# Sense-And-Trace: A Privacy Preserving Distributed Geolocation Tracking System

Eyüp S. Canlar[1,2], Mauro Conti[3], Bruno Crispo[4], and Roberto Di Pietro[5]

[1] Sapienza University of Rome, Italy - canlar@di.uniroma1.it
[2] VU University Amsterdam, The Netherlands - escanlar@cs.vu.nl
[3] University of Padua, Italy - conti@math.unipd.it
[4] University of Trento, Italy - crispo@disi.unitn.it
[5] Roma Tre University of Rome, Italy - dipietro@mat.uniroma3.it

**Abstract.** The capabilities of modern smartphones pave the way for a new collaborative usage of this technology. Several researchers already envisaged to use this technology for distributed sensing purposes. In particular, one of these purposes focuses on tracing devices (people) movement. Current solutions for distributed tracing (either based on information provided by the mobile nodes, or collected by the surrounding network) have some limitations: e.g. accuracy, privacy, cost of deployment, and cost of operation.
The aim of this paper is to highlight the open problems of distributed geolocation tracing and to propose a solution for some of the current problems. In particular, we propose Sense-And-Trace (SAT), which is a system that makes use of collaborative sensing to collect information about other mobile nodes with the final aim of tracking potential target nodes. In SAT, information is collected in a way such that the privacy of nodes that voluntarily collaborate is preserved, and the information of the mobility of a node is disclosed only to the authorized entity (e.g. a law enforcement agency with the appropriate permission). Our solution can be seen as an enhancement of the classical "neighborhood watching" concept, with fine-grained mobility information automatically-collected through the devices carried by humans.

## 1 Introduction

Contemporary smartphones are equipped with several sensors, such as: GPS, accelerometer, compass, microphone, camera, and proximity meter. These sensing capabilities, together with the fact that smartphones are currently widely distributed, pave the way for a new usage of this technology: for distributed sensing purposes, as already been envisaged by several researchers [2, 8, 20, 14, 6].

One of the usage of this distributed sensing capability is in tracking the movement of the devices in a geographical area (e.g. a metropolitan area), with the final aim of tracking people carrying those devices. From a high level point of view, current mobile device tracking systems periodically update an authorized entity about the current position of a mobile device. In this context, we can identify two types of mobile device tracking system, namely: i) *handset-based*, and ii) *network-based* mobile device tracking systems [23, 13]. The handset-based mobile device tracking systems relies on the geographic location (geolocation) information provided by the mobile device which is the "target" of the tracing. Network-based mobile device tracking systems compute the position of the target based on its distance from network infrastructure elements (e.g. base stations, access points, beacon nodes), which are not mobile. The most well known example of a network-based tracking system is the system used by cellular phone network operators, namely cell tower triangulation. The network operator knows the exact position of its cell towers. The network operator computes the relative position of the target mobile device using its distance to three or more cell towers.

**Motivation and Open Problems.** Both handset-based and network based solutions have limitations. In fact, in handset-based solutions, the target might not always be able (or willing) to provide the sensed information. Also, the target might provide the system with false information. As for network-based solutions, they are not very accurate (e.g. triangulation error ranges from $200m$ to $1000m$), and also they need a significant deployed infrastructure and/or cooperation from different entities (e.g. telco operators). Furthermore, in some specific circumstances (e.g. rural area, natural disaster, or terrorist attack) there might not be any deployed infrastructure available.

Researchers also tried to design hybrid solutions [16, 19, 11] to increase the quality of network-based solutions: the coarse information gained from the network infrastructure is refined with information provided by the target device itself. However, hybrid solutions also combine the disadvantages of both types of systems. Hence, there is still need for a tracking system that (all at once):

- it is distributed (does not rely on availability/information provided by the network infrastructure);
- it does not rely on the information provided by the target mobile device;
- it is cheap and easy to deploy and maintain;
- it preserves the privacy of the devices (users) that voluntarily participate as *witnesses*. In particular, the ideal system should: i) disclose information only to the authorized party, and ii) only about the user for which the authority requested information.

**Contribution.** In this paper, we discuss the open problems in distributed tracing solutions. Then, we address some of these problems proposing Sense-And-Trace (SAT): a system that makes use of collaborative sensing to track mobile devices within a geographic area, and for a given period of time. The tasks we consider in our collaborative sensing systems are: i) data acquisition, and ii) querying.

Our solution is inspired by the "neighborhood watching" concept. In fact, in the SAT system, neighbor nodes collaborate to achieve a common goal, which is to track the movements of a potential target mobile device. Indeed, SAT just improves the accuracy of traditional “neighborhood watching” by adding the capability of having fine-grained and automatically-collected information. Hence, the security protocol become “embedded” in the life of humans carrying mobile wireless devices.

