B.Comp. Dissertation

Adversarial Machine Learning: Exploration of Potential Privacy Risks In TraceTogether

By

Yang Shuqi

Department of Computer Science

School of computing

2021/2022

Project No: H0041450

Advisor: Assoc Prof Chang Ee-Chien

Project Objectives Description

COVID-19 is a highly contagious epidemic. In many countries, social distancing has been an important preventive measure (Tang, 2020). Contact tracing, as a means of alerting users who have come into close contact with infected people, has been proven to be effective in the situation of a contagious epidemic (Tang, 2020). Due to the rapid spread of the COVID-19 virus, digital contact tracing solutions which are based on mobile apps, have been developed to prevent the epidemic more efficiently (Tang, 2020). In the following discussions, we will focus only on *TraceTogether*, the contact tracing solution introduced by Singapore. Thus, the phrase 'contact tracing' refers only to the means of *TraceTogether* to collect the relative location among users and alert them when they are considered to be in danger.

Under such an extreme situation of COVID-19, the public may not put much attention on the privacy issue. An example is that China has applied Alipay Health Code for tracking the public's movement, a software which sends the person's location to a server controlled by the company and the government cooperatively (Mozur et al., 2020). Maya Wang, a China researcher for Human Rights Watch, has considered this as the starting point of mass surveillance in China (Mozur et al., 2020).

Inspired by the situation in China, we would like to look into the situation in Singapore, and thus *TraceTogether*. Given that only infected users need to upload their personnel information to the server, we will look into the privacy introduced by the third party. The purpose of this project is to explore what information malicious users can mine even with encrypted temporary pseudonyms and thus the potential privacy problems of *TraceTogether*. For the current version, the purpose is simplified as grouping pseudonyms, linking the trajectories, and thus tracking the pedestrians' rough walking paths, which can be used to explore some more personal information in the future study.

Related Work

Potential risks

Due to the COVID-19 situation, many countries have applied contact tracing solutions to control the spread of the epidemic. In 2020, Singapore Government Technology Agency (GovTech) has built *TraceTogether*, an app that uses relative location data to solve the contact tracing problem to avoid privacy problems introduced by the authority agencies mentioned above. The phones with *TraceTogether* installed keep broadcasting their temporary encrypted pseudonyms, and once any two phones are within the Bluetooth communication range, they will exchange the information pair (timestamp, Bluetooth signal strength, the phone's model, temporary identifier) with each other (GovTech, 2020). Each individual's data will only be shared with the government when he or she is tested positive for COVID-19 to further protect individuals' privacy (Tang, 2020). Although these studies have pointed out the potential privacy risk of sharing the location information of individuals with the government, they have failed to notice the potential risk of the third party, which can be malicious and uncontrollable.

TraceTogether was developed with the assumption that the Ministry of Health (MoH) of the Singapore government can protect the users' information properly, which means that snoopers should not be able to hack the government database for the information stored (Asghar et al., 2020). However, similar to vehicle transmissions, which can be easily collected by anyone with the radio range (Wiedersheim et al., 2010), any mobile phone with *TraceTogether* installed can collect the information sent to it from others passed within its Bluetooth communication range. Realising this potential problem, GovTech has applied encrypted periodically updated pseudonyms to protect each individual to be discovered. The idea is similar to mix zones that were proposed by Beresford and Stajano in 2003. The concept is to split paths into unlink-able segments to increase privacy (Beresford & Stajano, 2003). However, as Wiedersheim (2010) pointed out, simple pseudonym change is not enough for location privacy. It should be possible for malicious users to group different pseudonyms and target them to a single user.

In the following discussions, we simulate a situation where snoopers place stationary devices in crowded areas in a grid. The devices are expected to exchange the information with users who pass by and are tracked within their Bluetooth communication range. All of the devices are assumed to check if they can exchange information with users periodically, starting at the same time, following

a global clock. The details about how data is generated and collected are described under *Progress Made So Far*.

Previous Work

It is shown that detailed location profiles can be easily generated from the encrypted information by correlating the profiles (Wiedersheim et al., 2010). However, their solution is based on the precise GPS location and has considered the specific behaviours of vehicles, such as no collision and following traffic lights. Moreover, the vehicles change their pseudonyms every time they send out information, which suggests there will be no sub-paths and travel history to be used. The task can be considered as grouping points potentially from the same vehicle.

In 2005, Gruteser and Hoh have proposed a situation that is more similar to *TraceTogether*, where the vehicles reported their location more frequently, and thus the task will be linking trajectories instead of points. Multi-target tracking algorithms have been proven to be able to reconstruct paths from trajectories with periodically updated pseudonyms (Gruteser & Hoh, 2005). However, the solution becomes less reliable when the paths intersect and thus confused about which path to choose. One possible solution they proposed is that the attackers can reduce the sampling rate to avoid trajectories intersect with each other (Gruteser & Hoh, 2005). However, this improvement cannot be applied to protect privacy for *TraceTogether*, because the application keeps broadcasting for tracing enough and accurate distance information.

