Maximizing Anonymity of a Vehicle through Pseudonym Updation

Brijesh Kumar Chaurasia
Indian Institute of Information Technology
Allahabad, India
+91-9453701376
bkchaurasia@iiita.ac.in

Shekhar Verma
Indian Institute of Information Technology
Allahabad, India
+91-9450965336
sverma@iiita.ac.in

ABSTRACT
A vehicle can be tracked through its locatable transmission. The broadcast by a source contains its current identity and also allows estimation of its location by receivers of this transmission. This possibility of mapping between the physical entity and the estimated location through the communication broadcast is a threat to privacy. The changes in the location due to motion and the alteration in the temporal identifiers diminish the correlation between location and physical entity. However, such a mapping can still be recognized when an actively communicating node in relative isolation is observed for a sufficient interval of time. This paper addresses the challenges in providing anonymity to a moving vehicle that continually switches identifiers. As a vehicle moves on a road, its neighbors change in accordance to its relative speed with neighboring vehicles. This change in the nature and size of the neighborhood, i.e. the anonymity set, reduces the anonymity of a vehicle. It is shown that the effective size of anonymity set reduces drastically due to change in the neighbors and transmissions by vehicles. The work studies the possibility that a node may retain its anonymity by switching identities in the vicinity of other vehicles to decorrelate its location and identity relation. A heuristic that allows a vehicle to switch its identity at a time and place where the potential of anonymity preservation can be maximized by increasing the anonymity set is proposed. The performance of the proposed heuristic is evaluated in a highway environment with vehicle mobility and dynamic vehicle population. Results indicate that updating pseudonyms in accordance to the heuristic maximizes the size of the anonymity set and through it, the anonymity of a vehicle.

Keywords
Anonymity, vehicular networks, pseudonyms, anonymity set (key words).

1. INTRODUCTION
Transmission in the shared wireless medium reaches all the nodes in the communication range.

Conference name: WICON'08, November 17-19, 2008, Maui, Hawaii, USA.

An adversary can determine the identity and position of the source vehicle from the contents of the communication packets. The position can further be confirmed by gauging the position of the vehicle from the signal strength. Using this information, the physical vehicle and its communication identity can be traced and related. Thus, location tracking through eavesdropping of source transmission and physical observation can breach the location privacy of the user. This can be used to disclose personal data of a user and would potentially dissuade a user to join and reap benefits from such networks.

To obtain and sustain anonymity, a temporal identity is used for communication. Pseudonyms allow a vehicle to interact with other vehicles anonymously. Pseudonyms are ephemeral and distinct pseudonyms hide their relation to each other and to the user's identity. To preserve privacy, a pseudonym system must prevent credential forgeability and disallow to usage of false pseudonym by a user. Moreover, the transaction of obtaining and the process of switching pseudonyms should not reveal the identity of the user or link pseudonyms to each other. Continually changing pseudonyms conceal the real identity of a vehicle by de-linking the source of signals to its original identity. But the relation between a communicating vehicle and its estimated location can reveal the identity of a vehicle. This vehicle can, then, be physically traced and switching pseudonyms would be meaningless. A vehicle can be under sustained observation and transmissions at different intervals of time with the same pseudonym can reveal the relation between physical vehicle and its current pseudonym if the vehicle is relatively isolated. This relation can be established even when pseudonyms are updated when the time interval between transmission prior to and after the updating is short. There is, moreover, one more challenge that needs to be addressed. When a vehicle under observation moves from one cluster and enters another cluster and changes its pseudonym, it can be spotted with high probability as soon as it transmits. This can happen if the number of vehicles from the previous cluster to the current cluster is small and the pseudonyms of vehicles belonging to current cluster are known a priori. The anonymity of the vehicle under observation is limited by the number of vehicles that join the current cluster from the previous cluster.

