CLUSTERED CROWD GPS FOR PRIVACY VALUING ACTIVE LOCALIZATION

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

BENJAMIN C HURDINS

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering Kothamangalam

CLUSTERED CROWD GPS FOR PRIVACY VALUING ACTIVE LOCALIZATION

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

BENJAMIN C HURDINS

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering Kothamangalam

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MAR ATHANASIUS COLLEGE OF ENGINEERING
KOTHAMANGALAM

CERTIFICATE

This is to certify that the report entitled **Clustered Crowd GPS for Privacy Valuing Active Localization** *submitted by* **Mr. BENJAMIN C HURDINS**, Reg. No. **MAC15CS020** *towards partial fulfillment of the requirement for the award of Degree of Bachelor of Technology in Computer science and Engineering from APJ Abdul Kalam Technological University for December 2018 is a bonafide record of the seminar carried out by him under our supervision and guidance.*

....	....	....
Prof. Joby George	**Prof. Neethu Subash**	**Dr. Surekha Mariam Varghese**
Faculty Guide	***Faculty Guide***	***Head Of Department***

Date: Dept. Seal

ACKNOWLEDGEMENT

First and foremost, I sincerely thank the 'God Almighty' for his grace for the successful and timely completion of the seminar.

I express my sincere gratitude and thanks to Dr. Solly George T, Principal and Dr.Surekha Mariam Varghese, Head of the Department for providing the necessary facilities and their encouragement and support

I owe special thanks to the staff in charge Prof.Joby George, Prof.Neethu Subash and Prof.Joby Anu Mathew for their corrections, suggestions and sincere effort. I express my sincere thanks to staff members in the Department of Computer Science and Engineering who have taken sincere efforts in helping me to conduct this seminar.

Finally I would like to acknowledge the heart felt efforts, comments, criticisms, cooperation and tremendous support given to me by my dear friends during the preparation of the seminar and also during the presentation without whose support this work would have been all the more difficult to accomplish

ABSTRACT

Crowd sourcing-based approaches have recently emerged as a promising solution for localization of lost objects or individuals (e.g., children or elders). The proliferation of mobile devices having Bluetooth Low Emission capability and the introduction of Beacon technology are reasons for it. Crowd GPS service popularity has not extended beyond passive mode in which localization is achieved in the background without intruding the mobility of users. Objects of care could be tracked and localized by user devices in the proximity by attaching affordable Beacon tags to them. The localization of lost objects through the crowd GPS service is studied in an active manner. Clustering users in a Beacon tag network is proposed based on the benefits they can receive from each other in terms of the localization of their lost items. A new metric is developed to quantify this benefit and the users that can provide most of the total possible benefits to each other are then grouped together so that they can provide active localization service for only the users most beneficial to them. The clustering of users is achieved based on a greedy heuristic based algorithm.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATION

GPS	Global Positioning System
BLE	Bluetooth Low Energy
NFC	Near Field Communication
RFID	Radio Frequency Identification
QR	Quick Response
RSSI	Received Signal Strength Indication
POI	Point of Interest
STD	Social Tracking Distance

INTRODUCTION

As human beings, we lose or misplace various of our belongings every day such as mobile phones, keys and sunglasses. However, searching for them could be time consuming (i.e., 15-20 minutes per day) as shown by several reports[1]. When we lose things outside rather than indoors, the localization task could be more challenging since the search area gets larger. Moreover, when what we lose is our loved ones such as pets[2] and vulnerable individuals like children[3] and elders[4] ,the process of finding torments us more due to involvement of emotions.

Recently, with the widespread adoption of smartphones, crowdsourcing based solutions have been provided for this challenging task. Harnessing the power of a mass of geographically dispersed user devices with Internet connectivity, an invaluable crowd GPS service is developed towards the goal of localizing lost items. With the release of Bluetooth Low Energy (BLE) that allows low-cost nearrange[5] communication , several vendors have created tiny, battery-powered, BLE tags that can easily be attached to the objects of care. Today, there are communities of over several millions of these devices managed by several vendors[6][7].

Mobile devices in the vicinity via BLE send periodic updates (i.e., beacons) to the device to indicate its presence. However, such updates do not reach to the device when the distance between the device and tags is more than the limited range of BLE communication. With the crowd GPS service, collaboration of multiple user devices is targeted to achieve an enhanced coverage for the localization of lost items. That is, when an item with a tag attached is lost, any nearby user device in BLE range could detect it in a transparent way and notify the server and eventually the user who owns it with the current location information.

Clearly, the benefit of such a collaborative sensing system will be pronounced with increasing number of users participating in the system. Moreover, the success in areas with more user density will boost compared the other areas. Despite the popularity of crowd GPS, due to its design, its benefit to localization of lost items does not extend beyond passive localization by the participating devices. That is,the localization of lost tags is achieved in the background without intruding the mobility of users. If the user by chance passes by the lost tag, the observation is sent to the server transparent to the user. While such a design provides a service without disturbing users, the benefit stays limited as it is not controllable.

