# Cluster-Based Vehicular Data Collection for Efficient LTE Machine-Type Communication

Christoph Ide, Fabian Kurtz and Christian Wietfeld
Communication Networks Institute
TU Dortmund University
44227 Dortmund, Germany
e-mail: {Christoph.Ide, Fabian.Kurtz, Christian.Wietfeld}@tu-dortmund.de

*Abstract*—**Machine-Type Communication (MTC) poses an ongoing research topic in the development of cellular communication systems. In this context, the efficient collection of extended Floating Car Data (xFCD) via Long Term Evolution (LTE) is a major challenge. In this paper, we present cluster-based xFCD collection schemes in order to form clusters with a long lifetime. As a result, the proposed clustering algorithms reduce the occurring cellular communication traffic. For the performance evaluation of the presented algorithm, a novel system model is used. By means of the system model, the user mobility can be modeled realistically and a precise quantification of the utilization of the LTE network for xFCD transmission is possible. The results show that the LTE network utilization can be significantly reduced by the proposed clustering algorithms.**

## I. INTRODUCTION

The scalability of Machine-Type Communication (MTC) is a recent topic in the development of cellular communication systems [1]. One key requirement for this special type of application is the integration into the communication infrastructure (e.g. Long Term Evolution (LTE)) without or with very low service degradation of other cellular users. This paper is motivated by a novel application of MTC. Traffic forecast systems should be improved by additional information sources. So called extended Floating Car Data (xFCD) consists of sensor data that is collected by vehicles [2]. Different sources of sensor data are conceivable. This could be a smartphone of a vehicle occupant or a dedicated cellular transceiver installed in the car which collects location data (Global Positioning System (GPS), LTE cell ID or WiFi Service Set IDentifier (SSID)) or car related sensor data (e.g. via Controller Area Network (CAN) bus). These could come from a sensor monitoring the velocity, breaks, steering wheel position or light.

One possibility to make these data available at a traffic forecast server, is to transmit the data via public cellular communication systems as Vehicle-to-Infrastructure (V2I) communication [3]. However, in contrast to the communication traffic which is originated by transmissions of xFCD (many users with a low payload), the LTE network is designed for a relatively low number of users with large payload respectively peak data rates [1]. For this reason, MTC may interrupt the Human-to-Human (H2H) communication in the network [4]. In this paper, an approach to minimize the negative impact on H2H communication is presented. We propose efficient algorithms for cluster-based xFCD collection and transmission (cf. Fig. 1). Vehicles should form groups and only one member of each group (Cluster Head (CH) [5]) transmits the data via LTE. xFCD from Cluster Members (CM) is transmitted to the CH via IEEE 802.11p based Vehicle-to-Vehicle (V2V) communication. For the performance evaluation of cluster-based xFCD a novel system model is introduced. This consists of different components in order to model realistic environments as well as user mobility and to quantify the LTE cell utilization.

The paper is structured as follows: In Section II the related work regarding cellular enabled MTC is presented. Next, the coalition-based data aggregation is illustrated in Section III. The novel system model is described in Section IV. Finally, results regarding the performance evaluation of clustering algorithms is presented in Section V.
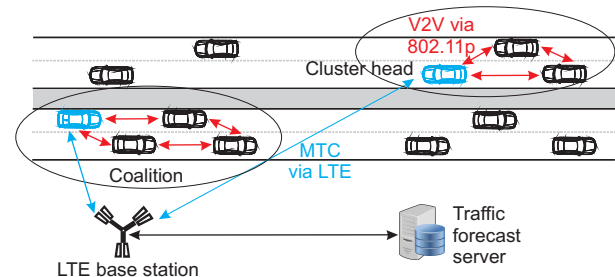


Fig. 1: Scenario with LTE MTC for xFCD Collection and V2V Coalitions.

