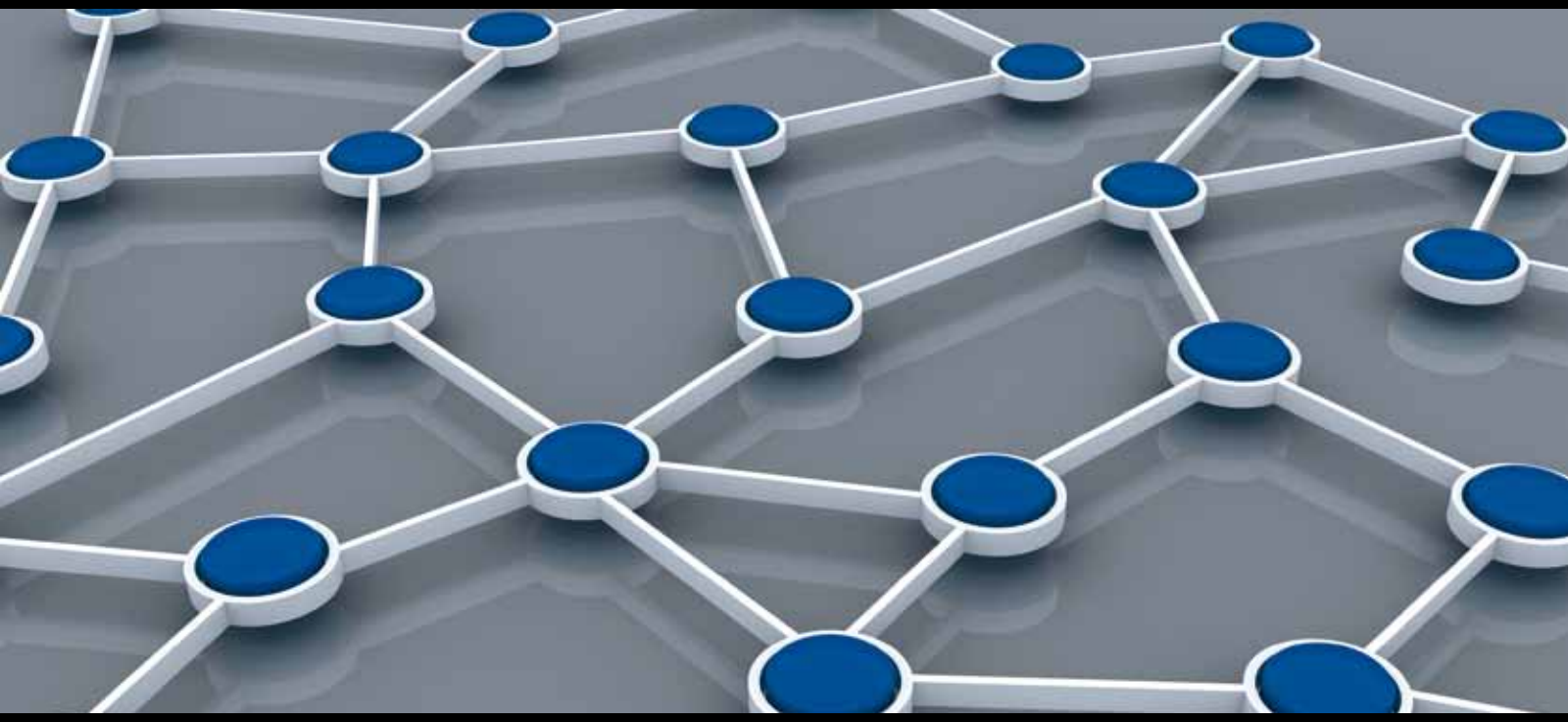
International Journal of Distributed Sensor Networks

Ubiquitous Sensing: Services and Applications

Guest Editors: Ming Liu, Yonggang Wen, Guihai Chen, and Chao Song



Ubiquitous Sensing: Services and Applications

International Journal of Distributed Sensor Networks

Ubiquitous Sensing: Services and Applications

Guest Editors: Ming Liu, Yonggang Wen, Guihai Chen, and Chao Song

Editorial Board

Contents

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 403857, 2 pages
http://dx.doi.org/10.1155/2013/403857

Editorial

Ubiquitous Sensing: Services and Applications

Ming Liu,[1] Yonggang Wen,[2] Guihai Chen,[3] and Chao Song[1]

[1] *University of Electronic Science and Technology of China, Chengdu 611731, China*
[2] *Nanyang Technological University, Singapore 639798*
[3] *Shanghai Jiao Tong Univeristy, Shanghai 200240, China*

Correspondence should be addressed to Ming Liu; wing.lm@gmail.com

Received 20 October 2013; Accepted 20 October 2013

Ubiquitous sensing has penetrated into all aspects of our everyday lives and has gained tremendous research attention in recent years. This issue on ubiquitous sensing compiles 18 exciting manuscripts which are approved by reviewers.

There are four manuscripts studying routing algorithm in sensor networks, two manuscripts proposing data distribution methods, two manuscripts describing carpooling methods, three manuscripts studying energy saving in WSN, two manuscripts discussing localization in WSN, two manuscripts studying human action recognition, and three miscellaneous manuscripts.

C. Ma and N. Liu propose a routing algorithm in vehicular sensor networks based on some statistical data, like traffic density and vehicle distribution which is not considered in existing works. H. Gong and X. Wang propose a routing protocol in Mobile Social Networks based on selfishness of nodes which is more related to reality. X. Zhang et al. propose a routing algorithm in Vehicle Sensor Networks which solves a tracking problem. The tracking problem is that the destination vehicles will move continuously, so that when the data packets arrive at the original destination point, they are not able to be delivered. The algorithm uses area epidemic and parking vehicles assisted methods to track the destination vehicles to ensure the data delivery. Y. Feng et al. firstly analyze the predictabilities of different types of vehicles and then propose a new driving path predication based routing protocol (DPPR).

H. Zhao and J. Zhu propose a novel data distributing method in VANETs using roadside parking vehicles, and the parking vehicles are formed to several clusters which provide data caching and distributing services. H. Gong et al. present a Latency Estimation based data Delivery (LED) scheme. According to the current location and the moving direction of the vehicles, LED calculates the expected delivery latency (EDL) of the individual vehicles and chooses the vehicle with the shortest EDL as the next hop to forward data.

J. Zhu et al. propose a parking-lot-assisted carpool over VANETs. It collects vehicle trajectories via accelerator sensor to sense vehicle's movement, establish a routing tree to deliver vehicle trajectory information to nearby parking lots, and design a suitable matching scheme to match the target vehicle in VANETs. N. Liu et al. present the idea of mobility crowdsourcing (MobiCrowd), which leverages private smartphone to collect individual trips for carpooling, without any explicit effort on the part of users. It acquires location information from smartphone, generates daily trips and mobility models for each user, and then makes carpooling zero effort by enabling travel data to be crowdsourced instead of tracking vehicles or asking users to input their trips. With prior mobility knowledge, one user's travel routes and positions can be predicted according to the current location, and then possible carpooling can be arranged to fit the mobility context.

J. Peng et al. present research on the existing clustering algorithm applied in heterogeneous sensor networks and then put forward an energy-efficient prediction clustering algorithm, which is adaptive to sensor networks with energy and objects being heterogeneous. This algorithm enables the nodes to select the cluster head according to factors such as energy and communication cost; thus the nodes with higher residual energy have higher probability to become a cluster head than those with lower residual energy, so that the network energy can be dissipated uniformly. O. Krejcar and R. Frischer present a modern and relatively cheap solution for powering some types of intelligent sensors in which

traditional battery power source is insufficient and other powering options are not applicable. Obtaining the energy to supply sensors is possible even from immediate sensors' environment. Sources (which are utilizing thermal gradient) are supplying energy from the surrounding environment and without the need for high intensity incident light (solar energy based). Y.-K. Kim et al. propose a novel Hilbert-curve based data aggregation scheme that enforces data privacy and data integrity for WSNs.

H. Wu et al. propose a method to make WSN localization suitable for the maritime search and rescue (MSR); an improved microelectromechanical systems (MEMS) aided algorithm on the basis of triangle and centroid algorithm is proposed to locate and track the search targets in real time and more precisely. Y. Lyu et al. propose a scalable location-sensing model based on RFID-sensing architecture for ephemeral social network (ESN) in consideration of four aspects of requirements, that is, the usability, QoS, scalability, and privacy. The model includes the perspectives of the privacy, architecture, deployment, and positioning algorithms, which can meet the four key requirements.

M. Kurz et al. study opportunistic activity recognition by enabling dynamic sensor configuration. In contrast to "traditional" applications where sensors, their modalities, locations, and working characteristics have to be defined at design time opportunistic systems do not rely on an initially defined and fixed sensing infrastructure. Sensors have to be utilized upon their spontaneous availability and activity recognition capabilities and dynamic sensor ensembles have to be configured at run-time with respect to maximized recognition accuracy and minimized energy consumption. M. H. Wang and P.-Y. Chen used palmprint with the extension method to design a low-cost personal recognition system.

X. Zhang et al. propose a novel traffic jam autoaware system which uses only sensors from mobile phones, and the mobile phones exchange information among each other in a crowdsourcing way and give traffic jam information to help drivers release anxiety for trapping in the traffic jam. B.-W. Jo et al. investigate the capability of wireless communication of sensor node embedded in reinforced concrete structure with a basic experiment on an electric wave permeability of sensor node by fabricating molding with variables of concrete thickness and steel bars that are mostly used in constructing structures to determine the feasibility of application to construct structures with ubiquitous sensor network (USN). M. Liu gives a survey of existing research works and applications on mobile phones.

Ming Liu
Yonggang Wen
Guihai Chen
Chao Song

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 652385, 13 pages
http://dx.doi.org/10.1155/2013/652385

Research Article

Enabling Dynamic Sensor Configuration and Cooperation in Opportunistic Activity Recognition Systems

Marc Kurz, Gerold Hölzl, and Alois Ferscha

Institute for Pervasive Computing, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria

Correspondence should be addressed to Marc Kurz; kurz@pervasive.jku.at

Received 5 December 2012; Revised 18 April 2013; Accepted 21 April 2013

Academic Editor: Ming Liu

Opportunistic activity recognition as research discipline is characterized by the fact that human activities (and more generally the context) shall be recognized with sensors that are initially unknown to the system. In contrast to "traditional" applications—where sensors, their modalities, locations, and working characteristics have to be defined at design time—opportunistic systems do not rely on an initially defined and fixed sensing infrastructure. Sensors have to be utilized upon their spontaneous availability and activity recognition capabilities and dynamic sensor ensembles have to be configured at runtime with respect to maximized recognition accuracy and minimized energy consumption. This requirement contains two research challenges that this paper tackles: (i) estimating the accuracy of an ensemble without being able to compare the output in the form of recognized activity classes to a (labeled) ground truth and (ii) optimizing the accuracy/energy trade-off by applying exact and heuristic methods adapted for cooperative sensor ensembles.

1. Introduction

Opportunistic activity and context recognition is a rather new research field [1, 2]. It is motivated by the fact that sensor systems of different kinds are nowadays integrated into things of everyday life, which become more and more present (e.g., smart phones, wearable computers, gadgets, etc.). This makes the explicit deployment of sensors for specific activity recognition applications gratuitous. Opportunistic activity and context recognition aims at utilizing sensors upon their spontaneous availability and configuring sensor ensembles that are best suited to execute a certain recognition goal. This recognition goal can be defined at runtime by an application or user [3] and defines what the systems shall recognize but does not specify how [2]. Sensor ensembles combining available sensor devices to recognize certain activities of interest (as formulated in the goal definition) have to be dynamically configured upon the spontaneous availability of sensor devices. An easy approach would be to include all sensors that are capable of recognizing a specific activity (that is requested within the goal formulation). Nevertheless, this "brute-force" approach does not necessarily mean that the accuracy can be optimized/maximized; the only effect that can be guaranteed is the fact that the energy consumption of the sensor ensemble is increased. An ensemble consisting of heterogeneous sensor devices shall meet two important requirements: (i) maximize the accuracy, whilst (ii) minimizing the overall ensemble energy consumption (e.g., to ensure a long lifetime of the sensor devices if operated on battery mode).

These two requirements could be conflicting; therefore, a trade-off function has to be found to define sensor ensembles utilizing currently available sensor devices. An important aspect that has to be considered is the real-time practicability; the sensing infrastructure at any point in time cannot be assured to be static and fixed; instead it is of highly dynamic nature [1]. Another crucial and nontrivial aspect is the fact that the accuracy of an ensemble cannot be easily estimated. Usually, the accuracy for a (sensor) signal classification system is based on a labeled ground truth, where the predicted activity classes can be compared to the actual classes. Since an opportunistic system operates in real time and cannot initially fix the set of input sensor devices, a ground truth cannot be taken for granted. This paper tackles two

research challenges that emerge from opportunistic activity and context recognition.

(I) *Unsupervised Accuracy Estimation*. The expected accuracy of a (initially unknown) sensor ensemble has to be estimated. Possible input parameters could be the accuracy of single sensor devices, diversity measures between sensor combinations, and the technical properties (e.g., sampling rate, sensor signal modality, etc.).

(II) *Ensemble Optimization*. Based on the expected accuracy and the expected energy consumption of the ensemble (calculated as sum of the sensors' individual energy requirements), an optimal trade-off (maximizing the accuracy whilst minimizing the energy demand) has to be found in order to configure ensembles at system runtime.

The remainder of the paper is structured as follows. Section 2 discusses related work in the broader fields of "traditional" activity recognition and utilizing sensors in a dynamic way. Recent achievements in the field of opportunistic sensing—which build the baseline for the research work within this paper—are also discussed in Section 2. The methodologies (i.e., how do we tackle the challenges of estimating the ensemble accuracy and optimizing the ensemble accordingly) are described in detail in Section 3. These methodologies are evaluated in Section 4 on a rich dataset that has been recorded including more than 70 sensors with 10 different modalities in a kitchen scenario (see [4, 5]). The paper closes with a conclusion and an outlook to future work in Section 5.

2. Related Work and Recent Achievements

Related work that matches the scope of this paper can be identified in two areas of research: (i) activity and context recognition and (ii) opportunistic sensing and opportunistic activity recognition. The following two Sections 2.1 and 2.2 contain overviews of related work (and recent achievements) within these two areas. Section 2.3 summarizes the findings and provides a discussion of the novelty of the targeted research challenges.

