PERFORMANCE COMPARISON OF GOOGLE GEOLOCATION API AND GPS DEVICE TRACKERS UNDER INDOOR AND OUTDOOR CONDITIONS

FOO JIA YUNG

UNIVERSITI KEBANGSAAN MALAYSIA

PERFORMANCE COMPARISON OF GOOGLE GEOLOCATION API AND GPS DEVICE TRACKERS UNDER INDOOR AND OUTDOOR CONDITIONS

FOO JIA YUNG

PROJECT SUBMITTED IN PARTIAL FULFILMENT FOR THE DEGREE OF MASTER OF CYBER SECURITY

FACULTY OF INFORMATION SCIENCE & TECHNOLOGY
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2023

PERFORMANCE COMPARISON OF GOOGLE GEOLOCATION API AND GPS DEVICE TRACKERS UNDER INDOOR AND OUTDOOR CONDITIONS

FOO JIA YUNG

LAPORAN PROJEK YANG DIKEMUKAKAN UNTUK MEMPEROLEH SEPARA IJAZAH SARJANA KESELAMATAN SIBER

FAKULTI TEKNOLOGI DAN SAINS MAKLUMAT
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2023

DECLARATION

I hereby declare that the work in this project is my own except for quotations and summaries which have been duly acknowledged.

17 August 2023

FOO JIA YUNG
P113856

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project supervisor Ts Dr Wan Fariza Paizi@Fauzi for her continuous patience and motivational support. Her guidance and knowledge input has helped me in successfully completing this project within the recommended timeframe.

Next, I would also like to express my appreciation to all the lecturers of Master of Cyber Security who have taught me throughout the seven modules, namely Dr. Umi Asma' Mokhtar, Dr. Rossilawati Sulaiman, Assoc. Prof. Dr. Khairul Akram Zainol Ariffin, Ts. Dr. Khairul Azmi Abu Bakar, Dr. Kok Ven Jyn and Assoc. Prof. Dr. Masnizah Mohd. They have assisted me a lot in the completion of all the required modules for this course. Their effort has inspired me to carry out this project regarding web client geolocation.

Moreover, I gratefully acknowledge the financial support provided by Ministry of Health under Hadiah Latihan Persekutuan 2021/2022, together with the fully paid study leave offered to me. This assistance has helped me to complete this degree as a full-time student in a timely manner.

Furthermore, to my dearest parents, Mr Foo Sin Yew and Mdm Yee May Yee, and my beloved sister, Ms Foo Jia Ning, thank you for all the moral support that is given to me throughout my study. It is such a blessing to be surrounded by positive energy all the time.

Last but not least, thank you to my fellow course mates for their cooperation who gave their best during all the group works and the organisation of the seminar for TTTX6134 Cyber Law and Ethics. I have learned tremendously from your industrial as well as personal experience regarding computer science and cybersecurity.

ABSTRAK

Penjejak peranti GPS telah digunakan oleh Agensi Penguatkuasaan Undang-undang di seluruh dunia untuk mendapatkan keputusan geolokasi suspek. Walau bagaimanapun, terdapat pelbagai isu mengenai pemasangan peranti ini, antaranya adalah suspek tidak diketahui/dikenali dan kos peranti. API geolokasi Google boleh diguna sebagai solusi alternatif di mana suspek boleh diperdaya untuk melancarkan API geolokasi Google supaya pihak agensi penguatkuasaan undang-undang mendapat maklumat lokasi suspek. Oleh itu, projek ini bertujuan untuk mengetahui sisihan prestasi maksimum antara API geolokasi Google dan penjejak peranti GPS daripada segi perbezaan jarak koordinat geografi yang dijana oleh dua teknologi ini. Berlandaskan pengetahuan ini, ia akan membantu pegawai Agensi Penguatkuasaan Undang-undang kerana mereka boleh mengukur radius carian maksimum daripada koordinat geografi API geolokasi Google. Untuk menjalankan eksperimen, kami mendapatkan longitud dan latitud bagi sepuluh lokasi menggunakan kedua-dua teknologi, dan mengira jarak antara koordinat geografi tersebut menggunakan formula Haversine. Keputusan menunjukkan sisihan prestasi maksimum antara kedua-dua kaedah adalah sekitar 2.5 km tanpa mengira dalam ataupun luar bangunan. Keputusan ini sepadan dengan tuntutan Google mengenai ketepatan API geolokasinya. Penemuan lain adalah kehadiran kumpulan orang yang besar akan menjejaskan ketepatan API geolokasi Google disebabkan oleh cara pengiraan geolokasi. Oleh itu, jika kita menggunakan API geolokasi Google, adalah disyorkan untuk mendapatkan maklumat risikan lain yang penting dan relevan tentang suspek untuk mencapai penjejakan geolokasi yang lebih berjaya.

ABSTRACT

GPS device trackers have been used by Law Enforcement Agencies (LEAs) around the world to get geolocation results of suspects. However, there are various issues surrounding the planting of these devices, some of which include unidentified suspects and device cost. Google geolocation API can be used as an alternative solution whereby suspects will be tricked into launching the Google geolocation API so that LEAs can get the suspects' geolocation information. Therefore, this project aims to find out the maximum performance deviation between Google geolocation API and GPS device trackers in terms of distance difference of their generated geographical coordinates. By having this knowledge, it would be of tremendous help to LEA officers because they can gauge the maximum search radius from the geographical coordinates of Google geolocation API. To conduct the experiment, we get the latitude and longitude of ten locations using both technologies, and calculate the distance between those geographical coordinates using Haversine formula. The results showed that the maximum performance deviation between both methods is around 2.5 km applicable for indoor and outdoor environment, which matched the claim by Google regarding the accuracy of its geolocation API. Another finding is that the presence of large crowds will affect the accuracy of Google geolocation API due to the way in which it works in geolocating places. Therefore, if we were to use Google geolocation API, it is recommended to get hold of other important and relevant intelligence about the suspects to achieve a more successful geolocation tracking.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Figure No. **Page**

LIST OF ABBREVIATIONS

AGPS	Assisted-GPS
API	Application Programming Interface
AV	Autonomous Vehicle
CEP	Circular Error Probability
COD	Cash-On-Delivery
FSPL	Free-Space Path Loss
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
INTERPOL	International Criminal Police Organization
IoT	Internet of Things
IP	Internet Protocol
LBS	Location-Based Service
LEA	Law Enforcement Agency
LoRA	Long Range
MAC	Media Access Control
MT	Mobile Terminal
OGA	Other Governmental Agency
PAEE	Physical Activity Energy Expenditure
PII	Personal Identifying Information
RFID	Radio Frequency Identification
TTFF	Time-To-First-Fix
UKM	Universiti Kebangsaan Malaysia
URL	Uniform Resource Locator

CHAPTER I

INTRODUCTION

1.1 RESEARCH BACKGROUND

The phenomenon of online shopping is becoming more prevalent in this digital age. The Covid-19 pandemic has just expedited this behaviour exponentially with all the movement restrictions imposed across the globe. It has transformed our lifestyle tremendously, from online shopping, online consultation, to online sales gallery. Online business has flourish to an unprecedented level as many organisations are moving their businesses online to keep in touch with their existing as well as potential customers to stay competitive in the market. On top of that, any person with an Internet-connected device is able to sell things online, thus relatively lowering the threshold of conducting business. As a result, there is a sudden surge of people entering this business field without the proper knowledge of relevant laws and regulations of the country. Not to mention job cuts in certain industries amid the pandemic as well as looming inflation and recession risk in the current situation, have forced people to seek different avenues to make ends meet.

The contemporary business platforms such as Shopee leverage the account information of both sellers and buyers to provide verification of their actual identities. When online transactions are done in the absence of verification authority, especially for those dubious websites, there would be instances whereby their identities become questionable. Sellers of illegal products would exploit this weakness as they can hide behind this regulatory loophole. The trend of selling unregistered products online has been increasing over the years, especially for health-related (pharmaceutical) products. It is a lucrative business considering terminally ill patients are willing to try various remedies alleged and offered by irresponsible sellers.

To incriminate a suspect, sufficient and complete evidence is required for a successful prosecution. The storage location of the illegal products is the most important information for a law enforcement agency (LEA) in order to achieve a more convincing case. When transactions are done through cash-on-delivery (COD), LEA officers are able to trace the sellers' location by following them. However, some tricky sellers would try to mislead the buyers, making it difficult for LEA officers to catch them. They would try to shrug potential followers off by moving quickly from one place to another, dodging their pursuit. In other instances, there are some sellers who refuse to conduct COD and will only post the products through regular or registered mail which will complicate the process of obtaining their actual location. The sellers can just put an arbitrary address on the postal documents in which LEA officers could not link that address back to the seller.

In the effort of combating these activities, Malaysia is one of the countries that has participated in Operation Pangea for several years. Operation Pangea is a global operation coordinated by the International Criminal Police Organization (INTERPOL) targeting the illegal online sales of pharmaceuticals. Since 2015 to 2022, the value of confiscated pharmaceuticals in Malaysia has increased from RM830,663 to over RM 2 million (Abdullah 2015, 2022). This worsening scenario necessitates prompt and effective countermeasures to be undertaken before it is deemed uncontrollable and brings about serious repercussions. Pharmaceutical crime is considered as one of the major public health issues affecting our societies globally. Consumption of unregistered pharmaceuticals is putting people's health at stake. The growth of online shopping has only hastened this worrying situation (INTERPOL 2022).

There are various conventional methods employed by LEA in tracking individuals' location, such as through profiling, social engineering, and following them furtively. These methods are only useful if the individuals are willing to share and divulge their personal identifying information (PII). With the cooperation from other governmental agencies (OGAs) like National Registration Department and Companies Commission of Malaysia, LEA officers are also capable of extracting some valuable information about them. However, the records in their databases may not be updated, and as such, the information on address might be less accurate to be used for LEA's purpose.

