DEMEDIA – DECENTRALIZED SOCIAL MEDIA PROTOCOL

2023 - 234

Perera B. S. S.
Bandara A. M. C. A.
Dhananjani G. G. S.
Abeykoon A. W. Y. I. K

B.Sc. Special (Honors) Degree in Information Technology Specialization in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2023

DEMEDIA – DECENTRALIZED SOCIAL MEDIA PROTOCOL

2023 - 234

Perera B. S. S.
Bandara A. M. C. A.
Dhananjani G. G. S.
Abeykoon A. W. Y. I. K

Dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science specializing in Information Technology

Department of Information Technology

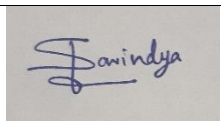
Sri Lanka Institute of Information Technology

Sri Lanka

September 2023

DECLARATION

We declare that this is our own work, and this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, we hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute our dissertation in whole or part in print, electronic or other medium. We retain the right to use this content in whole or part in future works (such as article or books).

Name	**Student ID**	**Signature**
Perera B. S. S.	IT20254698	
Bandara A. M. C. A.	IT20159726	
Dhananjani G. G. S.	IT20137496	
Abeykoon A. W. Y. I. K.	IT20157432	

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

…………………………..   …………………………

Signature of the supervisor: Date

(Mr. Kavinga Yapa Abeywardena)

ABSTRACT

This research introduces a decentralized social network architecture that addresses the privacy and security concerns associated with centralized social media platforms. It implies the establishment of a protocol referred to as "DeMedia," which would enable users to complete ownership over their data by keeping it locally on their devices instead of centralized servers. In order to ensure data privacy and integrity, the protocol utilizes peer-to-peer networking and cryptographic mechanisms. The DeMedia protocol aims to provide users with a more private and secure social media experience by using decentralized technologies such as peer-to-peer networking and distributed storage. It indicates user ownership and control of their data while excluding the need for expensive blockchain infrastructure. The platform also has tools for optimizing speed and preserving the integrity of user data, including data cache preservation and data integrity preservation. The proposed protocol offers a decentralized and secure social media network that provides users control and allows direct interaction without depending on centralized intermediary parties.

Keywords – *Peer-to-peer networking, Data integrity, Data caching, DeMedia protocol, Decentralization, Privacy, Security, Data ownership, Distributed networks*

ACKNOWLEDGMENT

The success of this research is due to the efforts of our supervisor, Mr. Kavinga Yapa Abeywardena. Additionally, we'd like to express our gratitude to the Department of Information Technology at the Sri Lanka Institute of Information Technology, as well as to CDAP academics and staff, for enabling this research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATION

MIT	Massachusetts Institute of Technology
TCP/IP	Transmission Control Protocol/Internet Protocol
IPFS	Interplanetary File System
TTL	Time To Live
DApps	Decentralized Applications
REST	Representational State Transfer
HTTP	Hypertext Transfer Protocol
AWS S2	Amazon Web Services
SOAP	Simple Object Access Protocol
GRPC	Google Remote Procedure Call
P2P	Peer to Peer
CDN	content delivery network

1.0 INTRODUCTION

1.1 Background literature

Social media platforms have witnessed exponential growth and have become an integral part of modern society. Popular platforms such as Facebook, Twitter, Instagram, and WhatsApp have connected billions of people worldwide. However, the centralized nature of these platforms has raised numerous concerns, including data privacy, misinformation, censorship, and the concentration of power. As a response to these challenges, decentralized social media networks have gained significant attention.

1.1.1 The Rise of Centralized social media

The early 2000s saw the advent of centralized social media platforms, with MySpace and Friendster leading the way. However, the launch of Mark Zuckerberg's Facebook in 2004 marked an important turning point. An important turning point in the development of social media was reached with that. Platforms like Twitter, Instagram, and WhatsApp further established the authority of these centralized digital platforms in the following years. [1].

Centralized social media platforms offer a range of compelling advantages that have contributed to their widespread popularity. One of the most noteworthy advantages is their user-friendliness. These platforms have been carefully designed to ensure that even individuals with minimal technical expertise can navigate them effortlessly. Communication on these digital platforms is seamless, allowing people to easily connect with friends and family, regardless of geographical distances. Moreover, centralized social media platforms extend their appeal beyond personal connections. Users can readily share various types of content, including photos, videos, and thoughts, with just a few simple clicks. These platforms have effectively harnessed the power of network effects, amassing vast user bases that make them powerful tools for sharing information, sources of entertainment, and hubs for social interaction [2].

In fact, the development of centralized social media has changed the way we engage and communicate in today's digital world. It has brought about a time where connectivity has no bounds, news and trends may spread quickly over the world, and the distinction between the actual and virtual worlds is becoming more and more unclear. These platforms have, in many ways, shaped how we express ourselves, share our experiences, and interact with the rest of the world in how we live every day. It becomes clear as we look more into the complex web of centralized social media that their influence goes well beyond screens and gadgets, influencing the very fabric of modern society. [3]

1.1.2 Challenges of Centralization

Centralized social media platforms, despite their widespread popularity, struggle with several pressing challenges, which have raised concerns among users, regulators, and policymakers. These challenges, which are caused by problems with data privacy, false information, content moderation, and power concentration, have a big impact on how these platforms work and what they can do. [10]

1.1.2.1 Data Privacy and Ownership

The handling of user data stands as one of the most prominent concerns associated with centralized social media platforms. These platforms gather considerable amounts of user data, often without clear and transparent consent mechanisms. The data collected is primarily employed for targeted advertising, but instances of data breaches and unauthorized access have fostered mistrust among users. Furthermore, users possess limited control over their own data, thereby prompting fundamental questions regarding the concepts of privacy and ownership.

User data represents a valuable asset for centralized platforms, enabling them to tailor advertisements and recommendations to individual users. However, the manner in which this data is obtained and utilized often raises ethical questions. Users may not be fully aware of the extent to which their data is collected, and the opacity surrounding data handling practices engenders mistrust.

In the context of ownership, users might reasonably expect that the content they create and share on social media platforms belongs to them. Yet, the reality often deviates from this expectation, as platforms assert ownership or licensing rights over user-generated content. This complex interplay between data privacy, ownership, and platform practices underscores the need for comprehensive regulatory frameworks to safeguard user rights and address these concerns effectively. [3]

1.1.2.2 Misinformation and Fake News

Centralized social media platforms confront significant criticism for their role in distribution misinformation and fake news. As a result of the viral nature of content sharing on digital platforms, misleading information can spread quickly, causing social conflict, and damaging the foundations of a well-informed society.

The mechanisms supporting the spread of misinformation on these platforms are complicated. Users can quickly share and amplify misleading content, often driven by emotional responses or confirmation bias. Algorithms employed by centralized platforms tend to prioritize engaging content, which may not necessarily equate to accurate or reliable information. As a result, misinformation can gain traction and reach a wide audience before corrective measures can be implemented.

The centralized nature of content moderation on these platforms poses a further challenge. While moderation is crucial to mitigate the spread of false information, it is a delicate balance to strike. Decisions regarding what content should be removed or labeled as misleading often entail subjective judgments, leading to accusations of bias and inconsistent enforcement of community guidelines. [4]

Addressing the issue of misinformation on centralized social media platforms necessitates a multi-pronged approach. This includes enhancing algorithmic accountability, promoting digital literacy, and fostering collaboration between platforms, fact-checkers, and regulators to curb the virulent spread of false information.

1.1.2.3 Censorship and Content Moderation

Over centralized social media platforms, the explosive subject of content moderation is significant. It can be difficult to reach a compromise between defending free expression and establishing social standards. Platforms are under a lot of pressure to remove inappropriate or damaging information while avoiding censorship that would violate users' rights to free communication. [5]

The transparency and consistency of platform policies is one of the primary problems with the filtering of content. Users frequently have trouble with a lack of understanding of what is appropriate and what pushes too far into the offensive zone. Furthermore, centralized platforms have been criticized for unfair censorship, with claims that some points of view are given priority over others.

In addition to addressing negative behaviours like harassment and hate speech, content moderation also presents challenges. Effective policies are not enough to address these problems; there also needs to be effective enforcement mechanisms. Given the huge quantity of user-generated content on these platforms, finding the ideal balance between automated content moderation and human monitoring continues to be a challenge.

Centralized platforms have attempted to refine their content moderation policies and invest in advanced AI systems to identify and mitigate harmful content. Nevertheless, the complexity of this endeavor persists, and ongoing dialogue with stakeholders, including users, is essential to shape more effective moderation practices.

1.1.2.4 Monopoly and Power Concentration

As certain large businesses control the majority of the marketplace, centralized social media platforms are frequently characterized by a concentration of authority. The impact on competition, innovation, and user choice is uncertain given this monopoly over the digital public sector.

The authority of a few significant companies in the social media sector can limit competition by establishing significant obstacles to entry for newcomers. Smaller or growing platforms struggle to compete with giant corporations' massive user foundation, significant resources, and network effects. The combination of these factors can lead to a lack of diversity and variety in the social media ecosystem, which makes it difficult for users to explore other platforms that correspond with their interests and values.

Furthermore, the concentration of power in the hands of a few corporations has significant effects on the free distribution of information and speech. These platforms operate as real gatekeepers of online communication, affecting conversation in society and the digital public area. Concerns regarding potential biases and decisions made by these companies can put doubt on the truthfulness and impartiality of online platforms as settings for free expression and idea exchange.

Competition issues have grown in prominence in recent years as regulators and policymakers evaluate the market power of centralized social media platforms. The power of these platforms in various areas of the digital ecosystem, from advertising to data gathering, has prompted investigations and calls for governmental involvement to create a more competitive and equal environment. [6]

Efforts to address the challenges posed by the concentration of power on centralized platforms include antitrust actions, regulatory oversight, and discussions on interoperability and data portability to empower users with greater control over their online experiences. These measures seek to strike a balance between allowing platforms to innovate and thrive while ensuring fair competition and safeguarding the interests of users and society as a whole.

1.1.3 The Road to Web 3.0

Moving from interactive web applications, or Web 2.0, to decentralized web applications, or Web 3.0, represents an important change in how we interact with the internet. With this evolution, a user-centric, private, and secure digital environment is proposed. Leading-edge research projects and efforts are reshaping this revolutionary path toward a decentralized internet experience. We examine these advancements and their critical contribution to this revolutionary change in this article. [7]

1.1.3.1 Ethereum and Smart Contracts

Ethereum is a significant blockchain technology, that is establishing itself as an essential component of the decentralized social media movement. Ethereum's main concept is smart contracts, which are self-executing agreements with predefined rules and conditions. These contracts allow for automated and trustless interactions, which makes them invaluable in the context of decentralized social media [8].

