This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2010 proceedings.

Location Privacy Protection in Contention Based Forwarding for VANETs

Qing Yang Alvin Lim Xiaojun Ruan and Xiao Qin
Computer Science and Software Engineering
Auburn University, Auburn, AL, USA 36849
Email: {yangqin, limalvi, xzr0001, xqin}@auburn.edu

***Abstract*—Compared to traditional wireless network routing protocols, geographic routing provides superior scalability and thus is widely used in vehicular ad hoc networks (VANETs). However, it requires every vehicle to broadcast its location information to its neighboring nodes, and this process will compromise user's location privacy. Existing solutions to this problem can be categorized into two groups: 1) hiding user's location or 2) preserving user's identification information in routing protocols, which drastically reduce network performances. To address this issue, we proposed a dummy-based location privacy protection (DBLPP) routing protocol, in which routing decision is made based upon the dummy distance to the destination (DOD), instead of users' true locations. In this scheme, users' true locations and identification information are preserved, so the user's location privacy is protected. Compared to existing solutions, simulation results show that while DBLPP provides similar network performances as other routing protocols, it achieves a higher level of location privacy protection on vehicles in networks.**

I. Introduction

To support geographic routing in vehicular ad hoc networks (VANETs), the location information of vehicles must be periodically exchanged among one-hop neighbors [1]–[3]. This broadcasting process will release a user's location and identity information to all its one-hop neighbors among which malicious nodes may exist. Simply by overhearing periodic beacon messages sent from vehicles, an adversary can identify locations visited by a certain car and then breach the privacy of the driver.

Location privacy protection in geographic routing for VANETs can be defined as: without losing the benefits of geographic routing, a network protocol should not reveal any user's (or vehicle's) current and historical locations to unauthorized nodes. The unauthorized node may be an malicious infrastructure (e.g a wifi access point), a laptop with a wireless interface, or a vehicle moving on roads.

In this paper, to avoid broadcasting users' locations, the Contention Based Forwarding(CBF) [4] is adopted. In CBF, only the nodes participating in routing need to broadcast their locations, so the location privacies of other nodes are preserved. To further preserve the location privacies of those nodes involved in routing, we propose a dummy-based location privacy protection (DBLPP) protocol. In DBLPP, a packet forwarder (vehicle) first broadcasts a dummy distance to the destination (called dummy DOD) to its neighbors. Based on the dummy DOD, receivers compete with each other and the one closest to the destination will be elected as the next hop. Then, data packets are sent to the chosen next hop which becomes the next forwarder and routes packets as its last-hop did. The dummy DOD has to be carefully chosen so that not only the adversary cannot infer a forwarder's true location, but also the geographic routing goals are achieved.

DBLPP protects user's location privacy by 1) completely hiding user's identity using pseudonyms, and 2) utilizing dummy DODs to protect user's location information. The major contribution of this paper is a routing protocol that lets users in VANETs benefit from location-based geographic routing (e.g. CBF), and retain their location privacies. Although DBLPP is designed for VANETs, it can be easily generalized and applied to other mobile ad hoc networks.

The rest of this paper is organized as follows. Section II discusses currently available location privacy protection protocols for VANETs. In Section III and IV, the DBLPP protocol and simulation results are presented, respectively. Section V gives the conclusions.

II. Related Works

Geographic routing has been widely used in vehicular ad hoc networks (VANETs) to achieve vehicle-to-vehicle and vehicle-to-roadside communications [2], [3], [5]. By exploiting location information, geographic ad hoc routing provides superior scalability compared to traditional reactive routing protocols. However, location information sharing among neighbors compromise location privacy. The location privacy issue in VANETs were first addressed in [6], in which the authors defined the location privacy problem, threat model and application framework. In VANETs system, vehicle's location privacy issue is addressed in [7]–[9].

