REDUCTION OF DATA LEAKAGE IN DISTRIBUTED CLOUD STORAGE SYSTEMS USING DISTRIBUTED CLOUD GUARD (DCG)

Meesala Sravani

Assistant Professor, Department of Computer Science and Engineering, GMRIT (JNTUK Affiliation), Rajam,
Vizayanagaram, Andhra Pradesh-532127, India
smeesala502@gmail.com
https://gmrit.edu.in/

Meesala Krishna Murthy

Assistant Professor, Department of Allied Health Sciences, Chitkara School of Health Sciences, Chitkara
University, Rajpura, Punjab - 140401, India
krishnameesala6@gmail.com
https://www.chitkara.edu.in/

Abstract

Cloud storage offers security, affordability, and global data access. Cloud storage is scalable, so organizations may simply add or delete storage. Cloud storage is convenient and safe. Dropbox, Google Drive, and Microsoft OneDrive allow cross-device data storage. Cloud storage providers (CSPs) also encrypt data. If data were dispersed across different CSPs, attackers would need to target numerous CSPs to retrieve the whole set. Hence, attackers struggle to obtain all the data. Email, cloud storage, and other methods can readily exchange data chunks without user consent. Data breaches can occur if security is inadequate. Because there are no access controls or insights into the data exchanged across clouds. Cyberattacks might disclose cloud data. Distributed Cloud Guard (DCG), a cloud security system, leverages advanced analytics to detect data flow irregularities like unlawful data exfiltration to solve this problem. We can then take immediate steps to prevent data leakage. Attackers would have to assault numerous clouds to access all semantically homogeneous data in the same cloud. DCG simplifies data leak detection and mitigation by centralizing data. This project uses Min-Hash and Bloom filter techniques to trademark data hunks for secure storage. Clustering lowers data leaks by distributing data hunks among clouds.

Keywords: **Cloud storage providers; numerous clouds; distributed cloud guard; data leakage**.

1. Introduction

As technology advances, the need for more powerful and capable storage solutions increases. This is because data-intensive applications, such as streaming media, require large amounts of storage, and cloud-based solutions are becoming increasingly popular [Buyya (2013)]. There are different cloud-based services and file sharing services available like Amazon servers, Drop Box, Microsoft One Drive, I Drive, Google Drive, Box, Apple iCloud Drive that are all becoming more popular for their easy end-user interfaces and cheap storage [Duffy and Muchmore (2023)] Centralized storage of data makes it vulnerable to cyberattacks from hackers, malware, and ransomware [Parker (2018); Buchanan (2020)]. It is possible for a single attack to compromise the data of all users, which could lead to the loss of data or the theft of private information [Parker (2018); Buchanan (2020)].

To overcome this problem, we have to spread user data across multiple clouds. This approach ensures that the data is stored redundantly across multiple locations and systems, making it less vulnerable to a single point of failure [Kumari (2021)]. It also makes it more secure since it would be much more difficult for a malicious actor to gain access to all of the user data in one go [Kumari (2021)]. In case of an attack on a cloud platform, the attacker will not get all the information about that user since they only store a small amount of data in one chunk of data [Mulazzani (2011); Prajapati (2022)]. This is because cloud platforms use encryption

DOI : 10.21817/indjcse/2023/v14i3/231403014

protocols to store data, which means that the data is fragmented and stored across multiple servers [Sanchez-Gomez (2017)]. This makes it difficult for an attacker to access all the data in one go, as they would need access to all the servers in order to do so [Sanchez-Gomez (2017)].

In order to locate the data of a single user, the attacker needs to know which files have been updated locally and where they were uploaded. This requires the attacker to use a hash function to compare the two files and determine which one is the most up-to-date [Mulazzani (2011)]. This is especially important when there are numerous files that need to be synchronized, as the hash line up function will help the attacker pinpoint the exact file that contains the user's data [Mulazzani (2011)].

2. Problem definition

Major headings should be typeset in boldface with the first letter of important words capitalized.