2.1. Activity Recognition. This section provides a discussion on related work across different types of activities and different scales of activities' complexity. The recognition of user's activities and more generally the context [6] of persons by utilizing different body-worn and environmental sensor devices has been well evaluated in previous work (e.g., [7–9]). There, the authors attached a set of acceleration, inertial, and microphone sensors on the body of subjects, utilized different signal processing and machine learning algorithms for classification, to investigate in the quality of the detected activities in terms of recognition rate, accuracy, and other performance metrics [10]. Again, as we are interested in the recognition of activities of persons, the choice of sensors to achieve an accurate estimation of the activities together with the requirement to be as unobtrusive for the user as possible is a crucial task. In [11], Tapia et al. focused on recognizing activities with a set of simple and ubiquitous sensors ("tape on and forget" sensors). Logan et al. [12] investigated in the question which sensors and modalities in general are well suited for highly accurate activity recognition, whereas the findings are that infrared motion detection sensors outperformed others (e.g., RFID, built-in wired sensors). A third contribution that investigates accurate recognition of activities in user's homes is presented in [13]. Again, the presented system relies on a set of specific wireless sensor nodes that have to be explicitly deployed in the home environment. Summarizing, the best sensor modality for on body activity recognition are accelerometers [14] due to their high accuracy and their unobtrusiveness. They are already integrated into things of everyday life like in smart phones or smart watches and can, therefore, be easily utilized.

2.2. Opportunistic Sensing. Opportunistic activity and context recognition as new research discipline draws from the fact that a plethora of sensor devices is already deployed in the world. Therefore, the application-specific sensor deployment is an effort that can be avoided. Important methods that have already been developed tackle the fact that an unknown and highly dynamic sensing infrastructure consists of heterogeneous devices, capable of delivering different functionality to an activity recognition system. Therefore, the concepts of *sensor abstractions* and *sensor self-descriptions* (see [1, 15, 16]) allow for the dynamic generation upon sensor availability of activity recognition chains (ARCs) dependent on the sensors capabilities as preserved in the self-descriptions. The OPPORTUNITY Framework [1] is a ready-to-use system combining different methods for enabling highly flexible, dynamic, and stable activity recognition. *ExperienceItems* as building blocks within the sensor self-descriptions memorize gathered experience in form of recognizable activities. The degree of fulfillment (DoF) value [1] is a metric (between 0 and 1) that is used to estimate the expected accuracy of a sensor/ensemble for a specific recognition goal. The DoF together with required machine learning and signal processing technologies (and accompanying parameters) is persistently stored in the ExperienceItems. These items can either be generated offline—in a traditional way by using a labeled ground truth—or (as shown in [17]) at systems runtime by applying the concept of *transfer learning*. There, a sensor (or sensor ensemble) is trained by another sensor/ensemble and the gathered experience is written into the corresponding self-description as new ExperienceItem. Still open in the research landscape of opportunistic activity recognition are two highly challenging issues: (i) estimating the accuracy for an initially unknown ensemble (by fusing single sensors using the MajorityVoting method [18]) and (ii) optimizing and shaping the ensemble according to the estimated accuracy and the expected costs (in form of energy consumption of the sensor devices); these items compose the core contribution of this paper.

2.3. Related Work Summary. Opportunistic sensing (and more specifically *opportunistic activity recognition*) is a rather

Table 1: Overview of possible ways of calculating the degree of fulfillment (DoF) for single sensors and ensembles.

	Single sensor	Sensor ensemble
(i) Offline/initial Ground truth available	Easy; calculation of accuracy by validation of initially trained classifier	Easy; calculation of accuracy by validation of (initially trained) fusion method
(ii) Online/runtime Ground truth unavailable	Difficult because no ground truth for validation is available	Difficult because no ground truth for validation is available

new research area. Particularly, the fact that no labeled and reliable ground truth is available to compute solid metrics to quantify an applications quality is a very challenging characteristic. The work presented in this paper tries to overcome this issue of not being able to reliably predict an ensembles accuracy and use this information subsequently to generate an ensemble that is very likely to be close to optimal in terms of expected accuracy and costs. To the best of our knowledge, these two challenges are novel in the area of sensor-based activity recognition and are identified as being important building blocks towards the vision of opportunistic activity recognition [1].

3. Methodology

3.1. Estimating the Accuracy of an Ensemble at Runtime. The definition of the quality criterion of having the "best" sensors configured in an ensemble (and not simply using all possible sensor candidates from the set S_C, which is a subset of the set S containing all available sensors) is a nontrivial task. On the one hand, the accuracy of an ensemble for a certain set of activities that have to be recognized has to be maximized, and on the other hand, the energy consumption of the sensor(s) (ensemble) should be as low as possible. Particularly, the first criterion—maximizing the ensemble's accuracy—is very difficult to predict. Having multiple ARCs from different sensors and the DoF of each, the accuracy of a fusion method that combines these sensors (e.g., MajorityVoting fusion) cannot be predicted easily (see [18–21]). Nevertheless, an estimation of the overall ensemble DoF is needed to be able to maximize the ensemble's performance in terms of accuracy and recognition rate for certain activities. Therefore, we present an approach how the diversity of multiple sensors can be used to approximate the accuracy of sensor combinations. The second criterion—minimizing the energy consumption—is important to ensure a long lifetime of a sensor network and the sensing, respectively infrastructure, which is seen as an equally important criterion. Therefore, we define the term "best ensemble" as being the sensor combination out of the set S_C that maximizes the estimated ensemble accuracy, whilst minimizing the energy consumption of the sensors in set S_E (the set of sensors that are included in the ensemble). Further criteria could also be included to provide a more complex definition of "the best ensemble" (e.g., a user-convenience factor or an initial-cost factor), but this is not in the scope of this paper.

Predicting the accuracy (or the degree of fulfillment (DoF)) of an ensemble, which consists of multiple sensors that are combined on the machine learning level by utilizing a decision fusion methodology, can be done in two different ways, respectively, processing stages of a system (see also Table 1):

(i) *offline* and *manual*: if a dataset is available that can deliver ground truth activity labels to compare the predicted activity classes from the fusion method with the actual classes and compute the overall accuracy of the ensemble. Obviously, this way of calculating the accuracy of an ensemble does not fit the opportunistic approach, since usually no ground truth and no predefined sensing infrastructure are available in a real-time, open-ended system;

(ii) *online* and *autonomous*: at runtime of the system. This fits the opportunistic approach, since neither a ground truth nor the available sensor infrastructure is available or known. Thus, the DoF of an ensemble has to be calculated (or rather estimated) at runtime of the system. This challenging task—constituting one major contribution of this paper—is discussed in the following.

As already described and analyzed in related work literature [19, 21], a correlation between the individual accuracies of the sensors together with their ARCs involved in an ensemble and the diversity between the particular sensors can be observed. This means that the performance of an ensemble is dependent on the accuracies (i.e., the degree of fulfillment (DoF)) and the diversities between the involved sensors. Meynet and Thiran propose in [21] an information-theoretic score (ITS) that expresses exactly this relation. The idea behind this approach is the fact that if two or more sensors have very low mutual information (MI) values (thus have a low diversity), they also agree on wrongly classified activity classes. Therefore, the converse argument comprises that if sensors in an ensemble that have high diversity values can avoid wrongly classified results. Chavarriaga et al. follow up this approach in [19]. They use the accuracy of individual classifiers and the available pairwise MI values to calculate an ITS value of an individual sensor. Then, the authors show in simulations that the information-theoretic approach for ensemble creation and recreation outperforms a random method. They subsequently add the classifier to an ensemble that has the highest ITS value. Therefore, the information-theoretic score (ITS) is proven to be a candidate to estimate the accuracy and DoF of an ensemble. Generally, the higher the ITS value of an ensemble is, the higher is the accuracy. This finding is essential and will build the fundamental for further optimization methods within this paper. The MI value between two discrete variables X and Y according to Shannon and Weaver [22] and Kraskov et al. [23] is defined as (see also Meynet and Thiran [21]):

$$\mathrm{MI}(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y)\log\left(\frac{p(x,y)}{p(x)\,p(y)}\right). \quad (1)$$

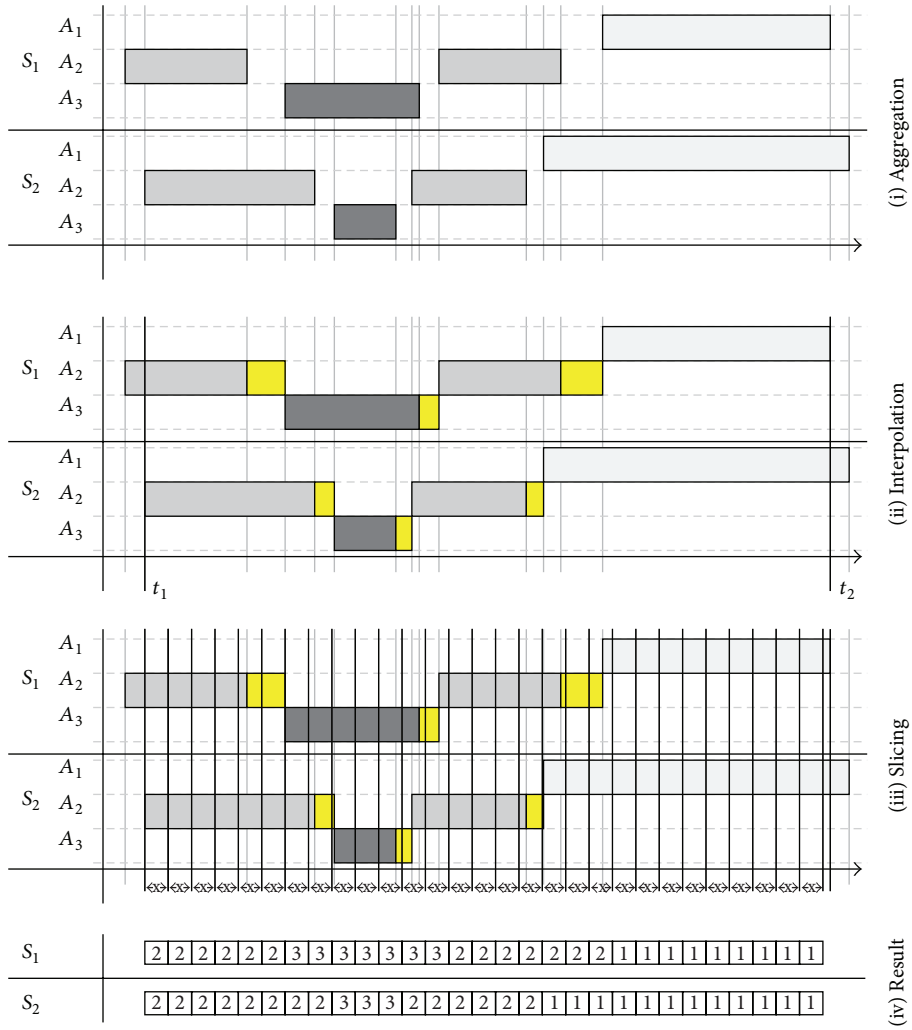


FIGURE 1: Illustration of the processing of the activity label sequences of two sensors (S_1 and S_2) for calculating the pairwise mutual information (MI) value.

According to Meynet and Thiran, the information-theoretic score (ITS) expressing an estimate of the accuracy can be calculated as [21]

$$\mathrm{ITS} = (1 + \mathrm{ITA})\,(1 + \mathrm{ITA})^2\,(1 + \mathrm{ITD})\,. \tag{2}$$

The information-theoretic accuracy (ITA) value is the average accuracy of K sensors. There, the mutual information (MI) value of a sensor compared to the ground truth (i.e., $\mathrm{MI}(S; S_i)$) is required, which can be approximated by the system in the same way as the pairwise diversity is calculated (see Figure 1), where an active ensemble with a high DoF is taken as reference for comparisons. The ITA value is calculated as [21]

$$\mathrm{ITA} = \frac{\sum_{i=1}^{K} \mathrm{MI}\left(S; S_i\right)}{K}. \tag{3}$$

The second required parameter for computing the ITS value is the information-theoretic diversity (ITD), which describes the average diversity between sensors [21]:

$$\mathrm{ITD} = \frac{\binom{K}{2}}{\sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \mathrm{MI}\left(S_i; S_j\right)}. \tag{4}$$

In [19], Chavarriaga et al. had the accuracies for the individual classifiers and the pairwise mutual information values ready within their simulations. This is not the case in an opportunistic system and has to be achieved depending on the current sensing infrastructure and the recognition goal (thus, these MI values have to be calculated in the running system at real time), which is not an easy task and represents a major contribution of this paper. The mutual information (MI) value that is needed for calculation of the ITS value can be calculated as reciprocal of the diversity between two classifiers (sensors). Figure 1 illustrates how the mutual information (MI) value between two sensors can be calculated at runtime. This process is composed of four

phases, that is, (i) aggregation, (ii) interpolation, (iii) slicing, and (iv) result.

(i) *Aggregation*: two sensors (S_1 and S_2 in Figure 1) deliver activity labels for three activities (A_1, A_2, and A_3 in Figure 1). The examples of the two label-streams as illustrated in Figure 1 are very similar but have differences in start- and end-times. The goal is to calculate the MI value for these two sensors. For a certain amount of time, the two sensors deliver their results, which are stored in an appropriate format for further processing.

(ii) *Interpolation*: once the activity label sequences from the two sensors that have to be used to calculate the MI value are available, the second step can be executed. Beforehand, the start- and end-times have to be calculated. This is important since the timespan has to be defined, where both sensors continuously delivered data. Therefore, the latest start-time and the earliest end-time (t_1 and t_2 in Figure 1) from sensor S_1 and sensor S_2 have to be determined. Then, the label sequences have to be interpolated. In the sequence, where no activity is stored (thus, even no null-class is detected), the prior activity remains active (at least for a predefined timespan to avoid result falsification). This is indicated in Figure 1 with the black slots. This is necessary to avoid an undefined activity class when calculating the MI values.

(iii) *Slicing*: after the preparation of the activity label sequences (thus defining start- and end-times and interpolating undefined activity slots), the sequences can be sliced into pieces of same size/time. The size of these slots (in Figure 1 defined as x) can be provided as parameter and can be varied as needed.

(iv) *Result*: the sliced activity label sequences are translated into vectors of integers, containing the activity-label identifier in each sliced slot. This is illustrated at the very bottom of Figure 1. These two integer vectors are then input for the equations above that calculate the MI value for two discrete variables.

This MI value (thus also the diversity value) between two sensors is an essential and important information to estimate the accuracy and the DoF in an ensemble. Therefore, this newly achieved information, that is calculated at runtime of the system, is directly preserved in the corresponding ExperienceItems. As it cannot be defined, when the calculation of the mutual information (MI) values is finished and stable, the computed value is subject to further changes. Therefore, not the result of the MI value calculation is stored, but the cumulative moving average [25]. Equation (5) shows how this value is calculated, where n is the number of iterations in the calculation process, MI is the currently calculated value, and MI_{i-1} is the old value, which can be found in the ExperienceItem:

$$\mathrm{MI}_i = \frac{\mathrm{MI}_{i-1} * n + \mathrm{MI}}{n+1}. \quad (5)$$

The calculation of the mutual information (MI) value between two sensors, the preservation, and continuous update at runtime are key elements to estimate the accuracy of an ensemble at runtime and therefore optimize the ensemble in terms of energy consumption and expected recognition rate. The information-theoretic score (ITS) value of an ensemble can be calculated by utilizing the pairwise diversity and MI values of the sensors by applying the aforementioned equations. Thus, in order to maximize the accuracy, the ITS value of an ensemble has to be maximized. The following Sections 3.2.1 and 3.2.2 present two approaches to optimize the ensemble by utilizing this MI and ITS values that are calculated by the system at runtime. Evaluations of the method to dynamically calculate the diversity values at runtime are presented in Section 4.