Location privacy issue can be solved in two different ways: hiding the information of who send the data and the information of where this data come from. For the first methods, although node's location information is released, the adversary cannot link the location to a certain user, thus protecting user's location privacy [6], [7], [10], [11]. Those approaches usually require periodically changing user's ID and such schedule is initialized or maintained by a third-party trustworthy infrastructure. The potential threat of this framework is that, the infrastructure component may not always be available and itself may be subject to security or privacy problems. Moreover, changing identifiers has detrimental effects on routing

978-1-4244-5638-3/10/$26.00 ©2010 IEEE

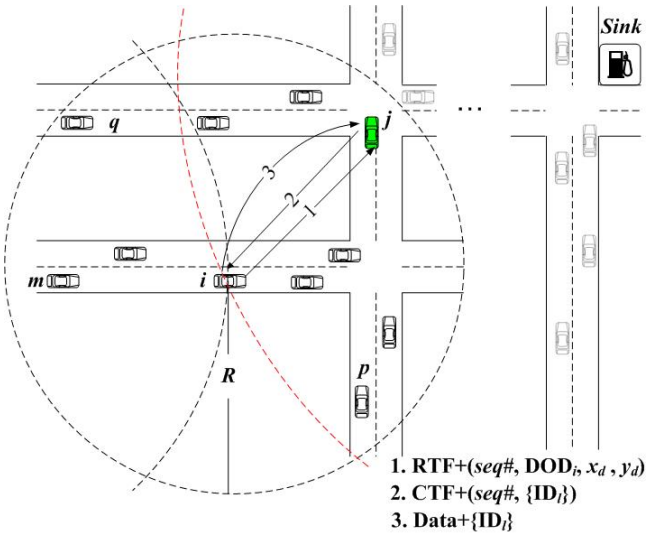

Fig. 1. Dummy based RTF/CTF exchanging among vehicles

efficiency and increases packet loss as shown in [12].

In the second method, packet forwarders will send out in an area or a set of dummy locations which hides the true locations. For instance, a node will send to a rectangle or circular area in which there exist at least $k-1$ other nodes [13]. Thus, *k-anonymity* is achieved since an adversary can only identify a user's location with the probability of no higher than $1/k$. Unlike the k-anonymity methods, dummy-based location privacy-protection algorithms were proposed [14], [15]. In [14], the network user generates several false position data (with one that contains the true position information) sent to the service provider. Because the service provider cannot distinguish the true position data from the dummies, the user's location privacy is protected. Similarly, authors in [15] hid user's real location by sending a set of dummy positions which are deliberately generated according to either a virtual grid or circle. In the above-mentioned methods, user's true location information is fully hidden within either an area or a set of dummies, so traditional geographic routing protocols will have a big problem in making routing decision as location information is not available.

Unlike previous work, we investigate user's location privacy issue through 1) replacing user's location information by dummy DOD during routing and 2) generating pseudonyms to preserve users identity information. Despite these changes, the geographic routing protocol will still work, with a slight modification. Both identity and location information of users is preserved in our DBLPP protocols, so it can achieve a higher level of location privacy protection in VANETs.

III. DBLPP Protocol Details

A. Next Hop Selection

As shown in Fig. 1, suppose the current packet forwarder is node i which just received a packet with sequence number $seq\#$ from node m, and the packet needs be transmitted to the destination located at (x_d, y_d). In the figure, we can see node j should be the next hop since it is the closest neighbor to the destination. Before routing a packet, node i first broadcasts a RTF packet which includes the $seq\#$ and (x_d, y_d) (the dummy DOD will used later in protecting its location privacy).

