Enhancing Vehicular Anonymity in ITS: A New Scheme for Mix Zones and Their Placement

Nirupama Ravi, C. Mani Krishna, and Israel Koren, *Fellow, IEEE*

***Abstract*—Intelligent transportation systems (ITS) achieve improved throughput and safety using periodic vehicle-to-vehicle and vehicle-to-infrastructure wireless communication. Vehicles use pseudonyms which are frequently exchanged to protect against eavesdroppers using such messages for tracking. Such exchange takes place in *mix zones* where all wireless transmission is forbidden in order to prevent matching the new pseudonym of a vehicle with its previous one. Mix zones are not free: in addition to infrastructure cost, they impose a cost in terms of reduced vehicular throughput and disruption to vehicular communication. We present a scheme to manage traffic within a mix zone to make it more resilient to attacks against privacy. Second, we introduce a heuristic to place mix zones appropriately so that the gain in privacy is balanced against the cost in reduced throughput. We evaluate our schemes assuming a powerful attacker who has access to all wireless transmissions and uses a simple but powerful machine learning algorithm. Our algorithms are evaluated using detailed traffic simulations in two US cities: New York, NY, and Cambridge, MA.**

***Index Terms*—Location privacy, intelligent transportation systems, mix zone placement.**

I. Introduction

INTELLIGENT Transportation Systems (ITS) propose to use vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications together with Cooperative Adaptive Cruise Control (CACC) to reduce the gap between vehicles, thereby increasing road system throughput while maintaining driver-assisted applications such as lane changing and forward collision warning. A recent US-DoT study estimated that CACC can more than triple throughput over traditional Adaptive Cruise Control (ACC) [1].

V2V communication creates risks to privacy; it involves broadcasting, typically every 100 ms, a Basic Safety Message (BSM) which includes vehicle ID, time, location and speed [2], [3]. To prevent tracking by eavesdroppers, pseudonyms rather than identifiable vehicle IDs are used in BSMs [6]; to be effective, these pseudonyms have to be changed frequently in such a way that the new pseudonym cannot easily be matched with the old one [11]–[13]. Several such schemes have been proposed [4], [5].

Manuscript received July 2, 2019; accepted July 23, 2019. This work was supported by National Science Foundation under Grant CNS-1329831. The review of this article was coordinated by Prof. A. l. Grieco. *(Corresponding author: Nirupama Ravi.)*

Q1

The authors are with the Department of Electrical and Computer Engineering, University of Massachusetts Amherst, Amherst, MA 01003 USA (e-mail: nirupamaravi@engin.umass.edu; krishna@engin.umass.edu; koren@engin.umass.edu).

Digital Object Identifier 10.1109/TVT.2019.2936529

To support pseudonym exchange, first, each vehicle needs a large pool of available pseudonyms. This is easy to implement [4]. Second, we require some location where vehicles congregate and where pseudonyms can be effectively exchanged: parking lots [20] and traffic light intersections [7], [16], [18], [19], [26]. Of these, the latter are preferable; vehicles rarely park partway through their journey.

Pseudonym exchange zones, aka mix zones, are spatial regions where BSMs are not transmitted (and therefore CACC cannot be supported). Road Side Units (RSUs) monitor traffic incoming to mix zones and facilitate pseudonym exchange.

Mix zones incur costs. First, there is the infrastructure cost of the RSU itself. Second, since CACC is not supported inside the mix zone, there is a reduction of traffic throughput. We cannot therefore place a mix zone around every traffic intersection; they have instead to be placed sparingly to balance privacy benefits against cost. Such a placement problem, as well as the management of vehicles within a mix zone, is the focus of this paper.

Our contributions in this paper are twofold. First, an algorithm, called the Anonymity Enhancing Mix Protocol (AEMP), is introduced. This attempts to proactively alter the exit order of vehicles from the mix zone to enhance privacy. Second, we study the placement of mix zones using AEMP to deal with the tradeoff between privacy and cost. The increased resilience of this approach against privacy attacks by a well-resourced attacker is demonstrated.

This paper is organized as follows. Section II reports on related work. Section III describes the baseline approach against which we compare AEMP. Section IV presents the AEMP protocol while Section V discusses the placement of mix zones. Section VI describes a powerful adversary, used to study the resilience of AEMP. Simulation results are presented in Section VII for traffic in New York and Cambridge; the paper concludes with Section VIII.

II. Related Work

In order to protect each vehicle's anonymity, its pseudonym needs to be updated sufficiently often. A popular way to facilitate this is by means of mix zones, where pseudonyms can be changed; a good recent survey can be found in [5]. Mix zones may be dynamically and opportunistically set up by a group of vehicles that happen to find themselves in close proximity to one another [34]–[40] or statically sited at suitable locations (which is the approach of the present paper) [14], [17], [18],

[21], [27], [33]. In the static approach, parking lots and traffic light intersections have been suggested as suitable locations for mix zones. To increase anonymity, virtual vehicles have also been suggested when mix zones have low traffic intensity: these vehicles emit wireless messages and change pseudonyms just like real vehicles [19]. The extent of the involvement of the RSU in pseudonym exchange varies from one approach to the next.

In all the above context-based and infrastructure-based schemes, an attacker can use the order of arrival and departure information to match old and new pseudonyms effectively. Opportunistically altering such order to weaken the correlation between entering and departing vehicles is central to the anonymity enhancing protocol introduced in this paper.

As all traffic light intersections cannot be mix zones, several works propose algorithms to place mix zones in a road network optimally. A survey by Primault [27] summarizes several schemes under various architectures to protect location privacy of mobile users using location-based services. Some of the mix zone placement algorithms from the survey are relevant to vehicular networks. Jadliwala *et al.* propose minimizing the mix zone placement cost as a criterion for selecting mix zones in a road network [26]. Liu *et al.* place mix zones which yield the least linkage between the point of interests corresponding to a mobile user's path [28]. Freudiger *et al.* propose to place mix zones based on traffic flow while limiting the disruption of communication in mix zones [29]. Palanisamy *et al.* recommend placing mix zones in gridded regions, which maximize the distance between two consecutive mix zones or place mix zones at high traffic density [30]. Sun *et al.* advocate placing mix zones per sub-region so that vehicles may choose to reroute to traverse through mix zones to gain anonymity [31]. Menon *et al.* employ a genetic algorithm to minimize the number of RSUs by placing mix zones in such a way that the connectivity of mobile users with an RSU is not disturbed [32].

Our approach recommends placing mix zones which benefit a majority of vehicles in gaining anonymity. We estimate anonymity contributed by each mix zone based on traffic flow statistics. We utilize well known genetic [41] and annealing [42] optimization algorithms to place mix zones at intersections with high traffic flows. We also propose dividing regions into sub-regions based on the concentration of privacy-sensitive points of interest, such as health service centers and hospitals. For a given budget of mix zones, we evaluate the tradeoff between privacy gain in the sub-region due to the placement of more mix zones with the loss of privacy for the rest of the region.

