A privacy leakage upper-bound constraint-based essay

Economics

However, preserving the privacy of intermediate datasets becomes a challenging problem because adversaries may recover privacy-sensitive information by analyzing multiple intermediate datasets. Encrypting ALL datasets in cloud Is widely adopted In existing approaches to address this challenge. But we argue that encrypting all intermediate datasets are neither efficient nor cost-effective because it is very time consuming and costly for data-intensive applications to en/ decrypt datasets frequently while performing any operation on them. In this paper, we propose a novel upper-bound privacy leakage constraint based approach to Identify which intermediate datasets need to be encrypted and which do not, so that rivalry-preserving cost can be saved while the privacy requirements of data holders can still be satisfied.

Evaluation results demonstrate that the privacy-preserving cost of intermediate datasets can be significantly reduced with our approach over existing ones where all datasets are encrypted. EXISTING SYSTEM: Existing technical approaches for preserving the privy-coy of datasets stored In cloud mainly include encryption and minimization. On one hand, encrypting all datasets. A straightforward and effective approach, is widely adopted in current research . However, processing on encrypted datasets efficiently is quite a challenging task, because most existing applications only run on unencrypted datasets. However, preserving the privacy of Intermediate datasets becomes a challenging problem because adversaries may recover privacy-sensitive information by analyzing multiple intermediate datasets. Encrypting ALL datasets in cloud is widely adopted in existing approaches to address this challenge. But we argue that encrypting all Intermediate

datasets are neither efficient nor cost-effective because It Is very time consuming and costly for data-intensive applications to en/decrypt datasets frequently while 1 OFF PROPOSED SYSTEM: In this paper, we propose a novel approach to identify which intermediate datasets need to be encrypted while others do not, in order to satisfy privacy requirements given by data holders.

A tree structure is modeled from generation relationships of intermediate datasets to Ana-laze privacy propagation of datasets. As quantifying loin privacy leakage of multiple datasets efficiently is chalk-leaning, we exploit an upper-bound constraint to confine privacy disclosure. Based on such a constraint, we model the problem of saving privacy-preserving cost as a con-strained optimization problem. This problem is then did-vided into a series of sub-problems by decomposing PRI-vacancy leakage constraints.

Finally, we design a practical heuristic algorithm accordingly to identify the datasets that need to be encrypted. Experimental results on real-world and extensive datasets demonstrate that privacy- preserving cost of intermediate datasets can be signifies-scantly reduced with our approach over existing ones where all datasets are encrypted. MODULE DESCRIPTION: Number of Modules After careful analysis the system has been identified to have the following modules: 1 . Data Storage Privacy Module. 2.

Privacy Preserving Module. 3. Intermediate Dataset Module. 4. Privacy Paperbound Module.

1 . Data Storage Privacy Module: The privacy concerns caused by retaining intermediate datasets in cloud are important but they are paid little

attention. A motivating scenario is illustrated where an on-line health service provider, e.

G. , Microsoft Health Vault has moved data storage into cloud for economical benefits. Original datasets are encrypted for confidential-TTY. Data users like governments or research centers access or process part of original datasets after minimization.

Intermediate datasets generated during data access or process are retained for data reuse and cost saving. We proposed an approach that combines encryption and data fragment-taxation to achieve privacy protection for distributed data storage with encrypting only part of datasets. .

Privacy Preserving Module: Privacy-preserving techniques like generalization can with-stand most privacy attacks on one single dataset, while preserving privacy for multiple datasets is still a challenge- ins problem. Thus, for preserving privacy of multiple datasets, it is sharing them in cloud. Privacy-preserving cost of intermediate datasets stems from frequent en/decryption with charged cloud services. 3. Intermediate Dataset Module: An intermediate dataset is assumed to have been anon-minimized to satisfy certain privacy requirements. However, putting multiple datasets together may still invoke a high risk of revealing privacy-sensitive information, resulting in violating the privacy requirements. Data provenance is employed to manage intermediate datasets in our research. Provenance is com-only defined as the origin, source or history of derive- Zion of some objects and data, which can be reckoned as the information upon how data was generated.

Re-productivity of data provenance can help to regenerate a dataset from its nearest existing predecessor datasets rather than from scratch 4. Privacy Paperbound Module: Privacy quantification of a single data-set is stated.