8th International Conference on Computer and Knowledge Engineering (ICCKE 2018), October 25-26 2018, Ferdowsi University of Mashhad

A Privacy Preserving Method for Crowdsourcing in Indoor Fingerprinting Localization

Nasim Alikhani [1], Vahideh Moghtadaiee[1], Amir Mahdi Sazdar [2], Seyed Ali Ghorashi[1, 2]
[1]: Cyberspace Research Institute, Shahid Beheshti University G. C. Tehran, Iran
[2]: Cognitive Telecommunication Research Group, Department of Electrical Engineering, Shahid Beheshti University G. C. Tehran, Iran
nassimalikhani@gmail.com; v_moghtadaiee@sbu.ac.ir; sazdar@gmail.com; a_ghorashi@sbu.ac.ir

***Abstract*—Localization services have gained popularity in recent years to facilitate the daily lives of users. With increasing people desire to use Location Based Services (LBSs), the privacy of users has become critical. Most of these services thus use an anonymizer between Location Service Provider (LSP) and the user to protect the user's identity from LSP. One of the localization techniques in indoor environments is Wi-Fi based location fingerprinting which uses received signal strengths (RSSs) at different locations. In this paper, we propose a method to preserve the privacy of users from anonymizer. Hilbert curve and double encryption technique are used. The simulation results indicate that by using the proposed privacy preserving method, the level of privacy preserving is increased.**

Keywords— Indoor Localization; Privacy Preserving; LBS; Hilbert Curve.

I. INTRODUCTION

By expanding the scope of localization systems' applications in indoor environments, the user's privacy in these systems has attracted more attention. Smartphones have a significant role in everyday activities. In spite of their high efficiency and ease of use, the users' important information including their habits, interests, activities and personal information may be exposed to the attackers. Therefore, users should be able to control their data and thus privacy policy should be considered in systems that are relevant to user's data in order to maintain the user's privacy.

Indoor localization has been developed by increasing the use of smartphones in these areas. Due to the failure of GPS/GNSS technologies for indoor environments, non-satellite-based navigation technologies are employed which utilize other existing signals such as Wi-Fi, Bluetooth, FM radio, radio-frequency identification (RFID), sound, light, magnetic field, etc. [1-4]. Among them, Wi-Fi signals have attracted more attention since they can easily be sensed by smartphones and used for localization purposes [5].

Traditional outdoor localization techniques such as time of arrival (TOA), time difference of arrival (TDOA), and angle of arrival (AOA) require line-of-sight (LOS) measurements. However, these methods do not work well in indoor areas, because of non-line-of-sight (NLOS) and multipath issues. One of the widely used methods for localization in indoor environments is location fingerprinting. There are two stages of training and location estimation in this method [1-4]. The received signal strengths (RSSs) are measured at some known locations called Reference Points (RPs) and sent along with their locations to the Location Service Provider (LSP) to build a radio map of the desired area. Then in the location estimation stage, test points' (TPs) locations can be estimated by measuring the RSS values at TPs and comparing that with the stored fingerprints in the radio map using different pattern recognition techniques. Creating a radio map in fingerprinting method is costly and time consuming. Therefore, crowdsourcing method has been suggested in order to collect data and build the database [6]. Crowdsourcing is a measurement process in which individuals participate to sense and send data using their smartphones. One of the benefits of crowdsourcing is the potential of involving a large number of users to measure and report data, which helps building a radio map with a very low cost. The most important challenge of crowdsourcing method however is violating users' privacy by exposing their personal information and locations to LSP [6].

There are some approaches for privacy preserving that have been suggested so far. Authors in [7] and [8] used Hilbert curve as a method to transform the real location of users to a new location, using Hilbert transformation. Another privacy preserving method has been proposed to use Bloom filter to anonymize the location of users [9]. Some other methods for preserving privacy in crowdsourcing networks have been mentioned in [10]. In addition, authors in [11] suggested a privacy preserving technique for outdoor environments using GPS signals, in which three different servers are considered to protect users' privacy. These servers are a Function Generator (FG), an anonymizer and a LSP, all of which operate independently for better preserving the privacy of user. They used Hilbert curve to protect the location privacy from anonymizer. They also assumed that no servers are trusted unlike other approaches that assumed anonymizer as a trusted party [12, 13]. In a Trusted Third Party (TTP)-based schemes, an anonymizer is located between LSP and a user as a trusted entity to hide the identity of the user from LSP [13]. In this paper we consider three server entities operating independently for better preserving the privacy of the user that use Hilbert curve in Wi-Fi fingerprinting indoor localization. To the best of our knowledge, this is the first time they are used in indoor positioning with Wi-Fi signals. Authors in [7] considered Hilbert curve, when they had a LSP and an anonymizer without FG, but in this paper we consider FG server for securing the identity of user from LSP.