Modus operandi of most online sellers include advertising products through e-commerce platforms such as Shopee and Lazada. Other avenues of advertisement include websites and social media platforms such as Facebook and Instagram. The e-commerce platforms are easier to regulate because they have liaison officers locally in Malaysia who are responsible in controlling and maintaining their company's wellbeing and reputation. These officers will strive to weed out illegal products being advertised in their platforms because the advertisements are under their jurisdiction. In accordance to Regulation 7(1)(a) of Control of Drugs and Cosmetics Regulation 1984, no person shall sell any product unless the product is registered under the Regulation (Legal Research Board 2021). The e-commerce company will be liable to the offence upon conviction. As the online sellers are dependent upon the platforms to reach out to potential customers, the platforms are capable of controlling and imposing various limitations and restrictions on the sellers who are less law-abiding. In fact, the platforms are holding valuable information of the sellers including active contact number and addresses which are useful for LEA to carry out their duties. This PII is submitted by the sellers during their account registration. Some of the information might be incomplete, such as no full names with dodgy addresses, but the bank account information is most likely be available for payment transfer. The address stored in the platforms' database might not be updated.

Websites, blogs, and forums can be set up by sellers but these methods are less common these days due to the success of social media and e-commerce platforms. Nevertheless, these sites are still favoured and adopted by sellers because less information is collected from them. Sellers with ill-intention will supply the least PII to conduct transactions with purchasers. In order to stay undetected, they will provide their mobile number to reach out to potential customers, and bank account number for payment. In terms of websites, WHOIS lookup provides address information of domain name being searched. If the location is within Malaysia, appropriate legal action can be taken against the address. Otherwise, LEA will liaise with Malaysian Communications and Multimedia Commission in order to block the access to that particular website in

Malaysia. This is because the registered address of the website is situated overseas, there will be cross-jurisdictional issues arising. Malaysia's law is not applicable in other countries, thus the only way to control this kind of websites is to restrict their access locally. However, website information from WHOIS database for these tricky sellers may not be accurate as they may forge them. The same can be said for blogs and forums, causing the information to be less reliable. Telco companies and banking companies may have their PII which might not be updated and sufficient to locate the sellers.

Advertisements in social media websites are trickier to handle because most of the websites are from overseas, and each of the social media has their own community standards. As long as the advertisements do not violate the standards, LEA cannot do much in terms of taking down the advertisements. In order to reach out to potential customers, online sellers on social media platforms have a different way of selling their products, unlike in e-commerce and websites whereby there are payment gateways for convenient checkouts. As social media are initially designed for online users to interact socially, they have recently been used to promote products with the advent of Facebook, Instagram and YouTube advertisements. The easiest way for potential customers to communicate with the sellers is through the contact number listed on the advertisement, in addition to bank account number and even address to buyers who opt for COD transaction. In order to gain trust and popularity from potential customers, sellers on social media would post photos of their track record and feedback from previous customers. There will be difficulty in locating those sellers who seldom post photos or are unwilling to conduct COD.

These efforts in attempting to locate individuals are resource-intensive and time-consuming, especially those involving unverified sellers. LEA officers have to spend much time going through social media posts and other sites in order to gather relevant open-source intelligence about the sellers. Responses and feedback from telco, banking, and OGA necessitates extensive procedures to obtain confidential information of their customers. Usually, contact numbers are sufficient for LEA to get hold of the targeted seller because the telco companies will keep a database of their customers which include their registered address. However, in some instances, this information might not be up-to-date, or that they might be deliberately misrepresented by the owner of the contact number. Same goes with other databases such as national registration identity card and bank accounts, the currency and accuracy of the information is not guaranteed. Completeness and updated location of the sellers is very crucial in ensuring smooth operation of LEA. Therefore, the reliability of the sellers' location based on these databases is in question.

Nowadays, there are various applications which capture the current location of online users in order to provide contextualized services. A variety of purposes to gauge web clients' location include content-related restriction, and delivery cost and time estimation (Sommers 2020). These applications also provide localised experience to customers depending on their current position in the map. Web clients in different regions will be customised to receive information relevant to that area. Geolocation is defined as "the use of technology to find the location of an internet or mobile phone user" (HarperCollins Publishers 2023). The term is sometimes used interchangeably with geopositioning (ISO 2018).

There are various ways in which geolocation can be done, namely through Internet Protocol (IP) address or Global Positioning System (GPS) coordinates. There are look-up tables for IP address which map to actual locations, but the database might not be updated regularly, as with the case with other databases seen earlier like telcos' and e-commerce platforms' (Abdou & van Oorschot 2019). Web server logs are demonstrated to be able to collect IP address of web clients as shown in Figure 1.1, whereby the IP address can be mapped to the conventional address of the device that visited the website. Therefore, it is possible to determine the physical address of any web clients in terms of country, state, and city. However, the IP addresses displayed in the logs are considered as public IP address in which they are assigned by Internet Service Providers to the network routers of the individuals. It would only provide an approximate location, which is insufficient to track down individuals effectively.

```
127.0.0.1 - - [03/Aug/2022:14:26:53 +0800] "GET /testget.php HTTP/1.1" 200 291 "http://localhost/" "Mozil
la/5.0 (X11; Linux i686; rv:7.0.1) Gecko/20100101 Firefox/7.0.1"
```

Figure 1.1 Example of web server log showing IP address of 127.0.0.1

Another geolocation technique is by using measurement algorithms. There are two main classes of measurement-based algorithms, which are delay-based and topology-aware (Chadha 2019). Delay-based algorithm correlates the round-trip delay with distances between destinations. However, this method is less reliable because the delay might be brought about by circuitous routes, queuing delay, and network congestions, resulting in distance measurement error (Youn 2013). Topology-aware algorithms overcome the problem by localising the intermediate routers in between the path repetitively, resulting in better delay estimates (Bendale & Ratanaraj Kumar 2014). It was shown that the highest average accuracy is achieved by using Octant algorithm, which is 35 to 40 kilometers. Nevertheless, this topology-aware geolocation algorithm does not work well in metropolitan areas (Chadha 2019).

GPS is deemed the most accurate geolocation technique because it is made up of 24 satellites orbiting the Earth (Roxin et al. 2007), as illustrated in Figure 1.2. In order to geolocate, GPS receivers detect and measure the time taken for signals travelling between the satellites and the mobile unit (Gentile et al. 2012). The receivers are able to calculate the mobile unit's position in terms of latitude, longitude, and altitude (Djuknic & Richton 2001). In the rest of this project, latitude and longitude will be collectively known as geographical coordinates. As satellite signals are required to geolocate, the accuracy of GPS is impeded by buildings and indoor conditions with a difference of few meters (Gentile et al. 2012).

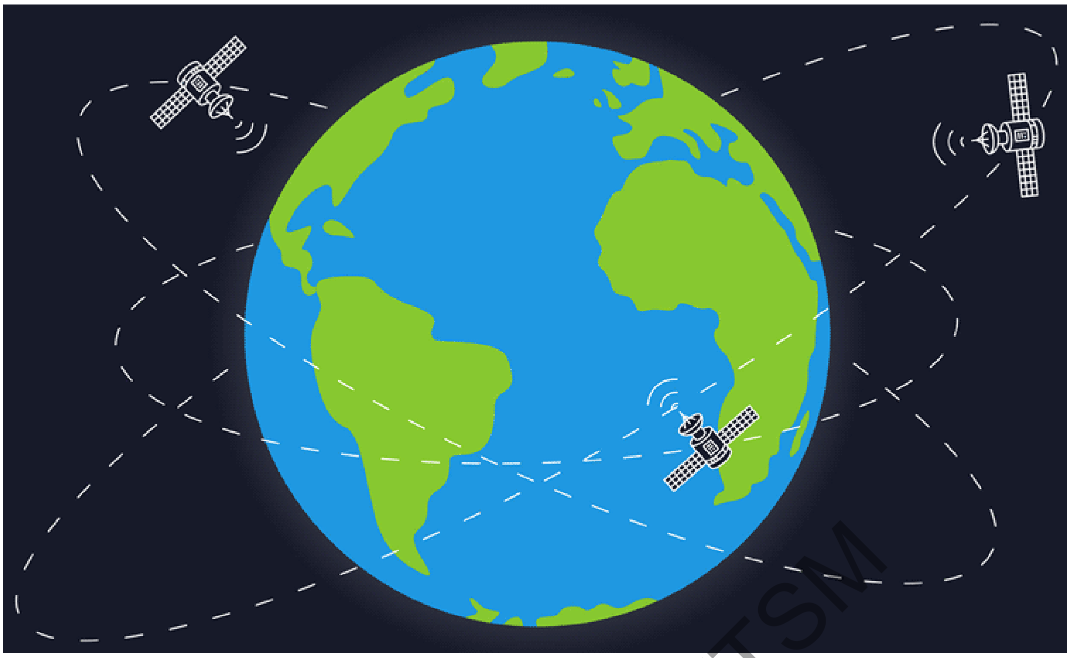

Figure 1.2 Illustration of how GPS works in terms of satellites orbiting the Earth (Geotab 2020)

Network-based geolocation technology provides some advantages over GPS such as quicker time-to-first-fix (TTFF), deployment, and continuous navigation. However, it is less accurate, and many base stations are required in comparison with that of GPS. This technology also raises privacy issues as network operators possess information in terms of time of arrival, multipath fingerprinting, and timing advance of network subscribers (Djuknic & Richton 2001).

The introduction of assisted-GPS (AGPS) overcomes the problems of GPS and network-based geolocation technology mentioned above. Every time geolocation is requested, GPS receivers calculate the TTFF, which is the time taken to establish reliable connection with the satellites. As this connection is refreshed upon request, it would take several minutes to fix the location (Djuknic & Richton 2001). AGPS provides satellite information from base stations to GPS receivers, improving the TTFF, leading to faster geolocation. As base stations calculate the receiver’s location using triangulation technique, removing the burden of mobile device, thereby speeding up TTFF to merely seconds (Gentile et al. 2012). Therefore, AGPS improves GPS’s limitation of indoor accuracy by utilising nearby base stations for geolocation purposes.