## II. RELATED WORK

xFCD transmission is one example of MTC, which could be integrated into public cellular networks. MTC is one of the main topics in the standardization process of LTE Rel. 12 [6] [7]. In contrast to H2H applications, MTC services pose very different requirements on a communication system [7] [8]. In addition, new MTC applications must have a very small or even no negative impact on the existing H2H Quality of Service (QoS). Therefore, in [9] a channel-aware transmission scheme is presented which reduces this negative impact. Several clustering algorithms have been proposed for CH section. In [5] a framework for centralized vehicular network organization using LTE is presented. Thereby, CHs collect data from neighboring vehicles by IEEE 802.11p and

transmit the data to a server via LTE. In [10] the CH is selected as function of the channel conditions and in [11] the CH is assigned as the car in the middle of the cluster.

## III. COALITION-BASED DATA AGGREGATION

A main goal of the xFCD collection is a low impairment of other LTE users. Hence, in this paper a coalition-based xFCD transmission is employed. Thereby, the vehicles communicate via V2V communication with each other. This hybrid communication architecture has two main advantages regarding the cellular communication system in contrast to a communication system in which each car transmits xFCD via LTE.

First, this aggregated data can much better be handled in a modern cellular communication system like LTE. These systems are designed for a relatively low number of users with high data rates resp. large payloads. This is reflected in the random access procedure, as well as in the multiple access strategy in terms of scheduled Resource Blocks (RBs). By aggregating the data, the number of active connections can be significantly decreased. Beside this, the total amount of xFCD effecting the same benefits for the dynamic traffic forecast also decreases. This is due to the fact that neighboring cars have similar information about the current traffic situation. Therefore, a CH can perform data compression in order to reduce the redundancy. It is a common approach to model the upper bound of the amount of compressed data as square root function of the number of uncompressed data units [12]. In order to model this data reduction, we approximate the compressed xFCD payload $PL_c$ of a coalition by an upper bound:

$$PL_c = \sqrt{\sum_{i=1}^{N}\{PL_i\}} \ [kByte]$$

Hereby, $PL_i[kByte]$ is the uncompressed xFCD payload of one vehicle collected over time and $N$ is the cluster size. We assume that 1 kByte xFCD is generated per second by each vehicle [4]. The data is transmitted in intervals of 10 s or if the CH changes. If no LTE or V2V connectivity is available, the data is stored in the vehicles for later upload.

### A. Proposed Clustering Algorithms

In order to ensure stable clusters, an efficient data aggregation and compression, the clusters should be as large as possible and the lifetime of a cluster should also be large. The latter serves to avoid a V2V overload caused by rapid cluster reorganisation. Furthermore, these V2V constraints restrict the range of values of clustering parameters as the desired minimal LTE traffic and utilization do not necessarily lead to a minimal cluster lifetime. Therefore, lifetime considerations take priority over purely LTE utilization based parameters and methods. The optimization problem, in terms of finding the vehicle with the most vehicles in communication range, can be seen as maximal covering location problem [13]. A solution of this problem is the Greedy algorithm [14]. This algorithm leads to clusters of a huge size, but due to the high dynamic of the

system, these clusters would exist very shortly. Hence, we use this algorithm as reference for a high cluster size.

The clustering algorithms are based on a two steps procedure. First, CHs from all possible vehicles have to be selected. In the second step, the CMs are assigned to connected CHs. We propose the following iterative cluster procedure which is applied each second:

1) Set up a list of all active vehicles.
2) Check V2V connectivity. For each vehicle the number of reachable vehicles is calculated.
3) Vehicles which cannot communicate with any other vehicle are CHs.
4) Calculate vehicles' driving angle $\alpha$ (quantified in steps of $10°$) and velocity $v$ (quantified in steps of 2 m/s).
5) Calculate cluster score $CS$ of each car: $CS = num_{V2V} + num_{CM}$. Thereby, $num_{V2V}$ represents the number of reachable vehicles with an angle difference smaller than $\alpha_{diff}$ and a velocity difference smaller than $v_{diff}$. $num_{CM}$, named CH bonus, is the number of CMs for vehicles which were CH a time step before. $num_{CM} = 0$ is used for vehicles that were CM in the last time step.
6) The vehicle with the highest $CS$ value becomes CH.
7) Assign CMs to a CH by selecting vehicles which are reachable and meet the specified velocity and angle requirements.
8) Erase grouped CH and CMs from vehicle list.
9) Repeat steps 2 - 8 until every vehicle is CH or CM.