978-1-5386-9569-2/18/$31.00 ©2018 IEEE

The rest of the paper is organized as follows. In section II the basic concepts of privacy and Hilbert curve are overviewed. Then in section III the proposed method is described. Simulation results are discussed in section IV. Section V concludes the paper.

II. BASIC CONCEPTS

The main algorithm employed in the proposed method is stated here.

A. Hilbert Curve

An effective algorithm to obscure the location of users is the use of Hilbert curve. As it is shown in [11], this curve transfers the coordinates of the user in 2D to the Hilbert curve with only the parameters associated with this curve. By using this transformation, the location can be considered as an encrypted coordinates. Two adjacent users in 2D space are likely to be in the vicinity of each other in the transformation Hilbert curve [11, 12]. The initial orientation, the starting point, the Hilbert square value and its scale for each square are necessary for creating this curve. As discussed in [7], the N^{th} order of Hilbert curve for d-dimensional is defined by ${H_d}^N$ where $N \geq 1$ and $d \geq 2$. The Hilbert curve ${H_2}^N$ divides the whole curve into 2^{2N} squares [11]. After mapping the entire environment into Hilbert curve, each of the squares will have a unique number.

Fig. 1 depicts an example for the first, second and third order of the Hilbert curve. As it is shown, the orders of central points are defined by drawing a line among central point of each square. For the first order curve, there are four squares and each square is subdivided to four sub-squares to get the higher orders of the Hilbert curve [8].Maintaining the Integrity of the Specifications.

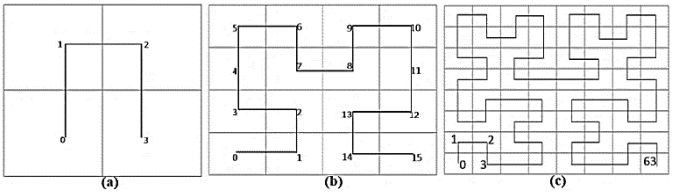

Fig. 1. An example of the Hilbert curve of (a) first, (b) second, and (c) third order.

An example of second order Hilbert curve is shown in Fig. 2. Based on the Hilbert curve ${H_2}^{2N}$, the whole area is divided into 2^4 square indices with the same intervals. Among the locations created by the Hilbert curve, some of them are considered by the LSP as a point of interest. In this figure, these points of interest are shown in English alphabet. For example, there are three points of interest (7, 5 and 0) for point (Q), which are marked by a, c and e [7].

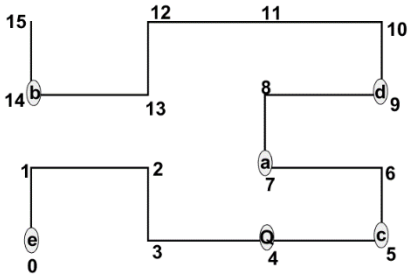

Fig. 2. An example of the second order Hilbert curve [12].

The pseudo-locations of the users are produced based on Hilbert parameters [11]. There are four parameters associated with the Hilbert curve function: **1)** The starting point of the curve (X_l, Y_l), **2)** The curve order (N), **3)** Starting orientation of the curve σ showing the curve is going clockwise or counters clockwise, **4)** The curve scale factor U. These parameters are referred to the spatial transformation parameters.

The Hilbert function is a one-way function which means that we are not able to calculate the real location with no information about the spatial transformation parameters. Therefore, if there is an attacker, it will not find the true location of the user without knowing the used Hilbert parameters. If we consider a user in a point $s = (x_s, y_s)$, the transformed point of s, $< x_s, y_s >$, using Hilbert curve is carried out as follows [11]:

$$< x_s, y_s >= \left\lfloor \frac{(x_s, y_s) - (x_0, y_0)}{U} \right\rfloor \quad (1)$$

where (x_0, y_0) is the coordinates of the left side of the Hilbert square which has coordinate $< 0,0 >$ in the Hilbert curve or the real coordinates $(0,0)$.