Smart contracts offer a wide range of uses in decentralized networks. For example, they enable decentralized content reward systems in which content authors receive payments automatically based on established criteria. Furthermore, smart contracts are used in identity verification processes on decentralized platforms to ensure the validity of users. Smart contracts also enable governance tools such as voting systems, allowing for community-driven decision-making.

1.1.3.2 Mastodon: Decentralized Microblogging

Mastodon is a brilliant example of a decentralized microblogging network that threatens the centralized social media companies' authority. Mastodon, which operates on a decentralized basis, decentralizes the social media experience by connecting numerous independently maintained servers known as instances. These instances interact with one another, establishing a network of distinct online communities [9].

Users on Mastodon can select instances that represent their values and preferences. Because users can join instances that match their interests and ideas, this federated method encourages variety in the digital environment. The decentralized system of Mastodon promotes a more inclusive and user-driven online environment in which control is distributed rather than centralized.

1.1.3.3 Decentralized Identity Solutions

Secure and user-controlled identity management is a top priority in shifting to decentralized social media. This essential component is addressed by decentralized identity systems such as uPort and Sovrin. These projects give users the ability to manage their digital identities across several channels while maintaining their privacy and control.

The goal of decentralized identity solutions is to eliminate dependency on centralized identity providers, who frequently gather and market user data. Users are able to prove their identity with uPort and Sovrin without revealing any personal information. This not only improves privacy but also empowers users to control their online personas, resulting in a more user-centric web experience [10].

1.1.3.4 Content Addressable Networks

Decentralized data storage is a fundamental challenge in the development of Web 3.0. Content Addressable Networks (CANs), exemplified by the InterPlanetary File System (IPFS), offer innovative solutions to this problem. IPFS allows content to be addressed and retrieved based on cryptographic hashes, ensuring data availability even if individual nodes go offline. The use of CANs enhances data resilience and availability in decentralized networks. Unlike traditional centralized servers, which are vulnerable to downtime and data loss, IPFS leverages a distributed network of nodes to store and retrieve data. This approach not only improves data reliability but also reduces the risk of data censorship or tampering [11].

The transition from centralized social media platforms to decentralized networks represents a significant paradigm shift in the online landscape. Pioneering projects and research initiatives have paved the way for this transformative journey, introducing concepts such as smart contracts, federated microblogging, decentralized identity solutions, and content addressable networks.

As these developments continue to shape the future of online interaction, they hold the promise of a more user-centric, private, and secure internet experience. It is imperative that users, developers, regulators, and policymakers closely monitor and support these initiatives to ensure that the principles of decentralization are upheld and that the web evolves in a manner that prioritizes the interests and rights of its users.

The rise of centralized social media platforms has revolutionized the way we communicate and share information. However, the numerous challenges associated with centralization, including data privacy, misinformation, censorship, and power concentration, have sparked interest in decentralized alternatives. Decentralized social media platforms, built on technologies such as blockchain, P2P networking, and cryptography, offer a path toward greater user empowerment, privacy, and control. As we move closer to Web 3.0, the evolution of decentralized social media promises to reshape the digital landscape, emphasizing user-centricity and data ownership.

1.1.4 Decentralized social media as a Solution.

The challenges associated with centralized social media platforms, such as data privacy concerns, misinformation spread, content moderation issues, and power concentration, have prompted the emergence of decentralized social media networks as a promising alternative. These decentralized platforms offer solutions that aim to distribute control, enhance privacy, and empower users. In this article, we explore the key concepts and technologies that underpin decentralized social media and their potential to address the challenges posed by centralized platforms. [13]

1.1.4.1 Blockchain Technology

Blockchain technology, originally developed for cryptocurrencies like Bitcoin, has found new applications in decentralized social media platforms. Blockchains offer transparency, security, and the ability to create decentralized applications (DApps). Platforms like Steemit and Minds, based on Ethereum blockchain, have harnessed blockchain technology to reward content creators and prioritize user ownership [7].

Blockchain's transparency ensures that transactions and data sharing on these platforms are recorded in a tamper-proof manner. Users can have confidence in the authenticity and integrity of the content they encounter, reducing the risk of misinformation and fraud. Additionally, blockchain-based reward systems provide content creators with a fairer distribution of earnings, increasing their incentive to produce quality content [14].

- Data Decentralization

Decentralized social media platforms prioritize data ownership as a fundamental principle. Users maintain control over their personal data, deciding who can access it and how it is used. This user-centric model seeks to eliminate the need for central servers, reducing the vulnerability to data breaches and unauthorized access [12].

By putting users in charge of their data, these platforms shift the balance of power away from centralized corporations. Users can choose to share their data selectively and revoke access when necessary, enhancing their privacy and control. Decentralized data storage solutions also reduce the risk of large-scale data breaches that have plagued centralized platforms [15].

- Peer-to-Peer Networking

Many decentralized social media systems use peer-to-peer (P2P) networking to allow users to communicate directly with one another. This method minimizes dependency on central servers, which improves privacy and reduces the chance of censorship. Mastodon, for example, uses peer-to-peer technology to establish decentralized microblogging communities.

P2P networks enable users to communicate directly without the need for intermediaries in order to making it difficult for third parties to monitor or regulate their interactions. This distributed communication strategy empowers users and promotes a community-driven responsible manner, in which community members enforce content restrictions collaboratively. As a result, decentralized platforms frequently provide users with a more democratic and censorship-resistant environment [16].

- Cryptographic Techniques

Cryptographic techniques play a vital role in ensuring data privacy and security on decentralized social media platforms. These techniques encompass encryption, digital signatures, and cryptographic protocols that protect user identities and the integrity of content. By leveraging cryptography, decentralized platforms enable secure and private interactions among users [17].

Encryption ensures that messages and data are unreadable to anyone without the necessary decryption keys, safeguarding user privacy and preventing unauthorized access. Digital signatures verify the authenticity of content, making it difficult for malicious actors to manipulate or forge information. Cryptographic protocols, such as secure communication channels, protect user identities and communications from eavesdropping and tampering.

Decentralized social media networks present a compelling solution to the challenges faced by centralized platforms. By embracing blockchain technology, data decentralization, peer-to-peer networking, and cryptographic techniques, these platforms offer users greater control over their data, enhanced privacy, and protection against misinformation and censorship.

As decentralized social media platforms continue to evolve and gain traction, they have the potential to reshape the digital landscape, fostering a more user-centric, transparent, and secure online environment. Policymakers, regulators, and users should closely monitor the development and adoption of decentralized social media to ensure that these platforms fulfill their promise of addressing the shortcomings of centralized counterparts.

1.2 Research Gap

Numerous issues have been recognized within the area of centralized social media platforms, generating significant research and intellectual exploration. However, the current landscape of digital communication reveals considerable research gaps that must be addressed. This section, based on an academic perspective, discusses these research gaps, and emphasizes the necessity of further research.

1.2.1 User Data Decentralization:

Current Progress: In recent years, there has been an increasing recognition of the importance of user data decentralization within the context of centralized social media platforms. This acknowledgment originates from a thorough appreciation of the need to provide users with more control and complete ownership rights over their personal data. In response to this crucial a variety of proposals, primarily based on blockchain technology and decentralized identity frameworks, have emerged as potential solutions to the multiple issues connected with user data management inside centralized social media ecosystems [1].

These proposals have established innovative frameworks and procedures in the current research environment, with the goal of redefining the usual dynamics controlling user data custody. These solutions employ blockchain technology to establish an immutable and transparent ledger where user data is securely kept and can only be accessed or modified by authorized users. Furthermore, decentralized identity models aim to give consumers granular control over the transmission and utilization of their personal information.

However, as we navigate through this landscape of rising possibilities and transformational concepts, it becomes clear that, despite the actual anticipation surrounding user data decentralization, certain crucial routes remain relatively unexplored.

- Practical Implementation Assessment:

While the theoretical basis of user data decentralization has been thoroughly established, there is a notable gap in comprehensive studies evaluating the pragmatic implementation of these decentralized data management solutions within the intricate fabric of centralized social media platforms. The move from theoretical constructs to practical implementation requires close examination. This involves the complex technical aspects of integration, potential bottlenecks, and the expenses associated with them. Furthermore, it is critical to comprehend the adaptability and scalability of such solutions within the dynamic and resource-intensive context of social media platforms. It demands practical investigation into the challenges of integration, ensuring that the suggested decentralization techniques will function with current infrastructures in a smooth manner.

- User Acceptance Dynamics:

Another critical dimension that is still relatively unexplored is the user acceptability of decentralized data management concepts. Beyond technological complexities, it is critical to understand how end-users, who are the ultimate beneficiaries of these breakthroughs, perceive and interact with decentralized systems. It is critical to investigate user attitudes, interests, and concerns regarding the transition from centralized to decentralized data governance. Research investigations that assess user satisfaction, usability, and confidence in decentralized data systems can help determine the feasibility and durability of these solutions. Furthermore, recognizing the methods that can bridge the knowledge gap and drive user education about the benefits and complexities of decentralized data management is an important but under-addressed aspect.

- Data Security and Accessibility Implications:

The shift towards user data decentralization entails a number of implications for data security and accessibility. While the blockchain's cryptographic foundations provide reassurance, they also pose new issues, such as managing private keys and minimizing the risk of unauthorized access. In this subject, research should delve into the complex issues of data security within decentralized systems, investigating potential weaknesses and safeguards. Furthermore, ensuring equal and barrier-free data access is a fundamental goal. Investigating the effects of user data decentralization on data availability, particularly in areas with inadequate internet connectivity, is an area that deserves concentrated effort.

While the current landscape is filled with anticipation of user data decentralization, these study gaps highlight the complex and diverse nature of this adverse approach. Comprehensive, theoretical, and multidisciplinary research is required to realize the full potential of decentralized data management within centralized social media networks. Such research efforts will lead the way for a future in which individuals have greater control over their digital identities and the data they generate and share.

1.2.2 Peer-to-Peer Communication:

In the constantly shifting context of digital communication, the concept of peer-to-peer (P2P) communication has gained a lot of attention as a viable alternative to the current centralized messaging systems. The direct exchange of messages and information between users, as opposed to the intermediaries that characterize centralized platforms, is the distinguishing aspect of this paradigm. In this field, recent advancements have seen the emergence and communication of encrypted P2P messaging applications. Such applications have grown in popularity as a result of their continuous commitment to

protecting user privacy and improving communication security [2].

