Chapter 10
Blockchain-Based Data Market (BCBDM) Framework for Security and Privacy: An Analysis

Shailesh Pancham Khapre
Amity University, Noida, India

Chandramohan Dhasarathan
https://orcid.org/0000-0002-5279-950X
Madanapalle Institute of Technology and Science, India

Puviyarasi T.
https://orcid.org/0000-0003-3668-3264
Madanapalle Institute of Technology and Science, India

Sam Goundar
https://orcid.org/0000-0001-6465-1097
British University Vietnam, Vietnam

ABSTRACT

In the internet era, incalculable data is generated every day. In the process of data sharing, complex issues such as data privacy and ownership are emerging. Blockchain is a decentralized distributed data storage technology. The introduction of blockchain can eliminate the disadvantages of the centralized data market, but at the same time, distributed data markets have created security and privacy issues. It summarizes the industry status and research progress of the domestic and foreign big data trading markets and refines the nature of the blockchain-based big data sharing and circulation platform. Based on these properties, a blockchain-based data market (BCBDM) framework is proposed, and the security and privacy issues as well as corresponding solutions in this framework are analyzed and discussed. Based on this framework, a data market testing system was implemented, and the feasibility and security of the framework were confirmed.

DOI: 10.4018/978-1-6684-7132-6.ch010

INTRODUCTION

The amount of data in today's world is increasing rapidly. Since the establishment of Facebook, it has collected more than 300 PB (petabytes) of personal data, and the scale of this is still expanding. Balazinska et al., Researcher from IBM have suggested that 90% of the data in the world today has been generated in the past two years, and with the emergence of new equipment and technologies, data growth will accelerate further. In the era of big data, data is continuously collected and analyzed, leading to technological innovation and economic growth. Companies and organizations use the data they collect to provide personalized user services, optimize company decision-making processes, and predict future trends. People are concerned about the security of personal data and process of extensive data used (Pang et al., 2017), worrying about whether Internet companies that provide services and collect data will protect users' data privacy, and do people have little control over the data they generate and how they use it (Balazinska et al., 2011). In recent years, many incidents related to violations of user data privacy have been reported. The most famous example is that of Facebook's 50 million user data been leaked, and user privacy has been greatly violated.

To ensure the normal circulation, use of data, and maximize the value of big data, in recent years, many new organizations have emerged regarding personal data sharing and transactions. In addition to the traditional method of data circulation (that is, the widespread data exchange service between companies and users), a big data sharing transaction market has emerged to facilitate data transactions by matching data needs with data sources (Zyskind et al., 2015). These data markets are already of considerable size. These data markets are valued at tens of billions of dollars and continue to grow (Zyskind et al., 2015). In the data market, data holders display their data information to attract potential data consumers; data consumers search and select the data sets they need, and obtain data usage rights by paying a certain fee; the data market gains revenue by facilitating data transactions. However, as the scale of data sharing transactions and the value of data increase, it is expected that fraud and leakage in the process of sharing transactions will gradually increase. The general architecture of a centralized data market is shown in Figure 1. In this architecture, the market platform operated by a centralized company or organization plays a vital role in the system.

The parties involved in the data-market, data buyers and market platforms, can obtain higher profits through collusive fraud, arbitrage purchase strategies and so on. In addition, according to (Zheng et al., 2018), the centralized data transaction model lacks effective information communication channels between data buyers and data sellers, resulting in inefficient data transactions (Goldfeder et al., 2017). Finally, the data market platform has more information advantages, i.e., the market platform knows the data content, but the data buyer cannot know the data content before buying the data, so the market platform can illegally obtain profits by constructing information barriers and controlling information disclosure. The centralized data market as highlighter by (Wang & Krishnamachari, 2019) has some inevitable problems such as data security, data privacy protection and data circulation performance bottlenecks. First of all, the intermediary of data transactions (usually the market platform) must be safe and reliable. The market platform needs to have credibility to ensure that it will not illegally use the data in the transaction and leak the privacy of the data holder (Dziembowski et al., 2018). However, the market platform does have such a motive, and even if it uses or sells data illegally, it is generally difficult to pursue this illegal activity. At the same time, the centralized data market can easily become the target of attackers. The user's sensitive information (such as location, chat history, etc.) is stored in a centralized database, and there is a risk of privacy leakage and data loss. Most existing data markets run on centralized servers,