Differentially Private Data Publishing and Analysis: a Survey

Abstract:

Digital information collected by corporations, organizations, and governments has resulted in a vast number of datasets, and the speed of such data collection has increased dramatically over the last few years. Typically, a data collector, also known as a curator, is in charge of publishing data for further analysis. Data publishing aims to share datasets or some query results to the public. In some literature, this scenario is known as data sharing or data release. Differential privacy is one such new and promising privacy model. It ensures that the ability of an adversary to inflict harm on any individual in a dataset is essentially the same, independent of whether any individual opts in to, or out of, the dataset.

Existing System:

Differential privacy is an essential and prevalent privacy model that has been widely explored in recent decades. This survey provides a comprehensive and structured overview of two research directions: differentially private data publishing and differentially private data analysis. We compare the diverse release mechanisms of differentially private data publishing given a variety of input data in terms of query type, the maximum number of queries, efficiency, and accuracy. We identify two basic frameworks for differentially private data analysis and list the typical algorithms used within each framework. The results are compared and discussed based on output accuracy and efficiency.

Disadvantages:

➢ Less Efficiency.
➢ Less Accuracy.

Proposed System:

This paper presents a multi-disciplinary survey of work on differential privacy including an overview of the huge amount of literature in two major differential privacy research streams: data publishing and data analysis. We identified different publishing mechanisms for data
publishing and compared various types of input and output data. In addition, we presented two basic dataset publishing methods: anonymized and learning-based. We discussed two basic frameworks for data analysis and illustrated their respective analysis scenarios. The basic techniques in differential privacy look simple and intuitively appealing, and when combined with specific problems, differential privacy demonstrates itself as a powerful and useful tool for a diverse range of applications.

Advantages:
➢ More Efficiency.
➢ More Accuracy.

Modules:
➢ Differentially Private Data Publishing.
➢ Differentially Private Data Analysis.

SYSTEM REQUIREMENTS

H/W System Configuration:-

- Processor : Pentium –III
- RAM : 256 MB (min)
- Hard Disk : 20 GB
- Key Board : Standard Windows Keyboard
- Mouse : Two or Three Button Mouse
- Monitor : SVGA

S/W System Configuration:-

- Application Server : Tomcat5.0/6.X
<table>
<thead>
<tr>
<th>Component</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front End</td>
<td>HTML, Jsp</td>
</tr>
<tr>
<td>Scripts</td>
<td>JavaScript</td>
</tr>
<tr>
<td>Server side Script</td>
<td>Java Server Pages</td>
</tr>
<tr>
<td>Database</td>
<td>MySQL 5.0</td>
</tr>
<tr>
<td>Database Connectivity</td>
<td>JDBC</td>
</tr>
</tbody>
</table>