CHAPTER 7

CONCLUSION AND FUTURE WORK

This thesis investigates a bloom filter based metadata service and discussed in detail all issues the primary issues leading to the design of MaaS model. Due to the presence of large amount of data, the proposed metadata model has a promising, prospect in this growing and challenging domain. The references made here are by no means exhaustive as many research works in this domain are still under way. This research proposes that the success of the data retrieval in large scale distributed systems like cloud is dependent on the metadata model, which improves the efficacy of data retrieval by means of reduction in latency, increase in throughput and improved security. The present study has identified “MaaS”, Metadata as a Service for the utilization of large scale distributed systems in cloud, which provides an efficient way of retrieving data in a secured manner from huge amount of data servers dispersed geographically.

First, the overall cloud metadata architecture model is studied. The architecture has three different layers. The top most is the Application layer, middle layer is the “MaaS” - Metadata as a Service layer and bottom layer is the Data server layer. The claim that metadata is the building layer of meta-information for filtering, search and retrieving data in a secured way from huge data servers is established.

A novel method of metadata management mechanism has been proposed using a specially designed bloom filter called cloud bloom filter
(CBF), a probabilistic data structure, which is used for efficient retrieval of data stored across various data storage servers located in geographically dispersed locations. While GBF efficiently handles scalable and decentralized metadata placement schemes, reducing latency effectively, LBF handles the dynamic metadata efficiently with effective update mechanism and hence reduces the presence of stale data at the MDS location. Experiments have been carried out and the results show that the LBF update holds good in enhancing the consistency of the metadata file stored by means of reducing the error rate and decreasing the false alarm. The effective use of overall CBF in the proposed model increases the efficacy of the model. The influence of metadata refines the searching of data in a cloud environment by reducing the latency. The role of metadata in further improving the efficiency of MaaS is carried by means of reducing the search space has been proposed next.

Clustering has been found to play a major role in analysing the metadata for data retrieval tasks. Two mechanisms viz. Base Cluster and Derived Cluster which represent the time slot based transactions and user based transactions have been used and statistical analysis have been applied. Use of BMF and finding frequent transaction set using MFT algorithm have been shown to improve efficiency. The use of BMF reduces the complexity of scanning large distributed database, by means of having an independent probabilistic lookup. Thus, by making use of BMF and MFT in the proposed model and the use of derived cluster increases the efficacy of the proposed model by means of reducing the search space dimension. The results of metadata analysis model show that the time taken for metadata lookup is reduced due to the search space reduction and thus results in reduction in latency during data retrieval. Apart from using metadata for latency reduction the thesis further explored the possibility of using metadata to security during data retrieval thereby providing a complete model.
A security scheme is proposed which is best suited for cloud storage systems with huge volume of data stored and guarantees data security and user privacy during the data retrieval process. The key generation and issuing protocol is handled at various levels, User level, MDS level and DS level. The model also makes the data owner, feeling confident about the complete security of the data stored, since the encryption and decryption keys cannot be compromised without the involvement of data owner. Our security model allows key based policies to be enforced on the encrypted data stored at cloud servers. Based on the proposed scheme, the thesis presents a secure metadata model architecture that allows the user to store data securely in a cloud scenario. Analysis of the proposed scheme shows that the keys are efficiently distributed and the proposed scheme is analyzed in terms of correctness, security, and efficiency.

Thorough testing of the “MaaS” model has been carried out to check the efficiency of the proposed model. The promising results show that the use of metadata leads to the success of fast retrieval of data in cloud environment. Further, the model performed better than the traditional models which are evident from the evaluation results.

Thus, this thesis explored the role of metadata which has a large impact on reducing latency and providing security to the data through metadata. There also exists much scope for the future work in the form of combining homomorphic encryption technique with the proposed metadata inspired retrieval model. The process of refining the security scheme is to perform data operations directly on cipher text in the cloud without the need to decrypt the data and hence improves the security. Homomorphic security in cloud storage needs a greater level of investigation, and has great future when combined with metadata.