In order to evaluate the impact of the clustering algorithms, four different algorithms are compared which each other:

- **No Clustering (NC)**. Each vehicle transmits xFCD via LTE.
- **Greedy-based Clustering (GC)**. Vehicles with the highest number of reachable vehicles via V2V communication are CHs. No velocity and angle requirements are met and the cluster bonus is deactivated ($num_{CM} = 0$). For this reference algorithm the cluster structure would change very quickly (e.g. due to different traveling directions on a highway) and it is very challenging to organize these clusters by V2V communication. This is due to the high packet error rate for high velocity differences [15].
- **Greedy-based Clustering with Velocity and Direction restriction (GC-VD)**. Vehicles with the highest number of reachable vehicles via V2V communication which fulfill the velocity and angle requirements (steps 4+5) are CHs. The cluster bonus is deactivated.
- **Greedy-based Clustering with Velocity and Direction restriction and cluster Bonus (GC-VDB) (cf. Fig. 2)**. Vehicles with the highest CH score which fulfill the velocity and angle requirements (steps 4+5) are CHs.

## IV. SIMULATION ARCHITECTURE

In order to analyze the cellular communication under cluster-based xFCD collection, a novel system model is presented in this paper (cf. Fig. 3). This model consists of the following components:
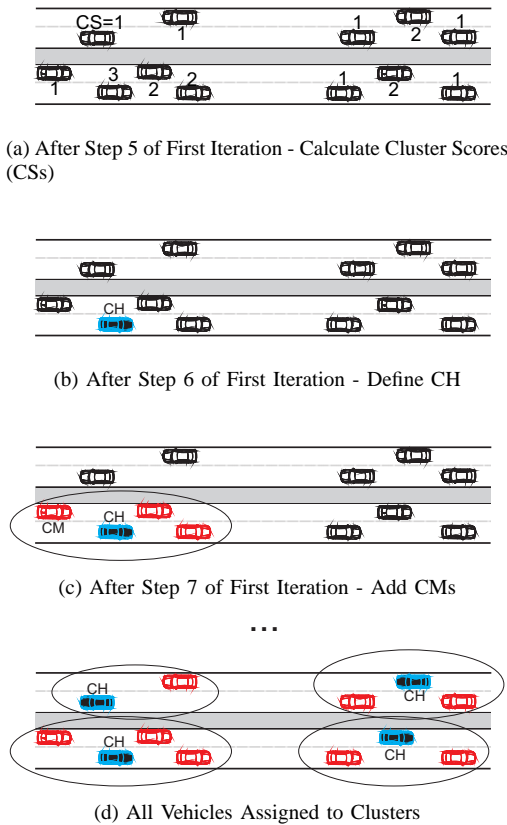
(a) After Step 5 of First Iteration - Calculate Cluster Scores (CSs)



(b) After Step 6 of First Iteration - Define CH



(c) After Step 7 of First Iteration - Add CMs

...



(d) All Vehicles Assigned to Clusters

Fig. 2: Visualization of GC-VDB Algorithm.

- **Building Environment and Road Network from Open-StreetMap (OSM)**. To guarantee realistic user mobility and to model realistic communication characteristics, the building environment and streets for a predefined scenario are used.
- **Mobility from Simulation of Urban MObility (SUMO) [16]**. The street map from OSM is imported to SUMO. In this component the trajectory for each vehicle is calculated.
- **Clustering Algorithm in MATLAB**. The actual clustering algorithm is calculated in MATLAB. For this

purpose, the buildings map from OSM and the user mobility from SUMO have to be imported. The V2V communication is IEEE 802.11p based. We approximate the V2V connectivity with two different cases. For a Line-Of-Sight (LOS) connection a communication range of 300 m [5][11] is assumed and for a Non-Line-Of-Sight (NLOS) connection the range is reduced to 76 m [17]. As result the clusters are formed and communication traffic of the CHs is calculated. This includes transmission times and payload sizes.

