**Enhancing File Sharing and Data Management Capabilities at Batangas State University Alangilan Faculty Through a Distributed File System**

RESEARCH PAPER PRESENTED

TO THE FACULTY OF THE COLLEGE OF ENGINEERING

AT ALANGILAN, BATANGAS CITY

IN FULFILLMENT OF THE REQUIREMENTS OF

CpE 422: CpE Practice and Design 1

**By:**

**ABAD, JEMILYN L.**

**OCTA, AR-JAY E.**

**VASQUEZ, JILBERT S.**

**TABLE OF CONTENTS**

[**CHAPTER I 2**](#_ecnrnvz6y34c)

[**INTRODUCTION 2**](#_kx5bj49dwwjl)

[1.1. BACKGROUND OF THE STUDY 2](#_fit79nxvkfoq)

[1.2. STATEMENT OF THE PROBLEM 3](#_lveuw34nqjzl)

[1.3. RESEARCH QUESTIONS 4](#_jea94yqk3gz8)

[1.4. RESEARCH OBJECTIVES 6](#_mqhr9j2erw6i)

[1.5. RESEARCH HYPOTHESIS 7](#_64op9ttqocgt)

[1.6 SCOPE AND DELIMITATIONS OF THE STUDY 7](#_1rttlomx8j5r)

[1.7. IMPORTANCE OF THE STUDY 8](#_5h44de1eeqsj)

[**CHAPTER 2 9**](#_kjuz6d2onivu)

[**LITERATURE REVIEW 9**](#_u05y43mj4r4)

[2.1. CONCEPTUAL LITERATURE 9](#_kj5kqd8a9bxo)

[2.2 RELATED LITERATURE 16](#_5sq9t79hv96n)

[2.3. SYNTHESIS 21](#_5g2rcyl0xgru)

[**CHAPTER 3 25**](#_5d6mnng1z8u8)

[**DESIGN AND METHODS 25**](#_k6b7u02a34g4)

[3.1. CONCEPTUAL FRAMEWORK 25](#_7u7e39ry0qhx)

[3.2. TECHNICAL DESIGN AND PROCEDURE 26](#_llk31ksqdxeg)

[3.2.1. Distributed Hash Table (DHT)-based Distributed File System 27](#_arpeu9syxxa8)

[3.2.2. Interconnected Distributed File System inspired by BitTorrent Protocol 29](#_iu4ho4cohvx3)

[3.2.3. Distributed Erasure-Coded File System 32](#_9t4ovxhlm2th)

[3.3. EQUIPMENT/FACILITIES/PROGRAMS 34](#_droeamot7lo1)

[3.4. EXPECTED OUTPUT 35](#_7ov7kkye89xd)

[3.5. GANTT CHART 37](#_kc98sx12yq1o)

[**REFERENCES: 40**](#_o0j6n7caoe8y)

# 

# CHAPTER I

# INTRODUCTION

## 1.1. BACKGROUND OF THE STUDY

Managing teaching, research, and administrative tasks efficiently is one of the problems facing 21st-century higher education institutions. According to a statement in Research data management services in academic libraries: a comparative study of South Asia and Southeast Asia (Subaveerapandiyan A, 2023) asserts that contemporary information technologies offer ways to improve school administration and facilitate decision-making. They concentrate on two of these technologies in this paper: data mining and data warehousing. They define data warehouses and data mining, discuss how to utilize them in classroom management in higher education, and outline the advantages of doing so. Based on the Anthropological Data in the Digital Age (Reily and Thompson, 2020) the researchers provide a thorough manual on responsible data management, emphasizing the need for standards, openness, and researcher communication. The writers address a variety of topics related to data management, including study design, data gathering, analysis, interpretation, and dissemination. They place a strong emphasis on data integrity since reliable and high-quality data are critical to the advancement of science, the formulation of public policy, and the maintenance of professional reputations. In light of technical advancements and legislative changes, the study explores ethical questions, intellectual property rights, and evolving data management techniques.

Managing and sharing the vast number of files produced by instructors, staff, and students is one of the issues that contemporary colleges must deal with. There are restrictions on storage capacity, security, and accessibility with traditional means like email attachments, USB drives, and local servers. Some institutions have implemented web-based file sharing solutions to get around these problems, enabling users to upload, download, and modify files online through a web browser. Numerous benefits are provided by these systems, including synchronization, backup, encryption, and cloud storage. Kabale institution in Uganda is one instance of an institution that has put in place a web-based file sharing system. Based on the research Web-Based File Sharing System: A Case Study Of Kabale University (Akandinda, 2022) the university created and implemented KU-FileShare, an open-source software system built on Nextcloud. One hundred users from various university departments and campuses tested the system; they could access their files from any location and on any device. According to the research, the system increased user happiness and trust in the university's IT services as well as their effectiveness, productivity, and teamwork. Along with identifying some obstacles, the study included suggestions for how to strengthen the system going forward, including user education, bandwidth optimization, and platform integration.

The Philippines' poor infrastructure for file sharing and data management is one of the problems. According to the research, Data Sharing at Bestlink College of the Philippines (Macula et al., 2020) Many institutions do not have the bandwidth, storage space, or technical assistance to allow faculty members to share data in an effective and dependable manner . This implies that teachers might not be able to communicate huge or sensitive data sets with their peers or pupils, or they might have sluggish or intermittent access to crucial files. The quality and security of the data may also be impacted by a lack of infrastructure since it may be more vulnerable to loss, corruption, or illegal access. The security threats connected to file sharing and data management in the Philippines provide another difficulty. Many faculty members are not aware of the best practices and procedures for safeguarding their data against malware, phishing, and cyberattacks (Macula et al., 2020). Faculty members might also store or exchange their data on unprotected devices or platforms, such flash drives, cloud services, or personal email accounts. Due to these activities, their data may be compromised, stolen, or misused by rival companies, hackers, or other nefarious parties. Additionally, there can be ethical and legal ramifications for the security issues, particularly if the data contains private or sensitive information about students, employees, or research participants.

The purpose of this study is to look into whether Batangas State University's Alangilan Faculty could use Microsoft DFS. The purpose of the study is to ascertain whether Microsoft DFS can improve data sharing and management capacities among faculty members by examining the present file sharing issues and weighing the advantages of the technology. Improved faculty productivity, expedited research procedures, and an all-around more effective workflow for BatStateU Alangilan Faculty can result from the effective deployment of DFS.