3.2. Ensemble Shaping and Optimization. The following two sections present two concepts and methods on how the ensemble DoF can be estimated (upon usage of the MajorityVoting fusion [18] method) and how the ensemble can be optimized and shaped with respect to the expected accuracy and the required energy for the sensor devices.

3.2.1. The Knapsack Metaphor for Ensemble Optimization. The knapsack problem is a known combinatorial optimization problem [26–28]. A set of items is given, each with a weight and a value, and the goal is to determine the combination of items that are packed in the knapsack, so that the total weight does not exceed a given maximum allowed weight, and that the sum of the values of the items in the knapsack are maximized. This problem is known to be in the complexity class *NP-complete* [29], which means that exact solutions can only be computed for very few items. Usually, if a solution for a bigger set of items shall be computed, this is done with mathematical approximations or heuristic algorithms, whereas it cannot be assured to find the optimal solution, but at least a rather good one (examples for such methods are simulated annealing [30], Branch and Bound [31], or Tabu Search methods [32]).

Nevertheless, the knapsack as metaphor for ensemble optimization is a promising approach, at least for a small set of available sensors, and if a maximum allowed total ensemble energy consumption is given. Thus, the degree of fulfillment (DoF) of the individual sensors can be seen as the value, the energy demand as the weight in the traditional definition of the knapsack problem. Then, by searching the complete solution space (i.e., all possible combinations of sensors, which is 2^n for n sensors), it is possible to find the set of sensors that maximizes the accuracy (by estimating the DoF with the information-theoretic score (ITS) value, see Section 3.1), whilst not exceeding the allowed total energy consumption of the ensemble. If the solution space becomes too big (i.e., too many sensor candidates in the set S_C), a heuristic solution has to be utilized (e.g., the linear efficiency analysis (LEA) optimization method, see Section 3.2.2).

The following example of a particularly available sensing infrastructure (see Table 2) illustrates the working principle of the knapsack optimization method. In the example, 9

sensors (i.e., $2^9 = 512$ combinations in total in the solution space), the available ITS values (for demonstration purposes, the sensors feature specific ITS values; in an actual system, the pairwise mutual information (MI) values are available and the ITS value for sensor combinations is calculated as shown in the equations in Section 3.1), and the energy demands (as defined in the technical sensor self-description) for each single sensor are given.

The mathematical formulation of the exact knapsack ensemble optimization method is as follows [26–28]:

$$\max \sum_{i=1}^{n} x_i * d_i \tag{6}$$

subject to

$$\sum_{i=1}^{n} x_i * e_i \leq L$$
$$x_i \in \{0, 1\} \quad \forall i = 1, \ldots, n, \tag{7}$$

whereas the following variables/constants are used:

(i) $x_i \ldots$ binary variable; 1 if sensor i is in the ensemble, 0 otherwise;

(ii) $d_i \ldots$ ITS value of sensor i;

(iii) $e_i \ldots$ energy demand of sensor i;

(iv) $L \ldots$ allowed energy limit of ensemble.

If, for example, the maximum allowed energy consumption of the resulting ensemble is set to 6.5 W (this could be defined in the goal description, where a parameter defining the maximum allowed energy value is provided), the method considers all possible combinations (in this case 512) and delivers the optimal solution: $\{S_4, S_5, S_7, S_8\}$, with a total information-theoretic score (ITS) value of 42 (again for demonstration purposes, the example is shown in Table 2 feature ITS values, and the method simply sums them up; in an actual system, the pairwise mutual information (MI) values are available and the ITS value for sensor combinations is calculated as shown in Section 3.1) and a total energy consumption of 6.4 W. Listing 1 displays the console-output for the exemplary sensor setting as defined in Table 2.

This method is feasible and applicable if two prerequisites are given:

(i) an upper limit for the total allowed energy consumption in the ensemble is given,

(ii) the number of ensemble candidates does not exceed approximately 10 sensors, since the ensemble definition and configuration has to be done in real-time.

In contrast to the knapsack optimization method, the linear efficiency analysis (LEA) method (see Section 3.2.2) needs neither a defined energy constraint, nor a number of sensors limitation.

```
Candidate List
   ID: S1; ITS: 14.0; Energy: 3.7
   ID: S2; ITS: 18.0; Energy: 4.2
   ID: S3; ITS: 10.0; Energy: 2.0
   ID: S4; ITS: 16.0; Energy: 2.4
   ID: S5; ITS: 12.0; Energy: 1.4
   ID: S6; ITS: 3.0; Energy: 0.8
   ID: S7; ITS: 4.0; Energy: 0.8
   ID: S8; ITS: 10.0; Energy: 1.8
   ID: S9; ITS: 12.0; Energy: 1.9
Sensor Ensemble:
   S4   S5   S7   S8
   ITS: 42.0
   Energy: 6.400000
```

Listing 1: Output for the knapsack ensemble optimization method for an exemplary sensor setting.

Table 2: Exemplary sensor infrastructure as input for the knapsack optimization method.

Sensor ID	ITS value	Energy (Watt)
S_1	14	3.7 W
S_2	18	4.2 W
S_3	10	2.0 W
S_4	16	2.4 W
S_5	12	1.4 W
S_6	3	0.8 W
S_7	4	0.8 W
S_8	10	1.8 W
S_9	12	1.9 W

3.2.2. The Linear Efficiency Analysis for Ensemble Optimization. The knapsack metaphor for ensemble shaping is an exact optimization method, which suffers from two major disadvantages: (i) the allowed total energy consumption has to be defined and provided to the system and (ii) the scalability (i.e., combinatorial solution) of the method is limited due to the real-time operability requirements by approximately 10 sensors. Therefore, heuristic solutions that do not expand the complete solution space (thus are not ensuring to find the optimal solution, but an acceptable one) have to be selected for ensemble optimization. The linear efficiency analysis (LEA) method operates on the efficiency of the sensor candidates that can be calculated by dividing the ITS value by the energy demand of the sensor. The approach operates in $O(n^2)$ (where n is the number of sensor candidates), by simply starting from an initial ensemble that contains all sensors and subsequently removing the one with the worst efficiency. The occurring ensembles (that are constructed until only one sensor is left) are examined with respect to their efficiency. As a result, the ensemble with the highest efficiency is calculated and configured. The LEA method consists of three steps: (i) initialization, (ii) ensemble reduction, and (iii) resulting ensemble. Table 3 shows an example of a sensor infrastructure (again, for demonstration

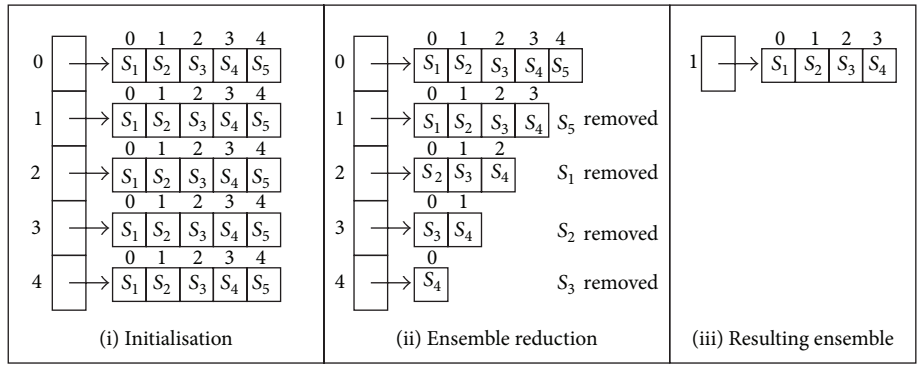


Figure 2: The working principle of the LEA optimization method consisting of three steps: (i) initialization, (ii) ensemble reduction, and (iii) resulting ensemble.

Table 3: Exemplary sensor infrastructure as input for the linear efficiency analysis optimization method.

Sensor ID	ITS value	Energy (Watt)	Sensor efficiency
S_1	21	0.37 W	(21/0.37) = 56.76
S_2	39	0.37 W	(39/0.37) = 105.41
S_3	78	0.54 W	(78/0.54) = 144.44
S_4	88	0.35 W	(88/0.35) = 251.43
S_5	15	3.20 W	(15/3.2) = 4.69

purposes, the sensors feature specific ITS values; in an actual system, the pairwise mutual information (MI) values are available and the ITS value for sensor combinations is calculated as shown in the equations in Section 3.1), and in the following Figure 2 and paragraph, the working principle of the LEA optimization is described (and demonstrated by application on the example shown in Table 3).

In the first phase the (i.e., initialization), the starting ensemble consisting of all available sensor candidates is calculated and stored in a list n times. Furthermore, the efficiencies for each single sensor candidates are calculated (see Table 3, column sensor efficiency). These values are important in the second phase (i.e., ensemble reduction), where in each step, the sensor with the lowest efficiency is subsequently removed from the ensemble. In this exemplary setting, the "worst" sensor is S_5, followed by S_1, S_2, and S_3. The $n-1$ worst sensors are removed step by step. For each resulting possible ensemble, the total efficiency is calculated (for that reason it is important to start with the initial ensemble that is stored in a data structure n times) and multiplied by the actual number of sensors and the ensemble with the highest efficiency is delivered as a result. The processing steps in the ensemble reduction phase in this example are listed in Table 4 (where E_0 stands for Ensemble$_0$).

Therefore, the resulting ensemble in this example would be $\{S_1, S_2, S_3, S_4\}$ with an efficiency value of 554.60. This very simple approach can be further improved by considering a more sophisticated calculation of the ensemble ITS values (in this example, they have been simply added), and by adding further constraints (e.g., minimal number of required sensors in ensemble, etc.). Even a limit for the ensemble energy consumption can be provided (much like it is necessary in the knapsack solution method), then the expanded ensembles have to be examined to find the resulting sensor setting. Next Section 4 presents evaluations of the introduced methods in a real-time system operating in an opportunistic way.

Table 4: Step-by-step working principle of the LEA optimization method.

Ensemble ID	Sensors	ITS	Energy (Watt)	Ensemble efficiency
E_0	$\{S_1, S_2, S_3, S_4, S_5\}$	241	4.83	(241/4.83) $*$ 5 = 249.48
E_1	$\{S_1, S_2, S_3, S_4\}$	226	1.63	(226/1.63) $*$ 4 = 554.60
E_2	$\{S_2, S_3, S_4\}$	205	1.26	(205/1.26) $*$ 3 = 488.10
E_3	$\{S_3, S_4\}$	166	0.89	(166/0.89) $*$ 2 = 373.03
E_4	$\{S_4\}$	88	0.35	(88/0.35) $*$ 1 = 251.43

4. Evaluation

The methods for estimating the ensembles accuracy based on MajorityVoting fusion (i.e., the activity class where most of the classifiers agree on is selected as fusion result) [18] and the ensemble optimization and shaping are tested and evaluated on a real-world dataset. In this rich dataset, recorded in a kitchen scenario, 12 subjects performed activities for about 14 minutes per run (whereas each subject performed 6 runs) surrounded by 72 sensors with 10 different modalities. The dataset has been labeled to have a reliable ground truth available for evaluation purposes (i.e., the modes of locomotion have been added to the dataset). Details about the recordings and the dataset can be found in [4, 5]. The OPPORTUNITY framework enables the utilization of the sensors in the dataset as if the sensors were currently present by abstracting them in form of PlaybackSensors [1, 15]. This means that, the framework performs real-time activity recognition with the sensors that are simulated by replaying the prerecorded sensor data. As this paper aims at presenting and evaluating methods (i) for estimating the ensemble accuracy by utilizing the pairwise diversity between sensors and (ii) for optimizing an ensemble in terms of this estimated accuracy and the expected energy demand, a

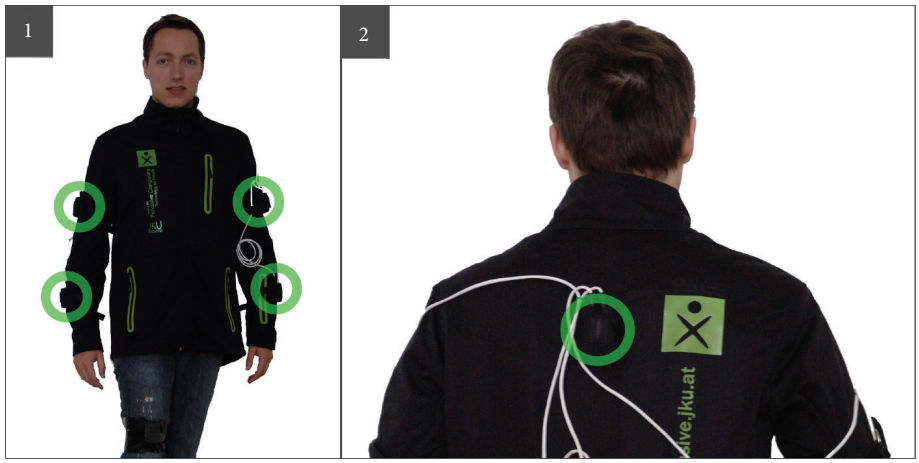


Figure 3: Illustration of the on-body sensor placement in the evaluation sessions (see also [17, 24]).

```
<swe: field name="Locomotion_motionjacket-
      ACC_LUA_ACCGYR_MEANVAR_DecisionTree_NOFUSE">
   <swe: DataRecord>
      <swe: field name="count">
        <swe: Quantity><swe: value>1588</swe: value></swe: Quantity>
      </swe: field>
      <swe: field name="miValue">
        <swe: Quantity><swe: value>1.15343</swe: value></swe: Quantity>
      </swe: field>
   </swe: DataRecord>
</swe: field>
```

Listing 2: Part of the sensor self-description for the BAC sensor storing the pairwise mutual information with the LUA sensor after 1588 iterations.

subset of the available sensors is picked for recognizing the modes of locomotion (i.e., WALK, SIT, STAND, and LIE) as activity of interest. The sensors that are used for evaluation purposes are located on the body of the subject. In detail, the upper body is equipped with 5 sensors of type Xsens MTx, which provides 3D acceleration. The sensors are located on both arms (right/left lower/upper arm—RLA, RUA, LLA, and LUA), and on the upper back (BAC). The on-body sensor placement is displayed in Figure 3 (see also [17, 24]).

The first method evaluation tackles the dynamic generation of the diversity measurements. Therefore, 10 independent runs involving the five on-body sensor have been conducted, whereas every 300 samples (the number of samples per calculation iteration is kept intentionally low due to scalability reasons) the method described in Section 3.1 calculated the pairwise diversity value for all 10 possible combinations and stored this value (by utilizing the cumulative moving average) into the corresponding sensor self-description. Listing 2 shows a clipping of the self-description for one sensor (BAC) and its pairwise diversity value with another sensor (i.e., LUA).

