Data Replication and Splitting in the Cloud for High Availability and Speed

Mrs. C. Sandhya[1], S. Sreejhari[2], C. Deekshitha[3], Ch. Sindhu Priya[4]

[1] Assistant Professor, Sridevi Women's Engineering College

[2,3,4,] Computer Science and Engineering, Sridevi Women's Engineering College, B.Tech IVYear Hyderabad, India

E-mail address: [1]chirrasandya@gmail.com

Abstract

The DROPS methodology was designed to address both security and performance issues in cloud computing. By breaking up a file into smaller pieces and replicating them across multiple nodes, the method ensures that even if one node is compromised, an attacker cannot access the entire file. This approach provides a higher level of security than traditional cryptographic techniques, which can be computationally expensive. To ensure optimal performance, the DROPS method uses centrality measurements to select nodes that are best suited for storing and retrieving data. Additionally, data duplication is used to speed up access to frequently accessed files. The effectiveness of the DROPS method was compared to 10 other replication strategies, and while it did result in a slight performance hit, it offered a higher level of security. Overall, the DROPS method provides a solution that balances both security and performance in cloud computing environments, ensuring that data remains secure while also allowing for efficient access and retrieval.

Keywords: Centrality, Cloud Security, Replication, Performance, Fragmentation.

1. INTRODUCTION

Cloud computing has significantly transformed the IT support structure and how it is utilized. The cloud computing model provides a range of features such as on-demand self-service, broad network connectivity, resource pooling, elasticity, and metered services, all of which have fundamentally changed the way organizations consume and manage IT resources. The cloud computing paradigm has enabled businesses to rapidly deploy and scale their applications and services, reduce capital expenditures, and benefit from more efficient use of IT resources, among other benefits. The benefits of cloud computing make it an attractive alternative for enterprises, groups, and individuals. Low cost, less maintenance (from the user's viewpoint), and enhanced adaptability come with new security risks. Security concerns are a major roadblock to the mainstream use of cloud computing. Security flaws in the cloud might originate in the underlying technology itself (virtual machine (VM) escape, session riding, etc.), Cloud-based service vulnerabilities (SQL injection, insufficient authentication, etc.), or in the unique properties of the cloud itself (data recovery vulnerability, Internet protocol vulnerability, etc.). All the nodes in a cloud need to be safe for the cloud to be secure.

Since there are several parts to a system, the most secure part can only be as secure as the least secure part. In a cloud environment, it is not only the user's responsibility to keep their data safe. The user's defenses might be compromised by the surrounding entities.

The use of off-site data storage clouds depends on the ability to transfer data in their virtualized and shared environment, which introduces a variety of security problems. Because of the pooling and adaptability of a cloud, physical resources may be shared between several users. Another issue is that even if data recovery methods are used, there is always the possibility that the shared resources would be redistributed to other users, which might lead to data corruption. It is also possible for a virtual machine in a multi-tenant virtualized environment to escape the control of its host machine (VMM). The freed virtual machine might cause havoc among other VMs and get access to sensitive information. Data privacy and integrity may also be at risk if virtualized networks are shared across tenants. Security measures must be taken to protect data stored in the public cloud from threats such as media sanitization failures that might compromise sensitive customers. Data must be protected against access by unauthorized parties, both accidentally and maliciously. A single vulnerable node may put the whole cloud in danger, as we've seen. For example, suppose

The security technique must make it very difficult for an attacker to get a usable quantity of data from the cloud once they've already breached the system. Hence, it's also important to minimize loss (caused by information disclosure).

The efficacy, reliability, and security of the cloud must be ensured. The throughput of a cloud storage service is heavily dependent on how quickly stored data can be retrieved. When used in large-scale systems, data replication processes improve data reliability, data availability, and response speed. Yet, data that is copied over several nodes are more vulnerable to attacks. The above statement highlights the trade-off between security and performance in cloud systems. Keeping multiple copies of a file in the cloud increases the chances of an attack on any one node, but it also improves performance by allowing for faster access and reduced network latency. Therefore, cloud systems must carefully balance security and performance to ensure both are adequately addressed. This is especially important given the massive scale of modern cloud systems, which makes them prime targets for cyber-attacks. As this is a difficulty with safely copying data, we approach it in the same way as the speed and security issues.

We present DROPS, a system that intelligently splits user data and then makes copies of them at strategic locations in the cloud to improve both speed and security. A file is split into smaller fragments that are intentionally meant to be useless on their own based on user input. The data encryption key is different for each cloud node (a node might be a computational, storage, physical, or virtual machine). Even if an attack on a single node is successful, the locations of further cloud components must be kept hidden. In order to boost security, we make sure that the nodes aren't too close together and are instead spread out across a certain distance from one another.

