

A Collaborative Bluetooth-Based Approach to Localization of Mobile Devices

Nicholas Mair and Qusay H. Mahmoud

School of Computer Science
University of Guelph
Guelph, ON, N1G 2W1 Canada
{nmair01,qmahmoud}@uoguelph.ca

Abstract— In this paper we present a collaborative Bluetooth localization method, which aggregates the location information about Bluetooth devices that is provided by multiple mobile devices. The method aims to take advantage of an enhanced version of the KNN algorithm in which the location of a mobile device can be determined based on a Bluetooth distance measure. The experimental results showed that there are a sufficient number of Bluetooth devices that are discoverable, which can be used collaboratively to assist with localization requests from mobile devices. In addition, it is shown that the Bluetooth localization method was able to successfully localize the mobile device using the Bluetooth radio. Furthermore, the response time of Bluetooth localization is shown to be less than GPS, and greater than Wi-Fi and Cellular localization methods.

Keywords: *Bluetooth, mobile device localization, crowdsourcing, location based services, location prediction*

I. INTRODUCTION

Location based services are becoming more and more popular each day. With the widespread adaptation of smartphones and the increased use of third party applications on smartphones, more and more applications are taking advantage of the information that is accessible from the sensors on the mobile devices. Of all the sensors on a mobile device, the most commonly used in third party applications is the GPS receiver. This is due to the fact that it can provide contextual information about the location of a mobile device, and using this information, the application can then perform actions based on the current location of a device. Due to the ease of incorporating this type of data into applications, it is becoming more common to see this type of information present in mobile applications.

This popularity however comes at a cost. The GPS receiver on a mobile device is one of the most power inefficient sensors. If used continuously, it has a severe effect on battery life whereby a fully charged battery can be drained in a mere 8.5 hours when a mobile device is being used regularly [12]. In a real life scenario, a mobile device is used for a lot more than GPS invocations, and as such, the battery resources should be partitioned for a number of different purposes. The battery should be budgeted in a way that all tasks that the device was designed for can be done while still allowing the device battery to last for a full day. Common tasks include making phone calls, sending and receiving SMS messages, emails, pictures, videos, etc.

Therefore, in a real life scenario, the energy budget for localization should be around 25% of the entire battery life, thus continuous GPS localization is only available for about 2.5 hours per fully charged battery [5].

In order to extend the battery life of mobile devices which continuously request localization information, duty cycling techniques using the GPS receiver have been proposed. This entails invoking the GPS receiver periodically at predetermined intervals in order preserve battery life. This method however decreases the accuracy as the location of the mobile device is only known at fixed points in time and unknown between subsequent GPS invocations. The use of more energy efficient sensors has also been used as a way of obtaining localization information at a fraction of the energy cost of GPS. The two most common uses the cellular radio and the Wi-Fi radio as a means of localization. While these methods do use a significantly lower amount of energy, their localization error is far greater than that of GPS which is around 10 meters. Another drawback is due to the fact that these other energy saving methods may not be usable everywhere. It would be hard to get a cellular location fix in a basement or some indoor locations. Similarly, at a crowded event outdoors where Wi-Fi is not available, you would have to rely on the inaccurate cellular location fix or the more accurate, but also more power hungry GPS receiver to obtain a location fix.

As a solution to some of these problems, mobile device localization using Bluetooth is explored in this paper as another energy efficient localization method. Due to the lower energy usage of the Bluetooth radio compared to the GPS receiver on mobile devices, it has the potential to provide localization information about a device at a fraction of the energy cost compared to GPS. To this end, the contributions of this work are:

- Analysis of energy usage on BlackBerry smartphones, and benchmarks for others to use.
- A technique for predicting the location of mobile devices using Bluetooth scan data.
- A data mining-based method for mobile device localization through recognizing patterns in mobile device data.

The rest of this paper is organized as follows. Section II discusses the related work previously completed. The proposed method is presented in Section III. Section IV

discusses the experiments and presents the evaluation results. Section V describes a case study that was performed using the proposed method. Some privacy concerns of the proposed method are addressed in Section VI. Finally, conclusions and future work are presented in Section VII.

II. RELATED WORK

There has been some interesting work already done in the area of energy efficient mobile device localization. Most of this work focuses on different ways of duty cycling the GPS in order to save energy, but as mentioned before this reduces the accuracy as there is a time period where the location is unknown. Enloc [5] is a method which was proposed to estimate the location of a mobile device between duty cycles. It does this based on the past information about a mobile device, and human mobility patterns as they were found to be predictable [15]. The drawback of this method however is that there is increased localization error as the direction and velocity of the mobile device can change at any time between the duty cycles.

Adaptive GPS duty cycling has also been explored, and is a method whereby the GPS is triggered at either a fixed interval [16] [24] or is adapted dynamically based on the users current activity [9] [6]. There are drawbacks to both methods. The drawback of the fixed interval method of adaptive GPS duty cycling is that the method doesn't consider if the user has actually moved or if the location of the mobile device is easily predictable. The problem with current methods of adaptive GPS duty cycling is the method of determining when to sense. There have been methods proposed which use other sensors such as the accelerometer [10] or the microphone to determine when to sense the location. In essence, when the device is stationary with a small amount of movement, location sensing should be suppressed, or if the sound pattern matches that of an outdoor environment, the location should be sensed. The drawback of using the microphone for suppression is that it is also a very power hungry sensor.