The pairwise diversity value perfectly converges towards a value by utilizing the moving average value, which can be further used to define the information-theoretic score (ITS) value which can be used to estimate accuracies for the resulting ensembles, as described in Section 3.1. Figure 4 shows a graphical visualization for 280 iterations (the two top-level images, boxplot and plot chart), and 2500 iterations (lower-level images). The figure displays the pairwise diversity values for each of the 10 possible combinations (i.e., BAC-LLA, BAC-LUA, BAC-RLA, BAC-RUA, LLA-RLA, LUA-LLA, LUA-RLA, LUA-RUA, RUA-LLA, and RUA-RLA).

For the evaluation of the optimization methods, the ongoing dynamic updates of the diversity measures have been stopped after approximately 2500 iterations to have a reliable (and comparable) sensor self-description base for comparing the two optimization methods. To be able to compare the optimization techniques and to highlight the benefits in terms of optimizing towards reduced energy consumption and maximized accuracy, one session (lasting approx. 14 min, involving the 5 on-body sensors described above) was conducted by combining all available sensors into an ensemble. The results for this session are shown as confusion matrices in Figure 5 (top left illustration), in Figure 6 (illustrating the accuracies and the efficiencies) and in Table 5 (method *ALL*).

One major benefit of using a prerecorded dataset is the fact that a labeled ground truth is available which makes the

Table 5: Overview of the evaluation result for the three different methods (the knapsack optimization method was started twice: one run with 5 W as boundary for the energy consumptions and one run with 7.5 W).

Method	Ensemble	Energy (Watt)	Accuracy	Efficiency
ALL	BAC, LUE, LLA, RUA, RLA	13.5 W	0.759	0.056
LEA	RLA	2.7 W	0.641	0.237
*KS*5	LLA	2.7 W	0.561	0.208
*KS*7.5	BAC, RLA	5.4 W	0.680	0.126

calculation of the actual accuracy possible. The accuracy (i.e., telling how many of the predicted activity classes are correctly classified, manifested as DoF value in the self-description) is the most important metric that quantifies the quality of a classifier or an activity recognition application. The accuracies for the four ensembles that have been generated by utilizing the optimization methodologies can be calculated out of the confusion matrices that display the number of correctly classified classes (the main diagonal of the matrix) and the false positives/negatives. The two optimization methods as described in Sections 3.2.1 (knapsack optimization) and 3.2.2 (linear efficiency analysis (LEA)) have been utilized in equal playback sessions within the dataset [4, 5] used by the system as PlaybackSensors (simulating sensors by using prerecorded data). The knapsack method was subject to evaluation twice, since the method needs a specific maximum allowed energy boundary parameter (therefore, this value was set to 5 W in the first run and 7.5 W in the second run). The results for the four sessions are displayed in the form of confusion matrices in Figure 5, whereas the top left matrix shows the results for the ensemble consisting of all five available sensor (i.e., BAC, LUA, LLA, RUA, RLA); the top right matrix shows the classification result for the LEA optimization methods, where the method configured an ensemble consisting of only one sensor (i.e., RLA); the two lower images show results for the knapsack optimization methods with different maximum allowed energy boundaries (lower left has an energy boundary of 7.5 W and an ensemble consisting of the BAC and RLA sensors; lower right has an energy boundary of 5 W and an ensemble consisting of the LLA sensor). The results in form of the accuracies and efficiencies are summarized in Table 5, where the accuracies are calculated out of the confusion matrices for all four evaluation ensembles. This summary highlights the benefits of optimizing ensembles instead of using all available and suitable sensors.

Figure 6 visualizes the calculated accuracies (gray bars) and the efficiency (black bars) of the four methods. As it can be seen, utilizing all sensors in the ensemble results in the highest accuracy (resp., degree of fulfillment (DoF) value), but has the lowest efficiency value. This is obvious, since the ensemble consisting of all sensors has the highest energy consumption. The linear efficiency analysis (LEA) method performs best in this experiment session in terms of efficiency. An interesting issue is that the knapsack method (configured with 5 W upper boundary) selected the LLA sensor (resulting in an accuracy of 0.561), and the LEA method selected the RLA sensor (resulting in an accuracy of 0.641). The reason for this discrepancy is the fact that both sensors have the same initial DoF value in their corresponding self-description, so they are equal in terms of energy consumption, expected accuracy, and thus also efficiency. Coincidentally, one method selected the RLA sensor, the other method the LLA sensor. The different values for the accuracy are caused by the fact that these accuracies are the actual ones, calculated by comparing the predicted classes to the actual classes from the available ground truth within the dataset.

The ensemble consisting of all sensors has an energy demand of 13.5 W (2.7 W for each single sensor) and reaches an accuracy for the modes of locomotion of 0.759. Compared to that, the ensemble generated by using the LEA method consists of only one sensor (i.e., RLA); therefore, it needs only 2.7 W but reaches an impressive accuracy of 0.641. The same holds for the knapsack optimization ensemble, where the accuracies are also very close to the original ensemble consisting of all sensors, but demand much less energy.

An additional evaluation of the two ensemble optimization methods is done without the dataset. A set of ExperienceItems is artificially generated, including a random value for the DoF of the activity WALK (between 0.0 and 1.0), a random value for the pairwise MI values (between 0.0 and 4.9, which is a realistic approximation), and a randomly generated energy consumption (between 0.5 W and 4.9 W) per sensor. Both methods are applied within the framework based on the set of constructed ExperienceItems, where the focus is on the aggregated energy consumption of the sensors selected for the ensemble and the ITS value per ensemble. As already discussed, the "best" ensemble minimizes the energy consumption whilst maximizing the accuracy. The information-theoretic score (ITS) value is used for estimating the expected accuracy; the higher the ITS value of an ensemble is, the higher is the expected accuracy. Both methods are compared against a random ensemble generation method and the selection of all candidate sensors (i.e, "select all" method). The results are illustrated in Figure 7.

The upper two illustrations in Figure 7 visualize the knapsack ensemble optimization method. On the left-hand side, the energy consumption for the method compared to a random and "select all" method is illustrated. The knapsack method needs an upper bound for the maximum allowed energy consumption (which is set to 15 W in this run); therefore, the right-hand side illustration (showing the ensemble ITS value in relation to the number of sensors in the candidate set) is more meaningful. In most of the cases, the knapsack method outperforms the random and "select all" methods in terms of the reachable ITS value. The two lower images show the results for the LEA ensemble optimization method for up to 40 sensors in the candidate set. As shown on the left-hand side, the energy consumption level of LEA is constantly kept low. The most impressive and convincing visualization on the lower right-hand side shows the calculated ITS values for the ensembles in relation to the

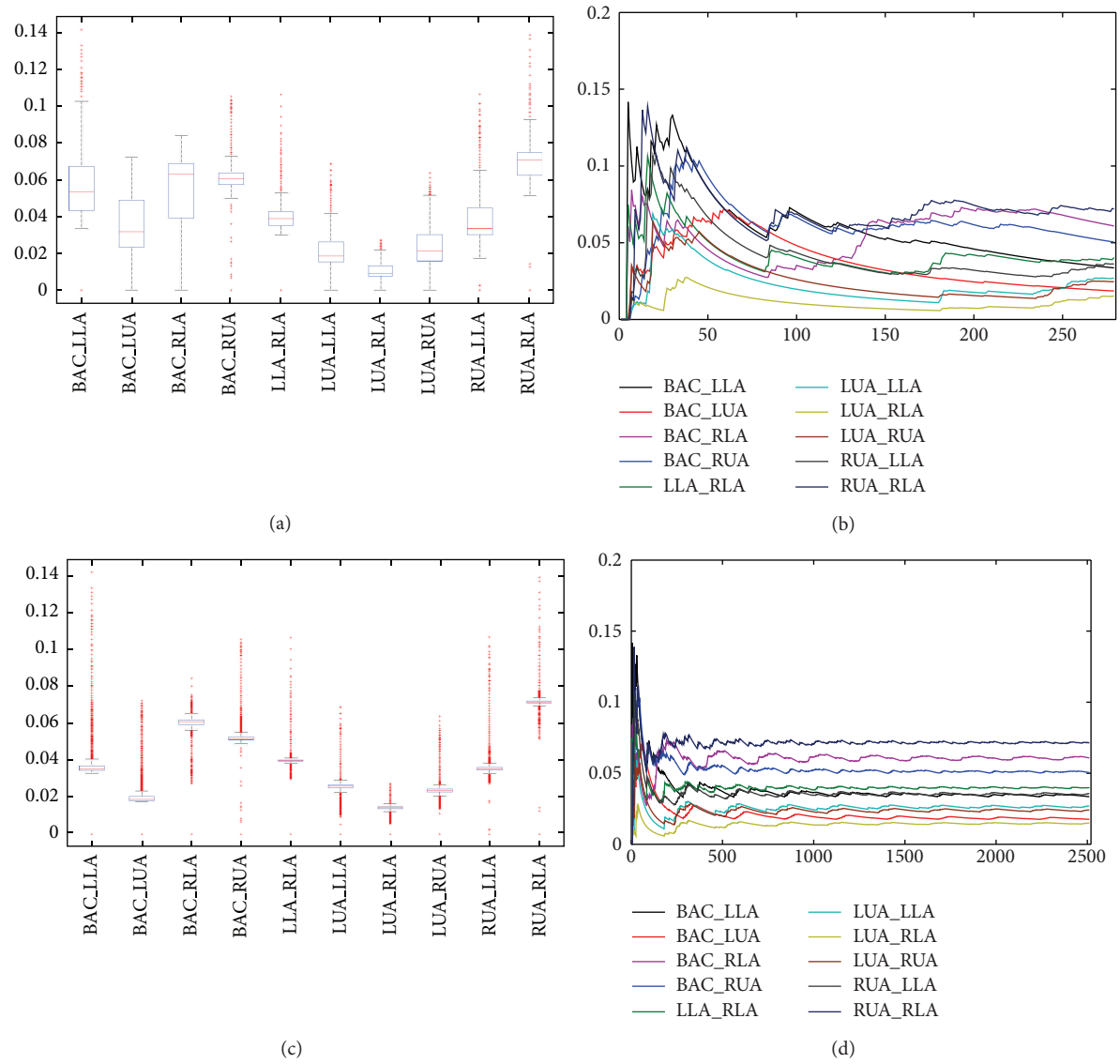


Figure 4: Development of the diversity values for 5 sensors (i.e., 10 combinations). The upper two images show the diversity values (as boxplot and plot chart) for approx. 280 iterations; the lower two images show the diversity values for approx. 2.500 iterations resulting in stable values written back into the sensors' self-descriptions.

number of sensors in the candidate set (the random method selects approx. 50% of the available candidates; the "select all" method combines all available candidates into the ensemble). It can be safely said that the LEA method outperforms both naive approaches in terms of energy consumption and expected accuracy.

5. Conclusion

The work presented in this paper tackles two challenging features in an opportunistic activity and context recognition system: (i) the estimation of the accuracy of a sensor ensemble without relying on a labeled ground truth and (ii) optimizing the ensemble (resp., the involved sensor systems) in terms of accuracy and expected energy consumption. Regarding the first research challenge, the method proposed in this paper utilizes the mutual information (i.e., the diversity) between two sensors to calculate an information-theoretic score. The basic idea is that if two sensors have a high diversity value, it is very unlikely that they do not agree on wrongly classified activities, thus resulting in a higher accuracy if such sensors are utilized together in an ensemble. This correlation can be broadened to the available sensors resulting in an estimate of the ensemble accuracy for a specific recognition goal in form of the degree of fulfillment (DoF). One major contribution presented in this paper is the ability of calculating this diversity value at system's runtime which results in a stable value for sensor combinations and can be persistently stored in the corresponding self-descriptions. Thus, the sensor (and the entire system) learns at runtime and is able to enhance its capabilities.

Optimizing and shaping sensor ensembles is the second major contribution of this paper. Particularly in large-scale

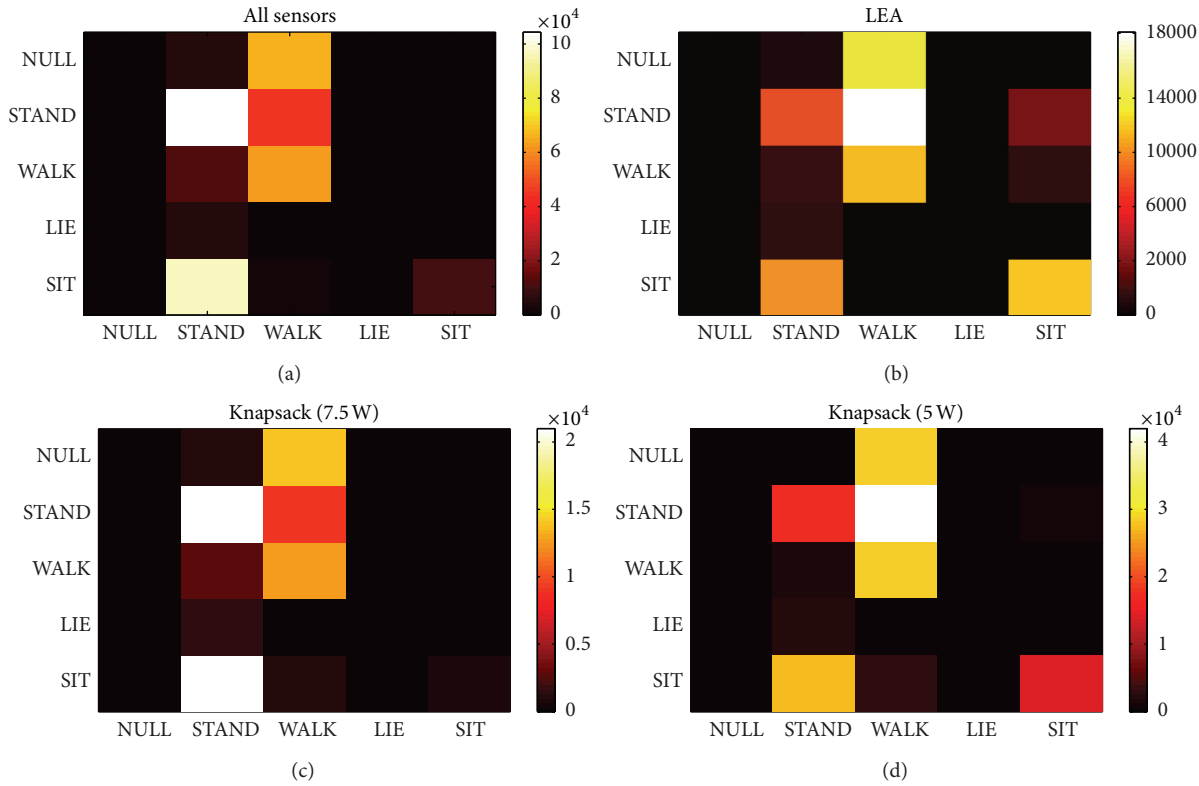


Figure 5: Confusion matrices generated by comparing the predicted activity classes to the actual classes received from a labeled ground truth in the reference dataset [4, 5] for the investigated optimization methods (i.e., LEA and knapsack) compared to no optimization (i.e., all sensors are integrated in the ensemble).