III. Baseline Algorithm

The following protocol is used as Baseline within this paper. Fig. 1 shows a mix zone within a traffic light intersection. An RSU broadcasts the mix zone dimensions to all the vehicles within range. Vehicles observe silence starting d_{before} meters before the intersection while entering, and until d_{after} meters after exiting the intersection. Each vehicle broadcasts its BSM before entering and after exiting the mix zone.

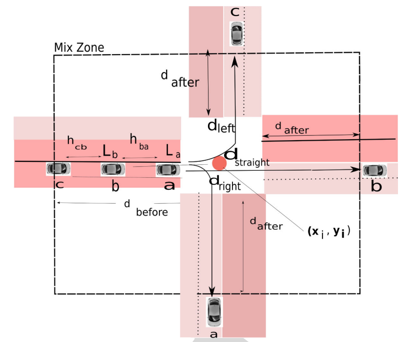


Fig. 1. Mix zone.

While maintaining silence, each vehicle changes its pseudonym and other physical and logical IDs of each layer of the protocol stack. Once the vehicle exits the mix zone, it starts broadcasting its BSM with its new pseudonym. An attacker would try to break anonymity by successfully matching the new pseudonym to the previous one.

The entry and exit timings to a mix zone can be used by an attacker to link the new pseudonym of a vehicle to its previous one. The simplest case is that of a unidirectional lane which permits no change; the exit order of vehicles is the same as their entry order, and matching new and old pseudonyms to the same vehicle is simple. When there are multiple turn options from a given lane, if the delays associated with making such turns are distinctive (e.g., a different delay for a left turn compared to a right turn or going straight), this information can be used by the attacker along with information on vehicle mix zone entry and exit times to improve the probability of correctly matching the new pseudonym of a vehicle to its previous one.

IV. Anonymity Enhancing Mix Protocol (AEMP)

System Model: We assume the following entities in the intelligent transportation system. RSUs located near mix zones monitor traffic at each intersection and broadcast mix zone information. We assume that each vehicle has: a) sufficient pseudonyms for the entire journey so that there is no repeated use of pseudonyms and b) capability to use CACC and ACC technologies as required.

AEMP Mix Zone: In contrast to traditional mix zones, AEMP takes active steps to make it more difficult to match vehicles going out of the intersection to those coming in. Prior to entering a mix zone, some vehicles may randomly switch lanes to obscure their intended outgoing direction. For example, a vehicle intending to go right may switch to a lane which turns left or goes straight. Once in the mix zone, vehicles also alter their speed. Each vehicle selects a target speed randomly from a given speed range, say [2,13] m/s. It then tries to maintain its actual speed as close to the target as possible, subject to traffic constraints. This random exit speed leads to some vehicles overtaking others, thereby scrambling the order of vehicles exiting the intersection. Such scrambling is not free: it affects traffic flow and tends to reduce vehicle throughput. We consider

TABLE I
PROBABILITY OF TRACKING FOR THREE STRATEGIES

Speed (m/s) Range	Lane Switching before Mix Zone	Probability of Tracking
[2,13]	Yes	0.65
[2,13]	No	0.75
13 (Baseline)	No	1

this cost when determining whether it is worth placing a mix zone at a particular location.

Table I shows the results for a right only turn in the intersection shown in Fig. 1: the speed range of [2,13] m/s was selected based on its performance in simulation experiments. Note that even for a lane allowing only right turns, our approach provides some privacy protection. Vehicles choose maximum allowed speed to exit the intersection. Therefore, Baseline uses 13 m/s speed (as permitted given the traffic) to exit the intersection.

Three factors determine the effectiveness of this approach at maintaining anonymity: the mix zone geometry, traffic intensity, and turn ratios.

The impact of the mix zone geometry on anonymity depends on the mix zone size and the number of lanes in and out. The bigger it is, the more lanes are coming in and going out, and the higher the number of potential output directions for vehicles flowing into it, the more difficult it is for an attacker to match vehicles coming into it with those emerging from it.

The traffic intensity and vehicle turn ratios are also significant factors. The higher the intensity, the greater the potential for scrambling the vehicle order coming out of the mix zone: there are more vehicles to mix. (At an extreme, if we only have one vehicle in the entire mix zone at any point, matching the emerging vehicle with the one coming in is trivial.) Equally, the more even the turn probabilities, the less well the attacker is expected to be able to do. For example, an intersection where 99% of the traffic goes straight will be easier to attack than one where a third of the traffic goes straight, right and left, respectively.

These factors, together with the range of speeds allowed within the mix zone and percentage of vehicles which can change lanes outside the mix zone, determine the number of vehicles which can gain anonymity.

Performance Metrics: We use the probability of successfully tracking pseudonyms of vehicles as a privacy metric to measure the performance of a mix zone. Note that to track successfully, the adversary needs to track it successfully across each mix zone that it passes through. Anonymity (or privacy) is the complement of tracking probability.

Let vehicle v traverse a route R that has n_v intersections $I_1, I_2, ... I_{n_v}$ with n mix zones. Denote the probability of successfully tracking vehicle v traversing a mix zone I_j by p_v^j. If there is no mix zone in intersection I_j then $p_v^j = 1$. The probability of successful tracking of a vehicle by the end of its route R is calculated as

$$p_v = \prod_{j=1}^{n_v} p_v^j \tag{1}$$

We measure traffic efficiency considering a) the average trip delay and b) the average throughput of an intersection. We measure trip delay as the lag accrued by a vehicle due to traversing mix zones. The throughput of a lane L_j during an i^{th} interval of Δt minutes with $N_{L_j}^i$ vehicles traversed through the lane is given by $N_{L_j}^i/\Delta t$. The average throughput q_{L_j} observed during a set of k equally spaced intervals of Δt minutes is given by Equation (2).

$$q_{L_j} = \frac{1}{k}\sum_{i=1}^{k}\frac{N_{L_j}^i}{\Delta t} \tag{2}$$

If an intersection, I, has m outgoing lanes then its average throughput is the sum of the throughput of each lane:

$$q(I) = \sum_{j=1}^{m} q_{L_j} \tag{3}$$

V. MIX ZONE PLACEMENT ALGORITHM

We cannot afford to place a mix zone in each traffic intersection. Deciding where to place a limited number of mix zones requires resolving the tradeoff between the capital cost of a mix zone, its impact on traffic delays, and its contribution to anonymity.

We use simulated annealing and genetic algorithms for placement. To be effective, such algorithms require a good initial starting point that they can then proceed to refine iteratively. We obtain such a starting point based on a measure of *mixability* which can be calculated as follows for each intersection.