Due to the fact that smartphones may run multiple application each which request localization information, research has been done to synchronize these requests in order to save energy [25]. Their method allowed existing application to put in requests for localization information, and these requests were synchronized on the OS level of an Android smartphone. For example, if an application requires localization information every 2 minutes, and another application is started that requires localization information, the request will be delayed until the 2 minute period for the other application is reached. At that point, the localization information is sent to both applications with a single invocation.

An indoor localization method using wireless access points was proposed in [4] due to the fact that cellular signal may not always penetrate through building walls and there may be no line of sight with GPS satellites. Their solution used existing wireless access points to determine the location of a mobile device by measuring the signal strength of the device from each access point, which was used to determine the devices distances from each access point. The drawback

of this method is interference and environmental factors can affect the signal quality, and thus degrade the accuracy of the prediction. To combat interference, two access points were positioned at locations where the numbers of walls between them were equal. Therefore, both access points would incur the same amount of interference and the difference between the shading and fading effect would be minimized.

III. PROPOSED METHOD

There have been many research projects which attempt use additional sensors on smartphones in order to either suppress the GPS or use the information obtained from these sensors in order to predict the location of the device. In addition to the projects discussed, there are other implementations not mentioned which focus on using the accelerometer, microphone or compass to predict location information; however no work has been done on how Bluetooth can be used for GPS suppression and prediction. In this section, we will propose a method of localization using the Bluetooth radio on a mobile device.

Driving safety has always been one of the top priorities in many cities and countries. As a result, more and more hands free driving laws are being passed throughout the world which prevents drivers from using mobile devices while they are driving. The answer to these new rules is using Bluetooth to make and receive phone calls. As a result, more and more mobile devices are being operated with Bluetooth turned on and discoverable. Almost anywhere you go, if there are many people around, mobile devices will be discoverable via Bluetooth. This presents a wide range of new information which can be used to the advantage of location based services.

In terms of power usage, Bluetooth uses much less power than that of Wi-Fi and GPS [20]. On average, the power usage of a Bluetooth Scan is around 150mW, while the power usage during a Wi-Fi scan is close to 1000mW. Because of the length of time a Bluetooth scan takes, the complexity of the protocol, it has the potential to use more energy than Wi-Fi. The energy can be kept below that of Wi-Fi by limiting the number of devices found on each Bluetooth scan simply because the longer the scan takes, the more energy it uses.

Bluetooth has a short range, which is around 10 meters. It can be assumed that a Bluetooth scan finds a device, you are within 10 meters of it. This localization range is smaller than that of Wi-Fi (60-300 meters) and cellular positioning (2000+ meters). This short vicinity coupled with the reduced power requirements makes Bluetooth an excellent candidate for energy efficient localization.

To this end, the main goal of our approach is to allow mobile devices to collaborate in discovering and providing the location of Bluetooth devices, so that this data can be used to localize other mobile devices. The proposed method focuses on taking advantage of the patterns which exist in humans [15] as it was found in our research that most discoverable Bluetooth devices are either smartphones or laptops, and will follow the same movement pattern of the individuals which own them. Therefore, a Bluetooth scan should discover the same group of devices in the same

location if the scan is performed at the same time each day. There may be special cases when the pattern of an individual may deviate from their normal behavior, such as when attending a special event. In terms of the overall pattern of an individual, this would account for a small fraction of their overall movement pattern; therefore, we still focus on using the patterns of an individual, and the Bluetooth devices surrounding them to predict their location. Using these patterns, the goal is to create a footprint about the locations of the found devices so that a scan footprint can be created based on past information about the locations of the devices.

The following steps will describe our proposed method in more detail:

- 1) Collect data about discoverable Bluetooth devices.
- 2) Preprocess and store the data in the cloud.
- 3) Use the enhanced KNN classifier to determine the location (class) of a mobile device using the collected information.
- 4) If the location cannot be determined, obtain mobile device location using traditional localization methods, and store the new information about the location of discoverable Bluetooth devices in the cloud for future use.

Figure 1 illustrates the process flow overview of our proposed method.

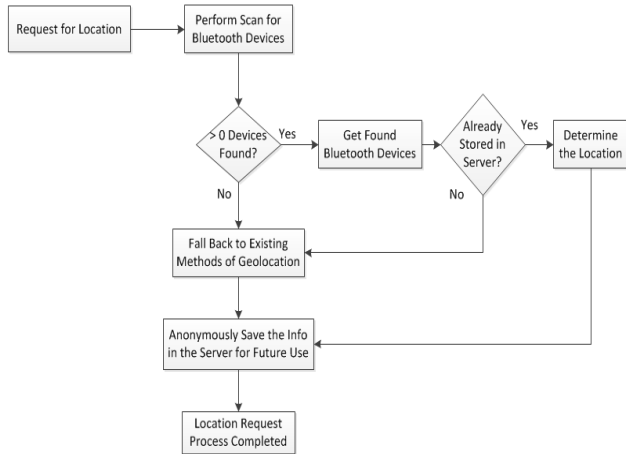


Figure 1. Bluetooth Localization Process Flow Overview

A. Bluetooth Location Services using Scan Footprint

People have patterns, students go to school every weekday, adults go to and from work every weekday, some might go to the gym in the mornings, etc. The motion of individual is not random; there is a pattern to the movement of each individual [15]. As a result, a Bluetooth scan result should find the same group of devices in the same location each time the scan is performed there. A scan footprint could be created based on past information, and past locations that those devices are found. This information could be stored on the local device or the cloud. The advantage of storing this information on the cloud is so that location information about the pattern of devices can be shared across the

localization requests of multiple devices. Some sort of data mining needs to be done to obtain patterns in localization requests based on the location of the requesting device, and the Bluetooth devices found around that device. Once a footprint is created, GPS can be suppressed as the location of a device can be predicted based on the group of Bluetooth devices that are found.

