Influence Maximized MCL based RSU Deployment

Jeonghee Whang[1], Soyoung Park[2] and Jeonghee Chi[3*]

[1]*Department of Computer Science,*
NamSeoul University, South Korea
jhwhang@nsu.ac.kr
[2,3]*Department of Internet and Multimedia Engineering,*
Konkuk University, South Korea
{soyoungpark, jhchi}@konkuk.ac.kr

Abstract

RSUs (Road Side Units) that can support seamless communications between vehicles on VANET environment are usually installed in the intersections. However, it cannot be installed on whole intersection because of installation costs. Therefore, it is essential to allocate RSUs on important positions. In this paper, we find RSU candidates by clustering intersections based on the connectivity of them by using influence maximized MCL algorithm, and then finally identify RSU allocation locations according to the size of cluster and the RSU priorities. In our experiment, we showed that our approach had outstanding performance compared to K-Means-based algorithm.

Keywords: Roadside unit allocation, Influence based Markov Clustering

1. Introduction

Vehicular Ad-hoc network environments (VANET) that allows wireless communications between vehicles-to-vehicles or vehicles-to-Infrastructure provide more secure and useful traffic services to drivers. Roadside units (RSU), a type of traffic message transmitter statically located on the roads, are one of the most essential components to establish the VANET environment for supporting seamless traffic message communication [1].

Recently many researches on efficient RSU deployment have emerged [2-5], and we have also proposed a couple of RSU allocation strategies [6-8] which can cover intersections based on the concept of intersection connectivity. The intersection connectivity means the probability of message transmission between any two intersections [2]. Based on the intersection connectivity, we can find several intersections with relatively big traffic inflow from other intersections. Then, we can cluster intersections centering for the point of intersections absorbing traffics. The existing MCL based allocation approaches select the intersections after the final convergence by many repetition, while do not consider the influential spread factor. In this paper, we improved the existing MCL based technique to consider the influential spread on roadway network.

First, in order to find the clusters efficiently, we propose a RSU allocation technique based on the influence maximized MCL (I-MCL). The I-MCL is used in a marketing analysis that uses the pre-constructed social networks to perform marketing with small cost while maximizing the spread [9]. As shown in Figure 1, the roadway network has a similar structure with social network services. We find the clusters using I-MCL technique that can find the most influential intersection in roadway network.
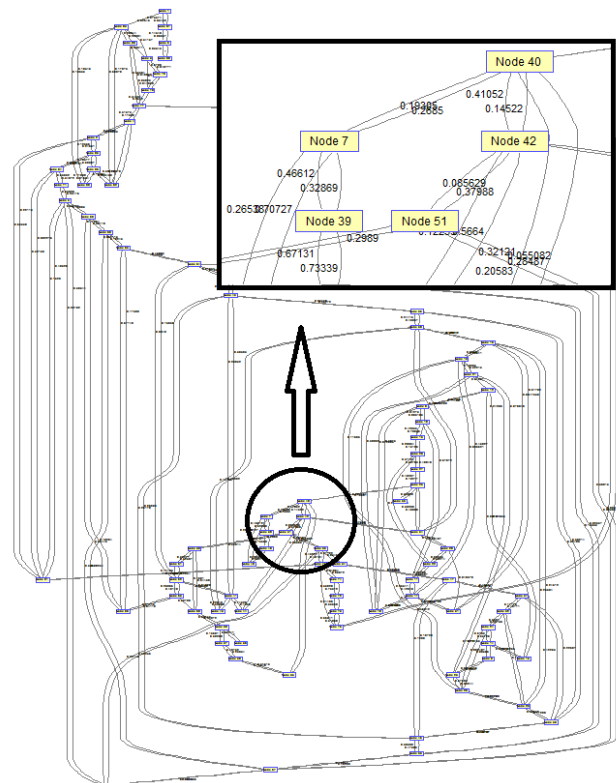
---

* Corresponding author

Figure 1. Roadway Network with Intersection Influences

We simulate the performance of our proposed algorithms using with real intersection and traffic data about Seocho-gu in Seoul in Korea. In addition, we analyze the performance of the proposed algorithm with k-means clustering based algorithm.