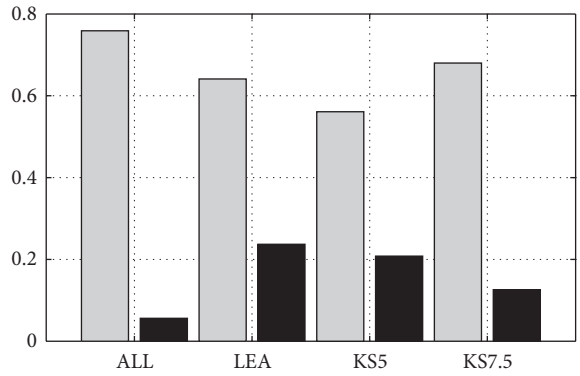


Figure 6: Overview of the accuracies (gray bars) and the efficiencies (black bars) of the optimization methods.

systems, utilizing all available sensors for activity recognition might not necessarily improve the recognition capabilities. However, the lifetime of the system is minimized since sensors can run out of energy. Therefore, optimizing the ensemble in terms of estimated accuracy and expected energy consumption is being identified as important aspect in an opportunistic system. Two methods are introduced: (i) the knapsack as metaphor for ensemble optimization and (ii) the linear efficiency analysis (LEA) method. The two solutions differ on the one hand in the required input parameters (i.e., the knapsack needs a maximum allowed energy consumption as boundary) and on the other hand in the algorithmic complexity. The knapsack expands the complete solution space (i.e., $O(2^n)$), which results in an optimal solution. The LEA method runs in $O(n^2)$ time and does not need a boundary for the allowed energy consumption. Thus, the solution is not necessarily an optimal solution (but good enough, as the results show).

These two challenging features are identified as being important building blocks that still have mission in the vision of opportunistic activity recognition. The results of the evaluation of the presented methods are promising,

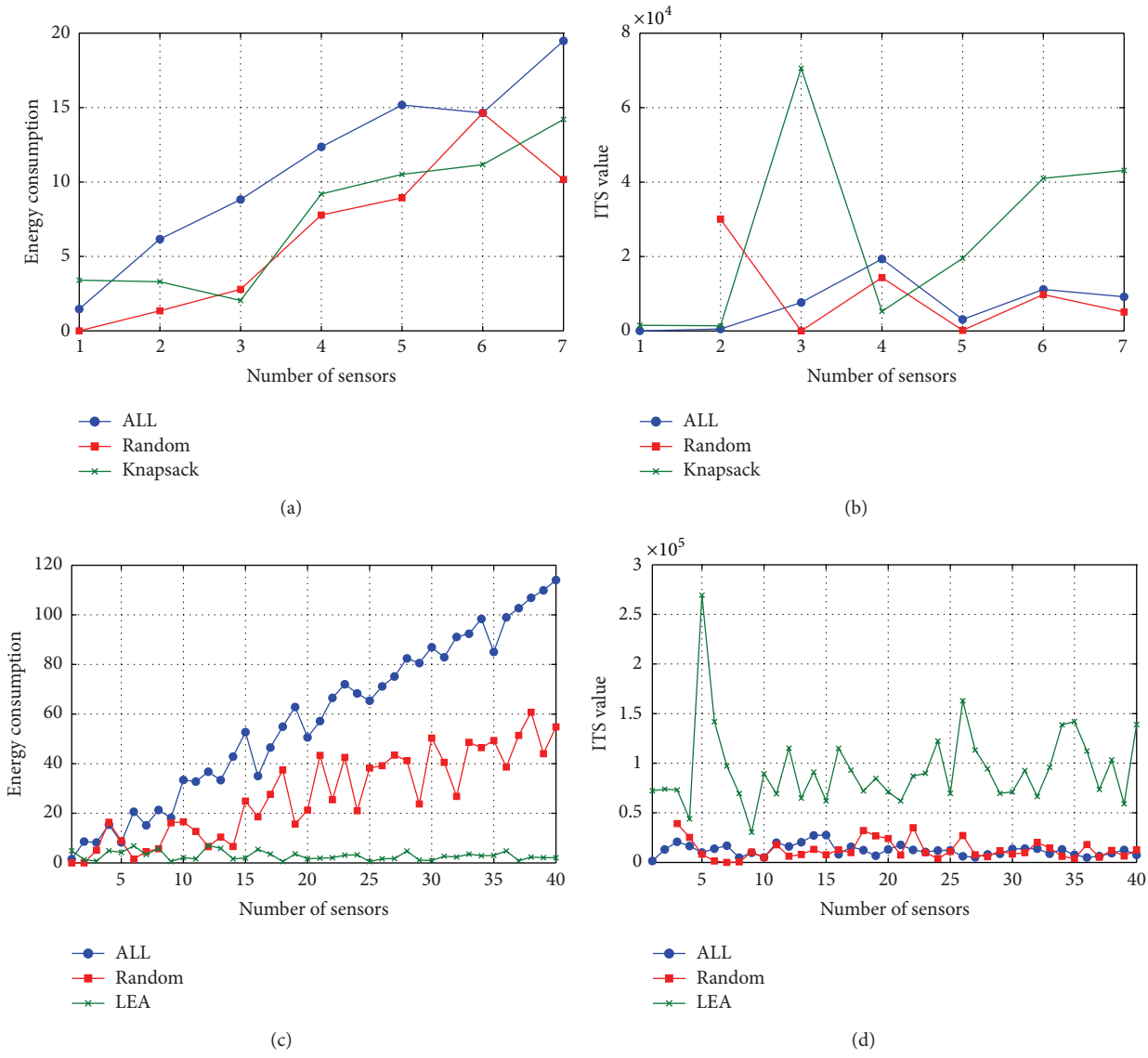


FIGURE 7: Illustration of the analysis of the knapsack and LEA ensemble optimization method compared to a random and "select all" method.

which means that the methods will be further expanded and investigated.

Acknowledgments

The project OPPORTUNITY acknowledges the financial support of the Future and Emerging Technologies (FET) Programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open Grant no. 225938.

References

[1] M. Kurz, G. Hölzl, A. Ferscha et al., "The opportunity framework and data processing ecosystem for opportunistic activity and context recognition," *International Journal of Sensors, Wireless Communications and Control*, vol. 1, no. 2, pp. 102–125, 2011.

[2] D. Roggen, K. Förster, A. Calatroni et al., "Opportunity: towards opportunistic activity and context recognition systems," in *Proceedings of the 3rd IEEE WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC '09)*, IEEE CS Press, Kos, Greece, June 2009.

[3] G. Hölzl, M. Kurz, and A. Ferscha, "Goal oriented opportunistic recognition of high-level composed activities using dynamically congured hidden markov models," in *Proceedings of the 3rd International Conference on Ambient Systems, Networks and Technologies (ANT '12)*, Niagara Falls, Ontario, Canada, August 2012.

[4] D. Roggen, A. Calatroni, M. Rossi et al., "Collecting complex activity data sets in highly rich networked sensor environments," in *Proceedings of the 7th International Conference on Networked Sensing Systems (INSS '10)*, IEEE Computer Society Press, Kassel, Germany, June 2010.

[5] D. Roggen, A. Calatroni, M. Rossi et al., *Walk-Through the Opportunity Dataset For Activity Recognition in Sensor Rich Environments*, 2010.

[6] D. Salber, A. K. Dey, and G. D. Abowd, "The context toolkit: aiding the development of context-enabled applications," in *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit (CHI '99)*, pp. 434–441, ACM, New York, NY, USA, 1999.

[7] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive Computing*, A. Ferscha and F. Mattern, Eds., vol. 3001, pp. 1–17, Springer, Berlin, Germany, 2004, Lecture Notes in Computer Science.

[8] H. Junker, O. Amft, P. Lukowicz, and G. Troster, "Gesture spotting with bodyworn inertial sensors to detect user activities," *Pattern Recognition*, vol. 41, no. 6, pp. 2010–2024, 2008.

[9] P. Lukowicz, J. Ward, H. Junker et al., "Recognizing workshop activity using body worn microphones and accelerometers," in *Pervasive Computing*, A. Ferscha and F. Mattern, Eds., vol. 3001, pp. 18–32, Springer, Berlin, Germany, 2004, Lecture Notes in Computer Science.

[10] J. A. Ward, P. Lukowicz, and H. W. Gellersen, "Performance metrics for activity recognition," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 1, pp. 1–6, 2011.

[11] E. Tapia, S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *Pervasive Computing*, vol. 3001, pp. 158–175, Springer, Berlin, Germany, 2004, Lecture Notes in Computer Science.

[12] B. Logan, J. Healey, M. Philipose, E. M. Tapia, and S. Intille, "A long-term evaluation of sensing modalities for activity recognition," in *Proceedings of the 9th international conference on Ubiquitous computing (UbiComp '07)*, pp. 483–500, Springer, Berlin, Germany, 2007.

[13] T. van Kasteren, A. Noulas, G. Englebienne, and B. Kröse, "Accurate activity recognition in a home setting," in *Proceedings of the 10th international conference on Ubiquitous computing (UbiComp '08)*, pp. 1–9, ACM, New York, NY, USA, 2008.

[14] J. Mantyjarvi, J. Himberg, and T. Seppanen, "Recognizing human motion with multiple acceleration sensors," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC '01)*, vol. 22, pp. 747–752, 2001.

[15] M. Kurz and A. Ferscha, "Sensor abstractions for opportunistic activity and context recognition systems," in *Proceedings of the 5th European Conference on Smart Sensing and Context (EuroSSC '10)*, K. K. G. Lukowicz and P. Kunze, Eds., Springer, Berlin, Germany, November 2010.

[16] M. Kurz, G. Hölzl, A. Ferscha, H. Sagha, J. R. Millan, and R. Chavarriaga, "Dynamic quantication of activity recognition capabilities in opportunistic systems," in *Proceedings of the 4th Conference on Context Awareness for Proactive Systems (CAPS '11)*, pp. 15–16, Budapest, Hungary, May 2011.

[17] M. Kurz, G. Hölzl, A. Ferscha, A. Calatroni, D. Roggen, and G. Troester, "Realtime transfer and evaluation of activity recognition capabilities in an opportunistic system," in *Proceeding of the 3rd International Conference on Adaptive and Self-Adaptive Systems and Applications (ADAPTIVE '11)*, pp. 73–78, Rome, Italy, September 2011.

[18] J. Kittler, M. Hatef, and R. Duin, "On combining classiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[19] R. Chavarriaga, H. Sagha, and J. d. R. Millan, "Ensemble creation and reconfguration for activity recognition: an information theoretic approach," in *IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC '11)*, 2011.

[20] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine Learning*, vol. 51, pp. 181–207, 2003.

[21] J. Meynet and J. P. Thiran, "Information theoretic combination of pattern classifiers," *Pattern Recognition*, vol. 43, no. 10, pp. 3412–3421, 2010.

[22] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Ill, USA, 1949.

[23] A. Kraskov, H. Stögbauer, R. G. Andrzejak, and P. Grassberger, "Hierarchical clustering using mutual information," *EPL*, vol. 70, no. 2, p. 278, 2005.

[24] M. Kurz, G. Hölzl, and A. Ferscha, "Dynamic adaptation of opportunistic sensor confgurations for continuous and accurate activity recognition," in *Proceedings of the 4th International Conference on Adaptive and Self-Adaptive Systems and Applications (ADAP-TIVE '12)*, pp. 22–27, Nice, France, July 2012.

[25] Y. l. Chou, *Statistical Analysis, with Business and Economic Applications*, Holt, Rinehart and Winston, New York, NY, USA, 2nd edition, 1975.

[26] S. Martello, *Knapsack Problems: Algorithms and Computer Implementations*, John Wiley & Sons, New York, NY, USA, 1990.

[27] J. Puchinger, G. Raidl, U. Pferschy, and editors, "The core concept for the multidimensional knapsack problem," in *Evolutionary Computation in Combinatorial Optimization*, J. Gottlieb and G. Raidl, Eds., vol. 3906, pp. 195–208, Springer, Berlin, Germany, 2006, Lecture Notes in Computer Science.

[28] G. Zäpfel, R. Braune, and M. Bögl, *Metaheuristic Search Concepts: A Tutorial with Applications to Production and Logistics*, Springer, Berlin, Germany, 2010.

[29] M. R. Garey and D. S. Johnson, *Computers and Intractability, A Guide To the Theory of NP-Completeness*, W. H. Freeman, New York, NY, USA, 1990.

[30] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[31] E. L. Lawler and D. E. Wood, "Branch-and-bound methods: a survey," *Operations Research*, vol. 14, no. 4, pp. 699–719, 1966.

[32] F. Glover and M. Laguna, *Tabu Search*, Kluwer Academic Publishers, Norwell, Mass, USA, 1997.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 520507, 10 pages
http://dx.doi.org/10.1155/2013/520507

Research Article

The Experimental Study on Concrete Permeability of Wireless Communication Module Embedded in Reinforced Concrete Structures

Byung-Wan Jo, Jung-Hoon Park, and Kwang-Won Yoon

Department of Civil Engineering, Hanyang University, Seoul, Republic of Korea

Correspondence should be addressed to Jung-Hoon Park; goolss@nate.com

Received 11 October 2012; Accepted 29 April 2013

Academic Editor: Yonggang Wen

Recently, as the information industry and mobile communication technology develop, their study is conducted on the new concept of intelligent structures and maintenance techniques that apply wireless sensor network, USN (Ubiquitous Sensor Network), to social infrastructures such as civil and architectural structures on the basis of the concept of Ubiquitous Computing, which invisibly provides human life with computing, along with mutually cooperating, compromising, and connecting networks to each other by having computers within all objects around us. The purpose of this study is to investigate the capability of wireless communication of sensor node embedded in reinforced concrete structure with a basic experiment on electric wave permeability of sensor node by fabricating molding with variables of concrete thickness and steel bars that are mostly used in constructing structures to determine the feasibility of application to constructing structures with USN. By installing wireless communication module inside the structures, it is possible to communicate to measure the pitch of steel bars and permeability of concrete, by measuring in both directions horizontally and vertically. The magnitude of an electric wave in the range of used frequencies was measured by using Spectrum Analyzer. This electric wave was numerically analyzed and the effective wavelength of frequencies was analyzed by the properties of a frequency band area. As a result of constructing structures with wireless sensors, Plain concrete showed 45 cm for depth of permeability. Reinforced concretes that has pitches of 5 cm showed 37 cm and pitches of 15 cm showed 45 cm for the depth of permeability. This suggests that if the pitch of steel bars was more than 15 cm, it would not affect wireless communication.

1. Introduction

Structural health monitoring and damage detection have been of great concern in the design, operation maintenance, and repair of many civil structures. In order to monitor the health of large infrastructures such as bridges, highways, buildings, and dams, the outputs from various kinds of sensors were installed at different places in civil structures data collected simultaneously and/or sequentially [1].

Besides, as the recent information industry and mobile communication technology are developed by using the wireless sensor network technique. USN (Ubiquitous Sensor Network), with the concept of Ubiquitous Computing, which invisibly provides information to human with computing along with mutually cooperating, compromising, and connecting networks to each other by having subminiature computers within all objects around us, is being tried mainly by the information and communication industry [2, 3].