Raw data was obtained from various mobile devices running an application which would perform a Bluetooth scan of the devices surrounding their location every 30 minutes when the device was within the confines of the University of Guelph. In addition to the found Bluetooth devices, the current location of the device that performed the scan was obtained as well. The goal was to create a dataset which could be fed into a classifier and determine how well the classification algorithm is able to predict the location of the devices just by using the Bluetooth scan data.

B. Data Preprocessing

The data that is used in this research is collected within the University of Guelph and contains 1,168 samples which were collected over the course of a few weeks, a snippet of which is presented in Table I. The collected data is real-world, and as a result, requires preprocessing. The main purpose of preprocessing, here, is to change the data into a format that can be used with the Enhanced KNN classifier. For this purpose, in the first column of Table I, the time is extracted from the date information in order to be used as a numerical attribute. Also, the hardware address of the found device in column 4 of the table is transformed into a nominal value for ease of use. In addition, the GPS coordinates in columns 5 and 6, are transformed into a single value representing the location of the device, which is referred to as the location class. The format of the transformed data is shown in Table II.

TABLE I. FORMAT OF RAW COLLECTED DATA

| Date/Time | Accuracy (m) | Found Devices | GPS Latitude | GPS Longitude |
|---|--------------|--|--------------|---------------|
| Tue Oct 25 11:35:56 America/New_York 2011 | 60.0 | 0125005E0314 ::Computer ::Laptop | 43.5313 | -80.2272 |

TABLE II. FORMAT OF THE TRANSFORMED DATA

| Time | Accuracy (m) | Found Device IDs | GPS Location Class |
|------|--------------|------------------|--------------------|
| 11 | 60.0 | 5 | 21 |

After cleaning the data and transforming it into a format suitable for the application of classification algorithms, it was further preprocessed by removing any noisy data. Instances where the recorded location had a location fix accuracy above 120 meters were not included in the training and testing of classification algorithms. This is due to the associated errors in the recorded location classes. For example, if during a Bluetooth scan only cellular geolocation is available on the mobile device, the accuracy of the location fix will be around 3000 meters. This presents a very

high probability that the location class recorded for that scan will be incorrect. These types of data generate noise, and as a result were omitted.

C. GPS Location Classes

In order to create location classes for a given area, a grid overlay can be used. The grid overlay will also help to reduce errors which may sometimes occur if a device has not moved, but consecutive GPS invocations are off by a few meters. Consider the map of the University of Guelph from Google maps which is shown in figure 2. If the map is split into equally spaced partitions with a grid overlay, the map will look similar to figure 3. The partitions can be adjusted to make the area of each grid smaller, and thus increase the preciseness of the location classes. While a Bluetooth scan is occurring, if devices are found anywhere within the bounds of location class 1, the location of the found devices will be recorded as location class 1. When the location is determined based on the found Bluetooth devices, the result will be a location class number.

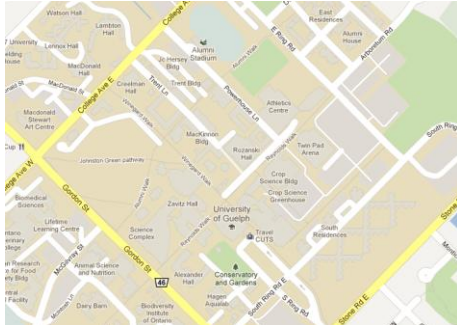


Figure 2. Map of the University of Guelph (UoG)

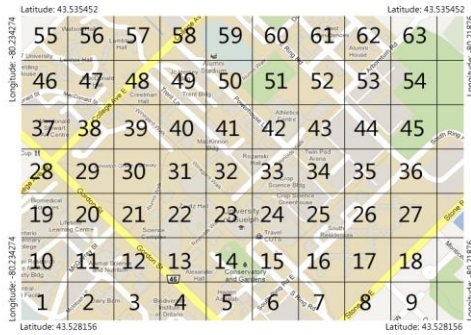


Figure 3. Map of the UoG with an Overlay of GPS Location Classes

D. Conversion of GPS Coordinates to Location Classes

The location class is obtained using the following steps. First, the GPS coordinates of the boundaries that is going to be used is determined. We defined the boundaries of the University of Guelph as 43.519801,-80.240207 for the starting and 43.545251, -80.218835 for the ending GPS coordinates. Also, we define an increment value of 0.001, which is used to determine the distance of separation between each location class (depending on the application, the increment value can be different). Next, the difference

between the starting and ending points of the GPS coordinates of the boundaries is calculated, which is referred to as $\Delta latitude_{map}$ and $\Delta longitude_{map}$. Afterwards, the number of partitions for the GPS coordinates is computed using equations 1 and 2. Then, the difference between the GPS coordinates of a sample and the starting points of the boundaries is computed, which is referred to as $\Delta latitude_i$ and $\Delta longitude_i$. These values are divided by the increment value, the result of which is referred to as $xValue$ and $yValue$. These values represent the latitude row and longitude column respectfully in the location map. Finally, the floors of these values are used in equation 3, to acquire the location class value of the sample.

