T-79.511 Special Course on Cryptology / Review of survey of Privacy Preserving Data Mining

Tatiana Issaeva

September 30, 2003

Overview

The survey under review is done for article "Privacy Preserving Data mining" by Agrawal and Srikant.

Survey provides sufficient amount of material to cover the topic and leaves detailed mathematical and algorithmic descriptions out of review, thus easing the understanding. Also sometimes some minor but important details are left out, which makes it difficult to follow the material and reference to original paper is needed.

Section 1

In survey: the given true value can be reveled. Correction: the given true value can be revealed.

In survey: The main problem with this methods ... Remark: either this or methods are incorrect

Section 2

In survey, in the end of the section 2: focus on this value disturbtion method. Correction: focus on this value distribution method.

Section 3

In first sentence: which all have the same distribution X. This is incorrect, since X is not a distribution but a random variable. Correct way to say is: which all have the same distribution as random variable X. The same goes for the next sentence about Y.

In the first formula in section 3, what is f_X?

In the third formula for $F_X^{'}$ in section 3: integral in denominator is taken by $dz^{'}$ not by dz.

In survey: Y_1 is independant of X_1. Correction: Y_1 is independent of X_1.

Do not use article the with $F_X^{'}$.

In all formulas on page 3 for $F_X^{'}$, w_i should be used instead of w_1.

Second sentence on page 3: With sufficient high number. Correction: With sufficiently high number.

Section 4

There are couple of times verb are is used with data noun. Data is allways singular, so is should be used.

Second paragraph in section 4: There are usually two phases ... Question: two phases for what?

In first formula in section 4, what does S mean?

Last sentence in section 4.1 should be checked for grammar and spelling.

On page 4 sentence about partitioning the data should be checked for spelling.

Conclusion

From the survey, as well as from the paper, it is a bit unclear what do we actually trying to protect and what are the metrics. One metric for privacy is defined but it is not clear how it is used to conclude that privacy is actually protected.