to know the presence or absence of web access from specific IP address such as 104.40.128.114. To avoid this, the count function pattern is understood using a decompiler tool [41] and applied the proposed algorithm. The results are presented in the Table 1.

IP	Crawler	Genuine Count	Count in Presence of Attacker
101.81.76.106	0	12455	12454
104.40.128.107	0	9546	9545
104.40.128.108	0	10339	10340
104.40.128.109	0	12350	12349
104.40.128.110	0	14452	14451
104.40.128.111	0	7480	7479
104.40.128.112	0	13987	13986
104.40.128.113	0	12894	12893
104.40.128.114	0	13654	13653

As the algorithm applied its differential privacy, it is able to provide different value when there is attacker presence or in the presence of malicious code. The adversary thus cannot identify the presence or absence of the target IP, 104.40.128.114 availability in the data. Thus the differential privacy algorithm can protect data from malicious map and reducer codes. The differential privacy related noise is applied to actual count. The application for the value 13654 for an IP address 104.40.128.114 is as follows.

$\boldsymbol{\varepsilon}$ = 8.85x10^{-12}

=count+ [(1+ $\boldsymbol{\varepsilon}$)+R]

=13654+[(1+8.85x10^{-12})-2.00000000001]

=13653

9. CONCLUSION AND FUTURE WORK

Privacy issues of big data are studied in the presence of untrusted mapper and reducer. The malicious code used for mapper and reducer can lead to sensitive data leakage and misuse of information. As cloud computing became reality, enterprises are moving their data to cloud where storage and processing takes place. With this phenomenal change in computing, cloud also brings about privacy challenges. Specifically, MapReduce paradigm in distributed programming frameworks like Hadoop can cause the disclosure of sensitive information when mapper or reducer is under an influence of attack. In this paper a methodology is proposed for secure and privacy preserving computations in MapReduce framework. The methodology is based on our differential privacy algorithm. The methodology is realized in the MapReduce framework of Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3). Our empirical study revealed that our methodology is useful in privacy preserving big data mining. This research can be extended to have further optimization of security and privacy to MapReduce programming in the presence of untrusted mapper and reducer.

REFERENCES

[1] Xiaokui Xiao, Guozhang Wang and Johannes Gehrke. (2009). Differential Privacy via Wavelet Transforms. IEEE, p1-15.

[2] Chao Liy, Michael Hayy, Vibhor Rastogiz, Gerome Miklauy, Andrew McGregor. (2010). Optimizing Linear Counting Queries Under Differential Privacy. ACM, p1-22.

[3] Cynthia Dwork,Moni Naor,Toniann Pitassi and Guy N. Rothblum. (2010). Differential Privacy Under Continual Observation. ACM, p1-10.

[4] Chao Li and Gerome Miklau. (2012). An Adaptive Mechanism for Accurate Query Answering under Differential Privacy. ACM, p1-13.

[5] Joshua Rosen, Neoklis Polyzotis,Vinayak Borkar, Yingyi Bu, Michael J. Carey,Markus Weimer, Tyson Condie and Raghu Ramakrishnan. (2013). Iterative MapReduce for Large Scale Machine earning. ACM, p1-9.

[6] Avita Katal,Mohammad Wazid and R H Goudar. (2013). Big Data: Issues, Challenges, Tools and Good Practices. IEEE, p1-6.

[7] Priya P. Sharma and Chandrakant P. Navdeti. (2014). Securing Big Data Hadoop: A Review of Security Issues, Threats and Solution. International Journal of Computer Science and Information Technologies. 5 (2), p1-6.

[8] Mircea Moca, Gheorghe Cosmin Silaghi and Gilles Fedak. (2011). Distributed Results Checking for MapReduce in Volunteer Computing. IEEE International Parallel & Distributed Processing Symposium, p1-8.

[9] Diogo A. B. Fernandes • Liliana F. B. Soares • João V. Gomes , Mário M. Freire and Pedro R. M. Inácio. (2014). Security issues in cloud environments: a survey. Springer , p1-58.

[10] Amresh Kumar,Kiran M , Saikat Mukherjee and Ravi Prakash G.. (2013). Verification and Validation of MapReduce Program model for Parallel K-Means algorithm on Hadoop Cluster. International Journal of Computer Applications. 72 (8), p1-8.

[11] Katarina Grolinger,Michael Hayes,Wilson A. Higashino and David S. Allison. (2014). Challenges for MapReduce in Big Data. Electrical and Computer Engineering , p1-10.

[12] Jiong Xie, Shu Yin, Xiaojun Ruan, Zhiyang Ding, Yun Tian, James Majors, Adam Manzanares, and Xiao Qin. (2010). Improving MapReduce Performance through Data Placement in Heterogeneous Hadoop Clusters. ACM, p1-10.

[13] HAI JIN, SHADI IBRAHIM, LI QI, HAIJUN CAO, SONG WU and XUANHUA SHI. (2011). THE MAPREDUCE PROGRAMMING MODEL AND IMPLEMENTATIONS. Springer , p1-16.

[14] Chu Huang, Sencun Zhu and Dinghao Wu. (2010). Towards Trusted Services: Result Verification Schemes for MapReduce. Springer , p1-8.