When a neighbor (e.g. node k) receives this RTF, it first checks if the packet was received before by comparing the $seq\#$ to those of buffered packets. If there is a hit, that means it received this packet before. It will simply drop the RTF since it cannot make any further propagation progress than that of node i. If this is a brand new RTF, node k saves this packet into buffer and sets a timer with the runtime of:

$$t(r_k) = f(1 - 1/r_k) \tag{1}$$

where r_k is the DOD of node k. As the timer's runtime of every node is proportional to it DOD, the one closest to the destination will first time out. As shown in Fig. 1, node j will first time out and send back the clear-to-forward (CTF) message along with $seq\#$ and a set of pseudonyms, denoted as $< ID_l >$. These l pseudonyms in $< ID_l >$ are randomly chosen from the set of L pseudonyms, which are pre-installed on each vehicle. Note that the pseudonyms will not be the same in different CTF messages even if they are sent from the same node. If we select the values l and L properly, the probability of two nodes choosing the same $< ID_l >$ will be extremely low. This CTF message also serves as a suppression message for all j's neighbors, so those nodes (neighbors of node j and i) receiving the CTF message will immediately cancel their timers. Since the CTF from j can only suppress its neighbors, there may exist other nodes which may send duplicated CTF messages, i.e. multiple CTFs may be received at the sender i.

After receiving the first CTF sent from j, node i immediately sends a data packet along with the previously received pseudonyms $< ID_l >$. If the data packet is sent before the second CTF is received, all its neighboring nodes will be suppressed by the data message. If another duplicated CTF message is receive before the data packet being sent, node i simply omits this CTF.

When a neighbor of node i receives the data packet, if it has not sent a CTF message, it drops this data packet. Otherwise, it checks whether the $< ID_l >$ in the data packet are the same as what it sent out. If so, the packet are decoded; otherwise, the packet is dropped. By exchanging one-pair of RTF and CTF messages, geographic greedy forwarding can be achieved between node i and j.

B. Duplicated Responses and Location Privacy Protection

By the above mentioned next hop selection strategy, there may exist several duplicated CTFs which may affect network performance. Moreover, from the timer information of neighboring nodes, the adversary can easily infer their DODs and that comprises user's location privacy. Therefore, it is important to set up a proper timer so that the location privacies of receivers and senders are preserved and the number of duplication responses are minimized.

In Equation (1), we note the runtime of timer on a certain node only depends on the node's DOD. Therefore, a simple

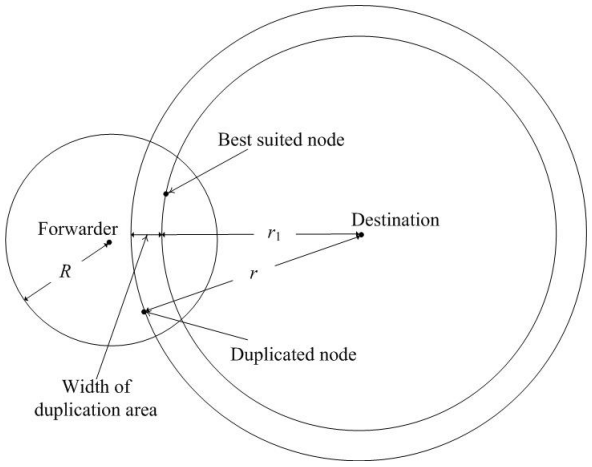

Fig. 2. Duplicated Responses and Duplication Area

way to calculate this interval with a DOD r will be:

$$t(r) = T \cdot (1 - 1/r) \tag{2}$$

where T is the maximal one-hop forwarding delay. In this case, duplicated responses can be generated by receivers in the duplication area shown in Fig. 2. Suppose there are a best-suited node with DOD of r_1 and anther node with DOD of r. If the runtime differences between those two nodes are too small (i.e. $t(r) - t(r_1) < \delta$ where δ is the minimal time interval needed for suppression), then duplicated messages will be sent from those two nodes. Therefore, the larger the width of the duplication area, the more duplicated responses will be generated.

