Abstract

Privacy preserving data mining(PPDM) is concerned about extracting useful information from a large amount of data in such a way that neither the result should reveal any private information nor any private information is revealed while data mining task is being performed. Privacy preserving data publishing (PPDP) is concerned about making a large amount of data publicly available in such a way that published data would not reveal any private information. We are concerned about the distributed data which means distributed data of single owner or data of multiple owners. Data mining task if performed on the union of data from all the owners, we will be able to find new hidden patterns. Available methods of PPDM for distributed data are centered towards any single data mining task. Available PPDP methods for distributed data, which are very few, are unable to work for n parties.

In this work, we provide a comprehensive and structured review of most available techniques for PPDM and PPDP. To provide a consistent description of the techniques, we also introduce a taxonomy of PPDM and PPDP techniques. We also propose a model as a solution to the problem using which any data mining task can be applied on distributed data preserving the privacy. Using this model distributed data can be published preserving the privacy.

Keywords: PPDM, PPDP, attack, SMC, unstructured data, distributed data, anonymization, slicing, cellular automata, secret sharing, threshold cryptography, enciphering.