An interesting approach, which could potentially extend the benefit of such a system, is proposed by Locus Pocus[8] , which aims to monetize the service of searching of the lost objects by charging a fee. Such an approach could potentially trigger active participation of users in this process. For example, users who are in the areas where the lost item has been recently seen, could change their mobility for sometime and look for it more actively depending on the price offered. The concept could be considered under the umbrella of a specific type of crowdsourcing called spatial crowdsourcing. Unlike traditional crowdsourcing such as image tagging, where tasks can be performed anywhere, in the context of spatial crowdsourcing, workers need to move to a specified location physically and perform a task (e.g., taking a photo of a point of interest (POI)) before the expiration time in order to successfully perform a task[9] .

In the spatial crowdsourcing context, the main focus is the optimal matching of workers and tasks [10] while giving priorities to several different factors (e.g., privacy, minimization of the cost). Thus, in the context of crowd GPS which has been considered in non-intrusive or passive manner so far, if an active localization model will be employed, similar to other spatial crowdsourcing models, incentives should also be provided to users for their efforts.

For some specific cases like finding missing children, users could be motivated to voluntarily participate in such a system. Otherwise, providing fee based incentives may contradict with the motivation of masses for adopting crowd GPS, as it is free of charge after the Beacon tags are purchased. In order to design such a complimentary service while utilizing active spatial crowdsourcing based localization of lost items, we propose to form groups of users depending on their historical visit patterns and let each member of the group benefit from other members freely. This can provide mutually similar benefit among group members and release the burden of providing incentives required in spatial crowdsourcing based task assignments.

Bluetooth beacons use Bluetooth low energy proximity sensing to transmit a universally unique identifier picked up by a compatible app or operating system. The identifier and several bytes sent with it can be used to determine the device's physical location,track customers, or trigger a location-based action on the device such as a check-in on social media or a push notification.Bluetooth beacons differs from some other location-based technologies as the broadcasting device (beacon) is only a 1-way transmitter to the receiving smartphone or receiving device, and necessitates a specific app installed on the device to interact with the beacons. This ensures that only the installed app (not the Bluetooth beacon transmitter) can track users, potentially against their will, as they passively walk around the transmitters.

Note that while users initiate tasks (e.g.,find my lost item in this area), they directly interact with other users in their group and notify them about their private location information. Thus, group sizes should be determined valuing the privacy of users at its maximum and inclusion of a user in a group should be allowed only if the improvement in network level benefit is worth the additional exposure of that user's privacy.

The nearest network node is determined using some kind of signal emitted by the node. The smartphone can sense this signal and determine its own position related to the known network node position.

RELATED WORKS

Localization of people's belongings through the sensors on mobile devices has recently been studied under different names such as people centric-sensing , participatory sensing and mobile crowd sensing . Especially with the proliferation of smartphones that are equipped with multiple sensors, the need for deploying and maintaining separate dedicated sensors for such kind of service is invalidated.

With the release of Bluetooth Low Energy (BLE) technology, the smartphones have received the capability of communication with nearby devices with lower consumption at lower data rates. As a result, high power savings (i.e., 60-80%) achieved compared to previous technologies . After Apple devised the iBeacon standard protocol in 2013 , BLE has become more popular and several BLE-enabled products are released by multiple manufacturers.

Beacons tags are BLE devices that can periodically (e.g., every 100 ms in Apple's iBeacon standard) advertise themselves to their surroundings to be discovered by other BLE capable devices. Thanks to the fexibility in packet format, it is also possible to send some limited data during the broadcasting of these advertisement packets without making an actual connection to nearby devices. Compared to the other nearby communication technologies such as QR codes and NFC, Beacons are also more convenient since they require the least interactions with users. Moreover, compared to RFID based localization , Beacons are also easy to deploy as most of the smartphones today support BLE technology.

NFC is the data transfer technique and it is the type of the wireless communication with short range, The data can be transferred in the form of the beam by touching the two things together and the single wave or the beam helps transfer the data between two devices within the range of (4 cm).NFC (Near field communication) consumes little power than Bluetooth technology, It helps to transfer the data at a faster rate, The devices connected using NFC must be close proximity to each other in the crowded locations to prevent the interference caused when the other devices are present and they are trying to communicate.While Bluetooth may have trouble dealing with the interference when trying to send the signals between two devices when many devices are in close proximity, Bluetooth requires the users to manually set up the connections between the smartphones and it takes several seconds.

Bluetooth offer the longer signal range for connecting during the data communication and transfers, NFC technology can connect two devices quickly, then it turns the signal over to Bluetooth, so, you can move further away without severing the connection.

But Bluetooth is limited in its transfer of data, It can not transfer the data between more than two devices, The data theft may occur because of the long distance for the data transfer, When it is installed, The viruses or the bugs might be downloaded with it.

To perform a real-time indoor navigation , we need a visual-tags system which is robust to rotations and deformations, and which is as fast as possible in detection/decoding phases.

While the GPS Technology has an excellent performance outdoor, it fails into the indoor environments, due to the fact that the GPS signal is not available inside the buildings. A lot of indoor navigation techniques were proposed in literature, with different accuracy and complexity levels, but there is not a unique solution today, and the problem is still matter of research.The massive diffusion of smartphone, which is a perfect gateway between real and digital life, enables a lot of services for the user.