There are several indoor geolocation techniques being developed for military, public welfare, and commercial purposes. The fundamental element of this technology involves sensors on mobile terminal (MT), a positioning algorithm, as well as a display unit for locating the MT. In order to achieve indoor geolocation, MT in the form of wearable tags have to be introduced separately from mobile devices or units. Examples of such application of indoor geolocation include tracking elderly in nursing homes, people with disabilities and need supervision, as well as those performing special task forces inside buildings (Pahlavan et al. 2002). For this technology, it is more suitable to be employed amongst known users of the system who will be wearing the MT. It is less relevant to geolocate online sellers who will be most likely strangers to LEA, because they would not be wearing the tags.

Google geolocation Application Programming Interface (API) has been employed in various domains around the world to overcome problems and issues faced in respective countries. The main function of this API is to capture the latitude and longitude of web clients, and their geographical address can be calculated based on those values (Sharma & Morwal 2015). It works by detecting nearby mobile towers and Wi-Fi nodes in order to determine the location of web clients (Google Developers 2023), as illustrated in Figure 1.3. This online service provides an easier way to get the current location of device's users comparatively to previously mentioned methods (Sharma & Morwal 2015). In Indonesia, Nasution and Samsudin (2018) has developed a mobile application utilising this API to locate any criminal incidents being reported by the public. With the use of this application, the police officers are able to get to the crime scene in a timelier manner. Whereas in Ghana, this API has been used to overcome their country's addressing system inadequacies. It provides digital address to users, allowing others to be able to locate them conveniently (Gah et al. 2018).

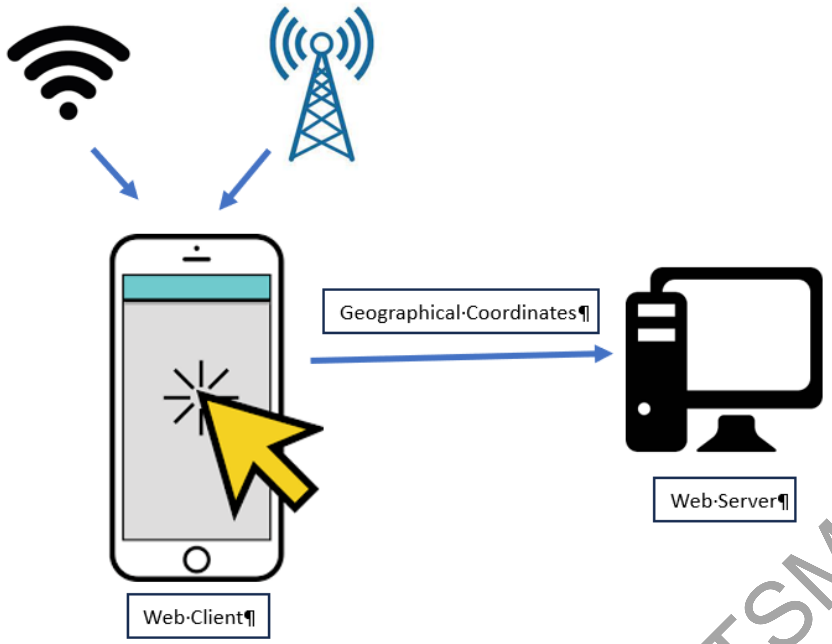

Figure 1.3 Once web clients click on URL link, location details (geographical coordinates) about nearby mobile towers and Wi-Fi nodes will be captured and sent to web server.

1.2 PROBLEM STATEMENT

One of the contemporary methods used in LEA domain is the adoption of GPS device trackers. However, there are various issues and considerations to be taken care of when choosing them, which are the person being tracked needs to be known to the tracker (not a stranger), poor indoor performance among others. As multiple researches have shown the implementation of Google geolocation API for the purpose of location tracking, it might be beneficial and useful to adopt this method in our domain to track the current location of online sellers. It is relatively convenient and conducive for LEA to employ Google geolocation API because all it takes is to trick the sellers to click on our given Uniform Resource Locator (URL) link, which will capture their geographical coordinates, rather than planting GPS device trackers to get the current location of online sellers.

Nevertheless, there has been no research conducted currently on how accurate and reliable is the Google geolocation API in detecting the current whereabouts of web clients. As GPS has been considered the most accurate geolocation technique to date, the performance of Google geolocation API is still unknown in comparison to GPS. The comparison between both geolocation techniques has not been done by other researchers. Should Google geolocation API give at least similar results compared to GPS, then LEA can confidently use this API to obtain seller/suspect's location. And if the results are different, it is of interest to know the difference in distance. Knowing the difference can provide an insight to the LEA officers in performing their search.

1.3 RESEARCH QUESTION

The research question for this project would be: If Google geolocation API is capable of tracking the current location of users, what is its maximum performance deviation from that of GPS device trackers in terms of finding the geographical coordinates of web clients under indoor and outdoor conditions? Performance deviation refers to the difference in results given by Google geolocation API and GPS device trackers. Maximum performance deviation means that the largest result difference between the two technologies.

1.4 RESEARCH OBJECTIVES

Based on the research question made for this project, the main objective is to determine the maximum performance deviation between implementation of Google geolocation API and GPS device trackers in terms of the geographical coordinates of web clients/device users under indoor and outdoor conditions. In order to measure the performance deviation, the geographical coordinates derived from both methods are obtained and the distance between those geographical coordinates are calculated. By having the knowledge of this distance, LEA officers can use it as a reference of the maximum search radius from the geographical coordinates of Google geolocation API. This distance serves as the maximum localisation error from the geographical coordinates of GPS device trackers, which is supposed to be the most accurate geolocation technique in the current era. An illustration of maximum search radius and maximum localization error is shown in Figure 1.4. This information is useful to LEA officers because it is capable of solving the problem stated in Section 1.2 above.

The specific objectives of this project are as follow:

1. To measure the distance between the geographical coordinates of web clients/device users obtained via Google geolocation API and GPS device trackers
2. To validate the results of the distance by doing statistical hypothesis testing

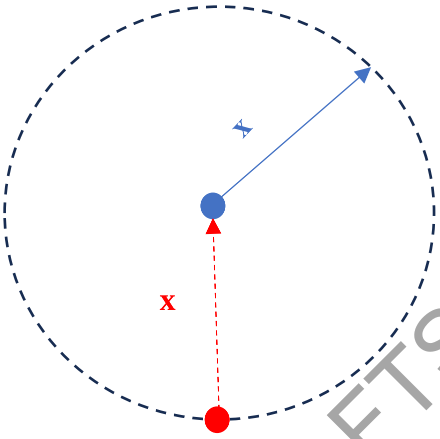

Figure 1.4 Maximum search radius (solid arrow) around the dotted circle perimeter from geographical coordinates of Google geolocation API (blue dot in the middle); maximum localisation error (dashed arrow) from the geographical coordinates of GPS device trackers (red dot on the dotted circle perimeter). Both maximum search radius and maximum localization error are equal in length (x kilometers).

1.5 RESEARCH SCOPE

The scope of this project will be conducted in the area of Kuala Lumpur, covering both indoor and outdoor environment, at 10 selected locations. The geographical coordinates will only be taken during non-rainy days.

As different web browsers will generate slightly different geographical coordinates (Almehmadi et al. 2022; Atencio et al. 2020), Microsoft Edge will be used to capture geographical coordinates incorporating Google geolocation API. This is to ensure the consistency of results from Google geolocation API.

1.6 REPORT ORGANISATION

The remainder of this thesis has been structured as follows: CHAPTER II reports on the literature review conducted; CHAPTER III reports on the research methodology used;

CHAPTER IV discusses the results obtained followed by conclusion and future works in CHAPTER V.

CHAPTER II

LITERATURE REVIEW

2.1 INTRODUCTION

GPS gets location information through satellites orbiting the Earth, while Google geolocation API gets information through nearby mobile towers and Wi-Fi nodes. Information about location has become an important data as well as asset nowadays. Knowing a customer's location means a lot to any business industry in order to provide relevant products and services in a timely and effective manner. Location-based services (LBSs) can be targeted towards nearby potential customers instead of the whole population, thereby contributing to a more optimal cost of production and subsequently saving resources and the environment (Vaughan-Nichols 2009). Advertisements can be localised to cater for the needs and requirements of people around that particular area. Less wastage and destruction are achieved when resources are channelled to the right people who are more welcoming and appreciate them.

Before the advent of computers, people's whereabouts were usually not recorded in detail relative to today's standards. Mode of communication that time included postal, telephone, and telegraph. In order to make a successful communication, the recipient's location must be known to the sender so that the message can be conveyed to the intended receiver accurately. Therefore, people had striven to identify and pinpoint locations as comprehensive as possible and matched them to the identities of individuals. Many efforts accomplished by predecessors include the introduction of global navigation satellite system (GNSS) which consists of American GPS, Russian Glonass, European Galileo, and Chinese Beidou (Szot et al. 2021a). The accuracy of these systems in providing location information is in the range of a few metres (Gauld et al. 2023a). The satellites emit microwave signals to the GPS devices via cellular or wireless network so that the exact location can be measured and calculated by triangulation. In order to get this location information, the line of sight between the satellites and GPS devices must be clear and should not be blocked because the way in which this system works depends heavily on radio signals (Mahanthesh et al. 2022). Any interference between these objects will affect the accuracy of location information.

The widespread use of Internet in the current era has given rise to globalisation, which enables people from opposite parts of the globe to communicate with each other easily. The huge geographical distance between them becomes almost non-existential in the cyber world. Barriers and borders between countries are difficult to stop the communication especially when virtual private network is used. Much information can be exchanged between web clients and web servers. Web clients are able to request information from web servers' database; web servers are also capable of capturing important data regarding web clients in terms of device, date and time, as well as IP address who visited their webpage. IP address reveals a great deal of valuable information regarding web clients' location. Most of the time, this information will be beneficial and useful to customise webpage contents according to country and regions. Webpage and applications which integrate Google geolocation API are also capable of getting hold of web clients' location in the form of geographical coordinates. This information has facilitated the growth and expansion of many LBS which rely heavily on precise geolocation and quick deliverance.