The measuring of sag as cracks in bridge structures is measured by the wireless measurement sensor in the field of bridges, and the ubiquitous bridge system that is monitoring erosion by the water flow on the lower part of pier is partly applied but the environment based on complete ubiquity is not established [4, 5]. In addition, there is the displacement measurement device that measures the sinking amount of tunnel works and displacement inside tunnels before and after an explosion using wireless optical fiber in tunnels, and the system for searching and maintaining underground structures by attaching RFID tags to underground structures having USN and the field application like the automatic sluice control system that monitors and controls the level of water is being attempted [6, 7].

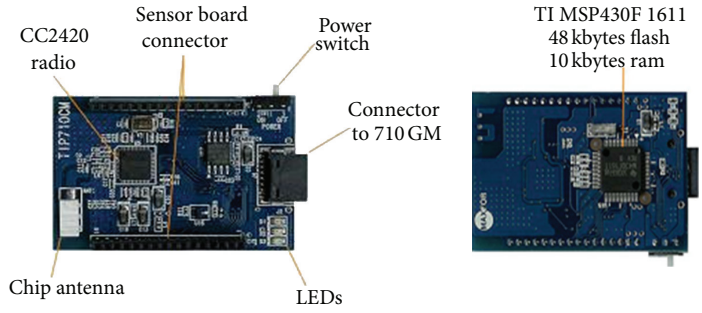


Figure 1: Structure detailed of cc2420.

The depth of permeability of concrete by frequencies and the effect of the pitch of steel bars with Active RF wireless communication module embedded into concrete for structuring ubiquitous structures as a basic experiment for an application to construction industry were looked into in this study.

For this, specimens were made of concrete and bars that are mostly used for construction of structures wile wireless communication modules were installed inside them, the possible communication distance and the intensity of electric waves were measured. The effect of the length of wavelength of frequencies was analyzed according to the properties of the frequency band. This is to suggest the basic data for constructing social infrastructure having the concept of ubiquity for the real-time monitoring of the level of displacement and damage inside the structure as well as external damage by embedding a wireless sensor in social infrastructure through this.

2. Background

In this study, we measured the transmittance and radio signal strength of a sensor which was embedded in the structure using USN. The fundamental experiments were conducted to prevent symptom like demolition, crack, corrosion, and fire through the determination whether or not the embedded sensor works in construction structure [8].

Recently studies that find corrosion and cracks and defects in the concrete structure using radio waves transmission characteristics, such as "Radiowave Propagation through Window Panes and Reinforced Concrete Structure" research [9], are in progress. There is no enough study about the radio wave attenuation but we did experimental work about the radio wave attenuation effects depending on depth, thickness, and amount of reinforcing steel using the test piece.

3. Devices for Experiments

How much steel bars and concrete, which are the main materials of the construction of structures, affect transmitting and receiving of electric waves was analyzed as a basic experiment for the field application of a wireless sensor network for civil and architectural structures. The possible communication distance by the thickness of wireless communications modules installed inside plain concrete and reinforced concrete and the communication distance by the pitch of steel bars were measured, the waveform of electric waves by thickness was measured, the phenomenon of diminution of electric waves was evaluated by using a Spectrum Analyzer, and the effect of the length of frequency wavelength according to the properties of the frequency band was analyzed [10].

3.1. Materials for Experiments. Various sized specimens (e.g, 600 × 600 × 100 (mm)) were made with variables of the concrete thickness, the pitch of steel bars, the direction of electric waves, and the intensity of RF power. When making specimens, OPC (Ordinary Portland Cement) that meets KS F 5201 and has specific gravity of 3.15 and fineness of 3,341 and the maximum size of aggregate having 13 mm Φ, specific gravity of 2.65, and fineness modulus of 2.9 were used. Small aggregates that have specific gravity of 2.62 and standard sand were used. Meanwhile, deformed bars with the 16 mm diameter were placed.

3.2. Equipment for Experiments

3.2.1. Module. By installing sensor module RF chip having frequency of 2.4 GHz inside the model and starting to receive wireless signals from the sensor while transmitting any change in temperature, humidity, and illuminance, the detailed structure of the sensor is as shown in Figure 1.

3.2.2. Spectrum Analyzer. Spectrum Analyzer used is the equipment indicating the result of measurement by the spectrum of signal or frequency band on a screen and shows what intensity range signals of certain frequency elements have [11]. The unit of horizontal axis is shown as kHz/Div and MHz/Div, which are called Span/Div (frequency span of horizontal axis). Spectrum Analyzer of model R3131A (Figure 2) was used in this experiment with frequency of 2.4 GHz and Span/Div of 50 MHz, and the result value was represented with dBM (the intensity of electric waves).

3.2.3. Chip Antenna. The pattern of radiation of Chip antenna that was attached to used module is the way of OMNI-DIRECTIONAL (Figure 3). The real size of Chip antenna is 8.0 × 4.0 × 1.5 mm, which is small and light, and it provides several data channels to a single transmission medium using FDMA (Frequency Division Multiple Access) as broadband and the output ratio against input such as

Figure 2: Spectrum Analyzer.

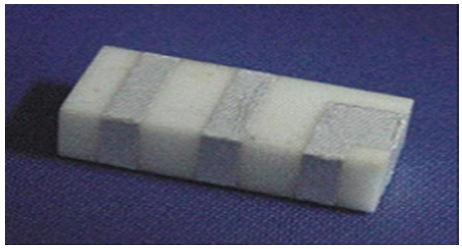


Figure 3: Chip antenna.

receiving an amplifier is high up to 1.5 dBi and it shows linear polarization [12, 13].

4. Methods of Experiments

4.1. Fabrication of Specimen. To experiment the capability of wireless communication in plain and reinforced concrete, the base plate that was grooved to put a module and plain concrete specimens with various thickness (Figures 4 and 5) were made and concrete was placed with steel bars in the pitches of 5, 7, 10, and 15 cm, respectively [14]. The mixture of concrete followed KS standard mixture, and when placing concrete, the dimension of a specimen was 600 × 600 (mm) in length and width in case of vertical penetration experiment and five of each 1, 3, 5, and 20 (cm) by the thickness were made, respectively. Besides, the length and width are 600×100 (mm) in horizontal direction and five specimens of each 1, 3, 5, and 20 (cm) by the thickness were made, respectively [15, 16].

4.2. Methods of Measurement of Receiving Distance. Aluminum foil and special case were used to block off effects of electric waves from the bottom and the side in case of an experiment of vertical-oriented electric waves permeability and from the bottom and upwards in case of an experiment of horizontal-oriented electric waves permeability. Besides, the transmission of data was confirmed by using monitoring system (Figure 6). After node number is designated for each module that was transmitted, the number of node is designated as 0 that can receive a reception. Once communication started, the state of receiving by each node could be checked by linearly connecting nodes (Figure 6) and in case that communication is cut, which could be known by the change of color of node number. Moreover, if each node number was chosen, the value of sensors for temperature, humidity, and illuminance that were installed in each selected node could be identified. The state of transmission of module installed inside concrete was checked through linear connection (Figure 6) on the basis of the baseplate with the change of concrete thickness. The exact intensity and loss of electric waves cannot be known since only the possibility of transmission can be checked and the intensity of transmission of electric waves is not measured in this program. For this, intensity and the loss of electric waves were analyzed by concrete thickness using Spectrum Analyzer.

4.3. Test of Distance of Transmission and Reception in Free Space. First, the capability of module was experimented and the capability of transmission in free space (Figure 7) was experimented for the comparison of the value of the experiment result. The experiment was executed in a play yard of H University in Seoul at 2 p.m. when there would be the least humidity in the air to minimize the effect of external environment.

After fixing transmission module on the starting point, the distance of transmission was measured by gradually making the distance be farther.

4.3.1. The Result of Test of Distance of Transmission and Reception in Free Space. As a result, it was possible to transmit and receive up to the end of the play yard with the length of 200 m in free space. Since it is not possible to experiment further, this experiment assumed that it was possible to receive from the distance more than 200 m in free space. The result shows that it is possible to use sensors with exposure to the outside without embedding them into the structure.

As a result of the experiment on communication distance (Figure 8), the quality of communication was 86% in case of 120 m and 56% in case of 300 m, which means it is more than 50% from both 120 m and 300 m (Figure 9), and it was determined to be satisfactory to apply to civil and architectural structures [17].

4.4. Permeability Experiment in Vertical Direction. For the experiment of permeability in vertical direction, after putting the baseplate on the far bottom in the specially fabricated case (Figure 10), R31 and R15 that have the intensity of RF with 31 dB and 15 dB, respectively, were embedded.

For more accurate experiment, transmission module S31 that was specially fabricated to enable only each other communication was embedded more than once and the experiment was conducted. Unlike R31, R15 was transmitted through a computer. If transmission module sent a signal, the lamp itself of reception module would be lit and indicate that transmission and reception are in progress. In addition, the experiment was executed by setting equally RF intensity of both R31 and S31 as 31 dB and the comparison between R31 received through a computer program and S31. The module of transmission and reception, that was specially fabricated, made it possible to experiment with more accuracy.

The possible distance of reception with piling up concrete plates (Figure 11) was measured in this experiment.

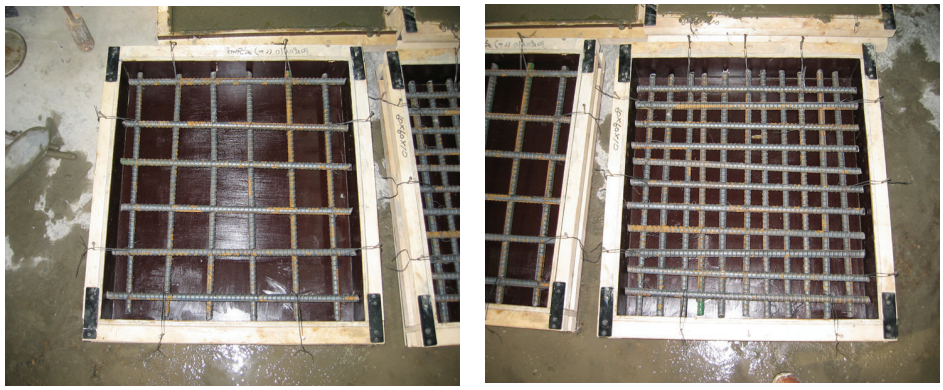


Figure 4: Placement of bars (D16@10, D16@5).

Figure 5: Concrete placement.

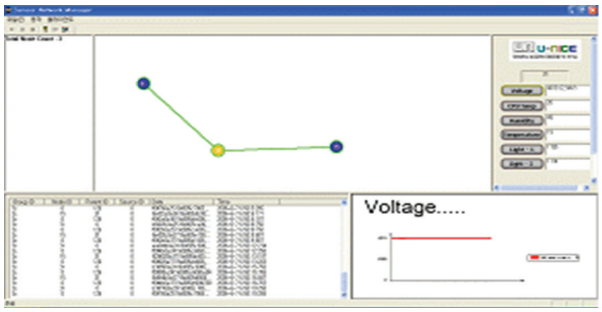


Figure 6: Monitoring system.

Figure 7: Free space transmission and receiving test.

4.5. Permeability Experiment in Horizontal Direction. In the same way with permeability experiment in vertical direction for the method of this experiment, the depth of permeability in horizontal direction was measured (Figure 12). The distance of transmission and reception was measured with the baseplate on the right side and piling up in horizontal direction on the left using variously fabricated concrete with thickness of 1, 3, 5, and 20 cm. The experiment was implemented by using the module of transmission and reception R15, R31, and S31 that were used in the experiment of permeability in vertical direction.

4.6. Test of Attenuation of Electric Waves. In the experiment of permeability in vertical and horizontal directions that was

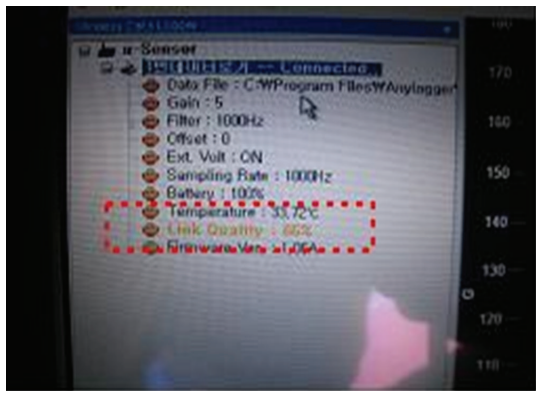


Figure 8: Measurement scene and results in the 120 m.

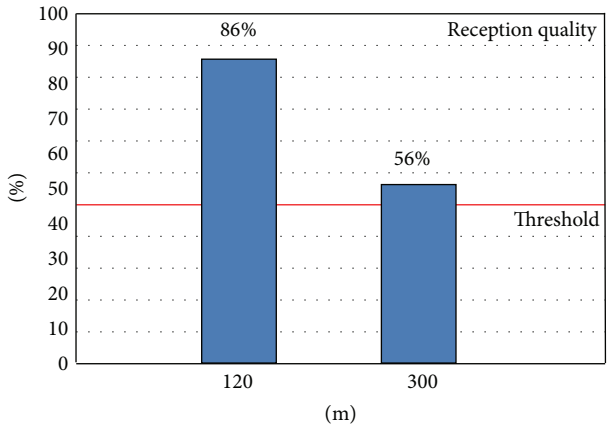


Figure 9: Communication results in 120 m and 300 m.

conducted above, since it was possible only to check the feasibility of transmission of wireless communication and the intensity of transmission of electric waves could not be checked, it was impossible to know the accurate intensity and loss of electric waves by the thickness of concrete. For this, the intensity and loss of electric waves by concrete thickness were measured and analyzed using Spectrum Analyzer (Figure 13). The object of this experiment was R31 that had the biggest intensity of RF. With the same way as the above, experiment spectrum in free space was measured to compare the value of the experiment. Furthermore, it was measured by every 10 cm in case of plain concrete and by changing the pitch of steel bars in reinforced concrete.

5. The Result and Analysis of the Experiment

5.1. Permeability Experiment in Vertical Direction. The result of permeability in vertical direction is shown in Table 1. In case of R15, it showed 39 cm for the depth of permeability in plain concrete and 31 cm for D16@5 which means that the pitch of steel bars is 5 cm, 34 cm for D16@7 which means that the pitch of steel bars is 7 cm, 36 cm for D16@10 which means that the pitch of steel bars is 10 cm, 38 cm for D16@15 which means that the pitch of steel bars is 15 cm in reinforced concrete. In case of R31, it penetrated up to 45 cm in plain concrete, 37 cm for D16@5, 40 cm for D16@7, 42 cm for D16@10, and 45 cm for D16@15. S31 showed the same results. As a result, when R15 and R31 were compared, it was suggested that the bigger the intensity of RF is, the more it penetrates.