In the second stage, we select nodes for data replication based on the frequency of read/write requests. The nodes that serve the most read/write requests are identified, and data is duplicated among them to improve access times. This is done to optimize the performance of the system by ensuring that the most frequently accessed data is readily available on multiple

nodes. Overall, the two-phase node selection procedure aims to improve both security and performance in cloud storage systems.

2. RELATED WORK

"Quantitative comparisons of the state-of-the-art data centre architectures"

The number of servers that are part of a network in a data centre has increased dramatically in recent years. Network infrastructure is a major factor in deciding how much money has to be spent on a data centre and how to measure its performance. Yet, the data center's growth pattern and aggregate bandwidth needs are above the capacity of the data center's legacy data centre network (DCN) architecture. Even when the most advanced corporate networking hardware is deployed, only about half of the total network bandwidth is made available at the network's periphery. In order to meet the expanding needs of the "cloud computing paradigm," new DCN designs are required to address the critical problems encountered by the existing architecture.

Using network monitoring We compared the effectiveness of the three most recent DCN models in terms of throughput (a), average packet delay (b), and hybrid (c) models. Further research on data centre networking (DCN) suited topologies and addressing protocols in large-scale data centres may utilize the aforementioned analyses as a benchmark. We have conducted extensive simulations under different network traffic patterns to evaluate the relative merits of the different DCN architectures. In addition, we provide a firm foundation for further research and development of DCN designs.

We compared the most popular DCN designs that deal with network scalability and oversubscription problems. We ran simulations to see how the most popular DCN designs fared in a wide range of realistic settings and with a variety of network topologies. In terms of average network performance and packet latency, The simulation findings showed that when comparing the DCell and three-tier DCN designs, the fat-tree-based DCN design performed the best.

A future goal of ours is to compare the DCell-optimized routing scheme to the shortest route routing-based protocols. In addition, we are enthusiastic about developing better routing strategies to make DCN systems more environmentally friendly. In addition, we want to implement some of the capabilities related to consolidating workloads. routing systems that use dynamic power management-based methods to take advantage of idle and under-utilized connections to save power.

"On the Characterization of the Structural Robustness of Data Centre Networks"

In the realm of information and communication technology (ICT), data centres play a crucial role as both a structural and operational building element of cloud computing. Research, data storage, and analysis are only some of the many uses for cloud computing in fields as diverse including but not limited to farming, nuclear power, smart grids, medicine, and internet search engines. A DCN is the communication backbone of a data centre and determines the limits of cloud computing in terms of performance. To provide the specified Quality-of-Service (QoS) and meet SLA requirements, the DCN must be resilient to failures and uncertainty (SLA). Our primary contributions are the following: We present multilayered graph models of various DCNs; we analyze classical robustness metrics by taking a variety of

failure scenarios into account; we demonstrate that traditional measures of network robustness are inadequate for gauging DCN resilience; and finally, we propose alternative methods for quantifying DCN resilience. There hasn't been a comprehensive look into DCN resilience yet. Hence, we believe this study will provide a firm foundation for future research into the robustness of DCNs.

The reliability of modern DCN designs in terms of their underlying infrastructure. In contrast to the Fat Tree and Three Tier designs, the DCell architecture shows smooth degradation across all failure scenarios. The findings showed that traditional robustness indicators underestimated the DCN's resilience because of its unique connection pattern, multi-layered design, and diverse components. This indicates and motivates the search for alternative robustness measures with which to quantify the DCN.

To quantify DCN hardiness, we introduced a degradation metric. Degradation measures the network's steadiness by calculating the extent to which its graph structure has changed over time. Based on the results of the degradation metrics, DCell was shown to be the most durable design of the DCNs that were studied. The DCN robustness research showed that conventional robustness measurements are inadequate for DCN layouts. DCN architectures are unusual in that they are hierarchical and pattern-based in their connections.

Metrics for robustness are required. In addition, it is important to analyze network traffic and performance under different failure conditions involving DCN components. The expenses increase as network resilience rises. To provide the requisite degree of resilience at the lowest possible cost, new cost-effective robust DCNs are required. For DCN performance quantification, it is also important to think about the network's performance characteristics, such as its bisection bandwidth and bottleneck degree.