## 1.2. STATEMENT OF THE PROBLEM

The ineffective and unreliable file sharing and data management system at Batangas State University Alangilan Faculty is the primary issue that this study seeks to address. The faculty shares data manually under the present system. These restrictions reduce the faculty members' productivity and quality of work, particularly when it comes to online learning and remote employment.

This paper suggests creating and implementing a distributed file system that makes use of the Microsoft DFS in order to address this issue. Users can distribute files through a web application where users can upload, download, search and encrypt files . The suggested distributed file system will improve Batangas State University Alangilan Faculty's ability to share files and manage data inspired by the BitTorrent protocol. Additionally, the system would include functions like file syncing across many platforms and devices and version control.

## 1.3. RESEARCH QUESTIONS

The Faculty of Batangas State University - Alangilan has its own record management system that handles the academic and administrative data of the faculty members. However, these systems are not integrated and may have issues of inconsistency, redundancy, and security. Therefore, this study aims to design and implement a distributed file system implementing the BitTorrent protocol that will allow the faculty members to share and manage their files more effectively that will address these problems and improve the efficiency and effectiveness of data management. The research questions that guide this study are:

1. What protocols can be implemented to make sure the ease of access of each file or data in the system?
2. How can the system ensure the security, reliability, and availability of the files stored and shared by the faculty members?
3. How can the system find the files in the distributed file system in an orderly manner?
4. What process will be used to create a reliable distributed file system for the Faculty of Batangas State University - Alangilan?

## 1.4. RESEARCH OBJECTIVES

The main objective of the study is to design and develop “Enhancing File Sharing and Data Management Capabilities at Batangas State University Alangilan Faculty Through a Distributed File System”

1. To develop a website that has the following capabilities:

1.1 Upload;

1.2 Download;

1.3 Share;

1.4 Search; and

1.5 Encryption of files

1. To incorporate web application and Distributed File System with the campus network's existing infrastructure for enhanced file management and access capabilities.
2. To create a DFS replication group to automatically synchronize and replicate data between designated servers for redundancy and data recovery.
3. To map the virtual DFS path to the actual physical location of the files in the server using Django Framework.

## 1.5. RESEARCH HYPOTHESIS

H1: Implementing a distributed file system using Microsoft DFS at Batangas State University Alangilan Faculty will significantly enhance file sharing and data management capabilities compared to the current system.

H2: The integration of the BitTorrent protocol into the distributed file system will result in improved file transfer speeds and overall system performance within the faculty network.

H3: The distributed file system using Microsoft DFS will exhibit scalability that aligns with the growing file sharing and data management needs of the faculty at Batangas State University Alangilan.

## 1.6 SCOPE AND DELIMITATIONS OF THE STUDY

The scope and delimitations of the study are the boundaries and limitations of the research project. The scope defines what the study will cover, while the delimitations specify what the study will not cover. The scope and delimitations of the study are based on the research objectives, questions, and hypotheses.

The scope of the study is to design and implement a distributed file system using Microsoft DFS that will allow the faculty members to share and manage their files more effectively. The study will also evaluate the performance and benefits of the proposed system in terms of speed, efficiency, reliability, security, and cost.

The delimitations of the study are the following:

1. The study will only focus on the faculty files and data of Batangas State University - Alangilan and will not include other campuses or departments of the university.
2. The study will only consider the files and data of full-time and part-time faculty members and will not include students of the university.

## 1.7. IMPORTANCE OF THE STUDY

The study aims to enhance the file sharing and data management capabilities at Batangas State University Alangilan Faculty by implementing a distributed file system using Microsoft DFS. This is important because it can improve the efficiency, security, and reliability of data transmission and storage among the faculty members. It can also reduce the cost and complexity of maintaining a centralized server and network infrastructure. Furthermore, it can foster collaboration and innovation among the faculty by enabling them to access and share various types of files and data easily and quickly.

**To Faculty**

Batangas State University faculty members are now allowed to access every public file distributed in the system assuming that faculty members' devices are connected to the local network. This will increase communication, easy to track, search, store, and retrieval of files. It will also be more convenient in managing different types of data such as student activities, records, as well as professional and personal information of faculty members.

**To the University**

Batangas State University will be the beneficiary of the management system using a distributed file system approach. It will help the university to manage personal and professional data and different types of files that are used in the accreditation program. In addition, it will help to utilize other existing management systems in the university.

# CHAPTER 2

# LITERATURE REVIEW

The literature review for this topic aims to explore the current state of the record management system, identify the challenges and opportunities for improvement, and propose a framework for developing and implementing a distributed file system that meets the needs and expectations of the faculty members.

## 2.1. CONCEPTUAL LITERATURE

**P2P Networks for DFS**

In *"Accelerating Content Routing with Bitswap: A multi-path file transfer protocol in IPFS and Filecoin"* by Alfonso de la Rocha, David Dias, and Yiannis Psaras (2022), investigates the design and performance of Bitswap, a block exchange protocol utilized in peer-to-peer (P2P) content addressable networks. The authors present the foundational design of Bitswap and introduce various enhancements to enhance its efficiency, effectiveness, and bandwidth usage. They conduct evaluations to assess the impact of these enhancements on the content discovery rate, a crucial metric for content routing in open networks. The study is motivated by the utilization of Bitswap in two notable projects: IPFS, a decentralized file system aiming to create a distributed version of the web, and Filecoin, a decentralized storage network that utilizes IPFS for data delivery. The authors leverage Merkle-linked graphs and data structures, enabling parallel retrieval and content integrity verification, as the basis for Bitswap. They also utilize a distributed hash table (DHT) as the primary content routing mechanism, leveraging the information acquired through DHT interactions to make informed decisions on the source of content retrieval. This study contributes to P2P content distribution by proposing and evaluating a novel multi-path file transfer protocol that expedites IPFS and Filecoin content routing.

The study authored by Telesphore Tiendrebeogo and Kabre Laciné, titled *"DISTRIBUTED E-LEARNING SYSTEM USING AN HYBRID P2P-BASED CONTENT ADDRESSABLE NETWORK ARCHITECTURE"* (2020), presents a distributed e-learning system that utilizes a hybrid P2P-based content addressable network architecture. This system aims to enhance the scalability and robustness of web-based training by employing mobile agents to distribute content and functions across computers. The system accommodates different levels of participation from computers, including active, passive, or intermittent, to offer convenience and flexibility. It is built upon a modified DHT Content Addressable Network (CAN) version to facilitate efficient and decentralized content retrieval and delivery.

