Parallel computing for preserving privacy using k-anonymisation algorithms from big data

Sharath Yaji and B. Neelima*

Department of Information Science and Engineering,
NMAM Institute Of Technology,
Nitte, Karkala (Taluk),
Karnataka, Post Box No. 574118, India
Email: sharathyaji@nitte.edu.in
Email: neelimareddy@nitte.edu.in
*Corresponding author

Abstract: Many organisations still consider preserving privacy for big data as a major challenge. Parallel computation can be used to optimise big data analysis. This paper gives a proposal for parallelising k-anonymisation algorithms through comparative study and survey. The k-anonymisation algorithms considered are MinGen, DataFly, Incognito and Mondrian. The result shows the parallel versions of the algorithms perform better than sequential counterparts, as data size increases. For small size dataset in sequential mode MinGen is 71.83% faster than parallel version. However, in sequential mode DataFly and in parallel mode incognito performed well. For large size dataset in parallel mode Incognito is 101.186% faster than sequential. However, in sequential mode MinGen and DataFly performed well. In parallel mode Incognito, DataFly and MinGen performed well. The paper acts as a single point of reference for choosing big data mining k-anonymisation algorithms. This paper gives direction of applying HPC concepts such as parallelisation for privacy preserving algorithms.

Keywords: big data; k-anonymisation; privacy preserving; big data analysis; parallel computing in big data.

Reference to this paper should be made as follows: Yaji, S. and Neelima, B. (2018) 'Parallel computing for preserving privacy using k-anonymisation algorithms from big data', *Int. J. Big Data Intelligence*, Vol. 5, No. 3, pp.191–200.

Biographical notes: Sharath Yaji is a Research Scholar at the Department of Information Science and Engineering at NMAMIT, Nitte. He received his Bachelor of Engineering, Master of Technology in Computer Science and Engineering from the Visveshwaraya Technological University – Belgavi, Karnataka, India. He is a student member of IEEE Computer Society and ACM. His areas of interest are privacy preserving in big data analysis, IoT and cyber security. He has experience in teaching and industry.

B. Neelima is working as Professor and Head of the Department of Information Science and Engineering at NMAM Institute of Technology, Nitte, Karnataka, India. She has completed her PhD from the National Institute of Technology Karnataka (NITK), Surathkal, Karnataka, India in the area of high performance computing. She is a life member of CSI, ISTE, IE-India, member of ACM-W and WIE and senior member of ACM and IEEE. She has completed an R&D project from DST, GoI and has around 50 publications in various international and national journals and conferences. She is the coordinator for GPU Education Centre at NMAMIT, awarded by NVIDIA, USA.

1 Introduction

In this big data era, many organisations facing challenges on security and privacy handling in cloud computing, big data analysis and in internet of things (IoT). Big data deals with massive amount of data in terms of zetta and yotta bytes. The major issue in data privacy is re-identification. Hackers can easily expose sensitive data through re-identification. To address this challenge, privacy preserving algorithms have been proposed and used for securing sensitive data (Agrawal and Srikant, 2000; Takabi et al., 2010) from big data.

Recently, research on privacy preserving in data mining is very active. Many researchers are working in different aspects of the same. Privacy preserving of sensitive data is addressed by anonymisation algorithms (Sweeney, 1997). There are different categories of anonymisation algorithms such as: k-anonymisation, L-diversity, T-closeness and differential privacy (Ayala-Rivera et al., 2014; Fung et al.,

Copyright © 2018 Inderscience Enterprises Ltd.