The University of Hong Kong
Faculty of Engineering
Department of Computer Science
COMP4801 Final Year Project 2018/19

Project Plan

Topic: Utilising Smartphone Application Data
Student: Wong San Yu (3035273022)
Supervisor: Dr T.W. Chim
# Table of Contents

- **Introduction** .................................................................................................................................................................................. 3
- **Problem Statement** .......................................................................................................................................................................... 4
- **Project Background** ........................................................................................................................................................................... 4
- **Related Works** .................................................................................................................................................................................. 5
- **Project Objective** ............................................................................................................................................................................... 6
  - Scope .............................................................................................................................................................................................. 6
- **Methodology** .................................................................................................................................................................................... 6
  - Network Sniffing ................................................................................................................................................................................. 6
  - Reverse Engineering ........................................................................................................................................................................ 7
  - The Showcase Application .................................................................................................................................................................. 7
- **Challenges** ....................................................................................................................................................................................... 8
- **Project Schedule** ............................................................................................................................................................................... 8
- **Conclusion** ....................................................................................................................................................................................... 9
- **References** ..................................................................................................................................................................................... 9
Introduction

It is now a decade after the Apple iPhone debut in 2007. In the 10 years, Apple and Google have put a lot of resources in developing smartphones. Their effort have reflected in customer reactions. Smartphone has become a very popular item across the globe. Many people are using smartphones on daily basis. What makes smartphones better than the old feature phones is the world of smartphone applications. Users can browse and obtain millions of different applications from official and unofficial application stores. In 2017, there were 178 billion application downloads worldwide [1]. With the applications, users can entertain themselves with mobile games, enhance productivity with scheduling apps, making use of small tools to make their lives easier, and there are so much more. Many applications are popular among users because they provide useful data. However, most of the companies do not provide open data and it is not convenient to get data from several applications in a short time. This project aims at relieving this situation by exploring methods to make data retrieval from multiple applications easier. In the project, the focus will be on data of Estimated Time of Arrival of buses, as it is a problem often faced by general public.
Problem Statement

The everyday commute has always been a problem for working people and students. In the past, passengers have to wait for unknown time at stations for buses. With the introduction of Estimated Time of Arrival (ETA) of buses in bus companies’ apps, commuters know when does the bus come in advance and that helps in time management. However, people often have to enquire ETAs from different apps for different bus routes. For example, a user checks ETA for a cross harbor route jointly operated by KMB and Citybus. Then the KMB app says it is now non-KMB hours. Therefore the user has to open Citybus app and check ETA again. Changing apps back and forth could be annoying. A universal tool gathering ETAs from different bus companies will be favourable for commuters.

![Figure 1. Users have to switch apps to get real time data](image1.jpg)

Project Background

Although users can get ETAs from the bus companies’ apps, the data are not open to public. Bus companies claim that the ETA data are their private property [2] so that they will not provide any means (API accessing real time data and/or static data) for users to retrieve ETA data other than using their app. However, the bus company apps are often slow and in some sense not very user friendly. As a heavy user of bus services, a transportation enthusiast and CS student, the abovementioned situation of having to switch apps always appear in daily life. Therefore this project will be interesting for me and it will benefit other passengers.

![Figure 2. Bus companies use standalone apps to provide ETA data](image2.jpg)
Related Works

There are few integrated transportation applications available on the market. They provide one-stop route suggestions and mixed real time and non-real time ETA. However, as a user, non-real time ETA is not preferred. Following is a general comparison of those apps:

  Google Maps is a popular way-finding app which provides route suggestions and basic schedules of bus routes and MTR lines. However, the information provided is often not accurate and it does not provide any real time ETA data for the user.

- Citymapper (https://citymapper.com/hong-kong)
  Citymapper provides route finding function in several locations over the world including Hong Kong. It integrates most of the public transport in Hong Kong including red minibuses into their service. However, users can only get real time ETA for KMB and LWB routes. For other bus companies, the app still provides ETA but they are only calculated based on the official schedule from bus companies. The calculated ETA are not accurate.

- Moovit (https://moovit.com)
  Moovit is a service similar to Citymapper. Moovit also provides route suggestions for most kinds of public transportation in Hong Kong. It also supports push notifications to alert users on incidents. The ETA system is the same as Citymapper with only KMB and LWB routes having real time ETA.

This project is not going to compete with the above apps, but try to improve their weakness, which is not providing real time ETA for companies other than KMB and LWB.
Project Objective

This project aims at finding ways to centralise the process of ETA data request and retrieval from different bus companies within one single application. As the ETA data are not public, different approaches have to be tried in order to retrieve data from companies’ servers. If the way is found, a platform will be built for users to view real time bus ETAs retrieved from different companies, eliminating the need of switching apps to get real time ETA. This will be a big step forward as this will be a new channel for users to get required data easily, which makes their life easier.