The study entitled *"A Distributed Content Addressable Network for Open Educational Resources”* (2019) authored by Stephen Downes aims to propose a novel architecture for storing and retrieving open educational resources (OERs) using a content addressable network (CAN). A CAN is a distributed system that maps data to unique identifiers based on their content rather than location. This allows for efficient and scalable data retrieval and enhanced security and privacy. The conceptual literature for this study would review the relevant literature on OERs, CANs, and their applications in education. It would also define the key terms and concepts used in the study, such as content addressability, distributed hash table, peer-to-peer network, and digital object identifier. The study would then present the study's main research question and hypothesis, which is how CAN can improve the access and quality of OERs in education. It also outlines the expected contributions and implications of the study for theory and practice.

**File Protocols for DFS**

The *"Engineering Resource-Efficient Data Management for Smart Cities with Apache Kafka"* by Raptis et al. (2022), introduces a modular architecture designed to process and distribute large-scale data streams in smart city applications. The authors leverage Apache Kafka and Apache NiFi as pivotal technologies for acquiring, managing, and disseminating data, proposing a multi-layer engineering approach to extract valuable insights from diverse real-time data flows. The paper addresses the primary challenges and provides general guidelines for designing and implementing intricate data flow tasks in smart city contexts. The study demonstrates the proposed data platform's effectiveness and scalability by showcasing multiple use cases with varying requirements, including situational awareness, video surveillance, and geo-localization. Overall, this study contributes to the field of smart city management by presenting a comprehensive and adaptable solution for data-driven services capable of handling the ever-growing volume and diversity of data sources and applications.

The study *“Distributed Network File Storage for a Serverless (P2P) Network”* retrieved by Asad I. Khan (2022), proposes an innovative method for implementing distributed network file storage in a serverless (peer-to-peer) network. The authors argue that their approach can achieve high availability, scalability, security, and fault tolerance without relying on centralized servers or intermediaries. The authors outline the design and implementation of their system, which comprises three key components: a distributed hash table (DHT) for metadata storage, a content-addressable network (CAN) for data block storage, and a cryptographic protocol for ensuring data integrity and confidentiality. Through simulations and experiments, the authors assess the performance of their system and compare it with existing solutions like IPFS and Sia. Their findings demonstrate that their system outperforms current methods in terms of performance and cost-effectiveness. This study contributes to distributed systems by presenting a practical and novel solution for distributed network file storage in a serverless environment.

**Security Protocols for DFS**

The study entitled *"An Implementation of User-PC Computing System Using Docker Container"* (2020) by H. Htet et al. proposes a novel system that allows users to access their personal computers (PCs) remotely through a web browser. The system utilizes Docker containers to create virtual environments for each user, which can run various applications and services. The system also provides a user-friendly interface and a secure authentication mechanism. The authors claim that their system can reduce the cost and maintenance of PCs and improve the user experience and productivity. The study evaluates the performance and usability of the system through experiments and surveys and compares it with other existing solutions. The study contributes to the field of cloud computing by presenting a practical and innovative way of using Docker containers for user-PC computing.

In *"A Distributed File-Based Storage System for Improving High Availability of Space Weather Data"* (2019) conducted by Yoga Andrian, Hyeonwoo Kim, and Hongtaek focuses on the importance of space weather data and the challenges associated with accessing and utilizing them effectively. Space weather refers to the variations in the solar wind, interplanetary magnetic field, and Earth's magnetosphere and ionosphere, which can impact space-borne and ground-based technological systems. However, space weather data is often dispersed across different sources and formats, making them challenging to access and vulnerable to loss or corruption.

To address these issues, the authors propose a distributed file-based storage system (DFSS) that enhances the high availability of space weather data. The DFSS employs erasure coding, which divides files into fragments and encodes them with redundant information. These encoded fragments are distributed across multiple storage nodes in different locations. This ensures that even if some nodes fail or become inaccessible, the original file can still be reconstructed from the remaining fragments. The DFSS also supports dynamic replication and migration of fragments based on popularity and availability. The authors evaluate the DFSS using real-world space weather datasets and compare its performance and reliability with existing storage systems. Their findings demonstrate that the DFSS achieves improved availability, reduced storage overhead, and faster data retrieval than traditional systems.

**Speed Optimization of Uploads and Downloads for DFS**

The study conducted by Md. Sajid Hossen et al. *"Enhancing Quality of Service in SDN based on Multi-path Routing Optimization with DFS"* (2019) introduces a novel multi-path routing optimization algorithm based on depth-first search (DFS) for software-defined networks (SDNs). The objective is to enhance the quality of service (QoS) parameters, including throughput, delay, jitter, and packet loss ratio, by identifying optimal paths for data transmission in SDNs. The authors compare the performance of their algorithm with existing approaches such as shortest path first (SPF), equal-cost multi-path (ECMP), and load balancing (LB). Simulation experiments assess the algorithm's effectiveness in various network scenarios and traffic conditions. The study concludes that their algorithm outperforms other methods in terms of QoS, demonstrating improved handling of network congestion and link failures.

In a study conducted by Aneta Moravčíková, *“Decentralized data distribution with financial incentives”* (2022), examines the concept of decentralized data distribution with the inclusion of financial incentives. The author argues that existing centralized data storage systems suffer from inefficiency, vulnerability, and high costs. The author proposes a novel solution that utilizes blockchain technology and smart contracts to address these issues to establish a peer-to-peer network of data providers and consumers. Data providers receive tokens as rewards for sharing their data, while data consumers utilize tokens to access the data. The author asserts that this system can enhance data distribution's security, scalability, and affordability while fostering a fair and transparent data market. The study presents a conceptual framework and a prototype implementation of the proposed system and conducts evaluations to assess its performance and feasibility.

According to Ibrahim Ayodeji who conducted a study entitled *“Increasing service capacity of peer-to-peer file sharing networks by using a decentralized reputation system”* (2022), the author introduces a novel approach to enhance the service capacity of peer-to-peer (P2P) file-sharing networks. The study proposes the implementation of a decentralized reputation system as a means to encourage peers to collaborate and share resources more efficiently. The reputation system assigns reputation scores based on contribution and reliability, thereby incentivizing cooperative behavior. The reputation system utilizes distributed ledger technology to ensure both security and scalability. The study evaluates the performance of the proposed system through simulations and compares it to existing solutions. The results indicate that the reputation system significantly improves the service capacity of P2P file-sharing networks by addressing the issue of free-riding and enhancing the availability and quality of shared files.