These encrypted peer-to-peer messaging apps are supported by cryptographic protocols that enable end-to-end encryption, protecting communication content from unauthorized access or interception. This strong encryption technique has emerged as a key selling point, assuaging customer concerns about data privacy and confidentiality. Encrypted P2P messaging apps have established their own niche for themselves in the digital communication landscape as users attempt to restore agency over their digital chats and wrap them in layers of security.

- User-Centric Exploration:

While previous research looks into the technical basis of P2P communication protocols and their cryptographic foundations, there is a noticeable gap when it comes to studies that take a user-centric approach. It is critical to clarify the user experience in the area of peer-to-peer communication, identifying the minute aspects that impact how people perceive and interact with these platforms. This includes research on the user-friendliness of P2P messaging applications, the simplicity of their interfaces, and general user satisfaction in comparison to centralized alternatives. Understanding user preferences and pain points is critical for improving P2P communication platform usability and acceptance.

- Adoption Challenges:

While peer-to-peer communication systems exhibit excellent potential, they also face a number of adoption problems that necessitate thorough research. It is critical to investigate the limitations and constraints that prevent the widespread adoption of P2P messaging applications. This includes variables like as network effects, which raise the value of the platform as the number of users increases, and interoperability between different communication platforms. Investigating methods that can help users migrate from centralized to P2P communication, as well as the associated problems, is an important aspect of future research.

- Scalability on a Larger Scale:

While encrypted P2P messaging apps have gained popularity, their scalability on a wider, social media platform scale remains a concern. It is quite a challenge to scale P2P communication systems to handle millions or even billions of users while keeping effective security and privacy features. Scalability research must analyze the technological aspects of scalability, studying how these systems may efficiently handle enormous user bases while maintaining essential fundamentals of P2P communication, such as privacy and security.

In conclusion, the emerging field of peer-to-peer communication, as represented by encrypted messaging applications, is characterized by both promise and complication. To fully realize the potential of peer-to-peer communication in the context of social media and digital interaction, the existing research gap must be bridged. A thorough understanding of the user experience, together with research into adoption issues and scalability issues, will pave the way for the seamless integration of P2P communication into the digital communication landscape.

1.2.3 Decentralized Data Integrity

Recent research efforts in the field of decentralized data integrity have put forth interesting solutions in the continuous effort to address the disruptive challenges of disinformation and fake news that exist within digital environments. These suggestions promote the use of novel methods, with blockchain-based solutions at the center. The primary idea of these solutions is to increase trust and validity in the information delivered across digital channels. [3].

Blockchain technology, known for its immutability and resistance to tampering, is at the core of these decentralized data integrity solutions. It provides an incorruptible ledger for securely recording and verifying data. The goal of wrapping the integrity of digital content in blockchain is to build a strong barrier against the spread of fraudulent information. The openness and traceability of blockchain technology allow users to determine the authenticity and sources of digital content.

- Real-World Efficiency:

While the theoretical foundations of decentralized data integrity systems, particularly those based on blockchain, are attractive, there is a significant research gap in determining their real-world usefulness. It is important to move from the theoretical world to research studies that look at the actual use of these systems. This needs further research into the influence of decentralized data integrity on content verification in the cluttered social media setting. Such investigations would determine whether these processes truly enable people to distinguish between trustworthy material and misinformation and fake news. Furthermore, the study should look into user trust dynamics, investigating how the deployment of these procedures affects users' belief in the accuracy of digital content.[19]

- Adversarial Manipulation:

Understanding the weaknesses and potential channels for opposed manipulation of decentralized data integrity methods is essential in the cat-and-mouse game between technology and adversaries. Potential flaws in these systems should be identified and mitigated through research. This involves looking into potential attack vectors that might compromise the integrity of data stored on blockchains. Examining these mechanisms' adaptation in the face of sophisticated attackers is critical for strengthening their resilience.

- Scalability Challenges:

Since centralized social media platforms frequently have massive user populations, the scalability of decentralized data integrity measures becomes crucial. While blockchain technology provides strong security and transparency, it has scalability constraints. The difficulties of achieving decentralized data integrity on networks with millions or even billions of users should be researched. Exploring new solutions and optimizations to ensure that these processes can handle the massive volume of digital content generated on centralized social media is a critical research subject.[20]

In conclusion, the pursuit of decentralized data integrity methods as a defense against false information and fake news signals a promising course in the digital era. Future research must bridge the existing gap by empirically measuring their effectiveness, reviewing vulnerabilities, and addressing scalability issues in order to fully realize their potential. This comprehensive strategy will be critical in ensuring the accuracy of digital information and the integrity of online conversation.

1.2.4 Decentralized Data Caching:

The context of research has seen an increase in pioneering approaches in the field of decentralized data caching in a constant effort to improve the efficiency of content delivery while simultaneously easing the demanding weight on centralized servers. These innovative concepts advocate for the reconfiguration of existing Content Delivery Networks (CDNs) through the use of decentralized nodes and distribution [4].

At the core of these efforts is the need to rethink the basic nature of content distribution. Decentralized data caching tries to establish a more resilient and efficient infrastructure by moving away from the traditional reliance on centralized servers, which can be subject to congestion and scalability concerns. This transition requires the distribution of content across a network of decentralized nodes actively placed to meet the increasing needs of consumers.[21]

- Practical Deployment in social media:

Decentralized data caching has strong theoretical foundations in the broader context of content delivery networks, but there is still a clear research gap about how to actually implement it in the dynamic environment of social media platforms. It is critical to move beyond theory and into actual studies that analyze the feasibility and implications of adopting these ideas. Importantly, this requires assessing how they actually affect important metrics like latency and content availability. It is critical to understand how these caching algorithms perform in the context of social media's dynamic and unpredictable content distribution landscape.

- Network Efficiency:

Network effectiveness is a problem that decentralized data caching has fundamentally. The total efficiency of data transmission is a topic that needs to be thoroughly researched in order to understand how these distributed caching networks affect it. This requires a careful analysis of elements including bandwidth usage, load balancing, and the capacity to quickly adjust to changes in user demand. Decentralized caching solution solutions would be improved due to the knowledge gathered from such research, which would help advance the general discussion on network optimization.

- Economic Feasibility and Incentive Structures:

The financial viability and incentive systems behind decentralized data caching are fascinating aspect that calls for further investigation. It frequently depends on individuals or companies voluntarily engaging in these networks for the move from centralized to decentralized caching. It is crucial to comprehend the incentives and motives behind such participation. A relatively unexplored area in the current study is examining the economic viability of these systems and their potential to develop sustainable models. [22]

In conclusion, the development of decentralized data caching indicates an innovation in the field of content distribution and holds up an opportunity of greater effectiveness and robustness. Experimental investigations to evaluate practical implementation, a focus on network efficiency, and a thorough investigation of the economic viability and incentives supporting these novel caching networks are all required to fill the research gap currently existing. Such research projects will be crucial in determining how the distribution of data will advance in the dynamic world of social media.

1.3 Research Problem

Decentralized social media systems can be made more secure, effective, and user-focused by addressing the problems with user data decentralization, peer-to-peer communication, decentralized data integrity, and decentralized data caching.

The main problem here is how to provide customers with total ownership and control over their data in a decentralized social media environment. In order to accomplish this, a strong framework that enables users to manage, access, and delete their data at will is required. This requires a thorough understanding of data collecting, processing, and storage procedures. A significant challenge is overcoming blockchain's limitations as a storage option for large amounts of data. It is essential to research innovative options to improve scalability and effectiveness while assuring secure data storage.

Peer-to-peer communication networks are currently facing issues that require considered solutions. The most efficient communication techniques must be identified in order to optimize communication inside peer-to-peer infrastructures. Another challenge with such networks is scaling them. It is important to find methods for growing peer-to-peer communication networks while resolving scalability problems. Furthermore, it is necessary to create a general-purpose peer-to-peer protocol that can accommodate different communication requirements. To enable smooth peer-to-peer communications, a flexible, open-source, and maintainable protocol must be developed.

Within decentralized social media protocols, maintaining the integrity of user data is a challenging problem. Protecting user data stored on devices against loss, corruption, or tampering is one of the most significant difficulties. It is essential to set up procedures for continuous information preservation and verification. Another challenge is deciding which approach is best for preserving data integrity in a decentralized system. Exploring alternative approaches while achieving the proper balance between security and speed is essential. Additionally, ensuring data integrity over multiple distant sites is essential and requires continuous validation in the context of IPFS (InterPlanetary File System).

Decentralized social media networks may benefit significantly from effective caching mechanisms. These technologies have an opportunity to improve response times, reduce network load, and increase scalability. However, efficiently developing caching mechanisms is necessary to avoid issues such as incorrect data, cache invalidation, storage limitations, security vulnerabilities, and scalability issues. Finding the best caching strategy for decentralized social media sites is essential. Although blockchain technology has been widely used, its limitations, such as resource consumption, limited storage capacity, slow transaction speeds, and expensive costs, must be solved. Improving the user experience requires developing a better caching technique that addresses these difficulties.

In conclusion, addressing each of these areas—user data decentralization, peer-to-peer communication, decentralized data integrity, and decentralized data caching—allows for the development of more secure, scalable, and user-centric decentralized social media platforms. By addressing these issues, we may develop environments in which users have more control over their data, efficient communication is possible, data integrity is ensured, and caching mechanisms improve platform reliability and performance.

1.4 Research Objectives

1.4.1 Main objective

The primary objective of this research is to develop a new decentralized social media protocol that provides users with a more secure, private, and decentralized alternative to standard social media platforms. The proposed protocol is intended to store user data on the user's device rather than on a centralized server, allowing the user complete control over their data.

To accomplish this goal, this research will require the creation of a user data decentralization protocol that specifies how user data will be stored, retrieved, and shared among network peers. This protocol ensures that user data is encrypted and protected while still remaining accessible and available to the user.

This research will include the design and implementation of a client application capable of successfully communicating with other network peers through a decentralized protocol, in addition to the development of the user data decentralization protocol. The client application will be developed to have an easy-to-use interface that allows users to simply create and share information with other network users. It also uses encryption, authentication, and verification techniques to protect the security and privacy of all user data and communications.

The overall goal of this research is to provide users with a decentralized social networking platform that gives them ultimate control over their data while still providing a smooth and user-friendly experience. The suggested platform provides a more secure, private, and decentralized alternative to standard social media platforms by building a user data decentralization protocol and a client application that uses this protocol.