We first estimate the traffic flows through the region. We then calculate the anonymity gained per vehicle per intersection along its route, as shown below. Let V be the set of vehicles in our traffic database. For each $v \in V$, the sequence of intersections visited is denoted by $(I_1^v, I_2^v, \ldots, I_{n_v}^v)$ where n_v is the number of intersections in the route of vehicle v. Let the action (e.g., *left_turn*) taken by a vehicle in intersection I_j^v be a_j^v. This action controls the direction in which the vehicle leaves that intersection. Note that I_{k+1}^v is determined by I_k^v and a_k^v, so the intersection and action sequences are not independent of one another.

Based on the routes taken by each vehicle in V, we calculate the fraction of vehicles passing through the intersection I which take a particular action a at that intersection; denote this fraction by $\pi(I, a)$. We denote by $h_{\text{in}}(v, I_k^v)$ the probability of successfully tracking a vehicle v up to just before entering the intersection I_k^v. $h_{\text{out}}(v, I_k^v)$ is the probability of successfully tracking a vehicle v just after it exits that intersection.

Now, define the following recursions:

$$\begin{aligned}
h_{\text{in}}(v, I_1^v) &= 1 \\
h_{\text{in}}(v, I_k^v) &= h_{\text{out}}(v, I_{k-1}^v) \text{ for } 1 < k \leq n_v \\
h_{\text{out}}(v, I_k^v) &= h_{\text{in}}(v, I_k^v)\pi(I_k^v, a_k^v) \\
\Delta_h(v, I_k^v) &= h_{\text{out}}(v, I_k^v) - h_{\text{in}}(v, I_k^v)
\end{aligned}$$

Algorithm 1: Optimal Mix Zone Placement Algorithm.

Input:
Budget b mix zones,
N Sorted list of mix zones based on mixability
Output: S Best mix zone set
$S1 \leftarrow$ b mix zones at $N[1:b]$
if $\|N\| \leq 2b$ **then**
  S $\leftarrow$ SimulatedAnnealing($b, N, S1$)
else
  $S2 \leftarrow$ b mix zones at $N[b+1:2b]$
  S$\leftarrow$ GeneticAlgorithm($b, N, S1, S2$)
end if
return S // The best mix zones

The *mixability* of intersection I, $M(I)$, is defined as

$$M(I) = \sum_{v \in V} \Delta_h(v, I) \tag{4}$$

$M(I)$ is used to determine the usefulness of initially placing a mix zone in that intersection. In other words, suppose we number the N available intersections in descending order of mixability: $(I_1, I_2, \ldots, I_N)$. If we aim to use μ mix zones in all, we will start our optimization process by initially placing mix zones at intersections $I_1, \ldots, I_\mu$. Simulated annealing or genetic algorithms are then used to iterate from this starting allocation.

The intuition behind the mixability metric is as follows. For a given vehicle, the gain in anonymity as it goes through a mix zone in intersection I is a function of the anonymity it already possesses and the factor by which this anonymity increases as a result of passage through I. An approximate proxy for the anonymity of vehicle v entering the first intersection on its path, I_1^v (i.e., as it starts its journey), is 0, its complement, measured by $h_{\text{in}}(v, I_1^v) = 1$. That is the basis of the recursion for each vehicle, v. The anonymity is assumed to change only at mix zones since those are the only places where the pseudonyms change. So, if the vehicle moves from the intersection I_k^v to I_{k+1}^v, its anonymity going out of I_k^v is the same as that going into I_{k+1}^v. We use $\pi(I_k^v, a_k^v)$ as a proxy for the factor by which the complement of the anonymity decreases. Calculating along these lines provides us an indication of how the anonymity of vehicle v changes as it progresses along its route. For each intersection, adding up the contribution to the anonymity of each vehicle that passes through it gives us a rough measure of how useful this intersection as a mix zone. This measure is then used as noted above to set up an initial allocation for the second, optimization step.

Our mix zones placement algorithm, shown in Algorithm 1 consists of three steps. In the first step, we calculate the mixability of all intersections and sort the intersections according to their mixability, N. In step 2 the initial placement is done by placing mix zones in the b intersections (input as budget) with the highest mixability. In the 3^{rd} step we use a standard optimization algorithm (e.g., simulated annealing or a genetic algorithm), to improve on the initial placement, thus obtaining the final placement, S. If the total number of intersections $N \leq 2b$, we select the simulated annealing algorithm otherwise the genetic algorithm (GA) as GA needs two sets of non-overlapping mix zones of size b [41].

VI. The Adversary

We assume a powerful attacker (adversary) who is capable of listening to every broadcast in the vehicular networks as the attacker's ability to track pseudonyms increases with the increase in the attacker's capability to listen to vehicular broadcasts [14]. Furthermore, we assume that the attacker knows the AEMP protocol, the location, and sizing of mix zones, the traffic signal timings, and overall traffic statistics.

The attacker is assumed to use the Random Forest (RF) algorithm [43] to track pseudonym changes. RF is a powerful algorithm able to effectively tease out relationships between parameters and is widely used in numerous fields, including banking, medicine, and e-commerce. We thus choose a formidable adversary against which to test our approach.

VII. Evaluation of AEMP

We evaluate the performance of AEMP at two levels and compare it to the Baseline algorithm in each case. First, we consider mix zones in isolation (i.e., without being part of a network) under the following physical and traffic factors: a) size of the mix zone; b) traffic arrival rate; c) traffic flows through an intersection and d) the intersection's physical configuration in terms of number of turns per incoming lane and number of available lanes into and out.

Second, we consider a road network consisting of many mix zones. We measure the cumulative anonymity gained, the cumulative loss of throughput for all the mix zones, and the average trip delay. Our results show that AEMP's performance on an average is significantly better and that a relatively small number of mix zones suffice to obtain desired anonymity, at only a marginal cost of an increase in trip delay. We also show the results of our mix zone placement algorithm by computing the maximum percentage of vehicles with desired anonymity for a range of 10 to 100 mix zones.

Simulation Setup

We use the widely used SUMO [23] simulator for generating vehicular traffic in a single intersection and in city road networks. We use the road networks of mid-town Manhattan, NY, and Cambridge, MA. We use DUAROUTER, available within SUMO, to generate mobility traces (routes) of vehicles. We use SUMO logs which consist of location, speed, and acceleration of each vehicle as recorded BSMs available for an attacker. The SUMO logs within a mix zone are deleted to maintain silence within a mix zone. This approach has been used widely elsewhere, e.g., in [8], [12], [14], [29], [30].