Following Equation 2, the value of T needs to be very large to achieve a reasonable small duplication area. For example, according to Equation 2, the interval of timer on the best-suited node is $t(r_1) = T \cdot (1 - 1/r_1)$. Then, duplicated messages will be generated by other nodes with DOD r satisfying the following condition:

$$t(r_1) < t(r) < t(r_1) + \delta = T(1 - \frac{1}{r_1}) + \delta \tag{3}$$

Therefore, to avoid generating duplicated message a node needs to set a timer with runtime being larger than:

$$t(r) = T\left(1 - \frac{1}{r_1}\right) + \delta = T\left(1 - \frac{1}{\frac{r_1 T}{T - \delta r_1}}\right) \tag{4}$$

Thus, the width of duplication area can be computed as:

$$\frac{r_1 T}{T - \delta r_1} - r_1 = \frac{\delta r_1^2}{T - \delta r_1} \tag{5}$$

Since δ is a fixed value, we can denote the maximal one-hop forwarding delay as $T = k \cdot \delta$. To achieve an acceptable duplication area with a width of Δ, the following equation must be satisfied:

$$\frac{\delta r_1^2}{T - \delta r_1} = \frac{r_1^2}{k - r_1} < \Delta \tag{6}$$

In other words, the delay T will be:

$$T = \frac{r_1^2 + r_1 \Delta}{\Delta} \cdot \delta \tag{7}$$

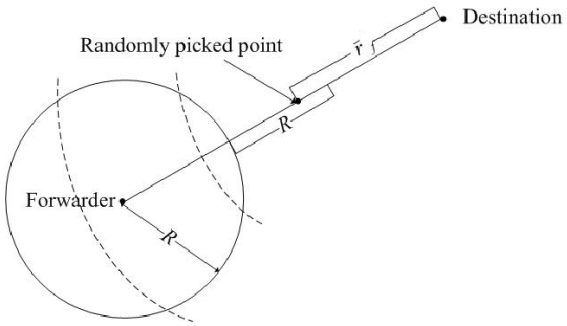

Fig. 3. Dummy DOD selection on a packet forwarder

From the above equation, we find that the longer the DOD of a node, the larger the value of T will be. However, the value of T must be very small to avoid huge communication delays. To avoid this issue, we modify the Equation (2) by considering the dummy DOD information obtained from the last-hop forwarder. Suppose a node (current packet forwarder) sends out a RTF packet, all receivers will set timers according to the following equation:

$$t(r) = T \cdot \left(\frac{r - \bar{r}_f}{3R}\right) \tag{8}$$

where r is the DOD of a receiver and $\bar{r}_f$ is the packet forwarder's dummy DOD which is sent along with the RTF. Since the value of $\bar{r}_f$ is random, the real locations of the forwarder and the next hop are preserved.

As shown in Fig. 3, the dummy DOD of the forwarder $\bar{r}_f$ is computed as:

$$\bar{r}_f = r_f - (1 + \rho) \cdot R \tag{9}$$

where r_f is the real DOD of this forwarder, ρ is real number randomly chosen from $(0, 1)$, and R is the wireless communication range. Since the value of ρ ranges from 0 to 1, the forwarder's real location is hidden within a range of R. That means the difference between the real and dummy DODs is $r_f - \bar{r}_f$ is within $[R, 2R]$, so the location privacy of the forwarder is preserved.

We also know that $r - \bar{r}_f$ is equal to $r - r_f + (1 + \rho) \cdot R$. Since the forwarder and the receiver are neighbors, the value of $r - r_f$ is within $[-R, R]$. Therefore, the value of $r - \bar{r}_f$ is within $[0, 3R]$, so the timer of this receiver is within $[0, T]$.

With these equations, we can calculate the width of duplication area as $3R \cdot \delta / T$. If $T = k \cdot \delta$, to achieve an acceptable duplication area, we must have:

$$T = 3R \cdot \Delta \cdot \delta \tag{10}$$

We note that the one-hop maximal delay in the above equation is a fixed value. This delay is much smaller than what is computed from Equation (7), and is acceptable in VANETs. Using the dummy DODs of forwarders and pseudonyms of receivers, the DBLPP provides a higher level of location privacy protection on vehicles in VANETs. Meanwhile, with the active selection of every next hop, geographic routing is achieved in networks and the number of duplication responses and average one-hop delay are minimized.