**Outline.** The remainder of this paper is outlined as follows. Section 2 reports the related work in this area. Section 3 presents the system model we consider and the assumptions we make. Consequently, in Section 4 we introduce our proposed solution: Sense-And-Trace (SAT). Section 5 provides a discussion about the properties of our proposed solution. Finally, in Section 6 we present our conclusions.

## 2 Related Work

There are two categories in the work related to location tracking of mobile nodes. The first category introduces systems that track mobile nodes within a closed environment, like in a building (indoor tracking) [12, 19]. The second one deals with geolocation tracking of mobile nodes outdoors [16, 24, 3, 11, 20, 4].

Let us start with discussing the related works about indoor target tracking. In [12], the authors propose the “*MoteTrack*” system which is a decentralized approach to radio frequency based indoor location tracking. The MoteTrack system computes the location of a mobile node based on the Received Signal Strength Signature Indication (RSSI) from a network of beacon nodes spread throughout a building. In contrast to other known solutions for indoor tracking, the MoteTrack system is decentralized and does not rely on any back-end server. However, this system requires additional infrastructure (network of beacon nodes) to work properly. Furthermore, preserving the privacy of the traced mobile nodes is completely out of the scope of MoteTrack.

In another work [19], the authors propose “*FindingMiMo*” which is also a decentralized indoor mobile node tracking system. In contrast to the MoteTrack system, FindingMiMo does not require the availability of beacon nodes to track mobile nodes. Instead, the FindingMiMo system makes use of the available WiFi access points to track a lost mobile node. More specifically, in FindingMiMo the target mobile device logs WiFi channel conditions. When the user looses her device, she tries to find it back using another “chaser” device. The chaser compares the logs it receives from the lost device with the WiFi channel conditions it is sensing. In this way, the chaser backtracks the path that the lost device has followed to the place it is currently located at. However, the FindingMiMo system still relies on the availability of existing infrastructure. Furthermore, also this system does not take privacy into consideration.

Other research focused on location tracking of mobile nodes in an outdoor environment. In [16], the authors propose “*Adeona*”, which is a privacy preserving decentralized system to locate and trace stolen or lost devices. Adeona consists of a client-side and a distributed storage. The client-side periodically collects location information. Then, the client-side performs some cryptographic operations on the collected location information, to make it anonymous and unlinkable. The client-side uploads the location updates to the distributed storage on pseudo-randomly determined times to overcome potential timing attacks. In contrast, the distributed storage is based on an open source distributed storage system OpenDHT [15], whose nodes run on PlanetLab [5]. Only the owner of the lost or stolen system can decode the location updates from the distributed storage. As said, this system relies on the cooperation of the mobile devices it is tracking. However, in the case that a thief applies counter measures (i.e. power off device immediately, erase and reinstall software at home), then the Adeona system fails to track the target device. Moreover, as of 1 July 2009 OpenDHT was taken down from PlanetLab. So, the Adeona system is not operational anymore.

Another outdoor approach [3] proposes to accurately detect the location of a smartphone combining data from two sensors. The first sensor is the camera integrated in the phone. The other sensor is the GSM modem of the smartphone. More specifically, the first step in this approach is computing the geolocation. This system computes the location by querying (i.e by sending an AT+CSQ [10] request) the GSM modem for the RSSI. In the second step, the smartphone sends a set of pictures of its surroundings to a server. The server uses Content Based Image Retrieval (CBIR) to analyze and match the pictures with known landmarks. Then, the server computes the more accurate geolocation of the smartphone by using a *time-forwarding algorithm.* In the last step, the server sends to the smartphone the geolocation as computed in the previous step. Unfortunately, also this work did not consider privacy. In fact, anybody that can gain access to the server is able to track any smartphone involved in this system.

The contributions exposed so far rely on the cooperation of the tracked device. As well as on the fact that the tracked device provides the tracking system trustworthy information. However, in [24, 25], the authors make the observation that malicious smartphone users can make their devices send falsified geolocation information to get illegitimate access to resources or provide bogus alibis. To overcome this problem, the authors propose "*APPLAUS*" which is a privacy preserving location proof updating system. In APPLAUS, Bluetooth enabled mobile (neighboring) devices generate location proofs, and it updates this to an untrusted server. An authorized verifier can query and receive location proofs from the server. The APPLAUS system makes use of statistically changing pseudonyms to guarantee the location privacy from every party. A drawback of the APPLAUS system is that it makes use of a centralized component, namely the untrusted server. When this untrusted server is unavailable due to DoS attacks, or hardware malfunctions, the system is useless.

