Final Year Project (CS4E2) Overview

(Arnav Malhotra, 17317424)

Overview:

The field of interest for this project would be a combination of Machine Learning and Data Security. Particularly, network analysis and data privacy in corporate sector.

Problem Statement:

In most of the companies, there is a constant threat of misuse of company network for either personal gains or corporate espionage. Hence, to tackle this, a software needs to be devised which monitors the company network for suspicious/malicious activities and reports them to concerned authority within the company.

Challenges Faced and How to tackle them:

There a couple of challenges which would be faced explained by the following examples-

1) A person circulates malware through email in the company. To tackle this, the software will learn the nature of the file from the provided dataset and regularly updated database of definitions of malware in order to isolate it and report the sender of that file.
2) A person in one department circulates sensitive information not related to their own department in the company. In this case, the software will learn and determine whether that person should be having that information in the first place. Then, determine whether the receiver deserves to have that information.
3) Differentiating between a simple task such as, say, checking personal social networking platform and circulating confidential company information on the network. The software will learn the difference between such tasks and will have a ranking system to ascertain the “threat level”.

All the datasets and databases required for this software would be regularly updated.

How to get there:

This can be achieved through supervised learning of daily activities on the network. As far as data sets are concerned, many open source repositories and websites such as https://vizsec.org/data/ can be used for this purpose.
Although making a software for the whole company network might not be feasible in the given time and resources, one which works for a comparatively limited set of activities is still within reach and hence, is the aim for this project.