Existing solutions address the problem of randomizing the location based relation between pseudonyms and physical identity by the enabling the vehicles to update their pseudonyms only at specific times [1] or pre-determined locations known as MIX-Zones. Continuous transmission before and after an identity switch can aid in forming of a relation between pseudonyms

Digital Object Identifier: 10.4108/ICST.WICON2008.4904
http://dx.doi.org/10.4108/ICST.WICON2008.4904

through the nature of transmission and signal properties. A random silent period technique [2] in which a node does not transmit for a random period during update of identifiers excludes the possibility of forging this relation. However, the location privacy provided by these solutions is limited by tracking methods that leverage the predictability of the movement of vehicles to correlate their locations before and after the switch [3]. The increase in the size of the MIX-Zone and silence time period mitigate the possibility of any relation formation. The formation of any such relationships based on the predictability of node movement and signal transmission can be diminished further by increasing the frequency of pseudonym switch. However, the switching time and the frequency of switching can be limited by routing [4] and other network needs [5]. To strike a balance, a vehicle should switch as soon as possible after such change is warranted by anonymity requirements. In this work, we focus on the problem of diminishing the possibility of forging a relationship between vehicle identity and its transmissions by determining the conditions that maximize anonymity during an identity switch.

The rest of the paper is organized as follows. Section 2 describes the problem along with a measure of anonymity for a moving vehicle and the proposed heuristic with analytical analysis in section 3. The simulation and results are given in section 4; section 5 concludes the work.

2. PROBLEM DESCRIPTION

2.1 Measuring Anonymity

A vehicle is under sustained physical observation and all communication emanating from a source (with the same pseudonym) in a region can be listened to by the adversary. The adversaries can also collude to obtain a rough estimation of the zone of the presence of communicating vehicle. This zone *Z* is the anonymity zone. The level of anonymity of a vehicle is the inability of the adversary to pinpoint a vehicle as the source of the communication in the set of vehicles *V* (anonymity set) in the region estimated from the communication. This anonymity set $V \subseteq V_{total}$ with V_{total} being the total number of vehicles in *Z*. The entropy of the anonymity set is the measure of the anonymity of a vehicle in the set.

If the all the vehicles can be the source of communication with equal probability, then the probability p_i that the vehicle V_i under observation is the target,

$p_i = P_r(V_i = V), \forall \in Z$ and $\sum p_i = 1$

The entropy, $H(p)$ [6] of the distribution of the anonymity set is

$$H(p) = -\sum_{i=1}^{|V|} p_i \log_2 p_i$$

The anonymity of a given vehicle is maximized when all the vehicles are equally likely to be the potential target (source of communication). Under this uniform distribution, the probability p_i that the vehicle V_i under observation is the target becomes $p_i = \frac{1}{|V|}$ for all the vehicles.

Following the definition of level of anonymity given in [7], we have,

$$A_l = 1 - 1/|V|$$

With entropy $H(p) = log_2 V$

The population of a zone is dynamic with vehicles joining and leaving a cluster. Let there be a cluster of vehicles (set of initial vehicles $V_{initial}$) at time instant t_i. Some vehicles are communicating while others are silent. At time, $t_i + \Delta t$, few other vehicles join this cluster. Some vehicles of these vehicles had been transmitting just prior to joining the cluster. Some are silent. Let the target vehicle V_i be one of the vehicles in the set of late entrants, V_{late}, in the cluster.

The level of anonymity of a target vehicle is dependent on the size of the crowd indistinguishable from the target vehicle and not the total number of vehicles in the zone *Z*. The vehicles that stay with the target vehicle when it joins or leaves a cluster determine the level of anonymity enjoyed by a target vehicle.

Following conditions may arise.

Condition 1: Some vehicles in the set V_{late} communicated immediately prior to joining the cluster. Once these vehicles join the cluster (time $t \geq t_i + \Delta t$), one of the vehicles of the set V_{late} again communicates without changing in its pseudonym. In this case, the anonymity set $V_{c(late)}$ is equal to the number of vehicles who entered cluster and had communicated just prior to entering the cluster.