**File and Data Query For DFS**

*“Blockchain-based healthcare management system with two-side verifiability”* (2022) by Tian Lim Tan, Iftekhar SalamID, and Madhusudan Singh is a study that proposes a novel framework for improving the security and efficiency of healthcare data management using blockchain technology. The authors argue that the current centralized systems are vulnerable to data breaches, privacy violations, and inefficiencies due to the stakeholders' lack of trust and transparency. A decentralized system based on the blockchain can overcome these challenges by providing two-sided verifiability, which means that both the data owners and the data consumers can verify the authenticity and integrity of the data without relying on a third party. The authors also present a prototype implementation of their framework using Hyperledger Fabric and demonstrate its feasibility and performance through experiments and simulations.

In *"Multi-agent-based hybrid peer-to-peer system for distributed information retrieval"* (2021) conducted by Abdel Naser Pouamoun and Ilker Kocabasx proposes a new approach to enhance distributed information retrieval (DIR) through a multi-agent-based hybrid peer-to-peer (P2P) system. DIR involves searching and retrieving information from diverse and distributed data sources. While P2P systems offer a decentralized structure for resource sharing and direct communication, they encounter challenges related to scalability, efficiency, reliability, security, and quality of service. The study introduces a hybrid P2P system that combines structured and unstructured elements, leveraging both benefits, to overcome these challenges. The system includes three types of agents: super peers, ordinary peers, and broker agents. Super peers maintain the global index and route queries to relevant sub-networks, while ordinary peers store and provide local data and metadata. Broker agents coordinate communication and collaboration among peers within a sub-network. The system incorporates techniques like semantic clustering, query expansion, relevance feedback, and trust management to enhance DIR performance and quality. The study assesses the proposed system through simulations and experiments, comparing it with existing P2P systems using metrics such as recall, precision, response time, network traffic, and robustness. According to the study, the proposed system outperforms existing systems in most metrics, demonstrating its feasibility and effectiveness for DIR applications.

## 2.2 RELATED LITERATURE

The study *“Dynamic Replication Management Scheme for Distributed File System”* by May Phyo Thu, Khine Moe Nwe, and Kyar Nyo Aye (2019), presents an innovative method to enhance the performance and dependability of distributed file systems. The authors propose a dynamic replication management scheme that can adjust to the system's evolving workloads and network conditions. This scheme employs a replication factor that varies based on the popularity and availability of files, along with a replication placement algorithm that considers nodes' network latency and bandwidth.

To assess the effectiveness of their approach, the authors conduct simulations and compare their scheme to existing replication methods. The results demonstrate that their system outperforms other response time, throughput, fault tolerance, and storage utilization approaches.

In the study entitled *"Dynamic Replication Policy on HDFS Based on Machine Learning Clustering"*, Motaz A. Ahmed et al. (2023), propose an innovative approach to optimize file replication in Hadoop Distributed File System (HDFS), a widely used Distributed File System (DFS) for big data applications. The authors employ machine-learning techniques to cluster files into different groups based on their importance and frequency of use. Subsequently, they apply distinct replication policies to each group to reduce storage consumption, enhance read and write operation times, and ensure the availability and reliability of HDFS as a High-Performance Distributed Computing (HPDC) system.

The authors compare their proposed method with the default replication policy of HDFS and other existing methods. The results demonstrate that their approach achieves superior storage efficiency, data availability, and fault tolerance.

According to a study entitled *"Scalable Distributed File Sharing System: A Robust Strategy for a Reliable Networked Environment in Tertiary Institutions,"* authored by Emmanuel N. Ekwonwune and Bright U. Ezeoha (2019), came up with a solution to address the limitations of centralized and single-server file-sharing systems. The authors propose a distributed system that operates on multiple servers and utilizes the Linux network operating system to synchronize user data and activities. The primary objective of this system is to achieve scalability, reliability, security, fault tolerance, and transparency in a distributed environment. The paper also explores the fundamental concepts and provides examples of distributed file systems, along with discussing the design and implementation of the proposed system.

The study titled *"Efficient I/O Merging Scheme for Distributed File Systems,"* conducted by Byoung Chul An and Hanul Sung (2023), provides a new method to enhance the performance of decentralized file systems by reducing the overhead associated with remote procedure calls (RPCs). According to the authors, decentralized file systems, which lack a metadata server, often experience significant RPC latency and traffic imbalances due to frequent metadata operations between clients and servers. To address this challenge, the authors propose a scheme that queues and merges multiple RPCs into larger ones, thereby mitigating excessive RPC delays. They evaluate the effectiveness of their scheme using a prototype implementation based on GlusterFS and demonstrate notable improvements, achieving up to 13% and 16% enhancements in write and read performance, respectively, compared to the original system. This paper contributes to distributed file systems by presenting a straightforward yet effective technique for optimizing I/O performance and scalability in decentralized systems.

In *"Interception of P2P Traffic in a Campus Network,"* authored by Merouane MEHDI (2019), investigates the issue of bandwidth saturation in university networks resulting from peer-to-peer (P2P) file sharing applications, with a particular focus on BitTorrent. The study utilizes open-source tools like Wireshark and Snorts to capture and filter P2P traffic. It addresses a significant 35% reduction in bandwidth consumption by implementing new rules to detect P2P packets. Additionally, the paper offers a comprehensive analysis of the BitTorrent protocol and its effects on network performance.

A study entitled *"Distributed File System for an Edge-Based Environment,"* authored by Rafael Neujahr Copstein and Fernando Luis Dotti (2020), examines the complexities and opportunities associated with constructing a distributed file system (DFS) for devices operating outside of dedicated infrastructures. The study includes personal computers, smartphones, tablets, and other edge devices. The authors contend that these devices possess significant computing power and storage capacity that could be effectively utilized for collaborative purposes. However, they also encounter challenges such as high mobility, network partitioning, heterogeneity, and security concerns.

The study presents a system model and architecture for a DFS capable of addressing these challenges and offering features similar to those found in cloud-based DFS. The article also delves into implementation details and outlines potential future directions for the advancement of such a system.