Moreover, on comparing all of R15, R31, and S31 it penetrated better when the pitch of steel bars was 15 cm rather than 5 cm and it did the best in plain concrete. And Figure 13 shows that the depth of penetration increases linearly as the pitch of steel bars get bigger, which means wireless communication is affected by the pitch of steel bars. In addition, Figure 14 shows that when the pitch of steel bars is more than 15 cm the pattern of permeability is the same as plain concrete, which means when the pitch of steel bars is more than 15 cm, it is not affected by steel bars [18].

Table 1: The Result of experiment of permeability in vertical direction (unit: cm).

Classification	Concrete	D16@5	D16@7	D16@10	D16@15
R15	39	31	34	36	38
R31	45	37	40	42	45
S31	45	37	40	42	45

Table 2: The result of experiment of permeability in horizontal direction (unit: cm).

Classification	RF15	RF31	S31
Penetration depth	43	47	47

Table 3: Measurement of electric waves in plain concrete (unit: dBm).

Classification	10 cm	20 cm	30 cm	40 cm	50 cm	60 cm
R31	−47	−52	−64	−67	−71	−75

5.2. Permeability Experiment in Horizontal Direction. The result of permeability in horizontal direction is shown in Table 2. It was measured as 43 cm in R15, 47 cm in R31 and S31, and as the permeability experiment in vertical direction, R31 and S31 with big intensity showed better permeability. Besides, when compared with the permeability experiment in vertical direction, the permeability in horizontal direction was deeper. This suggests effects of the direction by radiation pattern of antenna attached to module and Chip antenna shows similar radiation pattern with monopole antenna that is an omnidirectional one (shown as Figure 15) and it is also similar with the radiation pattern of the existing small antenna. What is a solid line is the exact radiation pattern and from a view of this radiation pattern, the reason of resulting in minutely deeper penetration in horizontal direction than vertical direction is that when the permeability experiment in horizontal direction was conducted, the part of sensors was laid on the right side with piling up concrete on the left side, which was confirmed with about 5 ~ 8 variance when the radiation pattern from the left side was compared with upwards radiation pattern in radiation pattern (Figure 15) [19].

5.3. Analysis Using Spectrum Analyzer. As a measurement result of attenuation of electric waves using Spectrum Analyzer, it showed −27 dBm in free space. In addition, as shown in Table 3 as concrete is getting thicker the loss value of electric waves was definitely disclosed in plain concrete with −47 dBm in the thickness of 10 cm, 64 dBm in 30 cm, and −71 dBm in 50 cm. It was considered that the electric waves were scarcely transmitted as −75 dBm in concrete thickness of 60 cm. From the result of experiment it was suggested that wireless communication of module was possible up to 45 cm.

Besides, Table 4, showed −60 dBm in D16@5 that has the pitch of 5 cm, −56 dBm in D16@7, −51 dBm in D16@10, and −50 dBm in D16@15 in reinforced concrete. From these results, it could be seen that the closer the pitch of steel bars is the more attenuation of electric waves happens. Furthermore,

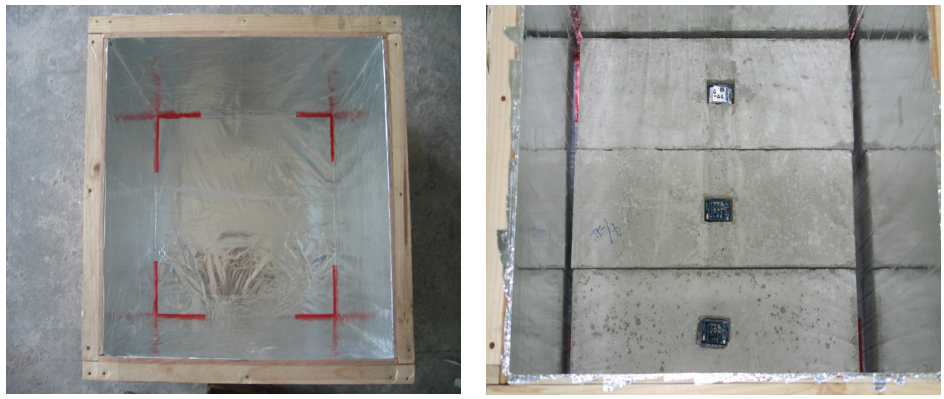


Figure 10: Left plates: Aluminum special case, right plates: base plate.

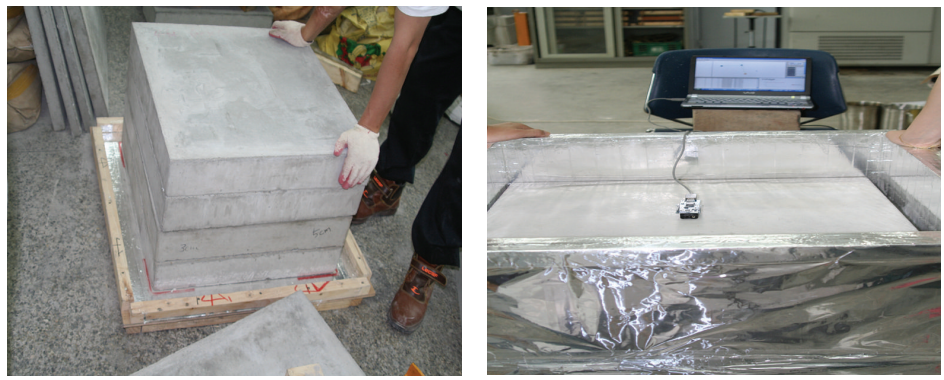


Figure 11: Permeability experiment in vertical direction.

Table 4: Measurement of intensity of electric waves in reinforced concrete (unit: dBm).

Classification	D16@5	D16@7	D16@10	D16@15
R31	−60	−56	−51	−50

there was −4 dBm attenuation of electric waves in the pitch between 7 cm and 10 cm. Finally they show that they are affected a little by steel bars up to 10 cm and from the fact that there is no big change in 15 cm, the Pitch of 15 cm affects penetration of electric waves as shown in the permeability experiment above (Figure 16).

5.4. Analysis by the Length of Wavelength of Frequency. The signal received in antenna with wireless communication module is generally affected by the length of wavelength of frequency for smooth communication. Frequency is how many oscillations for 1 second occur. As shown in Figure 17 if one oscillation happens per one second, it is called 1 Hz and one cycle is shown as ~, which is considered to oscillate one time. In addition, the wavelength means the distance of a single oscillation. In other words, the length of one cycle is wavelength [20]:

$$\text{Oscillation velocity of a single time (m/s)} = 3.0 * 10^8 \text{ (m/s)}/2.4 \text{ (number)} * 10^8 \text{ (number)} = 0.125 \text{ (m/s)}.$$

$$\text{Oscillation distance of a single time (m)} = 0.125 \text{ (m/s)} \times 1 \text{ (s)} = 0.125 \text{ (m)}.$$

The frequency band used in this experiment is 2.4 GHz that oscillates $2.4 * 10^8$ times for one second. In other words, the velocity of electric waves (the velocity of light) per one second should be divided by the times per one second 2.4 GHz oscillates to get the oscillation distance of one time. If the velocity of one time 2.4 GHz oscillation were calculated, oscillation distance of one time could be obtained by multiplying it by the time. The velocity of electric waves for one second is $3.0 * 10^8$ (m/s) and the oscillation frequency of frequency for one second is $2.4 * 10^8$. In other words, the oscillation velocity of one time is 0.125 m/s and the oscillation distance of one time is 12.5 cm by 0.125 m/s. Besides the absorption of electric waves of Chip antenna occurs in a conductor less than one wavelength (λ). Namely, the absorption of electric waves happens in the wavelength more than 12.5 cm. The result of experiment was analyzed by this.

In case that the pitch of steel bars both in perpendicular and diagonal directions is within λ as in Figure 18, since the maximum length that electric waves can go through becomes less than λ, steel bar which is a conductor absorbs electric waves [21].

Finally, it is determined that in reinforced concrete with the pitches of 5 cm and 7 cm that are less than 12.5 cm (λ), steel bars absorb so much electric waves of wireless

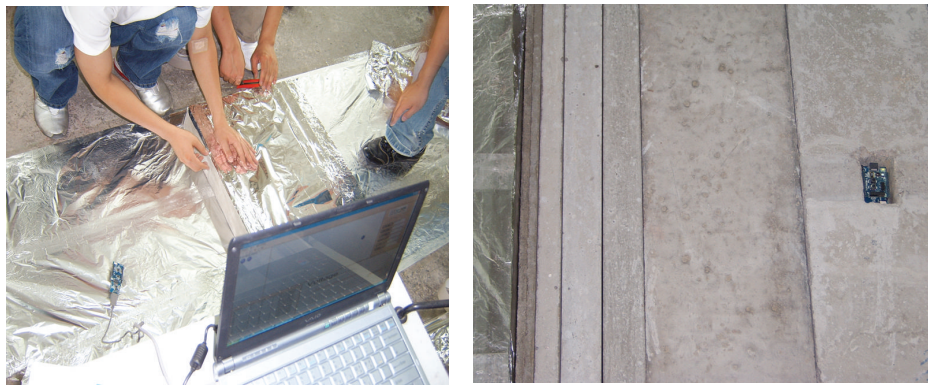


Figure 12: Permeability experiment in horizontal direction.

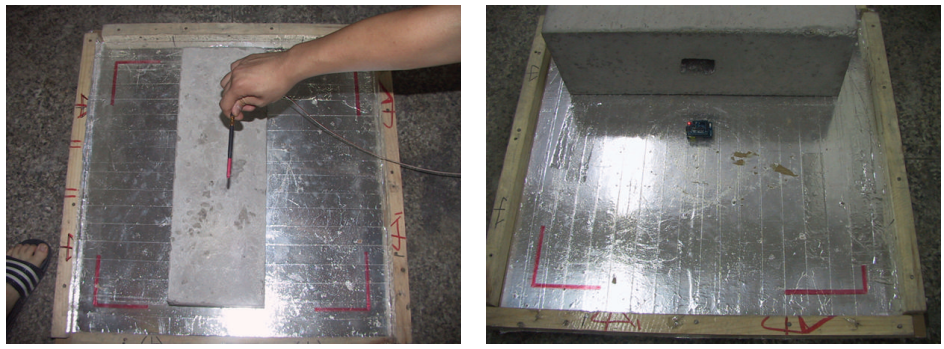


Figure 13: Experiment of measurement of electric waves attenuation.

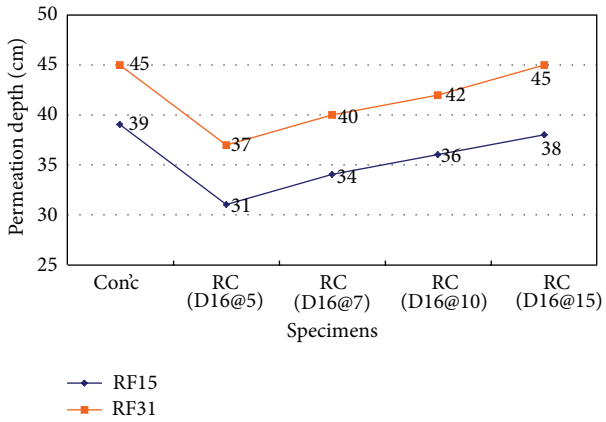


Figure 14: The result of experiment of permeability in vertical direction (unit: cm).

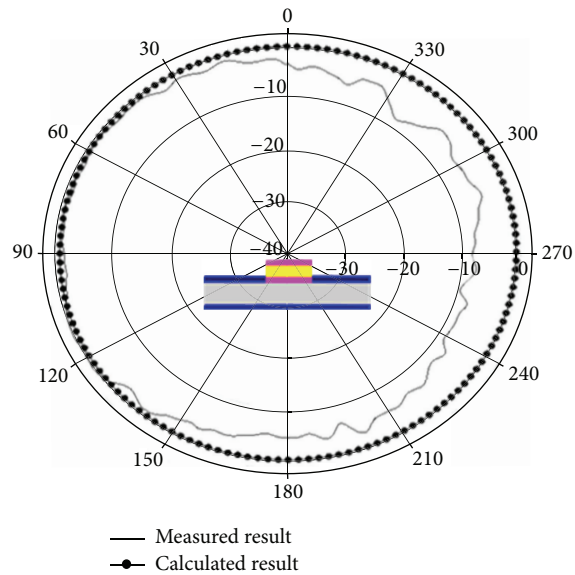


Figure 15: Radiation pattern of Chip antenna.

communication that takes place on communication barrier. Besides, reinforced concrete with the pitch of 10 cm is affected in perpendicular direction by the absorption of electric waves but in case of diagonal direction, the distance is 14.14 cm which is longer than 12.5 cm (λ) which showed that it was not affected. It is considered that reinforced concrete with the pitch of 15 cm is not affected by steel bars during wireless communication since it is more than 12.5 cm (λ) in both perpendicular and diagonal directions. These results agree with experiment of permeability depth and analysis results of attenuation of electric waves of Spectrum Analyzer.

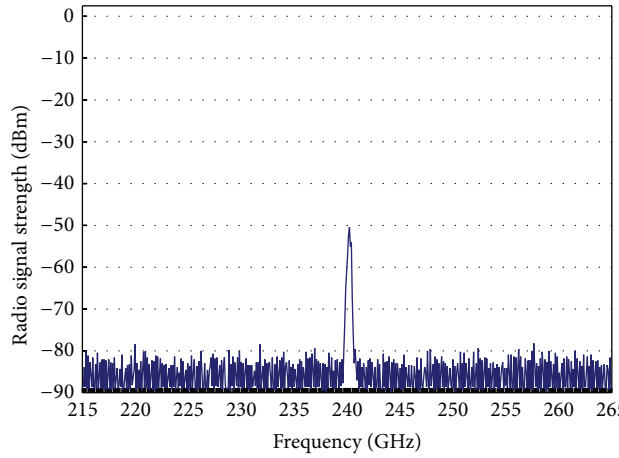


Figure 16: Spectrum of concrete with the thickness of 20 cm.

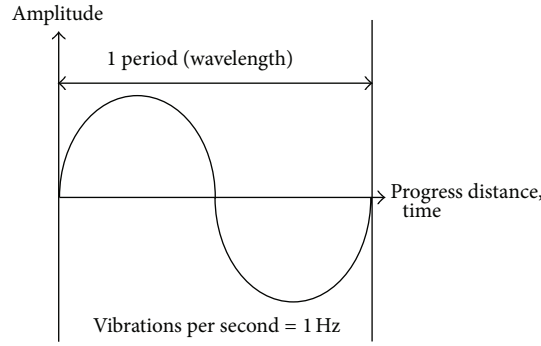
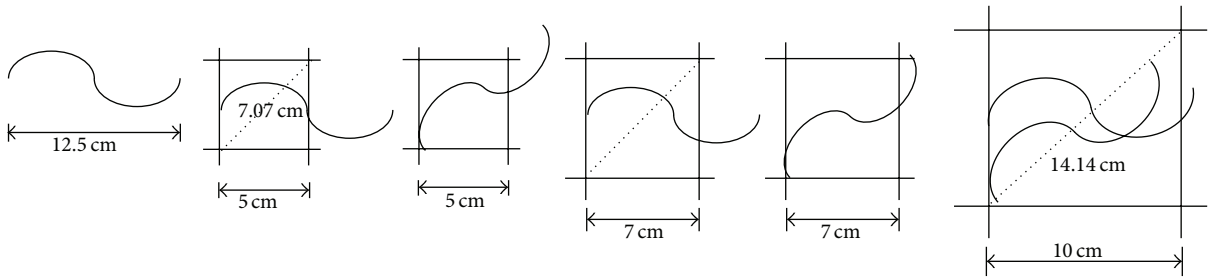


Figure 17: The concept of frequency.