$A_l = 1 - 1/|V_{c(late)}|$, $V_{c(late)} \subseteq V_{late}$

When a vehicle moves from one cluster to another without changing its pseudonym, its anonymity set is the number of vehicles that are common (move with the vehicle) through different clusters. For example, if some vehicles from cluster C1 join cluster C2 and a few vehicles from C2 join C3, then, the anonymity set for a target vehicle that moved from $C_1 \rightarrow C_2 \rightarrow C_3 \rightarrow \ldots \rightarrow C_n$ would be $C_1 \cap C_2 \cap C_3 \cap \ldots \cap C_n$ and $A_l = 1 - 1/|C_1 \cap C_2 \cap C_3 \cap \ldots \cap C_n|$. Thus, the anonymity set may be very small even though large number of vehicles present in *Z* is large.

Condition 2: All the vehicles in the set V_{late} are silent for a random period before joining the cluster. Communication is received from a source with a pseudonym not used before in the ongoing communication emanating from the cluster. In this case, the anonymity set is the set of all vehicles silent just prior to this communication. If the number of different communication (with different pseudonyms) emanating from the cluster at time *t* was V_{silent}, then the anonymity level is

$A_l = 1 - 1/|V_{c(late)} \cup V_{silent}|$, $V_{c(silent)} \subseteq V_{initial}$

If all the vehicles in the set $V_{initial}$ were communicating, then the anonymity set reduces to the vehicles that have just joined (assuming in this case that pseudonym switch is preceded by a silence period). The anonymity level is

$A_l = 1 - 1/|V_{c(late)}|$

Consequently, for a target vehicle under observation by an adversary or group of adversaries, the anonymity is maximized when the cardinality of the set is more than *k* (known as *k*-anonymity) and all the vehicles in the zone of anonymity have equal probability of being the target. The anonymity level becomes ($A_l = 1 \forall |V| \geq k$)

The target is hidden a crowd of *k* vehicles in the zone of uncertainty around the target, this enables anonymity even when the time of communication, extended position and pseudonym of a source are known.

Digital Object Identifier: 10.4108/ICST.WICON2008.4904
http://dx.doi.org/10.4108/ICST.WICON2008.4904

2.2 Proposed Heuristic for Maximizing Anonymity

The level of anonymity is a function of the cardinality of the anonymity set. Increasing anonymity is tantamount to increasing the number of vehicles in the intersection set close to *k*. The proposed heuristic aims to achieve this objective by changing its pseudonym at a time and place so that the size of the anonymity set becomes sufficiently large to hide the vehicle in the increased crowd. Moreover, this change in pseudonym itself does not point towards the vehicle. The heuristic identifies two cases.

Case1: To maximize its anonymity, a moving vehicle V_i continually observes the number of vehicles in its vicinity *Z* that are communicating.

After an updation, a vehicle does not change its pseudonym for a short period which is fixed. After this period, pseudonym is updated at an appropriate time and place. If sufficient time has elapsed since V_i had from the previous to the current pseudonym or the number of vehicles in the neighborhood from the last transmission becomes less than *k*; updation in pseudonym is imperative. The change is effected as soon as the number of vehicles that are transmitting becomes more than the critical mass, *k*. Till such time, the vehicle remains silent. If the vehicle has to transact with an RSU to obtain a pseudonym, it may obtain it a priori and update its identity when the aforesaid condition is satisfied. It can be further observed that the silence period can be before or after the identity change. Hence, if the pseudonym change is immediately warranted; the change must be followed by a time lag before next transmission is done.

Case 2: The above condition of critical mass of the anonymity zone can be relaxed if vehicles are able to observe all the vehicles in their vicinity (using radar). In this case, a pseudonym switch can be performed when the total number of vehicles (silent and communicating) is at least equal to *k*. All the other conditions remain identical to Case 1.