2.1. Existing System

The increasing amount of data being generated by businesses, organizations, and individuals [Kimble (2015)]. Cloud storage solutions offer a cost-effective way to store and access this data, as well as providing scalability and accessibility [Wu *et al.* (2010)]. The data providers could be biased and present biased data to influence the decisions they want to make. Furthermore, they may also have a conflict of interest in the data they provide and therefore may not be providing accurate information. Even though we spread our data across many CSPs, these storage systems follow simple algorithms [Koo *et al.* (2013)]. In some cases, when cloud services are unplanned and not carefully managed, they become vulnerable to security risks and data breaches, since any flaws in the system can be exploited. In addition, there is no way to guarantee that the data stored in the cloud is secure and that it will remain private [Kandukuri (2009)].

2.2. Proposed System

To overcome the drawbacks of traditional methods, including data leakage easily and data guessing in multiple clouds, we have proposed Distributed Cloud Guard (DCG). In DCG, we use breadth first search, min-hash, and sequential pattern clustering to achieve minimal leakage. As a result of these two mechanisms, we are able to reduce data leakage by 70 percent. This is extremely productive and well organized in terms of time and space compared to traditional methods.

3. Background and related work

A major priority is preserving our data, the Alexandrian imaging library. Traditionally, unauthorized access personnel created scandal lists. This was not possible today since gaps between the outline and the rest of the world were connected via the internet at no cost. Almost all computers and communication systems are both breakable and vulnerable, so the risk of data leakage and hacking is high. A single attack can remove all public data and expose private data to the world. In this paper the author considers the problem of providing flexibility against loss, and opposing unacceptable access as a dual problem. In order to obtain a better understanding of two obvious solutions to multiple technical problems, two different solutions can be combined into one [Crowcroft (2015)].

The rapidly growing proliferation of cloud storage services compiles the need for data storage based on their services, such as medical records, power systems, and financial databases. These data bases have critical data that could be moved to the cloud. Anyway, still reliability and security of data is the main problem. In this paper the author presents Epsky, a system that improves the availability, integrity and confidentiality of data stored in the cloud through the encryption, encoding and replication of data on diverse clouds and utilizes Planet lab to run clients accessing the service from different countries. We see that our protocols improved to recognize availability in major cases and the access latency when compared with cloud providers individually. The financial benefit of using DepSky in this case is twice the price of using a single cloud. This is optimal and seems to be a reasonable price, which provides the benefits [Bessani (2013)].

NC cloud was implemented on high of network-coding-based storage issue mention to the practical minimum storage make (FMSR) codes, which provide similar fault tolerance and knowledge duplication as in traditional codes like RAID-6.but use less repair traffic and, so need less financial value. Massive storage systems can be exposed to node failures. To provide failure tolerance, data is commonly encoded to prevent data duplication across numerous storage nodes. In the event of a failure of any node, it may be repaired by downloading data from living nodes and creating the last information on a completely new node. On the other hand, network cryptography is a subject that is investigated in theoretical [Chen *et al.* (2013)].

In this paper, the author presents SPAN Store, a key value store that exports a unified view of data providers in naturally distributed data centers. To reduce the cast of providers they combine three principals: 1) SPAN store spans multiple cloud providers to increase the naturally density of data centers and to minimize price by exploiting pricing discrepancies across providers, 2) by estimating application work load, and 3) lastly SPAN store minimizes the use of compute resources to implement tasks such as two-way locking and data propagation. The study shows that span stores can lower costs [Wu (2013)].

Using Scalia, the author introduced a cloud-breakage solution that continuously repositions information based on its access pattern and objective optimization. Scalia makes reference to selected objects that may be important, and it reduces the stage cost of the process by organizing the selected objects [Papaioannou (2012)].

4. System design

Figure. 1 showed the sequence diagram, which can be used to visualize how this system actually works and how actual communications occur between objects.

