T-79.511 Special Course on Cryptology / Review of Survey of Vulnerabilities in Similarity Search Systems

Vesa Vaskelainen

17.11.2003

1 Overview

The survey under review is done by Sami Vaarala based on the article "Vulnerabilities in Similarity Search Based Systems" by Tosun and Ferhatosmanoglu. The survey is written with good english and the topic is widely covered. Presentation is clear and proceeds logically which makes the survey easy to understand and also very readable. One general correction though, throughout the survey words in the titles should be written with capital initial letters.

2 Introduction

The first paragraph could be better to put under the section "Abstract" and the rest of the text under "Definitions". This section does not actually give any introduction to the Topic yet. Instead of saying that "defined in the usual way" would be more relevant to say that l_2 norm is used. In the sentence "some function $f(x)$ of the distance", x would be good add to add after the word "distance".

3 Problem Considered

In the first paragraph correct the word "discuessed" to "discussed".

4 Reply Model Vulnerabilities

Probably more fluent way to start is "Here we will examine what follows when we assume a distance function ...".

4.1 Basic Approach

Voronoi region of a query point is not clearly defined. An example is given but is it the definition also.

4.2 Schemes

The word "probe" appear first time in the text here, and meaning of it is unclear if the reader is not familiar with the original paper.

5 Score Model

Sentences between the formulas should not start new paragraphs. The formulas are part of the text so punctuation marks should be used in the normal way such as there should be a dot after the first formula since the new sentence starts after it.

6 Attack Detection

Figures 6,7 and 8 should have titles for the axes and captions could show the actual functions plotted in the figures.

7 Thoughts

In the first sentence, requires 100^{10} what. Figure 9. there should be one more black query point in the region where data points are clustered. Because of the silarity of the Figure 9. and Figure 4. would be good mention that in the Figure 9. no assumptions about the minimum distance between vectors are made.