APPENDIX A
PROOF OF EQUATIONS (5) AND (6)

As the adversary does not have knowledge about conditional mobility profiles for the users, their mobility profiles are independent of each other – formally, $\Pr(a_u(t) = r | a_{u'}(t) = r') = \Pr(a_u(t) = r)$, for any users u and u'. Using *Bayes' rule* it follows that, for any $\mathbf{r} \in \mathcal{R}^N$

$$\Pr(\mathbf{a}(t) = \mathbf{r}) = \prod_{i=1}^{N} \Pr(a_{u_i}(t) = r_i) \tag{A.1}$$

We start the proof of Equation (5) by proving its base case: $t = 0$.

$$\begin{aligned}
\alpha_0^{\mathcal{U}}(\mathbf{r}) &= \Pr(\mathbf{a}(0) = \mathbf{r} \,|\, \mathcal{K}) && \text{(A.2)}\\
&= \Pr(a_{u_1}(0) = r_1 \,|\, \mathcal{K}) \times \cdots \times \\
&\qquad \Pr(a_{u_N}(0) = r_N \,|\, \mathcal{K}) && \text{(A.3)}\\
&= \pi_{u_1}(r_1) \ldots \pi_{u_N}(r_N) && \text{(A.4)}\\
&= \pi_{\mathcal{U}}(\mathbf{r}) && \text{(A.5)}
\end{aligned}$$

In step (A.2)→(A.3) of the derivation, we use the independence assumption (A.1); in step (A.3)→(A.4), we use the fact that the probability of a user u being in some region r at time $t = 0$, given her mobility profile, is captured by the steady state vector, *i.e.*, $\pi_u(r)$, as there are no observations at, or before, $t = 0$. We now complete the proof for any $t > 0$.

$$\begin{aligned}
\alpha_t^{\mathcal{U}}(\mathbf{r}) &= \Pr(\mathbf{o}(1) \ldots \mathbf{o}(t), C_1 \ldots C_t, \mathbf{a}(t) = \mathbf{r} \,|\, \mathcal{K}) && \text{(A.6)}\\
&= \Pr(C_t \,|\, \mathbf{o}(1) \ldots \mathbf{o}(t), C_1 \ldots C_{t-1}, \mathbf{a}(t) = \mathbf{r}, \mathcal{K}) \cdot \\
&\qquad \Pr(\mathbf{o}(1) \ldots \mathbf{o}(t), C_1, \ldots, C_{t-1}, \mathbf{a}(t) = \mathbf{r} \,|\, \mathcal{K}) && \text{(A.7)}\\
&= \Pr(C_t \,|\, \mathbf{a}(t) = \mathbf{r}, \mathcal{K}) \cdot \\
&\qquad \Pr(\mathbf{o}(1) \ldots \mathbf{o}(t), C_1 \ldots C_{t-1}, \mathbf{a}(t) = \mathbf{r} \,|\, \mathcal{K}) && \text{(A.8)}\\
&= l_t(\mathbf{r}, C) \cdot \\
&\qquad \Pr(\mathbf{o}(1) \ldots \mathbf{o}(t), C_1 \ldots C_{t-1}, \mathbf{a}(t) = \mathbf{r} \,|\, \mathcal{K}) && \text{(A.9)}\\
&= l_t(\mathbf{r}, C) \cdot \Pr(\mathbf{o}(t) \,|\, \mathbf{a}(t) = \mathbf{r}, \mathcal{K}) \cdot \\
&\quad \Pr(\mathbf{o}(1) \ldots \mathbf{o}(t-1), C_1 \ldots C_{t-1}, \mathbf{a}(t) = \mathbf{r} \,|\, \mathcal{K}) && \text{(A.10)}\\
&= l_t(\mathbf{r}, C) \cdot f_{\mathcal{U}}(\mathbf{r}, \mathbf{o}(t)) \cdot \\
&\quad \Pr(\mathbf{o}(1) \ldots \mathbf{o}(t-1), C_1 \ldots C_{t-1}, \mathbf{a}(t) = \mathbf{r} \,|\, \mathcal{K}) && \text{(A.11)}\\
&= l_t(\mathbf{r}, C) \cdot f_{\mathcal{U}}(\mathbf{r}, \mathbf{o}(t)) \cdot \\
&\qquad \sum_{\boldsymbol{\rho} \in \mathcal{R}^N} \Pr(\mathbf{o}(1) \ldots \mathbf{o}(t-1), C_1 \ldots C_{t-1}, \\
&\qquad\qquad \mathbf{a}(t) = \mathbf{r}, \mathbf{a}(t-1) = \boldsymbol{\rho} \,|\, \mathcal{K}) && \text{(A.12)}\\
&= l_t(\mathbf{r}, C) \cdot f_{\mathcal{U}}(\mathbf{r}, \mathbf{o}(t)) \cdot \\
&\qquad \sum_{\boldsymbol{\rho} \in \mathcal{R}^N} \Pr(\mathbf{o}(1) \ldots \mathbf{o}(t-1), C_1 \ldots C_{t-1}, \\
&\qquad\qquad \mathbf{a}(t-1) = \boldsymbol{\rho} \,|\, \mathcal{K}) \cdot \\
&\qquad\qquad \Pr(\mathbf{a}(t) = \mathbf{r} \,|\, \mathbf{a}(t-1) = \boldsymbol{\rho}, \mathcal{K}) && \text{(A.13)}\\
&= l_t(\mathbf{r}, C) \cdot f_{\mathcal{U}}(\mathbf{r}, \mathbf{o}(t)) \cdot \\
&\qquad \sum_{\boldsymbol{\rho} \in \mathcal{R}^N} \alpha_{t-1}^{\mathcal{U}}(\boldsymbol{\rho}) \cdot p_{\mathcal{U}}(\boldsymbol{\rho}, \mathbf{r}) && \text{(A.14)}
\end{aligned}$$