If we assume that the origin of the coordinates is in a square with center coordinate equal to (x_c, y_c), the difference between the real coordinates of a user in a square to the center of the same square can be written as [11]:

$$(x_c, y_c) = U \times < x_s, y_s > + (x_0, y_0) \quad (2)$$

$$(x_s, y_s)' = (x_s, y_s) - (x_c, y_c) \quad (3)$$

III. PROPOSED METHOD

There are two stages for fingerprint base positioning, training and location estimation stage. We aim to protect the user privacy in both stages. In the training stage, when users are involved in the crowdsourcing network, their privacies are preserved by utilizing Hilbert curve. Then in the location estimation stage the privacy of users is protected by using both double encryption technique and Hilbert curve. There are three main parts for the proposed method, an anonymizer, a LSP and a FG, in order to protect the user privacy as it is shown in Fig 3.

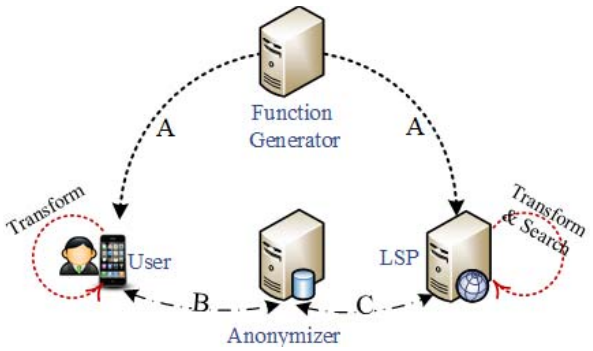

Fig. 3. The entities and steps of the proposed method.

The role of FG is to produce the parameters needed for the Hilbert curve for the users and LSP at any given time interval. The purpose of FG is to keep the real location of the user hidden from the anonymizer. By using these parameters, the pseudo-location of the user is constructed. The anonymizer's task is to hide the identity of the user from LSP and LSP calculates the estimated position of the user employing user's RSS vector. The significance of this method is that the anonymizer knows the identity of the user but never finds out the real location of him, and the LSP never realizes the identity of the user although it knows the estimated location of him.

Anonymizer always changes the ID of the user by a tag before sending data to the LSP. Therefore, not only the identity of the user is hidden but also the anonymizer could recognize whose data has been sending back from LSP in the location estimation stage.

A. Radio Map Construction Stage

Here we consider that the desired area has a square shape similar to the assumption in [10] and all users use the same spatial transformation parameters. Also we consider that anonymizer is reliable. If we desire a higher level of privacy preserving and want to not trust to anonymizer, we can change the parameters of Hilbert curve periodically so that the anonymizer could not know the Hilbert spatial transformation parameters.

Algorithm 1

Step A:
1) User and LSP: Both of them receive parameters of Hilbert curve from FG.
2) User: Finds his location.
3) User: Measures RSS vector.

Step B:
1) User: Sends his transformed location and RSS vector to anonymizer.
2) Anonymizer: Anonymized transformed location of the user by k-anonymity method.

Step C:
1) Anonymizer: Sends the anonymized transformed location and RSS vector to LSP.
2) LSP: Calculates the real location of user by parameters received from FG in Step A1.
3) LSP: Stores this real location with RSS vector in a radio map.

The steps of our suggested method are briefly shown in the Algorithm 1 and are explained as follows:

Step A:

1. User receives the spatial transformation parameters from FG. LSP also receives these parameters from FG to calculate the Hilbert curve at the same time.

2. User finds his location by tile counting (longitude and latitude) from a known point which has been formerly agreed between the user and LSP. Then he transforms his location to the Hilbert coordinates by using the Hilbert parameters.

