Survey on Improving Data Utility in Differentially Private Sequential Data Publishing

ABSTRACT

The rapid development of mobile, sensing, and communication technologies, massive amounts of sequential sensing data are being collected, stored, shared, and analyzed all the time, forming various comprehensive mobile sensing systems, sequential data is classified as mobile sensing data with time correlations. Examples include time-series data, real-time data, trajectory data, and others. Data publishing is referred to the process, in which various intelligent sensing devices contribute their sensing data to other third parties or applications. Mobile sensing systems, as the motivating application of this work, integrate aggregation, statistics, and analysis of big data to provide valuable information, and has brought significant improvements to human society [2] (e.g., e-Science, industry, business, social media, healthcare, etc.).

EXISTING SYSTEM

The massive generation, extensive sharing, and deep exploitation of data in the big data era have raised unprecedented privacy threats. To address privacy concerns, various privacy paradigms have been proposed to achieve a good tradeoff between privacy and data utility. Particularly, differential privacy has been well accepted as one of the *de facto* standards for privacy preservation, and numerous schemes guaranteeing differential privacy have been proposed. Nonetheless, most of the existing works claiming a superior utility-privacy tradeoff only present specific methods, with distinct perspectives, and a complete comparative analysis and evaluation study has not been fully investigated.

DRAWBACKS

- Extensive sharing, and deep exploitation of data in the big data era have raised unprecedented privacy threats
- Privacy-preserving schemes can be generally categorized into three strategies: (i) anonymity (ii) encryption and (iii) perturbation. Nonetheless, these schemes are unable to solve all of the problems at hand.

PROPOSED SYSTEM

To this end, in this paper we review and investigate existing schemes on providing differential privacy from a broad and encompassing perspective to provide a comprehensive survey with respect to both the privacy guarantee and the effectiveness and efficiency in utility improvement. We categorize the existing schemes into distribution optimization, sensitivity calibration, transformation, decomposition, and correlations exploitation, based on their mechanisms in improving data utility. We also conduct some analysis and comparison of their various concepts and principles, focusing on improvements to data utility. Finally, we outline some challenges and provide future research directions.

ADVANTAGES

- We review and investigate existing schemes on providing differential privacy.
- To provide a comprehensive survey with respect to both the privacy guarantee and the effectiveness and efficiency in utility improvement.

SYSTEM REQUIREMENTS

- **H/W System Configuration:-**

- Processor - Pentium –IV
- RAM - 4 GB (min)
- Hard Disk - 20 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA

- **S/W System Configuration:-**

- Operating System : Windows 7 or 8 32 bit
- Application Server : Tomcat5.0/6.X
- Backend coding : Java
- Tool : Virtual Box
- Environment : Ubuntu
- Technology : Hadoop