|  |  |
| --- | --- |
| Module: |  |
| Module Tutor: |  |
| Assignment Title: |  |
| Student Name: |  |
| Student University Identifier: |  |
| Course of Study: |  |
| Supervisor: |  |
| Second Marker: |  |
| Review Date: |  |
| Word Count (Only Text): |  |
| Word Count (Total): |  |

Table of Contents

[Title 3](#_Toc65997386)

[Aim 3](#_Toc65997387)

[Problem Statement 3](#_Toc65997388)

[Background 3](#_Toc65997389)

[Scope 4](#_Toc65997390)

[Objective 4](#_Toc65997391)

[Methodology 5](#_Toc65997392)

[References 5](#_Toc65997393)

# Title

Identifying Anomalies in 2-Factor Authentication with AI.

# Aim

To develop a third approval layer including arranged Machine Learning models prepared to perceive peculiarities after the main affirmation stage.

# Problem Statement

With the advent of technology, security and safety of our electronic lives isn’t a far-fetched. The only problem is that the development in security technology is a double-edged sword due to the freedom of access people have for technology. Thus, hackers also get their hands on this technology easily and find ways to get past it. Two factor-authentication (2FA) is one of the most innovative security measures to prevent hackers and malice approach our social accounts. 2FA requires a third-party app or device to verify the account details during login. Hackers have found a way to replace the third-party side with their own devices to get access to details required.

To prevent this, a novel solution of creating a third-layer in the 2FA that will detect anomalous attempts using machine learning will be implemented. This will be referred to as the 3FA or the 3-Factor Authentication.

# Background

When the Twitter account of its own chief executive, Jack Dorsey, was taken over by hackers in 2019, a stream of tweets with racial slurs, profanity and praise for Adolf Hitler were posted for 30 minutes. Weeks later, the food writer and campaigner Jack Monroe lost £5,000 from bank and payment accounts accessed from a hijacked phone. Both were victims of “sim-swap” fraud, a scam that has mushroomed in the last few years and has led to victims losing thousands, often before they even know anything is amiss. Fraudsters take control of a mobile phone account through a mixture of confidence tricks and online stalking, and then use those details to get access to the owner’s bank accounts.

Figures from Action Fraud, the national fraud reporting center, show the number of people falling victim to this type of scam has increased substantially since 2015 and that it has resulted in losses of more than £10m to UK consumers. So how can you ensure that your phone, and therefore your bank details, are safe? Variously called sim splitting, simjacking, sim hijacking and port-out scamming, the fraud focuses on moving control of someone’s phone account from their sim card to one controlled by the criminal. Although mobile phones and security measures have changed over the five years since the scam has come to prominence, the way the fraud works has remained consistent.

The tactic hasn’t changed significantly over the years, criminals obtain a victim’s personal information – bank details, address, etc. – by trawling through social networks or by mining data stolen during the breach of an online company’s systems. They then contact the victim’s mobile phone provider, pretend to be the victim, request a sim swap and change personal settings. In some cases, fraudsters work with an insider to assign the victim’s number to another sim. A more recent, tactic is to request a porting authorization code [PAC] to port the victim’s number to a different network. Once the hacker owns the victim’s number, they are able to intercept bank authorizations sent via SMS – or other codes that the mobile number is used for.

Often the fraudster will use information that has been put up on social networks, such as a mother’s maiden name, a birthday or the name of a pet, to help build up an information base on the victim. A victim was identified whose number was stolen by a criminal who used the victim’s identity to request a PAC to transfer it to the criminal’s phone. Payments of more than £1,000 were then made from the victim’s bank account to an online money transfer service. Since the scam emerged, the number of cases has risen rapidly.

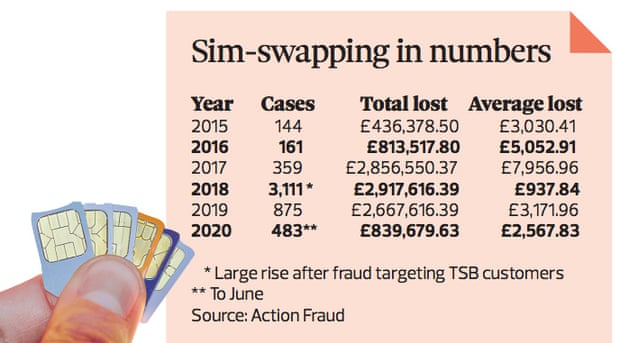


Figure 1 - Simswapping Cases in the UK for 2019-20

# Scope

* Research will be focused solely on the development of the Machine Learning layer required for detecting anomalies.
* A 2FA model will be constructed and the Machine Learning layer will be attached to it for a complete proof of concept software.
* A document explaining each detail of the project will be written.

# Objective

* Delivering a Machine Learning layer for the 2FA that can detect anomalies during login from the third-party side.
* Investigate the cause and the trend behind attacks on 2FA by the hackers using data visualization.

# Methodology

The methodology that would be followed is not inspired from any other existing project because to my knowledge, this project hasn’t been done, or if it has been approached, I couldn’t find any sufficient documentation for review. The approach would be fairly simple. The problem would be treated as an Anomaly detection problem. A dataset would be prepared either by downloading one from the internet or collection data through the internet from various hacking cases and compiling the results. Various machine learning algorithms would be trained and the model with the best performance would be selected for continuation.

## Project Modules

### Front-End

The project would have a proof-of-concept implementation of 2FA as a web app. To interact with this playground, there would be a front-end where the user can interact with 2FA and test it out. The Front-end would be built using Flask, a python-based front-end framework along with basic HTML and bootstrap.

### Back-end

The back-end of the project consists of many sub-modules. All of the back-end will be coded in python. The sub-modules are as follows:

#### Play-space 2FA

There would be a 2FA system implemented for testing and comparing how the addition of Machine Learning works. This would be implemented in either JavaScript using Node.js and Express or Python. This has yet to be decided since this is just a playground for testing and will act as a separate API that can be accessed. If during development it seems like a hassle to create this independently, this will be made using python and integrated with the project.

#### Machine Learning Ensemble

##### Algorithm

The machine learning algorithms/ensemble will be present in the back-end pre-trained. The training would be performed once and the instance would be saved in the back-end server hosted on heroku. The machine learning algorithms used would be from existing libraries like Tensorflow, Keras and Scikit-learn.

##### Dataset

The dataset used would be created from scratch using the internet news and incidents as a base of the research for collecting data. The research would then be compiled into a .csv file that would then be used as the dataset for training the machine learning ensemble. Python libraries like pandas would be utilized for data preprocessing. For data visualization, libraries like seaborn and matplotlib will be used. This would help gain more insight into the dataset and its inherent biases therefore making it easier for the reader to comprehend the results.

# References

1. Hickey, S., 2020. *Sim-swap fraud is on the rise. How can you stop it happening to you?* [Online] the Guardian. Available at: <https://www.theguardian.com/money/2020/sep/13/sim-swap-is-on-the-rise-how-can-you-stop-it-happening-to-you> [Accessed 4 March 2021].