In step (A.6)→(A.7) of the derivation, we apply the *chain rule*. In step (A.7)→ (A.8), we use *conditional independence*: given $\mathbf{a}(t) = \mathbf{r}$, the probability that the locations $\mathbf{r}$ can represent the reported C_t depends neither on the observations, nor on $\mathcal{K}$. In step (A.8)→(A.9), we use Definition (7). In step (A.9)→(A.10), we apply the chain rule and use conditional independence: given $\mathbf{a}(t) = \mathbf{r}$, $\mathbf{o}(t)$ does not depend on the past observations. In step (A.10)→(A.11), we use the fact that the location obfuscation process is applied independently for each user. In step (A.11)→(A.12), we apply the *law of total probability*, conditioning over all the possible actual locations $\boldsymbol{\rho}$ users could have been at, at time $t-1$. In step (A.12)→(A.13), we use the chain rule and conditional independence: given $\mathbf{a}(t-1) = \boldsymbol{\rho}$, $\mathbf{a}(t)$ does not depend on the past observations. In step (A.13)→(A.14), we use Definition (4). □

The proof of Equation (6) follows the same line of reasoning.

APPENDIX B
EFFECTS OF TRUE CO-LOCATIONS AND SPATIAL CLOAKING

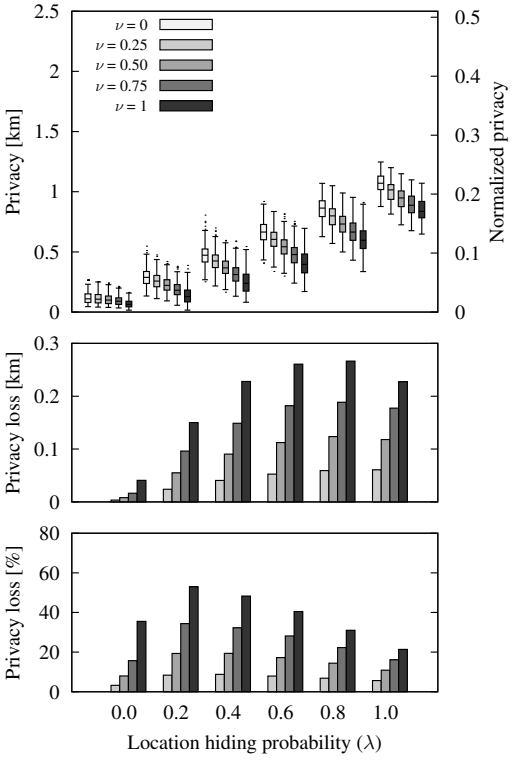


Fig. 10. Privacy (top), absolute privacy loss (middle) and relative privacy loss (bottom) for the limited user set attack with $N = 2$ users, when users do not report fake co-locations ($\mu = 0$) and use spatial cloaking or location hiding as protection mechanisms. The privacy *loss* is expressed w.r.t. the case where no co-locations are available ($\nu = 0$, $\mu = 0$); the histograms show median values.