Most of them are strictly related to the position so it is important to know where the user is in a certain moment and how he can navigate towards a destination. Thinking to the healthcare sector, it is common that patients and visitors, even if there are signage deployed on the environment, lost their way inside large hospitals. Moreover, patients with physical handicaps can experiment more difficulties than other people to orient themselves inside the building. An accurate smartphone indoor navigation system can be very useful for all these specific situations: it is possible to provide both visual and voice turn-by-turn navigation for people with visual or hearing deficiency, create specific POI and provide a way to reach them, find a doctor or locate a patient's room in a medical facilities, and lot more.

The BLE functionalities and Beacons have recently been used in several applications in different domains such as indoor localization , navigation , ticketing , proximity marketing and localization of missing and lost items. In indoor localization, utilizing the RSSI (Received Signal Strength Indication) value of BLE signals, the distance of the items are detected to be able to locate the items accurately. To the point localizations are shown to be possible in recent studies , with the integration of tags inside furniture that makes them searchable. A simple prototype that use beacons for localization of personal items is also implemented and tested.

In the commercial world, this concept has also attracted a lot of interest and several specific devices for various purposes (e.g., tracking of pets and children) have been developed.

There are also some work that study the security aspects of crowd GPS applications and provide efficient and privacy preserving designs. In Techu , a privacy preserving system is introduced for Beacon based tracking systems. Rather than a vulnerable centralized design that could be exposed to single point failures, a unique bulletin board based observation posting system is introduced. Users report their observations of tags to a server that could be untrusted, however the actual tag location information is stored locally only in the observers. Once the owner of the tags claims the ownership of the lost tag to the observer directly (i.e. server is not involved in this communication), the location information is disclosed.

THE PROPOSED METHOD

In this section, we elaborate on the design of the proposed system. We first develop a new metric to quantify the potential benefits of users to each other based on the relation of their visit patterns at a region (i.e., POI). Then, we study the clustering of users for building a complimentary active localization service for lost items.

3.1 Social Tracking Distance Metric

Let a=(t_s , t_e ,loc_{id}) denote a visit event by a user a with t_s and t_e denoting the start and end times of the visit and loc_{id} denoting the visited location id. All of the visits of a user at that location could be represented with a set V_a, in which the end time of the previous event is always smaller than the start of the next event.

$$\mathcal{V}_a = \{a_1, a_2, a_3, \ldots, a_n\} \text{ where}$$
$$a_i.t_e < a_{(i+1)}.t_s, \quad \forall i \in \{1..n\}$$

Since a Beacon attached item could only be detected by a mobile device within certain proximity (i.e., preset BLE range), we use a probability to denote the likelihood for the detection of the lost item by the user's mobile device that is in the same location at the current time unit. Moreover, in order to include the increased likelihood of finding the items by owners compared to other users (as they might remember exact spots they lost the item within the location), we use p_s for probability of finding by self, and p_o, for probability of finding by others.

To quantify the benefit of a user B to another user A in terms of finding his/her lost items, we propose a metric called Social Tracking Distance (STD), inspired by the metrics , used in analyzing contact patterns in DTNs. Consider the sample visit history of three nodes A, B and C in a location shown in Fig.3.1. The visits of each user is shown in a different timeline row. The i^{th} visit of a user, A, is labeled as a_i. We assume that the time is divided into equal time units and the durations of visits are denoted with $\delta(.)$ time units and the time passed since the visit of user A's i^{th} visit to the user B's j^{th} visit is denoted as $\Delta(a_i,b_j)$ time units. Without

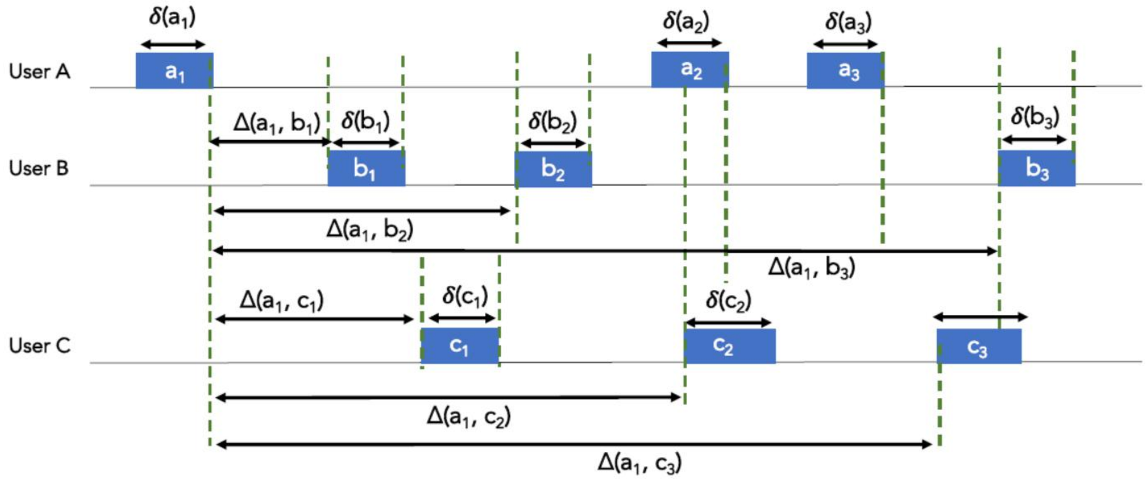

Figure 3.1: Sample visit patterns of three users.

loss of generality, assume that there are n visits of each user in a specific area.We define the $STD_{(A,B)}$ metric as the average delay that user B's device will sense the lost item (i.e., Beacon attached to the item) of user A. To calculate it, for each possible time unit during the visits of user A, we find the probability of finding in the upcoming visits of all users in the same location and corresponding delay. Then, we find the weighted average of these delays over all possible locations (i.e., POI).