Everything we discussed so far was related on how we can track a mobile node in either a centralized or decentralized approach. For the decentralized options it is also interesting to see how many sensor nodes we need, to track a mobile node in a given geographic area. To find an answer to this question, the authors of [9] conducted a simulation study. More specifically, in this study the authors assume a malnet (i.e. a malicious network consisting of smartphones, routers, and other

WiFi enabled devices). The nodes of this malnet cooperate to track a specific mobile node. The results of this study show that a small number (some 10%) of mobile devices can track the majority of users, during a significant fraction of their travel. Based on their research, the authors draw the following conclusions: i) in the current situation, voluntary networks with perceived benefits can probably achieve the usage rate necessary to track individual movements, and ii) the ubiquitous deployment of 802.11n in smartphones would make it possible for a malnet to track the geolocation of a specific mobile node.

Finally, we want to note that none of the above systems succeed in providing a robust distributed privacy-preserving trustworthy geolocation tracking system, which is the target of our research.

## 3 System Model and Assumptions

The main components of the SAT system model are: i) the *nodes*, ii) a *Distributed Data Store (DDS)*, and iii) a *Data Collector (DC)*.

The nodes are personally-owned mobile devices (e.g. smartphones, or tablets). These nodes are distributed over a geographic area (e.g. metropolitan area), and they move within this geographic area as their owners travel. They communicate using their short range radios (e.g. Bluetooth or WiFi). Furthermore, they can pinpoint their own geolocation (e.g. using GPS, AGPS, or WiFi Location Service) using their embedded hardware.

In the SAT system model, we have two types of nodes: i) *sensor nodes*, and ii) *target nodes.* The sensor nodes voluntarily participate in SAT i.e. they run an application to scan their environment, and log that other wireless devices are in their communication range. The sensor nodes store their logs on the DDS using any data packet service. That is by using a mutually authenticated TCP connection either via a WiFi access point that the sensor node has access to, or by using the cellular data service (e.g. UMTS or HSPDA).

In contrast, target nodes do not have to voluntarily participate in the SAT system. This is not surprising considering that target nodes are the objects that the SAT system wants to track. To clarify this consider the following: an individual involved in criminal activities does not want to be tracked, because its geolocation might incriminate its involvement in a criminal event. Even in this case the SAT system is able to track the target nodes, because it does not rely on any (geolocation) information provided by the target nodes. A final important note is that in this system model a node can be, at the same time, a sensor node and a target node.

In this system model, the DDS is a distributed database. More specifically, parts of the database (containing the geolocations of target nodes) are stored on multiple computers within a network. These multiple computers can reside at the same or at different geolocations. In any case, the DDS falls under the administrative domain of one authorized organization.

Furthermore, in this model the DC is the only authorized entity that can track the movements of target nodes. In addition, the communication between a DC and its DDS runs over a secure communication channel (e.g. private local area network, VPN, or SSL/TLS). We also have to note that the entity type of the DC depends on the use case scenario of the SAT system. For example, if we use the SAT system for surveillance purposes (e.g. tracking the movements of suspects), then the DC is a *Law Enforcement Agency (LEA)*. Another use case scenario is that of tracking lost or stolen devices. In this case, the DC would be the legitimate owner.

In Figure 1, we illustrate an example SAT system, where $W_0, W_1, \ldots, W_5$ are sensor nodes, and $S$ is the target node. The arrows between several components illustrates the communication links, and the messages (i.e. GLMs) that they exchange over these links.

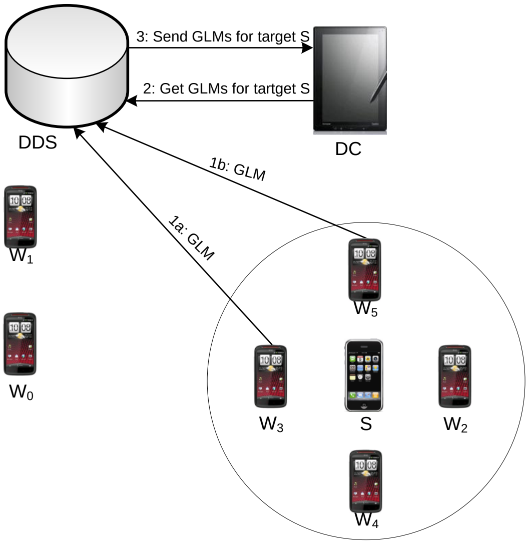

**Fig. 1.** Overview of our solution.

We also make the following explicit assumptions for the SAT system model. Similar to [9], we assume that both type of nodes (sensor and target nodes) have their short range radios enabled continuously. Moreover, as in [17, 24, 25], we assume a *Pseudonymous Public Key Infrastructure (PPKI)*. In which, the trusted *Certificate Authority (CA)* acts as an authentication and authorization service for the sensor nodes, DDS, and DC. For this reason, the sensor nodes, DDS, and DC have to register with the CA. As a result of the registration, the DDS and DC receive a public/private key pair. On the other hand, during the registration phase the CA preloads each of the $N$ sensor nodes $W_i$ (with $0 \leq i \leq N-1$) with $M$ pseudonyms $\{P_{i,k}\}_{k=1}^{M}$. For each one of these pseudonyms, the CA also preloads a public/private key pair $<K_{P_{i,k}}, K_{P_{i,k}}^{-1}>$ (with $1 \leq k \leq M$), and an associated public key certificate $sign_{CA}(K_{P_{i,k}})$ signed with the private key of the CA. We also have to note that the CA is the only entity that can link the real identity (e.g. BSSID or BD_ADDR) of a witness $W_i$ to one of its $M$ pseudonyms $\{P_{i,k}\}_{k=1}^{M}$. Finally, we assume that SAT replaces the source addresses, required by the underlying communication protocol, with the pseudonyms [22, 17]. This is done as a counter measure for the localization attack as described in [22].