When so much information of web clients is presented and readily available on the Internet, there will be concerns on privacy and safety issues, as well as the way in which that information is handled. People might be worried about location tracking violating and infringing on their personal space. However, this concern is overshadowed by the convenience and practicality of the webpages and applications (Hardy et al. 2018). Most web clients have been less resistant to location tracking after the companies updated their privacy policies regarding personal data protection. This policy exerts a sense of security amongst web clients when surfing the Internet, giving rise to the abundance of location tracking websites and applications.

This location tracking capability of websites will be very attractive to LEA officers because it is very useful and convenient in carrying out their tasks of suspect trailing. Instead of using a physical GPS device tracker, LEA officers might be able to switch to a more appropriate method of location tracking within the search radius, which serves as the maximum localisation error of the actual geographical coordinates. This research is carried out to determine this search radius, therefore a literature review on studies related to geolocation accuracy is performed.

To answer the research question stated in Section 1.2, a systematic literature review has been conducted following the "Procedures for Performing Systematic Reviews" (Kitchenham 2004). The following steps are taken:

1. development of the review protocol;
2. conducting the review using the protocol to identify relevant research;
3. download the relevant papers;
4. generate the results, and
5. write the review findings.

The review protocol developed and conducted starting with relevant literature searched on 7th May 2023 using the following keywords in Google Scholar, Web of Science, and Scopus electronic databases shown in Table 2.1. The reference period was between 2018 and the search date as mentioned above, and all articles except thesis or dissertation were included in the search. These articles are excluded from selection because they are not being peer-reviewed by experts of the field yet.

Table 2.1 Keyword/string used for literature searches

No.	**Keyword / String**	**Results from Google Scholar**	**Results from Web of Science**	**Results from Scopus**
1	"GPS device tracker" OR "google geolocation API" accuracy	99	27	36

From Google Scholar, the keyword gives rise to 99 results; from Web of Science, the keyword gives rise to 27 results; from Scopus, the keyword gives rise to

36 results. The articles were then filtered out based on the title that matched the selected keywords. Those articles that were found related to the study, its references were examined to identify other papers of interest. Duplicated publications were removed from the results of these searches. This resulted in 77 publications that were relevant to the study.

After that, the remaining 77 articles are downloaded and reviewed by reading the abstract, introduction, and conclusion parts to determine the relevancy of the papers to the study. 50 papers were then selected. Finally, the papers were examined and removed based on one of the following exclusion criteria:

1. GPS device or Google geolocation API was not clearly described
2. Prediction of location was utilised

The outcome and results of this final filtering is generated, resulting in 46 articles, because two of them did not express GPS device or Google geolocation API explicitly, while another two articles are focused on location prediction. These publications were then studied to answer the research question and write the review findings.

Both GPS device trackers and Google geolocation API have been used to track users' location over the years across different domains. Nowadays, most smartphones are integrated with GPS systems, enabling the cost of location tracking to reduce significantly while increasing its convenience (Stamatelopoulou et al. 2018). Due to privacy and security reasons, smartphone users have the choice to turn on or off the access to their location for each particular application. Nonetheless, most applications are compulsory to detect users' location in order to provide comprehensive services to the users. Such applications include food delivery, taxis, and road navigation.

Based on the articles selected above, the following subsections will review the domains which utilise GPS device trackers, Google geolocation API, as well as both GPS device trackers and Google geolocation API. This literature review also covers other geolocation technology because the article detailed out the way in which the experiment is carried out. This information is useful in designing this experiment later on, in terms of determining the number of locations.

2.2 DOMAINS WHICH UTILISE GPS DEVICE TRACKERS

Numerous studies have been conducted using GPS device trackers on different domains. The articles involving GPS device trackers are listed in Table 2.3. The following studies merely describe the incorporation of GPS module in various applications, and reports the accuracy of the GPS device in the applied domain. Accuracy here refers to how close the GPS tracker's measured location is to the actual location. For this project, readings from GPS device trackers are treated as benchmark and the focus is on the performance of Google geolocation API.

2.2.1 Wearable Fitness Trackers

One of the domains is the wearable fitness trackers. These trackers are mostly equipped with GPS and accelerometer to monitor physical activity energy expenditure (PAEE). In Adamakis (2020) and Dasa et al. (2022), their experiments were mostly focused on the accuracy of assessing EE between various wearable tracking devices. Adamakis (2020) even compared several freeware GPS-accelerometer-based applications and found out that Runtastic is the most accurate GPS app to measure PAEE during running. Another study is done to determine the heart rate accuracy being monitored by GPS-equipped fitness tracker (Budig et al. 2019). GPS devices are also being compared with accelerometers in detecting intermittent walking (Taoum et al. 2021). Garmin company which introduced the first handheld GPS device as well as GPS-equipped wearable fitness trackers, its different models of Forerunner (FR) has been compared on the accuracy of geolocation. Models of FR 735 and 945 have been repeatedly giving high accuracy results throughout the experiment (Szot et al. 2021). Based on the context of this project, Garmin brand of GPS device tracker can be considered but its wearable fitness trackers are less relevant. Qstarz BT 1000XT GPS device is another potential tracker to be used as it was comparably more accurate than Moves application for the purpose of route identification (Stamatelopoulou et al. 2018).

2.2.2 Population of Dementia Patients

Another group of population which will benefit tremendously from GPS device trackers is the assisted living or dementia patients. Dementia is a medical condition with symptoms of progressive memory loss, leading to difficulty in directional navigation as well as performing other daily activities. Various GPS tracker models have been proposed to assist and guide caregivers in monitoring these patients. Prototypes using GPS module are developed by Baugbog et al. (2020) and Cheriyan et al. (2022). Another study has tested different geolocation technologies such as Wi-Fi and Long Range (LoRA) in wearable Internet of Things (IoT) devices. GNSS produced a comparatively better average location accuracy than Wi-Fi and LoRA technologies (Rodrigues et al. 2020). A review on assistive devices providing location tracking is summarized in Figure 2.1. It has been proven that GPS is much more reliable than Wi-Fi and radio frequency identification (RFID) technologies, but not under indoor conditions (Tyagi et al. 2021). A compilation of commercially available GPS device trackers can be seen in Table 2.2 (Wojtusiak & Mogharab Nia 2021). It is shown that outdoor mostly use GPS while indoor mostly use Wi-Fi. Most of these devices are employing GPS technology due to the reliability results that they confer.

2.2.3 Logistics

Besides that, there has also been adaptation of GPS devices in logistics domain. A study by Leung et al. (2019) looked into the classification accuracy of ground transportation modes using GPS. Another study used GPS data to evaluate different modes of transportation and locomotion (Wang et al. 2021). Koubaa and Qureshi (2018) came up with DroneTrack which is an unmanned aerial vehicle to follow moving objects using GPS technology. Another prototype was also utilising GPS for tracking any type of faulty assets (Lu et al. 2022). These articles focus more on developing solutions with GPS as a supplementary element.

2.2.4 Innovative Tracking Devices

Furthermore, some innovative tracking devices have been designed with discreet mode in mind. To overcome the issues of missing luggage, Farooq et al. (2021) have designed an IoT-based luggage with GPS module. As its name implies, Easy-Weigh-Out also addresses the problem of oversize luggage by allowing constant weight monitoring from Android-based smartphones. Another state-of-the-art robot prototype dubbed The NextGen Bot is integrated with GPS module intended to act as human substitute during emergency and rescue operations (Ramanujam et al. 2022). In terms of security system, Swaroop et al. (2022) has proposed a GPS tracker in IoT to be brought along by girls. Their geolocation accuracy was not discussed in detail in both papers.

2.2.5 Traffic Assistance

Moving on, there are several other articles on traffic assistance domain which employs GPS technology. In addition to geolocation, Najmurrokhman et al. (2021) used GPS tracker to record vehicle speed on the road. The authors have tested the geolocation accuracy at nine (9) different locations, and compared the results with that from Traccar software, which is an open-source GPS tracker. As this device is placed inside a vehicle, the locations selected are most likely to be outdoor, which are Gerbang Tol Padalarang, SPBU Pertamina, Bale Pare KBP, Mason Pine KBP, Ikea KBP, Bundaran Ikea, RS Cahaya Kawaluyan, and Donat Madu Padalarang (Najmurrokhman et al. 2021). Although there are nine locations listed in the table, the location Bundaran Ikea is repeated twice with different geographical coordinates. This shows one of the location names is misrepresented.

Another study tried to extract human driving trajectories using GPS on infrastructure sensors rather than mobile ones. This is done to remove the sense of being observed and thus giving a less accurate result in terms of driving behaviour influence (Notz et al. 2020). A prototype motorcycle helmet was designed with GPS module in order to alert nearby health facilities should any motor vehicle accident happened to the wearer. In terms of location accuracy evaluation, five (5) different places have been selected for the experiment, namely APU University, APIIT University, Standard Charted, LRT Sri Petaling, and Home. The latitude and longitude of these locations are collected, then the results will be verified with Google Maps (Wijaya et al. 2020).