It is assumed that when user A loses an item during a visit, she will notice that she lost the item after she left the area and start a search process in the network. Thus, we assume that the item will not be found during the same visit it is lost. It is possible that when the owner visits the same location later, she may or may not find the item during that visit. The findings during visits by other users will only be considered if the owner or other users in earlier visits do not find the item. User A's device can potentially lose the item at any time during her visit (in range (0, $\delta(a_i)$)) and user B's device can potentially sense the lost item at any time during her visit (in range (0,$\delta(b_j)$)). This results in a range of ($\Delta(a_i$,b_j)= $\Delta(a_i$,$b_j)$ + $\delta(a_i)$ + $\delta(b_j)$) for the delay of finding the lost item since it is lost. However, the probability of each of these delays is different, and can be calculated using the visit durations and their temporal relations at the same location.

Assume that user A lost an item around x time units before her current visit, a_i, ends in that region. If user B can detect the presence of that item in that region, the average delay of finding it will be:

$$\mathcal{D}(a_i, x) = \sum_{j=s}^{n} \sum_{y=1}^{\delta(b_j)} \left((\Delta(a_i, b_j) + x + y) p_{(b_j,y)}^{(a_i,x)} \right)$$
$$\text{where, } s = \arg\min_k \{a_i.t_e < b_k.t_s\}$$

(Equ:1)

Here, p_y^x (a_i,b_j) denotes the probability of finding the item, that is lost x time units before the end of a_i visit, at y_{th} time unit of visit b_j. This probability should be calculated considering the visits of other users and the owner of the item in the same region before this time unit as well as the previous visits of user B and previous time units in the current visit of B. The item should not be found in previous time units and should be found in this exact time unit. That is,

$$p_{(b_j,y)}^{a_i} = p_o(1-p_o)^{y-1} \beta_y^{pre}\binom{a_i}{b_j}$$
$$\text{where, } \beta_y^{pre}\binom{a_i}{b_j} = (1-p_o)^{\beta_y'\binom{a_i}{b_j}} (1-p_s)^{\beta_y\binom{a_i}{b_j}}$$

Here, $\beta_y\binom{a_i}{b_j}(a_i,\ b_j\ )$ denotes the sum of visit durations of the owner (i.e., user A) and $\beta_y'\binom{a_i}{b_j}(a_i,\ b_j\ )$ denotes the sum of visit durations of the other users, between visits a_i and b_j . More formally:

$$\beta_y\binom{a_i}{b_j} = \left(\sum_{\forall k, i < k \le m} \delta(a_k) \right) + \delta_{b_j}'(a_{k+1}) \text{ where,}$$
$$m = \arg\max_k \{a_k.t_e < b_j.t_s + y\}$$
$$\delta_{b_j}'(a_{k+1}) = \max\{0, b_j.t_s + y - a_{k+1}.t_s\}, \quad \text{if } \exists a_{k+1} \in \mathcal{V}_A$$
$$\beta_y'\binom{a_i}{b_j} = \sum_{\forall u \in \mathcal{U}, u \ne A} \left(\left(\sum_{\forall k, l_u \le k \le m_u} \delta(u_k) \right) + \delta'(u_{k+1}) \right)$$
$$\text{where, } l_u = \arg\min_k \{a_i.t_e < u_k.t_s\}$$
$$\text{where, } m_u = \arg\max_k \{u_k.t_e < b_j.t_s\}$$

Note that, $\delta'(.)$ is used to denote the part of the visits which did not end yet and can still contribute to the finding of visits.

Having the expected delay formula for finding an item lost at a specific time unit in a visit of user A, we then iterate through all possible losing times to calculate the STD metric:

$$STD_{(A,B)} = \frac{\sum_{i=1}^{n}\left(\sum_{x=1}^{\delta(a_k)} \mathcal{D}(a_c, x)\right)}{P_{(A,B)}}$$

Note that the numerator in the formula above is the sum of products of probabilities and delays, and the denominator is the sum of all probabilities used. This average delay is the expected value of the delay assuming that A's lost item is found by B. The probability that it will be found by B is:

$$P(A,B) = \frac{\sum_{i=1}^{n}\sum_{x=1}^{\delta(a_i)}\left(\sum_{j=s}^{n}\sum_{y=1}^{\delta(b_j)} p_{(b_j,y)}^{(a_i,x)}\right)}{\sum_{i=1}^{n}\delta(a_k)}$$