$$\text{latitude partitions} = \frac{\Delta latitude_{map}}{\text{increment value}} \quad (1)$$

$$\text{longitude partitions} = \frac{\Delta longitude_{map}}{\text{increment value}} \quad (2)$$

$$\text{location class} = (\text{yValue} * \text{longitude partitions}) + xValue \quad (3)$$

The increment value is used to determine the distance of separation between each location class. This is achieved by adding the defined increment value to the GPS coordinates as each GPS class is constructed. This distance is determined using the Haversine formula [23] (4), also referred to as the great-circle distance, which is defined as:

$$a = \sin^2\left(\frac{\Delta lat}{2}\right) + \cos(latitude_1) * \cos(latitude_2) * \sin^2\left(\frac{\Delta long}{2}\right) \quad (4)$$

$$c = 4 * \tan^{-1}\left(\frac{\sqrt{a}}{1 + \sqrt{(1-a)}}\right)$$

$$d = R * c$$

Within the formula, R represents the radius of the earth which is 6,371 km. Δlat represents the difference between two latitude points and $\Delta long$ represents the distance between two longitude points. The result of the Haversine formula for the distance between two points can be used to determine how many decimal points to use in GPS location class construction. Tables III and IV show the distance between each latitude and longitude point depending on the number of decimal places used. In our experiment, we constructed our GPS location classes with 3 decimal places which have a distance of 112.2 meters between the latitude points and 81.32 meters between the longitude points. Therefore, the total distance inside each location class is $9042m^2$. This distance is a desirable as it provides a better accuracy level than cellular geolocation.

TABLE III. DECIMAL PLACE DISATNCE FOR LATITUDE

| Decimal Places | Distance |
|----------------|------------------|
| 0 | 111.2 kilometers |
| 1 | 11.12 kilometers |
| 2 | 1.112 kilometers |
| 3 | 111.2 meters |
| 4 | 11.2 meters |

TABLE IV. DECIMAL PLACE DISATANCE FOR LONGITUDE

| Decimal Places | Distance |
|----------------|-------------------|
| 0 | 81.32 kilometers |
| 1 | 8.132 kilometers |
| 2 | 0.8132 kilometers |
| 3 | 81.32 meters |
| 4 | 8.132 meters |

A visual example using an increment value of 0.01 (or 2 decimal places of separation) is shown in figure 4 for simplification. It should be noted that in order for each data point to be assigned to a specific location class, it must meet the boundary requirements for that class as the coordinate values for location classes do not overlap. For example, in order for a sample to be assigned to location class 1, the latitude must be greater than or equal to 43.54 but less than 43.55 and the longitude coordinate value must be greater than or equal to -80.12 and less than -80.11.

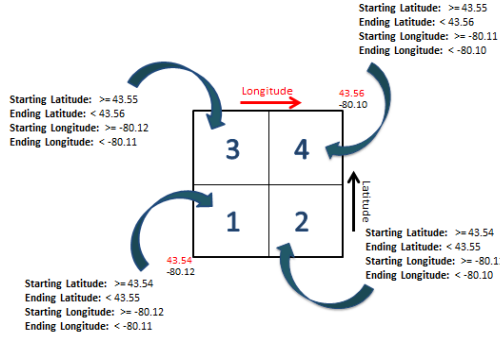


Figure 4. Sample Construction of Location Classes using 2 Decimal Places

E. KNN and the Enhanced KNN Classifier

The KNN classification algorithm [11] is a lazy learner or instance based algorithm which is used to assist with localizing a mobile device based on the Bluetooth devices which surround it. The algorithm is used extensively within the area of pattern recognition and works by comparing a new sample against all previously stored samples, choosing the closest K samples based on a distance measure. The closest K samples are then used to determine the class of the new sample, either by performing voting, or averaging the class values of the K nearest samples. Averaging is performed when the class attribute is defined as a numerical value, and voting occurs when the class attribute is defined as a nominal value.

We employ an enhanced version of the KNN algorithm for a few different reasons. The first reason has to do with how the algorithm calculates the distance of the attributes between samples. The original KNN algorithm considers the order of the attributes. However since we are dealing with Bluetooth device scan data, and in each subsequent Bluetooth scan, the devices may not be in the same order, considering the order of the device attributes is not appropriate. Therefore, the enhanced version of the KNN algorithm does not consider the order of the Bluetooth device

attributes when calculating the distance. The second reason has to do with how the predictions being made about a single, or group of device(s). It may not be possible to determine the location of a mobile device based on one sample about a surrounding Bluetooth device. Therefore, the enhanced KNN algorithm incorporates a nearest neighbor threshold to ensure that there are enough samples about the location of a Bluetooth device before using it to determine the location of a mobile device. In the research, a nearest neighbor threshold of 3 is used, which means that at least 3 samples about the location of a Bluetooth device is needed before it is used to determine the location of a mobile device. The third reason is when the algorithms determine the location of a mobile device. The traditional KNN algorithm will return a location class value whether or not there is information stored about the previous locations of Bluetooth devices found in a scan. The difference with the enhanced KNN algorithm is it will only make a location determination if there is previous data about device(s) found in a Bluetooth scan, and the number of samples currently stored about that device is above the nearest neighbor threshold. If the location of a device cannot be determined, the enhanced KNN algorithm will return a special value indicating that it cannot determine the location, at which point the mobile device will fall back to existing localization methods to determine the location of the device.

