

PROTECTION OF BIG DATA USING THIRD PARTY AUDITING AND DATA DE-DUPLICATION

Shubham Bhagwat¹, Prof Yashanjali Sisodia²

G. H. RAISONI COLLEGE OF ENGINEERING CHAS, AHMEDNAGAR¹²

bhagwatshubham@gmail.com¹, yashanjali.sisodia@raisoni.net²

Abstract: In recent years, huge information became a hot analysis topic. The increasing quantity of huge knowledge additionally will increase the prospect of cracking the privacy of people. Since huge private information need high procedure power and massive storage, distributed systems are used. As multiple parties are concerned in these systems, the danger of privacy contravention is accumulated. There are different types of privacy-preserving mechanisms developed for privacy protection at totally different stages (e.g., knowledge generation, knowledge storage, and knowledge processing) of a giant knowledge life cycle. The main goal of this paper is to provide a comprehensive summary of the privacy preservation mechanisms in huge knowledge and gift the challenges for existing mechanisms. Above all, during this paper, we tend to explain the infrastructure of massive knowledge and therefore the onward privacy-preserving mechanisms in every stage of the massive knowledge life cycle. What is more, we tend to discuss the challenges and future analysis directions associated with privacy preservation in huge knowledge

I INTRODUCTION

Cloud service providers manage an enterprise-class infrastructure that offers a scalable, secure and reliable environment for users, at a much lower marginal cost due to the sharing nature of resources. It is routine for users to use cloud storage services to share data with others in a team, as data sharing becomes a standard feature in most cloud storage offerings, including Drop box and Google Docs. Many mechanisms have been proposed to allow not only a data owner himself but also a public verifier to efficiently perform integrity checking without downloading the entire data from the cloud, which is referred to as public auditing. With evolution of computers the life of people became more and more easily. They were able to keep their data on their devices, and started finding ways to make them accessible to others, for example say by using floppy, writable disks, which was followed by portable hard-disk, all these were expensive in their own way during their time. The data was very much private on personal devices like PC, laptops, mobile phones etc., therefore sharing data with others was considered to be expensive. As the world of computing got more advanced the ways for sharing data started becoming cheaper and cheaper. In recent years a new term has evolved called "Cloud" which is provided by different providers, and which is nothing but facility or service of different resources or components like hardware, platform, storage's, software etc., and it is gaining importance because it frees the user from maintenance perspective on an investment of some money for the use of these services provided by cloud service providers.

II LITERATURE REVIEW

Protection Of Big Data Privacy AUTHORS: Abidmehmood, iynkarannatgunana than yongxiang Provide a comprehensive overview of the privacy preservation mechanisms in big data and present the challenges for existing mechanisms. Illustrate the infrastructure of big data and the state-of-the-art privacy-preserving mechanisms in each stage of the big data life cycle. • Big data security and privacy: A review AUTHORS: B. Maturdi, X. Zhou, S. Li, and F. Lin Reviewed the enormous benefits and challenges of security and privacy in Big Data & present some possible methods and techniques to ensure Big Data security and privacy. • Toward scalable systems for big data analytics: A technology tutorial AUTHORS: H. Hu, Y. Wen, T.-S. Chua, and X. Li Present a literature survey and system tutorial for big data analytics platforms, aiming to provide an overall picture for non-expert readers and instill a do-it-yourself spirit for advanced audiences to customize their own big-data solutions. • Big data: Issues, challenges, tools and good practices. AUTHORS: A. Katal, M. Wazid, and R. H. Goudar. Introduces the Big data technology along with its importance in the modern world and existing projects which are effective and important in changing the concept of science into big science and society too. The various challenges and issues in adapting and accepting Big data technology, its tools are also discussed in detail along with the problems Hadoop is facing. • Information security in big data: Privacy and data mining AUTHORS: L. Xu, C. Jiang, J. Wang, J. Yuan, Y. Ren We introduce View the privacy issues related to data mining from a wider perspective and investigate various approaches that can



Figure 5. File uploaded



Figure 6. TPA request to Cloud



Figure 7. Apply Decryption key



Figure 8. Final Output (Download File)

VIII CONCLUSION

We have seen how delegation of responsibility trusted 3rd party which provides security services secures user data. It relieves the client from maintaining any kind of key information and

allowing the client for using any browser enabled device to access the cloud services. It allows the client to verify the integrity of the data stored on download or retrieval of its own stored data in cloud. The client can share the data securely with specific band of people without any overhead of key distribution. We are providing concept of deduplication which operates at the file or block level. File deduplication eliminates duplicate files, but is not an efficient means of deduplication. Future Work. 1. To enhance the security more, a mechanism to secure the keys in security cloud can be a area of research. 2. To reduce the overhead of network traffic can be another area of research..

REFERENCES

1. Abidmehmood, iynkarannatgunanathanyongxiang (senior member, ieee), gaughua (member, ieee), Society and internet hsi'03, pp. 302-311,2016.
2. J. Manyika et al., Big data: The Next Frontier for Innovation, Competition, and Productivity. Zurich, Switzerland: McKinsey Global Inst., Jun. 2011, pp. 1137.
3. B. Maturdi, X. Zhou, S. Li, and F. Lin, "Big data security and privacy: A review," China Commun., vol. 11, no. 14, pp. 135145, Apr. 2014.
4. J. Gantz and D. Reinsel, "Extracting value from chaos," in Proc. IDC IView, Jun. 2011, pp. 112.
5. A. Katal, M. Wazid, and R. H. Goudar, "Big data: Issues, challenges, tools and good practices," in Proc. IEEE Int. Conf. Contemp. Comput., Aug. 2013, pp. 404409.
6. L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information security in big data: Privacy and data mining," in IEEE Access, vol. 2, pp. 11491176, Oct. 2014.
7. H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," IEEE Access, vol. 2, pp. 652687, Jul. 2014.
8. Z. Xiao and Y. Xiao, "Security and privacy in cloud computing," IEEE Commun. Surveys Tuts., vol. 15, no. 2, pp. 843859, May 2013.
9. C. Hongbing, R. Chunming, H. Kai, W. Weihong, and L. Yanyan, "Secure big data storage and sharing scheme for cloud tenants," China Commun., vol. 12, no. 6, pp. 106115, Jun. 2015.
10. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multikeyword ranked search over encrypted cloud data," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222233, Jan. 2014.
11. O. M. Soundararajan, Y. Jenifer, S. Dhivya, and T. K. P. Rajagopal, "Data security and privacy in cloud using RC6 and SHA algorithms," Netw. Commun. Eng., vol. 6, no. 5, pp. 202205, Jun. 2014