The study titled *"Assise: Performance and Availability via Client-local NVM in a Distributed File System"* by Anderson et al. (2020) presents Assise, a distributed file system (DFS) that capitalizes on client-local persistent memory modules (PMMs) to maximize performance and availability while guaranteeing linearizability for IO operations. In contrast to traditional DFSs that rely on remote storage devices with high latency, PMMs offer advantageous features such as low latency, high bandwidth, byte-addressability, and non-volatility, enabling new possibilities for applications and architectures. Assise incorporates a persistent, replicated coherence protocol that treats client-local PMMs as a crash-recoverable cache between applications and slower storage. It optimizes data locality by utilizing process-local, socket-local, and client-local PMMs for IO operations whenever feasible, and rather than maintaining consistency at fixed block sizes, Assise ensures consistency at the level of IO operations, thereby minimizing coherence overhead.

The authors compare Assise, a distributed file system (DFS), with other methods, such as Ceph/BlueStore, NFS, and Octopus, using Intel Optane DC PMMs and SSDs for various cloud applications and benchmarks. The findings show that Assise outperforms its counterparts by improving write latency, throughput, fail-over time, and scalability while offering stronger consistency guarantees. Assise is an innovative DFS that leverages PMMs to enhance the performance and availability of DFS-based applications. It introduces a novel coherence protocol that achieves strong consistency with minimal overhead.

The advantages, disadvantages, and performance traits of various DFS solutions are examined in a *“A Comparative Study on Distributed File Systems”* by De & Panjwani (2019). The experimental setting, DFS solution selection criteria, and performance evaluation measures all adhere to a systematic methodology in this study. A thorough comparison analysis is provided in the report, which rates each system's scalability, fault tolerance, data integrity, access latency, and throughput. To get unbiased results on read/write latency, throughput, data integrity, and resource utilization, performance evaluation and benchmarking are undertaken. The article highlights difficulties encountered and offers potential research avenues for the future, advancing the field of distributed file system research.

In "*Offline but still connected with IPFS based communication*" (2020) by Cristeaa et al. examines the potential use of the InterPlanetary File System (IPFS) protocol for enabling offline communication between devices. IPFS is a decentralized network that allows users to store and share files without relying on centralized servers. The authors propose a system that utilizes IPFS to establish local networks of devices capable of communicating with each other, even without an internet connection. Additionally, the system facilitates the transmission of messages to online devices by utilizing IPFS as an intermediary. The authors assess the performance and usability of their system through experiments and user feedback, concluding that it is both feasible and effective for offline communication. Furthermore, they note that the system has the potential to enhance online communication by offering increased privacy and resilience.

The study titled "*Routing performance of structured overlay in Distributed Hash Tables (DHT) for P2P*" by Rafiza Ruslan et al. (2019) examines the routing performance of three structured overlay algorithms (Chord, Kademlia, and Pastry) within DHT-based P2P networks, considering various network conditions and parameters. The study employs simulation experiments to assess and compare the three algorithms' routing efficiency, latency, and load. Additionally, it investigates the influence of network size, node degree, node churn, and message size on routing performance. By conducting a comprehensive and comparative evaluation of structured overlay algorithms in DHT-based P2P networks, the study contributes to the existing literature on P2P networks. Furthermore, it identifies challenges and offers future research directions to enhance the routing performance of structured overlay algorithms within DHT-based P2P networks.

## 2.3. SYNTHESIS

Both studies aim to improve the performance and reliability of distributed file systems through innovative replication techniques. The study by May Phyo Thu, Khine Moe Nwe, and Kyar Nyo Aye proposes a dynamic replication management scheme that can adapt to changing workloads and network conditions. This scheme involves adjusting the replication factor and employing a placement algorithm based on file popularity and availability. On the other hand, the study by Motaz A. Ahmed et al. focuses on optimizing file replication in the Hadoop Distributed File System (HDFS) using machine learning clustering. They utilize clustering techniques to group files based on their significance and frequency of usage and then apply distinct replication policies to each group.

While both studies share a similar objective, there are notable differences between them. The study by May Phyo Thu, Khine Moe Nwe, and Kyar Nyo Aye utilizes a dynamic replication management scheme, whereas the study by Motaz A. Ahmed et al. utilizes machine learning clustering for replication policy optimization. Furthermore, the first study compares its proposed scheme with existing replication methods, evaluating factors such as response time, throughput, fault tolerance, and storage utilization. In contrast, the second study compares its approach with the default replication policy of HDFS and other existing methods, focusing on metrics such as storage efficiency, data availability, and fault tolerance.

Both studies aim to improve file-sharing performance and efficiency in distributed environments. The first study which is conducted by Emmanuel N. Ekwonwune and Bright U. Ezeoha (2019) suggests a distributed file-sharing system to overcome the limitations of centralized systems, with an emphasis on scalability, reliability, security, fault tolerance, and transparency. In contrast, the second study conducted by Byoung Chul An and Hanul Sung (2023) focuses on enhancing the performance of decentralized file systems by introducing an I/O merging scheme to minimize overhead caused by remote procedure calls (RPCs) and improve the speed of read and write operations. The study by Emmanuel N. Ekwonwune and Bright U. Ezeoha (2019) emphasizes synchronizing user data and activities across multiple servers using the Linux network operating system. Whereas, the study by Byoung Chul An and Hanul Sung (2023) targets RPC latency and traffic imbalances in decentralized file systems. The first study comprehensively examines distributed file systems, encompassing fundamental concepts, practical examples, and implementation details to provide a robust strategy for reliable networked environments. On the other hand, the second study concentrates on optimizing I/O performance and scalability, assessing its approach through a prototype implementation based on GlusterFS and demonstrating significant performance improvements.

In "Interception of P2P Traffic in a Campus Network," authored by Merouane MEHDI (2019), and "Distributed File System for an Edge-Based Environment," authored by Rafael Neujahr Copstein and Fernando Luis Dotti (2020), both studies share the common objective of addressing challenges in network environments. The first study investigates the issue of bandwidth saturation in university networks caused by P2P file-sharing applications, specifically BitTorrent. It reduces bandwidth consumption by employing open-source tools for capturing and filtering P2P traffic. Similarly, the second study centers around constructing a distributed file system (DFS) for edge-based devices, including personal computers, smartphones, and tablets. It recognizes the complexities and potential benefits of utilizing these devices for collaborative purposes while considering challenges such as high mobility, network partitioning, heterogeneity, and security.