978-1-4244-5638-3/10/$26.00 ©2010 IEEE

TABLE I
SIMULATION SETUP PARAMETERS

Parameter	Value
Number of lanes	2 lanes per direction
Number of nodes	200
Velocity	15-35MPH
Period of traffic light	60 seconds red and green
Wireless communication range	250m
Beacon interval	5.0 second
Packet size	512 Bytes

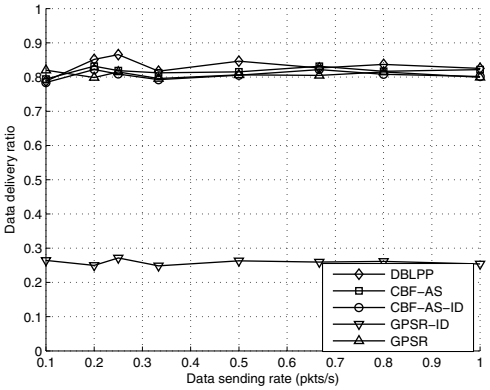

Fig. 4. Data delivery ratio vs data sending rate

IV. SIMULATIONS AND RESULTS

We implement the DBLPP protocol in ns-2.29 and compare it to original geographic routing protocols: GPSR and CBF-AS. To compare the location privacy protection, we implement the periodic changing-pseudonym scheme which is widely used in previous works. Therefore, besides the GPSR and CBF-AS, we compare another two protocols with the periodic changing-pseudonym scheme: CBF-AS-ID and GPSR-ID. Details of the simulation setup parameters are listed in TABLE I.

A. Data Delivery Ratio

Data delivery ratio is defined as the number of received packets at the destination divided by the number of sent packets from the source. As shown in Fig. 4, DBLPP, CBF-AS, CBF-AS-ID and GPSR achieve similar data delivery ratios. GPSR-ID gives the lowest data delivery ratio because chosen next hops often change their IDs and thus drop packets which supposed to be sent to them. In GPSR, every node selects the next hop based on the stored neighbors' information which is obtained from beacon messages. Since neighbors's location information is updated periodically, it is possible that out-of-date neighbors exist in one's neighbor list. Therefore, packets may be delivered to a neighbor which is already out of the communication range. However, in DBLPP, the next hop will be elected through competition and the winner will respond to the packet forwarder quickly. Then, packets will be immediately sent to this elected next hop. Therefore, the chance of forwarding packets to a out-of-date neighbor in DBLPP is very low, and this is why DBLPP delivers more packets than GPSR. Because contention-based forwarding is used in DBLPP and CBF-AS, their data delivery ratios are very similar. Although periodic changing ID is applied on CBF-AS-ID, its data delivery ratio performance is slightly worse than those of DBLPP and CBF-AS. This is because, after a next hop sends its ID in a CTF message, the sender immediately delivers the packet, the time difference between those two events is too small to allow the next hop to change its ID.

