PERSON VERIFICATION USING LINGUISTIC PROFILING FOR CONTINUOUS AUTHENTICATION BASED ON ML

Murtaza Anverali Esufali

A dissertation submitted in partial fulfilment of the requirement for

MSc (Hons) Cyber Security and Forensics Degree

Department of Computing

Informatics Institute of Technology, Sri Lanka
in collaboration with
University of Westminster, UK

2020

Abstract

Usage of mobile devices has increased exponentially over the past decade. The wide use has brought with it a variety of concerns including security. Currently, the most widely used method of authentication for devices is PINs, biometric fingerprint scanning and so on. These methods and their limitations have given rise to continuous authentication mechanisms. This involves verifying a user implicitly and continuously without causing a hindrance i.e. in the background while the user is using his/her device.

Person verification is the process that involves verifying a person uniquely using specific traits they have. Various advancements have been made to carry out this type of verification including usage of linguistic, behavioral profiling, tracking user location, and gait dynamics and many more. Linguistic profiling involves using the user's stylometry for verification i.e. verify using his/her writing style.

This research uses various machine learning techniques combined with dynamic feature extraction to create a platform that can be used by operating systems and chat application developers to implement continuous authentication using linguistic profiling. An additional use of this platform includes forensic analysis.

The platform developed accepts WhatsApp chats which it then pre-processes, extracts static and dynamic features, and builds relevant models. The models are built using the Support Vector Machines (SVMs), Multi-Layer Perceptron (MLPs) and Ensemble algorithms. This dynamic feature extraction process identifies the most common words, emojis, word extensions etc. used by the individual. Chat messages can then be posted to predict if the message is from an imposter or the actual user. The results obtained have been commendable. It includes an accuracy of 75.81%, FAR of 8.62% and FRR of 15.58% respectively.

Keywords: Continuous authentication, linguistic profiling, author verification, machine learning, WhatsApp, stylometric analysis, supervised learning