- **Impact on LTE System in OPNET**. In order to model the occurring LTE traffic in a realistic way, the protocol simulation must be aware of two components. The user mobility from SUMO is imported as user trajectories. Furthermore, the transmission times and the payload size for xFCD transmission are imported from MATLAB.

The modular architecture of this system model makes flexible usage of different scenarios possible. In this paper, two scenarios are used (cf. Tab. I for parameters). In Fig. 4 the highway scenario with the LTE communication infrastructure and in Table II the OPNET simulation parameters are illustrated.

TABLE I: Parameters of Scenarios.

| Parameter | Highway | Urban Streets |
|---|---|---|
| Number of vehicles | 400 | 608 |
| Average number of cars on streets | 124 | 256 |
| Simulation duration | 707 s | 1386 s |
| Number of eNBs | 5 | 14 |

TABLE II: LTE Simulation Parametrization.

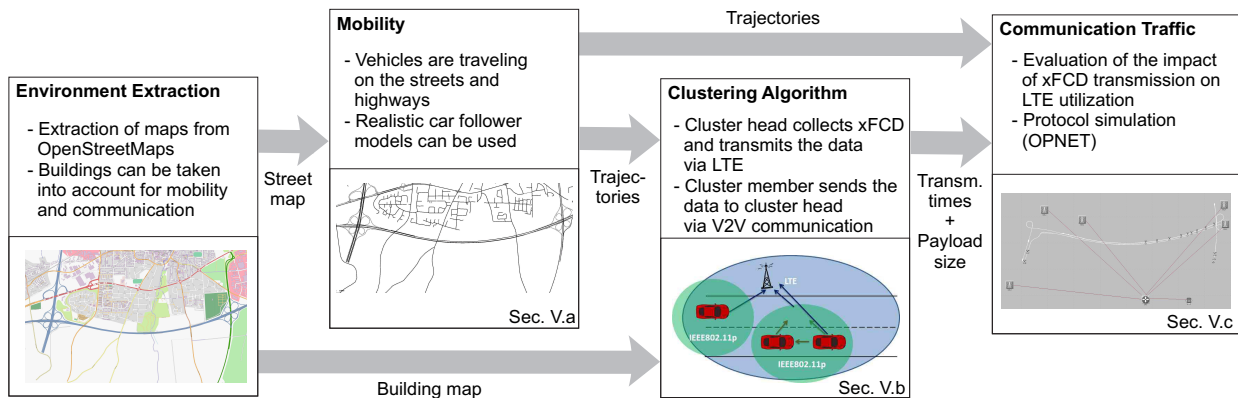| Parameter | Value |
|---|---|
| OPNET version | 17.5; 64 Bit |
| Carrier band | LTE band 1 |
| Channel bandwidth | 20 MHz |
| Duplex scheme | Frequency Division Duplex (FDD) |
| Pathloss channel model | ITU-R M.1225 (Vehicular) [18] |
| Multipath channel model | ITU Vehicular A [18] |



Fig. 3: Simulation Architecture for Performance Evaluation of Clustering Algorithms.
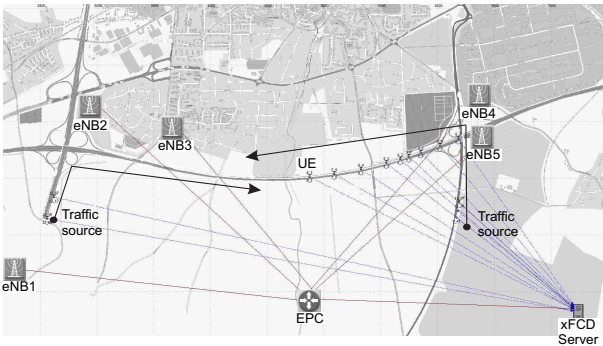
Fig. 4: Highway Scenario with LTE Communication Infrastructure and IP Traffic Flows.

## V. RESULTS

In this section, first the user mobility and then different cluster characteristics under cluster-based xFCD collection, are analyzed. At the end, the impact of clustering on the LTE communication system is shown.