The enhanced KNN classifier employs the Canberra distance function [19] (5) as a similarity measure for the numerical attribute (time), which is defined as:

$$dist(X, Y) = \sum_{i=1}^n \frac{|x_i - y_i|}{|x_i| + |y_i|} \quad (5)$$

where $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$

For nominal attributes (Bluetooth devices and location classes), we use our proposed Bluetooth distance measure which calculates the distance between multiple nominal values (attributes) while ignoring the order in which they appear so that if they match, the distance is 0; otherwise, the distance is 1. Therefore, when the enhanced KNN classifier is computing the closeness between two mobile device samples to predict the location, the order the device appears in the scan should not interfere. Table V can be used to show an example of the different results achieved using the KNN algorithm which considered the order of the nominal attributes and our proposed enhanced KNN algorithm. When attempting to determine which instance is closest to instance 0, the KNN algorithm will select instance 2 since it considers the order of the devices when calculating the distance (device B matches). However, the enhanced KNN classifier will select instance 1, due to the fact that both instance 0 and 1 contain all the same devices, but in a different order.

TABLE V. EXAMPLE DATA FOR KNN ALGORITHM COMPARISON

| Instance | Time | Device A | Device B | Device C |
|----------|------|----------|----------|----------|
| 0 | 11 | 24 | 22 | 23 |
| 1 | 11 | 22 | 23 | 24 |
| 2 | 11 | 26 | 22 | 27 |

F. Bluetooth Power Usage comparison

The research on power consumption in [20] presents the results that were obtained using an Android G1 mobile device; however BlackBerry devices have different power characteristics. BlackBerry devices are known to be very efficient in terms of power consumption, so the way they handle GPS, Wi-Fi, Cellular and Bluetooth invocations will be very different from that of an Android device. As a result, an experiment was set up which uses a BlackBerry Pearl 9100 device running an application which continuously uses one of the available localization methods until the battery is depleted. To ensure a completely fair test, the device was security wiped and no other third party applications were running on the device. The tests started at 100% battery level and ended at 5% battery level. The results of the test show that the BlackBerry Pearl 9100 device used the most amount of energy when the GPS receiver is active, lasting about 36.4 hours. The Bluetooth radio on the other hand was more energy efficient using about half the energy of the GPS receiver lasting 63.7 hours. Lastly, the Wi-Fi radio used the least amount of energy lasting 105.5 hours.

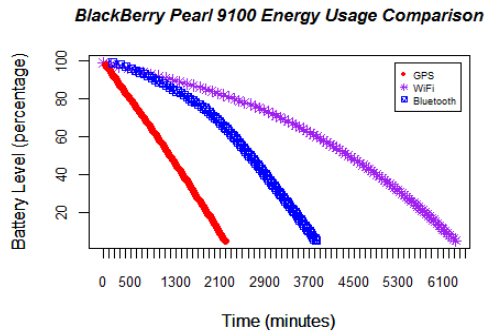


Figure 5. BlackBerry Pearl 9100 Energy Usage Comparison

The results of the experiment are presented in figure 5 and show that although Wi-Fi uses the least amount of energy on a BlackBerry device, the Bluetooth radio is still a very viable option for energy efficient localization as it still uses half the energy of the GPS receiver.

IV. EXPERIMENTAL RESULTS AND EVALUATION

In our experiments, we tested the effect of limiting Bluetooth scans to a specific number of devices. This is due to the fact that if a lower number is chosen, the scan time will decrease. In addition, it was found in separate tests within the University of Guelph that on average there are 3 devices that are discoverable when performing a Bluetooth scan. The results of the completed tests to evaluate the discoverability of mobile devices are shown in table VI. Each test was completed with the use of a BlackBerry 9800 smartphone which were carried by an individual in order to obtain a good variety of devices around the campus. In each test, the device performed Bluetooth scans every 30 minutes when the device was within the confines of the University of Guelph for a two week period.

TABLE VI. RESULTS OF DISCOVERABILITY TESTS

| | Total number of unique devices found | Maximum number of unique devices found in a single Bluetooth scan | Average number of unique devices found in a single Bluetooth scan |
|----------------|--------------------------------------|---|---|
| Test 1 | 179 | 26 | 5.12 |
| Test 2 | 234 | 20 | 2.24 |
| Test 3 | 90 | 18 | 2.60 |
| Average | 167.6 | 21.3 | 3.32 |

TABLE VII. RESULTS OF DISCOVERABILITY TESTS

| | Cellphones | Smartphones | Laptops | Tablets |
|---------------|------------|-------------|---------|---------|
| Test 1 | 10 | 26 | 142 | 1 |
| Test 2 | 34 | 84 | 110 | 6 |
| Test 3 | 19 | 25 | 46 | 0 |

The results obtained are broken down into specific device types in table VII. As can be seen, the majority of devices that were discoverable in each test are laptops, which makes sense given that tests were performed in a University environment. We are able to break these results down due to the 12 device type identification bits which are part of the Bluetooth protocol [2] [1], and are transferred whenever a Bluetooth device responds to an inquiry. The results shown prove that there are a sufficient number of Bluetooth devices that are discoverable.

