Blockchain in Data Science with Cryptography

Ishita Trivedi & Muskan Shukla
Pranveer singh institute of technology, Kanpur

----------ABSTRACT----------

Today, the wide use of the internet has made a huge quantity of data that needs high security to prevent its misuse. The data is collected at a large scale which is collectively termed Big Data. It is used for various administrative and business analytical purposes. Data science helps to accurately analyze and improve the weak areas. However, the problems of network security and data privacy threat have been a major hindrance in this path. The technique of Blockchain secures the data with the help of cryptography and makes it difficult for the attacker to break the security. Here in this paper, we will have a close look over the binding force of the trinity of data science, blockchain, and cryptography.

1. INTRODUCTION

The present era is data generating as well as data-consuming in work and practice. The internet is based on the isle of data which needs a high-security system to stand firm. The average data created everyday by a single person stands at 2.5 quintillion bytes per day. Hence there is a humongous amount of data generated by humans collectively, which is aptly stated as Big Data. Big data is analysed using Data Science. Data Science is the technique of making this data useful by using latest tools and techniques to find common patterns, thus concluding important conclusions and take better decisions. With the help of blockchain, we can significantly improve the quality of data, it has terminated the needless hassle of bringing data together. It rather proposes the new way of a decentralized framework where the data can be analysed right from the edge of end-user devices. The security protocols of blockchain use cryptography [1-9] which is a method of securing data from unauthorized access in wireless communication scenarios [10-15] and various network security applications [16-20]. Data science and blockchain are inseparable from building an efficient system.

- **BLOCKCHAIN**

Blockchain is a decentralized ledger that stores every transaction digitally. As it is decentralized, there is no one single authority, so nobody can alter the stored information, thus making it safe and transparent. The recorded information is decentral locked in such a way that no one can tamper with it, as the changes made in one block are reflected in the other blocks that follow it. As a result, even little changes in a single block cannot go unnoticed. Thus, blockchain technology is one of the most reliable approaches for internet confidentiality and protection

- **CRYPTOGRAPHY**

A way of safeguarding a point-to-point movement using a key is known as key encryption or cryptography. This key is a one-of-a-kind combination of numbers and letters that functions similarly to a password and is used to facilitate a transaction between two parties. The message or information to be transferred to the other node is encoded in an unreadable manner using key encryption. The other node will have to use the same or a different key (depending on asymmetric or symmetric) to decode the message to bring it into a readable format. Thus, we use the key cryptography method to ensure the identity of the sender and receiver and to secure the information from attack and misuse.

- **DATA SCIENCE**

Data Science is an interdisciplinary discipline that employs a variety of procedures, scientific methods, algorithms, and insights gleaned from a variety of noisy, unstructured, and structured data. It integrates data-driven knowledge and actionable insights to a wide range of application disciplines. Machine learning, data mining, and big data analysis are some of the techniques used in data science. Within the framework of statistics, mathematics, computer science, and domain knowledge, it employs techniques and theories from a variety of domains. Data science entails a wide range of disciplines and areas of knowledge in order to generate a comprehensive, complete, and sophisticated view of raw data. To develop models and make predictions using algorithms and other approaches, data scientists rely extensively on artificial intelligence, particularly its subfields of machine learning and deep learning.

- **BLOCKCHAIN AND DATA SCIENCE**

Currently, most of the data that organizations utilize is dispersed, requiring weeks or months of effort to sort out. Any type of human error can wreak havoc on the data's integrity, influencing the final analysis. When data is housed in a single centralized location, it runs the danger of being compromised. Data centres might also be tampered with and disclosed to the public. Everyone wants what they need, but ensuring that it is accurate and safe is a significant task. Data science requires a functional and reliable data set to do data analysis and predictive modelling. Data scientists can improve their abilities to manage data while also establishing a robust infrastructure using a decentralized blockchain.