[15] Konstantinos Chatzikokolakis, Miguel Andres, Nicolas Bordenabe, Catuscia Palamidessi. (2013). Broadening the Scope of Diferential Privacy Using Metrics. ACM, p1-34.

[16] Ashwin Machanavajjhala,Johannes Gehrke , Daniel Kifer Muthuramakrishnan and Venkitasubramaniam. (2012). ℓ-Diversity: Privacy Beyond k-Anonymity. ACM, p1-12.

[17] Marcelo D. Holtz, Bernardo M. David and Rafael Tim´oteo de Sousa J´unior. (2011). Building Scalable Distributed Intrusion Detection Systems Based on the MapReduce Framework. REVISTA TELECOMUNICAÇÕES. 13 (2), p1-11.

[18] Charu C. Aggarwal. (2011). On k-Anonymity and the Curse of Dimensionality. ACM, p1-9.

[19] Zhifeng Xiao and Yang Xiao. (2014). Achieving Accountable MapReduce in cloud computing. Future Generation Computer Systems. 30 , p1-13.

[20] Ninghui Li, Tiancheng Li and Suresh Venkatasubramanian. (2011). t-Closeness: Privacy Beyond k-Anonymity and -Diversity. ACM, p1-10.

[21] Chris Clifton. (2001). Privacy Preserving Distributed Data Mining. Computer Sciences p1-10.

[22] S.Selva Rathna, Dr. T. Karthikeyan. (2015). Survey on Recent Algorithms for Privacy Preserving Data mining. computer scince. 6 (2), p1-6.

[23] Cynthia Dwork, Vitaly Feldman, Moritz Hardt ,Toniann Pitassi, Omer Reingold and Aaron Roth. (2015). Preserving Statistical Validity in Adaptive Data Analysis. Computer Sciences p1-29.

[24] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann PitassiOmer Reingold and Aaron Roth. (2015). The reusable holdout: Preserving validity in adaptive data analysis. Computer Sciences. 349 ,p1-4.

[25] Chris Clifton, Murat Kantarcioglu, Xiaodong Lin, Michael and Y. Zhu . (2002). Tools for Privacy Preserving Distributed Data Mining. IEEE 4 (2) p1-7.

[26] V. Baby and N. Subhash Chandra. (2016). Privacy-Preserving Distributed A Survey Data Mining Techniques:. International Journal of Computer Applications. 143(10), p1-5.

[27] Pawel Jurczyk and Li Xiong. (2008). Privacy-Preserving Data Publishing for Horizontally Partitioned Databases. IEEE p1-2.

[28] BENJAMIN C. M. FUNG,KE WANG,RUI CHEN and and PHILIP S. YU. (2010). Privacy-Preserving Data Publishing: A Survey of Recent Developments. ACM. 42 (4), p1-53.

[29] V. V. Nagendra kumar and C. Lavanya. (2014). Privacy-Preserving For Collaborative Data Publishing. IJCSIT. 5 (3), p1-4.

[30] Rebecca N. Wright , Zhiqiang Yang and Sheng Zhong. (2006). Distributed Data Mining Protocols for Privacy: A Review of Some Recent Results. IEEE, p1-13.

[31] A N K Zaman and Charlie Obimbo. (2014). Privacy Preserving Data Publishing: A Classification Perspective. IJACSA. 5 (9), p1-6.

[32] The Apache Software Foundation. (2016). Welcome to Apache™ Hadoop. Available: http://hadoop.apache.org/. Last accessed 01 December 2016.

[33] Apache Software Foundation. (2016). MapReduce Tutorial. Available: https://hadoop.apache.org/docs/stable/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html. Last accessed 01 December 2016.

[34] NIST (2016). Final Version of NIST Cloud Computing Definition. Available online at: https://www.nist.gov/news-events/news/2011/10/final-version-nist-cloud-computing-definition-published. Accessed on 01 December 2016.

[35] Malhotra, L., Agrawal, D. and Jaiswal, A. (2014). Virtualization in Cloud Computing. Information Technology & Software Engineering, 4 (2), p1-3.

[36] Amazon Web Services. (2016). Amazon EMR. Available: https://aws.amazon.com/emr/. Last accessed 01 December 2016.

[37] R. Agrawal and R. Srikant. Privacy-Preserving Data Mining. In Proceedings of the ACM SIGMOD Conference on Management of Data. Dallas,Texas. May 2000. pp.439-450.

[38] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor.Our data, ourselves: Privacy via distributed noise generation. In EUROCRYPT, 2006.

[39] Securities and Exchange Commission. (2016). EDGAR Log File Data Set. Available: https://www.sec.gov/data/edgar-log-file-data-set. Last accessed 10th November 2016.

[40] Madhusudhan Reddy, N., and Nagaraju, C. (2015). Survey on Emerging Technologies for Secure Computations of Big Data. I-Manager's Journal of Cloud Computing, 2 (1), p1-6.

[41] Decompilers Onlne (2016). Available at: http://www.javadecompilers.com/jad [accssed on 10 November 2016]