

deobfuscation algorithm assumes that data is hidden. So if an attacker were to run the algorithm on any piece of text he finds, it will always return results. This ties into the chosen stego attack, as the attacker have access to both the algorithm and the stegotext but would not be able to tell if the data he was extracting was meaningful or not.

5.8.2 Human

Taking into account the user survey results and studying the output from the algorithm, the algorithm does provide a good level of protection from human analysis. In general, the words that are used for replacements fit in with the surrounding sentence. One thing that causes more problems to human analysers is errors in the original text, meaning that the strength of the algorithm is related to the quality of the original text.

There are cases where the word that is used as a replacement does not make sense. As this is a prototype algorithm, this was expected, and will require further work to improve (although it will be extremely difficult to remove all cases).

The implementation of the “bad words” list to remove words from the system which cause false positives to be found when deobfuscating is another improvement that can be made to the algorithm. This could be done either through user reporting on the live application or by using the test application and much larger sets of test data that was used in the evaluation. The algorithm will be stronger when used on English text in countries where English is not the native language. Even the test subject who has been speaking English for four years almost exclusively had trouble understanding some of the words used; not that they didn’t make sense in the sentence but they didn’t recognise the words themselves.