Below we provide an overview of all notations used in this paper in Table 1.

| Notation | Description |
|---|---|
| $W_i$ | A sensor node |
| $S_j$ | A target node |
| $\mathcal{W}$ | The set of all sensor nodes |
| $\mathcal{S}$ | The set of all target nodes |
| $N$ | The number of sensor nodes (i.e. $\lvert\mathcal{N}\rvert = N$) |
| $x\Vert y$ | $x$ concatenated to $y$ |
| $A \xrightarrow{(m)} B$ | An entity $A$ sends a message m to another entity $B$ |
| $L_n$ | The geolocation of a node $n$ |
| $I_n$ | The identity of a node $n$ |
| $P_{W_i,k}$ | kth pseudonym of $W_i$ |
| $\{P_{W_i,k}\}_{k=1}^{M}$ | The set of $M$ pseudonyms associated with $W_i$ |
| $<K_{P_{i,k}}, K_{P_{i,k}}^{-1}>$ | The public/private key pair of an entity $A$ |
| $E_{K_A}(m)$ | Encrypt message $m$ with public key $K_A$ of an entity $A$ |
| $D_{K_A^{-1}}(m)$ | Decrypt message $m$ with private key $K_A^{-1}$ of an entity $A$ |
| $Sign_A(m)$ | Sign the message $m$ with private key $K_A^{-1}$ of an entity $A$ |
| $Verify_A(m)$ | Verify the signature of the message $m$ with public key $K_A^{-1}$ of an entity $A$ |
| $G_{GLM_{T_x}}$ | A group of at time $T_x$ concurrently created location logs |

**Table 1.** Notation used in this paper

## 4 Sense-And-Trace

In this section, we introduce our solution: *Sense-And-Trace*. We start with giving a brief overview of our system (Section 4.1). Then, we provide the algorithmic description of our solution (Section 4.2).

### 4.1 Overview

To describe the overview of our system, let us assume a real use case: a surveillance scenario. In this scenario, the DC is a LEA, and it manages and operates a DDS. This LEA deploys a SAT application to collaboratively track any suspect $S_j$ within a given geographic area $GA$. The LEA distributes this SAT application via application markets, like Google Android Market, Apple App Store, etc. Individuals voluntarily download and use this application to help the LEA in tracking the movements of suspects. This collaborative behavior might be encouraged via incentives (e.g. the users that aids in capturing a criminal gets a financial reward, or gets a tax reduction). We assume that $N$ individuals in a given geographic area $GA$ downloaded the application. In other words, this means we have a collaborative sensor network with $N$ mobile sensor nodes (e.g. witnesses $W_0, W_1, \ldots, W_{N-1}$) in the given geographic area $GA$.

Now we have set the use case scenario, we continue with providing a brief overview of how SAT works on the witness side. First of all, a witness $W_i$ needs to start the SAT application. After this, the SAT application periodically senses its environment for any $S_j$. In other words, each $W_i$ checks whether $S_j$ is within its communication range. When a $S_j$ is indeed in the communication range of a $W_i$, then this $W_i$ logs the presence of $S_j$ in a GLM and stores it on the DDS. We provide the exact details of the GLM in Section 4.2.

In contrast, the LEA has to do the following to track the movements of any $S_j$. First of all, the LEA needs to query its DDS for all GLMs associated with a specific suspect $S_j$. After receiving this query, the DDS responds by sending all the GLMs associated with this $S_j$ to the LEA. The contents of these GLMs provide the LEA a history of all the geolocations visited by $S_j$. So, using the contents of the GLMs, the LEA reconstructs the movements of $S_j$.

To clarify the above, let us consider a concrete example. More specifically, the situation is as illustrated in Figure 1. So, in the specific geographic area $GA$, we have six witnesses, $W_0, W_1 \ldots W_5$, and one suspect $S$. These six witnesses cooperate with the LEA to track the movements of $S$. As illustrated in Figure 1, $S$ is in the communication range of $W_2, W_3, W_4$, and $W_5$. Consequently, $W_0$ and $W_1$ are not in the communication range of $S$. Equally important to note is that $W_2$ and $W_4$ are in the communication range of $S$. However, they do not collaborate with the LEA to track $S$ because they have not started their SAT application.

