Automatic Inference of Movements from Contact Histories

Pengcheng Wang Zhaoyu Gao Xinhui Xu Yujiao Zhou
Haojin Zhu* and Kenny Q. Zhu*
Shanghai Jiao Tong University
Shanghai, China
{wpc009, gaozy1987, xuxinhui08, yujiao.zhou}@gmail.com
*{zhu-hj, kzhu}@cs.sjtu.edu.cn

ABSTRACT

This paper introduces a new security problem in which individuals movement traces (in terms of accurate routes) can be inferred from just a series of mutual contact records and the map of the area in which they roam around. Such contact records may be obtained through the bluetooth communication on mobile phones. We present an approach that solve the trace inference problem in reasonable time, and analyze some properties of the inference algorithm.

Categories and Subject Descriptors

C.2.m [**Computer-Communication Networks**]: Miscellaneous

General Terms

Algorithm, Experimentation, Security

Keywords

Traces, Inference, Contacts, Location privacy

1. INTRODUCTION

Location privacy and, in particular, privacy of individuals' timestamped movements is receiving increasing interest. Recent research indicates that the widespread use of WiFi and Bluetooth enabled smartphones opens new doors for malicious attacks, including geolocating individuals by illegitimate means (e.g. spreading worm based malware) and even legitimate means (e.g. location based advertisement networks and theft locators) [4, 3, 1]. However, most existing geo-localization techniques require GPS information or additional control of hardware infrastructures such as WiFi access points or GSM base stations. This paper presents a new technique to infer individuals' complete movement traces using only their mutual contact histories and a map. This technique can be deployed both indoors and outdoors, so long as the area map is available. To the best of our knowledge, the only other work that attempts to infer traces from bluetooth contact histories is by Whitbeck et al. [5]. However, they did not make use of a map and therefore only produces rough moving trajectories. It also relies on the existence of various imaginary forces that are supposed to bias the person's movement, though it is not clear how to determine the parameters involved in the forces calculation. The technique in this paper, on the contrary, produces detailed movement traces according to a map.

Next we describe the *Trace Inference Problem*. We define a map M as a graph (V, E), where V is a set of road junctions each with geographic coordinates $(x,\ y)$, and E is a set of straight road segment. Note that a curved road can be approximated by a sequence of straight road segments. Let a set of nodes N move on the edges of the map at various but constant speeds. We assume that a node never backtracks unless it's at a dead end. Let the trace of node i be a location function on time $l_i(t)$, and a *contact* between node i and j be a 4-tuple: (i, j, t_{in}, t_{out}) where t_{in} is when the encounter of i and j begins, and t_{out} is when their encounter ends. Further, a *contact history* is a set of contacts. Given the set of traces of N, we can *induce* all the contacts by solving inequality $||l_i(t) - l_j(t)|| \leq r$ for all pairs of traces by node i and j, where r is the range in which two nodes are considered in contact. For simplicity, we assume $r = 0$ in the rest of this paper, and the contact induction can be computed in $O(|N|^2|V|^2)$ time.

Trace Inference Problem(TIP): Given a map M, a set of moving nodes N, their speeds $\{v_i\}$, their initial locations $\{l_i(0)\}$, and a contact history H, find the traces of N whose induced contacts $H_{ind} = H$.

Suppose the last contact in H is at t_{max}, the maximum speed is v_{max}, the length of the shortest edge in M is e_{min}, and largest degree of any vertex on M is d_{max}, then a naive search across all possible paths costs

$$O\left(|N|^2|V|^2(d_{max} - 1)^{\frac{|N|v_{max}t_{max}}{e_{min}}}\right).$$

2. OUR APPROACH

Our main idea is to decompose **TIP** into $|H|$ subproblems, each resolving a single contact in H. To solve a contact of i and j at time t, we consider the locations of their last contacts. The key observation is that even though there can be many possible traces spanning from the previous locations of i and j, only a small number of them can produce the contact, due to the distance constraints of the map.

Fig.1 illustrates this approach. Suppose nodes A, B and C starts their movements at t_0 from point A, B and C in the diagram. A and B contact at t_1. Given A's speed, we know by t_1, there are 5 possible traces for A ending at $A1$-$A5$ respectively. Similarly, B has 6 possible traces and 5 locations up to time t_1. Of all 15 pairs of traces between A

SIGCOMM'11, August 15–19, 2011, Toronto, Ontario, Canada.
ACM 978-1-4503-0797-0/11/08.

and B, only 4 pairs of traces satisfy the contact constraint, which may occur at locations marked by ($A1$, $B1$), ($A4$, $B3$) and ($A2/A3$, $B2$). The rest of the traces are pruned and not considered further. In the next round, B and C contact at time t_2. We repeat the above process, using $B1$, $B2$ and $B3$ as the B's initial locations for this round. As the figure shows, the only possible trace for B left is $B1 \rightarrow B6$ given the short time interval between t_1 and t_2.

