AimBrain Homework Project

1 Introduction

At AimBrain, we are developing a multi-model mobile focused biometric authentication platform. One of the modalities we are actively developing is behavioural biometrics.

Using behavioural biometrics we track how the user interacts with their mobile application in order to authenticate them. We monitor features such as touch pressure, touch location and timing. Because of physiological (finger length) and psychological (rhythm of typing) factors, these features can be used for authentication.

In this project you will be asked to propose and implement an authentication method using the provided dataset. The accuracy of the method is not as important as the approach you will take to solving the problem.

2 Scenario

In this project you are dealing with mobile banking application. The application has a pin code screen and you are asked to add an additional layer of biometric security, without prompting the user to do anything differently.

Behavioural biometric based on touch input is selected as the way to approach this. A dataset described below is collected to start developing the algorithms for enrollment and authentication.

The algorithm needs to produce a score that would represent how certain we are that it is the correct user entering the pin code.

3 Dataset

In this project you are given a dataset from 5 users using a LG Nexus 5 device to enter an identical for everyone 6 digit pin code and the confirmation key. Each instance consists of 7 touch events, which in turn consist of a touch down event when pressing the screen and touch up event when releasing. There are 50 instances for each user, split into a 40 instance training set and a 10 instance testing set.

Each instance consists of:

- user ID;
- device screen height;
- device screen width;
- unix timestamp;
- 7 instances of the following for each touch:
  - touch down time;
  - touch down x coordinate;
  - touch down y coordinate;
  - touch down size;
  - touch down pressure;
  - touch up time;
  - touch up x coordinate;
  - touch up y coordinate;
  - touch up size;
  - touch up pressure;

For this project the data has already been cleaned and we guarantee that the correct pin code with no errors has been entered each time. However, solutions that deal with users making a mistake and correcting it are highly encouraged.

When testing the solution, we will add 3 additional users that have not been observed to the testing set, so your algorithm needs to be able to deal with previously unseen users. This simulates an attacker trying to compromise and account.

4 Submission

You are provided with the following files:

- **main.py**
- **auth_algorithms.py**
- **dataset_training.csv**
- **dataset_testing.csv**

Please implement your solution in **auth_algorithms.py** and add any additional helper files you need.

We want you to submit the solution code based on the provided template and a brief report explaining the choice of the algorithm, how it could scale to millions of users, how it deals with previously unseen attackers. Again, the actual accuracy you get isn't as important as the approach you take.

Please use Python as the programming language and mention any 3rd party libraries you used (e.g. scikit-learn) in the report. Use the pdf format for the report.

Send the code and the report in an .zip archive to alesis@aimbrain.com, and I'll get back to you shortly to arrange a time to discuss your approach.

Feel free to ask any questions regarding the data or the problem via Skype (alesis_n.) or e-mail.