This paper is composed as follows. We review some related works in Section 2, and describe the concept of intersection connectivity in Section 3. Our proposed RSU allocation algorithm is explained in Section 4 in details, and Section 5 analyzes the simulated results of our method. Finally, we conclude our paper in Section 6.

2. Related Works

Recently, many algorithms that deploy the RSUs according to several factors related to traffic information has been suggested. K. Barrachina *et al* in [2] suggest deploying RSUs using an inverse proportion to the expected density of vehicles. However, this approach simply considers the density of vehicles, while neglecting the fact that a vehicle can carry and forward the information to the adjacent intersection. M. Rios *et al* in [3] suggested placing the RSUs so that a vehicle can communicate with the RSU within 2 hops in a bus transport network. This method only considers communication among the buses and thus cannot be applied to roads with high traffic volume. P. Patil *et al* in [4] proposed Voronoi-based RSU deployment algorithm, and C. M. Silva *et al* [5] suggested RSU deployment algorithm derived from partitioning the roads into grid cells. Using the migration ratios between the adjacent cells has been suggested. However, these approaches only consider the location of the roads without the connectivity of these roads.

We have introduced the concept of intersection connectivity, and we have suggested RSU allocating mechanism based on the intersection priority in [1]. However, the goal of these algorithms is to find RSU locations which can cover the entire intersections on the roads. In [6,7,8], the intersection connectivity has been redefined as considering the traffic flow, and proposed MCL-based intersection clustering mechanism. In this paper, we

improved the existing our MCL based RSU candidate selection algorithm in order to select a RSU candidate with more influential.

3. Definitions and Notations

In this section, we introduce the intersection connectivity. Table 1 summarizes notations used throughout the paper.

Table 1. Notations

Notation	Description
I	A set of all intersections
I_i	The ith intersection in I where $i = \{1,\ldots,n\}$
$N_{i,j}$	The jth intersection adjacent to I_i where $j=\{1,..,m\}$
$P_{i,j}$	The traffic flow probability from I_i to $N_{i,j}$
$D_{i,j}$	The distance ratio from I_i to $N_{i,j}$
$C_{i,j}$	The intersection connectivity from I_i to $N_{i,j}$
RH	A set of RSU candidates, which is a set of cluster heads
RC_i	A set of intersections belonging to the ith cluster
RS	A set of final RSU placement locations

We define the intersection connectivity as the probability that the traffic at I_i can be transferred to $N_{i,j}$, for any two adjacent intersections I_i and $N_{i,j}$. The intersection connectivity $C_{i,j}$ is determined by traffic flow probability $P_{i,j}$ and the distance $D_{i,j}$ between the two intersections and is defined as in (1).

$$C_{i,j} = P_{i,j} * D_{i,j} \qquad (1)$$

As shown in the Figure.2, intersection connectivity $C_{i,4}$ between I_i and $N_{i,4}$ is determined by the traffic flow probability $P_{i,4}$ and the distance $D_{i,4}$.
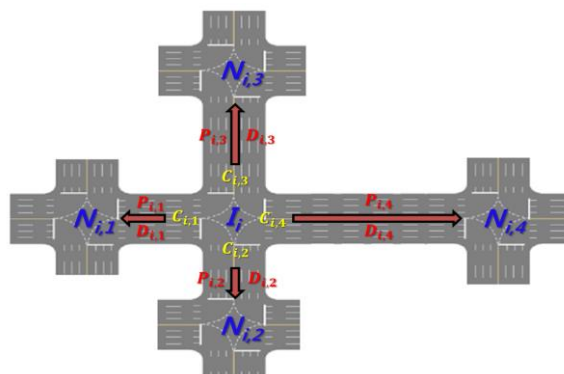