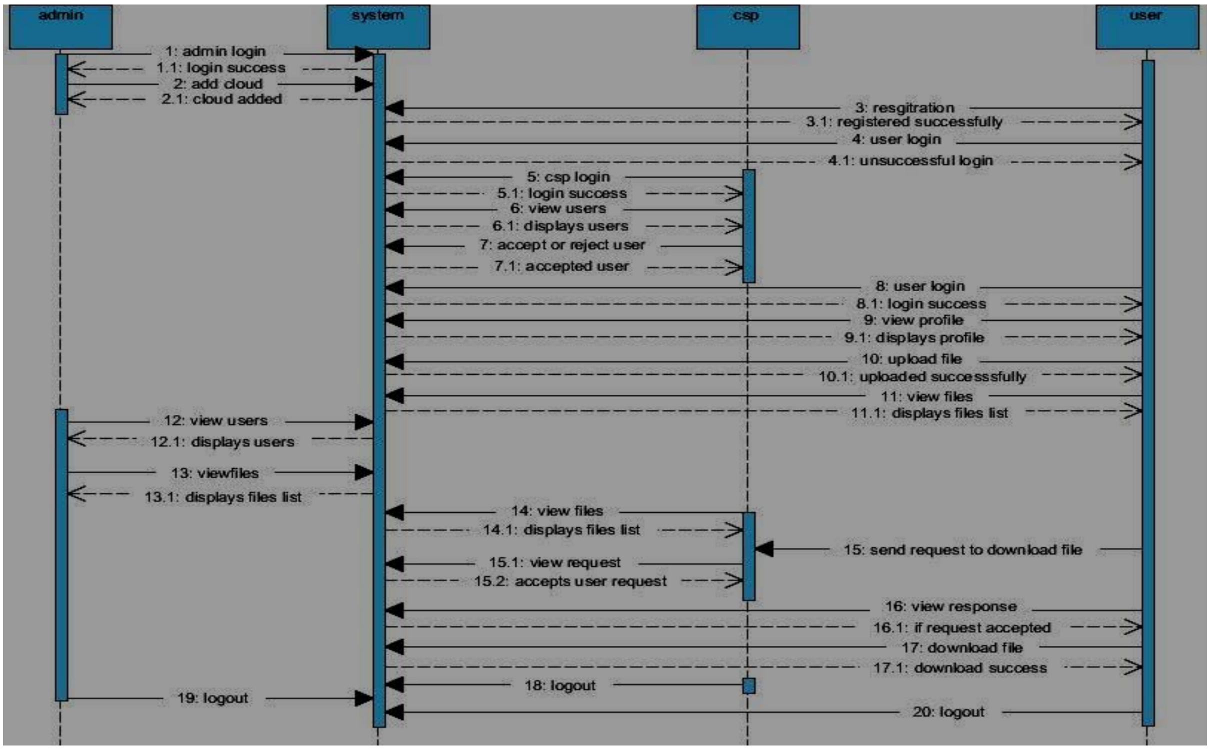
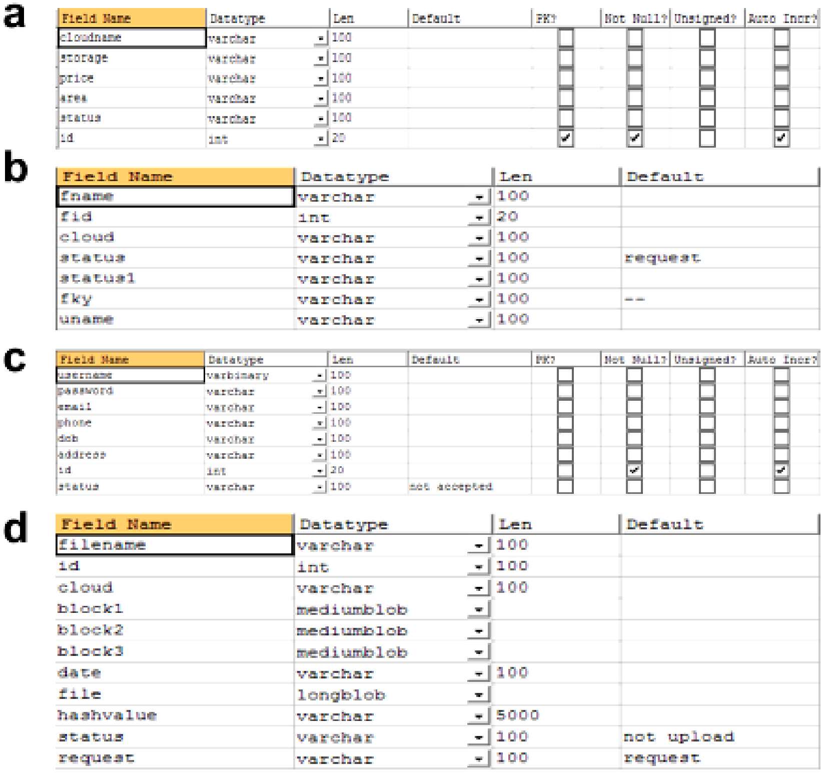


Figure 1: Illustration of system design

5. Database design

In figure 2, which cloud is available and what is the status for that cloud and how much data is already available in numerous clouds are shown.