Simplifying the equations, the general formula for the $STD_{(A,B)}$ could be rewritten in a more structured way as:

$$STD_{(A,B)} = \frac{\sum_{\forall a_i}\left(\delta(a_i)\left(f_1(a_i) + f_2(a_i) + \sum_{j=s}^{n} f_v(b_j)\right)\right)}{\sum_{\forall a_i}\delta(a_i)p(a_i)}$$

where,

$$\begin{aligned}
f_1(a_i) &= \sum_{\forall b_j.t_s > a_i.t_e} \Delta(a_i, u_j) p_f^{\delta(u_j)} \beta_0^{pre}\binom{a_i}{b_j} \\
p_f^d &= 1 - (1-p)^d \\
f_2(a_i) &= (\delta(a_i) + 1)p(a_i)/2 \\
p(a_i) &= \sum_{j=s}^{n} p_f^{\delta(b_j)} \beta_0^{pre}\binom{a_i}{b_j} \\
f_v(b_j) &= \left(p_f^{\delta(b_j)}/p + (1 - p_f^{\delta(b_j)})\delta(b_j)\right) \beta_0^{pre}\binom{a_i}{b_j}
\end{aligned}$$

Similarly, $P_{(A,B)}$ could be simplified as:

$$P_{(A,B)} = \left(\sum_{\forall a_i} \delta(a_i) p(a_i) \right) / \sum_{\forall a_i} \delta(a_i)$$

STD metric defines the expected delay of finding and is derived from over all scenarios that ends up with item's finding. However, it is possible that the item may not be found during all visits, thus, that probability should be considered in defining the benefit of user B to A. Moreover, the STD value for the same pair of nodes can vary at different locations. To accommodate the impact of such differences in the average benefit of users to each other, we define a weighted satisfaction value for user B's efforts in finding the lost items of user A in any of the locations visited by A.

$$\gamma_{(A,B)} = \sum_{\forall r} \left(w_r^A \left(\frac{P^r_{(A,B)}}{STD^r_{(A,B)}} \right) \right)$$

(Equ:2)

where, w_r^a denotes the weight of the region r (i.e., total visit durations by A in region r within all visit durations in all regions).

We also define the average delay of finding a user's item by any user in the network as follows:

$$STD_A = \frac{\sum_{\forall A} \left(STD_{(A,A')} P_{(A,A')} \sum_{i=1}^{n} \delta(a_i) \right)}{\sum_{\forall A} \left(P(A,A') \sum_{i=1}^{n} \delta(a) \right)}$$

(Equ:3)

Here, A' notes all other nodes except A in the network. The satisfaction of a user from all other nodes in the network can also be computed using a similar formulation to equation 2 .

3.2 Clustering of Users

Once the satisfaction values of each user from every other user in the network is found, we want to group them such that the users that can mutually benefit from each other similarly are in the same group. This is to ensure that the users in the same group will share the same eagerness for active localization of the other's lost items.

Let the set of users in the network be X=[u_1,u_2,u_3...u_N]. A group of users, G_i, is a subset of X and the set of all groups is denoted by:

$$\mathcal{G} = \{G_1, G_2, \ldots, G_r\},$$

where,

$$\bigcup_{i=1}^{r} G_i = X \text{ and } G_i \cap G_j = \emptyset \ (i \neq j)$$

Assume that there are R possible locations that these N users visit. These locations could be considered as all potential locations that users visit with some boundaries. Moreover, they can also be considered as the locations of the POIs that users visit more frequently or they have a high likelihood to lose their items. Each node visits all or some of these locations with different durations and frequencies.

Assume that the number of groups is | G| = r. The goal is to find the partitioning of N users into r groups such that the sum of average satisfaction values of each group will be maximized. More formally, the objective function is:

$$\tau(\mathcal{G}) = \max \sum_{\forall G_i \in \mathcal{G}} \left(\frac{\sum_{\forall i,j \in G_i} \gamma(i,j)}{|G_i|(|G_i| - 1)} \right) \quad \text{(Equ:4)}$$

This objective function can also be interpreted as maximization of total benefit per all user interactions in the network. As only the users in the same group interact with each other for active localization queries, there exists $|G_i|(|G_i| - 1)$ interactions in group i. Taking the average group benefit by dividing with this value and iterating it through all groups yields us the sum of all average group benefits.

Note that dividing a set of n labeled objects into r different non-empty unlabeled subsets is defined by Stirling numbers of the second kind and can be explicitly calculated as:

$$\left\{ {n \atop r} \right\} = \frac{1}{r!} \sum_{j=0}^{r} (-1)^{r-j} \binom{r}{j} j^n$$

It is possible to generate different number of groups between 1 and n, thus, we actually need to try $\Sigma_{r=1}^{r=n}$ $\left[{n \atop r} \right]$possible cases (which is defined as nth Bell number, B$_n$) to find the best group that maximizes the objective function. As this will take too long, in the next subsections, we propose two different clustering algorithms to find the best group.

3.2.1 Greedy Algorithm

In order to achieve an algorithm that runs fast, we design a clustering algorithm with a greedy heuristic. The steps of this algorithm is summarized in Algorithm 1.

