We propose SECRETA, a system for evaluating and comparing anonymization algorithms for relational, transaction, and RT-datasets.

Numerous anonymization algorithms have been proposed, to maintain data truthfulness while at the same time prevent disclosure of individuals’ private and sensitive information, such as:

- medical
- healthcare

The effectiveness and efficiency of these algorithms depends on:

1. data characteristics (e.g., the distribution of values in an attribute)
2. various input parameters which affect the level of privacy and utility (e.g., hierarchies that govern data transformation)
3. data utility requirements (e.g., the need to accurately answer a certain query workload, or to adhere to constraints on the way values are transformed)

SECRETA operates in two modes:

1. Evaluation mode can be used to configure and evaluate the effectiveness of a given algorithm, with respect to data utility and privacy, as well as its efficiency. SECRETA enables the use of 20 different combinations of algorithms to anonymize RT-datasets. The selection and management of these combinations is performed in an intuitive way that allows preserving different aspects of data utility.
2. Comparison mode offers data publishers the ability to design and execute benchmark comparisons, using a single algorithm with the specified configuration.

Programming frameworks, and it allows data publishers to apply and analyze the performance of one or more of these algorithms.

SECRETA integrates 9 popular anonymization algorithms under a common, benchmark-oriented framework, and it allows data publishers to apply and analyze the performance of one or more of these algorithms.

The system has a front-end and a back-end component.

The front-end consists of:

1. Dataset Editor enables users to select datasets for anonymization
2. Configuration Editor allows users to select hierarchies and to specify utility and privacy policies
3. Experimentation Interface Editor selects the operation mode of SECRETA
4. Queries Editor component allows specifying query workloads, which will be used to evaluate the utility of anonymized data in query answering
5. Plotting Module supports a series of data visualizations that help users analyze their data and understand the performance of anonymization algorithms
6. Data Export Module allows exporting datasets, hierarchies, policies, and privacy settings to a file

The back-end consists of:

1. Policy Specification Module invokes algorithms that automatically generate hierarchies [4], as well as the strategies in [3], which generate privacy and utility policies
2. Method Evaluation / Comparator implements the functionality that is necessary for supporting the interfaces of the Evaluation and the Comparator module
3. Anonymization Module is responsible for executing an anonymization algorithm with the specified configuration
4. Experimentation Module is responsible for producing visualizations of the anonymization results and of the performance of the anonymization algorithm

Evaluating a method for RT-datasets: The users will configure, apply, and evaluate a method, in a series of steps.

1. They will set the values for parameters \( \lambda \), \( \mu \), \( \delta \) (top-left pane)
2. They may select two algorithms, one for anonymizing the relational attributes, and one for the transaction attribute, and a bounding method for combining the selected algorithms
3. They will initiate the anonymization process. When this process ends, a message box with a summary of results will be presented and the anonymized dataset will be displayed in the output area (middle pane)
4. They will select a number of data visualizations

Comparing methods for RT-datasets: The users will compare multiple anonymization methods.

Using the “Methods comparison” interface, they will:

1. select algorithms for anonymizing each type of attributes, as well as a bounding method
2. set the values for parameters that will be fixed, as described above (top-left)
3. choose a varying parameter (top-mid), along with its start/end value and step (bottom 左)
4. These choices comprise a configuration, which will be added into the experiment area (top-right pane). Similar configurations will be created by the users for at least another method. After the methods are applied, the users will select various graphs, which will be displayed in the plotting area (bottom pane).

Acknowledgements

- G. Poulis is supported by the Research Funding Program Heraclitus II and the Research Project PublicaMundi.
- G. Loukides is partly supported by a Research Fellowship from the Royal Academy of Engineering.
- S. Skadopoulos is partially supported by EU/Greece the Research Funding Program: Thales.

References