3. ANONYMITY ANALYSIS

3.1 System Model

Vanets are characterized by a highly dynamic topology with vehicles moving at high speed in restricted geographical strait jackets (highways). When vehicles move on a road, they lateral motion is very restricted and the motion is unidirectional except at the junctions. A vehicle moving on the road in a particular direction can move at different speeds and also pause. Since, the speed of a vehicle can be variable; vehicles may overtake one another without any restriction. Since the transmission range of any vehicle is more than the total width of the road, this lateral motion has no effect on communications and can therefore be neglected. At the junctions or crossroads, the vehicle has a choice of taking one of options. At any junction, new vehicles may enter and continue on the road. Thus, if we consider a single unidirectional road or highway, then vehicles may enter the road at different junction points. A junction is also an exit point where vehicles from may depart from the road or continue their onward journey on the road. In the present model, a departed vehicle does not reenter or participate. Each vehicle on the road is either continually transmitting periodically, aperiodically or is silent. The aim is to find the distribution of the vehicle population on the road when the vehicles move from one end of the highway to the other end to estimate the size of the anonymity set of a target vehicle. To determine the anonymity set, the system model of the road with vehicle mobility is taken from [8].

3.2 Road Model and Input Traffic

A road with multiple unidirectional lanes is considered. The vehicles move in a single direction with different speeds and multiplicity of the lanes allows vehicles to overtake each other without any restriction in the number of lanes. The road has *S* segments of *D* meters each. The road length is *S.D* meters. Vehicles can enter and exit at the end points of a segment and there is no waiting period or queue required to enter the road. There are two end points of the road and there is no exit point at the start of the initial segment or entry at the end of the last segment. Vehicles arrive at the beginning of a segment follow a Possion distribution $p_Y(y)$ and travel towards the end point independently of other existing vehicles on the road. The arrival rate at the $(ij)^{th}$ segment at $(i)^{th}$ point is λ_i with the departure rate at the $(j)^{th}$ point being δ_j (λ_i =0 for i =0 and δ_j for j=*S*). Once it enters the road at an entry point, it cannot exit from the same point i.e. every vehicle must traverse at least one segment as soon as it enters the road. Hence, this system can be modeled as an $M/G/\infty$ queue [9]. This means that the jobs (vehicles) arrive at the system (road) in accordance with a Poisson process. The system has infinitely large number of servers, so that a job is serviced immediately on its arrival without any queuing, ie an arriving vehicle is admitted in the road and starts traversing on the road without any waiting period (on the side of the road). The service time is the time which a vehicle spends on the road. Job service times are assumed to be independent general random variables with a common distribution function $G(v)$.

The vehicles travel in one direction with variable velocity. The velocity of a vehicle is assumed to be piecewise constant ie a vehicle moves with a constant speed for a time period and then changes its speed [8]. The velocity of a vehicle v_i in during time periods follows a normal distribution with mean speed μ and variance σ^2; $N(\mu, \sigma^2)$. The time durations t_i form i.i.d. random variables with exponential distribution with distribution parameter $1/\alpha$. The distance covered in a time duration t_i is $d_i = v_i\, t_i$ and the total distance covered in i time periods is $TD_i = \Sigma d_i = \Sigma v_i\, t_i$.

4. SIMULATION AND RESULTS

A road of *S* segments is considered. The vehicles arrive at the start of each segment with a Poisson rate λ (ranging from) and their speeds are normally distributed. The vehicles travel along the road as per the mobility model and transmit in accordance to a Poisson distribution. The simulation is run independently multiple times and various statistics like the density of vehicles at different times and distances from the respective start segment is collected. The simulation has been performed in Matlab.

The total length of the road is taken as 18 km with segments of 3 km each as in Fig. 1. The input traffic can enter from *A*,…, *F* and exit from *B*…*G*. The vehicle mean arrival rates are taken as λ= (0.2, 0.5, 0.7) with departure rates as δ= (0.1, 0.2). The time period for constant velocity is exponentially distributed with mean 1, 2 and 3 seconds and the velocities are normal variates with mean μ (15, 20, 25 and 30 ms^{-1}) and σ (2 and 3 ms^{-1}). The mean transmission rate of a vehicle is taken as ρ_i =0.5.