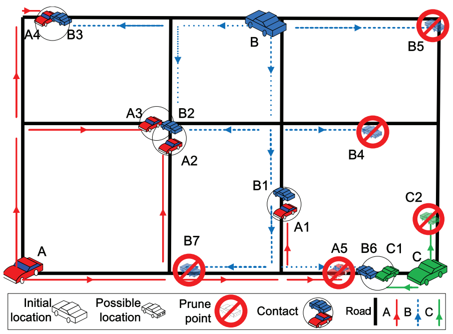


Figure 1: Inference of a Short Contact History

If we model the movements and contacts as a stochastic process, the time interval t between two successive contacts of i and j follows the exponential distribution [2]:

$$P\{t \leq x\} = 1 - e^{-\lambda_{ij}x}, x \in [0, \infty)$$

where λ_{ij} is the contact rate between i and j, the expected contact interval between i and j is $E[t] = \frac{1}{\lambda_{ij}}$. Let λ_{ij} be its mean value $\bar{\lambda}$, the number of contacts $|H|$ is

$$|H| = \frac{1}{2}\sum_{i}\sum_{j \neq i} t_{max}/\frac{1}{\lambda_{ij}} \approx \frac{1}{2}\bar{\lambda}t_{max}|N|^2$$

Let t_c be the expected time to infer each contact. If the contact interval of all nodes is bounded, then t_c is also bounded. Therefore, the time cost of our approach is linear to $|H|$, which is also linear to t_{max} and $|N|^2$.

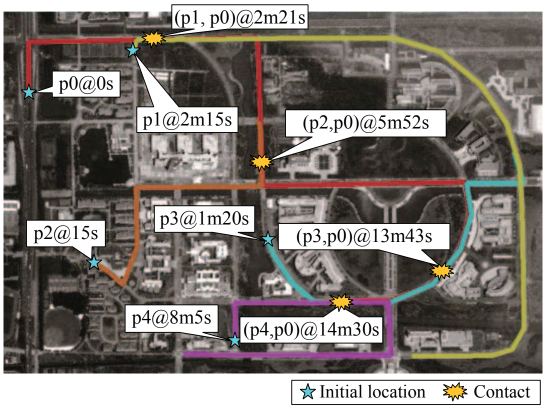
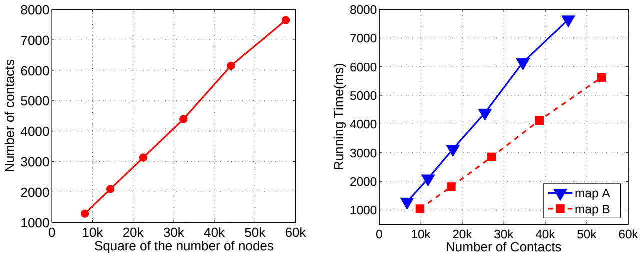


Figure 2: Five Inferred Traces on SJTU Campus



(A) (B)

Figure 3: Induced Contacts vs. Nodes (A) Scale-up on Contacts (B)

3. PRELIMINARY RESULTS

We implement the approach, run it on numbers of synthetic data sets and get the following results. Each trace is generated by randomly selecting an origin and a destination on a map of Shanghai Jiao Tong University campus and has a duration of 1 hour. Starting from the origin, we randomly select the next location among all adjacent junctions. A junction is more likely to be selected if it is closer to the destination. From the traces, we can induce a list of contacts and their locations like the following:

p0	p2	339.7877	154.0384	17.5236
p0	p2	41.2401	276.1339	87.4113
p0	p179	41.3582	284.7934	92.3681
...	...	...	...	...

Each row contains information for one contact. The columns represent the ids of two nodes in contact, the X and Y coordinates of the contact location, and contact time. The coordinates are not used in inference but in validation. We run 11 experiments in which the sizes of contact histories range from hundreds to 60,000. All traces are correctly inferred. Fig.2 shows 5 inferred trace fragments of nodes p_0 through p_4, along with their initial locations and contacts. Fig.3(A) shows the number of contacts induced in the data set is roughly proportional to the square of the number of nodes. Fig.3(B) shows the running times of the algorithm on various data sets. The solid line represents results for data on map A with 48 junctions which corresponds to the area in Fig.2. The dotted line represents the results for data on a smaller map B with 25 junctions. The running time is almost linear to size of contact histories. These preliminary results are in line with the discussion in Section 2.

4. ACKNOWLEDGEMENT

This work was partially supported by NSFC (Grant Nos. 61033002 and 61003218).

5. REFERENCES

[1] I. Constandache, X. Bao, M. Azizyan, and R. R. Choudhury. Did you see Bob?: human localization using mobile phones. In *MOBICOM*, 2010.

[2] W. Gao and G. Cao. User-centric data dissemination in disruption tolerant networks. In *IEEE Infocom*, 2011.

[3] N. Husted and S. Myers. Mobile location tracking in metro areas: malnets and others. In *ACM CCS*, 2010.

[4] C. Y. T. Ma, D. K. Y. Yau, N. K. Yip, and N. S. V. Rao. Privacy vulnerability of published anonymous mobility traces. In *MOBICOM*, 2010.

[5] J. Whitbeck, M. D. de Amorim, and V. Conan. Plausible mobility: Inferring movement from contacts. In *ACM MobiOpp*, 2010.