Figure 2. Intersection Connectivity at I_i

The traffic flow probability $P_{i,j}$ in (1) means the probability of traffic data transferring from I_i to $N_{i,j}$. We assumed that the traffic data delivery between two adjacent intersections can be carried out by vehicles passing through the two intersections. In order to acquire $P_{i,j}$, it is required to calculate how many vehicles move towards $N_{i,j}$ from I_i firstly, and to calculate the probability of how they are connected to vehicles going from I_i

to $N_{i,j}$. For this, we have defined traffic influence $\alpha_{i,j}$ and used Poisson distribution to calculate traffic flow probability.

$$\alpha_{i,j} = \frac{n \sum_{t=1}^{n} x_{i,j}^t y_i^t - \sum_{t=1}^{n} x_{i,j}^t \sum_{t=1}^{n} y_i^t}{n \sum_{t=1}^{n} (x_{i,j}^t)^2 - (\sum_{t=1}^{n} x_{i,j}^t)^2} \tag{2}$$
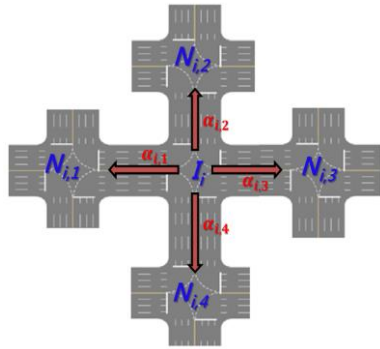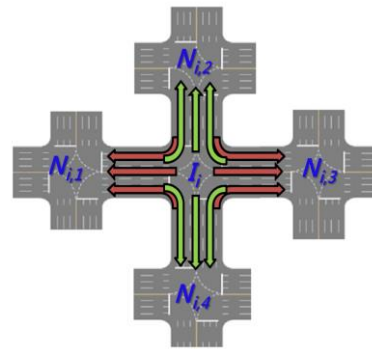

Figure 3. Traffic Influence at I_i   **Figure 4. Traffic Flow at I_i**

First, the traffic influence from I_i to $N_{i,j}$, denoted as $\alpha_{i,j}$, means the ratio of vehicles migrating from I_i to $N_{i,j}$. It can be obtained by a linear regression analysis about the traffic amount heading for $N_{i,j}$ over the total traffic volume at I_i and is as defined in (2). In (2), the $x_{i,j}^t$ is the total volume of traffic moving from I_i to $N_{i,j}$ at time t, which combines 3 directional traffic flows as in Figure 4, and y_i^t indicates the total traffic volume that pass through I_i at time t. For example, traffic influence $\alpha_{i,2}$ from I_i to $N_{i,2}$ in Figure 3 is determined by vehicles passing through I_i from $N_{i,1}$, $N_{i,3}$, and $N_{i,4}$ as shown in Figure 4.

The traffic flow probability can be defined as the probability of a vehicle V_i at I_i contacting with at least one other vehicle V_2 moving to $N_{i,j}$ within a given time θ. Consequently, the traffic data collected at I_i by V_i can be transferred to $N_{i,j}$ by V_2. When supposed that n number of vehicles heading for $N_{i,j}$ emerge at I_j by the Poisson distribution, as in (3), the traffic flow probability, denoted as $P_{i,j}$, can be defined as in (4).

$$f(\lambda_{i,j};\ n) = \frac{\lambda_{i,j}{}^n e^{-\lambda_{i,j}}}{n!}, \text{ where } \lambda_{i,j} = M_i * \alpha_{i,j}/\theta \tag{3}$$

$$P_{i,j} \geq 1 - f(\lambda_{i,j};\ n = 0)$$

$$\geq 1 - e^{-\lambda_{i,j}} \tag{4}$$

M_i indicates the total traffic volume measured at I_i during the time θ. Consequently, the probability can be calculated by subtracting the probability of contacting with no vehicles moving from I_i to $N_{i,j}$, in which case of $n = 0$, from 1. For example, as shown in the Figure 5, it indicates that probabilities of vehicle ① meeting vehicles ②, ③, ④ are $P_{i,2}$, $P_{i,3}$, and $P_{i,4}$, respectively.