To evaluate the effectiveness of the enhanced KNN classification algorithm over the traditional KNN classification algorithm, we used F-measure as a performance evaluation measure. F-measure is shown in equation 6 and is function which measures the ability of a classifier based on the precision (7) and recall (8).

$$F\text{-measure} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

The precision and recall equations contain True Positive (TP), False Positive (FP) and False Negative (FN). TP represents positive samples that are predicted correctly. FP represents negative samples that are predicted as the positive class. FN represents positive samples that are predicted as the negative class. In our case, the incorporation of F-measure as a performance evaluation measure is used rather than accuracy as our data set is not completely balanced among the classes, and as a result, is more suitable for real-world data which by nature may not be balanced.

Figure 6 shows the performance of the KNN classifier on the collected Bluetooth scan data. In our experiments, the K-values tested are 2, 4, 6, 8 and 10, and it was found that a K-value of 2 had the highest F-measure on the Bluetooth scan data. We also experimented with limiting the Bluetooth scan to find either 2, 4, 6, 7 and 10 device to see how it affects the F-measure.

The results show that when limiting the Bluetooth scan to a maximum of 2 devices, the highest F-measure is obtained

(42.37%) and the lowest F-measure occurs when the Bluetooth scan is limited to 10 devices (F-measure of 35.27%). There is a downward trend due to the design of KNN which considers the order in which the devices appear.

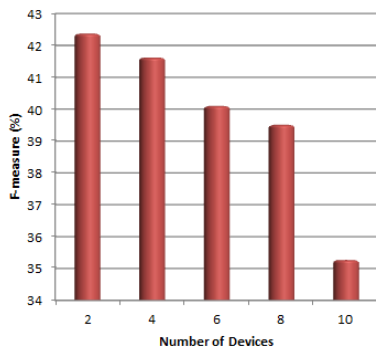


Figure 6. Performance of KNN with a K-value of 2

The enhanced KNN classifier is evaluated in two schemes. The first scheme is most similar to the KNN algorithm in which the algorithm will attempt to determine the location of a mobile device even if there is no previously stored information about the surrounding Bluetooth device. The second scheme of the enhanced KNN classifier experimented with does not make a prediction when there is no previous data about the Bluetooth devices surrounding the mobile device. This is due to the fact that in a real-world scenario, it is possible that when determining the location of a Bluetooth device based on a Bluetooth scan, none of the stored samples will match the device(s) found in the new scan. We refer to the second scheme as the enhanced KNN classifier which omits non-predictable samples.

The performance first scheme is shown in figure 7. Again, multiple K-values were experimented with and it was found that a K-value of 2 had the highest F-measure on the Bluetooth scan data. The results show that when limiting the Bluetooth scan to a maximum of 10 devices, the highest F-measure is achieved (73.8%) and the lowest F-measure occurs when limiting the Bluetooth scan to two devices (65.08%). This upward trend is due to the enhanced KNN classifier's ability to handle out of order Bluetooth scan data. In terms of overall improvement over the KNN algorithm, the average performance of the Enhanced-KNN classifier is 29.6% above the average performance of the KNN algorithm.

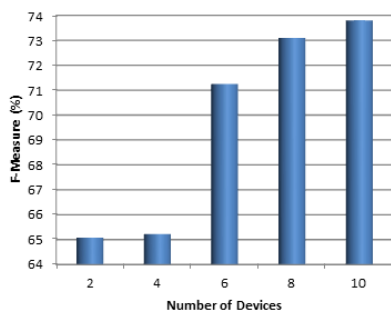


Figure 7. Performance of Enhanced KNN with a K-value of 2

The performance of the second scheme of the enhanced KNN algorithm which only determines the location of a device if previous data is stored about the surrounding Bluetooth devices is shown in figure 8. The experiments performed on multiple K-values showed that in this scheme, a K-value of 1 had the highest F-measure on the Bluetooth scan data.

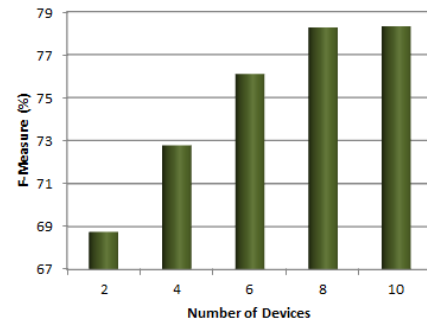


Figure 8. Performance of Enhanced KNN while Omitting Samples with a K-value of 1

The results obtained show a similar trend of the Enhanced-KNN classifier in which the highest F-measure occurs with a maximum of 10 Bluetooth devices per scan (78.31%), and the lowest F-measure occurs with a maximum of 2 Bluetooth devices per scan (68.76%). In terms of overall improvement, the average performance of the Enhanced-KNN classifier which omits samples over the original KNN algorithm is 34.6%.

Figure 9 shows a comparison of all classification algorithms experimented with for the classification of location classes using Bluetooth scan information. As can be seen in the figure, the enhanced KNN classifier outperformed the KNN classifier in all cases. This confirms that using an enhanced KNN is beneficial for the purpose of Bluetooth localization.