In SUMO, each vehicle uses a car-following model to control its speed and acceleration. We use the CACC car-following model [24]. Typically, in a connected vehicle, the CACC module of each vehicle broadcasts a BSM. Each BSM consists of speed, acceleration, and location of the vehicle. Each vehicle's control unit adjusts its speed and acceleration based on the BSMs of its

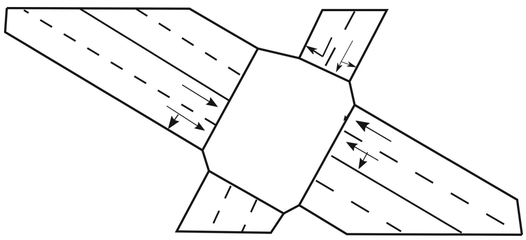
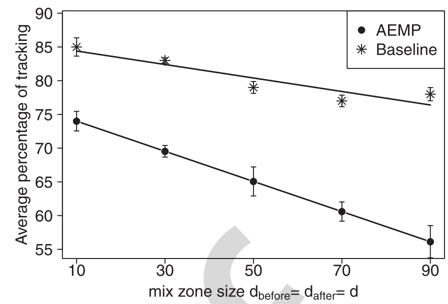
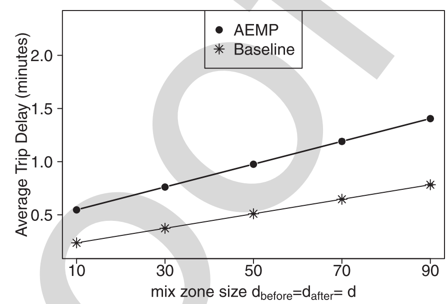


Fig. 2. A single intersection.



(a) Successful pseudonym tracking vs. mix zone size



(b) Trip delay vs. mix zone size

Fig. 3. Mix zone size.

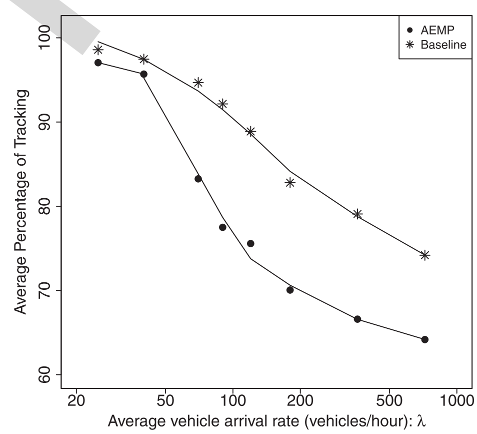


Fig. 4. Pseudonym trackability vs. arrival rate.

neighboring vehicles (we enhanced the CACC module in SUMO such that each vehicle obtains the speed, acceleration, and location of all the line-of-sight vehicles at each simulation step). The enhancement further enables the formation of platoons of cars. In order to form a platoon, each vehicle periodically determines a lead vehicle and a preceding vehicle. A vehicle becomes part of a platoon, when either a lead vehicle is available or when there is a preceding vehicle, which is already part of a platoon. Typical platoons are 2 to 5 vehicles long. The vehicles in the platoon maintain a headway time of 0.6 sec. Within a mix zone, since vehicles remain silent, there is no information from neighboring vehicles and platoons break down. Then, all vehicles follow the ACC car-following mode with a headway time of 2 sec. Once a vehicle comes out of a mix zone, it rebroadcasts BSMs and forms platoons when feasible. We do not simulate RSUs as we provide the mix zone information *a priori* to the vehicles.

Impact of Intersection Geometry and Traffic Factors

Consider an intersection with six incoming lanes and seven outgoing lanes, as shown in Fig. 2. The lane length of the intersection is 100 m on each side. We simulate a standalone intersection, to begin with, and then multiple intersection types. The assumed average vehicle arrival rate is 5 vehicles/sec/lane. For each data point, we simulated two data sets - a) training; and b) testing data sets. We used 90% of the simulated data for training and the remaining 10% for testing. We obtained test data by simulating vehicular traffic for 20 hours in total with nearly 14,000 vehicles.

Size: Fig. 3(a) shows the probability of pseudonym tracking as a function of mix zone size, assuming that $d_{\text{before}} = d_{\text{after}} = d$.

We can observe that for all mix zone sizes, AEMP provides higher anonymity compared to the Baseline. At smaller mix zone sizes such as 10 m $\times$ 10 m, AEMP is only 12% better than Baseline. However, as the mix zone size increases the entry-exit order of vehicles becomes less predictable for AEMP, resulting in AEMP anonymity around 25% higher than Baseline. In the Baseline approach, most vehicles retain their entry-to-exit order irrespective of an increase in size. For this reason, Baseline exhibits only small gains in privacy with the increase in mix zone size; these gains are because a small percentage (around 10%) of vehicles switch lanes within the mix zone.

Fig. 3(b) shows the average trip delay as a function of mix zone size. In AEMP, vehicles sometimes deliberately switch lanes to conceal their turn, which results in higher delays to merge back into the correct lane; this increases delays over the Baseline algorithm.

Impact of Vehicle Arrival Rate: Fig. 4 illustrates the impact of the average arrival rate λ (the average number of vehicles entering a mix zone from each direction within a given time) on the average anonymity of vehicles. The mix zone size is 50 m $\times$ 50 m positioned at the center of the intersection. The total simulated traffic time for the test data, at each arrival interval, for each protocol, is 20 hours.

At a low vehicle arrival rates with (50) vehicles/hour or less, there is not much mixing. Therefore, both AEMP and the Baseline protocol tend to retain FIFO order, and hence provide almost no anonymity. As the arrival rate increases to more than 50 vehicles/hour, for both AEMP and Baseline, successful tracking drops rapidly. When the arrival rate of vehicles is 200 vehicles/hour or more, the percentage of successful pseudonym tracking for both two protocols further decreases to 70% for AEMP and 80% for Baseline.

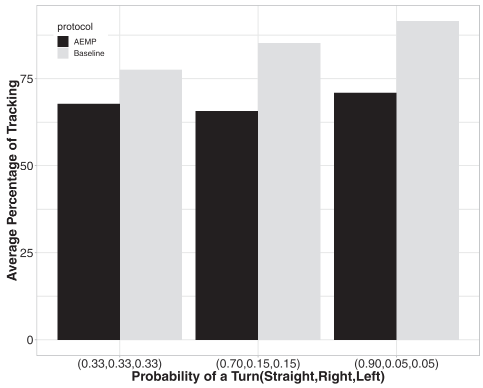


Fig. 5. Pseudonym trackability vs. traffic flow pattern.

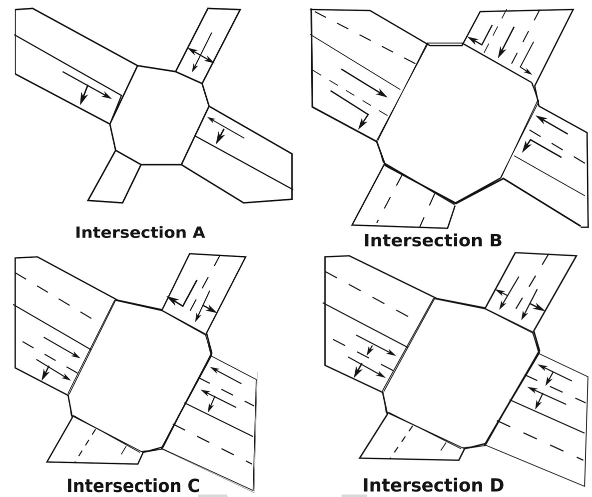


Fig. 6. Intersections simulated with varying physical characteristics.