In addition to the above two factors, we consider the geographical distance between two adjacent intersections. Suppose that there are two roadways I_i to $N_{i,j}$ and I_i to $N_{i,k}$ with different distances. If $P_{i,j}$ and $P_{i,k}$ are the same, there is no difference from the view of intersection connectivity. However, if the distance from I_i to $N_{i,k}$ is relatively smaller than the distance from I_i to $N_{i,j}$, the traffic at I_i can be delivered to $N_{i,k}$ faster with the smaller number of hop-by-hop propagations by vehicles.
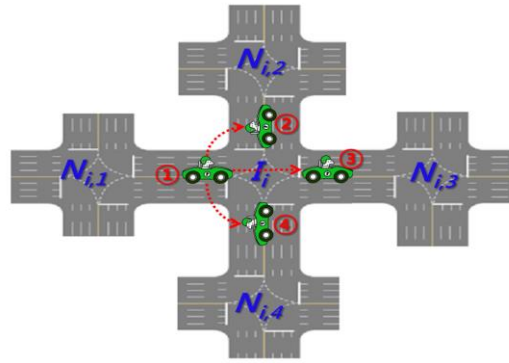
Figure 5. Traffic flow probability at I_i

For an extreme case as in Figure 6, if the distance from I_i to $N_{i,4}$ is within the transmission range of wireless communication, the traffic at I_i can be directly transferred to $N_{i,4}$ by a single hop propagation of vehicles located at I_i. Hence, we can say it definitely that the case from I_i to $N_{i,4}$ has better connectivity than the one from I_i to $N_{i,2}$ and $N_{i,3}$, even though both of them have the same traffic flow probability. Therefore, the distance between two adjacent intersections should be also reflected on determining the intersection connectivity. To achieve this, we define the distance ratio for every adjacent roadway to I_i as follows: For $j=\{1,...,m\}$, let the distance between I_i and $N_{i,j}$ be $DIS_i(N_{i,j})$. Then, the distance ratio of $D_{i,i}$ for $N_{i,i}$ is defined as in (5).

$$D_{i,j} = \frac{DMin(i)}{Dis_i(N_{i,j})},$$

$$\text{where } DMin(i) = \min_{j=\{1,...,m\}}\{DIS_i(N_{i,j})\} \tag{5}$$

Consequently, the intersection connectivity $C_{i,j}$ between I_i and $N_{i,j}$ in (1) is defined using $P_{i,j}$ in (4) and $D_{i,j}$ in (5).
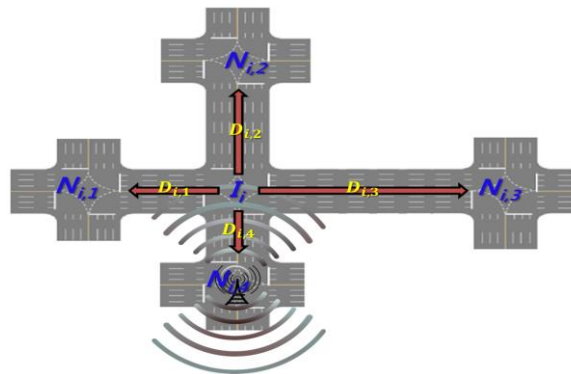

Figure 6. Distance Ratio of Di,j for Ni,j at Ii

4. I-MCL based RSU Selection

In this section, we describe our proposed method in detail. Our technique consists of two stages. The first is RSU candidate selection using I-MCL, and then select Top-k RSU based the priorities of RSU candidates.

4.1. RSU Candidate Selection

The MCL algorithm is which clusters nodes that arrive at the same node after k times of random walks based on the transition probability of a graph. In the MCL algorithm,

intersections and the intersection connectivity can be represented as a weighted-and-directed graph G=(V, E, W) where V is a set of all intersections, E is a set of roadways between two adjacent intersections, and W is a set of intersection connectivity for every edge in E.