Field Name	Datatype	Len	Default	PK?	Not Null?	Unsigned?	Auto Incr?
cloudname	varchar	100		□	□	□	□
storage	varchar	100		□	□	□	□
price	varchar	100		□	□	□	□
area	varchar	100		□	□	□	□
status	varchar	100		□	□	□	□
id	int	20		☑	☑	□	☑

Field Name	Datatype	Len	Default
fname	varchar	100	
fid	int	20	
cloud	varchar	100	
status	varchar	100	request
status1	varchar	100	
fky	varchar	100	--
uname	varchar	100	

Field Name	Datatype	Len	Default	PK?	Not Null?	Unsigned?	Auto Incr?
username	varbinary	100		□	□	□	□
password	varchar	100		□	□	□	□
email	varchar	100		□	□	□	□
phone	varchar	100		□	□	□	□
dob	varchar	100		□	□	□	□
address	varchar	100		□	□	□	□
id	int	20		□	☑	□	☑
status	varchar	100	not accepted	□	□	□	□

Field Name	Datatype	Len	Default
filename	varchar	100	
id	int	100	
cloud	varchar	100	
block1	mediumblob		
block2	mediumblob		
block3	mediumblob		
date	varchar	100	
file	longblob		
hashvalue	varchar	5000	
status	varchar	100	not upload
request	varchar	100	request

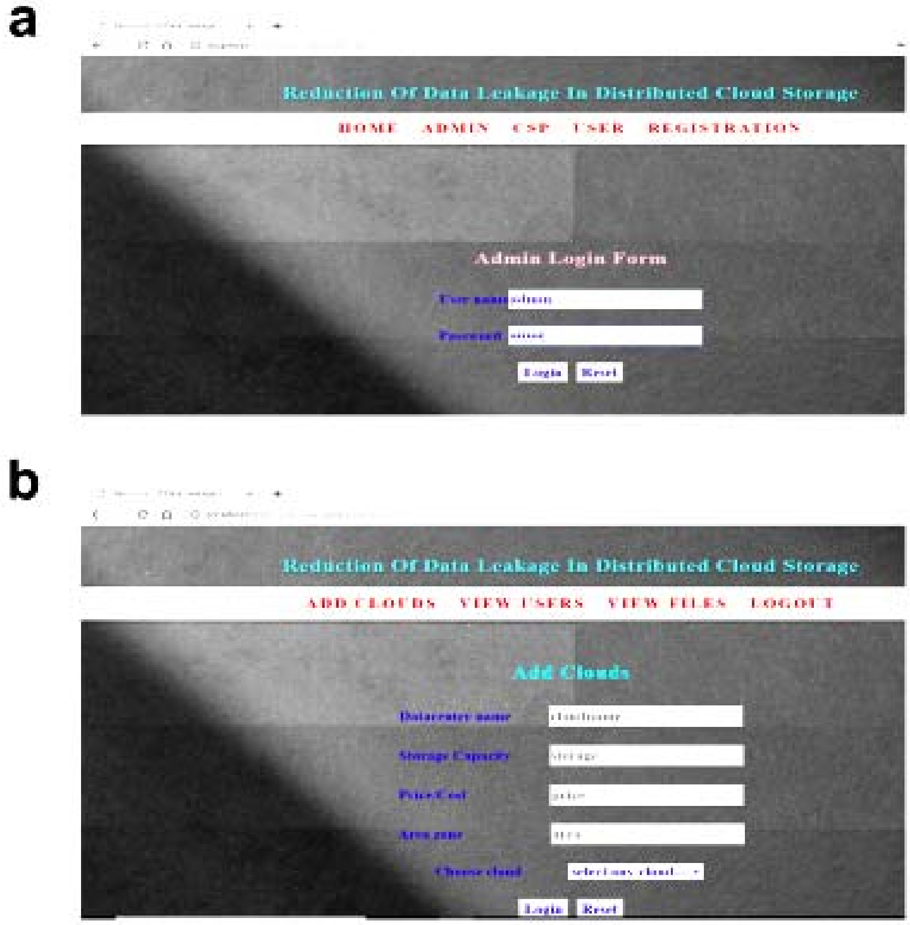
Figure 2: Database design. (a) Checking cloud, (b) Geniting file keys, (c) user registration, and (d) outcome after uploading the files

6. System implementation

To exciting this process, we are mainly taking three allotments that are Admin, CSP, and Users.

