CollabLoc: Infrastructure-free Privacy-preserving Localization via Collaborative Information Fusion

Vidyasagar Sadhu, Dario Pompili, Saman Zonouz, Vincent Sritapan*
Department of Electrical and Computer Engineering, *Cyber Security Division
Rutgers University, *Department of Homeland Security Science & Technology Directorate
{vidyasagar.sadhu, pompili, saman.zonouz}@rutgers.edu, *vincent.sritapan@hq.dhs.gov

1. INTRODUCTION

Motivation: Today's location-sensing techniques to enable location-based services still face open-research challenges, listed below. 1) *Time to market:* Indoor localization solutions are not ubiquitous yet due to the need for extra infrastructure or other requirements. 2) *Few hybrid solutions:* Majority of the localization solutions fall into either outdoor or indoor categories, where the former works only outdoor while the latter only indoor (and for one building at a time). 3) *Room-level granularity is sufficient:* In certain security applications such as [3], policies are defined based on the device's context (location, time, etc.); policies define allowed behavior, such as whether recording or making a call is allowed or not, and can depend on room-level location. Other examples are setting room temperatures based on human presence in the room. Achieving fine-grained accuracy comes at an additional cost of increased power consumption, which should be avoided unless absolutely needed. 4) *Privacy:* Many of the existing localization solutions ignore this important aspect; generally users are worried that the locations obtained by these solutions may be used in ways that could compromise their privacy. 5) *Reliability:* Location obtained through collaboration is likely more reliable than the one obtained by a single device due to lack of enough sensors in a single device, sensor noise, faulty data.

Our Approach: In order to address these challenges, we propose a unified location determination framework as a mobile application. Our framework can be readily deployed in the real world without the need for extra infrastructure—other than the already existing Wi-Fi Access Points (APs)—as we base it on users' smartphones. In addition to the Wi-Fi Received Signal Strength Indicator (RSSI), we make use of additional sensor data such as sound level, light level, cell signal level to distinguish among different regions/rooms within the area covered by a set of Wi-Fi APs. *Though we target CollabLoc mainly for room-level granularity*, the granularity can also be adjusted by a parameter called Wi-Fi similarity threshold (see Sect. 2). As mentioned above, having more-than-needed granularity comes at an additional cost. Unlike indoor localization solutions, CollabLoc is pervasive (i.e., not limited to a single building)—we do not build a map of the building or identify important fixtures (such as elevators). We tag different locations *in and around* buildings by a location label using the location-specific features (such as Wi-Fi APs and their strengths, sound, light, cell signal levels, etc.) as tags. To address privacy concerns and guarantee anonymous communications, CollabLoc is based on The Onion Router (ToR) network [2] composed of smartphones using CollabLoc and data perturbation techniques. To address the last issue on reliability, we obtain location through collaboration from multiple devices. Lastly, to reduce the energy footprint, we make use of Wi-Fi scan data most of the time, which is *automatically* generated by the mobile Operating System (OS) at no extra cost. For full details, see [1].

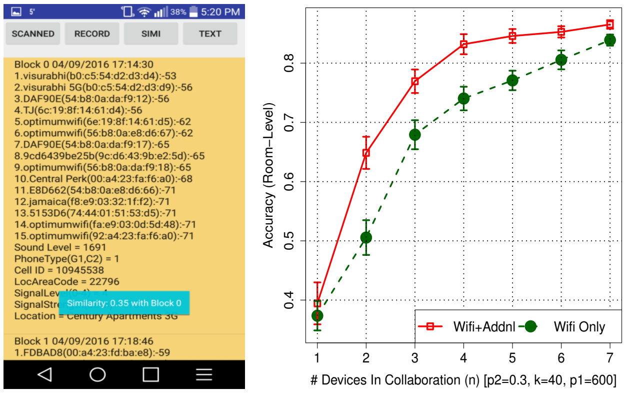
ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

2. OUR PROPOSED SOLUTION

Below we describe CollabLoc consisting of local database updating, obtaining the location label for a given location request and privacy-preserving communication between location requester and providers.

Local Learning: Each mobile phone maintains a local database with location features and the corresponding location labels. *Local learning* is a continuous phase (i.e., happens continuously) where the device updates its location database with *new* location features and labels. Location features consist of "list of Wi-Fi APs," "sound-level," "light-level," "indoor/outdoor," "cell tower ID," "Location Area Code (LAC)," "cell signal strength," etc. The method of populating this database with new features and labels is outlined below. CollabLoc needs Wi-Fi to be ON as it relies on Wi-Fi scan results. After receiving the scan results, the app checks to see if it is similar to any of the entries in its database. We use the concept of *similarity measure* (based on cosine similarity function) between two lists of APs to identify how significantly the two regions are similar to each other. If the similarity measure is found to be lower than a certain threshold value ($sim < sim_{th}$) (new location) or if $sim \geq sim_{th}$ but at least one of the location features is different compared to the entries in database, the app initiates a location request with all the location features (averaged over a small time window) collected at that moment. The request is forwarded to the appropriate location providers in a privacy-preserving manner (discussed below) and the received location label is used to make a new entry. If no response is returned from the collaboration, the application first waits for $t = t_{th} = 2$ hr. If no location change is detected, it recognizes the location as a significant location for the user and prompts at $t = t_{th}$ to enter manually the publicly-known label of that location (this happens only once per location and mostly only location-natives will be asked for this). We have a few techniques to reduce further the one-time manual entry and also to ensure uniformity among all labels using ontology-based methods.