The MCL carries out clustering by repeating two functions, expansion and inflation. By the expansion function, the transition probability of a random vehicle from each intersection arriving at a certain intersection is calculated based on the intersection connectivity. This function multiplies the transition matrix by itself. Next, the MCL uses the inflation function to further increase the probability of transitioning to a node of having higher transition probability and to further decrease the probability of transitioning to a node of having a lower transition probability. This process has the traffic flow converge faster to an intersection having maximum traffic inflow. If the transition probability between arbitrary two intersections does not reach the critical value ε, the connectivity of the corresponding edge is eliminated in the given graph. While repeating the above two functions, if the transition matrix does not change significantly, meaning that the difference between i^{th} and $i+1^{th}$ transition matrix is equal to 0, or less than ε, the algorithm ends.

However, the existing MCL based RSU allocation method tends to cause an issue that the higher the number of intersection, the more repetitive in order to converge. I-MCL-based RSU candidate selection in algorithm 1 has been suggested for determining the cluster based on when the influence in each intersection turned out to be the highest without waiting until they converged.

Algorithm 1: I-MCL based RSU Candidate Selection

Input: A transition matrix M and an inflation parameter r.
Output: (RH, RC_j) for 1≤j≤l where l is the total number of clusters, RH is a set of cluster headers, and RC_i is a set of cluster elements belong to the ith cluster

1:	M=M'
2:	Do {
3:	prennz = nnz(M)
4:	M0 = M;
5:	M = M²; // Expansion
6:	for ∀I_i ∈ I do { // Inflation
7:	for ∀I_j ∈ I do {
8:	M[i][j]= (M[i][j]^r) / ($\sum_{k∈v}$M[k][j]^r) ;
9:	if (M[i][j] < ε) then
10:	M[i][j]=0;
11:	} // end for
12:	} // end for
13:	} while (nnz(M) > prennz)
14:	RH = { I_i | ∃$_j$ (M[i][j] ≠ 0)};
15:	RC_i = { I_j | ∃$_i$(M[i][j] ≠ 0) in the jth column for I_i ∈RH};

In algorithm 1, intersection determined in the final convergence phase became cluster head, and we have selected such cluster head when the influence of intersection turned out to be the highest for affecting more intersections. When performing the existing MCL, the number of intersection with non-zero connection between intersections started increasing and decreasing again. When the number of non-zero connection turned out to be the highest, cluster head is selected.
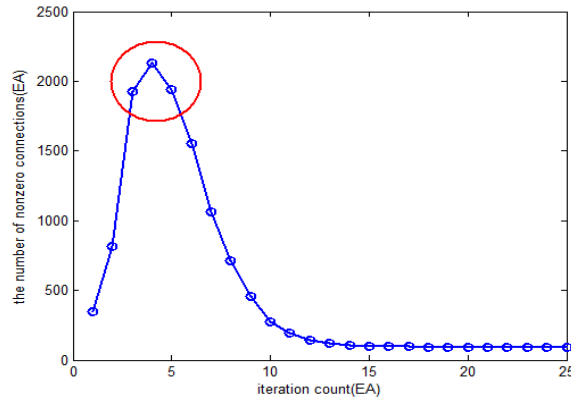
Figure 7. Non-zero Trend of Intersection Connection in Seocho-gu

Figure 7 represents the trend on the number of non-zero connection in Seocho-gu where 89 intersections are established as subjects in the experiment of this paper. When using the existing MCL, converged result was obtained after repeating the procedure for 25 times. If applying I-MCL, the number of non-zero connection became the highest when repeating the procedure for four times. Therefore, this tends to become when the highest level of connection in intersection is obtained. At this time, cluster head is determined with information of intersection, and they are selected as RSU candidates.