1.4.2 Sub Objectives

- User Data Decentralization

The main goal approach is to examine how user data is stored. We are aiming to create an evolution toward user-centric data control with our research. Our technique includes a thorough evaluation of existing decentralization technologies, examining both their advantages and disadvantages. By doing so, we gain essential perspectives that help us make conclusions. We will identify areas for improvement in order to improve the protocol's operation, ensuring that user data remains securely in the hands of its owner while benefiting from the increased security of decentralized storage. Best practices will guide the implementation phase, ensuring compatibility, scalability, and security. The outcome is a solution that allows users to take control of their data, securely keep it on their devices, and overcome the limits that currently restrict decentralized social media networks. Furthermore, we prioritize high data availability through comprehensive backup processes, effectively integrating with today's digital landscape's increasing demand for improved information security and control.

- Peer-to-Peer Communication

This objective's fundamental goal is to provide a scalable peer-to-peer communication protocol for generic applications. We prioritize the creation of a strong RPC-style full duplex communication method that allows developers to specify customized data structures and functions, making remote invocation easier. The use of a binary data format allows for more efficient directional communication between clients and servers. To ensure agility and adaptability, our protocol uses an adaptive key-value store to maintain a full peer register. Furthermore, to improve file-sharing performance, we highlight the integration of third-party storage providers, which isolate file-sharing from the protocol's core and allow for the exchange of various data types. Our ultimate goal is to create a protocol that is not only effective but also adaptable, allowing developers to innovate and adjust it to their specific peer-to-peer communication needs.

- Decentralized Data Integrity

Consistency plays a major role in this component. We emphasize the necessity of standard data preparation for hashing and signing, which acts as an insurance policy against data tampering or illegal changes. Organizations can ensure that every change, no matter how minor, results in significantly different hash values or eliminates digital signatures by adhering to these guidelines. This consistency is essential in reducing risks associated with cryptographic processes. Furthermore, our technique encourages sharing information across several nodes in distributed systems to improve security, ensure fault tolerance, enable scalability, and balance loads. Verifying data authenticity with digital signatures and other cryptographic techniques increases trust in digital communications and transactions by ensuring that data remains unchanged and comes from a trustworthy source. Furthermore, we prioritize data integrity throughout the storage lifespan by applying techniques such as cryptographic hash functions and checksums. Regular comparison and redundancy techniques are used to detect and correct not wanted changes or manipulation, particularly in long-term storage or archiving systems.

- Decentralized Data Caching

The focus here is on improving performance and resource management within decentralized social media protocols. We recognize the importance of caching systems, as well as their various types and compromises. An important component is a thorough examination of the impacts of caching inside decentralized social media protocols. We explore how caching can improve speed while identifying and addressing possible problems. Storage management takes priority, particularly when dealing with user-generated material and multimedia assets. Integration with IPFS provides a once-in-a-lifetime opportunity to increase efficiency and reduce resource use. Our main goal is resource efficiency, and caching technologies when combined with other resource-saving measures, play an important part. Finally, we prioritize thorough documentation to ensure that developers and stakeholders in the decentralized social media ecosystem have easy access to best practices and practical implementation guidelines for the efficient use of caching methods in combination with IPFS.

2.0 METHODOLOGY

2.1 Requirement Gathering

In order to identify the problems with the current research and the modifications that are required, evaluation is performed. Reviewing previously published research articles on the subject generally provides the research objectives. In order to understand more about comparable performances, we also looked up a number of websites, publications, and industry experts. Defining and understanding the research problem as well as any past methodology, approaches, procedures, and algorithms is the main objective of requirement collection.

2.1.1 Past Paper Analysis

Data collecting will be supported by the analysis of research articles. Analyzing previously published research articles provides for a review of current solutions as well as the positive and negative aspects of earlier techniques. Analyzing prior research might help you choose the best research method. The research publications looked at peer-to-peer networks, data integrity, caching mechanisms, and other specific fields.

2.1.2 Refer Official Documentations

The official documentation is an excellent source for staying current on the technology we plan to use in the development of our system. Official documents, as compared to earlier research publications that can contain outdated information, offer reliable and current data on the most recent technological advancements, features, and best practices. It's developed and updated by specialists who are familiar with the technology within, offers thorough protection, how-to guides, and crucial security updates, making it an essential asset when developing a functional and current system.

2.1.3 Identify Existing Methodologies

This is especially valuable when reviewing existing systems because it allows for the identification of research requirements. It also improves an analysis of existing systems' limitations and weaknesses and their application to the newly built system. Existing systems can be available online, and our research will assist in the development of a high-quality solution without existing faults and limitations.

2.2 Feasibility Study

2.2.1 Technical feasibility

In the first stage of the project, a client application is developed with the main objective of allowing smooth interactions with network nodes and peers. The main way that users will communicate with the network will be through this client application. Additionally, it is going to emphasize security by putting reliable protections in place for storing user data securely on the user's local device storage. This data storage option enhances user privacy and control while protecting sensitive user data.

2.2.2 Schedule feasibility

The recommended solution must be executed in the seven months allocated with about two months aimed at gathering and assessing requirements and the remaining three months to system development. The last two months have been allocated to testing the system because testing is critical to the success of any project, and it may be further improved during this period to deliver higher quality results.

Because the research was conducted utilizing a software-based technique, a requirement analysis was utilized to produce a timeline as well as technical and economic feasibility. This allowed the interaction of research objectives, ensuring that neither technological capabilities nor planned time frames were exceeded.

2.2.3 Economic feasibility

During the initial project phase, there are no budgeted tasks as it centers around crucial requirement-gathering. This phase relies on in-depth analysis and reviews of past research and field investigations, making it cost-effective by skillfully utilizing open-source technology and tools without incurring direct expenses. However, as we transition into the development phase, costs become unavoidable. This stage necessitates the utilization of services from cloud providers, essential for building infrastructure and applications, offering scalability and flexibility. These expenditures will be the primary focus of the upcoming Economic Feasibility analysis in the thesis, ensuring a comprehensive evaluation of the project's financial implications.

2.3 Requirement Analysis

This phase played an essential role in this research, primarily by facilitating the identification of multiple elements that needed careful consideration during the project's implementation. It involved a thorough examination of data collected from various sources during the requirement gathering phase. As a result, this phase rapidly presented critical insights pertaining to potential challenges, the roadmap for research progression, insights into potential devices and technologies, and a comprehensive understanding of their utility.

Furthermore, this phase also allowed for the evaluation of the research's scope and feasibility. As the research journey continued, the requirement analysis became essential in recognizing current gaps in existing research and provided invaluable guidance for emphasizing the core research problem.

2.4 System Architecture

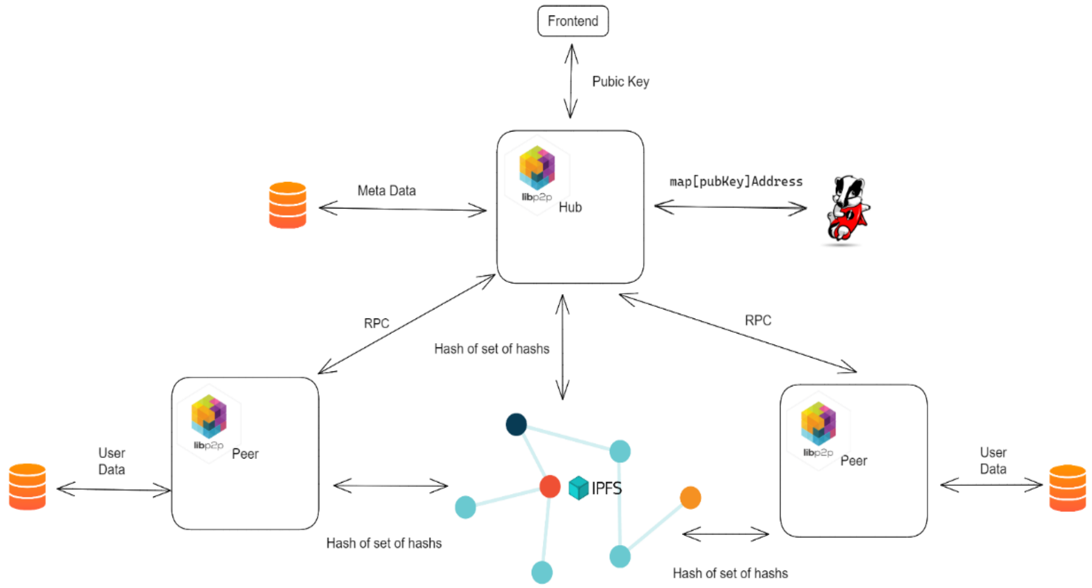


Figure 2.4.1: System Overview Diagram

The provided explanation outlines the components of the DeMedia architecture, which is a proposed infrastructure for a decentralized social media platform. The diagram accompanying the explanation provides a high-level overview of the architecture, which includes four major components: data decentralization protocol, peer-to-peer communication, decentralized data storage, and data integrity in a decentralized network. The first component of the DeMedia architecture is the data decentralization protocol, which will enable the platform to store user data within the users' devices rather than on centralized servers. This decentralization of data will enhance user control over their personal data, as well as increase data privacy and security.

The second component of the DeMedia architecture is peer-to-peer communication, which will allow users to communicate with each other through a hub which will act only as a communicator. This will increase the speed and efficiency of communication while reducing the dependence on centralized servers. 14 The third component of the DeMedia architecture is decentralized data storage, which will be accomplished using the InterPlanetary File System (IPFS). IPFS is a protocol that enables the creation of a decentralized file-sharing network, which is more secure and fault-tolerant than centralized file-sharing networks. The fourth and final component of the DeMedia architecture is data integrity in a decentralized network. This involves ensuring that the data stored on the decentralized network is secure, reliable, and tamper-proof. This is accomplished through the use of cryptography and other security measures to ensure that the data cannot be tampered with or compromised.

Overall, the DeMedia architecture aims to create a decentralized social media platform that offers increased privacy, security, and user control over personal data. The infrastructure will consist of hubs, peers, and a decentralized data storage network, which will be illustrated in the high-level architecture diagram

2.5 Commercialization

DeMedia will provide two paid models in addition to its free model: a membership-based subscription model and an advertising-based income decentralized social media platform.

In addition to the free model, DeMedia will also offer two paid models.

- Subscription-based membership model
- Advertising-based revenue model

The host can govern these paid models, allowing them to generate revenue from their platform. Therefore, one could describe DeMedia as a research project focused on commercializing the development of decentralized social media platforms. The project is developing a range of monetization models that hosts can use to generate revenue from their platforms, thereby enabling the commercialization of the technology.