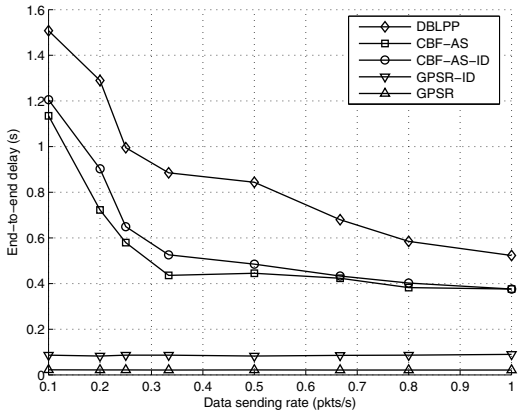

Fig. 5. End-to-end delay vs data sending rate

B. End-to-end Delay

The end-to-end delay is defined as the average time taken for a packet to be transmitted across a network from source to destination. As shown in Fig. 5, GPSR and GPSR-ID gives smaller end-to-end delays compared to others because there are no delays in the selections of next hops. However, in DBLPP, CBF-AS and CBF-AS-ID, after a packet forwarder sends out a RTF message, its neighbors will start timers which cause delays. That means an extra delay for the next hop election is added to every hop, so the end-to-end delays of CBF-AS, CBF-AS-ID and DBLPP are higher. Using the same contention based selection mechanism, DBLPP generates a larger end-to-end delay compared to those of CBF-AS and CBF-AS-ID. This is because in DBLPP there are more duplicated responses which make the networks more congested and thus increase the end-to-end delay. This delay can be further reduced by using a smaller maximal-runtime of timers.

C. Network Throughput

Network throughput is defined as the number of packets delivered to the destination per second. As shown in Fig. 6, besides GPSR-ID, all the other protocols give the similar network throughput and these throughputs increase when the data sending rate increases. Since the link quality of every hop in DBLPP is better than that of GPSR, DBLPP achieves a slightly larger network throughput than GPSR.

D. Location Privacy Protection

Entropy was first introduced in *Information Theory* to quantify the uncertainty of a system. The higher the privacy entropy value is, the more uncertain attackers will be of their user location inference, and hence the better privacy protection our system offers. In the simulations, we tracked

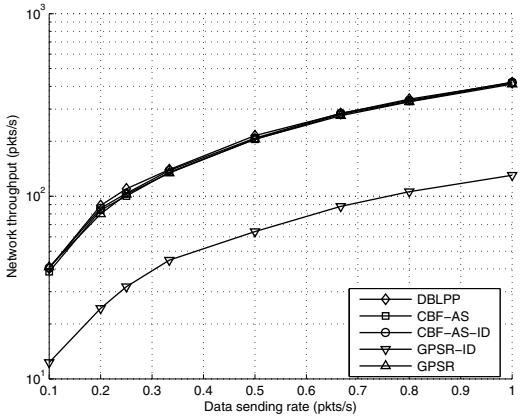

Fig. 6. Network throughput vs data sending rate

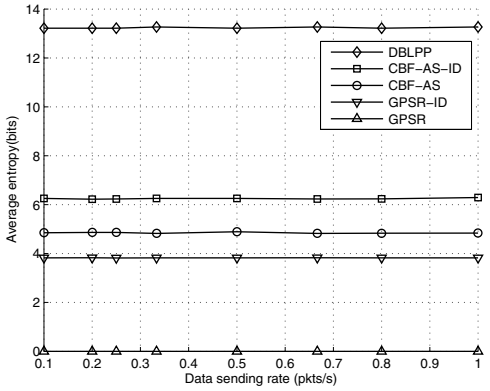

Fig. 7. Average entropy of location privacy protection

all communication events (RTF, CTF and data packets) and computed the average probability of predicting the location and identity information of a node involved in routing. Based on the definition of entropy, we then calculate the average entropy required for the adversary to predict a user's location and identity.

As shown in Fig. 7, in order to attack a vehicle's location privacy, more bits are required in DBLPP compared to others. In GPSR, every node periodically beacons its location and ID to neighbors, so the entropy of computing every vehicle's location is zero. In CBF-AS, every packet forwarder sends its location (not ID) in RTF messages to neighbors. When the self-elected next hop sends a CTF message, its ID (not location) is put into the packet, so the CBF-AS provides a higher entropy value. In DBLPP, dummy DODs and pseudonyms are used by packet forwarders and next hops, respectively. Therefore, it requires more bits to attack even one node's location privacy. Although CBF-AS-ID and GPSR-ID can provide a certain degree of location privacy protection, they are not as good as DBLPP because it preserved both the identity and location information. In summary, the location privacy protection in DBLPP is much better than others.