Protecting Requester/Provider Privacy: Below, we explain how the collaboration between location *requester* and *providers* happens anonymously and how the final location distribution is generated.
1) Area Level Privacy: We will first introduce the concept of "Phone Masters" (*PMs*). PMs are present at different levels akin to Domain Name System (DNS) forming an overlay network on top of TCP/IP: *Country, State, County, City, Cell Tower*. Each device (requester/provider/PM) in CollabLoc has an ID that uniquely identifies it. Each PM stores the IDs of its children in its repository. The last level of PMs, namely the Cell Tower PMs (CTPMs) store the IDs of the location providers that have opted to remain anonymous in that area (we call this database, the repository of that CTPM). A requester needs to contact the corresponding CTPM to get the location label. A repository is used by CTPM to pick providers to respond to location requests it receives. The objective of this privacy option is to anonymize the provider's history of locations to an area level of his/her choice (cell tower, city, etc.). ***2) ToR routing:*** In order for the communication between a requester and a CTPM to remain anonymous, we adopt the technique of ToR/Onion routing [2], where the ToR network is the overlay network of PMs mentioned before. A requester conceals its location features (corresponding to a cell tower) in the onion packet with the final destination being the PM of that cell tower. Requester also generates a public-private key pair (Pu_R, Pr_R) for *each* request and includes the public key along with the onion packet in the request. It sends this request to a topmost level PM and each PM then forwards the request to the appropriate children until it reaches the CTPM it is destined to. Once the CTPM receives the request, it queries the providers in its repository to find the final location distribution (discussed below). It then encrypts this distribution with Pu_R and sends back to the requester. ***3) Final Distribution Generation:*** A CTPM upon receiving the location request first picks j devices at random from its repository to answer the location request. Each provider runs a two-step classification procedure (based on Wi-Fi AP list followed by additional features) to return the correct distribution of location labels for the given request and to find if it knows anything about the request; if not, it returns "NA". This distribution is then enlarged (by adding random labels), perturbed (random noise is added to probabilities), sorted (in decreasing probabilities), and then truncated (top few) before sending back to CTPM. Until the desired number of non-NA responses are obtained, the CTPM repeats the process iteratively doubling j each time. These non-NA responses are averaged by the CTPM and the final distribution is sent back to the requester as in 2) above.



Block 0 04/09/2016 17:14:30
1.visurabhi(b0:c5:54:d2:d3:d4):-53
2.visurabhi 5G(b0:c5:54:d2:d3:d9):-56
3.DAF90E(54:b8:0a:da:f9:12):-56
4.TJ(6c:19:8f:14:61:d4):-56
5.optimumwifi(6e:19:8f:14:61:d5):-62
6.optimumwifi(56:b8:0a:e8:d6:67):-62
7.DAF90E(54:b8:0a:da:f9:17):-65
8.9cd6439be25b(9c:d6:43:9b:e2:5d):-65
9.optimumwifi(56:b8:0a:da:f9:18):-65
10.Central Perk(00:a4:23:fa:f6:a0):-68
11.E8D662(54:b8:0a:e8:d6:66):-71
12.jamaica(f8:e9:03:32:1f:f2):-71
13.5153D6(74:44:01:51:53:d5):-71
14.optimumwifi(fa:e9:03:0d:5d:48):-71
15.optimumwifi(92:a4:23:fa:f6:a0):-71
Sound Level = 1691
PhoneType(G1,C2) = 1
Cell ID = 10945538
LocAreaCode = 22796
SignalLev
SignalStre
Location = Century Apartments 3G

Similarity: 0.35 with Block 0

Block 1 04/09/2016 17:18:46
1.FDBAD8(00:a4:23:fd:ba:e8):-59

(a) (b)

Figure 1: (a) App screenshot. (b) Accuracy (with and without additional features) vs. no. of collaborators.

3. DEPLOYMENT

Experimental setup and few results. Since CollabLoc works across buildings, we collected data from two separate regions (university campus, downtown area), totaling 15 locations/buildings, separated by a cell tower distance using 4 phones and 3 tablets. Within each of these locations, whenever the app detected that the received Wi-Fi scan results are dissimilar (we used $sim_{th} = 0.05$ to get a room-level granularity of 0.5 meters) with the entries in the phone's database (initially empty), the application recorded the following location features—list of Wi-Fi APs/strengths (upto 15), sound level, cell tower ID, Location Area Code, cell signal strength (in dBm). For evaluation purposes we have entered the label manually, else an attempt will be made to acquire it through collaboration. Figure 1a shows a screen-shot of the app showing the location features recorded. Each block is an entry in the database. Each device is able to record an average of 50 entries corresponding to different rooms/places *in* and *around* the above 15 buildings (not exhaustive coverage). To test CollabLoc, we collected location features at 15 places (out of 50 places in training data). For each of these test places, a location label is found from collaboration. Accuracy *(different from granularity)* is defined as the percentage of test cases that are classified correctly against the training dataset. Fig. 1b shows the accuracy of our approach with and without additional features with four collaborating devices and certain privacy levels.

Deployment Requirements: Wi-Fi APs and access to different rooms are needed to test the solution. We will bring multiple mobile devices to realize and demonstrate the collaboration benefits.

Acknowledgments

We thank the Dept. of Homeland Security Science & Technology Directorate (DHS S&T) Cyber Security Division.

4. REFERENCES

[1] Technical report. http://bit.ly/collabloc.
[2] Tor project. https://www.torproject.org/.
[3] G. Salles-Loustau, L. Garcia, K. Joshi, and S. Zonouz. Don't just BYOD, Bring-Your-Own-App Too! Protection via Virtual Micro Security Perimeters. In *IEEE/IFIP International Conference on Dependable Systems Networks*, 6 2016.