In the initial situation we sketched above, only $W_3$ and $W_5$ sense and log the presence of $S$. Then, both witnesses ($W_3$ and $W_5$) create their respective GLMs. As a final action, they upload the GLMs on to the DDS (Arrows 1a and 1b in Figure 1).

As the suspect $S$ moves within $GA$, it moves out the communication range of some witnesses, and into the communication range of other witnesses. This movement of $S$ caused a new situation (illustrated in Figure 2). In this current situation, $W_0$ and $W_1$ are in the communication range of $S$. In a similar way as in the initial situation, both of these witnesses ($W_0$ and $W_1$) create and upload the GLMs (Arrows 1a and 1b in Figure 2) to log the presence of $S$.

Finally, when the LEA queries the DDS (Arrow 3 in Figure 2), it will receive four GLMs (Arrow 4 in Figure 2). After analyzing the contents of these GLMs, the LEA observes that $S$ was first in the vicinity of $W_3$ and $W_5$. The LEA also observes that $S$ moved from this initial position to a geolocation in the vicinity of $W_0$ and $W_1$.

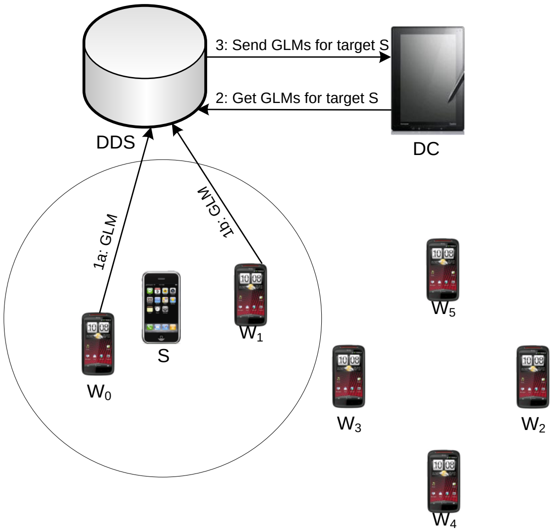

**Fig. 2.** The suspect $S$ moved to vicinity of $W_0$ and $W_1$.

## 4.2 Protocols

The SAT system uses two main protocols, namely the "*Geolocation Logging Protocol (GLP)*" and the "*Geolocation Tracing Protocol (GTP)*". A witness $W_i$ running the SAT application makes use of the GLP to log the presence of suspects $S$. Whereas, the DC runs the GTP to reconstruct the movements of a suspect $S$ within the specified geographic area $GA$. Below we provide the details of both protocols.

**Geolocation Logging Protocol (GLP)**
Let us start the discussion of the protocols with the GLP (detailed in Algorithm 1). This is the protocol that the SAT application runs on behalf of its witness $W_i$ to log the presence of potential suspects. After a witness starts the SAT application, the GLP periodically (i.e. every *timeOut* period of time) checks whether there are suspects in its communication range (line 1–4). For the sake of this discussion, let us assume a specific suspect $S_j$. When $S_j$ is within the communication range of a $W_i$, then the GLP logs the presence of $S_j$ in an encrypted log denoted as "*Location Update (LU)*" (line 6). This LU contains the following pieces of information: i) the current (i.e. kth) pseudonym $P_{i,k}$ of a witness $W_i$; ii) the identity of the suspect $I_{S_j}$; iii) the current geolocation of this witness $L_{W_i}$; iv) the Received Signal Strength Indication $RSSI$; and v) the timestamp $TS$. Furthermore, these pieces of information are encrypted with the DC's public key $K_{P_{DC}}$.

This protocol uses the LU to construct the GLM message as follows. First, the GLP creates a message by concatenating the necessary pieces of information together, namely: i) this witness' current (i.e. kth) pseudonym $P_{i,k}$, ii) the identity of the sensed suspect $I_{S_j}$, iii) the timestamp $TS$, and iv) the location update $LU$. Then, the GLP signs the concatenated message with the kth private key of the witness $K^{-1}_{P_{i,k}}$ (line 7). Finally, this protocol uploads the GLM to the DDS (line 8).