However, there are notable distinctions between the two studies. The first study addresses bandwidth saturation and implements new detection and filtering rules for P2P packets, analyzing the BitTorrent protocol's impact on network performance. On the other hand, the second study emphasizes developing a DFS tailored for edge devices operating outside dedicated infrastructures. It delves into the specific challenges associated with edge computing, presents a comprehensive system model and architecture for the DFS, and outlines potential future directions.

The study conducted by Anderson et al. (2020) titled "Assise: Performance and Availability via Client-local NVM in a Distributed File System", and in a study by De & Panjwani (2019) titled “A Comparative Study on Distributed File Systems”, both studies share the common objective of evaluating and comparing different distributed file system (DFS) solutions. The first study focuses on Assise, a specific DFS that utilizes client-local persistent memory modules (PMMs) to improve performance and availability. It introduces a unique coherence protocol and assesses Assise's performance against other methods through specific hardware configurations and benchmarks. In contrast, the second study takes a broader approach by conducting a comprehensive comparative analysis of various DFS solutions. It follows a systematic methodology, considering multiple evaluation measures and benchmarking criteria to evaluate scalability, fault tolerance, data integrity, access latency, and throughput. Additionally, the second study discusses challenges encountered during the evaluation and suggests future research directions.

The study entitled “Offline but still connected with IPFS based communication" (2020) by Cristeaa et al., and "*Routing performance of structured overlay in Distributed Hash Tables (DHT) for P2P*", a study conducted by Rafiza Ruslan et al. (2019), both studies evaluate and compare different protocols or algorithms in their respective domains. The first study explores the application of the IPFS protocol for offline communication among devices. The second study investigates the routing performance of structured overlay algorithms in DHT-based P2P networks. Experimental methodologies are employed in both studies to assess and compare the performance of the protocols or algorithms across diverse conditions.

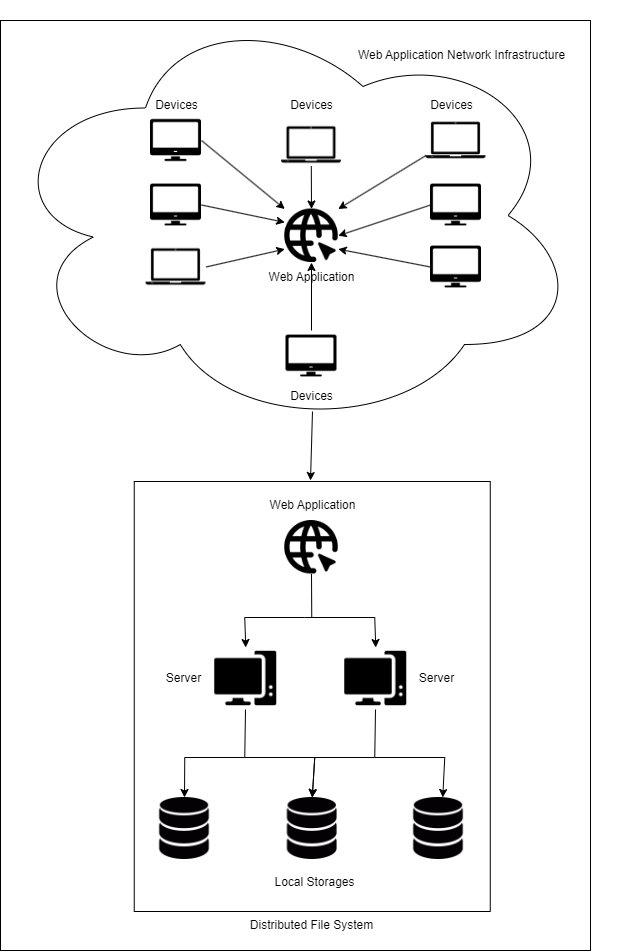
However, there are also differences between the two studies. The first study proposes a system that utilizes IPFS to establish local networks for offline communication, with IPFS serving as an intermediary for transmitting messages to online devices. Through experiments and user feedback, the study evaluates the system's feasibility, effectiveness, and potential benefits for online communication, such as improved privacy and resilience. In contrast, the second study evaluates the routing performance of structured overlay algorithms in DHT-based P2P networks. It analyzes three algorithms' efficiency, latency, and load characteristics (Chord, Kademlia, and Pastry) under varying network conditions and parameters. The study presents a comprehensive comparative analysis of these algorithms and identifies challenges and avenues for future research to enhance their routing performance.

# CHAPTER 3

# DESIGN AND METHODS

## 3.1. CONCEPTUAL FRAMEWORK

Shown below is the representation of the research design that the researchers intended to implement. The figure describes the flow of the process of the study in order to develop a working system.



*Fig. 1. Conceptual framework*

Shown in Figure 1 is the conceptual framework for this study. A system intended to transform data management and file sharing for faculty members at Batangas State University Alangilan Faculty is described in the conceptual framework. With the use of an intuitive web application, faculty members can securely share and collaborate on digital files by using it as the main hub for file upload. Additionally, the web application communicates with a Distributed File System (DFS), which serves as the central hub for managing and storing data. By distributing files among several servers inside the current campus network, the DFS provides benefits including automatic data replication and effective storage and retrieval in response to user demands. The web application can be accessed by faculty devices and other components over the communication route provided by the campus network. The web application and the DFS may communicate seamlessly because of the Django Framework, which enables developers to design an intuitive faculty interaction interface. This helps to close the gap between technological complexity and user-friendliness. The web application receives files, communicates with the DFS, and manages the storage of files as part of a well-defined data flow path within the system. The findings are shown back to the user via the interface, and Django generates a virtual DFS path, providing an easy-to-use method of accessing files. The current restrictions on file sharing and data management at the Alangilan Faculty of Batangas State University are addressed in this thorough conceptual framework.

## 3.2. TECHNICAL DESIGN AND PROCEDURE

The design of the study uses different designs that can be produced for the study, each of the designs uses a different approach on how the system works.

## 

## 3.2.1. DFS Namespace for Centralized File Access

## 

*Fig. 2: File Replication with DFS Namespace and Failover Cluster*

In Figure 2, demonstrates an innovative architecture built to handle uploaded files in a web application context with efficiency. Fundamentally, the web application acts as a file upload gateway, starting a smooth distribution and replication process. The system provides users with a unified picture of the file system hierarchy by abstracting the physical file locations through the usage of Distributed File System (DFS) Namespace. In addition, a failover cluster minimizes downtime by automatically rerouting traffic to servers that are in good health in the event of a breakdown. By replicating files over many local storage systems dispersed among cluster servers, replication algorithms further enhance the system's resilience. This method not only improves data redundancy but also makes load balancing easier, guaranteeing steady performance even during periods of high usage. For the most part, this architecture provides a reliable, scalable, and fault-tolerant way to manage and duplicate uploaded files—a critical component that keeps online applications running smoothly.