3. User measures the RSS values from all available APs for his RSS fingerprint.

Step B:

1. User sends his transformed location along with his fingerprinting measures to the anonymizer.

2. In this step we used k-anonymity approach. The main idea of k-anonymity is to hide the private information of the user between 1-k users. The probability of identifying the user's identity or location is $1/k$ [7]. So in this step, Anonymizer knows the indices of Hilbert squares of some previous users because anonymizer can save the Hilbert squares indices for users that had location query with anonymizer and LSP before. Therefore, the anonymizer finds the closest user to the main user's location using NN method by calculating Euclidean distance between the coordinates of each two users in the Hilbert square. NN method is a deterministic method that calculates the shortest distance and finds the nearest location [14]. The anonymizer then selects the location of the nearest user as a fake location of the user. The Euclidean distance between the main user and the other i users in the same Hilbert square is calculated as follows:

$$d_i = \sqrt{(\hat{x}_{TP} - x_i)^2 + (\hat{y}_{TP} - y_i)^2} \quad (4)$$

where $(\hat{x}_{TP}, \hat{y}_{TP})$ is the location of the main user (TP) and (x_i, y_i) is the coordinates of the i^{th} user. If we assume that we have h number of users in each Hilbert square, anonymizer can estimate the fake location of the main user in each Hilbert square as shown below:

$$index_{estimated} = \arg\min_{i,1\le i\le h} d_i \quad (5)$$

where $index_{estimated}$ shows the estimated index of location in the estimated Hilbert square for a fake one.

Using this technique, the anonymizer is able to hide the identity of the user from the LSP by a probability that is proportional to the inverse number of users in the same Hilbert's index.

Step C:

1. Anonymizer sends this fake pseudo-location along with the RSS vector measured by user to LSP.

2. After receiving this information from the anonymizer, the LSP is able to transform the location to the real coordinate of the user by utilizing the Hilbert parameters that agreed with user by interacting with FG.

3. Finally LSP stores the real location along with the fingerprint vector to build a radio map in the training stage

anonymously. For the location estimation stage, the following sub-section explains how to preserve the users' privacy.

B. Location Estimation Stage

For privacy preserving of the user, it runs a double encryption technique [15] on its RSS vector. By this method user first encrypts its RSS vector by the public key of LSP so that no one else could decrypt that. Then it encrypts that again by the anonymizer's public key. The user also encrypts its ID with the public key of the anonymizer. Then it sends its data to the anonymizer.

Anonymizer decrypts the received data by its private key and obtains an encrypted data with the ID of the user. It then changes the user's ID with a tag and sends the rest of the information to LSP. LSP decrypts the RSS vector by its private key and then estimates the location of the user. It also transforms the estimated location into Hilbert space and sends it back to the anonymizer. Therefore, LSP cannot find out who is in this location. The anonymizer gets the transformed location and sends it to user. Then, user transforms the location to a real location.

Therefore, by using the double encryption technique, the anonymizer cannot find out the RSS vector and LSP cannot recognize the identity of the user and a double protection layer will be achieved.

IV. SIMULATION RESULTS

In order to evaluate the proposed method, we consider a 250 × 250 m2 environment with no walls and simulate it using MATLAB. We assumed four sets of radio maps with 200, 400, 1000 and 2000 users with a 0.5m step in both x and y axis, 20 TPs and 100 APs. We use the path loss equation which has been extensively used for constructing the radio map indoor environments [16, 17].

The Hilbert parameters in our simulation are shown in Table I. As can be seen, Hilbert curve order is 2, so the whole space is partitioned into $2^{2\times2}$ square indices. We use 200, 400, 1000 and 2000 users to simulate this method and consider randomly chosen number of users in each Hilbert square for the analysis of anonymizer to estimate the fake location of the user. These numbers are 5, 10, 30 and 60 when there are 200, 400, 1000, and 2000 users, respectively. The anonymizer can compute the fake location using equation (5). The value of h in Equation (5) is then 5, 10, 30 and 60 in each radio map representing the number of users in each of 16 squares of Hilbert curve.

TABLE I. THE PARAMETERS OF THE HILBERT CURVE IN THE SIMULATION

Scale factor (U)	6.25
Order orientation	Clockwise
Starting point	(0,0)
Hilbert curve order	2

As it is illustrated in Fig.4, by increasing the number of users, the fake point that anonymizer selects (using NN method) is closer to the actual location of the user, since the density of the users is getting higher in one Hilbert square. Therefore, the average location distance is getting less than before. However, the anonymization level of identity of the user increases because the identity of the user can be identified by a probability equal to $1/k$ in k-anonymity.

