Tracking of Mobile Devices through Bluetooth Contacts

Nikodin Ristanovic
EPFL, Switzerland

Dang-Khoa Tran
EPFL, Switzerland

Jean-Yves Le Boudec
EPFL, Switzerland

ABSTRACT

We investigate if it is possible to reconstruct a mobile phone's mobility using its Bluetooth contacts with other mobile devices, some of which are equipped with GPS receivers. Our data mining analysis, based on two different data sets, shows that in certain environments coarse grained mobility of a significant portion of mobile phones can be obtained using this technique. For this reason, anyone capable of collecting Bluetooth and, if available, GPS logs for a large population of mobile phones, represents a serious privacy threat for all users with discoverable Bluetooth devices in the area.

1. INTRODUCTION

Low power consumption and security solutions built into Bluetooth led to the current situation where significant fraction of cell phones have their Bluetooth interfaces on all the time [1]. With the possible emergence of applications that involve mobile-to-mobile exchanges, the number of mobile phones with Bluetooth interfaces in discoverable mode could further increase. But, are we putting the little privacy that we have left at risk, by carelessly leaving our devices discoverable?

Mobile operators store customers' location records in CDRs. It is commonly believed that some phone manufacturers also collect location information. But, while we put a certain amount of trust in mobile operators, most of us would feel uncomfortable knowing that somebody unauthorized collects similar data. That would allow someone malicious to learn when we are outside our homes; advertising companies could get insight into our interests. The ability to reconstruct someone's mobility can also be beneficial. A solid estimate of skier's last location can save his life in case of an avalanche.

But how can anybody record my mobility, especially if I have a non-GPS phone? Let's assume for a second that someone managed to infect a population of mobile phones with a piece of malicious software. Such software can force phones to log Bluetooth contacts with other phones [3]. Additionally, it can force phones with GPS receivers to log their locations. The recorded information can then be uploaded to a central location via data connection. A cell phone bot like this one could allow reconstruction of mobility, not only for the phones with GPS, but also for other phones present in the trace (infected or not) as their locations can be estimated from the contacts with the GPS equipped phones.

We perform data mining analysis of two collected data sets. Our initial results show that even with simple algorithms (which do not involve advanced techniques, such as the training of a Bayesian network) it is possible to recreate coarse grained mobility of significant fraction of mobile devices present in the data sets.

2. RELATED WORK

Previously proposed approaches for wireless nodes localization exploit contacts with fixed reference points and knowledge of history, and their focus is location prediction. In [5] localization of nodes in 802.11 wireless networks is studied. The authors use information available at the link-layer of the 802.11 networks to train a Bayesian network and to calculate location. In [4, 6] the authors focus on wireless sensor networks.

In contrast, our focus is mobility reconstruction. We use no learning or prior knowledge and we obtain nodes' location from contacts with mobile reference points.

3. SYSTEM AND METRICS

The existing malware for mobile phones is capable of stealing data and logging GPS coordinates or Bluetooth neighbors [2]. It propagates via Bluetooth, e-mail, MMS, memory cards or infected applications. We assume that a certain population of mobile phones is infected by malicious software that forces the Bluetooth interface to stay visible and to periodically run the neighbor discovery mechanism. The recorded contacts are logged and uploaded to a central server via the phones' data connections. Some of the infected phones are equipped with GPS receivers; we refer to them as *anchors* or *reference devices*. In addition to logging contacts, they are required to log and upload their locations.

Our goal is to determine if the opportunistically collected contacts are sufficient to reconstruct the mobility of the *tracked* or *non-anchor devices*, i.e. the infected non-GPS phones. We are interested in two measures:

- **Mean localization error** is the average distance between real and estimated locations of tracked devices.
- **Mean inter-anchor time** is the average interval between two consecutive contacts, which occur between a tracked device and any of the anchors.

4. LOCALIZATION ALGORITHM

We design an algorithm that does not require training and knowledge of contact history. The algorithm exploits the fact that a contact between a tracked device and an anchor roughly determines the location of the tracked device, as the location of the anchor is known and the Bluetooth range is $\sim 10m$. Additionally, we assume that all the devices move at moderate speeds.

For these reasons, we consider the location of a tracked device d_1 to be known at time t iff: (i) a direct contact between d_1 and an anchor is recorded within the time window $[t - \Delta t, t + \Delta t]$ or (ii) a contact between d_1 and a tracked device d_2 is recorded at time t, another contact between d_2 and an anchor is recorded at time t_2 and $t \in [t_2 - \Delta t, t_2 + \Delta t]$.