Assistive Navigation Devices and Technologies	Advantages	Limitations	Technologies used in	Images of Assistive Devices
Map Reading System based on RFID	Bluetooth, RFID tag grids Help to offer technical solution	Cost is high, Interference in heavy traffic	Tactile Acoustical Navigation and Information Assistant(GPS+RFID)	
GPS, voice, and ultrasonic sensors	Alert user's recent location and provide verbal directions	Difficult to detect an obstacle and provide warning alert	Kapten Mobility	
Hybrid technology, Smartphone, camera for background motion detection	It does not require any prior knowledge	It requires a lot of processing power, bulky, costly	Tyflos Navigator, NavBeltis	
Ultrasonic Sensors	High-frequency ultrasonic beam to detect obstacles	Less accurate localization due to reflection or blockage of ultrasound signals	Ultrasonic Assistive Headset, white cane	
Laser-based	Beneficial for tracking and detects a targeted object at a long distance	Noise, stray light, and speckle interference create a problem for the laser beam	C5 Laser Cane	
Sonar-based sensors	Contain information of a detected object (geometric shape, size, orientation, and surface material properties)	Targets of low density and echo may be challenging to sense at long-range	Ultra cane, Guide cane, Palm Sonar, iGlass, Pathsounderis	

Figure 2.1 List of Assistive Devices with Location Tracker (Tyagi et al. 2021)

Table 2.2 List of Commercially Available Tracking Devices (Wojtusiak & Mogharab Nia 2021)

Device	Technology*	Indoor/Outdoor	Style
TRX Systems	Beacons, Ultra-wideband	Indoor	Wearable
Yepzon One	Bluetooth, GPS, GSM	Indoor, Outdoor	Necklace
Amcrest	GNSS/GPS, GPRS/GSM	Outdoor	Portable
Trax	GPS, GLONASS	Outdoor	Carriable
Americaloc GL300W	GPS, GSM	Outdoor	Carriable
AngelSense	GPS, GSM	Outdoor	Attachable
GPS SmartSole	GPS, GSM	Outdoor	Shoe sole
Mindme	GPS, GSM	Outdoor	Pendant
Safe Link	GPS, GSM	Outdoor	Carriable
Spy Tec	GPS, GSM	Outdoor	Attachable
Project Lifesaver	RF	Outdoor	Bracelet
Q-Track NFER RTLS	RF, Bluetooth, WLAN	Indoor	Attachable
Accuware	Wi-Fi, Beacons	Indoor	Carriable/Attachable
infsoft	Wi-Fi, Beacons, Ultra-wideband, RFID	Indoor	Carriable/Attachable
iTraq	Wi-Fi, GPS, GSM	Indoor, Outdoor	Attachable
MX-LOCare	Wi-Fi, GPS, GSM	Outdoor	Watch
PocketFinder	Wi-Fi, GPS, GSM	Indoor, Outdoor	Attachable
Mini A9 GPS Tracker	Wi-Fi, LBS, GPS, GPRS/GSM	Indoor, Outdoor	Necklace

*[Beacons - small physical devices, which broadcast a radio signal that can be detected by smartphones; Ultra-wideband - radio technology which uses very low energy level for short-range, high-bandwidth communications; Bluetooth - short-range wireless communication; GPS – global positioning system; GSM – global system for mobile communications; GNSS – global navigation satellite system; GPRS - general packet radio service; RF – radio frequency; WLAN – wireless local area network; RFID - radio frequency identification; LBS - location-based service]

2.2.6 Biologging Technologies

Other than humans, location tracking has been applied to animals as well. Biologging activities are carried out on a variety of animals to study their behaviours which otherwise could not be done by conventional observational methods. In order to secure GPS trackers onto the animals' body, the weight of those trackers must be comparable to each animal's body weight so as not to disturb their normal daily activities. As these lightweight GPS tracking devices tend to be costly and easy to drop off from the studied animals, numerous studies have designed relatively low-cost devices with similar accuracy. Gauld et al. (2023) have designed a prototype for griffon vultures; Kauth et al. (2020) have built a bio-logger for game birds; Yu et al. (2022) developed a GPS tracker for Pacific black ducks. Whereas Lok et al. (2023) studied the prey ingestion rates of spoonbills by comparing the results from various GPS/accelerator devices; Ozsanlav-Harris et al. (2022) studied the incubation duration of Arctic nesting goose species using those devices; Sirotek and Hart (2019) delved into localization technologies to monitor cattle.

Table 2.3 Compilation of Articles which Utilises GPS Device Trackers

	Domains					
Articles	**Wearable Fitness Trackers**	**Population of Dementia Patients**	**Logistics**	**Innovative Tracking Devices**	**Tracking Assistance**	**Biologging Technologies**
(Adamakis 2020)	✓					
(Dasa et al. 2022)	✓					
(Budig et al. 2019)	✓					
(Taoum et al. 2021)	✓					
(Szot et al. 2021b)	✓					
(Stamatelopoulou et al. 2018)	✓					
(Baugbog et al. 2020)		✓				
(Cheriyan et al. 2022)		✓				
(Rodrigues et al. 2020)		✓				
(Tyagi et al. 2021)		✓				
(Wojtusiak & Mogharab Nia 2021)		✓				
(Leung et al. 2019)			✓			
(Wang et al. 2021)			✓			
(Koubaa & Qureshi 2018)			✓			
(Lu et al. 2022)			✓			
(Farooq et al. 2021)				✓		

to be continued …

… continuation

(Ramanujam et al. 2022)				✓		
(Swaroop et al. 2022)				✓		
(Najmurrokhman et al. 2021)					✓	
(Notz et al. 2020)					✓	
(Wijaya et al. 2020)					✓	
(Gauld et al. 2023b)						✓
(Kauth et al. 2020)						✓
(Yu et al. 2022)						✓
(Lok et al. 2023)						✓
(Ozsanlav-Harris et al. 2022)						✓
(Sirotek & Hart 2019)						✓

2.3 DOMAINS WHICH UTILISE GOOGLE GEOLOCATION API

In terms of Google geolocation API, various studies looked into its contribution in terms of geolocation and geographical coordinates extraction. Some of them include the logistics industry (Gupta et al. 2020), particulate matter air sampling (Suarez-Bagnasco 2021), as well as emergency scenarios (Bilgi et al. 2022). Another study incorporated this API into its electronic form known as dCollective to capture interviewees' geolocation information (Eiamboonsert et al. 2018). None of them measured and evaluated the accuracy of the location captured by the API.

There is a statement saying that the geolocation accuracy of Google geolocation API can be up to a few thousand meters (Google 2023). However, there is no further details regarding this accuracy measurement. It is unknown how Google determines the accuracy of its geolocation API which is in the range of thousand meters. This project aims to verify and justify this information in order to provide an alternative avenue in carrying out location tracking duties by LEAs.

2.4 DOMAINS WHICH UTILISE BOTH GPS DEVICE TRACKERS AND GOOGLE GEOLOCATION API

In addition, there are also articles which combine both GPS technology and Google geolocation API in providing industrial solutions. Nyo and Hein (2019) designed an autonomous vehicle (AV) while Kavitha and Ravikumar (2021) looked into speed bumps detection with both geolocation technologies installed in AV. Another study on AV construction also incorporated the technologies as shown in Adnan et al. (2022). It was about a deep learning-based AV on unstructured road conditions. In order to come up with real-time traffic information for all road users, Chen and Yang (2020) developed an instant traffic assistant agent using both technologies. Another study on monitoring Covid-19 safety measures also utilises the technologies to detect ride-sharing drivers' live location. Any driver who violates the standard operating procedures, their locations will be sent to their companies (Rakshit et al. 2023).

2.5 OTHER GEOLOCATION TECHNOLOGY WHICH ASSISTS IN CARRYING OUT THE EXPERIMENT - WI-FI SIGNAL

An unprecedented fabric printed with tracking device's circuit layout and antenna was tested to yield localisation accuracy of up to 8 metres while being worn. Unlike other studies, this article mentioned that a total of 10 different locations in the authors' campus were selected in the experiment to determine the accuracy of GPS between in-air and on-body fabric. The way in which the accuracy evaluation is done by getting the geographical coordinates of the selected ten locations, then the tracking device is placed exactly on these coordinates (Krykpayev et al. 2017). There is no mentioning of how the authors get the actual geographical coordinates of the selected ten locations, either by Google Map or other ways. Subsequently, the device scans its environment for WiFi signals. As each WiFi access point has a distinct Media Access Control (MAC) address, this MAC address is sent to a localisation server computer to get the estimated coordinates from Google geolocation API. The distance difference between the actual and estimated coordinates represents the accuracy of the device (Krykpayev et al. 2017). The smaller the deviation, the higher the device localisation accuracy.

2.6 FACTORS THAT INFLUENCE THE ACCURACY OF BOTH GPS DEVICE TRACKERS AND GOOGLE GEOLOCATION API

There are some external factors which could potentially influence the accuracy of both geolocation technologies. Figure 2.2 lists out six determinants of accuracy, namely device model together with its operating system, the presence of nearby Wi-Fi, Bluetooth- and mobile data-enabled, user's altitude (floor), as well as cloudiness (Hovorushchenko et al. 2021). Plus sign indicates that the factor is turned on; minus sign indicates that the factor is turned off. For example, Samsung Galaxy M20, Android 9 with different factors being turned on, will give different geographical coordinates as shown in rows one to four of Figure 2.2. Considering that these factors are capable of influencing the accuracy and results of geolocation technologies, they should be kept constant throughout this project.

Model of device, operation system	Wi-Fi	Bluetooth	Mobile Data	Floor	Cloudiness	Coordinates in the app	Number of satellites	Assessment of the quest
Samsung Galaxy M20, Android 9	+	+	+	4	low	49.4365334, 27.0120489	9	100
Samsung Galaxy M20, Android 9	-	-	+	1	low	49.436552, 27.012048	8	70
Samsung Galaxy M20, Android 9	+	-	-	4	low	49.4365457, 27.0120476	9	100
Samsung Galaxy M20, Android 9	-	+	-	1	low	49.436747, 27.012059	4	0
Samsung Galaxy J7 2016, Android 8.1	+	+	-	4	low	49.436502, 27.0120723	9	85
Samsung Galaxy J7 2016, Android 8.1	-	-	-	1	low	49.436791, 27.012340	3	0
Samsung Galaxy J7 2016, Android 8.1	+	-	-	3	low	49.436500, 27.012063	8	100
Samsung Galaxy J7 2016, Android 8.1	-	+	-	2	low	49.436623, 27.012270	3	0
Doogee X5MAX PRO, Android 6.0	+	+	-	4	high	49.436572, 27.012173	8	40
Doogee X5MAX PRO, Android 6.0	-	-	-	3	high	49.436690, 27.012228	2	0
Doogee X5MAX PRO, Android 6.0	+	-	-	1	high	49.436501, 27.012093	7	97
Doogee X5MAX PRO, Android 6.0	-	+	-	2	high	49.436702, 27.012231	2	0

Figure 2.2 Impact of External Factors on Geolocation Accuracy

2.7 RESEARCH GAP

Despite their relatively accurate geolocation results, there has not been any studies comparing how accurate and reliable is the Google geolocation API to GPS device trackers in detecting the current whereabouts of web clients. Research on geolocation technologies has been done on various domains such as wearable fitness trackers, population of dementia patients, logistics, innovative tracking devices and so on. However, there is no research being done on law enforcement domain.