Similarly to our experimental setup presented in Figure 5b, we evaluate user privacy for a different LPPM, namely location hiding (with probability λ) or spatial cloaking (with probability $1 - \lambda$). When using cloaking, a user does not report the region corresponding to her actual location, but instead a meta-region consisting of four regions, one of which is the actual location. In Figure 10 we present our results. We conclude that the proportion of reported true co-locations consistently decreases the location privacy of

the users (as was the case for the other LPPM based on location hiding and location obfuscation), but in this case the privacy loss is more evident. This could be explained by the fact that in the case of cloaking, when observing a meta-region of size four regions, the adversary has to explore four possible regions as candidates for the user's actual location; whereas in the case of obfuscation, five possible candidates for the actual location have to be explored (one of the four neighboring regions of the observed (obfuscated) region and the observed region itself).

APPENDIX C
EFFECTS OF THE DIFFERENCES OF INDIVIDUAL LPPM SETTINGS

In this section, we analyze the effect of the differences, in the users' LPPM settings, on the location privacy (loss) due to co-locations. To do so, we focus on the case of two users, a target and her co-target, both who obfuscate their locations but with different hiding probabilities λ_{target} and $\lambda_{\text{co-target}}$. We perform a joint optimal localization attack. The results are depicted in Figure 11 under the form of heat-maps that represent the target user's location privacy (a) as well as her absolute (b) and relative (c) privacy loss (with respect to the case $\nu = 0$) as functions of the respective LPPM settings $\lambda_{\text{co-target}}$ (x-axis) and λ_{target} (y-axis).

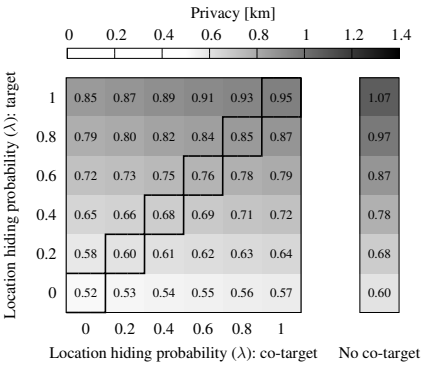
A first observation is that co-locations always decrease the privacy of the target (*i.e.,* all values in Figure 11b are positive) and that the more information the co-target discloses, the worse the privacy of the target is (*i.e.,* the cells of the heat-map depicted in Figure 11a become lighter, when going from right to left on a given row).

The diagonals of the heat-maps correspond to the case $\lambda_{\text{co-target}} = \lambda_{\text{target}}$, which is depicted in more detail in Figure 5. The region of the heat-map above the diagonal corresponds to the case where the target is more *conservative*, in terms of her privacy attitude, than her co-target (*i.e.,* $\lambda_{\text{co-target}} < \lambda_{\text{target}}$). It can be observed that the information disclosed by the target herself compromises her privacy more than the information disclosed by her co-target, *e.g.,* the cell (0.6,0) is lighter (which means that the target's privacy is lower) than the cell (0,0.6).

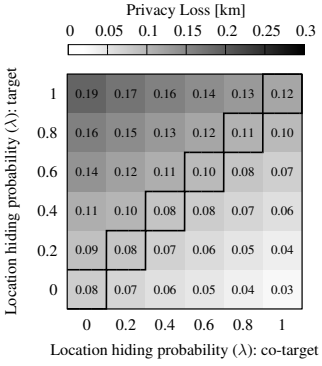
By comparing the columns "$\lambda_{\text{co-target}} = 1$" and "no co-target" (two right-most columns in Figure 11a), we can observe the privacy loss stemming from the use, through the co-location information, of the co-target's mobility profile alone (as the co-target never discloses her location). This is substantial. The intuition behind this result is that co-located users are likely to be at a place that is often visited by *both* of them, which narrows down the choice of locations the adversary needs to explore when localizing both users.

Finally, in the extreme case where the target never discloses location information and her co-target always does so (top-left cell of the heat-maps in Figures 11b and 11c), the privacy loss for the target is 190m, which corresponds to a decrease of 18%. This case (and in general the cases where the target never discloses location information, *i.e.,* the top row of the heat-maps) highlights the fact that, as reported co-locations involve two users, users lose some control over their privacy: Without revealing any information about herself, a user can still have her privacy decreased by other users, due to co-location information.

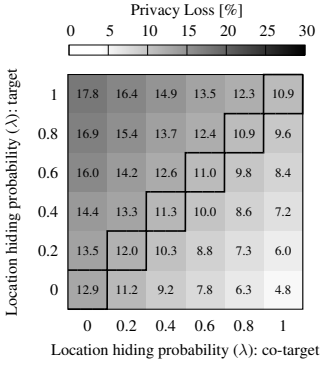
For the rest of the evaluation, we focus on the case where all users have the same LPPM settings (*i.e.,* same values of λ).