To further elaborate on the commercialization aspect of DeMedia, it's essential to understand that the project is focused on creating technology that can enable the development of decentralized social media platforms. By providing a free base model, DeMedia is making it easier for individuals or organizations to create their own social media platforms that are not controlled by a centralized authority. However, to sustain and grow these platforms, there needs to be a way to generate revenue.

The subscription-based membership model allows hosts to charge users for access to premium features or content on their platform. This revenue can be used to cover the costs of hosting and maintaining the platform.

On the other hand, the advertising-based revenue model enables hosts to generate revenue by displaying ads on their platforms. Hosts can charge advertisers to display their ads on the platform, and this revenue can be used to cover the costs of hosting and maintaining the platform and generate profits.

As a summary of this proposed system, DeMedia is enabling the commercialization of decentralized social media platforms by providing a technology that allows anyone to create their own platform and monetization models that enable hosts to generate revenue and sustain their platforms. This could lead to a more diverse and decentralized social media ecosystem, with a more excellent range of platforms catering to specific sectors and communities.

2.7 Implementation

2.7.1 Background Establishment for Implementation

The team went through a careful planning and decision-making process on the way to putting the decentralized social media protocol into implementation. This phase was thought to be essential for laying a strong basis for the next development and testing phases. In-depth discussions of the decisions made regarding the project's programming language, database system, communication tools, cloud provider, and containerization technologies will be discussed in this section. Each of these choices was crucial in determining the project's ultimate outcome.

2.7.1.1 Programming Language Selection: Go Lang

After thorough consideration, the decision was made to utilize the Go programming language, commonly known as Go Lang. There were multiple solid factors that led to this decision.

Go Lang is popular for its efficiency, simplicity, and scalability, making it an ideal choice for building decentralized systems. Its concurrency support allows multiple operations to be executed concurrently, enabling the protocol to handle a large number of users and interactions efficiently. Moreover, Go Lang offers a strong standard library and a rich ecosystem of packages, significantly expediting the development process.

The inherent support for robust error handling was considered advantageous, as it enhances the protocol's resilience and fault tolerance. In essence, Go Lang not only met the technical requirements but also aligned with the goal of creating a stable and reliable decentralized social media platform.

2.7.1.2 Database System: PostgreSQL

The selection of an appropriate database system is critical in ensuring the efficient storage and retrieval of data in any project, including the decentralized social media protocol. PostgreSQL was chosen as the database management system, a decision influenced by several key considerations.

First and foremost, PostgreSQL is recognized as a robust open-source relational database system. Its support for complex data types and advanced indexing mechanisms makes it well-suited to handling the diverse data structures prevalent in social media applications. Additionally, strong data integrity and security features offered by PostgreSQL ensure the privacy and reliability of user data.

Furthermore, PostgreSQL's extensibility through user-defined functions and a vibrant community of contributors made it adaptable to evolving project needs. The database's ability to perform efficiently even under high loads was a crucial factor in the decision, given the typically rapid and unpredictable user activity on social media platforms.

2.7.1.3 Collaboration and Version Control: GitHub

Effective collaboration and version control are indispensable for a team working on a complex project like a decentralized social media protocol. To achieve this, GitHub was adopted as the primary platform for code collaboration and integration. The reasons behind this choice were straightforward yet compelling.

GitHub provides an intuitive and user-friendly interface, simplifying the processes of code sharing, reviewing, and merging. This streamlined the development workflow, ensuring that team members could collaborate seamlessly. Its version control capabilities allowed changes to be tracked, conflicts to be managed, and a comprehensive history of the codebase to be maintained, enhancing transparency and accountability.

Moreover, GitHub's robust issue tracking system enabled effective task management and prioritization, particularly valuable in a project of this scale, where numerous features and components needed to be developed and integrated.

2.7.1.4 Cloud Deployment: Amazon Web Services (AWS)

In the modern era of software development, cloud computing has emerged as a game-changer, offering unparalleled scalability, reliability, and flexibility. The benefits of deploying development and production environments in the cloud were recognized, and after careful evaluation, Amazon Web Services (AWS) was selected as the cloud provider.

AWS's vast array of services and global infrastructure ensured that the decentralized social media protocol could scale seamlessly to accommodate increasing user loads. The pay-as-you-go pricing model allowed costs to be optimized while benefiting from world-class infrastructure and support.

Furthermore, AWS offered a range of tools and services for tasks such as server provisioning, load balancing, and auto-scaling. These features simplified the management of cloud infrastructure, freeing up valuable time and resources that could be redirected toward enhancing the protocol's functionality and performance.

2.7.1.4.1 Deployment Using Amazon EC2 Instances

For deploying Docker containers and the PostgreSQL database, an Amazon Elastic Compute Cloud (Amazon EC2) instance was utilized. Amazon EC2 instances offered several advantages, including scalability, flexibility, and ease of configuration. Below are the specifications of the Amazon EC2 instance used for this deployment:

Type	t2.micro
Number of vCPUs	1
RAM (GiB)	1.0
Storage (GiB)	30
Operating System	Ubuntu

Table 2. 7.1.4.1.1: Specifications of EC2 instance

2.7.1.5 Containerization: Docker

To achieve consistency in deployment and streamline the process, Docker, a containerization technology, was leveraged. Docker provided several advantages that significantly facilitated the implementation.

Docker's fundamental benefit lies in its ability to encapsulate applications and their dependencies within lightweight containers. Each component of the decentralized social media protocol was integrated and packaged as a Docker image. This approach not only simplified deployment but also ensured that each component could run consistently across various environments.

One of Docker's key strengths is its isolation capabilities, allowing conflicts between different components of the system to be avoided. This isolation reduced the risk of compatibility issues and contributed to the overall stability of the protocol.

Additionally, Docker's portability made it possible to develop and test components independently before seamlessly integrating them into the larger system. This modular approach enhanced development agility and minimized disruptions during the implementation phase.

2.7.1.5.1 Deployment as Docker Containers

The decision to employ Docker containers for deployment resulted in the successful deployment of the implemented demo social platform. This approach offered several tangible benefits:

- **Consistency**: Docker containers ensured that the demo platform ran consistently across different environments, eliminating the notorious "it works on my machine" problem.
- **Scalability**: Docker's scalability features allowed adaptation to varying levels of user demand effortlessly. As user activity increased, the application could be scaled horizontally by adding more containers.
- **Resource Efficiency**: Docker containers, being lightweight and sharing the host OS kernel, resulted in minimal overhead. This translated to efficient resource utilization, reducing infrastructure costs.
- **Quick Deployment**: Docker's quick start-up time enabled the rapid deployment of new updates and features, minimizing downtime and user disruption.

In conclusion, the strategic choices made in setting up the background for implementation were instrumental in ensuring the success of implementation of the decentralized social media protocol. The adoption of Go Lang, PostgreSQL, GitHub, AWS, and Docker laid a strong foundation for the subsequent phases of development and testing. These decisions, rooted in practical considerations and a deep understanding of the requirements unique to the project, enabled the creation of a robust, scalable, and efficient platform that met the goals and expectations.

2.7.2 Setting Up CI/CD Pipeline

Due to the frequent releases during the development process, the team realized the necessity for a CI/CD (Continuous Integration/Continuous Deployment) pipeline. As a result, it was decided to create a reliable CI/CD pipeline that incorporated both CI and CD components. This section describes the pipeline's implementation, which was crucial in accelerating the development and deployment procedures.

2.7.2.1 Continuous Integration (CI) Pipeline

For the CI portion of the pipeline, GitHub Actions was chosen as the preferred tool. This decision was well-founded for several reasons.

GitHub Actions offers a seamless and integrated approach to automating our software development workflows. It allowed us to define, customize, and automate various tasks, such as code compilation, testing, and code quality checks, directly within our GitHub repository.

Moreover, GitHub Actions integrates seamlessly with our codebase hosted on GitHub, making it a natural choice for our CI needs. The ease of configuration and extensive library of pre-built actions simplified the setup of our CI workflow.

2.7.2.2 Continuous Deployment (CD) Pipeline

In parallel with the CI pipeline, a CD pipeline was established using GitHub Actions and Watchtower. Each was chosen for distinct but complementary reasons.

GitHub Actions was extended to serve as our CD tool, ensuring the seamless deployment of our application. GitHub Actions' continuous deployment capabilities allowed us to automate the deployment process, ensuring that every successful CI build was automatically deployed to our production environment. This reduced manual intervention and minimized deployment errors.

Watchtower, on the other hand, played a crucial role in automating container updates. It continuously monitored our Docker containers for new image versions and automatically updated running containers to the latest versions when changes were detected. This ensured that our application was always running with the most up-to-date code, enhancing security and reliability.

Direct integration to the GitHub repository and the user-friendly configuration of GitHub Actions made them the best option for both CI and CD. It eliminated the need for third-party tools, simplifying our development workflow and ensuring that code changes were rapidly and reliably integrated and deployed.

The watchtower’s selection was driven by its ability to automate container updates, reducing the need for manual intervention, and ensuring our application's consistent and secure operation.

The implementation of this CI/CD pipeline provided several benefits including:

- **Automation:** The CI/CD pipeline automated critical aspects of the development and deployment processes, reducing the manual effort required.
- **Speed and Efficiency:** Rapid and automated testing, integration, and deployment improved the overall speed and efficiency of the development cycle.
- **Consistency:** With CI/CD, every code change was subjected to the same automated testing and deployment process, ensuring consistency and reliability.
- **Reduced Errors:** Automation minimized the potential for human errors during deployment, enhancing the overall quality of our application.
- **Frequent Releases:** The CI/CD pipeline facilitated frequent and reliable releases, crucial for the development requiring rapid iterations and updates.

The establishment of the CI/CD pipeline using GitHub Actions and Watchtower significantly enhanced the development and deployment processes, aligning with the unique demands of the project. This automation not only improved efficiency but also contributed to the consistency and quality of our work, enabling to meet project milestones and deliver robust outcomes.

2.7.3 Development

The development phase of the project involved breaking down the entire project into four key components, with each component addressing a critical aspect of the decentralized social media system. Following sections briefly represents the development effort in each component:

Data Decentralization:

The first component, Data Decentralization, focused on creating a Data Decentralization Client. This client empowers users to securely store and retrieve data from their devices, reducing reliance on centralized servers. The development effort included key phases:

- **Initialization and Configuration**: The client initializes with the necessary environment variables, offering users the flexibility to choose between PostgreSQL and Elasticsearch as their preferred database systems.
- **Peer Initialization and Communication:** The client becomes a network peer for decentralized communication. A unique private key is generated for secure communication and network identification, a network port is selected for data exchange, and crucial communication channels with other peers are established.
- **Data Decentralization**: The core goal is to hand control of user data to users themselves. The client facilitates on-device storage and retrieval, eliminating the need for centralized servers. It integrates seamlessly with the user's chosen database and handles data storage and retrieval.
- **Integration and Compatibility:** The client smoothly connects to PostgreSQL or Elasticsearch, depending on the chosen database, ensuring efficient data storage and retrieval. It actively engages with other parts of the decentralized social media system, such as sharing messages and monitoring peers for network stability.