5. USED DATA SETS AND RESULTS

We evaluate our algorithm and assumptions using two data sets. The first one was collected by 20 nodes, during a three day hiking trip. The hikers were moving through a large region in disorderly fashion. Their devices periodically logged Bluetooth contacts and GPS coordinates. The second data set is a two-week contact trace collected at the university campus by 60 mobile devices (for 10 of them GPS records exist).

We present only a part of our results for the first data set. The availability of GPS records for all devices in this data set allows us to estimate the localization error (we are able to compare the real with the estimated locations of tracked devices).

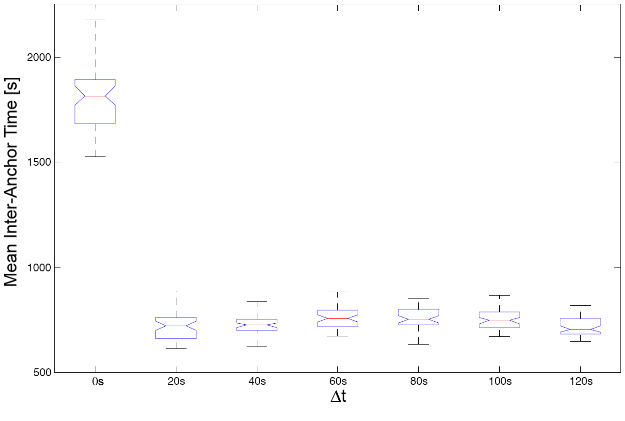

Figure 1: The average period during which the location of a tracked device is unknown.

When calculating the estimated locations of tracked devices we use only the anchors' GPS records. More precisely, we apply the following strategy: We select 5 out of 20 devices to be the anchors whose positions are known. The positions of other 15 tracked devices are reconstructed from the contact files using our algorithm. Anchor's positions between the logged locations are inferred using interpolation. The estimated positions are then compared with the logged positions of tracked devices, in order to obtain the localization error. We run 50 simulations, each with 5 randomly selected anchors.

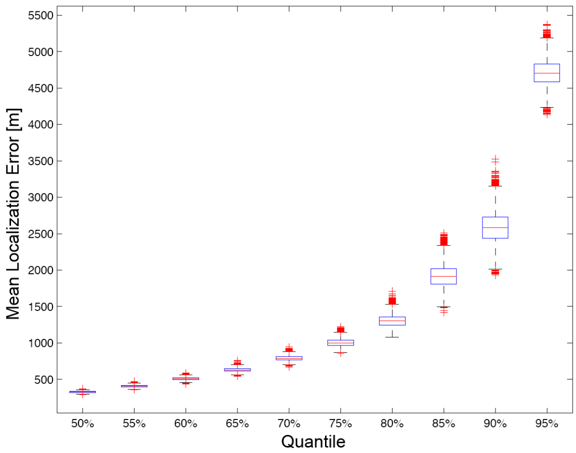

Figure 2: The localization error quantiles.

The average inter-anchor times for different values of parameter Δt are shown in Figure 1. Increasing the value of Δt from $0s$ to $20s$ allows for "indirect" contacts between tracked and anchor devices (as defined in Section 4) and significantly increases the number of location entries for the tracked devices. Further augmentation of Δt brings little improvement.

Some tracked phones can be localized with higher precision than the others. For this reason, we calculate the localization error quantiles. Figure 2 contains the quantiles for $\Delta t = 80s$. We see that 50% of tracked devices can be localized within $300m$ and 70% within $800m$.

6. REFERENCES

[1] G. Ananthanarayanan and I. Stoica. Blue-fi: enhancing wi-fi performance using bluetooth signals. In *Proc. MobiSys '09.*

[2] Symbianpoint. http://symbianpoint.com/types-latest-list-mobile-viruses.html.

[3] BlueAttacks. http://gcn.com/articles/2005/07/20/a-menu-of-bluetooth-attacks.aspx.

[4] S. Guha, R. Murty and E. Gun Sirer. Sextant: a unified node and event localization framework using non-convex constraints. In *Proc. MobiHoc 05.*

[5] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach and L. E. Kavraki. Practical robust localization over large-scale 802.11 wireless networks. In *MobiCom 04.*

[6] G. Mao, B. Fidan, D. Towsley and B. D. O. Anderson. Wireless sensor network localization techniques. Computer Networks, 2007.