



Transport Simulation Platform as Synthetic Data Generator

Octavio del Ser Perez - Supervised by: Konstantinos Markantonakis and Raja Naeem Akram
Information Security Group, Smart Card and IoT Security Centre



The Smart Card and Internet of Things Security Centre

Objectives

The purpose of this project is to construct a simulation based on a public transport system using open source information to later extract any required synthetically generated data.

- Explore methods of obtaining available public transport data in a systematic way.
- Created techniques to apply the above data in a way relevant to this project.
- Finally use the above to create a simulation that approximates real world transport as closely as possible given the data obtained.

Introduction

Public transport represents a complex network of routes aimed at efficiently transferring people or material from one location to another over a large surface area. Networks can span over a village, city, country or the whole world. There are many methods of simulating transport with very different use cases, usually a compromise has to be made between scale and accuracy of the simulation. Simulating a single road you should take a single person's reaction time into account and change in temperature of the road's surface as well as many other details, however increase the simulation to the scale for a city or country and there is simply not enough computational power to do the same. I decided to simulate subway systems, they are much smaller in scale than rail networks and less complex than bus transport systems which need traffic to be simulated.

Materials

The following materials were required to complete the research:

- Computer, Pycharm, Internet.
- Google API key for static maps (Free).
- Json from Overpass turbo for data on subway routes.
- Python with some graphical plugin such as pygame for display of data.

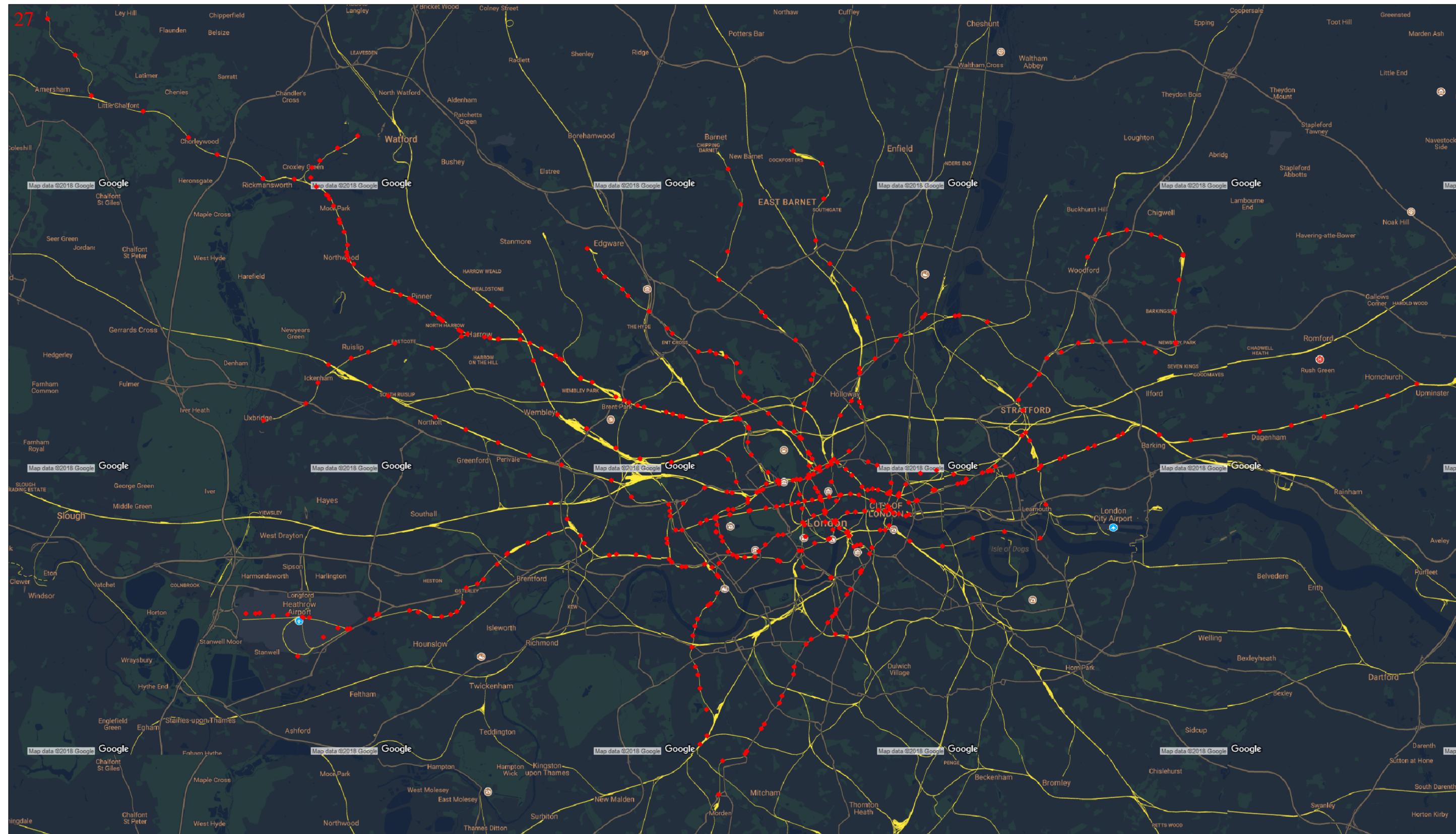


Figure 1: Transport Simulation Representing the Tube Network of London Managed by Transport for London (TfL)

Important Result

Results from the simulation can be streamed live to any platform. Data can be obtained on traffic on any line or node in real or accelerated time. Furthermore, population distribution and their travel pattern can be generated to simulate an operational transport network.

The materials were prepared according to the steps outlined below:

- I used my computer with a linux environment.
- Internet access is everywhere nowadays
- Google API key can be obtained by creating an account, it's free and you have 25000 free api map requests daily.
- Overpass turbo is open source as an extension of Open street maps.

Methods

I had continuous loop for the simulation from which I called various methods for calculating and displaying relevant data.

- A method mapLoader() was used to detect when map data needed to be loaded and threads were started accordingly storing the information offline for reduction in repetitive data use. The method also displays the relevant tiles.
- The class Train() generates a train object given two lines for it to switch between. the train will automatically follow these lines given the acceleration and max velocity given as input.
- trainLoader() is a method for scheduling trains
- Many other methods are used to simulate the environment such as key handlers, the camera object, sprite sheet controller and sprite animator e.t.c. these take care of appropriate object scaling and location during the simulation.

Conclusion

I found that open source data is relatively unreliable for more accurate simulation so a lot of dedicated code has to be written to make it useful. The relevance of produced synthetic data heavily relies on the accuracy of a simulation. The use of open-source data should thereby be used and tested carefully in order to prevent unnecessary complications.

Additional Information

This project is related to the EPSRC funded project "Data to Improve Customer Experience (DICE)". The project is particularly interested in personal data, and is using rail passengers as a specific focus of interest. The overall aims of the project are:

- Understand the role that personal data plays in enhancing the user experience of rail passengers
- To develop technical solutions to data privacy
- To develop an evaluation framework that can be implemented so passengers can understand how their data is used and how they can control and verify its use.

The project started in October 2016, and runs for three years to September 2019. For more information about the project, please visit <http://www.dice-project.org>

Acknowledgements

We acknowledge the support of the ISG-SCC for the summer internship program and EPSRC funded project. The views and opinions expressed in this poster are those of the authors and do not necessarily reflect the position of DICE project or any of partners associated with this project.

Contact Information

- Web: <https://scc.rhul.ac.uk/>
- Email: octaivio.delser@gmail.com