## 3.3. EQUIPMENT/FACILITIES/PROGRAMS

The following section outlines the research equipment, facilities and programs that the researchers plan to employ to this study.

**Equipment**: Personal computers, servers, and other storage-capable computing equipment are needed by participating nodes. The intended size and capacity of the distributed file system determines the number of nodes.

**Facilities**: The interconnected design is compatible with LANs, WANs, and even the internet as network environments. For seamless file sharing and access among nodes, a stable network architecture is required.

**Program**: Creating web applications for each node is necessary to implement the interconnected file system. These applications offer file sharing and node-to-node synchronization. Additionally, user-friendly file system access may be provided by client programs or online interfaces.

## 3.4. EXPECTED OUTPUT

The expected outcome for the topic "Enhancing File Sharing and Data Management Capabilities at Batangas State University Alangilan Faculty through an Distributed File System" is a comprehensive software system made to meet the file sharing and data management requirements of the faculty members. The faculty members will be able to upload, download, search and encrypt files for security. The file system's high availability, fault tolerance, and scalability are all guaranteed by this centralized strategy.

Faculty members will be able to safely save their data on the software system's user-friendly interface and access them from a variety of gadgets, including desktop computers, or laptops. With this solution in place, faculty members won't have to rely on a certain computer or location to access their data anymore. Additionally, the technology would make it simple for faculty members to share files, encouraging teamwork and knowledge sharing at the university.

The web app will include technologies to optimize file sharing and data management. Sensitive data will be shielded from unauthorized access by ensuring the security of the system. In order to conserve storage space, the system will only keep distinct files. Additionally, the system will provide file synchronization, enabling faculty members to maintain the most recent versions of their files across several devices. Additionally, version control features will let faculty members monitor and manage various iterations of their work, making it simple to access earlier iterations and encouraging collaboration.

Faculty members' productivity and collaboration will increase as a result of the implementation of this networked distributed file system. They will be able to exchange information, work together on research projects, share resources, and give input on shared documents all without difficulty. As a result, faculty collaboration and communication will be significantly improved. The system will allow features like file comments and notification alerts.

The system's stored files will be protected with security measures. Only faculty members with the proper authorization will be able to view certain files thanks to access controls, user authentication, and authorization processes. In order to monitor file access and identify any illegal activity, audit trails and logging systems will be put in place. This will improve the system's overall security. The robust and feature-rich software solution that improves file sharing and data management capabilities at the Alangilan Faculty of Batangas State University is the projected outcome of this thesis project. The peer-to-peer design, advanced functionality, and improved security measures of the system will considerably increase faculty members' productivity, collaboration, and overall performance on academic and administrative activities.

## 

## 3.5. GANTT CHART

*GANTT CHART*

The illustration above clearly shows the intended timeframe and the precise breakthrough for monitoring progress of this study. The researchers will be able to make sure the study maintains on track and complies with the planned timeframe. The first phase would involve extensive research and planning to understand the study requirements thoroughly, which would be conducted from April to June 2023. By conducting thorough study and precise preparation, the research and planning phase establishes a framework for the succeeding stages of the thesis project. It provides an in-depth understanding of the faculty's record management needs and expectations while offering a well-defined framework to guide the development and implementation of the record management system. The thesis project's successful completion ultimately depends on the accomplishment of this phase in developing the appropriate foundation.

The requirement analysis would follow next, occurring between June and July 2023. In order to make sure the record management system meets the specific requirements and issues of the faculty, this phase focuses on gathering and analyzing the demands and expectations of the faculty. The researchers will have in-depth talks and consultations with faculty, IT staff, and other relevant stakeholders. The main goal is to understand their current record management system, issues, and expectations for the suggested solution. The researchers will conduct consultations, inquiries, and discussions to get detailed information and identify all of the requirements of the faculty.

This software design phase, scheduled from July to September 2023 seeks to implement the established requirements into a detailed design that provides the faculty members accessibility to an effective and accessible record management system. The main goal is to specify the system's overall architecture, covering a range of concerns, including database design, user interface design, and system components. Following that, development and coding would occur from July to October 2023, this stage is crucial for developing the software design into a record management system that can function efficiently. It will apply the system's functionalities and features to actual use, as stated during the software design phase.

During the integration phase, which will be regulated from September to November 2023, the researchers will work to assemble all of the developed modules and components to ensure their functions are effectively combined. Establishing the record management system and the chosen system design involves configuring the required configurations, protocols, and routes for communication.

System testing and debugging, and finalizing the system would come next, which would both occur from October to November 2023. In the system testing and debugging phase, researchers will thoroughly test the system to ensure its proper functionality, performance, and compatibility with LAN cable and wireless connections. They will evaluate various aspects, including reliability, accessibility, efficiency, and security, to identify and address issues or errors through debugging, code modifications, and retesting. This iterative process will continue until the system meets the specified requirements. In the finalization phase, the researchers will focus on completing the system by conducting final performance and security tests. They will address minor bugs and make adjustments based on user feedback, ensuring the accuracy and relevance of system documentation. The system will be prepared for deployment by setting up the necessary hardware, software, and network components. Information transfer sessions may be conducted to familiarize the faculty with the system's structure and usage.

Finally, the last phase, which would be the final editing that will be conducted in December 2023. The researchers thoroughly evaluate every task stated in the illustration and verify how long it is during the last editing phase to ensure it meets the study's requirements and objectives. Additionally, they ensure that all crucial outputs and milestones are. The dependencies between tasks are extensively investigated, mentioned, and placed appropriately within the timeframe. To ensure that they accurately demonstrate the logical order and interdependencies between activities. The researchers also consider resource reliability, technical difficulties, and restrictions relating to the LAN cable, wireless connection, or combination components of the system while evaluating the feasibility of the timeline. The Gantt chart would be revised or updated according to the need to accommodate any alterations that might have taken place throughout the execution of the project, such as task additions, updates, or removals.