Algorithm 1 Greedy Clustering Algorithm

$continue = true$
while $continue$ **do**
  $maxIncrease = 0$
  $maxGroupIndices = \{0, 0\}$
  $\mathcal{G}_{max} = \emptyset$
  for $\forall i \in \mathcal{G}$ **do**
    for $\forall j \in \mathcal{G}$ *and* $j > i$ **do**
      Create a new group $G_{new} = G_i \cup G_j$
      $\mathcal{G}_{temp} = \mathcal{G}$ - $\{G_i, G_j\} \cup G_{new}$
      if $(\tau(\mathcal{G}_{temp})$ - $\tau(\mathcal{G})) > maxIncrease$ **then**
        $maxIncrease = \tau(\mathcal{G}_{temp})$ - $\tau(\mathcal{G})$
        $maxGroupIndices = \{i, j\}$
        $\mathcal{G}_{max} = \mathcal{G}_{temp}$
      end
    end
  end
  if $maxIncrease > 0$ **then**
    $\mathcal{G} \leftarrow \mathcal{G}_{max}$
  else
    $continue = false$
  end
end

It is assumed that initially there are N different clusters and each cluster consists of a single user. G=[G_1,G_2,G_3...$G_{|N|}$],with G_i=[u$_i$] $\subset$ X. Then, we try every possible merging of two different groups in current set of groups, G, and find the one that will provide with the maximum possible increase in the objection function, $\tau(G)$. The new set of groups is obtained by merging these individual groups and next iteration of the same operation is run. If there is no merging possible that will give a positive increase in the objective function, the algorithm stops.

3.2.2 Evolutionary Genetic Algorithm

For the clustering of the users, we also design an evolutionary genetic algorithm, which can provide near-optimal grouping results with comparably fast running times . Each chromosome consists of N numbers where each number indicates the group of the user up to N. For example, a sample chromosome (1, 1, 2, 3, 3) indicates that users (indexes) 1 and 2 will be in the group 1, user 3 will be in group 2 by itself and users 4 and 5 will be in group 3. The crossover function is achieved through standard single point crossover at random locations of two random chromosomes.

In artificial intelligence, an evolutionary algorithm (EA) is a subset of evolutionary computation,a generic population-based meta heuristic optimization algorithm. An EA uses mechanisms inspired by biological evolution, such as reproduction, mutation, recombination, and selection. Candidate solutions to the optimization problem play the role of individuals in a population, and the fitness function determines the quality of the solutions. Evolution of the population then takes place after the repeated application of the above operators.

Evolution is change in the heritable characteristics of biological populations over successive generations.These characteristics are the expressions of genes that are passed on from parent to offspring during reproduction. Different characteristics tend to exist within any given population as a result of mutation, genetic recombination and other sources of genetic variation.

Step One: Generate the initial population of individuals randomly. (First generation)

Step Two: Evaluate the fitness of each individual in that population .

Step Three: Repeat the following regenerational steps until termination:

(1)Select the best-fit individuals for reproduction. (Parents)

(2)Breed new individuals through crossover and mutation operations to give birth to child.

(3)Evaluate the individual fitness of new individuals.

(4)Replace least-fit population with new individuals.

Evolution occurs when evolutionary processes such as natural selection (including sexual selection) and genetic drift act on this variation, resulting in certain characteristics becoming more common or rare within a population.Evolutionary biologists have continued to study various aspects of evolution by forming and testing hypotheses as well as constructing theories based on evidence from the field or laboratory and on data generated by the methods of mathematical and theoretical biology. Their discoveries have influenced not just the development of biology but numerous other scientific and industrial fields, including agriculture, medicine and computer science.For mutation operation, we update the group number of a random node with another random group number between 1 and N. The fitness function is set to τ(G) defined in objective function equation.

3.3 Performance Analysis

We create a synthetic data set of user visits, then we use two different real user traces from location based social network platforms. Next, we provide the details of these data sets. We generate visits of the each node independently from other nodes. In every 10 minutes, a visit is created for a node with a probability of 0.05 (i.e., one visit per 200 minutes in average). The duration of the visits is assigned uniformly and randomly from the range of 5-15 minutes. Then, we decide the region of the visit. Each node has a home region, which is the region it most frequently visits. The probability of the visit being in the home region is 0.7, while for all other regions, the probability is (0.3/(R-1)). This visit generation process continues for a week time (appr. 10,000 minutes, meaning 50 visits for each node in average).

We first run simulations to see the impact of the average delay of finding lost items of a

Parameter	Value
Time frame	7 days
Number of users	100, 250, 500
Number of POIs	10
Visit duration	5-15 min
Average number of visits per user per region	5
Probability of self finding (p_s)	0.2
Probability of finding by other users (p_o)	0.05
Probability of visits in home region	0.7
Probability of visits in other regions	0.3

user by other nodes. To this end, we let each user lose an item at a random time in randomly selected one of his/her visits at a randomly selected POI he/she visits. Using the probabilities, we then find the user who can detect this item (which could be either other users or him/herself). The results show the average for all nodes in the network generated by synthetic data. As the number of users increase, the average delay decreases as expected. Moreover, due to the dominance of other users, changes in p_s can only slightly affect the results.The comparison of simulation and analysis results for different p_o and p_s values are shown in figure3.2.