6.1. Admin

Figure 3 showed that the admin side admin maintains the web page that is related clouds and admin also having authentication for that web page. The admin only attaches empty clouds as well as checking how many users are available for a particular cloud and how many files are uploaded. Admin can login using admin credentials after that admin profile will open in which admin can add cloud and view users and view which files are updated. After that admin logout form profile.



Reduction Of Data Leakage In Distributed Cloud Storage

Admin Login Form

User name admin

Password •••••

Reduction Of Data Leakage In Distributed Cloud Storage

Add Clouds

Datacenter name cloudname

Storage Capacity storage

Price/Cost price

Area zone area

Choose cloud select any cloud...

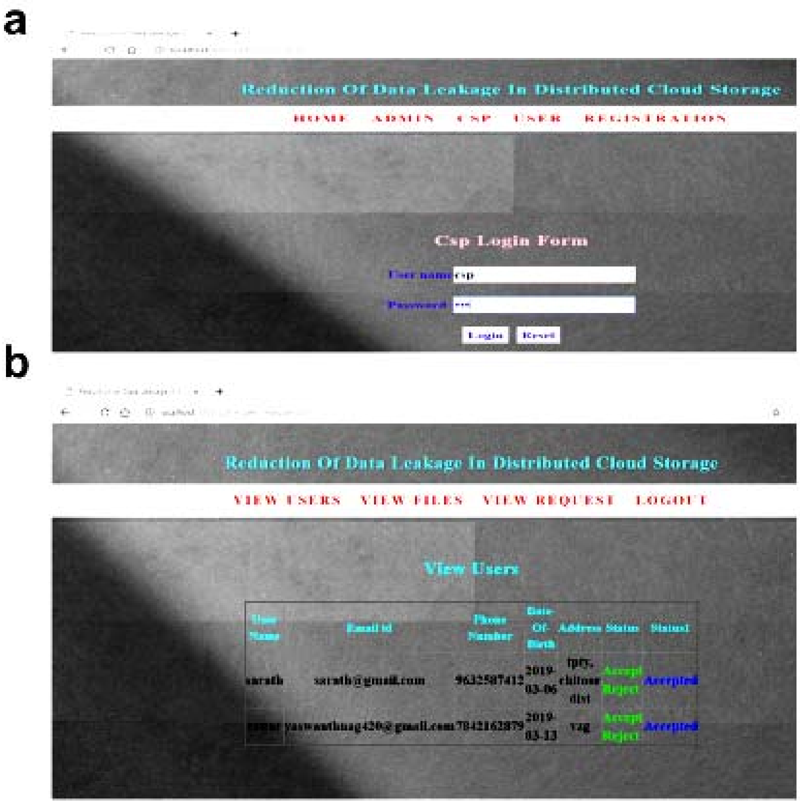
Figure 3: Admin system. (a) Admin login and (b) Admin Panel to do various operations

6.2. Cloud service provider (CSP)

Figure 4 showed that cloud service provider side is also where we need to use credentials. After that it checks which user request is presented and it maintains all the files which are uploaded by users. If the user request is accepted then only in CSP file we are able to see the data.