Digital Object Identifier: 10.4108/ICST.WICON2008.4904
http://dx.doi.org/10.4108/ICST.WICON2008.4904

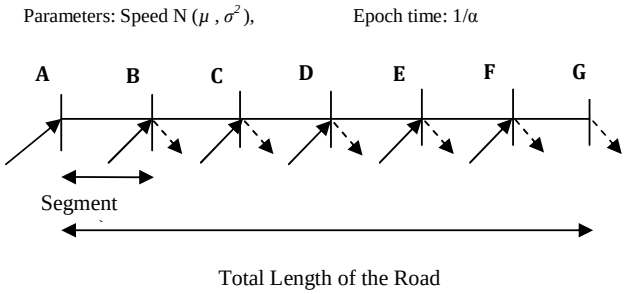

Fig. 1. Road Model

Fig.4 depicts the number of vehicles with distances. Since vehicles enter at different junctions, the number of vehicles increase with distance. The total number of vehicles at different distance becomes sparse at higher speeds (μ = 25 and 30 ms^{-1})) but interestingly, higher deviation (σ = 3 ms^{-1}) at these speeds results in more stable population of vehicles on the road. This might result in a stable cluster size but to determine whether is neighborhood itself is stable; the number of vehicles that stay with the vehicle under observation needs to be traced. To determine the number of vehicles around a vehicle of interest, a vehicle from the initial segment is considered. As this vehicle travels towards the destination, some vehicles form a part of the cluster around this vehicle. The size of the cluster and the neighborhood is a function of the transmission range and the speed of different vehicles. In the present study, we consider that the vehicles have a fixed transmission range of 500 *meter* radius and observe the cluster size around the vehicle for different speeds (μ = 20 and 30 ms^{-1}) and variations in speed (σ =2 ms^{-1}). The mean number of vehicles around the vehicle of interest is shown in Fig. 2 and Fig. 3. To study the effective size of the anonymity set, the number of vehicles in the neighborhood cluster that transmit are also determined and shown in Fig.2 and Fig. 3 along with the total number of vehicles in the cluster.

As shown in Fig. 2 and 3 (μ =20 ms^{-1} and σ =2 ms^{-1}), the vehicle under observation departs after the third segment. Till this segment, the total number of vehicles in its neighborhood is large (10-80). Even the number of vehicles accompanying it from its initial segment is sufficient. Thus, the cardinality of the anonymity set (defined as the total number of vehicles in the neighborhood) is 10-80, which is large crowd. When an adversary is tracking the

vehicle, the cardinality of the anonymity set reduces to the number of vehicles that remain with it from the initial segment. This is also sufficient. Similar nature is observed at higher speeds and higher deviations in speed. The neighborhood of the vehicle under observation is large but the neighbors keep on changing. However, as the mean speed increases, the size of the neighborhood decreases significantly with maximum number reducing to 40-50 at high speeds with large variation in speeds. The observation in earlier figures is substantiated, higher speeds decreases the cardinality of the anonymity set. Furthermore, the change in the neighborhood is also significant and an adversary tracking a vehicle along the road will be able to track the vehicle successfully if the communication identifier remains constant in all the transmissions as it moves along the road. Continual changes in the identifier is warranted i.e. pseudonyms are necessary for privacy of a vehicle. As described earlier, place and time of change of pseudonyms determines the effective cardinality of the anonymity set. The number of neighborhood vehicles form a significant portion of the total number of vehicles for mean transmission rate of ρ_i =0.5. The vehicles in communication from each segment are a small part of the total number of vehicles from each segment.