### A. User Mobility

In Fig. 5 the Probability Density Function (PDF) of the user velocity is illustrated for the highway scenario. It can be seen that many vehicles are traveling with a velocity between 100 km/h and 130 km/h. The scenario also includes freeway on-ramps and exit-ramps, which are responsible for the peak at 25 km/h. If only vehicles with a similar velocity are grouped into clusters, vehicles which are entering or leaving the highway are not connected to vehicles with a high speed. Hence, a single car which enters the highway accelerates and joins a cluster when it adapted its velocity to a common travel speed.

### B. Cluster Characteristics for Different Clustering Strategies

One main goal of the GC-VDB algorithm is to increase the cluster lifetime (time that a cluster stays unmodified - no car may enter or leave the cluster) in order to keep the V2V overhead as low as possible. We evaluated by a sensitivity analysis that a maximum cluster lifetime of 1.9 s can be achieved for a velocity difference of $v_{diff} = 22m/s \approx 80km/h$ and an angle difference of $\alpha_{diff} = 70°$. These values arise from the
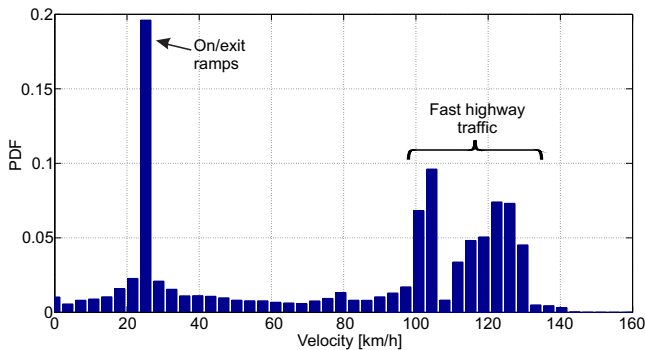


Fig. 5: PDF of User Velocity of all Vehicles for Highway Scenario (both directions).

two scenarios presented in this paper and do not necessarily represent universally optimal parameters. In addition, a very low total xFCD payload results for this parameterization. This parameterization of GC-VDB is used for the following results.

In Fig. 6 cluster characteristics are shown for the different clustering algorithms including no clustering. It can be seen that by applying GC many vehicles are grouped together. This leads to a high cluster size (cf. ① in Fig. 6a) and a low number of CHs (cf. ② in Fig. 6b) for both scenarios. This algorithm can be seen as reference related to the cluster size, because the vehicles that can reach the highest number of CMs are CHs. However, this algorithm leads to a problem. It does not consider the driving direction of the vehicles. Therefore, vehicles traveling in opposite directions would be grouped together. This would lead to a huge challenge (especially on the highway) for the V2V communication in order to organize those rapidly changing, short lived clusters (cf. ③ in Fig. 6c).

To overcome this problem, GC-VD is introduced. Here, only vehicles with similar velocity and driving direction are grouped into clusters. This makes the V2V communication much more manageable, because the speed and the angle difference between CH and CMs is limited. However, for this algorithm the CH is also changing very often so that the CH lifetime (time a car stays CH; the CMs can change) barely increases compared to GC (cf. ④ in Fig. 6c). Hence, we introduce the CH bonus. By applying GC-VDB the duration a car stays CH (cf. ⑤ in Fig. 6d) and the cluster lifetime increases (cf. ⑥ in Fig. 6c). The CH lifetime for the urban scenario is much higher than for the highway scenario because the vehicles stay longer in communication range due to a lower velocity. For NC the number of CHs equals the mean number of vehicles in the scenario, because each vehicle is CH. Hence, the cluster size is always 1 (cf. ⑦ in Fig. 6a).

As a result, all presented clustering algorithms lead to a smaller average (cf. Fig. 6e) and total xFCD payload (cf. Fig. 6f) due to data aggregation. By GC-VDB, the total xFCD payload can be reduced by 97 % in contrast to no clustering for the urban scenario (cf. ⑧ in Fig. 6f). The cost for this reduction is the V2V communication overhead. However, this overhead is lower than for GC, because the proposed GC-VDB algorithm leads to high CH lifetimes and only vehicles with a low velocity and angle difference are grouped as a cluster.

