IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 10, October 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910024 |

DC Store: A Deduplication-Based Cloud-of-Clouds Storage Service

Athyam Meghana Gupta, Mr.V.Mallesh

M. Tech Student, Dept. of Computer Science and Engineering, Bheema Institute of Technology and Science

Alur Road, Adoni Kurnool, Andhra Pradesh, India

Professor (HOD), Dept. of Computer Science and Engineering, Bheema Institute of Technology and Science

Alur Road, Adoni Kurnool, Andhra Pradesh, India

ABSTRACT: The increasing popularity of cloud storage is leading many organizations to move their data into the cloud. However, putting all data in one cloud causes problems such as vendor lock-in, increased service costs, and data availability. In this paper, we introduce DCStore, a Cloud-of- Clouds storage service designed for an organization to outsource their data into the clouds. To achieve the goal of cost-efficient and high-available, we combine three key techniques.

First, DCStore eliminates the redundant data at client-side to save storage cost via applicationaware chunking method. Second, DCStore uses an inner-chunk based erasure coding scheme to distribute unique chunks across multiple clouds for high availability. Finally, a container-based share management strategy is used for performance optimization.

I. INTRODUCTION

The increasing popularity of cloud storage is leading many organizations to move their own data into the cloud. Typical usage examples include off-site backup, storing digital media and content distribution. Even organizations that handle critical data, such as medical record and financial data, are adopting cloud computing as a way to reduce costs. With such large amounts of critical *data* stored in the cloud, data availability and storage cost have become two major concerns for an organization. Although cloud service providers abide by strict service-level agreements (SLAs) with impressive up-times and response delays, experience tells us that even the best providers suffer from occasional outages, which can result in severe financial losses. Besides, the cloud service providers always offer different prices for market competition. Therefore, it is wise for customers to leverage multiple cloud storage services, called Cloud-of-Clouds, to improve data availability and minimize cost.

In a Cloud-of-Clouds storage system, data is distributed across multiple clouds to achieve the goal of high-available and cost-efficient. Reliability-enhanced technologies (e.g., replication or erasure coding) are first used so that data can be recovered from a subset of clouds even if the remaining clouds are unavailable. For example, DuraCloud utilizes replication to achieve the goal of availability at very high storage overhead. Some recent studies (e.g., RACS and NCCloud) leverage erasure coding to tolerate cloud failures with much less storage overhead. Previous studies also have shown that data redundancy is moderate to high in the storage system for an organization. For instance, about 69%-97%, 42%-68%, and 20%-30% of the data in secondary storage, primary storage, and HPC data centers, respectively, are redundant. Since those unnecessary redundancies incur additional cost, organizations would benefit a lot from clientside data deduplication before outsourcing their data into the clouds.

However, it is not easy in real life. Several factors impede this goal. First, data deduplication technology splits large files into a sequence of small chunks, which leads to degraded Gets/Puts performance and additional expense due to the extra network I/Os. For example, Amazon S3 has both a per-request and a per-byte cost when storing a file. If an 8MB file was split into chunks with an average size of 8KB, the request costs will increase 1,000 times.

Second, Data deduplication reduces the reliability of the storage systems because the loss of a few critical data chunks can lead to many referenced files being lost. While applying erasure coding after deduplication improves data availability, existing works that operate erasure encoding on multiple fixed-size objects are inefficient. This problem will be specified in Section II-B. To handle the above-stated two challenges, we design and implement a Cloud-of-Clouds storage service called DCStore. DCStore is designed for an organization to outsource a large group of users' data into multiple clouds with cost efficiency and availability guarantees. It combines three main techniques: client-side data deduplication, inner-chunk encoding, and container-based share management strategy. Our experimental

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910024 |

evaluations show that DCStore can improve the performance and cost efficiency significantly, compared with some existing Cloud-of-Clouds storage systems (i.e., DAC and CYRUS).
In designing and evaluating this system we make three contributions. First, through trace-driven simulation, we show that it is possible for an organization to minimize cost and tolerate outages using data deduplication and erasure coding techniques. Second, we propose an inner-chunk based erasure coding scheme and use on-demand chunk reconstruction to reduce storage overhead upon failures. Third, we design and implement a container-based share management strategy to aggregate small data shares into a single larger unit for performance optimization.