"Appreciating the Risks of Cloud Computing"

It is challenging to construct a well-founded appraisal of the security implications of cloud computing in light of the current discourse about cloud computing security concerns for two key reasons. Risk, danger, and vulnerability are often used interchangeably without giving each concept the attention it deserves in discussions about risk. Second, keep in mind that not all worries are particular to cloud computing. The "delta" in security issues raised by cloud computing may be approximated by looking at how cloud computing impacts certain threats. When it comes to security flaws, cloud computing exacerbates the effects of well-known ones while introducing new ones. Here, we learn to recognize four vulnerabilities unique to the cloud, are introduced to a cloud reference architecture designed with security in mind, and experience actual cases of cloud-specific weaknesses across the architecture's many levels.

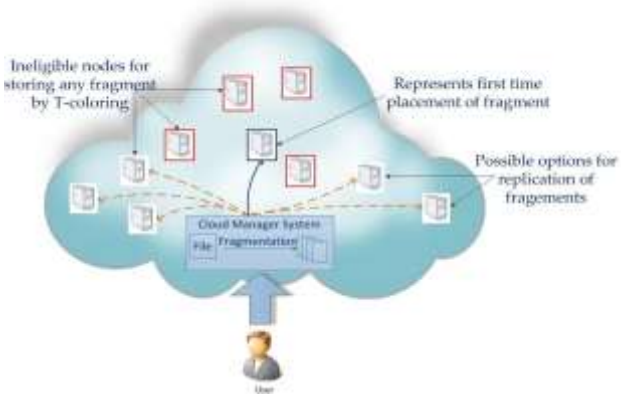
"Intrusion tolerance in distributed computing systems"

A distributed system that is intrusion-tolerant is one that is built to ensure that confidentiality, integrity, and availability won't be compromised in the event of an intrusion into any one of its components. This strategy is appropriate for distributed systems because distribution allows for element isolation, limiting the physical access that an intrusion can grant to the system's components.

The authors describe how intrusion-tolerant authentication and authorization servers enable the implementation of a consistent security policy on a collection of untrusted sites managed by non-conspiring individuals. They explain how some distributed system components can be designed to withstand intrusions. A prototype of the persistent file server presented in this

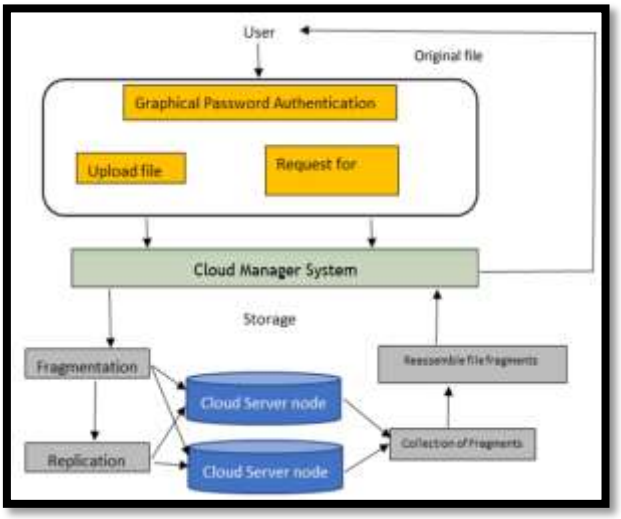
paper was successfully developed and deployed as part of the Delta-4 project under the European ESPRIT program.

3. METHODOLOGY



In order to assess the effectiveness of the DROPS method, we conducted experiments using 10 different replication mechanisms based on heuristics. These mechanisms included various fine-grained replication strategies, such as DRPA-star, WA-star, A-star, SA1, SA2, SA3, and GRA, which aimed to optimize replica numbers and locations while considering system performance. The experiments were carried out on three different Data Center Network (DCN) topologies, namely Three-tier, Fat tree, and DCell, which are commonly used in cloud computing. Since the DROPS method is designed to function within the cloud computing paradigm, these structures were used to ensure that the evaluation was conducted under realistic conditions.

4. SYSTEM DESIGN ARCHITECTURE



A cloud client is either a data owner or a data user. A data owner is a person who uploads files to the cloud. Data owner knows fragment placement with their node numbers. Anyone who downloads or views the files that others have uploaded is referred to as a data user. If a

person attempts to access a file without first providing authentication, that person is regarded as an intruder or attacker.

When a client uploads a file, the cloud server handles a variety of tasks, including encryption, fragmentation, nodal allocation, replication, and further nodal allocation of replicas. The cloud server combines and decrypts file fragments when a client asks to download an uploaded file.

In the cloud, data centre network (DCN) is used for communication. Many DCN architectures exist, including the three-tier, Dcell, fat tree, and others. In this study, the three-tier architecture is primarily used.

5. MODULES

1. Authentication - User login and registration have been completed.

2. Upload File and Generate Fragments - We create Cloud and User entities. A user can upload a new File and update previously uploaded File blocks in the user entity.