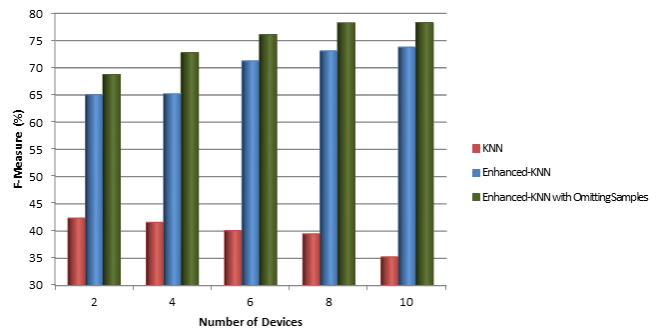


Figure 9. Performance Comparison of the KNN Classifiers

Based on the results obtained with the enhanced KNN classifier, and the discoverability tests, we experimented with our proposed method in a real-world scenario within the University of Guelph. In our experiments, a maximum of 2 devices is chosen which to limit Bluetooth scans, which also ensures the fastest response time. We employed an accuracy threshold of 400 meters in

our real-world testing. This was done to ensure only samples which have a recorded localization accuracy of 400 meters or less are used for location determination purposes. By incorporating this threshold, we also ensure that recorded localization data about Bluetooth devices using cellular localization is not used for determining the location of a mobile device as it will be extremely inaccurate.

Table VIII presents the results achieved when experimenting within real-world locations. As can be seen in the table, for 6 separate locations, the proposed collaborative Bluetooth localization method was able to localize the mobile device using Bluetooth with an average of 32.61 meters error from the actual GPS location.

Table IX presents the average response time of traditional methods we obtained in comparison to Bluetooth localization. The response time of each method is retrieved using 5 different tests in 5 separate geographical locations. In each test, specifically with GPS, localization is done with a cold start. This means no prior information about satellites is known before the location of the device is requested, which resulted in a response time of around 15 seconds for GPS localization. As can be seen, the response time of Bluetooth localization is less than GPS, but greater than Wi-Fi and Cellular localization methods.

V. CASE STUDY ON WIRELESS PRINTING

A case study was also performed to evaluate the use of the described method in a practical application. Within the University of Guelph, we setup test for the task of location-based printing. The test was completed with the use of BlackBerry devices running a third party application which takes advantage of the Bluetooth localization framework. Within the application, a user simply needs to select a document on their device to print, and using the collaborated information about the Bluetooth devices which surround it, the document will print to the closest printer. In all cases tested, the BlackBerry devices were able to determine its location successfully and printed the selected document to the correct printer. Screenshots of the application are shown in figure 10 and 11.

TABLE VIII. REAL-WORLD RESULTS OF BLUETOOTH LOCALIZATION

| Location using Bluetooth | | GPS Location | | Additional Information | |
|--------------------------|-----------|--------------|-----------|------------------------|-------------------|
| Latitude | Longitude | Latitude | Longitude | Error in Meters | Test Location |
| 43.53116 | -80.22798 | 43.53144 | -80.22773 | 36.59 | Library |
| 43.52996 | -80.22878 | 43.53041 | -80.22853 | 53.15 | Science Complex |
| 43.53066 | -80.22638 | 43.53068 | -80.22625 | 10.18 | University Centre |
| 43.53056 | -80.22908 | 43.53077 | -80.22905 | 22.53 | Reynolds Bldg |
| 43.53234 | -80.22723 | 43.53264 | -80.22726 | 32.92 | MacKinnon Bldg |
| 43.53104 | -80.229 | 43.53126 | -80.22860 | 40.29 | MacLachlan Bldg |

TABLE IX. RESPONSE TIME OF ALL LOCALIZATION METHODS

| Localization Method | Average Response Time (seconds) |
|------------------------|---------------------------------|
| Bluetooth Localization | 4.0882 |
| Wi-Fi Localization | 0.244 |
| Cellular Localization | 2.271 |
| GPS Localization | 15.238 |



Figure 10. Main Menu of the Location Based Printing Application



Figure 11. Closest Printer Found Within the Location Based Printing Application

VI. PRIVACY CONCERNS

The proposed method takes advantage of the recorded location of discoverable Bluetooth devices in a given environment in order to determine the location of a mobile device. Due to the fact that we are dealing with location data, there are some privacy concerns. The main focus of the work was not on addressing these concerns, however there were steps taken which took the privacy of the users and devices into consideration.

As MAC address data was collected, it was transformed into a unique identifier, and used within the proposed method. In addition, MAC address data was never linked to an individual, therefore, the only way to determine which individual is associated with each MAC address would be through direct contact, which was never done. The location data of all Bluetooth devices was also never continuously tracked. Data only exist about devices

at fixed points in time when Bluetooth scans were performed, and each device was discoverable.

VII. CONCLUSION AND FUTURE WORK

Location services are becoming more and more common, and with the increased popularity of mobile devices, the demand for these services will increase. Currently, traditional methods of localization are prevalently used. As the Bluetooth radio on mobile device is very common, fast and energy efficient, we proposed a collaborative Bluetooth localization method. We found that by using data about the Bluetooth devices which is obtained a priori, future localization requests could be completed based on the current Bluetooth devices found.