While there are several studies on location privacy, as mentioned above, most of them focusing on the reconstruction of the paths from GPS location and are based on vehicles. The relative location may not have been widely used before COVID-19, and thus did not draw much attention from the researcher. Moreover, compared with vehicles, humans show more social activities, and therefore more information may be used for prediction.

As Gruteser and Hoh (2005) and Beresford and Stajano (2003) mentioned, the location traces may implicitly suggest privacy information of an individual, such as medical problems, political preferences, as well as the location of their residence. While *TraceTogether* protects the users from leaking their precise location by storing the Bluetooth signal strength instead of GPS position, by putting enough devices and signal boosters in the area under surveillance, their actual position can be estimated with reasonable errors. In this case, any third party may reconstruct the movement of the public in an area.

Progress Made So Far

Simulation

- **Pedestrian Identifier**: Simulated *TraceTogether* allocates the encrypted temporary identifier $Pid_{t_{i,j}}$ to a pedestrian periodically at time $t_{i,j}$. $t_{i,j}$ represents a timestamp that starts from different $t_{i,0}$ for each pedestrian and have passed j time interval $TIME\_INTERVAL$ of 15 minutes.

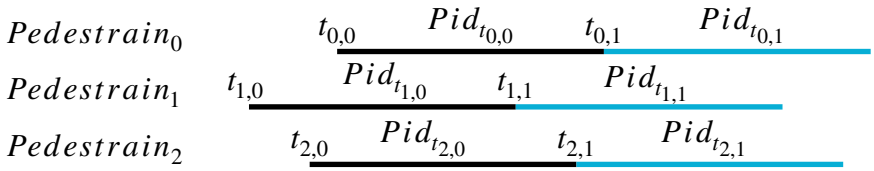

Figure 1. Life Cycle of Pedestrian Identifier

- **Device Identifier**: $Did_{x,y}$ is used to refer to the device put at the position x, y without considering the periodic update. This simplification can be done because the devices are stationary and thus even if they are allocated temporary pseudonyms, we can identify them by the positions.

- **Position of Devices**: Devices are placed in a 20 * 20 grid, covering the whole area $SCALE\_METERS * SCALE\_METERS$ under surveillance.

- **Walking of Pedestrian**: BLE contact tracing sniffer PoC implemented by Seiskari (2020) is used as the basis for simulating the walk of pedestrians. Modification has been done for fitting our problem. The simulation of a walking path is described below:

  1. Initiate a random starting point $position_0$ with standard normalisation with deviation of hyper-parameter $SCALE\_METERS$.

  2. Initiate a random initial velocity v_0 with standard normalisation with deviation of $\frac{1}{10} SCALE\_METERS$ in both x and y directions.

  3. Generate a random velocity v_i with standard normalisation with mean v_{i-1} and deviation of $\frac{1}{100} SCALE\_METERS$ in both x and y directions.

  4. Add point $position_i = position_{i-1} + v_i * TIME\_INTERVAL$ to the path.

Sense Data

- All of the devices are assumed to exchange the information with users periodically at time T_k starting from **exactly the same time** T_0, following a global clock. As specified by *TraceTogether*, the information will only be exchanged when the devices and the pedestrians are within each other's Bluetooth communication range.
- To simplify the situation, the distance $D_{T_k,Did_{x,y},Pid_{t_{i,j}}}$ to represent the information (Bluetooth signal strength, the phone's model). The Bluetooth signal strength can represent the real distance between any two endpoints with reasonable errors, as long as the Bluetooth signal strength difference between the two phones are known (GovTech, 2020).
- Snoopers will thus collect a database of information pair (T_k, $D_{T_k,Did_{x,y},Pid_{t_{i,j}}}$,$Pid_{t_{i,j}}$, $Did_{x,y}$)

Methodology

In order to link the trajectories, position of each pedestrians are needed. Since the simulator is following the logic from *TraceTogether*, only relative location is provided. However, there are enough devices putting in the area under surveillance, and thus the position of the devices are used as the approximated location of the user. The devices are put in grid, therefore, instead of precise position, we represent the position of each user as the grid index $Grid_{x,y}$. When a pedestrian is tracked by multiple devices, the one with highest confidence to determine the next step will be chosen as the representation of that pedestrian.