(a) Privacy



(b) Absolute privacy loss (w.r.t. to $\nu = 0\%$)



(c) Relative privacy loss (w.r.t. to $\nu = 0\%$)

Fig. 11. Median values of the target's location privacy (loss), for the limited user set attack with $N = 2$ users, when the target and her co-target have different values of λ (with obfuscation, $\nu = 0.5$, $\mu = 0$). The diagonals correspond to the values of Figure 5b.

APPENDIX D
CO-LOCATION INFORMATION ON A LARGER SCALE

In Section 7 and Section 5, we considered a small dataset of users, due to the high complexity of the optimal solution. We denote this small dataset by $\mathcal{U}_s$. Here, we evaluate our belief propagation solution on a larger dataset, in order to quantify location privacy loss when co-locations from a larger set of users are available. To this end, we select a subset $\mathcal{U}_l$ of users in the GeoLife dataset, such that each selected user

must have at least one *real* co-location[7] with any other user in $\mathcal{U}_l$ (across their full traces). This results in 38 users being selected. Note that $\mathcal{U}_s \subset \mathcal{U}_l$. We emphasize that due to the low availability of real co-locations across the GeoLife users, this represents a weaker constraint of minimum desired co-locations, compared to that which we use when sampling the users in our small dataset $\mathcal{U}_s$. The low availability of co-locations, coupled with the sparsity of the location information available, also motivates sampling 10 short *individual* collections of actual traces in the following way: For each u, a target user in $\mathcal{U}_l$, we generate actual traces for all the users in $\mathcal{U}_l$ such that (1) u has at least 10% of valid samples (*i.e.,* different from $r_\perp$) and u has at least 1 co-location with her co-target$_1$.

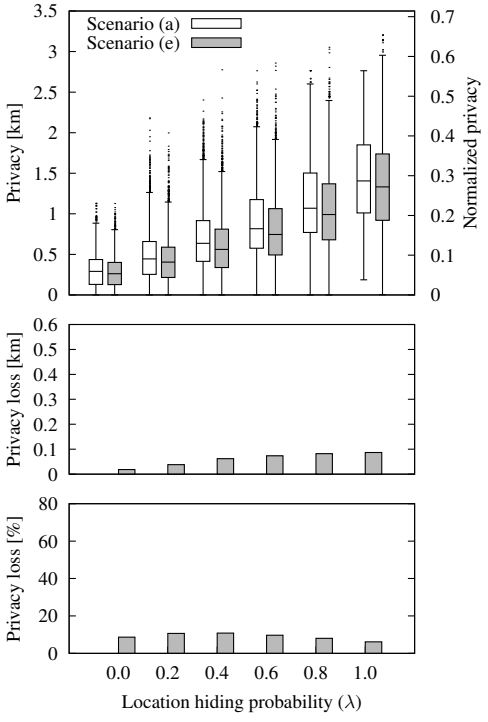


Fig. 12. Comparison of the localization attacks for target users in $\mathcal{U}_l$ on Scenarios (a) and (e), as depicted in Figure 6, with obfuscation. The privacy loss (middle and bottom) is evaluated w.r.t. Scenario (a). In Scenario (e), we consider users report true co-locations with probability $\nu = 0.5$ and that they do not report fake co-locations ($\mu = 0$).

We perform an individual localization attack by optimal inference for Scenario (a), considering, in turn, each user in the set $\mathcal{U}_l$ as the target user (using only their own reported locations and no co-locations). We then consider Scenario (e), the case of an adversary that exploits co-locations between any of the users in $\mathcal{U}_l$. We assume users report only a limited proportion of their true co-locations, with probability $\nu = 0.5$, and no fake co-locations ($\mu = 0$). We perform an approximate joint inference algorithm, by using the belief propagation algorithm with at most 20 iterations. We then compare the privacy in Scenario (e) to that in Scenario (a), in the case where all users use the same LPPM settings, *i.e.,* same value for λ and disclose only their obfuscated locations. Figure 12 shows the results of our comparison. It can be observed that, unsurprisingly, the users' privacy decreases with the amount of considered co-locations. The privacy loss can seem somewhat modest, in comparison to the one observed in our previous experiments using $\mathcal{U}_s$. This can be explained by the fact that users in $\mathcal{U}_s$ have more real co-locations than those in $\mathcal{U}_l$ (a user has a median number of real co-locations in their actual traces of 5.5 and 2, respectively). We further compare the privacy of only the target users from $\mathcal{U}_s$ (but still using all the co-locations in the larger dataset $\mathcal{U}_l$) with that when using co-locations among users from $\mathcal{U}_s$. Figure 13 shows the results of this comparison. It can be observed that the availability of co-locations with a larger number of users can further reduce privacy (privacy loss is as much as 31% when $\lambda = 0$).