There are many other potential applications that could be possible when using Bluetooth for localization. Our future work includes the use of Bluetooth to suppress GPS invocations. By analyzing the Bluetooth devices found in a given area, it can be determined whether or not a mobile device has moved from its current location. In essence, if the same devices are found in subsequent Bluetooth scans, it can be assumed that the mobile device has not moved from its previously obtained location, and thus an invocation of the GPS can be suppressed. As privacy is a big concern in our society, future work of this research aims also to determine how to further transform the data into a more anonymous state. In terms of the data used, because it is time sensitive, it has the potential to expire or become irrelevant over time. Research needs to be completed in determining the appropriate amount of time to keep the collected data about specific Bluetooth devices so that the information used to complete localization requests is always current and accurate.

ACKNOWLEDGMENT

This work is supported in part by RIM, through the Centre for Mobile Education and Research (cmer.uoguelph.ca).

REFERENCES

- [1] Bluetooth Baseband. [Online]. <https://www.bluetooth.org/Technical/AssignedNumbers/baseband.htm>.
- [2] Bluetooth Core Specifications. [Online]. <https://www.bluetooth.org/Building/HowTechnologyWorks/CoreSpecifications.htm>.
- [3] W. Chan, "DealFinder: A Collaborative, Location-Aware Mobile Shopping Application," *MIT Media Laboratory*, 2001.
- [4] N. Chang, R. Rashidzadeh, and M. Ahmadi, "Robust Indoor Positioning using Differential Wi-Fi Access Points," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 3, pp. 1860-1867, 2010.
- [5] S. Chawathe, "Low-latency Indoor Localization using Bluetooth Beacons," *12th International IEEE Conference on Intelligent Transportation Systems, 2009. ITSC '09.*, pp. 1-7, 2009.
- [6] R. Cheng, K. Rothermel, and T. Farrell, "Energy-Efficient Monitoring of Mobile Objects with Uncertainty-Aware Tolerances," *11th International Database Engineering and Applications Symposium*, pp. 129-140, 2007.
- [7] R. Cheng, K. Rothermel, and T. Farrell, "Energy-Efficient Monitoring of Mobile Objects with Uncertainty-Aware Tolerances," *Database Engineering and Applications Symposium, 2007. IDEAS 2007. 11th International*, pp. 129-140, 2007.
- [8] K. Cheung, S. Intille, and K. Larson. An Inexpensive Bluetooth-Based Indoor Positioning Hack. [Online]. http://stuff.mit.edu:8001/afs/athena/dept/cron/group/house_n/documents/CheungIntilleLarson2006.pdf.
- [9] H. Chu et al., "Impact of Sensor-Enhanced Mobility Prediction on the Design of Energy-Efficient Localization," *Ad Hoc Netw.*, vol. 6, pp. 1221-1237, 2008.
- [10] I. Constandache, R. Choudhury, and I. Rhee, "Towards Mobile Phone Localization without War-Driving," *INFOCOM 2010 Proceedings IEEE*, pp. 1-9, 2010.
- [11] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," *IEEE Transactions on Information Theory*, vol. XIII, no. 1, pp. 21-27, 1967.
- [12] G. Fischer, B. Dietrich, and F. Winkler, "Bluetooth Indoor Localization System," *1st Workshop on Positioning, Navigation and Communication*, vol. 1, pp. 147-156, 2004.
- [13] J. Hallberg, M. Nilsson, and K. Synnes, "Positioning with Bluetooth," *10th International Conference on Telecommunications, 2003. ICT 2003.*, vol. 2, pp. 954-958, 2003.
- [14] J. Han and M. Kamber, *Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)*.: Morgan Kaufmann, 2000.
- [15] M. Heger. (2008, June) Human Travel Patterns Surprisingly Predictable. [Online]. <http://spectrum.ieee.org/telecom/wireless/human-travel-patterns-surprisingly-predictable>.
- [16] D. H. Kim, Y. Kim, D. Estrin, and M. B. Srivastava, "SensLoc: sensing everyday places and paths using less energy," *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pp. 43-54, 2010.
- [17] D. Kim, Y. Kim, D. Estrin, and M. Srivastava, "SensLoc: Sensing Everyday Places and Paths using Less Energy," *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pp. 43-54, 2010.
- [18] Kowoma - GPS System. [Online]. <http://www.kowoma.de/en/gps/>.
- [19] G. N. Lance and W. T. Williams, "Mixed-Data Classificatory Programs. I. Agglomerative Systems," *Australian Computer Journal*, pp. 15-20, 1967.
- [20] K. Lin, A. Kansal, D. LyMBERopoulos, and F. Zhao, "Energy-Accuracy Trade-off for Continuous Mobile Device Location," *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pp. 285-298, 2010.
- [21] Skyhook Wireless. [Online]. <http://www.skyhookwireless.com/howitworks/>.
- [22] United States Updates Global Positioning System Technology. [Online]. <http://www.america.gov/st/washfile-english/2006/February/20060203125928lcniirellep0.5061609.html>.
- [23] C. Veness. Calculate Distance, Bearing and More Between Latitude/Longitude Points. [Online]. <http://www.movable-type.co.uk/scripts/latlong.html>.
- [24] Y. Wang et al., "A Framework of Energy Efficient Mobile Sensing for Automatic User State Recognition," *Proceedings of the 7th international conference on Mobile systems, applications, and services*, p. 2009, 179-192.
- [25] Z. Zhenyun, K. Kim, and J. Singh, "Improving Energy Efficiency of Location Sensing on Smartphones," *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pp. 315-330, 2010.