# REFERENCES:

De la Rocha, A., Dias, D., & Psaras, Y. (2021). Accelerating Content Routing with Bitswap: A multi-path file transfer protocol in IPFS and Filecoin. Protocol Labs. <https://research.protocol.ai/publications/accelerating-content-routing-with-bitswap-a-multi-path-file-transfer-protocol-in-ipfs-and-filecoin/>.

Tiendrebeogo, T., & Laciné, K. (2019). *DISTRIBUTED E-LEARNING SYSTEM USING AN HYBRID P2P-BASED CONTENT ADDRESSABLE NETWORK ARCHITECTURE*. International Journal of Distributed and Parallel Systems (IJDPS). <https://airccse.org/journal/ijdps/current2020.html>.

Stephen, D. (2019). A DISTRIBUTED CONTENT ADDRESSABLE NETWORK FOR OPEN EDUCATIONAL RESOURCES. Research Gate. <https://www.researchgate.net/publication/338496223_A_DISTRIBUTED_CONTENT_ADDRESSABLE_NETWORK_FOR_OPEN_EDUCATIONAL_RESOURCES>.

Ruslan, R., Zailani, A. S. M., Zukri, N. H. M., Kamarudin, N. K., Elias, S. J., & Ahmad, R. B. (2019). Routing performance of structured overlay in Distributed Hash Tables (DHT) for P2P. Bulletin of Electrical Engineering and Informatics. <https://beei.org/index.php/EEI/article/view/1449>.

Raptis, T. P., Cicconetti, C., Falelakis, M., Kalogiannis, G., Kanellos, T., & Lobo, T. P. (2022). *Engineering Resource-Efficient Data Management for Smart Cities with Apache Kafka*. MDPI. <https://zenodo.org/record/7560129>.

Cristea, A. G., Alboaie, L., Panu, A., & Radulescu, V. (2020). Offline but still connected with IPFS based communication. Science Direct. <https://www.sciencedirect.com/science/article/pii/S1877050920320858>.

Retrieved by Khan, A. I. (2022). Distributed network file storage for a serverless (P2P) network. Research Gate. <https://www.researchgate.net/publication/4055324_Distributed_network_file_storage_for_a_serverless_P2P_network>.

Htet, H., Funabiki, N., Kamoyedji, A., Kuribayashi, M., Akhter, F., & Kao, W. C. (2020). An Implementation of User-PC Computing System Using Docker Container. International Journal of Future Computer and Communication. <http://www.ijfcc.org/vol9/568-F1010.pdf>.

Andrian, Y., Kim, H., & Ju, H. (2020). A Distributed File-Based Storage System for Improving High Availability of Space Weather Data. Research Gate. <https://www.researchgate.net/publication/337450101_A_Distributed_File-Based_Storage_System_for_Improving_High_Availability_of_Space_Weather_Data>.

M. S. Hossen, M. H. Rahman, M. Al-Mustanjid, M. A. Shakil Nobin and M. A. Habib, "Enhancing Quality of Service in SDN based on Multi-path Routing Optimization with DFS," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2019, pp. 1-5, <https://ieeexplore.ieee.org/abstract/document/9068057>

Moravčíková, A. (2021). Decentralized data distribution with financial incentives. Muni. <https://is.muni.cz/th/o0nuq/Diploma_thesis.pdf>.

Ayodeji, I. (2022). Increasing service capacity of peer-to-peer file sharing networks by using a decentralized reputation system. National College of Ireland. [https://norma.ncirl.ie/5995/.](https://norma.ncirl.ie/5995/.%5C)

Tan TL, Salam I, Singh M (2022) Blockchain-based healthcare management system with two-side verifiability. PLoS ONE 17(4): e0266916. <https://doi.org/10.1371/journal.pone.0266916>.

Pouamoun, A. N., & Kocabasx, I. (2021). Multi-agent-based hybrid peer-to-peer system for distributed information retrieval. Journal of Information Science. <https://journals.sagepub.com/doi/abs/10.1177/01655515211010392>.

Thu, M. P., Nwe, K. M., & Aye, K. N. (2019). Dynamic Replication Management Scheme for Distributed File System. Springer Link. <https://link.springer.com/chapter/10.1007/978-981-13-0869-7_16>.

M. A. Ahmed, M. H. Khafagy, M. E. Shaheen and M. R. Kaseb, "Dynamic Replication Policy on HDFS Based on Machine Learning Clustering," in IEEE Access, vol. 11, pp. 18551-18559, 2023, <https://ieeexplore.ieee.org/document/10049393?denied=>.

Ekwonwune, E. and Ezeoha, B. (2019) Scalable Distributed File Sharing System: A Robust Strategy for a Reliable Networked Environment in Tertiary Institutions. International Journal of Communications, Network and System Sciences, 12, 49-58. <https://www.scirp.org/journal/paperinformation.aspx?paperid=92383>.

An BC, Sung H. Efficient I/O Merging Scheme for Distributed File Systems. Symmetry. 2023; 15(2):423. <https://doi.org/10.3390/sym15020423>.

Mehdi, M. (2019). Interception of P2P Traffic in a Campus Network. Research Gate. <https://www.researchgate.net/publication/334213518_Interception_of_P2P_Traffic_in_a_Campus_Network>.

Copstein, R. N., & Dotti, F. L. (2020). Distributed File System for an Edge-Based Environment. ArXiv. <https://arxiv.org/pdf/2006.09182.pdf>.

Anderson, T. E., Canini, M., Kim, J., Kostic, D., Kwon, Y., Peter, S., Reda, W., Schuh, H. N., & Witchel, E. (2020). Assise: Performance and Availability via Client-local NVM in a Distributed File System. King Abdullah University of Science and Technology. <https://repository.kaust.edu.sa/handle/10754/670570>.

De, S., & Panjwani, M. (2021). A Comparative Study on Distributed File Systems. Research Gate. <https://www.researchgate.net/publication/351116448_A_Comparative_Study_on_Distributed_File_Systems>.

Cristea, A. G., Alboaie, L., Panu, A., & Radulescu, V. (2020). Offline but still connected with IPFS based communication. Science Direct. <https://www.sciencedirect.com/science/article/pii/S1877050920320858>.

Ruslan, R., Zailani, A. S. M., Zukri, H., & Kamarudin, N. K. (2019). Routing performance of structured overlay in Distributed Hash Tables (DHT) for P2P. Research Gate. <https://www.researchgate.net/publication/333860645_Routing_performance_of_structured_overlay_in_Distributed_Hash_Tables_DHT_for_P2P>.