**Algorithm 1** Witnesses sensing for potential suspects within their communication range

```
1: $timeOut = \mathit{\Delta}$
2: while $true$ do
3:   if $\mathit{\Delta}$ seconds passed then
4:     populate $\mathcal{S}$ with all wireless devices in communication range
5:     for each item in $\mathcal{S}$ do
6:       $LU = E_{K_{DC}}(P_{i,k}||I_{S_j}||L_{W_i}||RSSI||TS)$
7:       $GLM = sign_{W_i}(P_{i,k}||I_{S_j}||TS||LU)$
8:       $W_i \xrightarrow{GLM} DDS$
9:     end for
10:  end if
11: end while
```

### Geolocation Tracing Protocol (GTP)

The DC uses the GTP (as detailed in Algorithm 2) to trace the movements of a specific target node $S_j$ within a geographic area for a specified period of time. The DC provides the GTP with the required parameters, namely: i) the identity of a target node $I_{S_j}$; ii) the starting date and time for the trace $T_{start}$; and iii) the ending date and time for the trace $T_{end}$. GTP starts by using these parameters to construct and send a request to the DDS (line 1). On its turn, the DDS returns all GLMs associated with $S_j$ with $T_{start} \leq TS \leq T_{end}$ (line 2). For each returned GLM, this protocol does the following. First, it verifies the signature of a GLM (line 4). Then, assuming that the signature is correct, the GTP decrypts the encrypted part (i.e. $LU$) of the GLM and adds it to the set of all decrypted GLMs $\{Dec\_GLMs\}$ (line 8 and 9). In contrast, if the signature is not correct for a specific GLM, then GTP discards it. In what follows, the GTP sorts $\{Dec\_GLMs\}$ in a chronologically ascending manner (line 13). After this, the GTP uses a time interval $\mathit{\Delta}$ to group several GLMs together. For each one of these GLM groups $G_{GLM'_{T_{x+\mathit{\Delta}}}}$, the GTP does the following. First, it computes the estimated geolocation of $S_j$ (line 15). In practice, this can be done using one of the already proposed positioning algorithms (e.g. triangulation, trilateration, multi-lateration) as described in [1]. Then the GTP updates the time line that illustrates the estimated geolocation of the target node at a specific time $T_x$ with $T_{start} \leq T_x \leq T_{end}$ (line 16).

**Algorithm 2** The DC queries its DDS for the geolocation data of a specific suspect

```
INPUT: $ID_{S_j}$, $T_{start}$, and $T_{end}$
OUTPUT: $visitedGeolocationsTimeLine$
$DC \xrightarrow{I_{S_j}||T_{start}||T_{end}} DDS$
$DDS \xrightarrow{\{GLMs\}} DC$

for each item $GLM_l$ in $\{GLMs\}$ do
    $signOK = Verify_{W_i}(GLM_l)$
    if $!signOK$ then
        discard $GLM_l$
    else
        $GLM^{'} = P_{i,k}||I_{S_j}||TS||D_{K_{W_i}}(LU)$
        add $GLM^{'}$ to $\{Dec\_GLMs\}$
    end if
end for

interval $= \Delta$
$\{Dec\_GLMs\} = sort(\{Dec\_GLMs\})$
for each group $G_{GLM'_{T_{x+\Delta}}}$ of concurrent $GLM^{'}$ do
    $computeEstimatedSuspectLocation(G_{GLM'_{T_{x+\Delta}}})$
    $update(visitedGeolocationsTimeLine)$
end for

$return\ visitedGeolocationsTimeLine$
```

## 5 Discussion

In this section, we analyze our proposed solution by discussing the following points: i) feasibility; ii) communication overhead, iii) deployability and costs; and iv) privacy preservation.

**Feasibility.** The feasibility of this system depends on the community participation rate. As a matter of fact, researchers showed that a WiFi based community sensing system needs to achieve a community participation rate of 10% to track the movements of any individual within a metropolitan area [9].

To achieve the necessary community participation rate, our proposed system can either deploy a reward based scheme, or solely depend on voluntary participation. A reward based scheme could be as simple as giving a fixed price for the provided sensor data. Another option could be that the government could grant tax reductions to individuals that provide sensor data, which are of importance for the government. However, this topic—deserving a line of investigations on its own—is out of the scope of this paper.

Another approach is that individuals voluntarily participate to the systems to achieve a common goal. This approach has been proved to be successful in several other initiatives [18, 21]. Considering this, it is not hard to see that community sensing systems that are used for the common good (e.g. security, safety) of a community would easily achieve the necessary participation rate.

Up to this point, we discussed feasibility in terms of community participation rate. Now, we discuss technical feasibility in terms of: the short range radio technologies, and their characteristics (e.g. communication range, device discovery).

In contemporary mobile devices there are two ubiquitously available short range radio technologies, namely Bluetooth and WiFi. Our proposed solution can use either one of these technologies.

The communication range of these technologies has the biggest influence on the performance of our proposed solutions. In fact, having a bigger communication range decreases the necessary community participation rate [9]. Bluetooth has a communication range of 10 meters as opposed to the 100 meters of WiFi. This implies that Bluetooth needs a higher community participation rate (i.e. $>>$ 10%) as opposed to WiFi (i.e. 10%).