3. Replicate Fragments - Data replication is carried out by storing copies of the data in various clouds.

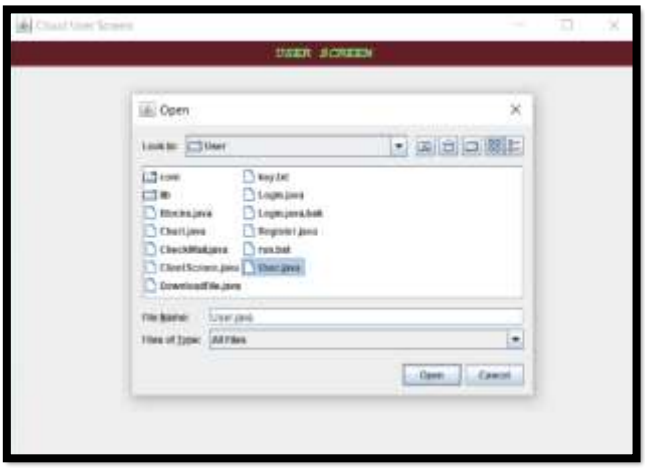
4. Download Fragments - When necessary, the user can download the file.

5. RC Versus File Fragments Graph - The graph shows that replication time is reduced when compared to greedy time.

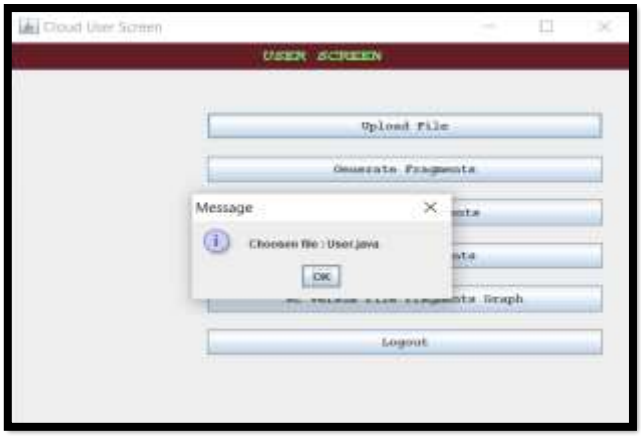
6. RESULT AND DISCUSSION

The user must register if the user is new to the server to access the features and modules present in the cloud user screen.

The user must log in with the registered details for accessing the modules present in the cloud user screen.



On clicking the upload file module, a file is selected to upload to the cloud.



Chosen file is successfully uploaded.

On clicking generate fragments module, the uploaded file is successfully generated into fragments.

The downloaded file is saved in D-Drive.

7. CONCLUSION

The use of the DROPS methodology enabled the creation of a cloud storage security plan that focused on both security and retrieval time performance. To achieve this, the data file was divided into fragments and dispersed across multiple nodes. This ensured that in the event of a successful attack, the adversary would not be able to access significant information due to the fragmentation and dispersal of the data. The evaluated system ensured that each file was stored on a single cloud node, and both full-scale replication techniques and the DROPS methodology were evaluated for their performance. The simulation results indicated that when security and performance were given equal priority, the security level of the data was enhanced, albeit with a slight decrease in performance.

8. FUTURE SCOPE

A user must download the file, make any necessary updates, and then re-upload it using the DROPS methodology. Creating a system for automatic updates that can only recognize and update the necessary fragments is a wise move. Similar work to that described above will

help conserve time and resources that are typically expended in the process of downloading, updating, and re-uploading files.

References

[1] A. Juels and A. Opera, "New approaches to security and availability for cloud data," Communications of the ACM, Vol. 56, No. 2, 2013, pp. 64-73.

[2] W. A. Jansen, "Cloud hooks: Security and privacy issues in cloud computing," In 44th Hawaii IEEE International Conference on System Sciences (HICSS), 2011, pp. 1-10.

[3] B. Grobauer, T.Walloschek, and E. Stocker, "Understanding cloud computing vulnerabilities," IEEE Security and Privacy, Vol. 9, No. 2, 2011, pp. 50-57.

[4] Y. Deswarte, L. Blain, and J-C. Fabre, "Intrusion tolerance in distributed computing systems," In Proceedings of IEEE Computer Society Symposium on Research in Security and Privacy, Oakland CA, pp. 110-121, 1991.

[5] L. M. Kaufman, "Data security in the world of cloud computing," IEEE Security and Privacy, Vol. 7, No. 4, 2009, pp. 61-64.

[6] S. U. Khan, and I. Ahmad, "Comparison and analysis of ten static heuristics-based Internet data replication techniques," Journal of Parallel and Distributed Computing, Vol. 68, No. 2, 2008, pp. 113-136.