The vehicle under observation is silent at time $t_{(i-1)}$ and communicates at $t_{(i)}$. Even if it transmits in a zone, its effective cardinality of anonymity set is large since it had not transmitted before and pseudonym change increases the effective size of the anonymity set. The transmission with a new identifier makes it indistinguishable from other vehicles in the neighborhood even though they are from different segments, i.e. size = 80 at a distance of 9 km (Fig. 2). However, if there was a transmission from the vehicle (say at a distance of 6 km as shown in Fig. 3) and then again at 9 km without change in pseudonym, the effective cardinality reduces to around 15 (the number of vehicles that remain with the vehicle under observation from the initial segment). If the required cardinality for anonymity is less than 15, a vehicle may not change its pseudonym, lest a change should be imperative. However, if the vehicle continues to transmit without change in pseudonym, the cardinality progressively becomes smaller and ultimately, the adversary would be able to map the vehicle to its pseudonym. Thus, if a vehicle is able to monitor the traffic in its neighborhood, its can update its pseudonyms to remain hidden in a crowd of effective size. However, if none of vehicles in the neighborhood are communicating, the vehicle may not be able to decide upon updation. The actual effective size cannot be gauged by the vehicle itself as it knows only of the existence of the vehicles that are communicating. When a vehicle can sense the communication and observe the vehicles in is proximity physically, then, it can measure the effective size of the neighborhood and take decision of pseudonym updation to maximize its anonymity with minimum number of updations.

5. CONCLUSIONS

The paper dwelt on the issue of maximization of anonymity of a vehicle through enhancement of the effective crowd size around the vehicle. The transmissions of the vehicle prior to becoming a part of a zone and the other communications emanating from the zone entail that all the vehicles present in the zone of anonymity do not contribute to the effective size of the anonymity set. In accordance with the number of vehicles observed along with the target vehicle and the need for transmission by this vehicle, it may change its pseudonym. This change offsets the correlation of the target vehicle with its a priori group and makes its indistinguishable in the larger crowd by increasing its entropy i.e. its effective anonymity set. An analytical as well as simulation based determination of increase in effect size of anonymity set through pseudonym changes. This confirms the efficacy of the proposed heuristics for switching the pseudonym at specified time given a critical mass of neighbor vehicles to maximize anonymity with minimum updations. Further research is required to determine the effect of node velocity and direction of a vehicle after a pseudonym change to avoid predictive correlative tracking.

Digital Object Identifier: 10.4108/ICST.WICON2008.4904
http://dx.doi.org/10.4108/ICST.WICON2008.4904

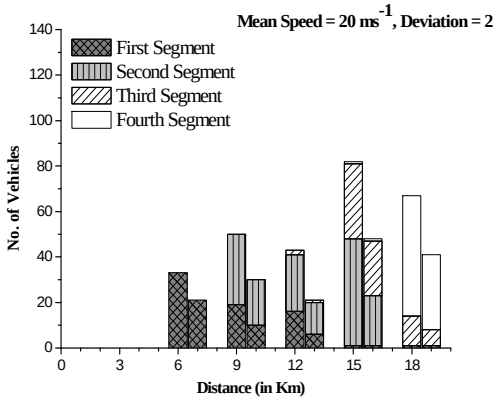

Fig. 2a: Mean population around the vehicle under observation.

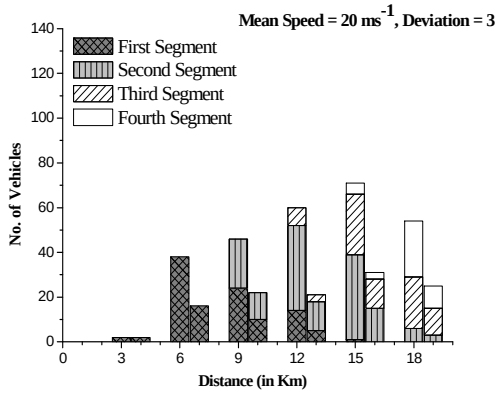

Fig. 2b: Mean population around the vehicle under observation.

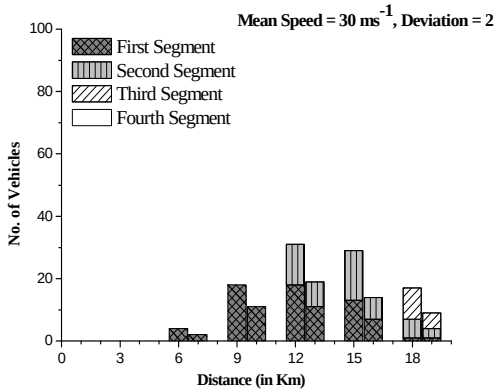

Fig. 3a: Mean population around the vehicle under observation.

