interface layer, administration layer, and application layer are the four levels that make up Web 3.0. Data management is essentially handled by the infrastructure layer. Data from the real world is mapped to the digital realm by the interface layer. The Web 3.0 ecosystem is governed by the management layer. Moreover, the actual use cases for Web 3.0 are being created and implemented at the application layer. We explain why Web 3.0 is organized into these discrete layers, what each layer does, and why artificial intelligence is so important in this setting.

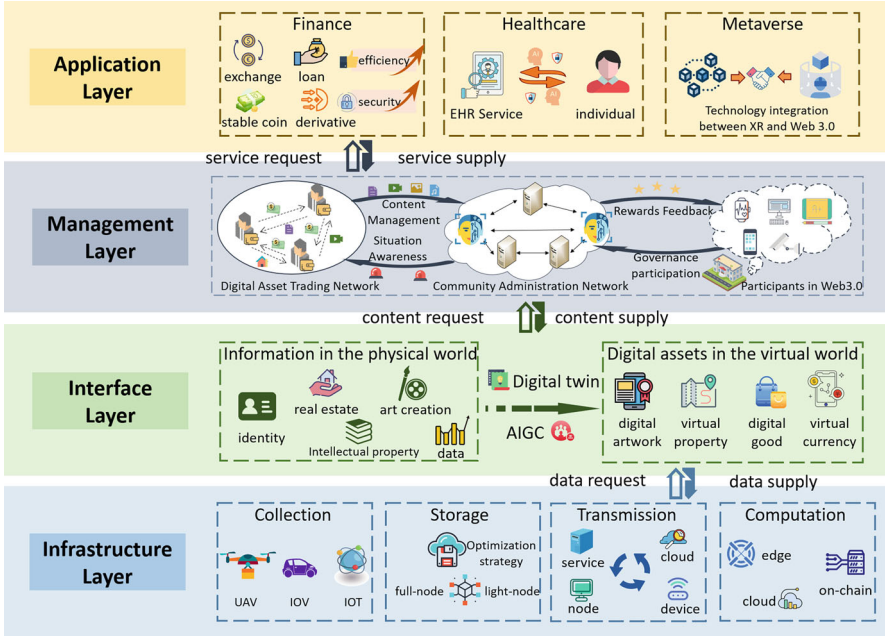


Fig. 1.1 The hierarchical architecture of Web 3.0

1.2.1 Infrastructure Layer

The infrastructure layer is in charge of data collection, processing, transmission, and storage. The deployment of Web 3.0 technologies, which prioritize decentralization and co-governance [7, 28, 33, 53], has resulted in a significant expansion of data sources, including real-time user feedback and the utilization of Internet of Things terminal devices. To guarantee the efficacy and transparency of the data, it may be stored using an on-chain approach alone or in conjunction with off-chain techniques. The data is sent to edge devices or nodes for analysis.

Two issues—the scalability issue and the security issue—can be used to sum up the issues with the infrastructure layer. The primary function of AI in the scalability problem is optimization, which RL is a popular method in it. The primary function of AI in the security issue is detection and prediction. Web 3.0 systems can become smarter thanks to AI technology, which can also greatly increase Web 3.0 data management efficiency in a variety of situations.

AI can help in enhancing the Web 3.0 infrastructure layer in various ways. AI can optimize data dissemination, collecting, and computing resource allocation procedures, which lowers energy costs and boosts Web 3.0 system throughput. Additionally, AI has the ability to create an intelligent Internet design that facilitates collaboration, enabling the Web 3.0 network’s intelligence to overcome resource constraints and enhance performance in challenging settings [35]. AI simulta-