- **BLOCKCHAIN AND CRYPTOGRAPHY**

By combining cryptography and encryption keys, blockchain applications make use of the concept of real-world signatures. For the storing and transfer of data values in secure formats, cryptography systems employ

advanced mathematical codes. As a result, it ensures that only the people who need the transaction or data may get it, read it, and process it, as well as verify the participants' and transaction's legitimacy. Cryptography and blockchain are in full swing.

2. BENEFITS OF BLOCKCHAIN IN DATA SCIENCE WITH CRPTOGRAPHY

- **DATA TRACEABILITY**

The huge bytes of generated data are often untraceable alone. Blockchain facilitates the data to be interlinked with each other forming a chain of data that can be easily traced. If a published account, for example, fails to adequately explain any methodology, any peer can analyse the entire process and determine how the results were reached. Anyone can learn whether data is reliable to use, how to store it, how to update it, where it originates from, and how to utilize it properly thanks to the ledger's transparent channels. To summarise, blockchain technology will allow users to track data from start to finish.

- **DATA PRIVACY AND SECURITY**

The data shared over the internet is prone to various cyber and network attacks. The surveys find that there is a cyber-attack every 39 seconds. Thus, the security and privacy of data are the quintessential need of the hour. Here comes the crucial role of cryptography. In the presence of malevolent third parties, known as adversaries, cryptography ensures safe communication. Encryption transforms a plaintext input into an encrypted output using an algorithm and a key (i.e., ciphertext). If the same key is utilized, a given algorithm will always turn the same plaintext into the same ciphertext. If an attacker can't deduce any properties of the plaintext or key from the ciphertext, the algorithm is considered secure. Given a significant number of plaintext/ciphertext combinations that used the key, an attacker should not be able to deduce anything about it. The key is in an algorithmically locked plain text-looking format, it is a daunting task for the attacker to identify the key.

- **PROTECTION OF DATA SOVEREIGNTY**

Currently, there is no single authority that solely controls the user data, once the data is created it travels infinitely. As the data is untracked, it is a tedious task to prevent its misuse. Here Blockchain suggests a smart way to protect its authorization with help of smart contracts. A smart Contract is a self-automating contract that acts as an agreement between the buyer and seller that is digitally stored on a blockchain that is converted into code lines. This computer code implements itself when the required conditions are met. They automate the workflow by implementing predefined subsequent actions. Using smart contracts allows anonymous parties to carry out transactions and agreements without the need for an intermediary, legal system, or external enforcement mechanism, thus, cutting out any fees owed to these third parties while promoting frictionless transactions. It saves a lot of time spent on different tedious tasks.

3. DATA SCIENCE IN DECISION MAKING

In today's scenario a lot of data is generated and it is so vast that it could not be analyzed manually. To analyze these data we require Data Science even the different fields of data science (data analytics, data statistics etc.) to convert these data into meaningful information.

- **DECISION MAKING**

When it comes to decision making, sometimes we humans are unable to make decision a lot of time and money is spent to identify to identify the root cause. But data science helps the organization to make decision, fast and accurate.

- **SPEED**

Sometimes decision need to be taken quickly in real time .With the help of human it may not be possible but with the use of data science we can decide and analyze it quickly.

- **DATA SCIENCE IN SPEECH RECOGNITION**

Speech recognition is a capability which enables a program to process human speech into a written format. Speech recognition identifies the speaker whereas speech analysis analyzes each word. In recent years we have seen a sudden rise in speech recognition and analysis software, due to introduction of latest technologies such as machine learning, big data, deep learning, data analytics etc.

- **AUDIO DATA**

With the use of big data huge amount of data can be collected and stored for use. Similar way audio is also collected and stored for future use, to recognize the user.

- **MEDICAL IMAGE ANALYSIS**

Previously medical images like X-rays, MRI's, CT-Scans etc. are examined by doctors only ,but now with the use of data science and machine learning it is possible for machines to examines the images and automatically detect the defect or problem in the image.

- **DRUG DISCOVERY**

In drug discovery new candidate medicines are examined. With the help of data science we can simplify this process and can easily tell the success rate of newly discovered medicine.