6.3. USER side operations

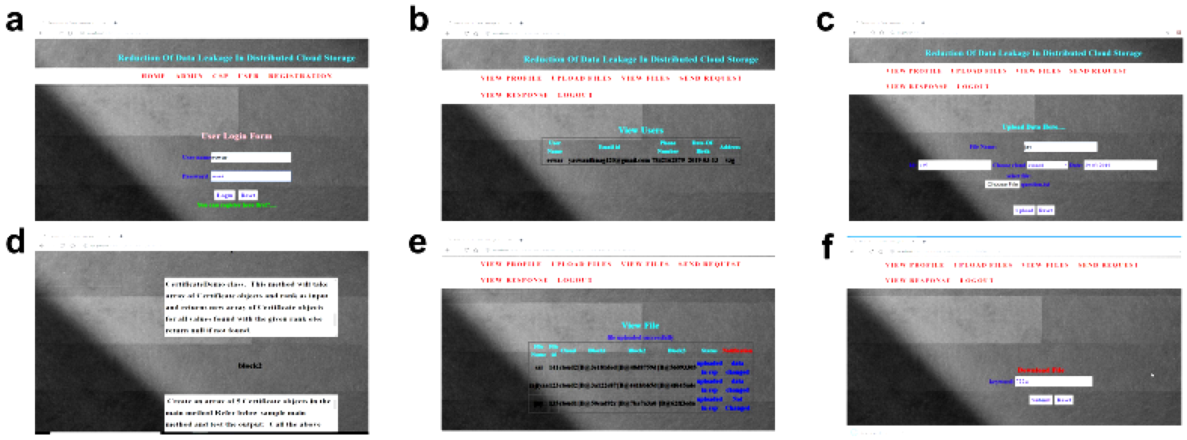
Figure 5 showed that user side web application resides in that user also need to login using user’s name and password after that user profile will open in that user checks all related user data which is need for basic verification after that user sends request for cloud storage and wait for response. After getting response user upload his data in appropriate manner and after that it stores in different hunks, user can also check file status is that file uploaded properly or not and any changes are made in that or not. If user want to download appropriate related data user need to send request to cloud service provider and after getting response user using his key only able to download his corresponding file successfully.


a
Reduction Of Data Leakage In Distributed Cloud Storage
HOME ADMIN CSP USER REGISTRATION
Csp Login Form
Login
Reset
b
Reduction Of Data Leakage In Distributed Cloud Storage
VIEW USERS VIEW FILES VIEW REQUEST LOGOUT
View Users

User name: csp
Password: •••

User Name	Email id	Phone Number	Date-Of-Birth	Address	Status	Status1
sarath	sarath@gmail.com	9632587412	2019-03-06	tpty, chittoor dist	Accept Reject	Accepted
[illegible]	yaswanthnag420@gmail.com	7842162879	2019-03-13	vzg	Accept Reject	Accepted

Figure 4: Cloud service provider (CSP). (a) user login and (b) user credentials after login


a
Reduction Of Data Leakage In Distributed Cloud Storage
HOME ADMIN CSP USER REGISTRATION
User Login Form
Login
Reset
b
Reduction Of Data Leakage In Distributed Cloud Storage
VIEW PROFILE UPLOAD FILES VIEW FILES SEND REQUEST
VIEW RESPONSE LOGOUT
View Users
c
Reduction Of Data Leakage In Distributed Cloud Storage
VIEW PROFILE UPLOAD FILES VIEW FILES SEND REQUEST
VIEW RESPONSE LOGOUT
Upload Data Here...
Choose File
Upload
Reset
d
block2
e
VIEW PROFILE UPLOAD FILES VIEW FILES SEND REQUEST
VIEW RESPONSE LOGOUT
View File
f
VIEW PROFILE UPLOAD FILES VIEW FILES SEND REQUEST
VIEW RESPONSE LOGOUT
Download File
Submit
Reset

User Name	Email id	Phone Number	Date Of Birth	Address
[illegible]	yaswanthnag420@gmail.com	7842162879	2019-03-13	vzg

CertificateDemo class. This method will take array of Certificate objects and rank as input and return new array of Certificate objects for all values found with the given rank else return null if not found.

Create an array of 5 Certificate objects in the main method Refer below sample main method and test the output. Call the above

Figure 5: User side operations. (a) user credential form for user login, (b) user profile 291 after login, (c) user files for uploading files, (d) user files in different hunks, (e) user 292 status, and (f) key for file download

Acknowledgement

Authors declared that the grate thanks towards the departmental deans to support the carried out this work.