The other characteristic that we want to discuss is device discovery. In Bluetooth, during the inquiry phase a master device broadcasts inquiry packets to discover neighboring devices. This inquiry phase lasts on average 5 seconds [7]. In contrast, the WiFi passive scan procedure lasts on average of $50ms$ for each channel [7]. For 11 channels, this means that the WiFi passive device discovery phase lasts for $550ms$. This implies that WiFi is suitable to track fast moving targets (e.g. cars), and Bluetooth is more suitable to track static or relatively slow moving targets (e.g. pedestrians, cyclists).

**Communication Overhead.** The readers might have noticed that the protocols as described in Section 4 suffers from communication overhead. More specifically, the $GLP$ makes witnesses to create and send $GLM$s for every wireless device within its communication range. As a consequence SAT can potentially flood the network. To overcome this problem, the SAT system should reduce the volume of $GLM$s send to the $DDS$. One solution would be to use a probabilistic mechanism in the $GLP$ to reduce the amount of generated traffic. However, this solution has the potential risk of not logging a wireless device that the $DC$ might be interested in tracking. A solution that does not have this problem and reduces the volume of $GLM$s send to the $DC$ is: a system in which the witnesses only track target nodes that the $DC$ is interested in. However, providing the details of such a system is out of the scope of this paper.

**Deployability and Costs.** Unlike some other systems, we do not need to install additional hardware to trace the movements of a target mobile device. Our proposed system solely depends on the information provided by the mobile devices owned by the voluntary participating individuals. This means that for our system to work, we do not have to explicitly deploy anything. In addition, due to the fact that everything we need is personally owned by the voluntarily participating individuals, there are also no high costs involved in deploying and using this system. This makes our proposed system easily deployable and not expensive.

**Privacy Preservation.** As a final point we discuss how effective the SAT system is in preserving the privacy of the witnesses. To do this, we first have to clarify what we consider being possible for the adversary. For the sake of this discussion, we assume a global passive eavesdropper. This means that the attacker, possibly colluding with malicious nodes, is capable to overhear all the messages exchanged within our system. However, the attacker can not modify and/or inject messages.

First of all, let us look at what information is transferred in clear text when transferring a GLM to a DDS. The unencrypted part of the GLM contains the current pseudonym of a witness, the identity of the target mobile device, and the timestamp of when the target node was observed. From these pieces of information, the eavesdropper can not learn the real identity of the witness. This because the pseudonym is a temporary identifier that changes over time. Only the CA is able to associate a given pseudonym to a real identity. However, the CA only reveals this association to an authorized third party. The timestamp and the identity only reveals when a specific device was observed. So, it does not reveal where this specific device was observed. This shows that our system preserves the identity and location privacy of the witness mobile devices. Furthermore, our system also preserves the location privacy of the target mobile device from the global passive eavesdropper.

## 6 Conclusion

In this paper, we presented Sense-And-Trace (SAT): a privacy preserving distributed geolocation tracking system. In contrast to other proposed solutions, SAT does not rely on: i) proprietary infrastructure, ii) on the availability of existing infrastructure, or iii) on geolocation information provided by the target device. The SAT system tracks the movements of a target mobile device using geolocation information provided by (voluntarily participating) neighboring mobile devices (witnesses). Furthermore, the SAT system preserves the privacy of the witnesses, and only discloses information about a specific target mobile device to an authorized third party.

In the future, we want to to perform a simulation study to evaluate the SAT system in terms of: i) effectiveness, ii) power consumption overhead, and iii) privacy preservation under several attack scenarios. Based on the results of the simulation study, we want to develop a forensic suspect tracking system based on the idea presented in this paper.

## References

1. Saif Al-Kuwari and Stephen D. Wolthusen. A survey of forensic localization and tracking mechanisms in short-range and cellular networks. In *Digital Forensics and*

*Cyber Crime*, volume 31 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 19–32. Springer Berlin Heidelberg, 2010.