Scope

This project manipulates ETA data from bus companies’ apps and combines them into one single application. As the focus shall be on the techniques to get data, not the breadth of data, only ETA data from Kowloon Motor Bus (KMB), Long Win Bus (LWB), New World First Bus (NWFB) and Citybus (CTB) will be covered as there are actually just two data sources. If time allows, ETA data from New Lantao Bus (NLB) will also be covered.

Methodology

For this project, the underlying request/response cycle of bus companies’ apps have to be understood first with the help of some existing tools. There are two main approaches to understand the process, namely Network Sniffing and Reverse Engineering.

Network Sniffing

The project will start off by trying network sniffing. Network Sniffing means studying the real time activity of a computer network [3]. The apps must send and receive data packets to/from bus company servers in order to get the ETA data. If the content of the data packets can be exploited, we can study the logic behind, send similar requests to bus companies and get the ETA data needed. For network sniffing, the software Wireshark will be used. Wireshark is a free open source software for analysing data packets in a computer network [4]. With the help of Wireshark, the data packets transmitted by apps can be traced. Contents, or at least destinations, of the packets can then be explored. The sniffing will be run on a
computer with Android emulator. Bus companies’ apps will be installed in the emulated Android environment and their network activity will be logged by Wireshark on the host machine. The logged data can then be studied later.

Reverse Engineering

If network sniffing does not work, or the results gathered are not sufficient, reverse engineering will be tried. Reverse Engineering means cracking down the application back to lines of code. Android APKs are just simple ZIP files containing binary codes. Binary codes can then be converted back to source codes for modification. With the source code, the ultimate logic of the app can be understood clearer. Non-public APIs might also be found. Binary codes can be converted to different other forms using tool like apktool and dex2jar [5]. After understanding the program logic, the ETA data request/retrieve process can be emulated for later use.

The Showcase Application

After studying and successfully obtained ETA data from both companies, the showcase application will be built. There will be two main functions to be showcased:

1. ETA of jointly operated routes
   Currently, the apps will show “it is another company’s operating cycle” (or equivalent) if the coming departures are from another company. In the showcase application, this message will not show up anymore and ETA of both companies can be shown in one single app.

2. Showing ETA of routes operated by different companies in same bus stop
   Similarly to above, for bus stops having routes operated by different companies, ETA of all routes stopping at that station will show up.

   As the network sniffing and the reverse engineering are done under Android environment, the showcase application will also be developed as an simple Android application for the ease of development. Users can enter router numbers to see ETA data.
Challenges

1. After converting APK into source codes, the original naming of variables might be changed, might need to study the whole program to understand what do the variables store and their function.
2. ETA data collected from different companies might have different formats. Need to find a way to unify them.
3. With the limited time, the showcase application might not have a lot of functions, or the routes have to be hard-coded.

Project Schedule

<table>
<thead>
<tr>
<th>September 2018</th>
<th>Research on tools used in the project</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 September 2018</td>
<td>Submit deliverables of Phase 1 (Inception)</td>
</tr>
<tr>
<td></td>
<td>• Detailed project plan</td>
</tr>
<tr>
<td></td>
<td>• Project web page</td>
</tr>
<tr>
<td>October 2018 – January 2019</td>
<td>1. Study the logic behind the apps</td>
</tr>
<tr>
<td></td>
<td>2. Try getting ETA data without official app</td>
</tr>
<tr>
<td></td>
<td>3. Work on reports</td>
</tr>
<tr>
<td>7-11 January 2019</td>
<td>First presentation</td>
</tr>
<tr>
<td>20 January 2019</td>
<td>Submit deliverables of Phase 2 (Elaboration)</td>
</tr>
<tr>
<td></td>
<td>• Preliminary implementation</td>
</tr>
<tr>
<td></td>
<td>• Detailed interim report</td>
</tr>
<tr>
<td>January 2019 – April 2019</td>
<td>1. Work on the showcase application if study on getting ETA data is successful, continue study on extracting ETA data otherwise</td>
</tr>
<tr>
<td></td>
<td>2. Work on final reports</td>
</tr>
<tr>
<td>14 April 2019</td>
<td>Submit deliverables of Phase 3 (Construction)</td>
</tr>
<tr>
<td></td>
<td>• Finalized tested implementation</td>
</tr>
<tr>
<td></td>
<td>• Final report</td>
</tr>
<tr>
<td>Mid - Late April 2019</td>
<td>Final presentation &amp; Project Exhibition</td>
</tr>
</tbody>
</table>
Conclusion

This project focuses on how to extract and make use of non-public, non-open data from smartphone applications, using ETA data from different apps from bus companies as an example. The project tries to gather the ETA data together in one place in order to benefit users in their daily lives. Network sniffing and reverse engineering techniques will be used and a showcase application will be built. With this project, we hope that more and more applications or companies will open their data to the public and make everyone’s life better.

References