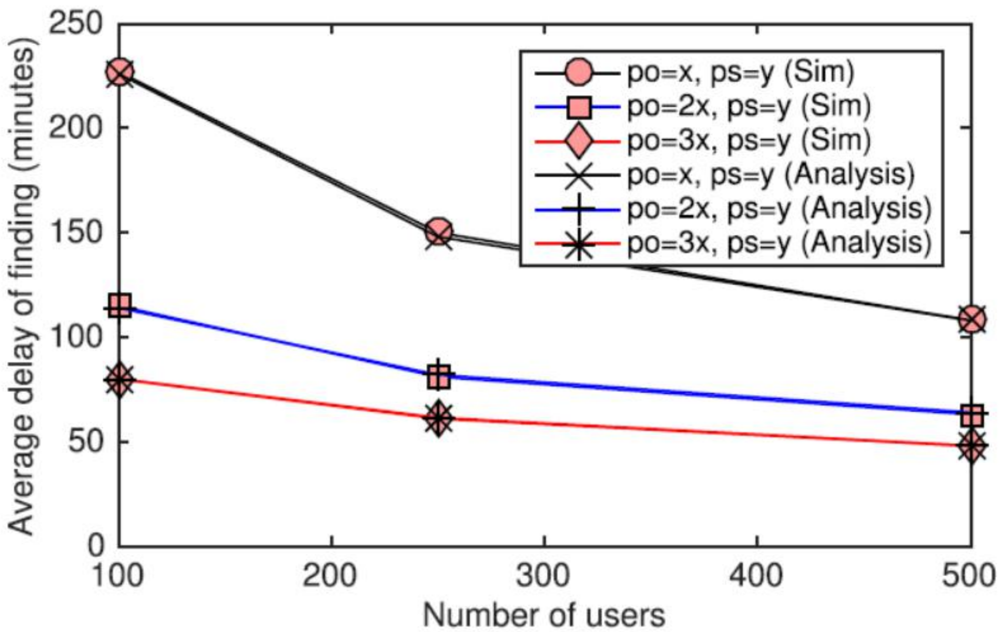

Figure 3.2: Comparison of simulation and analytical results for average delay of finding lost items

As the plots clearly show, the results almost perfectly match. We also obtain results using real datasets, and confirmed the match between simulation and analytical results. The corresponding figures are not shown here for brevity.Next, we present results regarding the performance of proposed clustering algorithms.

The greedy algorithm is designed to stop when the new group merge operation do not provide an increase in the objective function, thus it stops when it reaches a maximum value.

However, to observe the decrease after this maximum point, we let the algorithm run for some time. We observe that at the iteration which is around the half number of users in the dataset, the maximum value for the objective function is obtained. While this could be considered expected due to the nature of the greedy algorithm, how the grouping of users is done will matter and may not have provided such smooth curves. Using the genetic algorithm proposed, we also calculate the objective (i.e., fitness) function value through the generations.

But genetic algorithm can provide similar maximum objective function value as in greedy algorithm.However, to understand the performance differences between these clustering algorithms as the user count increases in the network, we compare their running time and the maximum values they achieve for objective function.While both of these algorithms provide similar value with smaller number of users, we observe that greedy algorithm can achieve better objective function value with larger number of users.On the other hand, we compare the running time of these algorithms , the advantage of greedy algorithm in terms of providing better objective function value at larger user counts is mitigated due to its longer running time.

Finally, using the greedy algorithm, we compare the benefits obtained and reductions in user interactions with the proposed grouping model.We plot the percentage of benefits obtained with proposed grouping idea compared to max benefits that could be obtained when all users are in a single group. While the grouping strategy decreases the overall benefit for the users a bit, it decreases the interactions between users a lot.

Privacy is the ability of an individual or group to seclude themselves, or information about themselves, and thereby express themselves selectively. The boundaries and content of what is considered private differ among cultures and individuals, but share common themes. When something is private to a person, it usually means that something is inherently special or sensitive to them.The right not to be subjected to unsanctioned invasion of privacy by the government, corporations or individuals is part of many countries' privacy laws, and in some cases, constitutions.

In the figure3.3 , we show the percentage of the reduction of these interactions in proposed grouping model compared to the single group model. As the figure shows, there is a huge benefit with introduced grouping model in preserving the privacy of users.