4.2. RSU Allocation

Once RSU candidate set are determined, we select Top-K RSUs according to the size of cluster and priorities using final RSU deployment algorithm in [7]. The size of cluster mean that it is possible to cover a large number of intersection traffic information, and the intersections with high priority mean that they should be covered because of the intersections are an important position. We used the combined weight TWx for each RSU candidate Ix using equation (6).

$$TWx = CWx + PXx \qquad (6)$$

Where the CWx is a cluster size-based weight and the PWx is the priority-based weight for each RSU candidate Ix. The intersection priority PWx of Ix represents the importance of each RSU candidates. The priority can be measured by various traffic factors such as traffic volume, intersection popularity (located within popular or famous areas), intersection particularity (located within attention regions such as an accident-prone area), and so on. In this paper, we have considered two traffic factors of 1) traffic volume and 2) the intersection popularity in order to measure the intersection priority of each intersection. The traffic volume is computed by the total number of vehicles passed through an intersection for an hour. In order to measure the popularity of each intersection, we have used public bus trajectories. Public buses move along the most commonly and mainly used roadways by people, and go through popular and famous areas. Hence, the popularity of an intersection is measured by the number of different bus lines passing through an intersection.

5. Experimental Results

We have simulated the performance of our proposed algorithm with real intersections and their traffic data for Seocho-gu in Seoul. We have used 89 intersections and the traffic data collected at each intersection for a day in our experiment. We compared our method with K-means based clustering algorithm that is a distance based clustering technique.

To evaluate the connectivity between the intersection and the RSU within one cluster, we utilized the Dijkstra algorithm to find the shortest path between the intersection and the RSU, and intersection connectivity at this time. In other words, the connectivity within the cluster was calculated for the path with the lowest intersection connectivity. When $path_{i,j}$ is defined as the shortest path within a cluster from intersection I_i to I_j, where the RSU is located at, the average connectivity $Cost_{i,j}$ from I_i to I_j was calculated as shown in (7). The length ($path_{i,j}$) indicates a length of the path from I_i to I_j.

$$Cost_{i,j} = (\prod_{k=i}^{j-1} P_{k,k+1})/length(path_{i,j})$$

(7)

We anlysis the results performed by each method for different number of RSUs. Figure 8 shows the average connectivity among cluster members. As the length of waiting time in intersection becomes longer, the probability of confronting vehicles that move towards the intersection increases. In order to consider the case that a connectivity is low, we set the waiting time in an intersection to 3sec. The results of the experiment showed the proposed method has outstanding performance compared to K-means. K-means have been clustered according to the length without considering the connection between intersections, and hence representing the lowest level of performance.
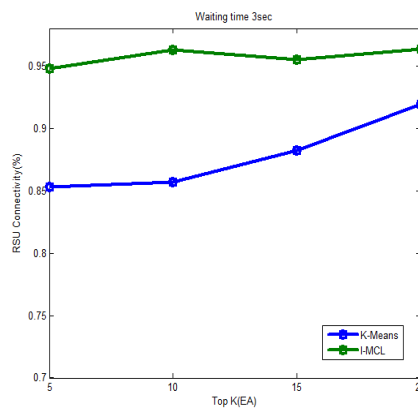

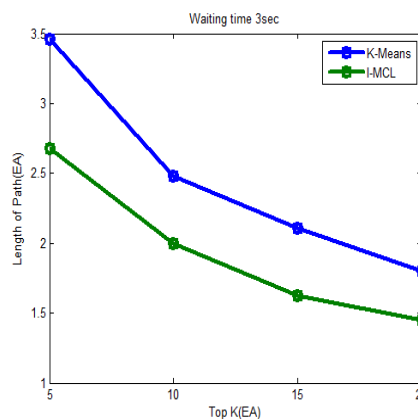**Figure 8. The Average Cluster Connectivity**

Figure 9. The Average Path Length to the Cluster Head

In addition, Figure 9 shows the average path length between RSU (cluster head) and cluster members. As the number of RSU increases, the length of path tends to be shorter.