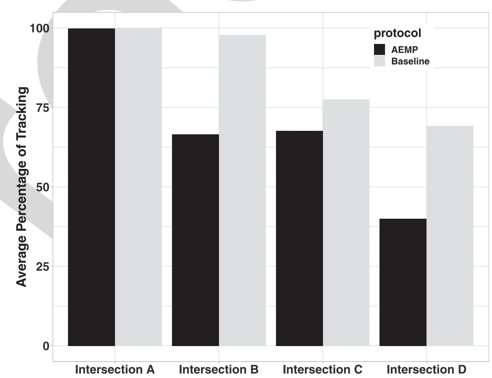


Fig. 7. Pseudonym trackability vs. physical characteristics of an intersection.

Impact of Traffic Flow Pattern: Fig. 5 shows how traffic flow patterns influence the probability of tracking of vehicles. The mix zone size is 50 m $\times$ 50 m positioned around the center of the intersection. The speed range allowed on the outgoing lanes is [2, 13] m/s. The average arrival rate is one vehicle every 5 seconds per incoming lane. The total simulated time for the test data is 20 hours per traffic pattern per protocol. The turn pattern (s, r, l) represents the probability of vehicles turning to an outgoing lane; s probability of going straight, r probability to turn right and l probability to turn left.

In the case of (0.33, 0.33, 0.33), traffic flows uniformly through all the outgoing lanes from an oncoming lane; this results in a lower probability of tracking compared to all the other turn patterns. As the traffic flows skew towards one direction, the probability of tracking increases much more for Baseline compared to AEMP. AEMP still performs better in all the three cases as tracking becomes more difficult due to the random delays of vehicles exiting the mix zone irrespective of their turn.

Impact of Intersection Physical Characteristics: The physical characteristics of an intersection include three primary parameters: a) an average number of possible turns from a lane (e.g., a right-turn-only lane has only one possible turn), b) number of incoming lanes, and c) number of outgoing lanes. We consider four intersections, as shown in Fig. 6.

Intersection A has an average of $(3 + 2 + 2)/3 = 2.33$ turns per lane with a single incoming and single outgoing lane. The remaining intersections all have multiple incoming and outgoing lanes. Intersection B has only a single turn per lane, Intersection C has an average of 1.5 turns per lane, and Intersection D has an average of 2 turns per lane. The mix zone size is 50 m $\times$ 50 m, positioned around the center of the intersection. The average arrival rate is one vehicle every 5 seconds per incoming lane. The total simulated time for the test data is 20 hours per traffic pattern per protocol.

Fig. 7 shows the average probability of tracking for both the protocols for each of the intersections.

For Intersection A, based on the exit times on the outgoing lanes, the Random Forest algorithm can guess the turns very effectively. In each outgoing lane, the FIFO order is retained for both Baseline and AEMP. In the case of AEMP, the lack of multiple outgoing lanes results in no mixing, and thus we see zero anonymity. Intersection B is a commonly seen geometry. As the entry and exit order here is again FIFO, Baseline provides no anonymity. However, AEMP retains its advantage over Baseline due to the intentional mixing that the protocol enforces. For Intersection C, AEMP's performance is better than the Baseline. For Intersection D, both the protocols have lower tracking compared to their respective tracking in the other intersection types, but AEMP still has the lowest tracking. It is primarily due to the ideal physical attributes of an intersection suitable for a mix zone which include a high number of turns per lane and multilane availability on both inlet and outlet of the mix zone.

Mix Zone Placement

As mentioned earlier, we use the mixability metric to guide us in the initial placement of mix zones, as an input to standard optimization algorithms such as simulated annealing and genetic. The problem is how to best place a limited budget of mix zones among the intersections of a city. In this simulation study, we consider two road networks: a mostly gridded road network (midtown Manhattan, NY) and a non-gridded road network (Cambridge, MA). A gridded road network typically consists of intersections with similar physical characteristics. The non-gridded road network contains intersections with varied

(a) MidTown Manhattan (b) Cambridge, MA

Fig. 8. Mix zones in road networks: high impact (★) and low impact (⬟).

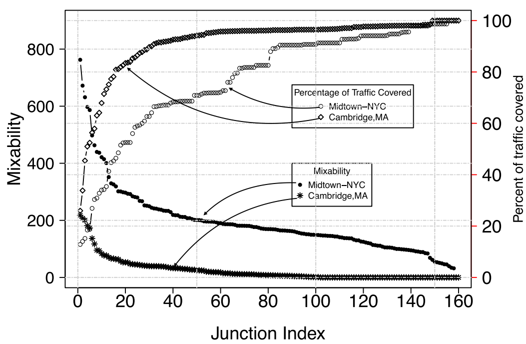
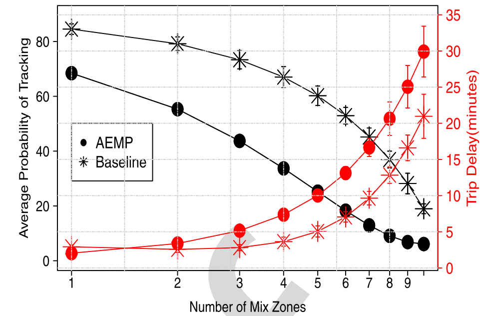
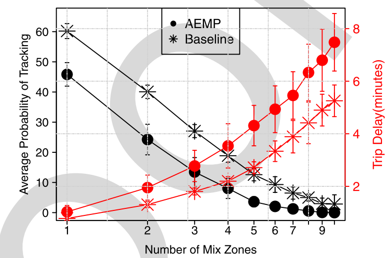


Fig. 9. Intersection selection based on mixability of intersections.



(a) Midtown Manhattan



(b) Cambridge, MA

Fig. 10. Probability of tracking and trip delay vs. number of mix zones traversed.

physical characteristics. Our purpose was to test the placement algorithm for diverse road network scenarios.

We first executed the optimization algorithms to obtain optimal mix zones for various budgets. We then study various tradeoffs which help determine the desired number of mix zones for a given region: a) privacy vs. trip delay, b) percentage of anonymous vehicles vs. road network throughput, and c) privacy in a sub-region to focus anonymity in the privacy-sensitive region.