The average location distance differences for 200, 400, 1000, and 2000 users are 2.73, 2.06, 1.61 and 1.2 m, respectively. If we have less number of users in each Hilbert square, the privacy of the user also has been less preserved in terms of identity.

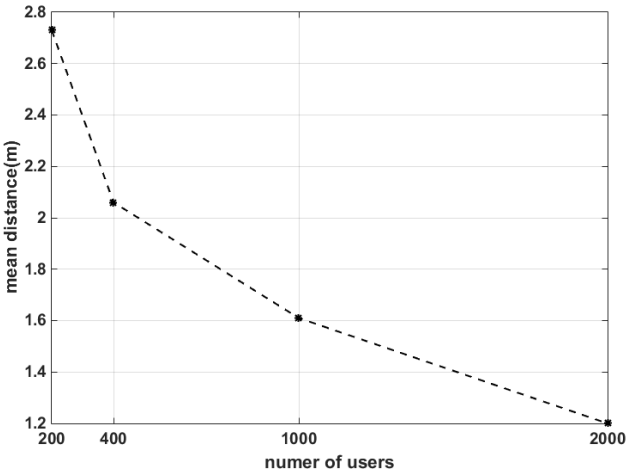

Fig. 4. The mean distance calculated for 200 to 2000 users.

Fig. 5 depicts the cumulative distance calculated for all four radio maps. The highest distance calculated in the radio maps with 200, 400, 1000 and 2000 users are 10.67m, 6.36m, 4.98m and 3.08m. As shown in Fig. 5, by increasing the number of users in each Hilbert square, the highest distance difference is reduced since the distance between the true location and the anonymized location reduces.

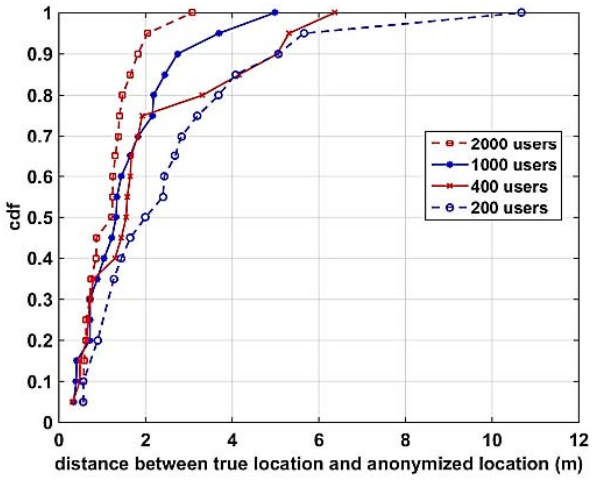

Fig. 5. Calculated cumulative distance for 200 to 2000 users.

Fig. 6 also shows the required amount of time for the anonymizer to find the nearest user in the Hilbert square rises by increasing the number of users. The spending times in the anonymizer are 0.036, 0.037, 0.039, and 0.047 seconds for the 200, 400, 1000, and 2000 users, respectively.

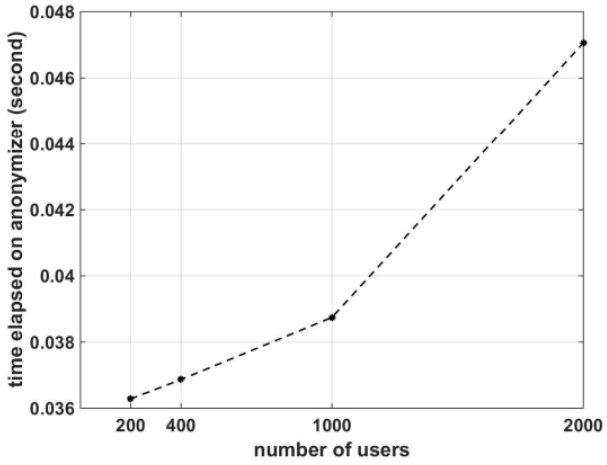

Fig. 6. The time (sec) spent on the anonymizer with respect to the number of users.