As for K-means, the length of path turned out to be longer than our technique. Through this experiment, we have shown that as the length of path became longer, RSU connectivity became lower. Therefore, this results mean that our technique can reflect the connectivity between intersections well.
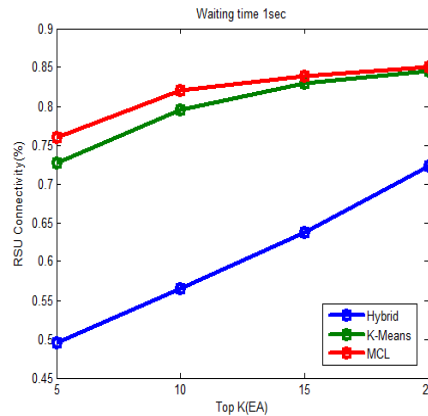

Figure 10. The Case of Worst Connectivity

Figure 10 represents the case of worst RSU connectivity. When installing five RSUs and having one second of waiting time, the worst level of connectivity was derived. In the worst case, I-MCL turned out to represent the highest performance as 76% connectivity, while MCL has 72% connectivity, and K-means indicated the lowest performance as 50%. It means that the proposed method has outstanding performance, even when passing through the intersection at a high speed.

6. Conclusions

In this paper, we proposed the influence maximized MCL-based RSU allocation technique using the intersection connectivity that considers relative distance between two intersections. The proposed technique selects the RSU candidates using the situation when the influence spread is maximized, and then choose Top-k RSUs considering intersection priority. We have analyzed the performance of our I-MCL based RSU allocation mechanism for various simulation criteria with real intersection and traffic data of Seocho-gu. We have also compared the simulated performance of the proposed model with the K-means clustering method that is representative clustering technique base on the distance. Through simulated analysis, we can conclude that our I-MCL based RSU allocation finds RSU locations that can collect traffic data from the more intersections.

Acknowledgment

References

[1] J. Chi, H. Park, T. Hwang and S. Park, "An Extended Intersection-Priority based RSU Allocation for VANET", Information Journal, vol. 17, no. 4, **(2014)**, pp. 1419-1432.
[2] K. Barrachina, P. Garrido, J. Cano and C. T. Calafate, "Road Side Unit Deployment : A Density –Based Approach", IEEE Intelligent Transportation System Magazine, **(2013)**, pp. 30-39.
[3] M. Rios, V. Marianov and M.Perez, "Locating fixed roadside units in a bus transport network for maximum communications probability", Transportation Research Part C, **(2015)**, pp. 35-47.

[4] P. Patil and A. Gokhale, "Voronoi-based Placement of Road-side Units to Improve Dynamic Resource Management in Vehicluar Ad Hoc Networks", IEEE Collaboration Technologies and Systems (CTS), **(2013)**.

[5] C. M. Silva, A. L. L. Auino and W. Meira, "Deployment of roadside units based on partial mobility information", Computer Communications, vol. 60, **(2015)**, pp. 28-39.

[6] J. Chi, S. Do, Y. Kim and S. Park, "Deployment of Roadside Units based on Markov Clustering Algorithm", International Journal of Applied Engineering Research, **(2015)**, pp. 33558-33561.

[7] J. Chi, S. Do and S. Park, "Traffic Flow-based Roadside Unit Allocation Strategy for VANET", International Conference on Big Data and Smart Computing, **(2016)**, pp. 245-250.

[8] Y. Kim, S. Park and J. Chi, "Absorbing Markov Chain-based Roadside Units Deployment", Contemporary Engineering Sciences, vol. 9, **(2016)**, pp. 579-586.

[9] C. Kim, S. Lee, S. Park and S. Lee, "Influence Maximization Algorithm Using Markov Clustering", LNCS 7827, **(2013)**, pp. 112-126.

Authors

Jeonghee Hwang, she got her Ph.D at Chungbuk National University, Korea. She is and Assistant Professor, Namseoul University, Korea.

Research Interests: data mining, semantic web, ubiquitous computing and big data processing.

Soyoung Park, she got her Ph.D at Ewha Womans University, Korea. She was a Post-doctoral Researcher at the University of Central Florida, U.S.A. She is anAssistant Professor, Konkuk University, Korea

Research Interests: Cryptography, Network security, Vehicular network

Jeonghee Chi, she got her Ph.D at Chungbuk National University, Korea. She is an Assistant Professor, Konkuk University, Korea

Research Interests: Data mining, Big data processing, VANET