The key details of our two simulation studies; a) the regions covered for Manhattan (3 km $\times$ 3 km) and Cambridge (2.6 Km $\times$ 2.9 Km) are shown in Figs. 8(a) and 8(b) b) Traffic Light Intersections are 184 and 350 c) Number of vehicles simulated are 10,400 and 5000 and d) Total simulation time is 180 min and 120 min for Manhattan and Cambridge respectively. Traffic intensities for arterial roads in Manhattan and Cambridge were 200–300 and for non-arterial roads 10–50 vehicles per hour. We generated congestion-free traffic at most intersections using dynamic user placement (DUAROUTER) available as part of SUMO tools. Training data was four times the test data. The default intersection size was 50 m $\times$ 50 m; if an intersection had lanes shorter than 50 m, we reduced the mix zone size accordingly.

Mixability and Traffic Covered: Fig. 9, which plots the mixability of each intersection, helps to determine the number of mix zones needed to cover the desired percentage of traffic and achieve privacy.

Fig. 9 shows similar broad trends for the two road networks: a) intersections with high mixability are few, b) the value of mixability decreases rapidly with the increase in intersection rank, and c) The percentage of traffic covered increases rapidly with the number of intersections. To cover 90% or more of the traffic, Cambridge network needs at least 50 intersections and Midtown Manhattan network needs 100 intersections.

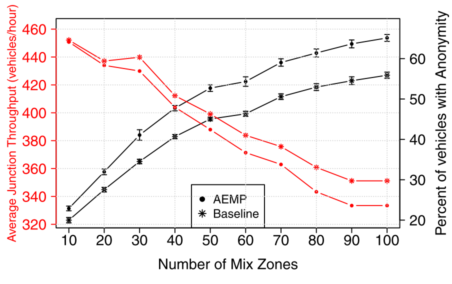
Based on these data, it is reasonable to place mix zones at the top 100 intersections as the remaining mix zones have low mixability.

Placement of Top Mix Zones: Fig. 8(a) shows the top and bottom mix zones in the Midtown-Manhattan map. The top mix zones are at the intersections at 9th avenue and 47th street, and 9th avenue and 42nd street as they have the desired level of traffic with physical characteristics of an ideal mix zone. Similarly, Fig. 8(b) shows the top and bottom mix zones in the Cambridge map. The top mix zone is at the intersection of River Street and Broadway.

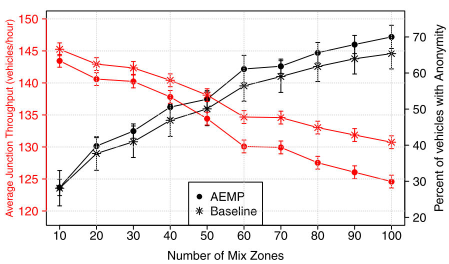
Other top mix zones include intersections placed at the beginning of traffic flows, with multiple inlets and outlets and optimal delay characteristics. As also seen in midtown Manhattan, the low mixability intersections are along the side roads with low traffic, imperfect intersection characteristics such as unidirectional traffic flow, and a low number of available turns per inlet.

Privacy vs. Trip Delay: Fig. 10 shows the tradeoff between cumulative anonymity gained by vehicles, computed using Equation (1), and additional trip delay accrued due to traversing multiple mix zones for both the cities considered.

The following trends (the first two are expected trends) are common to both the road networks: a) the average probability of successful tracking decreases (anonymity increases) while traversing more mix zones, b) the additional trip delay increases with an increase in the number of mix zones traversed, c) in AEMP, mix zones gain anonymity significantly more than in Baseline and d) the gain in anonymity in AEMP comes at the cost of additional trip delay.



(a) Midtown Manhattan



(b) Cambridge,MA

Fig. 11. Average anonymity and intersection throughput vs. number of mix zones.

Figure 10(a) shows that in midtown Manhattan, after traversing a single mix zone using AEMP, fewer than 25% of the vehicles are tracked as compared to about 50% for the Baseline algorithm. Since tracking probability drops geometrically with each mix zone traversed, it takes passing through just a handful of mix zones for AEMP to provide a significant level of privacy. This gain in anonymity costs only a four-minute increase in average trip time: from 22 minutes to 26 minutes. For Cambridge, results are shown in Figure 10(b); here, again, AEMP performs notably better than the Baseline.

Privacy vs. Throughput: Mix zones degrade throughput. It is therefore important to understand the tradeoff they present between the gain in the anonymity of vehicles and a loss of average throughput.

We compute the throughput of a mix zone using Equation (3). Fig. 11 highlights this tradeoff for both midtown Manhattan and Cambridge, MA. As the number of mix zones increases, the average probability of successful tracking drops, with AEMP providing better anonymity than the Baseline. However, AEMP loses somewhat more throughput than does the Baseline. The trends for both cities are very similar, despite their quite different road network topologies.

Mix Zone Placement in Subregions: A city road network can be divided into sub-regions which have sensitive points of interest such as cancer treatment centers, recovery centers, religious centers, consulates or any other space which exposes personal information by the knowledge of an individual's visit to such a location. It is desirable to have very high privacy for individuals visiting such privacy-sensitive points of interest. The mix zone placement algorithm can be used to increase the privacy of vehicles within such sub-regions by increasing mix zone density in such subregions.

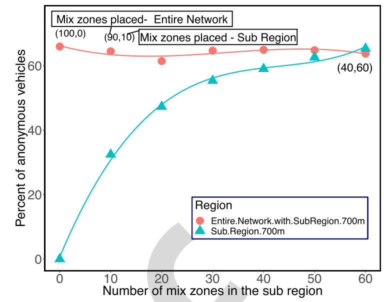
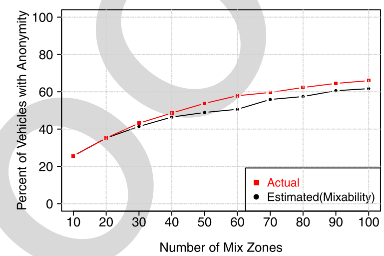
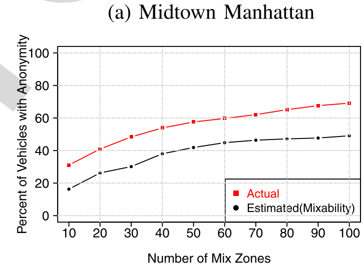


Fig. 12. Percent of anonymous vehicles vs. number of mix zones with and w/o focusing on the sub-region.



(a) Midtown Manhattan



(b) Cambridge, MA

Fig. 13. Estimated and actual impact vs. number of mix zones.

As an example of this process, we select a sub-region of 700 m located at the center of busy midtown Manhattan. We divide the 100 mix zones between the sub-region and the entire network. First, we optimally place 10 to 60 mix zones within this subregion. Then we place the remaining (90 to 40) mix zones in the rest of the road network. Fig. 12 compares the percent of anonymous vehicles with optimal placement of mix zones within this sub-region with that of mix zone placement in the entire area disregarding the sub-region.

