the data sets which incur less cost but disclose more privacy-sensitive information. Thus, higher PLAVG means more data sets in F_i should be encrypted to preserve privacy from a global perspective. Based on the above analysis, the heuristic value of the search node SNi can be computed by the formula: f SNi ð Þ¼Ccur=ð" _ "iþ1Þ þ ð"iþ1 _ Cdes _ BFAVGÞ=PLAVG

9. Results

The comparison of privacy preserving cost for encrypting all the intermediate datasets [11] in existing system and encrypting only part of intermediate datasets in our approach shows that we are reducing the privacy p reserving cost by using our approach as shown in the figure7

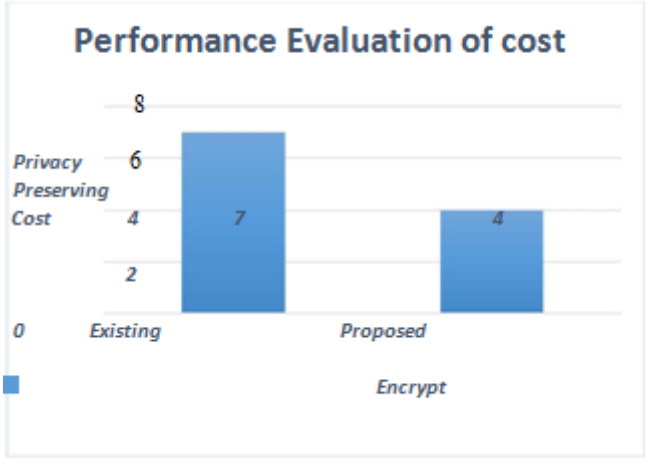

Figure 7: Reducing the privacy preserving cost by our approach

In the figure7 the vertical axis represents the cost required to encrypt the datasets and the horizontal axis shows the two categories existing and proposed, the shaded bars in the graph shows the encryption cost required. By observing the existing and proposed encryption costs we can evaluate the performance of our approach. By using our approach we can also prove that the time consuming is very less for encrypting only part of intermediate datasets compared with the existing approaches can be shown in the figure 8

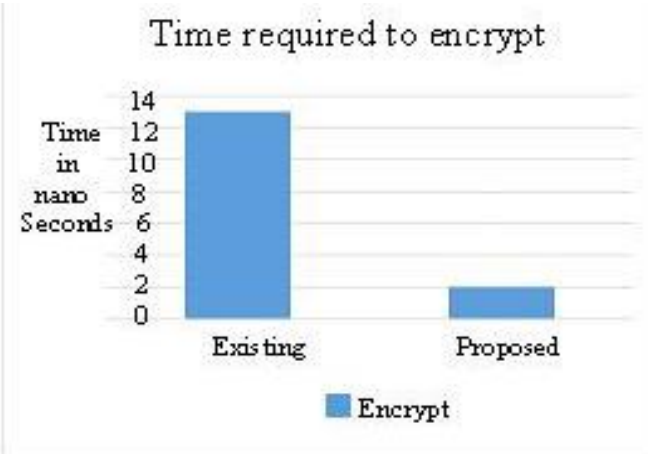

Figure 8: Result of time comparison for existing and our approach

In the figure8 the vertical axis represents the time required to encrypt the dataset in nanoseconds and the horizontal axis has two categories existing and proposed and t he shaded bars shows the encrypt time. We can easily analyze the time required to encrypt the datasets in the existing and our approach.

By comparing the cost for encrypting all the intermediate datasets and only part of intermediate datasets in the cloud we are saving the privacy preserving cost it can be shown in the following equation.

CSAV=CALL-CHEU

Here CSAV is the privacy preserving cost saved, CALL is the privacy preserving cost for encrypting all the intermediate datasets and CHEU is the privacy preserving cost for encrypting only part of intermediate datasets in the cloud. The resultant of our approach shows that the saving cost should be increases going on increasing the threshold value.