II. EXISTING SYSTEM

Cloud service providers abide by strict service-level agreements (SLAs) with impressive up-times and response delays, experience tells us that even the best providers suffer from occasional outages, which can result in severe financial losses
Besides, the cloud service providers always offer different prices for market competition. Therefore, it is wise for customers to leverage multiple cloud storage services, called Cloud-of-Clouds, to improve data availability

Disadvantages of Existing System :
However, it is not easy in real life. Several factors impede this goal. First, data deduplication technology splits large files into a sequence of small chunks, which leads to degraded Gets/Puts performance and additional expense due to the extra network I/Os.

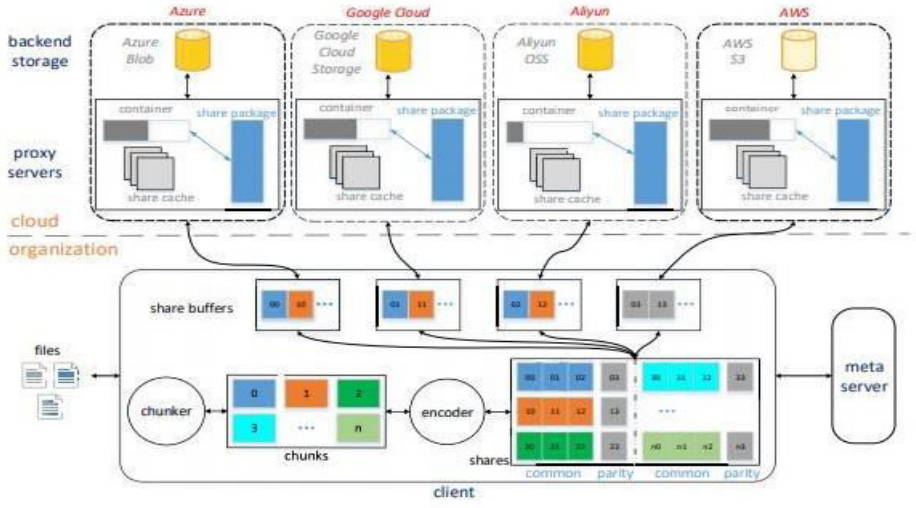
III. PROPOSED SYSTEM

We show that it is possible for an organization to minimize cost and tolerate outages using data deduplication and erasure coding techniques. Second, we propose an inner-chunk based erasure coding scheme and use on-demand chunk reconstruction to reduce storage overhead upon failures. Third, we design and implement a container-based share management strategy to aggregate small data shares into a single larger unit for performance optimization.

Advantages of proposed System:
We design and implement a Cloud-of-Clouds storage service called DCStore. DCStore is designed for an organization to outsource a large group of users' data into multiple clouds with cost efficiency and availability guarantees.
Our experimental evaluations show that DCStore can improve the performance and cost efficiency significantly, compared with some existing Cloud-of-Clouds storage systems

IV. SYSTEM ARCHITECTURE OF THE PROJECT

Below architecture diagram represents mainly flow of request from the users to database through servers. In this scenario overall system is designed in three tiers separately using three layers called presentation layer, business layer, data link layer. This project was developed using 3-tier architecture.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910024 |

3-Tier Architecture:

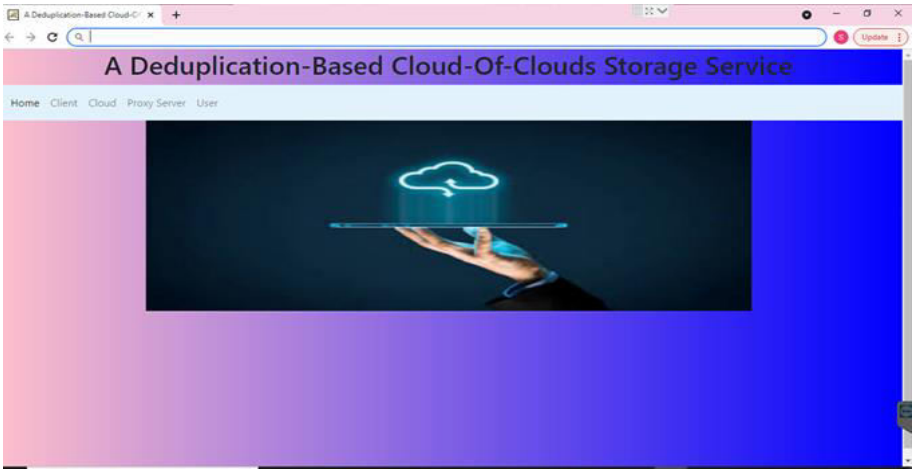
The three-tier software architecture (a three layer architecture) emerged in the 1990s to overcome the limitations of the two-tier architecture. The third tier (middle tier server) is between the user interface (client) and the data management (server) components. This middle tier provides process management where business logic and rules are executed and can accommodate hundreds of users (as compared to only 100 users with the two tier architecture) by providing functions such as queuing, application execution, and database staging.

The three tier architecture is used when an effective distributed client/server design is needed that provides (when compared to the two tier) increased performance, flexibility, maintainability, reusability, and scalability, while hiding the complexity of distributed processing from the user. These characteristics have made three layer architectures a popular choice for Internet applications and net-centric information systems.

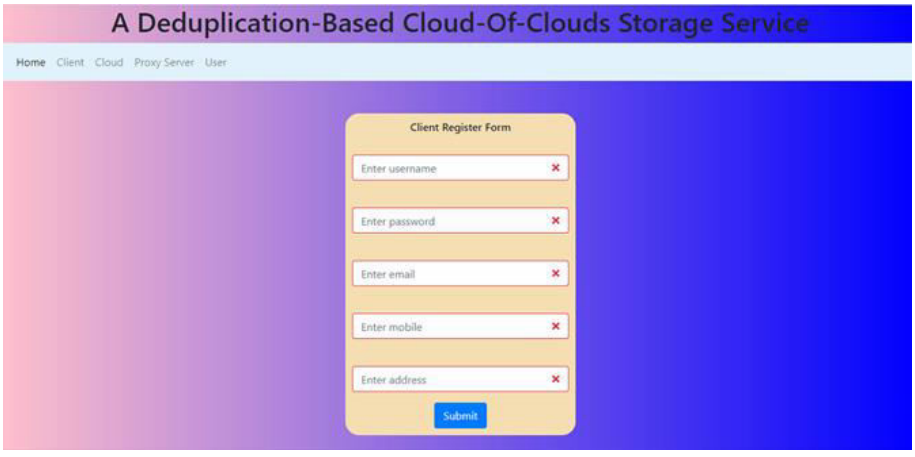
Advantages of Three-Tier:

Separates functionality from presentation.
Clear separation – better understanding.
Changes limited to well define components.
Can be running on WWW.
Effective network performance.

Screenshots:
Homepage of the website



Client register:

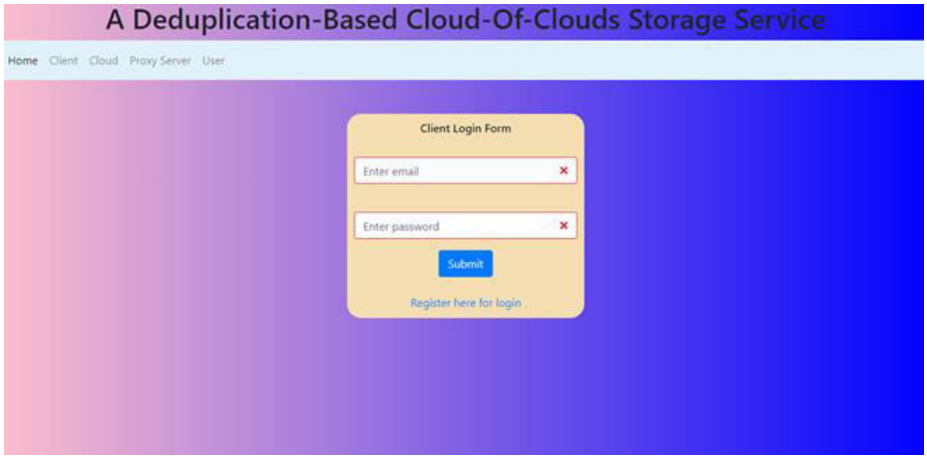


| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

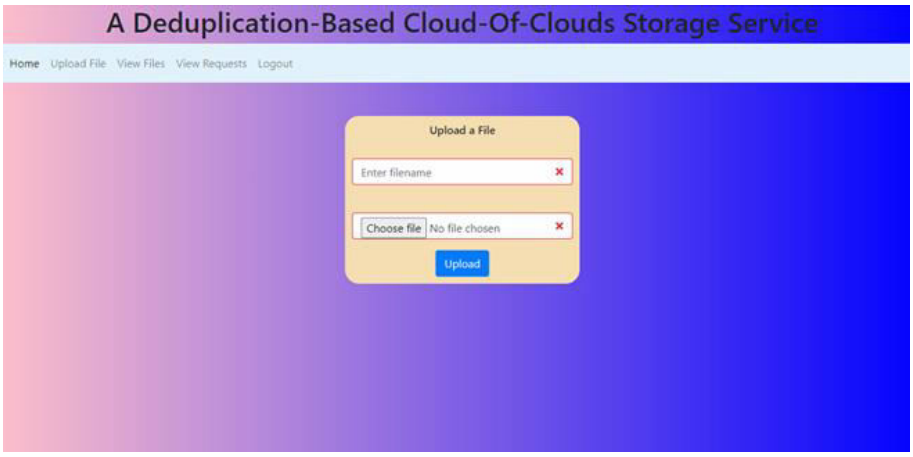
|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910024 |

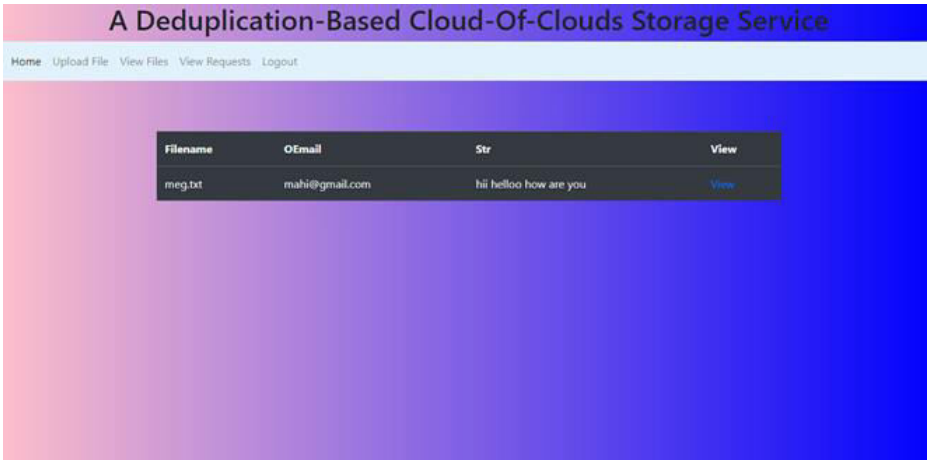
Client login:



Upload file:



View Files:

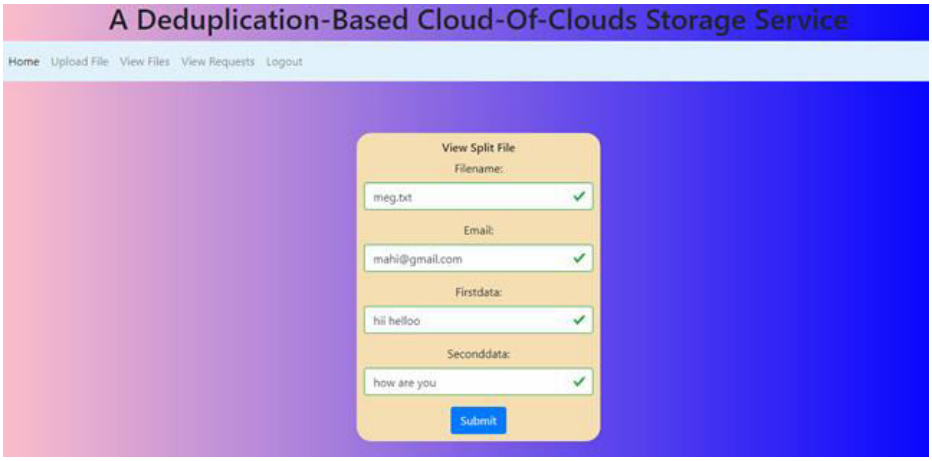


| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

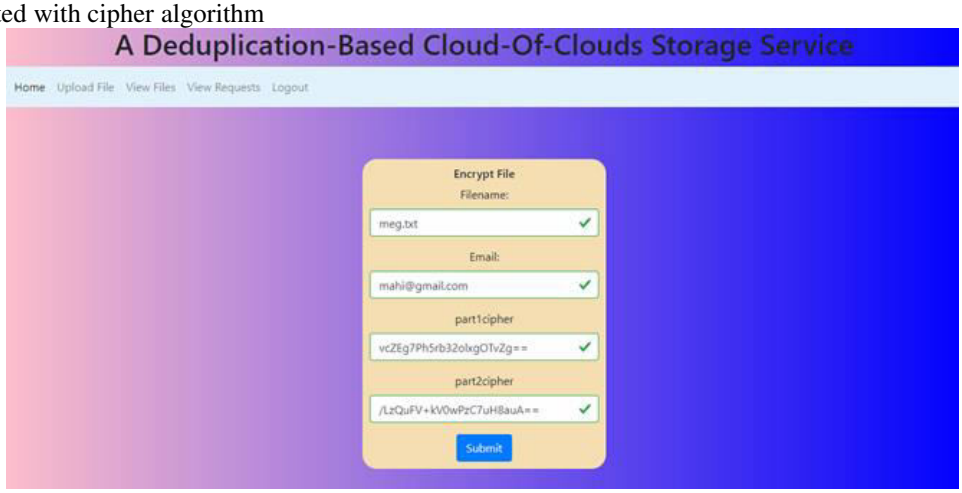
|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910024 |

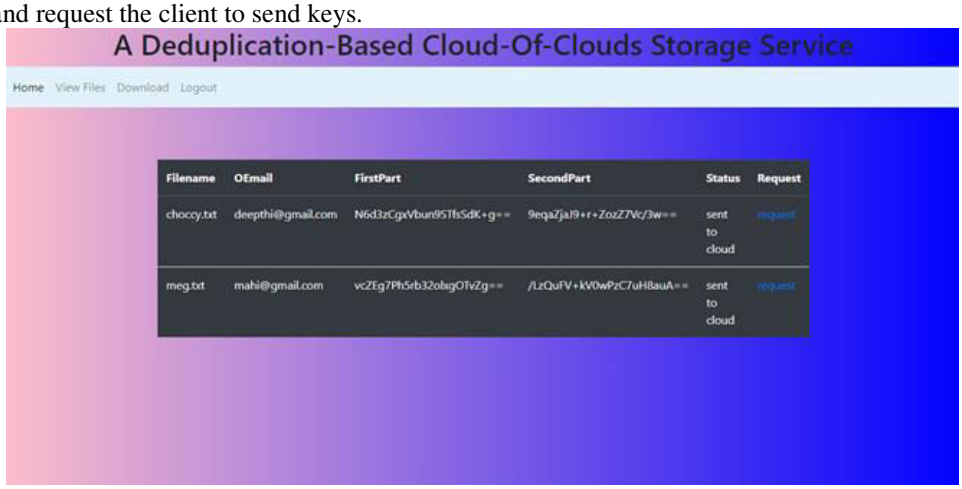
Divide data



Data is encrypted with cipher algorithm



Login as user and request the client to send keys.

