|  |  |
| --- | --- |
| **17ноября**вторник | **Коллоквиумфакультета компьютерных наук НИУ ВШЭ** |
|  | **Panagiotis Karras**Skoltech**k-Anonymization by Freeform Generalization** |
| Syntactic data anonymization strives to (i) ensure that an adversary cannot identify an individual’s record from published record attributes with high probability, and (ii) preserve data utility. These mutually conflicting goals can be expressed as an optimization problem with privacy as the constraint and utility as the objective function. Conventional research on *k*-anonymity, a popular privacy model, has resorted to generalizing data values in homogeneous groups. However, such grouping is not necessary. Instead, data values can be recast in a heterogeneous manner that allows for higher utility. Nevertheless, previous work in this direction did not define the problem in the most general terms; thus, the utility gains achieved are limited. In this paper, we propose a methodology that achieves the full potential of heterogeneity and gains higher utility. We formulate the problem as a network flow problem and develop an optimal solution therefor using Mixed Integer Programming, an *O*(*kn*2) greedy algorithm that has no time-complexity disadvantage vis-á-vis previous approaches, an *O*(*kn*2 log *n*) enhanced version thereof, and an *O*(*kn*3) adaptation of the Hungarian algorithm; these algorithms build a set of k perfect matchings from original to anonymized data. Our techniques can resist adversaries who may know the employed algorithms. Experiments with real-world data verify that our schemes achieve near-optimal utility (with gains of up to 41%), while they can exploit parallelism and data partitioning, gaining an efficiency advantage over simpler methods. |
| **17 ноября, 18:10-20:00Кочновский проезд, 3 Лекционный зал Декарт, 3 этаж****Заказать пропуск на проход в зданиеможно на computerscience@hse.ru** |  |