10. Conclusion

In accordance with various data and computation intensive applications on cloud, intermediate data set management is becoming an important research area. Privacy preserving for intermediate data sets is one of important yet challenging research issues, and needs intensive investigation. With the contributions of this paper, we are planning to further investigate privacy aware efficient scheduling of intermediate data sets in cloud by taking privacy preserving as a metric together with other metrics such as storage and computation. Optimized balanced scheduling strategies are expected to be developed toward overall highly efficient privacy aware data set scheduling. We have proposed an approach that identifies which part of intermediate data sets needs to be encrypted while the rest does not, in order to save the privacy preserving cost. A tree structure has been modeled from the generation relationships of intermediate data sets to analyze privacy propagation among data sets. We have modeled the problem of saving privacy-preserving cost as a constrained optimization problem which is addressed by decomposing the privacy leakage constraints.

References

[1] M.Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A.Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M.Zaharia, “A View of Cloud Computing,” Comm. ACM, vol. 53,no. 4, pp. 50-58, 2010

[2] D. Yuan, Y. Yang, X. Liu, and J. Chen, “On-Demand Minimum Cost Benchmarking for Intermediate Data Set Storage in Scientific Cloud Workflow Systems,” J. Parallel Distributed Computing,vol. 71, no. 2, pp. 316-332, 2011

[3] D. Zissis and D. Lekkas, “Addressing Cloud Computing Security Issues,” Future Generation Computer Systems, vol. 28, no. 3, pp. 583-592, 2011.

[4] "Encryption Basics | EFF Surveillance Self-Defense Project." Encryption Basics | EFF Surveillance Self-Defense Project. Surveillance Self-Defense Project, n.d. Web. 06 Nov. 2013. https://ssd.eff.org/tech/encryption.

[5] Goldreich, Oded. Foundations of Cryptography: Volume 2, Basic Applications. Vol. 2. Cambridge university press, 2004
[6] Bellare, Mihir. "Public-Key Encryption in a Multi-user Setting: Security Proofs and Improvements." Springer Berlin Heidelberg, 2000. Page 1.
[7] Damien Stehle; Ron Steinfeld (2010-05-19). "Faster Fully Homomorphic Encryption" (PDF). International Association for Cryptologic Research. Retrieved 2010-09-15.
[8] Microsoft HealthVault, http://www.microsoft.com/health/ww/products/Pages/healthvault.aspx, July 2012.
[9] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam "L-Diversity: Privacy Beyond K-Anonymity," ACM Trans. Knowledge Discovery from Data, vol. 1, no. 1, article 3, 2007.
[10] S.Hemalatha, S.Alaudeen Basha "Enabling for Cost-Effective Privacy Preserving of Intermediate Data Sets in Cloud" International Journal of Scientific and Research Publications, Volume 3, Issue 10, October 2013
[11] C. Lakshmi, " An Approach for Privacy Preserving Cost of Intermediate Data Set in Cloud", International Journal of Advanced Research in Computer Science and Software Engineering 4(5), May-2014,pp107-111
[12] W. Du, Z. Teng, and Z. Zhu, "Privacy-Maxent: Integrating Background Knowledge in Privacy Quantification," Proc. ACMSIGMOD Int'l Conf. Management of Data (SIGMOD '08), pp. 459-472, 2008.

Author Profile

Mr. Ravindra Suresh Kamble student of MLR Institutions of Technology, Hyderabad pursuing M. Tech degree in Computer Science and Engineering.

Sheikh Gouse B.Tech, M.Tech CSE. He is currently working in the Department of Computer Science and Engineering, MLRIT, Telangana, India. He is having 8 years of teaching experience**.** He is Certified in Oracle 9i: SQL & Java SE 6. His research interesting areas Programming (C&DS, C++, and JAVA), Data Mining, Software Engineering, Network Security & Computer Networks.