Peer-to-Peer Communication:

The second component, Peer-to-Peer Communication, focused on utilizing LibP2P, a versatile library renowned for its adaptability in P2P communication. The development effort included:

- **LibP2P Implementation:** LibP2P's capabilities were harnessed to build a robust P2P communication protocol for the decentralized social media network. The Go implementation, known for its reliability, was chosen.

- **Multiaddr for Peer Identification:** Peer identification was critical, and the multiaddr protocol was employed for this purpose, enhancing readability and peer discovery via DNS resolution.

- **Connection Establishment and RPC:** Establishing connections and implementing RPC clients using "go-libp2p-gorpc" ensured seamless communication, enabling request payload transmission.

- **Peer-Side Handling:** On the peer's side, RPC servers were initiated, registering required RPC methods. The RPC server was then launched to respond to incoming queries effectively.

Decentralized Data Integrity:

The third component, Decentralized Data Integrity, tackled the critical task of enhancing the security and reliability of user data. Key development aspects included:

- **Data Integrity through Signature Generation and Verification:** A process was implemented where user data is signed before storage and verified upon retrieval. This involved generating unique digital signatures and checking them for data authenticity.
- **Data Security through Hashing**: Data security was paramount, so hashing was employed to convert data into unique digital fingerprints, ensuring data remained secure and unaltered.
- **Consistency in Event Representation:** To maintain data consistency across user devices, event representation was standardized, ensuring uniformity, and enhancing overall data integrity within the decentralized social media protocol.

Decentralized Data Caching:

The final component, Decentralized Data Caching, aimed to optimize network performance and address challenges associated with large media object storage. Key development efforts included:

- **IPFS Integration for Media Object Caching:** To begin with the storage of large media objects, the InterPlanetary File System (IPFS) was integrated to store media objects off-chain, relieving the blockchain network.
- **Data Caching for Network Performance:** Implementing a data caching mechanism

on IPFS allowed for quicker media object retrieval, significantly improving network performance.

- **Cost-Effective Large Object Handling:** Offloading large media object storage to IPFS and involving users in maintaining the decentralized storage infrastructure made the solution cost-effective and sustainable.
- **Network Performance Enhancement**: The caching mechanism significantly enhanced network performance, reducing latency, and accommodating a larger user base by minimizing direct blockchain interaction for media retrieval.

The development phase was a collaborative effort that developed four key components. These components together built a protocol that provide users more control, improve security, and boost network performance in decentralized social media platform.

2.8 Testing

In terms of testing, the approach used was intended to guarantee the accuracy and dependability of the research project. Each development effort was verified through a methodical testing process before it was integrated into the project's codebase hosted on GitHub and after the deployment.

2.8.1 Unit Testing

Unit tests played an important role in the testing process. These tests were conducted before the code was pushed to GitHub, serving as a initial validation step. Unit tests are small, focused tests that verify the correctness of individual components or functions within the code.

Through the conduct of unit tests, potential issues were detected and rectified at an early stage, preventing them from propagating into the codebase. This approach ensured that each development effort underwent validation for correctness and functionality, contributing to the overall stability of the project.

2.8.2 Continuous Integration (CI) and Deployment (CD)

The testing process was tightly integrated with the CI/CD pipeline, streamlining the testing and deployment of developed features. Here's how it worked:

1. **Continuous Integration (CI):** Upon completion of development efforts, the CI pipeline automatically integrated these features. It compiled the code, ran unit tests, and ensured that the new code did not introduce regressions or errors.
2. **Docker Image Building:** Following successful CI, the pipeline proceeded to build Docker images. Docker images are like packaged containers that encapsulate the application and its dependencies.
3. **Continuous Deployment (CD):** A crucial aspect of the testing process was the CD pipeline. This pipeline utilized Watchtower, an automated container updating tool. When new Docker images were created, Watchtower was triggered as part of the CD process.

4. **Watchtower Deployment:** Watchtower's role was to pull the latest Docker images with updated code and deploy them on Amazon Elastic Compute Cloud (Amazon EC2) instances. This automated process ensured that the application consistently ran the latest code changes.

2.8.3 Smoke Testing

After each deployment, a critical step was the execution of smoke tests. Smoke tests are a set of initial tests designed to verify that the newly deployed release is stable and functional. They are meant to ensure that the release is "smoke-free," indicating that it is ready for further testing and use.

The testing approach offered several notable benefits to the project:

- **Early Issue Detection**: Unit tests allowed for the early detection and resolution of issues in the development process, reducing the likelihood of critical errors in the final product.
- **Quality Assurance**: Through thorough unit tests and continuous integration, a high level of code quality and reliability was maintained.
- **Efficiency**: Automation within the CI/CD pipeline reduced manual testing efforts, enabling faster development cycles and quicker deployment of new features.
- **Consistency:** The automated deployment process with Watchtower ensured that the application consistently operated with the latest code changes, enhancing overall reliability.
- **Rapid Updates**: The testing and CD approach facilitated swift updates and deployments, essential for an academic research project requiring frequent iterations and enhancements.

In conclusion, the testing approach, included unit testing, seamless integration with the CI/CD pipeline, and smoke testing, was important in maintaining the quality, reliability, and efficiency of the project. It enabled the validation of each development effort, early error identification, and the delivery of a robust and continuously evolving research platform.

2.8.4 Validation using practical implementations.

In order to assess the practical implementations of the project, the development of a demo social media platform and a separate benchmarking application was initiated. These efforts allowed valuable insights into the performance and functionality of the built protocol.

2.8.4.1 Benchmarking with Infrastructure as Code (IaaC) Approach

To deploy the benchmarking application, along with the required infrastructure and dependent applications, an Infrastructure as Code (IaaC) approach was followed. This approach brought significant advantages:

- **Reproducibility**: IaaC allowed the definition and recreation of the entire infrastructure consistently, minimizing discrepancies between deployments.
- **Version Control**: Infrastructure configurations were version-controlled, enabling the tracking of changes, effective collaboration, and maintenance of a comprehensive history.
- **Scalability**: With IaaC, easy scaling of the infrastructure up or down in response to varying workloads was possible, ensuring optimal performance.

2.8.4.1.1 Terraform as the IaaC Tool

For implementing the IaaC approach, terraform was selected as the tool of choice. Terraform provided numerous benefits, including:

- **Infrastructure ambiguity**: Terraform supports multiple cloud providers and infrastructure types, giving flexibility in choosing the best-suited resources.
- **Declarative Syntax**: Terraform' s declarative syntax made it easy to define and manage infrastructure configurations, enhancing readability and maintainability.
- **Modularity**: Terraform allowed the creation of reusable modules, simplifying the deployment of complex infrastructure components.

Along with these 2 practical implementations, testing is carried out to validate the functionality, usability, and reliability of the protocol.

3.0 RESULTS AND DISCUSSIONS

3.1 Results

The implementation of both the demo social media platform and the benchmark application played an important role in validating various aspects of the project. Here, present the results obtained from these implementations, highlighting the project's key achievements and insights.

3.1.1 Demo Social Media Platform Results

The development of the demo social media platform on top of the implemented decentralized social media protocol produced below noticeable results:

- **User Engagement**: The platform successfully facilitated user engagement, demonstrating the protocol's effectiveness in creating a user-friendly social media environment.
- **Feature Integration**: Various features, such as user profiles, posts, and interactions, were seamlessly integrated and validated, showcasing the versatility of the protocol.
- **User Experience**: Users were provided a positive and intuitive experience while navigating and interacting with the demo social media platform, highlighting its user-centric design.
- **Stability**: The platform established stability and robustness, with minimal downtime or disruptions during usage, ensuring a reliable user experience.