GPS device trackers are conventionally being used to track suspects by most LEAs around the world. Most of these trackers are costly and there poses various issues in mounting it to the suspect, such as legal complications. Therefore, alternative method to track them is suggested by means of luring them to click on our given URL link so that we are able to locate them in terms of geographical coordinates. Social engineering tricks and tactics will be used to convince suspects in clicking those links. For the purpose of this research, social engineering will not be discussed further.

In order to use this alternative method of Google geolocation API which is much more convenient and cost saving, it should be compared with the conventional method of GPS device trackers in terms of distance deviation between the geographical coordinates obtained from both methods. As the current practice is using GPS device trackers, its results should be treated as benchmark because generally LEAs have been using them all these while in locating suspects successfully. Thus, if we want to replace GPS device tracker with Google geolocation API, we need to ensure that the results obtained from Google geolocation API is comparable with that of GPS device tracker. Literature review done thus far has helped and guided me in carrying out this project to answer the research question stated in Section 1.2.

2.8 JUSTIFICATION OF EXPERIMENTAL SETUP

In order to reach the research objectives stated in Section 1.3 above, an experiment will be carried out systematically with reference from previous works. Justification of this experimental setup can be seen in the following subsections.

2.8.1 Choice of GPS Device Tracker

Qstarz BT 1000XT GPS device has been proven to be more accurate in identifying routes taken by the research subject (Stamatelopoulou et al. 2018). In another study, this model is said to have many strengths including high accuracy, long battery life, large data storage, as well as decent signal acquisition rate (Schipperijn et al. 2014). Vorlíček et al. (2021) also found out that the location accuracy of this GPS device tracker is comparable to Garmin Forerunner 35 smart watch and Holux RCV-3000. Even though Vorlíček et al. (2019) showed that Holux RCV-3000 is better in terms of accuracy, the merits of Qstarz outweigh that of Holux from the perspective of law enforcement domain, especially the battery life and data storage. This is because the device needs to have longer battery time and subsequently bigger data storage during covert location tracking.

However, this brand has been discontinued from the market, and it is no longer available to be used in this experiment (Qstarz International Co. 2013). Holux RCV-3000 is also not available in Malaysia. As Garmin Forerunner 35 is a watch rather than handheld device, its relevancy in law enforcement domain is less. TKStar GPS tracker is another alternative device which is comparably quite accurate. This brand is available in Malaysia and it has been touted to be quite accurate in location tracking (Shomer 2023). There is also a Youtube video reviewing TKStar brand of GPS device tracker as the best vehicle GPS tracker in the market in 2020 (great gadgets G.G 2020).

2.8.2 Number and choice of locations

From the literature above, there were five, nine, and ten different locations being suggested to gauge the accuracy of geolocation technologies. However, only the location names were listed out without detailing the specifics regarding the surrounding environment. Duncan et al. (2013) also carried out an experiment to assess GPS accuracy in six (6) different environment conditions which will influence the results. These conditions include unobstructed open sky, under a survey beacon, residential, mixed land use, high-rise buildings, and obstructed under a metal canopy. Another article also selected six geodetic locations with detailed description as shown in Table 2.4 (Vorlíček et al. 2021).

However, there is no article suggesting indoor environment to carry out the GPS tracker accuracy assessment. This is because GPS signals will be impeded indoors, therefore the geographical coordinates obtained would not be accurate as compared to outdoor location tracking. This project aims to verify whether Google's claimed accuracy of up to a few thousand meters is also applicable to indoor condition by measuring the distance deviation between geographical coordinates of GPS device trackers and Google geolocation API.

Table 2.4 Six Geodetic Locations with Detailed Description (Vorlíček et al. 2021)

Location	Environment Description
Historic Center	Historic buildings, narrow streets, multi-story houses (usually 4 floors), no vegetation cover, poor sky visibility;
Residential (Family Houses)	Single-family houses, mostly multi-story (usually 2 floors) that line the roadway with sidewalks, vegetation in the form of hedges, good sky visibility;
Open Space	Gravel road on the edge of the city, no vegetation cover, no buildings close by, excellent sky visibility;
Residential (Periphery)	Isolated family houses, mostly multi-story (usually 2 floors), lower housing density, roadway without pavements, generally less compact development with more green space (e.g., gardens), vegetation also in the form of hedges and trees, good sky visibility;
Housing Estate	Multi-story blocks of flats (usually 5 floors) with the usual spacing between buildings, complemented by tall coniferous and deciduous trees, good sky visibility; park—park in historic center, close to medieval city walls (height approx. 5 m), large number of tall deciduous trees, poor sky visibility

2.8.3 Data Collection and Processing

There are various methods in which data of GPS device trackers are being collected. In Schipperijn et al. (2014), four modes of transportation are employed in each four designated routes to record data on 300 trips. The devices are worn around the waist and concealed by clothing. The geographical coordinates are then collected using open source bt 747 GPS software. Spatial join function in ArcGIS is then used to group coordinates for each trip according to 2.5, 5, and 10m buffers of the lane polygon. Percentage, mean and median error in meters were calculated for each trip, mode, and for each of the four environmental types.

In Vorlíček et al. (2019), the GPS device trackers are placed inside the top pocket of a backpack. A 2-km route which included various types of environment was selected. A total of 30 trips (10 walking trips (5 km/h), 10 running trips (10 km/h) and 10 cycling trips (17 km/h)) were done following the centreline of the sidewalk or bike lane on one side of the street. The experiment was performed under different atmospheric conditions and at different times of the day. Similar as previous article, percentage of GPS points within the 2.5-, 5-, and 10-m buffers, is calculated, together with differences between proportions of points in each of the buffers, using the "N−1" Chi-squared test. Mann-Whitney U test is used to calculate the differences between distances of GPS points from the edge of the base lane polygon. IBM SPSS (Version 22; IBM, Armonk, NY, USA) with alpha ≤ .05 is used for all statistical analysis.

In Adamakis (2020), two GPS device trackers of each type were placed on a pad at each six different geodetic point for 60 minutes, recording location updates every second. The difference between GPS device results and the coordinates of each geodetic site was calculated using the Haversine equation. The geographic information system Esri ArcGIS for Desktop 10.6.1 was used for the visual interpretation of the data. Circular error probability (CEP) was calculated to determine the horizontal positioning accuracy, as well as mean, standard deviation, and median of accuracy error.

In Duncan et al. (2013), the GPS devices were placed in a flat opaque plastic tray (40 cm X 30 cm X 6.25 cm) centred directly over six geodetic sites of different environmental conditions. The devices were kept 0.5m above ground level by placing the tray on a small box. After being placed in the tray, the devices were turned on, and set to collect data for 50 minutes. The same test was repeated twice within 9 days. The difference between GPS device results and the coordinates of each geodetic site was calculated using the Haversine formula. CEP was also measured.

These information on choice of GPS device tracker, number and choice of locations, as well as data collection and processing are summarised in Table 2.5.

Table 2.5 Summary of Experimental Setup from Literature Review

Article	Positioning of GPS Device Tracker	Number and Location Choice	Data Collection and Processing
(Schipperijn et al., 2014)	Worn around the waist and concealed by clothing	▪ Four routes in Copenhagen, Denmark, predetermined protocol of walking, cycling, driving, and bussing on all routes, in both directions ▪ The routes were on different bearings, passing through a variation of environmental conditions ▪ Four modes of transportation are employed in each four designated routes to record data on 300 trips.	▪ The geographical coordinates are then collected using open source bt 747 GPS software ▪ Spatial join function in ArcGIS is then used to group coordinates for each trip according to 2.5, 5, and 10m buffers of the lane polygon. ▪ Percentage, mean and median error in meters were calculated for each trip, mode, and for each of the four environmental types.
(Vorlíček et al., 2019)	Placed inside the top pocket of a backpack	▪ 2-km route with various types of environments and at different times of the day ▪ A total of 30 trips (10 walking trips (5 km/h), 10 running trips (10 km/h) and 10 cycling trips (17 km/h)) following the centreline of the sidewalk or bike lane on one side of the street	▪ Percentage of GPS points within the 2.5-, 5-, and 10-m buffers, is calculated, together with differences between proportions of points in each of the buffers, using the "N–1" Chi-squared test ▪ Mann-Whitney U test is used to calculate the differences between distances of GPS points from the edge of the base lane polygon ▪ IBM SPSS (Version 22; IBM, Armonk, NY, USA) with alpha ≤ .05 is used for all statistical analysis
(Adamakis, 2020)	Two trackers of each type were placed on a pad	Six different geodetic point for 60 minutes, recording location updates every second	▪ Difference between GPS device results and coordinates of each geodetic site was calculated using the Haversine equation. ▪ Geographic information system Esri ArcGIS for Desktop 10.6.1 was used for the visual interpretation of the data ▪ Circular error probability (CEP) was calculated to determine the horizontal positioning accuracy, as well as mean, standard deviation, and median of accuracy error
(Duncan et al., 2013)	Placed in a flat opaque plastic tray (40 cmX30cmX6.25 cm), 0.5m above ground level using a small box	▪ Six geodetic sites of different environmental conditions for 50 minutes ▪ The same test was repeated twice within 9 days	▪ Difference between GPS device results and the coordinates of each geodetic site was calculated using the Haversine formula ▪ CEP was also measured

CHAPTER III

METHODOLOGY

3.1 MATERIALS INVOLVED IN THE RESEARCH

Literature review conducted in previous chapter has facilitated the planning of this experimental setup. As the requirement for this project is applied research, most of the elements from previous works are adopted and adapted to this research domain which is law enforcement to help carry out the experiment for the purpose of answering the research questions in Section 1.2. Some of the elements are chosen due to the fact that they are more suitable and relevant to be implemented in this current context. These elements will be detailed out in the following subsections.

