Privacy-Preserving Indoor Localization on Smartphones (Extended Abstract)

Andreas Konstantinidis*, Georgios Chatzimilioudis*, Demetrios Zeinalipour-Yazti*,
Paschalis Mpeis†, Nikos Pelekis‡ and Yannis Theodoridis‡

* University of Cyprus, 1678 Nicosia, Cyprus
† University of Edinburgh, Edinburgh EH8 9AB, United Kingdom
‡ University of Piraeus, 18534 Piraeus, Greece

***Abstract*—Predominant smartphone OS localization subsystems currently rely on server-side localization processes, allowing the service provider to know the location of a user at all times. In this paper, we propose an innovative algorithm for protecting users from location tracking by the localization service, without hindering the provisioning of fine-grained location updates on a continuous basis. Our proposed *Temporal Vector Map (TVM)* algorithm, allows a user to accurately localize by exploiting a k-*Anonymity Bloom (kAB)* filter and a bestNeighbors generator of camouflaged localization requests, both of which are shown to be resilient to a variety of privacy attacks. We have evaluated our framework using a real prototype developed in Android and Hadoop HBase as well as realistic Wi-Fi traces scaling-up to several GBs. Our study reveals that *TVM* can offer fine-grained localization in approximately four orders of magnitude less energy and number of messages than competitive approaches.**

I. Introduction

People spend 80-90% of their time in indoor environments[1], including shopping malls, libraries, airports or university campuses. The omni-present availability of sensor-rich mobiles has boosted the interest for a variety of indoor location-based services, such as, in-building guidance and navigation, inventory management, marketing and elderly support through Ambient and Assisted Living. To enable such indoor applications in an energy-efficient manner and without expensive additional hardware, modern smartphones rely on *Internet-Based Indoor Navigation (IIN)* services [4], which provide the accurate location (position) of a user upon request. There are numerous IIN services, including *Skyhook, Google, Indoo.rs, Wifarer, Navizon, IndoorAtlas, ByteLight* and our open in-house *Anyplace* system[2]. These systems rely on geolocation databases (DB) containing wireless, magnetic and light signals, upon which users can localize.

Particularly, IIN geolocation DB entries act as reference points for requested localization tasks. In summary, a smartphone can determine its location at a coarse granularity (i.e., km or hundreds of meters) up to a fine granularity (i.e., 1-2 meters), by comparing against the reference points, either on the service or on the smartphone itself. One fundamental drawback of IIN services is that they receive information about the location of a user while servicing them, generating a variety of location privacy concerns (e.g., surveillance or data for unsolicited advertising). These concerns do not exist with the satellite-based *Global Positioning System (GPS)*, used in outdoor environments, as GPS performs the localization directly on the phone with no location-sensitive information downloaded from any type of service. Although in this work we are mainly concerned with fine-grained Wi-Fi localization scenarios in indoor spaces, our discussion is equally applicable to other types of indoor fingerprints (e.g., magnetic, light, sound) and outdoor scenarios (e.g., cellular).

Location tracking is unethical in many respects and can even be illegal if it is carried out without the explicit consent of a user. It can reveal the stores and products of interest in a mall we've visited, doctors we saw at a hospital, book shelves of interest in a library, artifacts observed in a museum and generally anything else that might publicize our preferences, beliefs and habits. Somebody might claim that telecoms and governments are already tracking smartphone users outdoors, on the premise of public and national safety[3], thus there is no need to care about indoor location privacy either. Clearly, there is a lot of controversy on whether this is right or wrong, which has to do with different cultural, religious, legal and socio-economic dimensions.

We consider that IIN services are *fundamentally untrusted entities* and, as such, develop hybrid techniques that on the one hand exploit the IIN services utility, but on the other hand also offer controllable location privacy to the user. Particularly, we tackle the technical challenge of *enabling a user u to localize through an IIN service s, without allowing s to know where u is.* More formally, our desiderata is summarized by the following goal.

Research Goal. *Provide continuous localization to a mobile user u that can measure the signal intensity of its surrounding Access Points, with minimum energy consumption on u, such that a static cloud-based server s cannot identify u's location with a probability higher than a user-defined preference p_u.*

II. The TVM Algorithm

In [2], we devise the *Temporal Vector Map (TVM)* algorithm[4], which camouflages the location of some user u from s, by requesting a subset of k entries from s, where k is a user-defined constant. With the proposed method, s cannot identify u's location with a probability higher than a user-defined preference p_u.

To understand the operation of *TVM*, at a high level, consider the illustration of Figure 1 (left). An arbitrary user

[1]US Environmental Protection Agency, http://epa.gov/iaq/
[2]Available at: http://anyplace.cs.ucy.ac.cy/
[3]Dec. 4, 2013: The Washington Post, http://goo.gl/0jJcrL
[4]Available at: http://tvm.cs.ucy.ac.cy/

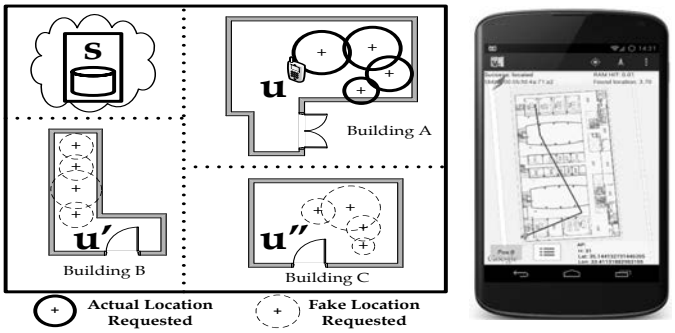

Fig. 1. (Left) Indoor localization of user u using the IIN service s. During the localization, u requests $k-1$ camouflaged locations using the *TVM* algorithm, such that s can know the location of u only with probability $1/k$. (Right) Our *TVM* prototype implemented in Android OS.