Linking the trajectory with $Pid_{t_{i,j}}$ with another trajectory, is the same as finding the next position of (T_k, $D_{T_k,Did_{x,y},Pid_{t_{i,j}}}$,$Pid_{t_{i,j}}$, $Did_{x,y}$), and grouping $Pid_{t_{i,j}}$ and $Pid_{t_{i_{next},j_{next}}}$. Given T_k representing the timestamp when the information pair is tracked, only (T_{k+1}, $D_{T_{k+1},Did_{x',y'},Pid_{t_{i',j}}}$,$Pid_{t_{i',j}}$, $Did_{x',y'}$) where $Pid_{t_{i,j}} \neq Pid_{t_{i',j'}}$ should be considered as candidates.

$$Pid_{i_{next},j_{next}} = \arg\max_{i',j'}(P(Pid_{t_{i',j'}}))$$

In order to fully cover the area under surveillance, the Bluetooth communication range of multiple devices can cover each other. Therefore it is possible for a pedestrian with $Pid_{t_{i',j'}}$ to be tracked by multiple devices at different position. Average is used to make use of the information from all of them:

$$P(Pid_{t_{i',j'}}) = Average_{\forall x',y'}(P((T_{k+1}, D_{T_{k+1},Did_{x',y',Pid_{t_{i',j'}}}}, Pid_{t_{i',j'}}, Did_{x',y'})))$$

The probability of moving from $Grid_{x,y}$ to $Grid_{x',y'}$ is the key point for grouping Pid together. Three aspects are considered in order to compute the probability:

(1) $P_{(x,y)\to(x',y')}$: the probability of moving from $Grid_{x,y}$ to $Grid_{x',y'}$;

(2) $P_{(x'-x,y'-y)}$: the probability of moving toward direction $Grid_{x',y'} - Grid_{x,y}$;

(3) $\alpha Distance(Grid_{(x,y)}, Grid_{(x',y')}, \Delta S_{x,y})$: the weighted difference between the estimated next position $Grid_{x,y} + \Delta S_{x,y}$ and $Grid_{x',y'}$.

Therefore,

$$P((T_{k+1}, D_{T_{k+1},Did_{x',y',Pid_{t_{i',j'}}}}, Pid_{t_{i',j'}}, Did_{x',y'})) = (P_{(x,y)\to(x',y')} + P_{(x'-x,y'-y)} + \alpha Distance(Grid_{(x,y)}, Grid_{(x',y')}, V_{(x,y)}))$$

To estimate the next position $Grid_{x,y} + \Delta S_{x,y}$, we compute $\Delta S_{(x,y)}$ by assuming that the pedestrian always travel through a similar distance in $T_k - T_{k-1}$ for any k. Therefore,

$$\Delta S_{(x,y)} = D_{T_k,Did_{x,y},Pid_{t_{i,j}}} - D_{T_{k-1},Did_{x'',y''},Pid_{t_{i,j}}}$$

Evaluation

I. Statistical

Instead of testing the different between the actual paths and the generated ones, the accuracy of grouping of Pid is used for current version. For each predicted group, the largest sub-group that is in fact from the same $Pedestrain_i$ is considered as the correct assignment, the remaining is incorrect assignment. The equation

$$\frac{\sum_i \max_{g_j \in G_i} num(g_j)}{\sum_i \sum_{g_j \in G_i} num(g_j)}$$

is used for calculating the accuracy of assigning $Pid_{t_{i,j}}$ into the correct group.

The statistical prediction accuracy with 10 pedestrians walking for 150 step each is 96.47%.

II. Visual

A webpage showing the linked paths with the grouping of Pids are implemented for better visualisation. The dashed line with the same colour represent the paths from a single pedestrian in fact. The lines with the same colour represent the reconstructed paths from a group of sub-paths that we believe travelled through by a single pedestrian. The blue dots represents the sub-paths that we did not successfully assigned to a group.