### C. Impact of Clustering Algorithms on the LTE Communication System

The mean LTE Physical Uplink Shared Channel (PUSCH) utilization over all eNBs and the utilization of the LTE cell with the highest load are illustrated as box plots in Fig. 7. It can be seen from the figure that NC leads to a high utilization and high peaks (especially for the urban scenario), because it can occur that many vehicles transmit xFCD at a similar point in time. In addition, the median of the average PUSCH utilization can be reduced from 1.7 % to 0.3 % for the highway scenario. This is due to the lower total xFCD payload (cf. Fig. 6f).
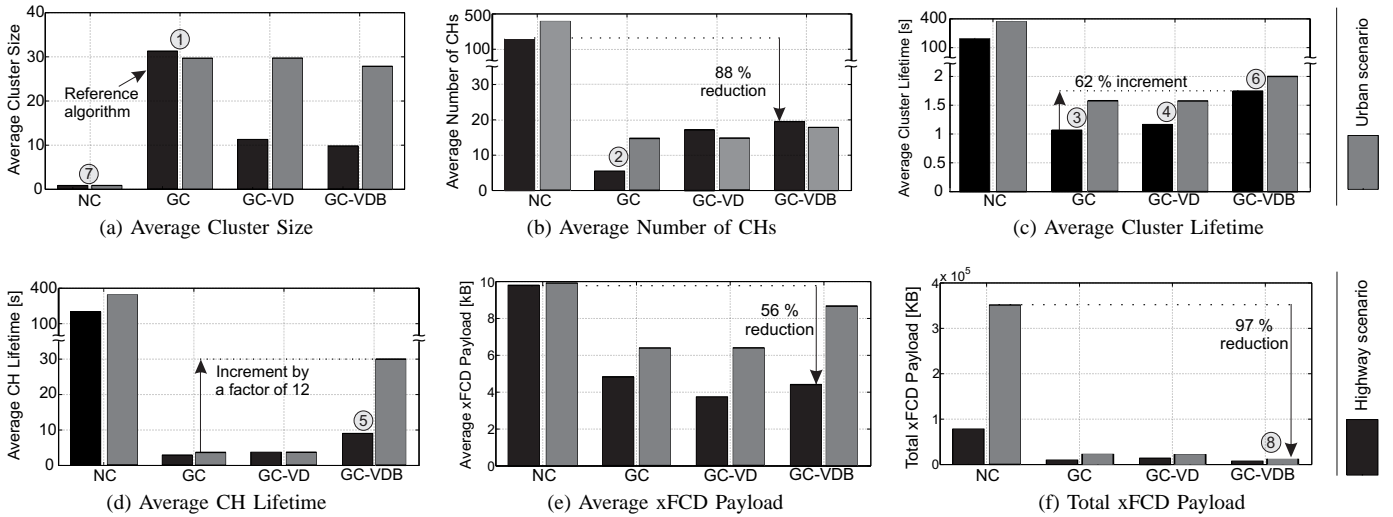
Fig. 6: Cluster Characteristics for Different Clustering Strategies.
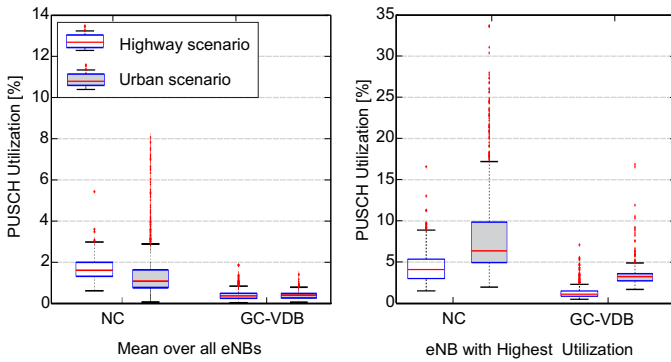


Fig. 7: LTE PUSCH Utilization for No Clustering (NC) and under Cluster-Based xFCD Collection (GC-VDB Algorithm).

## VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed clustering algorithms for the efficient xFCD collection. The performance of the algorithms has been evaluated by a novel system model, which allows for modeling of LTE cells under realistic user mobility. The results have shown that the GC-VDB algorithm enables an enhancement of the cluster lifetime and a decrement of the total xFCD payload. Hence, the LTE utilization has decreased significantly. In the future, the costs of cluster-based xFCD collection, in terms of V2V communication overhead with special focus on small connection durations, will be analyzed.

### REFERENCES

[1] S.-Y. Lien, K.-C. Chen, *Massive Access Management for QoS Guarantees in 3GPP Machine-to-Machine Communications*, IEEE Communications Letters, vol. 15, no. 3, Mar. 2011.

[2] B. S. Kerner, et al., *Traffic State Detection with Floating Car Data in Road Networks*, 8th International IEEE Conference on Intelligent Transportation Systems, Vienna, Austria, Sep. 2005.

[3] C. Ide, et al., *Efficient Floating Car Data Transmission via LTE for Travel Time Estimation of Vehicles*, IEEE 76th Vehicular Technology Conference (VTC-Fall), Quebec City, Canada, Sep. 2012.

[4] C. Ide, et al., *Influence of M2M Communication on the Physical Resource Utilization of LTE*, 11th Wireless Telecommunications Symposium, London, UK, Apr. 2012.

[5] G. Remy, et al., *Lte4v2x: LTE for a Centralized VANET Organization*, IEEE Globecom 2011, Houston, Texas, USA, Dec. 2011.

[6] 3GPP TR 22.888 V12.0.0, *Study of Enhancements for Machine-Type Communications*, Mar. 2013.

[7] 3GPP TS 22.368 V12.2.0, *Service Requirements for Machine-Type Communications*, Mar. 2013.

[8] K. Zheng, et al., *Radio Resource Allocation in LTE-Advanced Cellular Networks with M2M Communications*, IEEE Communications Magazine, pp. 184 - 192, Jul. 2012.

[9] C. Ide, et al., *Channel Sensitive Transmission Scheme for V2I-based Floating Car Data Collection via LTE*, IEEE International Conference on Communications (ICC), Ottawa, Canada, Jun. 2012.

[10] G. Zhioua, et al, *A Multi-Metric QoS-Balancing Scheme for Gateway Selection in a Clustered Hybrid VANET Network*, IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Barcelona, Spain, Oct. 2012.

[11] A. Benslimane, T. Taleb, and R. Sivaraj, *Dynamic Clustering-Based Adaptive Mobile Gateway Management in Integrated VANET - 3G Heterogeneous Wireless Networks*, IEEE Journal on Selected Areas in Communications, vol. 29, no. 3, Mar. 2011.

[12] A. D. Ker, *Steganographic Strategies for a Square Distortion Function*, In Proc. of Security, Forensics, Steganography and Watermarking of Multimedia Contents X, vol. 6819, 2008.

[13] R. Church and C. ReVelle, *The Maximal Covering Location Problem*, Papers in Regional Science, vol. 32, no. 1, pp. 101-118, Jan. 1974.

[14] T. H. Cormen, et al., *Introduction to Algorithms*, 4th ed. The MIT Press, 2001, vol. 1.

[15] R. A. Saeed, et al., *Evaluation of the IEEE 802.11p-based TDMA MAC Method for Road Side-to-Vehicle Communications*, International Journal of Network and Mobile Technologies, vol. 1, no. 2, Nov. 2010.

[16] D. Krajzewicz, M. Bonert, and P. Wagner, *The Open Source Traffic Simulation Package SUMO*, RoboCup 2006 Infrastructure Simulation Competition, Bremen, Germany, Jun. 2006.

[17] H. Schumacher, et al., *Vehicle-to-Vehicle IEEE 802.11p Performance Measurements at Urban Intersections*, IEEE International Conference on Communications (ICC), Ottawa, Canada, Jun. 2012.

[18] International Telecommunication Union, Recommendation ITU-R M.1225 *Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000*, 1997.