3.1.1 Tools

TKStar GPS tracker model TK905 is chosen as the GPS device tracker for this experiment. Size of this model is 90mm x 72mm x 22mm, weighing around 165 grams. The colour of this device is black, and its GPS accuracy is claimed to be 5 metres. Its power input includes rechargeable 3.7V 5000mAh lithium battery. It takes 8 to 10 hours to fully charge the device and its tracking can last up to 90 days (TKStar n.d.). A picture of this device is shown in Figure 3.1. Lenovo laptop model IdeaPad 3 15ITL6 is used to get Google geolocation API coordinates via Microsoft Edge browser, and it is connected to mobile data. Microsoft Edge browser is chosen to see whether the accuracy claimed by Google is applicable to Microsoft's browser.

Figure 3.1 TKStar GPS Tracker

3.1.2 API Script

Google geolocation API is a service by Google to provide geographical coordinates such as latitude and longitude of web clients by means of HTTPS request via WiFi access points and cell tower (Google Developers 2023). It has a built-in function *getCurrentPosition()* to obtain the current position of web clients of interest (MDN 2023). The code or script for getting the current geographical coordinates of web clients are shown in Figure 3.2. It is adapted from the web application which utilises Google geolocation API in accurately determining postal address (Kwabla et al. 2018).

```
<html>
<body>
	<button onclick="getLocation()">Click me to get location</button>
	<div id="demo">

	</div>
	<script>
		var x = document.getElementById("demo");
		function getLocation() {
		if (navigator.geolocation) {
			navigator.geolocation.getCurrentPosition(showPosition);
		} else {
		x.innerHTML = "Geolocation is not supported by this browser.";
		}
		}

		function showPosition(position) {
		x.innerHTML = "Latitude: " + position.coords.latitude +
		"<br>Longitude: " + position.coords.longitude;
		}
	</script>
</body>
</html>
```

Figure 3.2 Code for Geolocation API

3.2 CHOICES OF EXPERIMENTAL SETUP

3.2.1 Study Area

As mentioned in Section 2.6, there are a number of factors to be controlled in order to determine the accuracy of Google geolocation API in relation to GPS device trackers. As these public infrastructures of WiFi access points and cell towers are capable of influencing the accuracy of the results obtained, one way of maintaining control of these factors is by choosing one particular location to carry out the experiment. This location should have indoor and outdoor conditions in nearby vicinity so that the altitude difference of web clients is minimised. The location of Kuala Lumpur City Centre (KLCC) is chosen because it is made up of multiple buildings as well as outdoor parks within the same area. It is known that different environment conditions will influence the results of geolocation as described in Section 2.6, therefore this piece of KLCC location area confers a variety of selections to be picked as potential indoor and outdoor conditions. Map of KLCC can be seen in Figure 3.3 (“About Us - KLCC The Place” 2023).

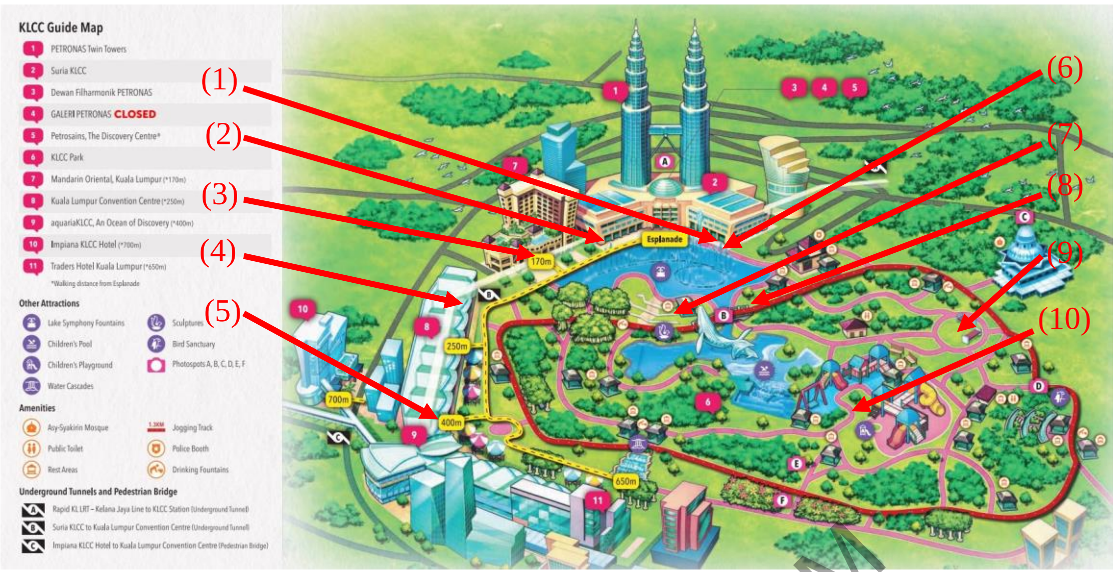

Figure 3.3 Map of KLCC (“About Us - KLCC The Place” 2023)

10 different locations which comprise of 5 indoor and outdoor conditions respectively are selected, as shown in Table 3.1 and Figure 3.3. These locations are selected based on the surroundings, suitability and relevancy to this domain. Since there is no research being conducted on geolocation accuracy in terms of indoor condition, the choices of locations are selected based on justifiable places where a person might be. All the locations are in ground floor to ensure the consistency of altitude which might affect the geolocation accuracy.

Table 3.1 Indoor and Outdoor Locations

Indoor	Outdoor
(1) Ground Floor of Suria KLCC (East)	(6) Outside of Suria KLCC
(2) Ground Floor of Suria KLCC (West)	(7) Rest Area in front of Sculptures
(3) Ground Floor of Mandarin Oriental	(8) Bridge in the KLCC Park
(4) Ground Floor of Kuala Lumpur Convention Centre (North)	(9) Open Area in front of Asy-Shakirin Mosque
(5) Ground Floor of Kuala Lumpur Convention Centre (South)	(10) Children’s Playground

As GPS signals will also be affected by high-rise buildings, several outdoor locations are selected encompassing adjacent to mall (6), shaded area (7), on the bridge (8), unobstructed area (9), and playground equipment (10). Pictures of both indoor and outdoor locations are shown in Figure 3.4 to Figure 3.13.

(1) The geolocation is taken in front of Capitano Caffe.

Figure 3.4 Ground Floor of Suria KLCC (East)

(2) The geolocation is taken in front of Sakana Japanese Dining.

Figure 3.5 Ground Floor of Suria KLCC (West)

(3) The geolocation is taken in front of a staircase.

Figure 3.6 Ground Floor of Mandarin Oriental

(4) The geolocation is taken beside an electronic bulletin board.

Figure 3.7 Ground Floor of Kuala Lumpur Convention Centre (North)

(5) The geolocation is taken in front of ticketing counter.

Figure 3.8 Ground Floor of Kuala Lumpur Convention Centre (South)

(6) The geolocation is taken in front of fountain.

Figure 3.9 Outside of Suria KLCC

(7) The geolocation is taken in front of sculptures.

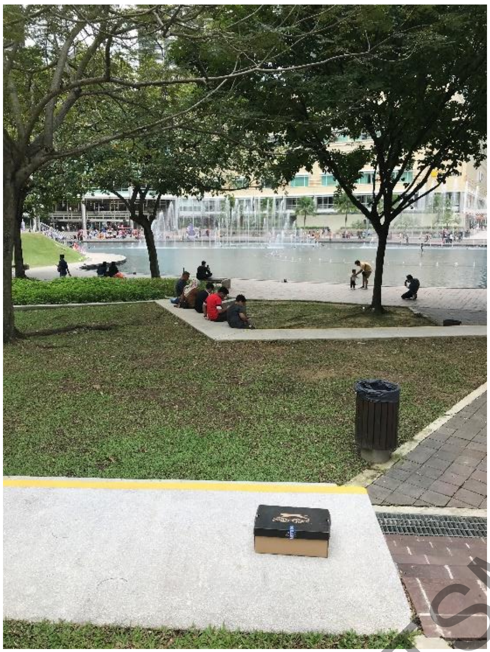

Figure 3.10 Rest Area in front of Sculptures

(8) The geolocation is taken on top of a bridge.

Figure 3.11 Bridge in the KLCC Park

(9) The geolocation is taken in front of Asy-Shakirin Mosque.

Figure 3.12 Open Area in front of Asy-Shakirin Mosque

(10) The geolocation is taken in the vicinity of children's playground.

Figure 3.13 Children's Playground

3.2.2 Study Duration

All GPS and Google geolocation API coordinates for all 10 different locations are taken on the same day for two (2) days, which were 2nd June 2023 and 9th June 2023 from 3pm to 6pm. Two sets of data are collected, whereby the geographical coordinates of each location are taken every minute for five (5) minutes for each GPS device tracker and Google geolocation API technology. Weather on both days was constant throughout the experiment.

A third set of data has to be taken on the third day, which was on 24th June 2023 as the second day's results seem to be unreliable due to identical geographical coordinates obtained from Google geolocation API across multiple locations. Careful monitoring and implementation of Google geolocation API on the third day is done to avoid the same error from happening again. Browsing history, cache and cookies are cleared every time when moving from one location to another.

3.2.3 Sampling Design

Step 1: Install SIM card into TKStar GPS device tracker. When indicator light is flashing blue light, it means that GPS connection is established. Turn on Lenovo laptop and connect to mobile data using hotspot.

Step 2: Go to designated locations as stated in Section 3.2.1 above. Place the TKStar GPS device tracker on top of a shoe box (27.8 cm x 18 cm x 10.4 cm).

Step 3: Send a message of "G123456#" (G<password># - G is the code to get location information followed by password and # to indicate end of message) to the SIM card number inserted inside the TKStar GPS device tracker. You will receive a message stating the latitude and longitude of the TKStar GPS device tracker. The conceptual diagram is shown in Figure 3.14. Note down this information.

Step 4: Repeat Step 2 to 3 every minute for 5 minutes.

