

Privacy-Preserving Big Data Analytics

Pr. Mohand-Tahar KECHADI

Academic Affiliation: University College Dublin, Dublin (UCD)

Mail: tahar.kechadi@ucd.ie

Guest Speaker—



BIOGRAPHY:

Professor M-Tahar Kechadi was awarded PhD and Masters degree - in Computer Science from University of Lille 1, France. After his PhD he won one of the prestigious EU post-doctoral fellowships under TMR programme, which is did at UCD, Ireland. He joined UCD in 1999 as a permanent staff member of the School of Computer Science (SCS). He is currently Professor of Computer Science at SCS, UCD. His research interests span the areas of Data Mining, distributed data mining, heterogeneous distributed systems, Big Data, Cloud Computing, and digital forensics and cyber-crime investigations. Nowadays we live in digital world, we produce more data than we can analyse and exploit. This 'big data' will continue to grow at rapid pace, will underpin new waves of innovation in nearly every sector of the world economy, and will reshape the way we build and use computers (hardware and software). The core and central focus of his research for the last decade is how to manage and analyse data quickly and efficiently to gain useful insights. Currently, his research interests are primary in:

- Big Data Analytics techniques and models and their execution environments and applications (Privacy preserving, distributed, cloud/Grid type services, etc.)
- Cloud/Grid computing and services for supporting data access, management, and mining processes,
- Data Privacy, Digital Forensics and cybercrime investigations.

Data-intensive (or driven) real-world applications are all behind what we call today 'big data'. The problem then is how to manage and process all these huge volumes of data. Both the analysis of large datasets and the computing environments created new problems and challenges for efficient execution and optimal system performance. This brought him to look at the challenges of the data mining in the heterogeneous, complex, and distributed environment. For more effective and efficient analysis of large datasets we now recognise that data analytics (DA) is a multi-step process and the user is at the centre of some steps. One of his research objectives is to study these steps and their interrelationships with regards to the overall process performance. We have started to understand the importance of some of these steps and their impact on the overall process. Another key objective is to design and develop DA techniques that delineate a careful division of work between the user and computer. One way to tackle this challenge is to provide constant feedback to the user and engage with the user only when it is required. And finally the scalability and privacy issues, as the datasets are becoming extremely large containing data of various types and pertaining to different systems or users. Recently, he is specifically looking at these issues from privacy, ethics, and healthcare perspectives.

Prof Kechadi has published over 350 research articles in refereed journals and conferences. He serves on the scientific committees for a number of international conferences and he organised and hosted one of the leading conferences in his area. He is an editor of the Journal of Future Generation of

Computer Systems. Prof M-Tahar Kechadi has been involved in international collaborations, in particular with researchers at the Universities of Tennessee, Purdue, Liverpool, Lille and Artois (France) and CERN. He is currently full member at CERN. He has been a visiting professor at the University of Artois – Bethune since 2002. He is currently visiting professor at the Fuzhou University, China. Prof Kechadi is a Principal Investigator in Insight Centre for data Analytics; the biggest research Data Centre in Ireland, and CONSUS project, one of the biggest project in Big data analytics applied to agriculture.

KEYNOTE ABSTRACT

We live in the era of Big Data and big data analytics has emerged as a promising alternative solution for gaining insights and understanding any phenomenon governing the data and therefore the corresponding application. analysing these very large sets of data. While collecting huge amounts of data from various sources is becoming commonplace, the overall process of big data analytics complex process as we have to deal with all the challenges of big data (Volume, Variety, Velocity, and Veracity). These challenges have direct impact on the accuracy of the results, computational complexity, and data management and storage. Big data analytics requires much more advanced machine learning techniques and sophisticated computing systems to face all these challenges at once, as traditional analytics techniques fail too address them. In addition, data analytics, in general, is prone to privacy violations and attacks. In this presentation, we review the challenges of big data privacy-preserving analytics and various techniques that have employed to prevent privacy threats. We will look also into the types of attacks against big data analytics and how can we avoid such attacks. We conclude by research challenges in the area of privacy preservation and their effects on this era of Big Data.

KEYWORDS:

Big Data, Data Analytics, Data Privacy, Sensitive Data Analysis, Big Data Analytics Security.