2. Saurabh Amin, Steve Andrews, Saneesh Apte, Jed Arnold, Jeff Ban, Marika Benko, Re M. Bayen, Benson Chiou, Christian Claudel, Coralie Claudel, Tia Dodson, Osama Elhamshary, Chris Flens-batina, Marco Gruteser, Juan carlos Herrera, Ryan Herring, Baik Hoh, Quinn Jacobson, Toch Iwuchukwu, James Lew, Xavier Litrico, Lori Luddington, Jd Margulici, Ali Mortazavi, Xiaohong Pan, Tarek Rabbani, Tim Racine, Erica Sherlock-thomas, Dave Sutter, and Andrew Tinka. Mobile century using gps mobile phones as traffic sensors: A field experiment. In *Proceedings of the 15th World Congress on Intelligent Transportation Systems*, 2008.
3. Marco Anisetti, Claudio Ardagna, Valerio Bellandi, Ernesto Damiani, Mario Dller, Florian Stegmaier, Tilmann Rabl, Harald Kosch, and Lionel Brunie. Landmark-assisted location and tracking in outdoor mobile network. *Multimedia Tools and Applications*, pages 1–23, 2011.
4. James Biagioni, Tomas Gerlich, Timothy Merrifield, and Jakob Eriksson. Easytracker: automatic transit tracking, mapping, and arrival time prediction using smartphones. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, SenSys '11, pages 68–81. ACM, 2011.
5. Brent Chun, David Culler, Timothy Roscoe, Andy Bavier, Larry Peterson, Mike Wawrzoniak, and Mic Bowman. PlanetLab: An Overlay Testbed for Broad-Coverage Services. *SIGCOMM Computer Communication Review*, 33:3–12, July 2003.
6. Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 1797–1806. ACM, 2008.
7. E. Ferro and F. Potorti. Bluetooth and Wi-Fi wireless protocols: a survey and a comparison. *IEEE Wireless Communications*, 12(1):12–26, 2005.
8. Juan C. Herrera, Daniel B. Work, Ryan Herring, Xuegang (Jeff) Ban, Quinn Jacobson, and Alexandre M. Bayen. Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4):568–583, 2010.
9. Nathaniel Husted and Steven Myers. Mobile location tracking in metro areas: malnets and others. In *Proceedings of the 17th ACM conference on Computer and communications security*, CCS '10, pages 85–96. ACM, 2010.
10. European Telecommunications Standards Institute. AT Command Set for 3G User Equipment (UE). In *Digital Cellular Telecommunications Systems (Phase 2+);Universal Telecommunications System (UMTS)*, volume ETSI TS 127 007: 3GPP TS 27.007 version 5.2.0 Release 5 of *Technical Specification*. 2002.
11. Branislav Kusy, György Balogh, János Sallai, Ákos Lédeczi, and Miklós Maróti. InTrack: High Precision Tracking of Mobile Sensor Nodes. In *Proceedings of the 4th European conference on Wireless sensor networks*, EWSN'07, pages 51–66, 2007.
12. Konrad Lorincz and Matt Welsh. Motetrack: a robust, decentralized approach to rf-based location tracking. *Personal Ubiquitous Comput.*, 11:489–503, 2007.
13. Robert P. Minch. Privacy issues in location-aware mobile devices. In *Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences - Track 5 - Volume 5*, HICCS '04. IEEE Computer Society, 2004.

14. Eric Paulos, Richard E. Honicky, and Elizabeth Goodman. Sensing atmosphere. In *Proceedings of the 5th ACM Conference on Embedded Networked Sensor Systems*, SenSys '07, 2007.
15. Sean Rhea, Brighten Godfrey, Brad Karp, John Kubiatowicz, Sylvia Ratnasamy, Scott Shenker, Ion Stoica, and Harlan Yu. Opendht: a public dht service and its uses. In *Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM '05, pages 73–84. ACM, 2005.
16. Thomas Ristenpart, Gabriel Maganis, Arvind Krishnamurthy, and Tadayoshi Kohno. Privacy-preserving location tracking of lost or stolen devices: cryptographic techniques and replacing trusted third parties with dhts. In *Proceedings of the 17th conference on Security Symposium*, SS '08, pages 275–290. USENIX Association, 2008.
17. Krishna Sampigethaya, Mingyan Li, Leping Huang, and Radha Poovendran. Amoeba: Robust location privacy scheme for vanet. *IEEE Journal on Selected Areas in Communications*, 25(8):1569–1589, 2007.
18. SETI@Home. SETI@Home Project Website. http://setiathome.berkeley.edu/.
19. Hyojeong Shin, Yohan Chon, Kwanghyo Park, and Hojung Cha. FindingMiMo: Tracing a Missing Mobile Phone Using Daily Observations. In *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, MobiSys '11, pages 29–42. ACM, 2011.
20. Arvind Thiagarajan, Lenin Ravindranath, Katrina LaCurts, Samuel Madden, Hari Balakrishnan, Sivan Toledo, and Jakob Eriksson. Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, SenSys '09, pages 85–98, New York, NY, USA, 2009. ACM.
21. Wikipedia. Wikipedia Project Website. http://www.wikipedia.org/.
22. Ford-Long Wong and Frank Stajano. Location Privacy in Bluetooth. In *Proceedings of the 2nd European Workshop on Security and Privacy in Ad-Hoc and Sensor Networks*, ESAS '05, pages 176–188. Springer, 2005.
23. Yilin Zhao. Mobile phone location determination and its impact on intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 1(1):55–64, 2000.
24. Zhichao Zhu and Guohong Cao. APPLAUS: A Privacy-Preserving Location Proof Updating System for Location-based Services. In *Proceedings of the 30th IEEE International Conference on Computer Communications*, IEEE INFOCOM '11, 2011.
25. Zhichao Zhu and Guohong Cao. Towards privacy-preserving and colluding-resistance in location proof updating system. *IEEE Transactions on Mobile Computing*, 99(PrePrints), 2011.