When we place all the 100 mix zones optimally for the entire network, 66% of vehicles are anonymous. With the increase of the mix zones in the sub-region from 10 to 60 mix zones, the privacy of vehicles within the sub-region increases from 30% to 65%. It did not adversely impact the privacy of vehicles in the rest of the region. As the sub-region is a busy region of the network, it contributed to the privacy of vehicles passing through this sub-region.

Improvement due to Search Algorithms: Fig. 13 compares the percentage of anonymous vehicles for the top n intersections selected based solely on the mixability criterion vs. the top n intersections resulting from optimal search algorithms. The data

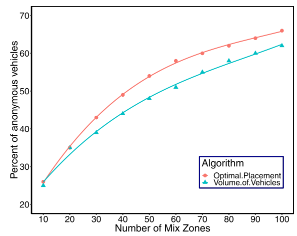


Fig. 14. Percent of anonymous vehicles vs. mix zone selection.

indicate that mixability, a heuristic measure, is a good starting point for optimal search.

Anonymous Vehicles vs. Mix Zone Selection Criteria: Fig. 14 compares the percentage of anonymous vehicles in the Midtown-Manhattan road network for two algorithms: 1) the top n intersections resulting from our optimal search of mix zones and 2) the top n intersections selected from intersections ordered based solely on the volume of vehicles passing through the intersection per hour.

We observe that at a low mix zone budget of 10 or 20, selection based on traffic volume is about as good as from optimal search. Optimal search shows a clear advantage only for a larger budget of mix zones. It is likely that this is because traffic volume ignores traffic flow patterns: an intersection positioned towards the end of most vehicle journeys (by which time most vehicles have already been anonymized) will not contribute much to anonymity even if it experiences high traffic.

VIII. Conclusion

The AEMP algorithm presented in this paper gains anonymity by intentionally disrupting FIFO order of vehicles at mix zones. The price for this is reduced traffic throughput; however, simulation studies have indicated that such a reduction is quite small.

For placing mix zones appropriately among traffic light intersections, we have introduced a mixability metric that is easy to compute and which captures how well a given intersection will perform as a mix zone.

This work is currently being extended in two directions. First, vehicle behavior may change in response to mix zone placement; the impact of that needs to be studied. Second, traffic patterns change with time-of-day or day-of-week, and so mix zones may have to be dynamically switched on or off based on detected changes in the traffic pattern.

Acknowledgment

The authors would like to thank the referees for their careful reading of the draft manuscript and their valuable suggestions.

References

[1] Cooperative Adaptive Cruise Control: Human Factors Analysis, Washington, DC, USA: U.S. Department of Transportation.

[2] "Intelligent Transport Systems (ITS); Security; Security Header and Certificate Formats," ETSI, Sophia Antipolis, France, Tech. Rep. ETSI TS 103 097 v1.1.1, 2013.

[3] "Dedicated Short Range Communications (DSRC) Message Set Dictionary," SAE Int., Warrendale, PA, USA, Tech. Rep. SAE j2735 v1.1.1, 2009.

[4] J. Petit, F. Schaub, M. Feiri, and F. Kargl, "Pseudonym schemes in vehicular networks: A Survey," in *Proc. IEEE Commun. Surv. Tut.*, Jan.–Mar. 2015, pp. 228–255.

[5] Abdelwahab Boualouache, Sidi-Mohammed Senouci, and Samira Moussaoui, "A survey on pseudonym changing strategies for vehicular ad-hoc networks," *IEEE Commun. Surv. Tut.* vol. 20, no. 1, pp. 770–790, Jan.–Mar. 2018. [Online]. Available: https://arxiv.org/pdf/1704.00679.pdf

[6] A. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive Comput.*, vol. 2, no. 1, pp. 46–55, Jan. 2003.

[7] A. R. Beresford and F. Stajano, "Mix zones: User privacy in location-aware services," *Proc. IEEE Workshop Pervasive Comput. Commun. Secur.*, Mar. 2004, pp. 127–131.

[8] K. Emara, W. Woerndl and J. H. Schlichter, "Vehicle tracking using vehicular network beacons," *IEEE 14th Int. Symp. "World Wireless, Mobile and Multimedia Netw.*," 2013, pp. 1–6.

[9] V. T. Kilari, S. Misra, and G. Xue, "Revocable anonymity based authentication for vehicle to grid (V2G) communications," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2016, 351–356.

[10] M. Gerlach, "Assessing and improving privacy in VANETs," in *Proc. 4th Workshop ESCAR*, Nov. 2006, pp. 1–9.

[11] M. Gerlach and F. Gutler, "Privacy in VANETs using changing Pseudonyms ideal and real," in *Proc. IEEE 65th Veh. Technol. Conf., Dublin*, 2007, pp. 2521–2525.

[12] L. Buttyan, T. Holczer, and I. Vajda. "On the effectiveness of changing pseudonyms to provide location privacy in VANETs," in *Proc. 3rd Eur. Workshop Secur. Privacy Ad hoc Sensor Netw.*, vol. 4572, pp. 129–141, Jul. 2007.

[13] M. Gerlach, "Assessing and improving privacy in VANETs," in *Proc. 4th Workshop ESCAR*, Nov. 2006, pp. 1–9.

[14] L. Buttyan, T. Holczer, A. Weimerskirch, and W. Whyte, "SLOW: A practical pseudonym changing scheme for location privacy in VANETs," in *Proc. 1st IEEE Veh. Netw. Conf.*, Oct. 2009, pp. 1–8.

[15] J.-H. Song, V. Wong, and V. Leung, "Wireless location privacy protection in vehicular ad-hoc networks," in *Proc. IEEE Int. Conf. Commun.*, 2009, pp. 1–6.

[16] R. Lu, X. Li, T. H. Luan, X. Liang, and X. Shen, "Pseudonym changing at social spots: An effective strategy for location privacy in VANETs," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 86–96, Jan. 2012.

[17] J. Freudiger, M. Raya, M. Felegyhazi, P. Papadimitratos, and J.-P. Hubaux, "Mix-zones for location privacy in vehicular networks," in *Proc. ACM Workshop WiN–ITS*, Aug. 2007, pp. 1–7.

[18] A. Boualouache and S. Moussaoui, "Urban pseudonym changing strategy for location privacy in VANETs," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 24, no. 1/2, pp. 49–64, Jan. 2016.

[19] C. Vaas, M. Khodaei, P. Papadimitratos, I. Martinovic, "Nowhere to hide? Mix-zones for private pseudonym change using chaff vehicles," in *Proc. IEEE Veh. Netw. Conf.*, 2018, pp. 1–8.

[20] A. Boualouache, S. Senoucib, and S. Moussaouia, "VLPZ : The vehicular location privacy zone," *Procedia Comput. Sci.*, vol. 83, pp. 369–376, 2016.

[21] A. Boualouache and S. Moussaoui, "S2SI: A practical pseudonym changing strategy for location privacy in VANETs," in *Proc. Int. Conf. Adv. Netw. Distrib. Syst. Appl., Bejaia*, 2014, pp. 70–75.