Funding

No funding is provided for the preparation of manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare**.**

References

[1] Buyya, R., Vecchiola, C. and Thamarai Selvi, S., 2013. Chapter 8–Data-Intensive Computing: MapReduce Programming. Mastering Cloud Computing; Buyya, R., Vecchiola, C., Selvi, ST, Eds, pp.253-311.

[2] Duffy, J. and Muchmore, M., 2023. The best cloud storage and file-sharing services for 2023. https://www.pcmag.com/picks/the-best-cloud-storage-and-file-sharing-services.

[3] Parker, C. and Parker, C., 2018. Cybersecurity 101. Firewalls Don't Stop Dragons: A Step-by-Step Guide to Computer Security for Non-Techies, pp.17-67.

[4] Buchanan, B., 2020. The hacker and the state: Cyber attacks and the new normal of geopolitics. Harvard University Press.

[5] Kumari, P. and Kaur, P., 2021. A survey of fault tolerance in cloud computing. Journal of King Saud University-Computer and Information Sciences, 33(10), pp.1159-1176.

[6] Mulazzani, M., Schrittwieser, S., Leithner, M., Huber, M. and Weippl, E., 2011. Dark clouds on the horizon: Using cloud storage as attack vector and online slack space. Pp.1-11.

[7] Prajapati, P. and Shah, P., 2022. A review on secure data deduplication: Cloud storage security issue. Journal of King Saud University-Computer and Information Sciences, 34(7), pp.3996-4007.

[8] Sanchez-Gomez, A., Diaz, J. and Arroyo, D., 2018. Encrypted Cloud: A Software Solution for the Secure Use of Free-Access Cloud Storage Services. In International Joint Conference SOCO’17-CISIS’17-ICEUTE’17 León, Spain, September 6–8, 2017, Proceeding 12 (pp. 683-692). Springer International Publishing.

[9] Kimble, C. and Milolidakis, G., 2015. Big data and business intelligence: Debunking the myths. Global Business and Organizational Excellence, 35(1), pp.23-34.

[10] Wu, J., Ping, L., Ge, X., Wang, Y. and Fu, J., 2010, June. Cloud storage as the infrastructure of cloud computing. In 2010 International conference on intelligent computing and cognitive informatics (pp. 380-383). IEEE.

[11] Koo, D., Hur, J. and Yoon, H., 2013. Secure and efficient data retrieval over encrypted data using attribute-based encryption in cloud storage. Computers & Electrical Engineering, 39(1), pp.34-46.

[12] Kandukuri, B.R. and Rakshit, A., 2009, September. Cloud security issues. In 2009 IEEE International Conference on Services Computing (pp. 517-520). IEEE.

[13] Crowcroft, J., 2015. On the duality of resilience and privacy. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 471(2175), p.20140862.

[14] Bessani, A., Correia, M., Quaresma, B., André, F. and Sousa, P., 2013. DepSky: dependable and secure storage in a cloud-of-clouds. Acm transactions on storage (tos), 9(4), pp.1-33.

[15] Chen, H.C., Hu, Y., Lee, P.P. and Tang, Y., 2013. NC Cloud: A network-coding-based storage system in a cloud-of-clouds. IEEE Transactions on computers, 63(1), pp.31-44.

[16] Wu, Z., Butkiewicz, M., Perkins, D., Katz-Bassett, E. and Madhyastha, H.V., 2013, November. Spanstore: Cost-effective geo-replicated storage spanning multiple cloud services. In Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles (pp. 292-308).

[17] Papaioannou, T.G., Bonvin, N. and Aberer, K., 2012, November. Scalia: An adaptive scheme for efficient multi-cloud storage. In SC'12: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (pp. 1-10). IEEE.

Authors Profile

Mrs. Meesala Sravani Profile

Meesala Sravani is pursuing her PhD degree in Computer Science and Technology from GITAM University, India. She is now an assistant professor in GMRIT Engineering College at JNTU University. Her research interests are distributed storage systems including cloud storage systems and Big Data. He is now the group leader for several projects.

Dr. Meesala Krishna Murthy

Dr. Meesala Krishna Murthy received his PhD degree in Bioinformatics from Mizoram Central University, India in June 2022. He is now an assistant professor in Department of Allied Health Sciences, Chitkara School of Health Sciences, Chitkara University, Punjab, India. His research interests are genomics data analysis and Big Data.