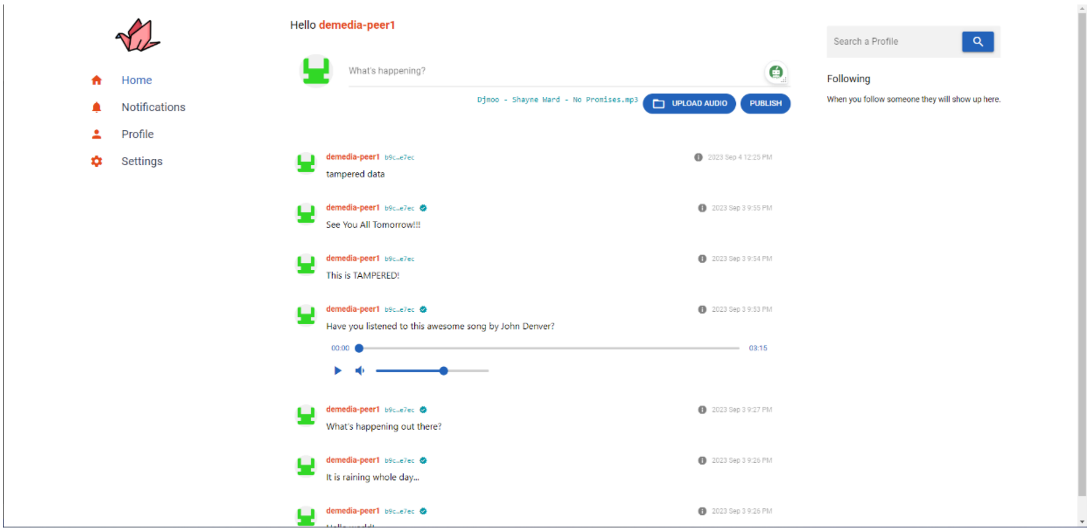


Figure 3.1.2.1: Screen capture of demo social media platform

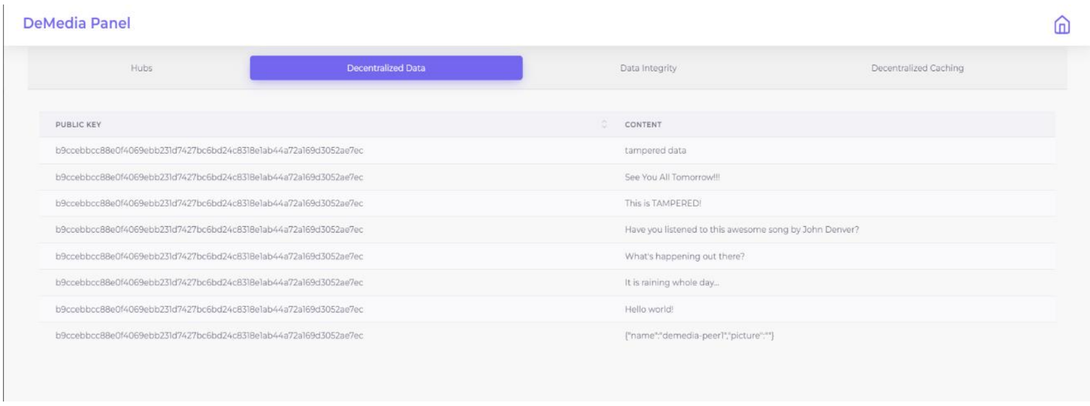


Figure 3.1.2.2: The panel of Decentralized Data

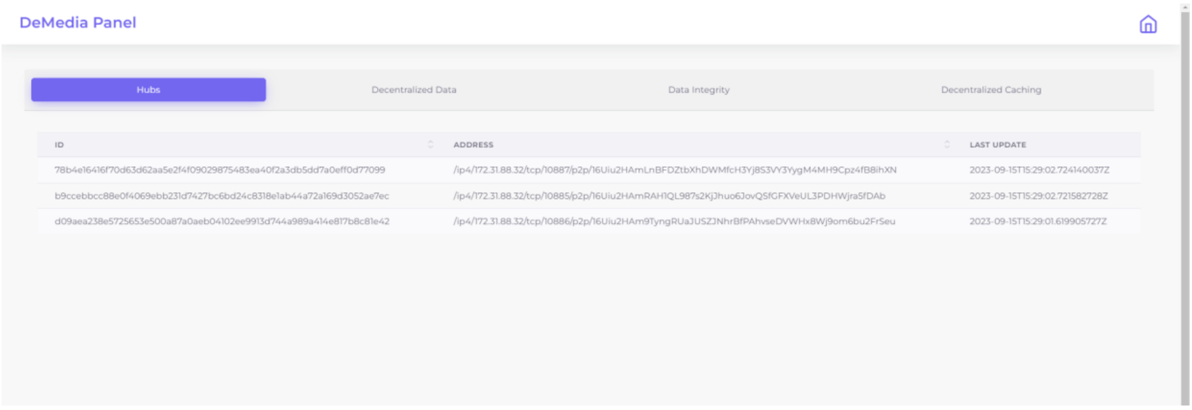


Figure 3.1.2.3: The panel of peer-to-peer communication

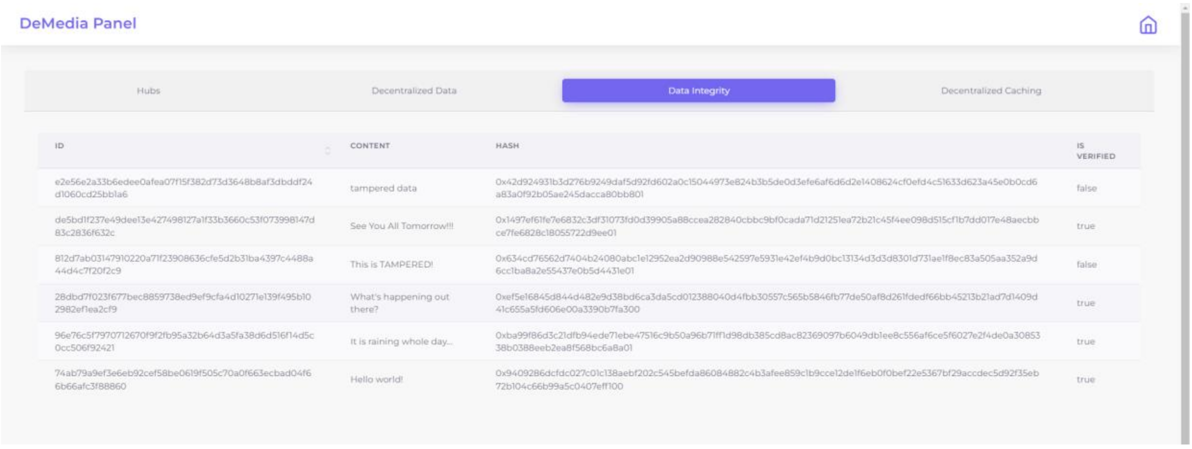


Figure 3.1.2.4: The panel of Decentralized Data Integrity

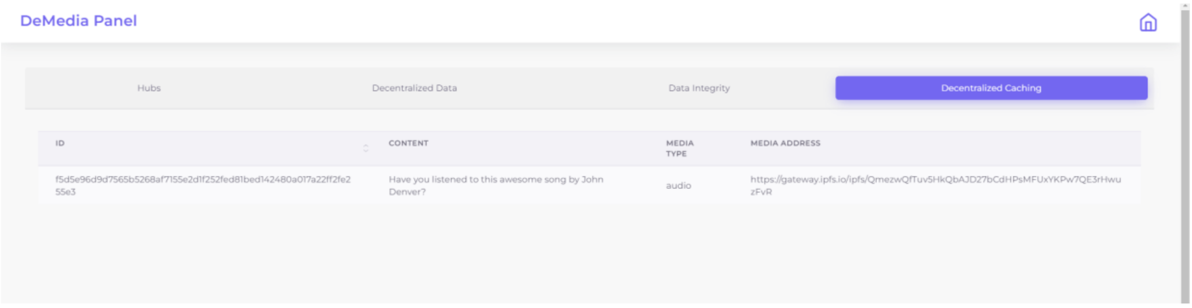


Figure 3.1.2.5: The panel of Decentralized Data Caching

3.1.2 Benchmarking Results

The benchmarking application provided insightful results on the performance and capabilities of the built protocol. These results, which are detailed below, are integral to the project's evaluation and future improvements:

```
+---------------------------------+--------------+---------------+
|             METHOD              |     REST     | IPFS/DEMEDIA  |
+---------------------------------+--------------+---------------+
| REST API Call vs DeMedia API    | 182.883684ms | 432.263605ms  |
| Call                            |              |               |
| Direct DB Fetch vs DB Fetch     | 3.353261ms   | 432.263605ms  |
| Through DeMedia                 |              |               |
| Direct DB Fetch vs REST API     | 3.353261ms   | 182.883684ms  |
| Call                            |              |               |
| Time Elapsed To Fetch Image 01  | 36.62113ms   | 23.011138ms   |
| Time Elapsed To Fetch Image 02  | 4.819432ms   | 9.295224ms    |
+---------------------------------+--------------+---------------+
```

Figure 3.1.2.1: Output from benchmark application

3.1.3 Results from overall practical implementations

The development of the demo social media platform and the benchmark application allowed for the validation of several critical aspects of the project, including:

- **Functionality**: Through rigorous testing and usage of the demo social media platform, validated the core functionality of the implemented decentralized social media protocol.
- **Scalability**: By deploying the demo platform and benchmark application at scale, assessed the protocol's ability to handle a substantial user load and interactions effectively.
- **Security**: Security features and measures were thoroughly examined and validated through the development of the demo platform, ensuring the protection of user data and interactions.
- **Performance**: Performance metrics were gathered and analyzed to evaluate the speed and efficiency of the decentralized social media protocol in real-world scenarios.

3.2 Research Findings

Recent research on decentralized social media protocols has revealed the significant advantages of peer-to-peer (P2P) communication in enhancing user privacy, reducing data latency, and increasing resistance to censorship. P2P systems distribute data between nodes, reducing vulnerability to single points of failure and censorship efforts. However, challenges like scalability and maintaining a consistent user experience persist as the user base grows.

While P2P communication may not match the performance of traditional REST APIs due to the nature of its design, its focus on decentralized communication aligns with the goals of data governance, privacy, and data commercialization. To address issues related to data availability, scalability, security, portability, and user acceptance, innovative approaches such as data sharing, incentive systems, learning federation, blockchain integration, and user-friendly interfaces are being explored. These advancements aim to maximize the positive aspects of data decentralization while ensuring high data availability and user-controlled data access.

One key aspect of decentralized social media protocols is data integrity. The decentralized structure readily identifies unauthorized data modifications, enhancing user trust. By distributing data across multiple nodes and implementing cryptographic signatures, the protocol minimizes single points of failure and provides users with effective means to verify data authenticity and integrity. As the network expands, the protocol's ability to adapt to growing demands while maintaining data integrity and ensuring smooth performance becomes crucial for long-term success.

Additionally, research into caching mechanisms within these protocols has shown that they can significantly reduce data retrieval times, improving user experiences and reducing latency. Caching mechanisms, particularly when integrated with IPFS, optimize storage by reducing redundancy and offer efficient content updates and retrieval in a decentralized environment. This research not only enhances storage management and resource efficiency but also provides valuable documentation for others seeking to implement similar solutions, ultimately contributing to the improvement of data access and storage in decentralized social media platforms. In summary, the combination of P2P communication, data decentralization, data integrity, and caching mechanisms presents a promising path toward more secure, private, and efficient decentralized social media protocols.

3.3 Discussion

This section presents the outcomes of the research and the practical application of the protocol, followed by a detailed discussion highlighting the findings' significance and their implications for the future of social media architecture.

Protocol effectiveness and usability assessment

The development of a demo social media platform provided a practical framework to evaluate the effectiveness and usability of the DeMedia protocol. Through the deployment and testing of this platform, the protocol's features and mechanisms were put to the test in a real-world setting. Users engaged with the platform interacted with peers, and shared content, thereby simulating the actual usage scenario. This usability assessment served as a validation of the protocol's design principles and its applicability to real-life social media interactions.

Panel for Hub and Peer Data Representation

An admin panel was created within the demo platform to showcase the inner workings of the protocol. This panel visually represented the decentralized communication architecture, demonstrating how hubs serve as intermediaries between peers and facilitate content sharing. Users could view hub details and the peer connections established through the hub. This visual representation enhanced user understanding of the protocol's distributed nature and emphasized the absence of centralized control.

Data Integrity Verification

The protocol's commitment to data integrity was demonstrated through a verification mechanism built into the demo platform. Hubs can verify the integrity of the data stored on user devices by checking the cryptographic signatures associated with users' content, generated when the content is stored on the device and published to the network. This verification process provided the platform users with a tangible assurance of shared data's authenticity and unaltered nature. This feature provided users control over their data and demonstrated the cryptographic techniques that enable the protocol's data integrity strategies.