[22] R. Yu, J. Kang, X. Huang, S. Xie, Y. Zhang, and S. Gjessing, "MixGroup: Accumulative pseudonym exchanging for location privacy enhancement in vehicular social networks," in *Proc. IEEE Trans. Dependable Secure Comput.*, Jan./Feb. 2016, pp. 93–105.

[23] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO–simulation of urban Mobility," in *Proc. Int. J. Adv. Syst. Meas.*, Dec. 2012, pp. 128–138.

[24] M. Segata, S. Joerer, B. Bloessl, C. Sommer, F. Dressler, and R. L. Cigno, "PLEXE: A platooning extension for veins," *Proc. 6th IEEE Veh. Netw. Conf.*, Dec. 2014, pp. 53–60.

[25] W. Leaf and D. Preusser, *Literature Review on Vehicle Travel Speeds and Pedestrian Injuries*, Washington, DC, USA: U.S. Department of National Highway Traffic Safety Administration, Oct. 1999. [Online]. Available: http://www.nhtsa.dot.gov/people/injury/research/pub/HS809012.html

[26] M. Jadliwala, I. Bilogrevic, and J. Hubaux, "Optimizing mix zone coverage in pervasive wireless networks," *J. Comput. Secur.*, vol. 21, no. 3, pp. 317–346, 2013.

[27] V. Primault, A. Boutet, S. B. Mokhtar, and L. Brunie. "The long road to computational location privacy: A survey," *IEEE Commun. Surv. Tut.*, vol. 21, no. 3, pp. 2772–2793, Jul.–Sep. 2019.

[28] X. Liu, H. Zhao, M. Pan, H. Yue, X. Li, and Y. Fang, "Traffic-aware multiple mix zone placement for protecting location privacy," in *Proc. IEEE Conf. Comput. Commun.*, 2012, pp. 972–980.

[29] J. Freudiger, R. Shokri, and J.-P. Hubaux., "On the optimal placement of mix zones," in *Proc. 9th Int. Symp. Privacy Enhancing Technol.*, 2009, pp. 216–234.

[30] B. Palanisamy and L. Liu, "Attack-resilient mix zones over road networks: Architecture and algorithms," *IEEE Trans. Mobile Comput.*, vol. 14, no. 3, pp. 495–508, Mar. 2015.

[31] Y. Sun, B. Zhang, B. Zhao, X. Su, and J. Su, "Mix-zones optimal deployment for protecting location privacy in VANET," *Peer-to-Peer Netw. Appl.*, vol. 8, no. 6, pp. 1108–1121, 2015.

[32] Memon, Imran, and Qasim Ali Arain. "Optimal placement of mix zones in road networks." 2017, *arXiv:1705.11104*.

[33] L. Zhang, "OTIBAAGKA: A new security tool for cryptographic mix-zone establishment in vehicular ad hoc networks," in *Proc. IEEE Trans. Inf. Forensics Secur.*, Dec. 2017, pp. 2998–3010.

[34] Q. Li, H. Wu, L. Liu, B. Pan, and L. Dong, "A group based dynamic mix zone scheme for location privacy preservation in VANETs," in *Proc. 3rd Int. Conf. Secur. Smart Cities, Ind. Control Syst. Commun., Shanghai*, 2018, pp. 1–5.

[35] Q. A. Arain *et al.*, "Privacy protection with dynamic pseudonym-based multiple mix-zones over road networks," *China Commun.*, vol. 14, no. 4, pp. 89–100, Apr. 2017.

[36] B. Ying, D. Makrakis, and Z. Hou, "Motivation for protecting selfish vehicles' location privacy in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5631–5641, Dec. 2015.

[37] A. Boualouache, S.-M. Senouci, and S. Moussaoui, "Towards an efficient pseudonym management and changing scheme for vehicular ad-hoc networks," *IEEE Global Commun. Conf.*, 2016, pp. 1–7.

[38] I. Memon, Q. Ali, A. Zubedi, and F. A. Mangi, "DPMM: Dynamic pseudonym-based multiple mix-zones generation for mobile traveler," *Multimedia Tools Appl.*, vol. 76, no. 22, pp. 24359–24388, 2017.

[39] Y. Pan and J. Li, "An analysis of anonymity for cooperative pseudonym change scheme in one-dimensional VANETs," in *Proc. IEEE 16th Int. Conf. Comput. Supported Cooperative Work Des.*, 2012, pp. 251–257.

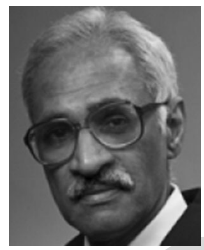
[40] Y. Pan and J. Li, "Cooperative pseudonym change scheme based on the number of neighbors in VANETs," *J. Netw. Comput. Appl.*, vol. 36, no. 6, pp. 1599–1609, 2013.

[41] J. H. Holland. *Adaptation in Natural and Artificial Systems.* Ann Arbor, MI, USA: Univ. of Michigan Press., 1975 (2nd ed. MIT Press, 1992.)

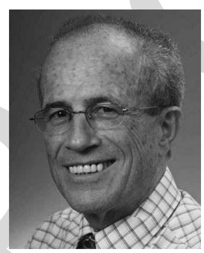
[42] Modern Heuristic Techniques for Combinatorial Problems, C. R. Reeves, Eds. New York, NY, USA: McGraw-Hill, 1995.

[43] L. Breiman, "Random forests," *Mach. Learn.*, Springer, vol. 45, pp. 5–32, 2001.

Nirupama Ravi is currently working toward the Ph.D. degree from the University of Massachusetts Amherst, Amherst, MA, USA. She worked in the telecommunications industry for over 15 years. Her research interests include security and privacy in connected vehicles, mobile networks and Internet of Things.



C. Mani Krishna received the Ph.D. degree from the University of Michigan, Ann Arbor, MI, USA. He is currently with the Faculty of the Department of Electrical and Computer Engineering, the University of Massachusetts Amherst, Amherst, MA, USA. His research interests include cyber-physical systems, real-time systems, performance/reliability evaluation, and distributed computing.



Israel Koren (M'76–SM'87–F'91) is currently a Professor with the Emeritus of Electrical and Computer Engineering, the University of Massachusetts Amherst, Amherst, MA, USA. He has been a Consultant to numerous companies including IBM, Analog Devices, Intel, AMD, and National Semiconductors. He has authored and coauthored more than 300 publications in refereed journals and conferences. He is the Author of the textbook *Computer Arithmetic Algorithms*, 2nd ed. (A.K. Peters, 2002) and a Co-Author of *Fault Tolerant Systems* (Morgan-Kaufman, 2007). His research interests include fault-tolerant systems, computer architecture, VLSI yield and reliability, secure cryptographic systems, and computer arithmetic. He was a General Chair, Program Chair and Program Committee Member for numerous conferences.