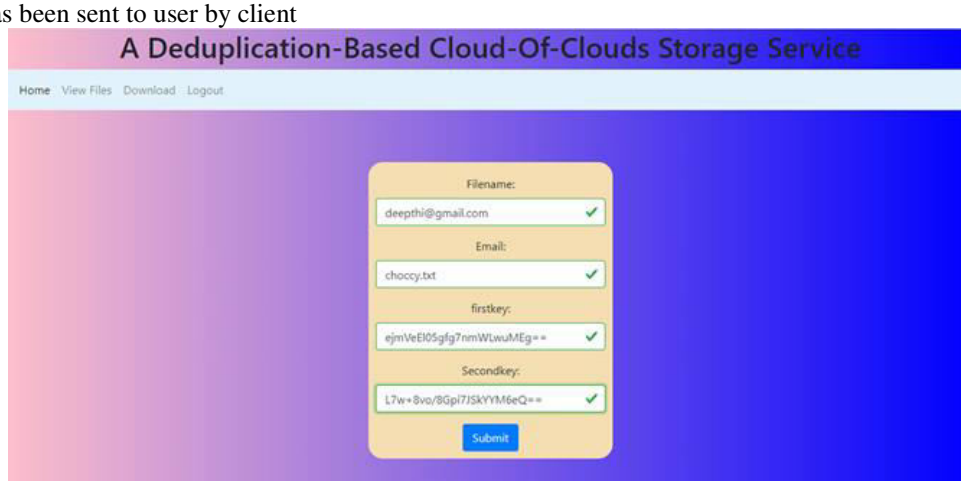


| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910024 |

Secret Keys has been sent to user by client



After entering the keys, users can see the message sent by the client.

V. CONCLUSION

In this project, we present DCStore, a deduplication-based Cloud-of-Clouds storage service that practically addresses the reliability of cloud storage service. DCStore not only achieves fault tolerance of storage, but also achieves cost savings via client-side data deduplication, inner- chunk encoding, and container-based share management strategy. Our lightweight prototype implementation of DCStore shows that DCStore improve the performance and cost efficiency significantly compared with existing Cloud-of-Clouds storage system.

REFERENCES

1. J. Crowcroft, “On the duality of resilience and privacy,” in Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 471, no. 2175. The Royal Society, 2015, p. 20140862.
2. Bessani, M. Correia, B. Quaresma, F. Andr´e, and P. Sousa, “Depsky: dependable and secure storage in a cloud-of-clouds,” ACM Transactions on Storage (TOS), vol. 9, no. 4, p. 12, 2013.
3. H. Chen, Y. Hu, P. Lee, and Y. Tang, “Nccloud: A network-coding-based storage system in a cloud-of-clouds,” 2013.
4. T. G. Papaioannou, N. Bonvin, and K. Aberer, “Scalia: an adaptive scheme for efficient multi-cloud storage,” in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis. IEEE Computer Society Press, 2012, p. 20.
5. Z. Wu, M. Butkiewicz, D. Perkins, E. Katz-Bassett, and H. V. Madhyastha, “Spanstore: Cost-effective geo-replicated storage spanning multiple cloud services,” in Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles. ACM, 2013, pp. 292–308.
6. G. Greenwald and E. MacAskill, “Nsa prism program taps in to user data of apple, google and others,” The Guardian, vol. 7, no. 6, pp. 1–43, 2013.
7. T. Suel and N. Memon, “Algorithms for delta compression and remote file synchronization,” 2002.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details