

EDBT 2014 Test of Time Award

In 2014, EDBT is starting a test-of-time (ToT) award with the goal of recognizing one paper presented at EDBT from at least a decade earlier that have best met the “test of time”. The EDBT ToT award for 2014 will be presented during the EDBT/ICDT 2014 Joint Conference, March 24-27, 2014 in Athens, Greece (<http://www.edbticdt2014.gr/>). The EDBT 2014 Test of Time Award committee was formed by Gustavo Alonso, Torsten Grust and Walid Aref, all members of the EDBT 2014 PC, and chaired by Sihem Amer-Yahia, the EDBT 2014 PC chair. The committee was charged with selecting a paper or a small number of papers from the EDBT 2004 proceedings that has had the most impact in terms of research, methodology, conceptual contribution, or transfer to practice over the past decade. After careful consideration, the committee has decided to select the following paper:

A Condensation Approach to Privacy Preserving Data Mining

by Charu C. Aggarwal and Philip S. Yu

The paper addresses the topic of privacy-preserving data mining and proposes a framework for anonymization in contrast to perturbation approaches that introduce noise. The approach advocated in this paper takes into consideration the correlated nature of multi-dimensional data and also that is not problem-specific, i.e., the framework is generic and many mining operations can be formulated and solved given the same framework.

The paper is seminal and covers a topic which has not lost relevance today. It constitutes a prime example of how far a simple idea (in this case condensation: reduce a group of records to a number of carefully selected statistical properties of that group) can carry a whole line of original research. Condensation comes with a naturally built-in “privacy dial” (namely the group size) and is practical (easily implemented, maintainable under updates). It established general techniques that have been taken up by many people in the area and outside the area. The results in this paper can be (and have been) generalized in many ways as they tackled a more general problem of grouping, aggregation, and sampling.

According to Google Scholar, the paper has been constantly cited in the period from 2005 until today (2013) with a remarkable peak in 2008 when privacy-preserving data mining and database security in general became research topics of core interest. Since 2009 until today, the number of citations has only gone up. Papers citing this paper are also highly cited, indicating the relevance of the contribution.

A number of follow-up papers by the same authors, quite clearly demonstrate that the proposed condensation approach has interesting aspects:

- 2005, Proc. of the 5th SIAM International Conference on Data Mining, “On variable constraints in privacy preserving data mining”
- 2008, Data Mining and Knowledge Discovery 16 (3), “A framework for condensation-based anonymization of string data”
- 2008, ACM Transactions on Database Systems (TODS) 33 (1), 2, “On static and dynamic methods for condensation-based privacy-preserving data mining”
- 2008, Handbook of Database Security, “Privacy-Preserving Data Mining: A Survey”

Also in 2008, the Springer book “Privacy-Preserving Data Mining”, edited by Aggarwal and Yu, was published.

Work on condensation-based privacy in data mining also had an impact on industry and industrial practice. At least the following US patents (or patent applications) are based on the idea of condensation:

- US Patent 8,010,541 – Systems and methods for condensation-based privacy in strings
- US Patent 7,885,941 – Methods and apparatus for privacy preserving data mining using statistical condensing approach
- US Patent 7,904,471 – Method, apparatus and computer program product for preserving privacy in data mining
- US Patent App. 12/119,766 – Method and Apparatus for Variable Privacy Preservation in Data Mining
- US Patent 7,305,378 – System and method for distributed privacy preserving data mining
- US Patent 7,302,420 – Methods and apparatus for privacy preserving data mining using statistical condensing approach
- US Patent App. 11/249,647