- **USE OF DATA SCIENCE IN TRANSPORT**

Nowadays data science is widely used in transport sector. Data science is making safer driving environment for the drivers; it also optimizes the vehicle performance. It also helps to analyze the fuel (as how much fuel is used) and it also analyzes the behaviour of driver (whether he is drunk or sleepy). The self driving car is completely using data science technology.

- **FUTURE OF DATA SCIENCE**

As more and more people are connected to mobile devices and internet ,huge amount of data is generated which is to be store for future use and day by day users are increasing continuously hence use of technologies such as big data, IoT, AI, ML , blockchain etc. will mark remarkable growth in the future.

The future of data science is very bright. Along with AI and ML, data science will contribute towards a higher level of intelligent decision making.

- **HOW BLOCKCHAIN CAN HELP BIGDATA**

As we know that big data refers to quantity of data and blockchain refers to quality of data. With the use of blockchain data can be handled more efficiently. Likewise data generated by blockchain is validated, structured and immutable, hence it enhances big data. Currently data generated by businesses is mostly scattered. Therefore most businesses are searching for an option to store their data more efficiently and secured or not accessible to any other person. Blockchain can store this big data and also provide security to keep this data safe and private.

- **CHALLENGES IN THE CONVERGENCE OF BLOCKCHAIN AND TECHNOLOGY.**

Even if blockchain can bring many benefits to data management, there are still some challenges. They are as:

- **SCALABILITY**

It is the ability of a system, network, or process to expand its potential by handling increasing workloads. Blockchain has some issue with scalability, the larger the size of the blockchain the longer time it will take to copy the data to the new nodes. Hence participants cannot download and verify the history of the entire slice. Therefore we should continue to study the blockchain solution for data science.

- **DIFFICULTY IN ACCURATE ANALYSIS**

As the anonymity and privacy protection are the main characteristics of blockchain, it makes difficult for researchers to obtain valuable data. The more anonymous the data is, the more it becomes difficult for researchers to predict and analyze the data. The combination of blockchain and data science forms a very complex network. With the anonymity of blockchain, it is very difficult to deeply mine data on chain.

- **INTENSIFIED COMPETITION**

In future there will be competition on how to deal with data that is to say who can provide more accurate and valuable content and tools is the key to winning blockchain big data. Every enterprise wants to collect more and more data to improve its competitiveness in future. Data acquisition is the first barrier to competition. In many areas whether data can be mined or not determines the survival of the enterprise. In blockchain there is no limit for big data companies to excess data. Thus, blockchain is designed in an ingenious way to immune against fraud and piracy.

REFERENCES

[1]. V. Shukla, A. Chaturvedi, N. Srivastava, Nanotechnology and cryptographic protocols: issues and possible solutions, Nanomaterials and energy, volume 8, issue 1, 2019, 1-6,
DOI: 10.1680/jnaen.18.00006

[2]. M.K. Misra, A. Chaturvedi, S.P. Tripathi, V. Shukla, A unique key sharing protocol among three users using non-commutative group for electronic health record system, Journal of discrete mathematical sciences and cryptography, volume 22, issue 8, 2019, 1435–1451,
DOI: 10.1080/09720529.2019.1692450

[3]. A. Chaturvedi, N. Srivastava, V. Shukla, A secure wireless communication protocol using Diffie-Hellman key exchange, International journal of computer applications, volume 126, number 5, 2015, 35-38,
DOI: 10.5120/ijca2015906060

[4]. V. Shukla, A. Chaturvedi, M.K. Misra, On authentication schemes using polynomials over non commutative rings, Wireless personal communications, volume 118, issue 1, 2021, 1-9,
DOI: https://doi.org/10.1007/s11277-020-08008-4

[5]. V. Shukla, A. Mishra, S. Agarwal, A new one time password generation method for financial transactions with randomness analysis, Innovations in electrical and electronic engineering (Part of the lecture notes in electrical engineering book series (LNEE, volume 661)), 2020, 713-720,
DOI: https://doi.org/10.1007/978-981-15-4692-1_54

