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.

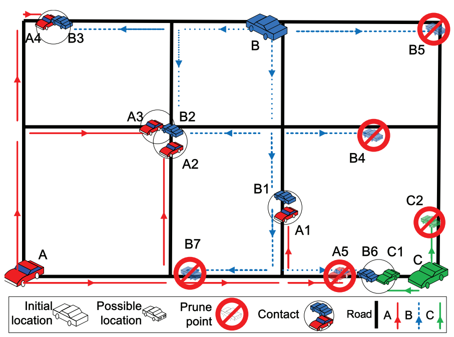

Figure 1: Inference of a Short Contact History

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 $|N|^2$.

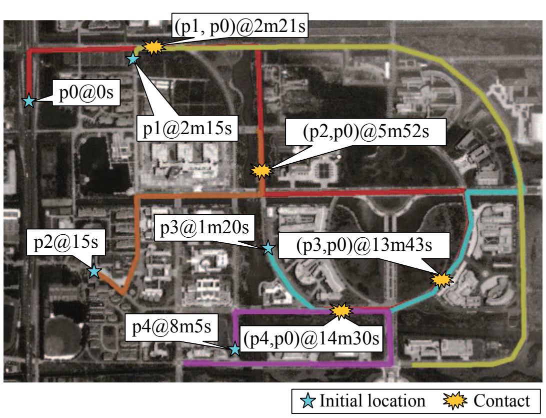

Figure 2: Five Inferred Traces on SJTU Campus

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.

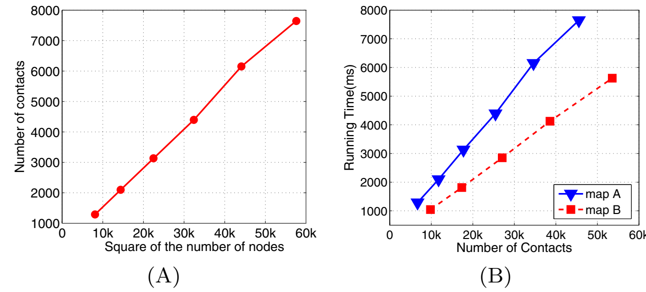

Figure 3: Induced Contacts vs. Nodes (A) Scale-up on Contacts (B)

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.