Step 5: Place the Lenovo laptop on top of the same shoe box as above. Run the script as stated in Section 3.1.2 and you will receive the latitude and longitude of the Lenovo laptop. The conceptual diagram is shown in Figure 3.15. Note down this information.

Step 6: Repeat Step 5 every minute for 5 minutes.

Step 7: Repeat Step 1 to 6 another time within nine days. The nine days duration is based on experimental design of Duncan et al. (2013).The difference between GPS device results and that of Google geolocation API was calculated using the Haversine formula.

Figure 3.14 Conceptual Diagram of Experimental Setup for GPS Device Tracker

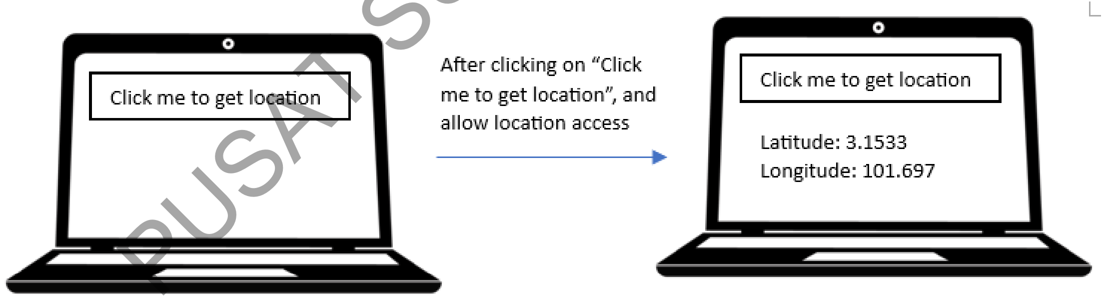

Figure 3.15 Conceptual Diagram of Experimental Setup for Google Geolocation API

Since this project is measuring how close the geographical coordinates generated from Google geolocation API in relation to that of GPS device tracker, the Haversine formula can be used to determine the distance between the two coordinates obtained. Its formula is shown as follow (Dauni et al. 2019):

$$\Delta long = (\text{long2} + \text{long1})\ \cos\left(\frac{lat1 + lat2}{2}\right) \quad \ldots(3.1)$$

$$= \cos a \cos b - \ \sin a\ \sin b$$

$$\Delta lat = (lat2 - lat1) \quad \ldots(3.2)$$

$$a = \sin 2\ \left(\frac{\Delta lat}{2}\right) + \cos(lat1)\cos(lat2)\sin 2\left(\frac{\Delta long}{2}\right) \quad \ldots(3.3)$$

$$d =\ \sqrt{(a)}R \quad \ldots(3.4)$$

Equations …(3.1) until …(3.4) are used to calculate the distance in kilometres (km) between two sets of geographical coordinates. Some of the definitions are as follow: R = the radius of the earth is 6371 (km) (1 degree = 0.0174532925 radians); Δlat = amount of change in latitude (km); Δlong = magnitude of change in longitude (km). Haversine formula is applied with spherical earth shape and disregarding that the earth is slightly elliptical (Dauni et al. 2019).

There are several open-source calculators for Haversine formula such as KurtHeckman (2023) and "Distance Calculator" (n.d.). The results obtained from both calculators are almost similar, therefore justifying the accuracy of these calculators in computing the distance between the geographical coordinates from GPS device tracker and Google geolocation API. Figure 3.16 shows the website that is being used to calculate the distance difference between two geographical coordinates using Haversine formula.

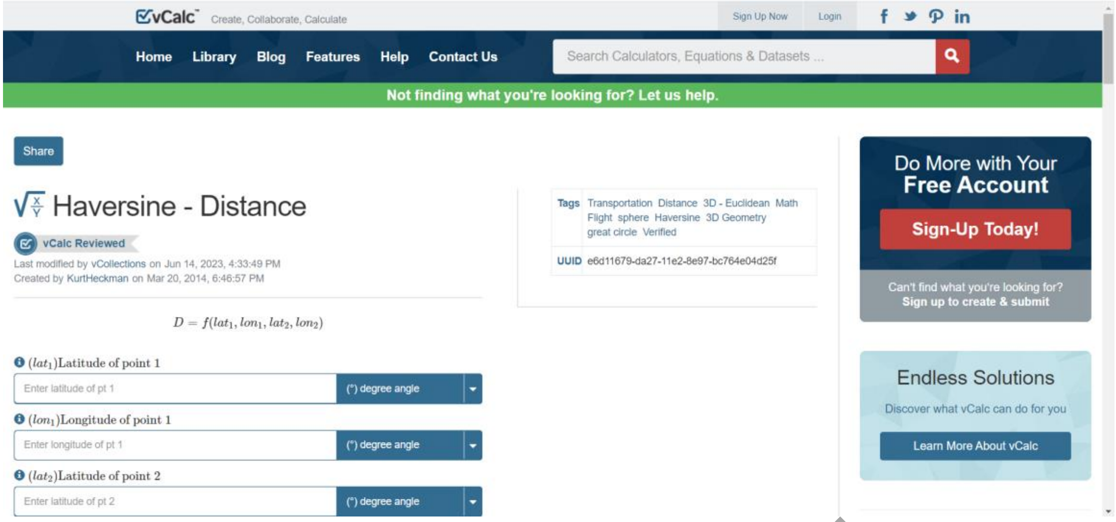

Figure 3.16 Webpage of calculator for Haversine formula (https://www.vcalc.com/wiki/vCalc/Haversine+-+Distance)

3.3 STATISTICAL TESTS

After conducting the experiment, the result of maximum performance deviation between Google geolocation API and GPS device trackers in terms of geographical coordinates of web clients under indoor and outdoor conditions are obtained. The maximum search radius from the geographical coordinates of Google geolocation API is determined and can be used as a reference for LEA officers to extend their pursuit. It can be seen as an inherent localisation error of Google geolocation API. In order to ensure the reliability and consistency of this sample results, statistical tests are then carried out so that the sample results are able to be inferred to the larger population.

Firstly, Shapiro-Wilk test is carried out to determine the normality of the sample results obtained. It is used when the sample size is less than 50 (Merry & Bettinger 2019). Each day's results are tested individually and were found to be not normally distributed. Mann-Whitney U test which is a non-parametric test can be used when assumptions of normalisation is violated, and it is appropriate for small sample size experiment. This test is conducted to compare differences between two independent groups of data. It is an alternative to the independent samples t-test (LaMorte 2017). This statistical analysis is to test the following hypotheses:

- H_0: The data obtained from both days are not significantly different.

- H_1: The data obtained from both days are significantly different.

At 95% confidence level, if the significance result of the Mann-Whitney U test is more than 0.05, null hypothesis cannot be rejected, therefore it is accepted that the data obtained from both days are almost similar. This is the intended result because it is to ensure that the results are consistent in both days. The maximum search radius which is represented by the longest distance obtained between geographical coordinates of GPS device tracker and Google geolocation API is considered constant in any situation.

Descriptive statistics involving mean, standard deviation, minimum and maximum are also calculated using IBM SPSS Statistics 28.0. Results for each day have their respective values. Descriptive statistics also involves figures to illustrate the patterns generated from these results.

CHAPTER IV

RESULTS AND DISCUSSION

4.1 INTRODUCTION

The experiment is conducted for two days to collect geographical coordinates using both GPS device tracker and Google geolocation API. Using the recorded geographical coordinates, the distances between each particular location are calculated to determine the performance deviation of Google geolocation API from GPS device tracker using Haversine formula. The smaller the distance difference, the closer the geographical coordinates obtained between GPS device tracker and Google geolocation API, thus lower performance deviation observed. The larger the distance difference, the further away the geographical coordinates obtained between GPS device tracker and Google geolocation API, leading to huge performance deviation. In this experiment, it is to determine the maximum performance deviation that can happen at any situation, so that the LEA officers can estimate how far away they should extend their search radius from the obtained geographical coordinates of Google geolocation API.

Due to unforeseen circumstances, another set of results has to be collected on the third day, which was on 24th June 2023 because there were some readings from the second day that were identical in different locations. These same readings across different locations suggests that there is some error occurred during the experiment. Thus, the experiment on the second day has to be repeated in order to collect a more reliable and plausible geographical coordinates. This occurrence will be discussed further in Section 4.3 below.

4.2 DATA ANALYSIS

4.2.1 Data Preparation

The latitude and longitude for the selected ten locations are obtained using both GPS device tracker and Google Geolocation API as shown in Table 4.1 for the first day; Table 4.2 for the second day; Table 4.3 for the third day. The distance differences between each geographical coordinate obtained from GPS device tracker and Google Geolocation API are also calculated and tabulated. Mean and standard deviation are also reported for each location, whereby mean shows the average distance, while standard deviation shows how consistent the distance measurement for each location.

Figure 4.1 depicts an example of the interface to get geographical coordinates through Google Geolocation API. Once we click on “Click me to get location”, the latitude and longitude of that particular location will be shown on the screen. Figure 4.2 portrays an example of SMS messages sent to (green) and received from (white) the SIM card of GPS device tracker. As soon as we send an SMS message of “G123456#” to the phone number stated on the SIM card in the GPS device tracker, we will receive a reply message stating the current latitude, longitude, speed of the device, date and time, battery level of the device, as well as the device’s ID number after a few seconds. The Google Map link provided will pinpoint the exact location based on the given geographical coordinates. The GPS device tracker is utilising GPS satellite signals to give us real-time geographical coordinates information.

Table 4.1 Geographical Coordinates Obtained Using Both GPS Device Tracker and Google Geolocation API on Day 1 as well as the Distance Difference

Location	GPS Device Tracker		Google Geolocation API		Distance (m)	Mean, Standard Deviation (m)
	Latitude	Longitude	Latitude	Longitude		
1	3.15716	101.71333	3.1533	101.697	1860	1844, 8.94
	3.15716	101.71316	3.1533	101.697	1840	
	3.15716	101.71316	3.1533	101.697	1840	
	3.15716	101.71316	3.1533	101.697	1840	
	3.15716	101.71316	3.1533	101.697	1840	

to be continued …