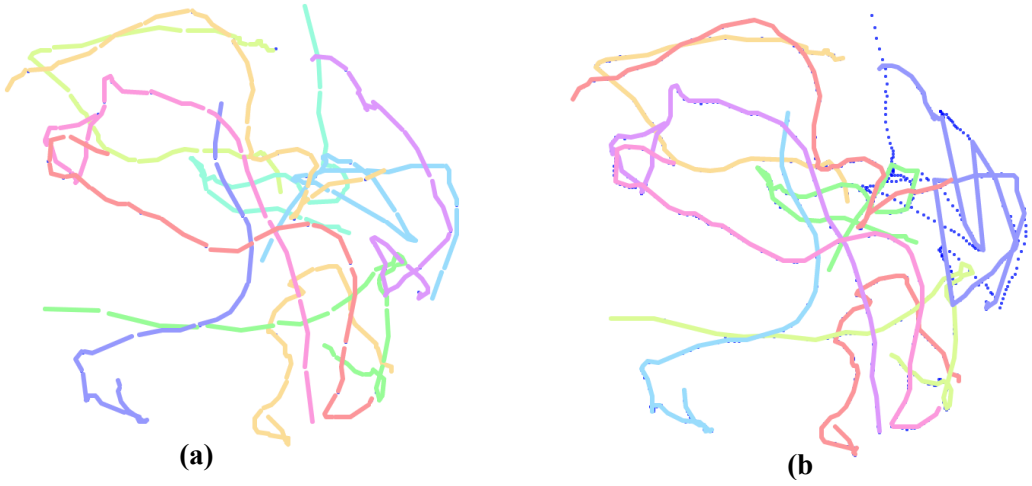

Figure 2. The visualisation prediction result with 10 pedestrians walking for 150 step each: (a) Visualisation of unlinked trajectories, the trajectories with the same colour are from the same pedestrian; (b) Visualisation of predicted linked trajectories, the trajectories with the same colour are from the same predicted group, i.e. the model predict that they are from the same pedestrian.

Future Plan

The current version of the simulator has the following unreasonable assumptions:

1. Pedestrians keep walking without any stop.
2. All devices exchange information at exactly the same time, following a global clock.
3. All walkers are independent, there is no relationship between them.
4. The problem is Markov Decision Problem.
5. Timestamp is not considered.

Assumption 1 estimated the next position more accurate than it is expected. This assumption, together with the way we randomise the velocity, we can newly achieve the accurate next position, which should not be that case in reality. Assumption 2 further simplifies the selection of candidates. With this assumption, we can select data for candidates by choosing the data with T_{i+1}. However, when this assumption is released, the information exchange will happen once a pedestrian becomes within the communication range of a device with some reasonable delay. This will predict the next step related to time as well, and therefore the prediction for each candidate should be different, which will make the computation more complex. By releasing assumptions 1 and 2, we may notice that instead of a single predicted next position for all candidates, each candidate will need to generate a range of predictions. Kalman filter has been proven to be accurate in predicting the next position for vehicles (Wiedersheim et al., 2010; Gruteser & Hoh, 2005), it may transfer to this problem as well.

Moreover, the current methodology relaxed the reconstruction problem to Markov Decision Problem, which means the travel trajectory does not make any difference. The relaxation is reasonable given the data is generated from the simulator. However, when it comes to real data, people from point A may tend to pass by point B and go to point C, while those from point D are more likely to end up with point E passing by point B. Assumption 3 is also related in this case. The relationship is a typical attribute for pedestrians. Compared with vehicles, pedestrians' behaviours can be examined, and their relationship suggested by that can also help in improving the prediction. Assuming two pedestrians keep moving together for some time, they tend to move together later as well. Multiple Hypothesis Tracking (MHT) proposed by Wiedersheim in 2020 can

be applied in this case. The method keeps generating a tree, and only decide which nodes to go to when it reaches to leaves.

The timestamp is also not considered in the current version of the implementation, however, it is natural for people to move towards the canteen at lunchtime and towards home when a day finishes. The dimension of time can be added in for better estimation. Though in this case, we may need to generate a map with functional facilities marked as well.

References:

Asghar, H., Farokhi, F., Kaafar, D., & Rubinstein, B. (2020, June 03). Privacy of the COVID-19 Tracing App - Everything you need to know! Retrieved October 29, 2020, from https://www.mq.edu.au/about/about-the-university/offices-and-units/optus-macquarie-university-cyber-security-hub/news-and-events/news2/news/covid-tracing-app

Beresford, A., & Stajano, F. (2003). Location privacy in pervasive computing. *IEEE Pervasive Computing, 2*(1), 46-55. doi:10.1109/mprv.2003.1186725

Government Technology Agency. (2020, March 25). TraceTogether - behind the scenes look at its development process. Retrieved October 29, 2020, from https://www.tech.gov.sg/media/technews/tracetogether-behind-the-scenes-look-at-its-development-process

Gruteser, M., & Hoh, B. (2005). On the Anonymity of Periodic Location Samples. *Security in Pervasive Computing Lecture Notes in Computer Science,* 179-192. doi:10.1007/978-3-540-32004-3_19

Mozur, P., Zhong, R., & Krolik, A. (2020, March 02). In Coronavirus Fight, China Gives Citizens a Color Code, With Red Flags. Retrieved October 31, 2020, from https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html

Seiskari, O. (2020, September). BLE contact tracing sniffer PoC [Computer software]. Retrieved from https://github.com/oseiskar

Tang, Q. (2020, April 25). Privacy-Preserving Contact Tracing: Current solutions and open questions. Retrieved October 29, 2020, from https://arxiv.org/abs/2004.06818

Wiedersheim, B., Ma, Z., Kargl, F., & Papadimitratos, P. (2010). Privacy in inter-vehicular networks: Why simple pseudonym change is not enough. *2010 Seventh International Conference on Wireless On-demand Network Systems and Services (WONS)*. doi:10.1109/wons.2010.5437115