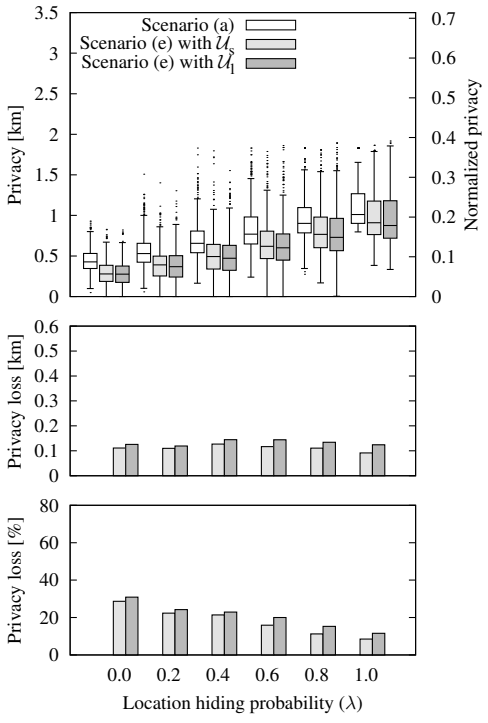


Fig. 13. Comparison of the localization attacks for target users in $\mathcal{U}_s$ on Scenario (a), Scenario (e) considering co-locations only with and among users in $\mathcal{U}_s$ and Scenario (e) considering co-locations with and among all users in $\mathcal{U}_l$. The privacy loss (middle and bottom) is evaluated w.r.t. Scenario (a). We consider users report true co-locations with probability $\nu = 0.5$, do not report fake co-locations ($\mu = 0$) and use obfuscation.

APPENDIX E
COMPARISON METRICS FOR THE ACCURACY OF THE DIFFERENT INFERENCE ALGORITHMS

We compare the approximate localization attack to the optimal localization attack, and we measure its accuracy by the average Hellinger and statistical distance between their output region distributions. Specifically, if h denotes the output of the optimal localization attack $\hat{h}$ that of the

7. Note that by real co-locations, we mean that the users are at the same location (i.e., their actual locations at a given time instant are the same), regardless of the fact that the co-location is reported or not.

approximate localization attack, then

$$\frac{1}{N \cdot T} \sum_{u \in \mathcal{U}} \sum_{t \in \{1,\ldots,T\}} \frac{1}{\sqrt{2}} \sqrt{\sum_{r \in \mathcal{R}} \left(\sqrt{h_t^u(r)} - \sqrt{\hat{h}_t^u(r)} \right)^2}$$

$$\frac{1}{N \cdot T} \sum_{u \in \mathcal{U}} \sum_{t \in \{1,\ldots,T\}} \frac{1}{2} \sum_{r \in \mathcal{R}} \left| h_t^u(r) - \hat{h}_t^u(r) \right|.$$

APPENDIX F
TABLE OF NOTATIONS

Table 1 summarizes the main notations used in our formalization throughout the paper.

TABLE 1
Table of notations.

$\mathcal{U}$	Set of mobile users
$\mathcal{R}$	Set of regions that partition the whole area
N	Number of users ($N = \lvert\mathcal{U}\rvert$)
M	Number of regions ($M = \lvert\mathcal{R}\rvert$)
T	Number of time instants
$p_u(\cdot,\cdot)$	Mobility profile of user u
$\pi_u(\cdot)$	The stationary distribution of p_u
$f_u(\cdot)$	Obfuscation function employed by user u
$g_{u,v}(\cdot,\cdot)$	Co-location reporting function for users u and v
$\mathcal{K}$	Adversary's background knowledge
$a_u(t)$	Actual location of user u at time t
$\mathbf{a}(t)$	Actual locations of all the users at time t
$u \,@_t\, r$	User u reports being in r at time t
$o_u(t)$	Obfuscated location of user u at time t
$\mathbf{o}(t)$	Obfuscated locations of all the users at time t
$u \leftrightarrow_t v$	A co-location was reported between u and v at time t
$c_{u,v}(t)$	Binary variable incorporating whether $u \leftrightarrow_t v$
C_t	Set of all reported co-locations at time t
C	Set of all reported co-locations