V. CONCLUSION

This paper proposed a method to protect the user privacy both in training and location estimation stages of location fingerprinting. The proposed method consists of an anonymizer which is unaware of the user's real location and also a LSP that knows nothing about the identity of the user. The simulation results indicate that increasing the number of users makes the error distance lower, because users' density is higher in each Hilbert square. In addition, the time spent on anonymizer entity becomes longer, therefore, finding the fake location for the user in the anonymizer takes more time than before. By increasing users in crowdsourcing, we have higher level of anonymization for user's identity and LSP can recognize the identity of the user with a lower probability in the anonymizing process.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support provided by the Iran National Science Foundation (INSF) for this work.

REFERENCES

[1] S. He and S.-H. G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," IEEE Communications Surveys & Tutorials, vol. 18, pp. 466-490, 2016.

[2] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern WLAN Fingerprinting Indoor Positioning Methods and Deployment Challenges," IEEE Communications Surveys & Tutorials, 2017.

[3] V. Moghtadaiee and A. G. Dempster, "Indoor location fingerprinting using FM radio signals," IEEE Transactions on Broadcasting, vol. 60, pp. 336-346, 2014.

[4] Z. E. Khatab, V. Moghtadaiee, and S. A. Ghorashi, "A fingerprint-based technique for indoor localization using fuzzy Least Squares Support Vector Machine," in Electrical Engineering (ICEE), 2017 Iranian Conference on, pp. 1944-1949, 2017.

[5] Y. Liu, M. Dashti, M. A. A. Rahman, and J. Zhang, "Indoor localization using smartphone inertial sensors," in Positioning, Navigation and Communication (WPNC), 2014 11th Workshop on, pp. 1-6, 2014.

[6] B. Wang, Q. Chen, L. T. Yang, and H.-C. Chao, "Indoor smartphone localization via fingerprint crowdsourcing: Challenges and approaches," IEEE Wireless Communications, vol. 23, pp. 82-89, 2016.

[7] A. Khoshgozaran and C. Shahabi, "Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy," in International Symposium on Spatial and Temporal Databases, pp. 239-257, 2007.

[8] J. K. Lawder, "Calculation of mappings between one and n-dimensional values using the hilbert space-filling curve," School of Computer Science and Information Systems, Birkbeck College, University of London, London Research Report BBKCS-00-01 August, 2000.

[9] A. Konstantinidis, G. Chatzimilioudis, D. Zeinalipour-Yazti, P. Mpeis, N. Pelekis, and Y. Theodoridis, "Privacy-preserving indoor localization on smartphones," IEEE Transactions on Knowledge and Data Engineering, vol. 27, pp. 3042-3055, 2015.

[10] I. J. Vergara-Laurens, L. G. Jaimes, and M. A. Labrador, "Privacy-preserving mechanisms for crowdsensing: Survey and research challenges," IEEE Internet of Things Journal, vol. 4, pp. 855-869, 2017.

[11] T. Peng, Q. Liu, and G. Wang, "Enhanced location privacy preserving scheme in location-based services," IEEE Systems Journal, vol. 11, pp. 219-230, 2017.

[12] A. Khoshgozaran and C. Shahabi, "Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy," Advances in Spatial and Temporal Databases, pp. 239-257, 2007.

[13] G. Ghinita, K. Zhao, D. Papadias, and P. Kalnis, "A reciprocal framework for spatial k-anonymity," Information Systems, vol. 35, pp. 299-314, 2010.

[14] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, "Introduction to algorithms mit press," Cambridge MA, 1990.

[15] I. J. Vergara-Laurens, D. Mendez, L. G. Jaimes, and M. Labrador, "A-PIE: An algorithm for preserving privacy, quality of information, and energy consumption in Participatory Sensing Systems," Pervasive and Mobile Computing, vol. 32, pp. 93-112, 2016.

[16] S. Y. Seidel and T. S. Rappaport, "914 MHz path loss prediction models for indoor wireless communications in multifloored buildings," IEEE transactions on Antennas and Propagation, vol. 40, pp. 207-217, 1992.

[17] A. Szajna, M. Athi, A. Rubeck, and S. Zekavat, "2.45 GHz near Ground Path Loss and Spatial Correlation for Open Indoor and Snowy Terrain," in Vehicular Technology Conference (VTC Fall), 2015 IEEE 82nd, pp. 1-5, 2015.