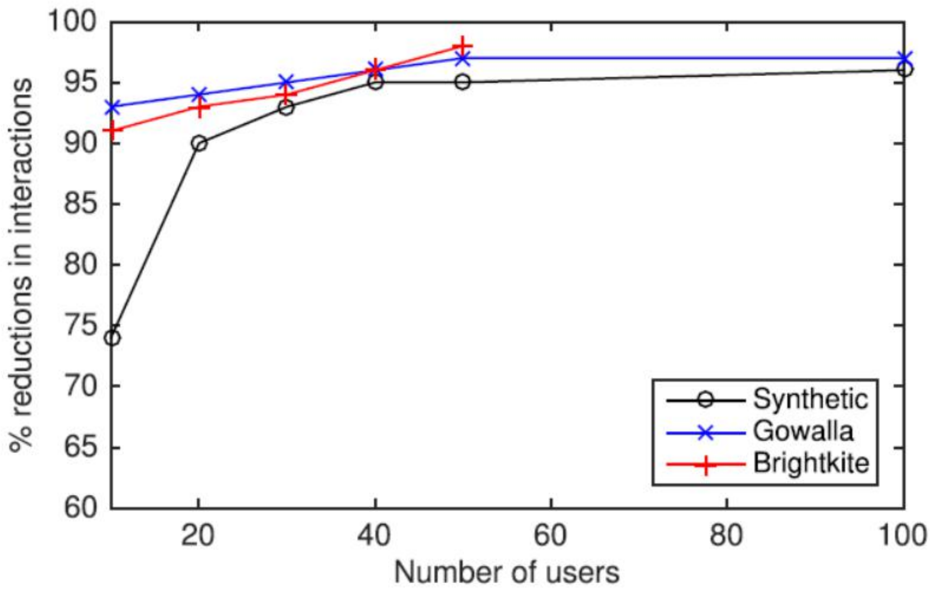

Figure 3.3: Percentage of reductions obtained with proposed greedy clustering of users compared to the interactions between all users in single grouping

Limited access refers to a person's ability to participate in society without having other individuals and organizations collect information about them. Various types of personal information are often associated with privacy concerns. Information plays an important role in the decision-action process, which can lead to problems in terms of privacy and availability. First, it allows people to see all the options and alternatives available. Secondly, it allows people to choose which of the options would be best for a certain situation. An information landscape consists of the information, its location in the so-called network, as well as its availability, awareness, and usability.

Yet the set-up of the information landscape means that information that is available in one place may not be available somewhere else. This can lead to a privacy situation that leads to questions regarding which people have the power to access and use certain information, who should have that power, and what provisions govern it. For various reasons, individuals may object to personal information such as their religion, sexual orientation, political affiliations, or personal activities being revealed, perhaps to avoid discrimination, personal embarrassment, or damage to their professional reputations.

CONCLUSION

To approach the problem of lost object tracking with a clustered crowd GPS service through recently trending spatial crowdsourcing context, an idea of clustering of users in the Beacon tag network that can provide most of the total possible benefits to each other in terms of the localization of their lost items was proposed. To this end, it was analyzed that the visit patterns of users at the same location and present a new metric called Social Tracking Distance (STD) that quantifies the benefit of users to each other in terms of the capability of finding each other's lost objects. Once the potential benefit of each user to every other user is calculated, then divide all users into clusters such that the users in the same group provide high benefit to each other and the privacy of users is valued at its maximum. For clustering there is a greedy algorithm and a genetic algorithm. In simulations, the performance of the proposed clustered crowd GPS system for active localization is evaluated. The results show that with the grouping strategy proposed, the user interactions drop drastically without sacrificing from the maximum possible benefits. A comparison of the performance of the clustering algorithms used was made and it shows that genetic algorithm can provide faster calculation time compared to greedy approach when the number of nodes in the network increases. However, the greedy algorithm provides better results in terms of achieving maximum total benefit per shared location privacy. As a future work,it would be suitable to enhance the proposed metric with the integration of mobility prediction algorithms and network community structure, for a better accuracy in understanding the future benefits of users to each other.

REFERENCES

[1] Pixie. (Mar. 2017). The Nation's Biggest Lost and Found Survey. [Online]. Available: https://getpixie.com/blogs/news/lostfoundsurvey

[2] N. Guy. (Oct. 2017). The Best GPS Trackers for Cats and Dogs. [Online]. Available: https://thewirecutter.com/reviews/best-gps-pet-trackers/

[3] R. Edwards. Best Wearable GPS Trackers for Kids Buyers Guide. Accessed: Apr. 2018. [Online]. Available: https://www.safewise. com/resources/wearable-gps-tracking-devices-for-kids-guide

[4] K. Bishop. (2017). Growth Business: GPS Tracking The Elderly. [Online]. Available: https://www.cnbc.com/2014/03/11/growth-businessgps- tracking-the-elderly.html

[5] C. Gomez, J. Oller, and J. Paradells, "Overview and evaluation of bluetooth low energy: An emerging low-power wireless technology," Sensors, vol. 12, no. 9, pp. 1173411753, 2012.

[6] thetileapp.com. (2016). Tile Bluetooth Tracker Device. [Online].Available: https://www.thetileapp.com

[7] TraxFamily.com. (2016). A Really Small and Light GPS Tracker for Kids and Pets. [Online]. Available: https://traxfamily.com

[8] S. Cohen. (2015). Locus Pocus Can Track all Your Bluetooth DevicesAnd Other People's, Too, VentureBeat. [Online]. Available: http://venturebeat.com/2015/06/06/locus-pocus-can-trackall-yourbluetooth- devices-and-other-peoples-too

[9] J. Zhang, F. Tang, L. Barolli, Y. Yang, andW. Xu, "Information gain based maximum task matching in spatial crowdsourcing," in Proc. IEEE 31st Int. Conf. Adv. Inf. Netw. Appl. (AINA), Mar. 2017, pp. 886893.

[10] L. Kazemi and C. Shahabi, "GeoCrowd: Enabling query answering with spatial crowdsourcing," in Proc. 20th Int. Conf. Adv. Geographic Inf. Syst., 2012, pp. 189198.