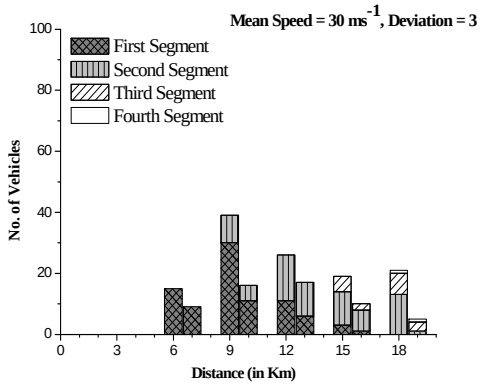

Fig. 3b: Mean population around the vehicle under observation.

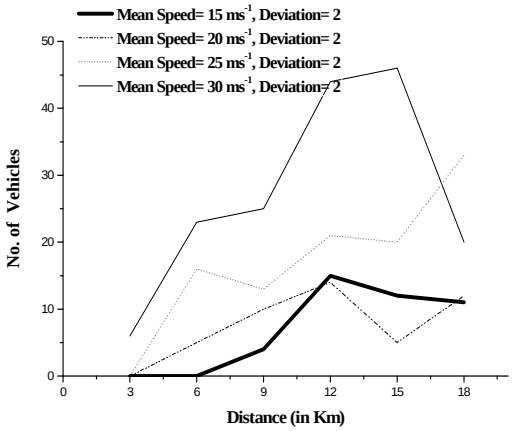

Fig. 4 : Mean node population on the road at various distances from the

6. REFERENCES

[1] Li , M., Sampigethaya, K., Huang, L. , Poovendran, R. 2006. Swing & swap: user-centric approaches towards maximizing location privacy. WPES '06: Proceedings of the 5th ACM workshop on Privacy in electronic society (Oct. 2006), 19-27.

[2] Bettstetter, C. , Resta, G. and Santi, P. 2003. The node distribution of the random waypoint mobility model for wireless ad hoc networks. IEEE Trans. on Mobile Computing, 2(3):257–269.

[3] Huang, L. , Matsuura, K. , Yamane, H. and Sezaki., K. 2006. Silent cascade: Enhancing location privacy without communication qos degradation. In Proc. of Security in Pervasive Computing (SPC), 165–180.

[4] Sampigethaya, K., Huang, L., Li, M. , Poovendran, R., Matsuura, K. and Sezaki, K. 2005. CARAVAN: Providing

Digital Object Identifier: 10.4108/ICST.WICON2008.4904
http://dx.doi.org/10.4108/ICST.WICON2008.4904

location privacy for VANET. In Proc. of Embedded Security in Cars(ESCAR).

[5] Gruteser, M. and Grunwald, D. 2003. Anonymous usage of location-based services through spatial and temporal cloaking. In Proc. of ACM MobiSys, 31–42.

[6] Samarati, P. and Sweeney, L. 1998. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. In Tech. Report SRI-CSL-98-04, CS Lab, SRI International.

[7] Wu, Xiaoxin. and Bertino, Elisa.2007. An Analysis Study on Zone-Based Anonymous Communication in Mobile Ad Hoc Networks. IEEE Trans. On Dependable and Secure Computing, vol. 4, no. 4, (Oct-Dec 2007), 252-264.

[8] Khabazian, M. and Ali, M. K. 2007. Generalized Performance Modeling of Vehicular Ad Hoc Networks (VANETs). ISCC-2007, 51-56.

[9] Trivedi, K. S. 2002. Probability and Statistics with Reliability, Queuing, and Computer Science Applications. John Wiley & Sons, 2nd ed.

Digital Object Identifier: 10.4108/ICST.WICON2008.4904
http://dx.doi.org/10.4108/ICST.WICON2008.4904