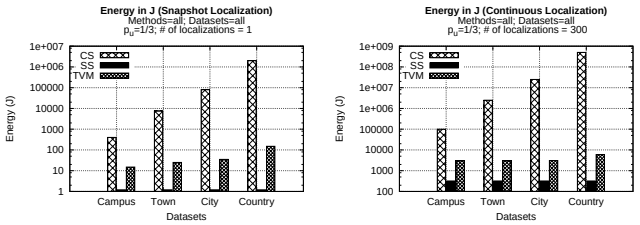

Fig. 2. **Performance Evaluation:** snapshot (left) & continuous (right) localization scenarios, respectively.

u moves inside building A, using the *TVM* smartphone application shown in Figure 1 (right). While u requests reference locations from s pertinent to building A, it also requests reference locations related to arbitrary other buildings B and C. Particularly, u uses a hashing scheme, whch ensures that for a given user-preference $k = 3$, s will not be able to distinguish u's request from requests made by $k-1$ arbitrary other users u' and u''. Under reasonable assumptions about the scope of IIN services, we show that s can know u's location only within p_u, even while u is moving. Particularly, the *TVM* algorithm operates in two phases outlined next.

In *Phase 1* of *TVM*, u computes a *k-Anonymity Bloom (*kAB*)* filter structure, which provides location privacy for *snapshot* localization tasks using a bloom filter [1]. When u needs *continuous* localization (e.g., as u moves), the *kAB* of Phase 1 itself is not adequate to preserve the privacy of u, since by issuing k independent requests, s can realize by exclusion that there are $k-1$ invalid requests (as one of the requests will always relate to the real building A). This allows s to deterministically derive u's real location.

To circumvent the above problem, in *Phase 2* of *TVM*, u uses the *bestNeighbors* algorithm to issue a set of camouflaged localization requests that follow a similar natural movement pattern to that of u (i.e., dotted circles in Figure 1, left). This provides the illusion to s that there are k other users moving in space, thus camouflaging u among k other users. Since our *TVM* algorithm transfers only a partial state of the database from s to u, it requires less network traffic and smartphone-side energy than current approaches that transfer the complete database to u prior the localization task.

When it Works. We consider a service that is fundamentally *untrusted*. As such, the service is operating in one of the following modes: i) it is compromised by the adversary owner of the service; or ii) it is compromised by some adversary third party (e.g., hacker). In both cases, the adversary can operate in the following two modes: i) an *active attacker* mode, in which the adversary attempts to alter system resources or actively combine background knowledge in order to infer where the users are; and ii) a *passive attacker* mode, during which the adversary attempts to learn from whatever data is available on the system (e.g., log files, wiretapping network sockets, etc.) without having additional information about the users. The *TVM* algorithm presented in this work, is sound under a *passive attacker* model, for which the following high-level characteristics apply to s (as explained in [2]): *i) No Low-Level Attacks; ii) No Modified Responses; iii) No Access to User Identifier*; and *iv) No Background Knowledge.*

III. Experimental Evaluation

We provide an extensive experimental evaluation with four different realistic datasets on our SmartLab cluster [3] comprising of over 40 real smartphones. We use four datasets: Campus, Town, City and Country, which have a size of: ≈20 MBs, ≈100 MBs, ≈1 GB and ≈20 GBs, respectively. The performance of our *TVM* approach is evaluated in terms of energy (in Joules) consumed by the smartphone device during the localization process. For measuring the performance of consecutive localizations we have defined a fixed route for each dataset, where a user localizes itself every 30 seconds for a total of 300 consecutive localizations. In our experiments we measure the cumulative cost of the whole route.

In our experiments we evaluate the performance and scalability of our *TVM* approach with respect to the following two counterparts: i) *Server-Side (SS)* solutions (i.e., Cell_ID, WiFi_ID or Server-side RadioMap), which are privacy-invasive, but consume minimal energy; and ii) *Client-Side (CS)* solution offering optimal privacy guarantees, but consuming the maximum possible energy.

For snapshot localization, Figure 2 shows that *TVM* performs around one to four orders of magnitude better than the *CS* approach as the dataset size increases. This is due to the fact that *CS* downloads the whole RadioMap (RM) and performs localization on the smartphone. Furthermore, the energy cost of *TVM* is not constant for all datasets, due the fact that larger datasets have a higher number M of access points, and therefore the required energy cost per message slightly increases. Similar results also apply to continuous localization.

References

[1] B. H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Commun. ACM*, 13(7):422–426, July 1970.

[2] A. Konstantinidis, G. Chatzimilioudis, D. Zeinalipour-Yazti, P. Mpeis, N. Pelekis, and Y. Theodoridis. Privacy-preserving indoor localization on smartphones. *Knowledge and Data Engineering, IEEE Transactions on*, 27(11):3042–3055, Nov 2015.

[3] G. Larkou, C. Costa, P. G. Andreou, A. Konstantinidis, and D. Zeinalipour-Yazti. Managing smartphone testbeds with smartlab. In *Proceedings of the 27th Intl. Conference on Large Installation System Administration*, LISA '13, pages 115–132, 2013.

[4] D. Zeinalipour-Yazti, C. Laoudias, K. Georgiou, and G. Chatzimiloudis. Internet-based indoor navigation services. *IEEE Internet Computing*, DOI: 10.1109/MIC.2016.21, Available Online: Feb. 18, 2016.