V. Conclusions

To address the location privacy leakage in geographic routing, we propose a dummy-based location privacy protection protocol. The location information shared among vehicles is required by all kinds of geographic routing protocols. However, the proposed DBLPP does not need vehicles to exchange their locations but only dummy DODs of forwarders. In addition, elected next hops respond to forwarders with a group of pseudonyms, so the ID information is hidden too. To verify the design and evaluate the performance of DBLPP, we used ns-2 and VanetMobiSim to simulation the networking activities and vehicle mobilities, respectively. Simulation results show that DBLPP not only protects user's location privacy but also achieves similar network performances as others. Moreover, the protocol is implemented in the IEEE 802.11 protocol, and thus can be easily applied to currently widely used wireless interfaces with IEEE 802.11 chips.

References

[1] B. Karp and H. T. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," in *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, 2000, pp. 243–254.

[2] C. Lochert, M. Mauve, H. Füssler, and H. Hartenstein, "Geographic routing in city scenarios," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 9, no. 1, pp. 69–72, 2005.

[3] J. Zhao and G. Cao, "VADD: Vehicle-assisted data delivery in vehicular ad hoc networks," in *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, April 2006, pp. 1–12.

[4] H. Fler, J. Widmer, M. Ksemann, M. Mauve, and H. Hartenstein, "Contention-based forwarding for mobile ad hoc networks," *Ad Hoc Networks*, vol. 1, no. 4, pp. 351 – 369, 2003.

[5] Q. Yang, A. Lim, and P. Agrawal, "Connectivity aware routing in vehicular networks," in *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, 2008, pp. 2218–2223.

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

[7] K. Sampigethaya, M. Li, L. Huang, and R. Poovendran, "AMOEBA: Robust location privacy scheme for VANET," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 8, pp. 1569–1589, Oct. 2007.

[8] F. Dötzer, "Privacy issues in vehicular ad hoc networks," in *Privacy Enhancing Technologies*, 2005, pp. 197–209.

[9] R. Lu, X. Lin, H. Zhu, P.-H. Ho, and X. Shen, "ECPP: Efficient conditional privacy preservation protocol for secure vehicular communications," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, April 2008, pp. 1229–1237.

[10] K. Sampigethaya, L. Huang, M. Li, R. Poovendran, K. Matsuura, and K. Sezaki, "CARAVAN: Providing location privacy for VANET," in *Proceedings of Embedded Security in Cars (ESCAR)*, 2005.

[11] J. Freudiger, M. Raya, M. Flegyhzi, P. Papadimitratos, and J.-P. Hubaux, "Mix-Zones for Location Privacy in Vehicular Networks," in *ACM Workshop on Wireless Networking for Intelligent Transportation Systems (WiN-ITS)*, 2007.

[12] E. Schoch, F. Kargl, T. Leinmuller, S. Schlott, and P. Papadimitratos, "Impact of Pseudonym Changes on Geographic Routing in VANETs," in *Third European Workshop on Security and Privacy in Ad hoc and Sensor Networks (ESAS 2006)*, 2006, pp. 43–57.

[13] L. Sweeney, "k-Anonymity: A model for protecting privacy," *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 10, no. 5, pp. 557–570, 2002.

[14] H. Kido, Y. Yanagisawa, and T. Satoh, "Protection of location privacy using dummies for location-based services," in *Data Engineering Workshops, 2005. 21st International Conference on*, April 2005, pp. 1248–1248.

[15] H. Lu, C. S. Jensen, and M. L. Yiu, "PAD: Privacy-area aware, dummy-based location privacy in mobile services," in *Proceedings of the 7th International ACM Workshop on Data Engineering for Wireless and Mobile Access*, 2008, pp. 16–23.