IPFS Media Object Caching

Integrating the InterPlanetary File System (IPFS) for media object storage was an important architectural aspect of the protocol. In the demo platform, media objects were cached on IPFS, allowing users to access and share content efficiently. This caching mechanism showcased the protocol's ability to leverage decentralized data storage while ensuring seamless content retrieval. The media object caching capability shows the protocol's scalability and potential to reduce demand for network resources and control the cache handed over to the user.

Benchmark Tests and Performance Evaluation

By conducting thorough benchmark testing and rigorous performance evaluation, significant insights have been obtained regarding the operational efficiency of the DeMedia protocol. It was observed that certain aspects of DeMedia's performance were marginally below initial expectations, warranting attention for further optimization in pursuing enhanced network efficiency.

In the benchmark testing, DeMedia API calls and traditional REST API calls were compared. Additionally, the performance of data retrieval was examined, comparing the speed of fetching data directly from the DeMedia database to traditional database retrieval methods. The outcomes showed that although DeMedia performed effectively, there was a deviation from the expected efficiency expectations.

This variation indicates a significant field for future research and development. Optimization of network efficiency remains a critical path for DeMedia to reach its full potential in real-world applications. To address this issue, the protocol's complexities will be refined, procedures will be streamlined, and network interactions will be fine-tuned to achieve adequate performance and responsiveness.

The benchmark test results, together with the practical development and testing of the demo social media platform, highlight the transformational potential of the DeMedia decentralized social media protocol. These findings emphasize its viability and usability in real-world scenarios. At the same time, the observations gained through benchmark testing and performance evaluation provide a degree of realism to the assessment of operational efficiency.

The usability assessment demonstrated through the demo platform validated the protocol's design principles and emphasized its applicability in authentic social networking contexts. The integration of the hub panel and data integrity verification mechanisms not only facilitated the transparency but also highlighted the protocol's commitment to enabling user control and ensuring data authenticity.
Integrating the InterPlanetary File System (IPFS) for media object caching effectively addressed the data storage and retrieval challenges. It provided the scalability and efficiency of protocol in storing media objects.

4. 0 SUMMARY OF EACH STUDENT'S CONTRIBUTION

Student Number	Name	Tasks
IT29254698	Perera B.S.S.	Implement RPC-style full duplex peer-to-peer communication protocol. Implement a key-value store to keep track of peers. Document the protocol. Test the protocol.
IT2159726	Bandara A. M. C. A.	Conduct a comprehensive review of user data decentralization protocols and associated research. Assess the opportunities and challenges present in existing user data decentralization protocols. Determine the potential enhancements and new capabilities that could be integrated into a mechanism for data caching in a peer-to-peer network. Determine the potential enhancements and new capabilities that could be integrated into a protocol for storing user data within the

		user's device. Implement a protocol with the identified modifications while adhering to best practices.
IT20137496	Dhananjani G.G.S.	Conducted a comprehensive review of data integrity preservation mechanism and associated research. Developed a cryptographic mechanism to save data in the user's device. Generated hash value for user data using the hashing algorithm. Signed hash values using digital signature and encoded it. Maintain data integrity of stored set of signed values in IPFS network.

IT20157432	Abeykoon A.W.Y.I.K.	Conduct a comprehensive review of data caching mechanisms and associated research. Assess the opportunities and challenges present in existing data caching mechanisms. Determine the potential enhancements and new capabilities that could be integrated into a mechanism for data caching in a peer-to-peer network. Implement a mechanism to cache data in a peer-to-peer network with the identified modifications while adhering to best practices.

Table 4.1: Student's contribution

5.0 CONCLUSION

In conclusion, this research introduces the DeMedia protocol, a decentralized social network protocol that addresses the privacy and security concerns inherent to centralized social media platforms. The protocol's innovative approach enables users to retain full ownership of their data by storing it locally on their devices, mitigating vulnerabilities associated with centralized servers. Employing peer-to-peer networking and Cryptographic techniques ensure data privacy and integrity.

By leveraging decentralized technologies like peer-to-peer networking and IPFS, the DeMedia protocol offers users a more secure and private social media experience. This approach empowers users with data ownership and control while negating the necessity for costly blockchain infrastructure. The protocol also features tools for optimizing data storage and maintaining data integrity, such as data cache and integrity preservation. Ultimately, the proposed protocol establishes a secure and decentralized social media network that grants users autonomy and facilitates direct interactions, freeing them from reliance on centralized intermediaries. This advancement holds substantial potential for revolutionizing the landscape of social media architecture.

REFERENCES

[1] d. m. boyd and N. B. Ellison, "Journal of Computer-Mediated Communication," vol. 13, no. 1, pp. 210-230, 2007.

[2] S. J. Dixon, "Number of worldwide social network users 2027," Statista, 29 August 2023. [Online]. Available: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/. [Accessed 13 September 2023].

[3] Z. Tufekci, "Big questions for social media big data: representativeness, validity and other methodological pitfalls," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 505-514, 2014.

[4] S. Vosoughi, D. Roy and S. Aral, "The spread of true and false news online," vol. 359, no. 6380, pp. 1146-1151, 2018.

[5] T. Gillespie, "Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media," in *Yale University Press*, 2018.

[6] E. Hafermalz, "Book review – the age of surveillance capitalism: The fight for a human future at the New Frontier of Power," *M@n@gement*, 2021.

[7] W. t. t. r. t. WEB3, "Alchemy Docs," [Online]. Available: https://docs.alchemy.com/docs/welcome-to-the-road-to-web3. [Accessed 10 Septermber 2023].

[8] S. B, A. Sh, S. K. E, S. N. K and N. S, "Blockchain Industry 5.0: Next generation smart contract and Decentralized Application Platform," *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, 2022.

[9] "Decentralized social media," Mastodon, [Online]. Available: https://joinmastodon.org/#how-does-mastodon-work. [Accessed 13 September 2023].

[10] "What is Sovrin?," Sovrin, 14 December 2018. [Online]. Available: https://sovrin.org/faq/what-is-sovrin-2/. [Accessed 13 September 2023].

[11] "How IPFs works," IPFS Docs, [Online]. Available:

https://docs.ipfs.io/concepts/how-ipfs-works/. [Accessed 13 September 2023].

[12] D. Garg, "The rise of decentralized social media platforms," UX Planet, 02 January 2023. [Online]. Available: https://uxplanet.org/the-rise-of-decentralized-social-media-platforms-b7e5025e9186. [Accessed 10 September 2023].

[13] ukg, "7 major problems of existing centralized social media platforms and *how-doo* address them," Steemit, [Online]. Available: https://steemit.com/socialmedia/@ukg/7-major-problems-of-existing-centralized-social-media-platforms-and-how-doo-address-them. [Accessed 10 September 2023].

[14] N. Upadhyay, "Blockchain-based social media enterprise," *Transforming Social Media Business Models Through Blockchain*, pp. 55-75, 2019.

[15] J. Benet, "IPFS - Content Addressed, Versioned, P2P File System," vol. 1407, no. 3561, 2014.

[16] "Decentralized social media," Mastodon, [Online]. Available: https://joinmastodon.org/#about. [Accessed 13 September 2023].

[17] A. Narayanan, J. Bonneau, E. Felten, A. Miller and S. Goldfeder, "Bitcoin and cryptocurrency technologies," no. 8, p. 4, 2016.

[18] "Empowering users: The rise of decentralized social media platforms on Blockchain," [Online]. Available: https://www.financemagnates.com/cryptocurrency/education-centre/empowering-users-the-rise-of-decentralized-social-media-platforms-on-blockchain/. [Accessed 10 September 2023].

[19] K. He, C. Huang, J. Shi, X. Hu and X. Fan, "Enabling decentralized and dynamic data integrity verification for secure cloud storage via T-Merkle Hash Tree based blockchain," Hindawi, 01 November 2021. [Online]. Available: https://www.hindawi.com/journals/misy/2021/9977744/. [Accessed 16 September 2023].

[20] S. Hirsh and S. W. Alman, "What is Decentralization in Blockchain?," ALA Neal-Schuman, 2020. [Online]. Available:

https://aws.amazon.com/blockchain/decentralization-in-blockchain/. [Accessed 10 September 2023].

[21] R. T. a. T. C. C. A. Sengupta, "Decentralized caching with secure delivery," in *2014 IEEE International Symposium on Information Theory*, Honolulu, HI, USA, 2014.

[22] R. Han, Z. Yan, X. Liang and L. Yang, "How can incentive mechanisms and blockchain benefit with each other? A survey," *ACM Computing Surveys*, vol. 55, no. 7, pp. 1-38, 2022.

GLOSARY

- Blockchain - A decentralized, distributed digital ledger that records transactions across multiple devices and networks.
- Decentralized - Something is not controlled or managed by a single central authority or entity.
- Hub - A central point or place where things come together or connect.
- Peer - Entities that have equal or similar within a group or network.
- Peer-to-peer (P2P) caching - a decentralized caching mechanism in which nodes in a network cache data for each other, reducing the load on the network and improving performance.
- Distributed caching: a caching mechanism in which data is stored across multiple nodes in a decentralized network.
- Content-addressable storage: a mechanism in which data is stored based on its content rather than its location, allowing for efficient retrieval and distribution.
- Content Delivery Network (CDN): a distributed caching system that stores and delivers content to end-users based on their geographic location, reducing latency, and improving performance.
- Data Privacy: The protection of personal data from unauthorized access, ensuring that only authorized users can access and make use of the information.
- Access Control: The system gives consumers complete control over data sharing by allowing them to specify who can view and interact with their data.
- Data Integrity: The assurance that user data is protected and kept private within the decentralized network, is frequently accomplished through cryptographic methods.

APPENDICES

DEMEDIA – DECENTRALIZED SOCIAL MEDIA PROTOCOL - group

ORIGINALITY REPORT

3%	1%	2%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Partha Pratim Ray. "Web3: A comprehensive review on background, technologies, applications, zero-trust architectures, challenges and future directions", Internet of Things and Cyber-Physical Systems, 2023 Publication	1%
2	5dok.net Internet Source	<1%
3	Submitted to University of Exeter Student Paper	<1%
4	roboticsandautomationnews.com Internet Source	<1%
5	Submitted to Coventry University Student Paper	<1%
6	Submitted to National School of Business Management NSBM, Sri Lanka Student Paper	<1%
7	Submitted to Symbiosis International University	<1%

Appendices -A: Plagiarism Report