[6]. V. Shukla, A. Mishra, A. Yadav, An authenticated and secure electronic health record system, IEEE international conference on information and communication technology, 2019, 1-5,
DOI: 10.1109/CICT48419.2019.9066168

[7]. A. Chaturvedi, V. Shukla, M.K. Misra, Three party key sharing protocol using polynomial rings, 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2018, 1-5,
DOI: 10.1109/UPCON.2018.8596905

[8]. V. Shukla, A. Chaturvedi, N. Srivastava, Secure wireless communication protocol: to avoid vulnerabilities in shared authentication, Communications on applied electronics, volume 7, number 6, 2017, 4-7,
DOI: 10.5120/cae2017652680

[9]. V. Shukla, M.K. Misra, A. Chaturvedi, Journey of cryptocurrency in India in view of financial budget 2022-23, Cornell university arxiv, 2022, 1-6,
DOI: https://doi.org/10.48550/arXiv.2203.12606

[10]. V. Shukla, N. Srivastava, A. Chaturvedi, A bit commitment signcryption protocol for wireless transport layer security (wtls), IEEE international conference on

electrical, computer and electronics engineering , 2016, 83-86,
DOI: 10.1109/UPCON.2016.7894629

[11]. V. Shukla, A. Chaturvedi, N. Srivastava, A new secure authenticated key agreement scheme for wireless (mobile) communication in an EHR system using cryptography, Communications on applied electronics, volume 3, number 3, 2015, 16-21,
DOI: 10.5120/cae2015651903

[12]. A. Chaturvedi, N. Srivastava, V. Shukla, S.P. Tripathi, M.K. Misra, A secure zero knowledge authentication protocol for wireless (mobile) ad-hoc networks, International journal of computer applications, volume 128, number 2, 2015, 36-39,
DOI: 10.5120/ijca2015906437

[13]. V. Shukla, A. Chaturvedi, N. Srivastava, Authentication aspects of dynamic routing protocols: associated problem & proposed solution, International journal of recent technology and engineering, volume 8, issue 2, 2019, 412-419,
DOI: 10.35940/ijrte.B1503.078219

[14]. V. Shukla, A. Kushwaha, S.S. Parihar, S. Srivastava, V. P. Singh, Authenticated wireless information display system using GSM module, Communications on applied electronics, volume 5, number 3, 2016, 7-11,
DOI: 10.5120/cae2016652251

[15]. V. Shukla, A. Chaturvedi, N. Srivastava, Double layer cryptographic protocol for mobile ad-hoc networks (MANETs) by commitment scheme, Communications on applied electronics, volume 7, number 9, 2017, 32-36,
DOI: 10.5120/cae2017652716

[16]. V. Shukla, A. Chaturvedi, N. Srivastava, A secure stop and wait communication protocol for disturbed networks, Wireless personal communications, volume 110, 861-872, 2019,
DOI: https://doi.org/10.1007/s11277-019-06760-w

[17]. V. Shukla, A. Mishra, A new sequential coding method for secure data communication, IEEE international conference on computing, power and communication technologies, 2020, 529-533,
DOI: 10.1109/GUCON48875.2020.9231252

[18]. A. Chaturvedi, V. Shukla, M.K. Misra, A random encoding method for secure data communication: an extension of sequential coding, Journal of discrete mathematical sciences and cryptography, volume 24, issue 5, 2021, 1189-1204,
DOI: https://doi.org/10.1080/09720529.2021.1932902

[19]. V. Shukla, A. Chaturvedi, N. Srivastava, A new one time password mechanism for client-server applications, Journal of discrete mathematical sciences and cryptography, volume 22, 2019, 1393-1406,
DOI: 10.1080/09720529.2019.1692447

[20]. V. Shukla, M.K. Misra, A. Chaturvedi, A new authentication procedure for client-server applications using HMAC, Journal of discrete mathematical sciences and cryptography, volume 24, issue 5, 2021, 1241-1256,
DOI: https://doi.org/10.1080/09720529.2021.1932908