5 cm: $\sqrt{(5^2 + 5^2)} = 7.07 < 12.5$ cm

7 cm: $\sqrt{(7^2 + 7^2)} = 9.89 < 12.5$ cm

10 cm: $\sqrt{(10^2 + 10^2)} = 14.14 > 12.5$ cm

15 cm: $\sqrt{(15^2 + 15^2)} = 21.21 > 12.5$ cm

Figure 18: The analysis of wavelength of frequency.

6. Conclusion

The test of capability of wireless communication of sensor node embedded in reinforced concrete structure was conducted to apply new Ubiquitous Computing techniques and networking techniques to construction industry in this experiment and the results are as below.

(1) When the test of transmission and reception in free space was implemented using transmission and reception module of wireless communication it showed more than 200 m for the possible distance of wireless communication. This shows that wireless communication module can be used with attachment to bridges or the outside structure.

(2) As a result of measurement of distance of wireless communication in vertical and horizontal directions in plain concrete it penetrated 45 cm in vertical direction and 47 cm in horizontal direction. These results seem to be possibly criteria when wireless communication module is embedded in civil and architectural structures.

(3) The fact that the horizontal permeability was deeper in permeability experiment of plain concrete in vertical/horizontal direction suggests that it penetrates better since the pattern of antenna shows higher value in the horizontal direction than in the vertical direction with influence of direction of antenna pattern of module used in this experiment.

(4) As a result of experiment of attenuation of electric waves in wireless communication through spectrum of reinforced concrete when the pitch of steel bars was 5 cm, it showed −60 dBm, −51 dBm for 10 cm, and −50 dBm for 15 cm. It showed attenuation of electric waves by 1 dBm between the pitches of 15 cm and 10 cm, 4 dBm between 10 cm and 7 cm, and 4 dBm between 7 cm and 5 cm. Finally, the narrower the pitch of steel bars was, the more attenuation of electric bars occurred, and this result indicated that steel bars affect wireless communication.

(5) It was considered that the phenomenon of attenuation did not occur since attenuation of −1 dBm between the pitches of 10 cm and 15 cm of steel bars was a delicate difference. Despite this result, when wireless is applied to inside structure, if the pitch of steel bars is 10 cm, it is not affected by steel bars.

(6) Since the absorption of electric waves of Chip antenna with 2.4 GHz frequency occurs in a conductor with wavelength more than 12.5 cm (λ), steel bars in reinforced concrete with the pitches of 5 cm and 7 cm absorb much electric waves in wireless communication, which impedes communication, and in case of 10 cm, the absorption of electric waves occurs in perpendicular direction not in diagonal direction. Besides, reinforced concrete with the pitch of 15 cm has more than 12.5 cm (λ), it is considered that it is not affected by steel bars in wireless communication.

(7) The epitome of the above results is that it is possible to apply ubiquitous Active RF wireless transmission and reception to civil and architectural structures made of plain concrete and reinforced concrete.

References

[1] B. W. Jo, J. H. Park, and B. C. Shin, "Design of intelligent bridge monitoring system for ubiquitous," in *Proceedings of the 34th Conference of the Korean Society of Civil Engineers regular CIVIL EXPO*, pp. 2513–2516, 2008.

[2] M. E. A. Bashir, H. S. Shon, D. G. Lee, H. s. Kim, and K. H. Ryu, "Real-time automated cardiac health monitoring by combination of active learning and adaptive feature selection," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 7, no. 1, pp. 99–118, 2013.

[3] K. J. Jo and H. J. Lee, "Development of real time monitoring system based on context-awareness for wireless sensor networks," *Journal of the Korea Contents Association*, vol. 11, no. 4, pp. 101–111, 2011.

[4] D. K. Hwang and S. H. Kim, "Design of bridge monitoring system based on sensor network," in *Proceedings of the KFIS Autumn Conference*, vol. 16, pp. 332–335.

[5] B. W. Jo, J. H. Park, K. W. Yoon, Y. K. Kim, and Y. G. Kim, "Study on developing program for ubiquitous bridge monitoring system," in *Proceedings of the 36th Conference of the Korean Society of Civil Engineers regular*, pp. 957–996, 2010.

[6] J. S. Kim and T. H. Shin, "An application strategy of EM simulation technology for optimizing readability of RFID/USN," *Architectural Institute of Korea Proceedings*, vol. 29, no. 1, pp. 685–688, 2009.

[7] J. H. You, A. H. Joon, and H. O. Kim, "Development of pipe management system using RFID," in *Proceedings of the 34th Conference of the Korean Society of Civil Engineers Regular*, pp. 3742–3745, 2008.

[8] S. W. Hong and Y. S. Jo, "Study of the structural internal assessment of concrete slab using the ground penetrating radar exploration," *Journal of the Architectural Institute of Korea*, pp. 51–58, 2007.

[9] J. V. R. Rodriguez, "Radiowave Propagation through Window Panes and Reinforced Concrete Structures—An application of Frequency Selective Surface techniques," Department of Electroscience, Lund University, July 2001.

[10] J. K. Lee, Y. H. Lee, P. J. Hyun, and M. J. Son, "Experimental study on wall transmission loss of electric wave for the RTLS application of building construction project," *Korea Institute of Building Construction*, no. 1, Chapter 9, pp. 95–101, 2009.

[11] C. D. An, J. H. Lee, H. S. Kim, C. S. Lee, and G. H. Shin, "Study on RF part of spectrum analyzer," in *Proceedings of the General Conference of Korea Information and Communications Society (Summer)*, pp. 2327–2329, 2008.

[12] L. C. Lee and K. S. Min, "Wide bandwidth RFID tag antenna design for protection of connection part between chip and antenna," *Journal of the Korea Electromagnetic Engineering Society*, vol. 20, no. 2, pp. 154–160, 2009.

[13] N. S. Soo and L. H. Min, "Multi-band chip slot antenna for mobile devices," *Journal of the Korea Electromagnetic Engineering Society*, vol. 20, no. 12, pp. 1264–1271, 2009.

[14] J. H. An, S. I. Jun, and S. A. Kwan, "Experimental study on application of KS F, 2456 using shear wave," *Journal of the Korea Society of Road*, vol. 14, no. 5, pp. 57–65, 2012.

[15] Korea, "The Monthly technology & standards," Technical Standards 89, 50-51, 2009.

[16] J. H. Park and B. W. Jo, "Development of ubiquitous median barrier system in the highway," *Journal of the Korean Society of Civil Engineers D*, vol. 29, no. 4, pp. 499–507, 2009.

[17] S. K. Song, *Bridge Safety Management with ubiquitous sensor node based on intelligent research [M.S. thesis]*, Hanyang University, 2007.

[18] N. S. Soo and L. H. Min, "Multi-band chip slot antenna for mobile devices," *Journal of the Korea Electromagnetic Engineering Society*, vol. 20, no. 12, pp. 1264–1271, 2009.

[19] J. H. Jang, *Study on USN sensor node efficient inspection and intelligence for Ubiquitous [M.S. thesis]*, Hanyang University, 2007.

[20] D. H. Kim, "Sub-wavelength focusing lens composed of a dual-plate metamaterial providing a negative refractive index,"

Journal of the Korea Electromagnetic Engineering Society, vol. 12, no. 1, pp. 26–31, 2012.

[21] M. J. Kim and H. S. Yoon, "Hydrodynamic characteristics for flow around wavy wings with different wave lengths," *International Journal of Naval Architecture and Ocean Engineering*, vol. 4, no. 4, pp. 447–459, 2012.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 217876, 14 pages
http://dx.doi.org/10.1155/2013/217876

Research Article

Hilbert-Curve Based Data Aggregation Scheme to Enforce Data Privacy and Data Integrity for Wireless Sensor Networks

Yong-Ki Kim,[1] Hyunjo Lee,[2] Min Yoon,[2] and Jae-Woo Chang[2]

[1] *Korea Institute of Science and Technology Information, Daejeon 305-806, Republic of Korea*
[2] *Department of Computer Engineering, Chonbuk National University, Chonju, Chonbuk 561-756, Republic of Korea*

Correspondence should be addressed to Jae-Woo Chang; jwchang@jbnu.ac.kr

Received 6 November 2012; Revised 25 March 2013; Accepted 16 April 2013

Academic Editor: Chao Song

Data aggregation techniques have been proposed for wireless sensor networks (WSNs) to address the problems presented by the limited resources of sensor nodes. The provision of efficient data aggregation to preserve data privacy is a challenging issue in WSNs. Some existing data aggregation methods for preserving data privacy are CPDA, SMART, the Twin-Key based method, and GP2S. These methods, however, have two limitations. First, the communication cost for network construction is considerably high. Second, they do not support data integrity. There are two methods for supporting data integrity, iCPDA and iPDA. But they have high communication cost due to additional integrity checking messages. To resolve this problem, we propose a novel Hilbert-curve based data aggregation scheme that enforces data privacy and data integrity for WSNs. To minimize communication cost, we utilize a tree-based network structure for constructing networks and aggregating data. To preserve data privacy, we make use of both a seed exchange algorithm and Hilbert-curve based data encryption. To support data integrity, we use an integrity checking algorithm based on the PIR technique by directly communicating between parent and child nodes. Finally, through a performance analysis, we show that our scheme outperforms the existing methods in terms of both energy efficiency and privacy preservation.

1. Introduction

With the proliferation of advanced technologies of mobile devices and wireless communication, wireless sensor networks (WSNs) are increasingly attracting interest from both industry and research institutes [1–3]. Because sensor nodes have limited resources (i.e., battery and memory capacity), data aggregation techniques have been proposed for WSNs [4–9]. However, the wireless communication can be overheard, and consequently data privacy in sensor networks is a crucial issue. Although data aggregation schemes that preserve data privacy have been proposed, they have the following limitations. First, the communication cost for network construction and data aggregation is considerably expensive. Second, the existing schemes do not support data integrity due to communication loss. Since the existing privacy-preserving schemes do not support privacy preservation and integrity protection simultaneously, it is necessary to carefully design an effective data aggregation scheme for recent applications of WSNs, such as military and environmental monitoring, where both privacy and integrity of the sensed data should be provided [10].

To resolve these problems, we propose a new energy efficient and privacy preserving data aggregation scheme in WSNs. To reduce the communication cost for preserving data privacy, we propose a seed exchanging algorithm for data aggregation. The seed generated by this algorithm is used not only to conceal the sensed data but also to preserve data privacy without additional message exchanges during the data aggregation step. For data privacy preservation, we also utilize a Hilbert-curve based technique, where it is difficult to obtain the actual sensed data, even if attackers try to overhear it, because the data being sent can be changed by using a unique Hilbert value. For providing data integrity, we propose an integrity checking algorithm based on a private information retrieval (PIR) technique. Upon receiving aggregated data from child nodes, a parent node starts an integrity checking algorithm in which the parent node generates a message based on the PIR technique by multiplying two large prime numbers. By sending a PIR

message to child nodes, the parent node can verify the aggregated data. Our integrity checking algorithm is more efficient than the existing schemes since it checks the data integrity between child and parent nodes, instead of checking all data during the communication. Therefore, our scheme requires low communication cost and yields an accurate aggregate result even in reasonably dense networks.

This paper is organized as follows. In Section 2, we present related work on privacy preserving aggregation schemes in WSNs. In Section 3, we provide both considerations and attack models for designing an efficient privacy preserving aggregation scheme. In Section 4, we propose a new privacy preserving data aggregation scheme including a seed exchange algorithm in WSNs. In Section 5, we present a performance analysis of our scheme. Finally, we draw conclusions and suggest future work in Section 6.

2. Related Work

In this section, we present the existing data aggregation schemes for supporting data privacy and data integrity in WSNs. Privacy preserving data aggregation schemes include CPDA, SMART, Twin-key, and GP2S. He et al. [11] proposed a Cluster Based Private Data Aggregation (CPDA) method in which a cluster header aggregates data from cluster members. The CPDA method first constructs clusters to perform intermediate aggregations. All nodes include a head node within a cluster and then share M public seeds, where M is the number of cluster members. Next, each node generates $M - 1$ private seeds and sends M messages generated by using the public and private seeds together with sensed data. Finally, the cluster head calculates their aggregate value by using its own private numbers and received information. However, the CPDA method has high communication cost to perform data aggregation. He et al. [11] also proposed a Slice Mix AggRegaTe (SMART) method to achieve private data preservation by using a data slicing technique. For this, each node randomly selects a set of nodes within h hops and slices its private data into J pieces randomly. One of the sliced data is kept on the node which sensed the data, and the remaining $J - 1$ pieces are encrypted and sent to pre-selected nodes. When a node receives the sliced data from neighbors, it aggregates the received data and sends the result to the sink node. The SMART method also suffers from high communication cost, however, because each node should share its divided data among neighboring nodes. Conti et al. [12] proposed a keys-based private data preservation method called Twin-key. Because the Twin-key scheme can prevent leakage of the sensed data during the data aggregation process, it is robust to data loss. For providing the robustness to data loss, they set up Twin-key, during constructing clusters [13], where two neighboring nodes share at least one common key corresponding to a hash value. Data aggregation is thus performed twice along with the Hamiltonian circuit in which each node adds its sensed value to the partial aggregate value. At the same time, for each live twin-key it adds or removes a corresponding shadow value in accordance with the live announcement. As a result, each cluster head obtains the correct aggregate for the cluster. The cluster head then passes the aggregated value to the sink node by following a tree aggregation structure. However, the Twin-key method has high communication cost due to the process of live announcement and data aggregation. Finally, Zhang et al. [14] proposed the Generic Privacy Preservation Scheme (GP2S) for perturbed histogram-based aggregation. This scheme supports data aggregation for a variety of queries since it provides both individual data and aggregated data. For this, each sensor node is preloaded with a secure one-way hash function that maps a bit string to a value between 0 and $N - 1$, where N is a system parameter. A sink node then sends out a query message with a threshold σ (i.e., data duration). After receiving the query, each sensor node sends its data composed of a hash function. If the sink node receives aggregated data from all child nodes, it determines the distribution of sensed data readings. However, the accuracy of the aggregated value of the network data is low and the data privacy can be broken by a data aggregator (parent node) having leaf nodes.

He et al. [15] also proposed iPDA and iCPDA schemes to support integrity checking in WSNs, by extending their previous schemes, SMART and CPDA, respectively. To the best of our knowledge, the schemes are the first to address both privacy preservation and integrity protection for data aggregation in WSNs. The iPDA scheme utilizes the data slicing and assembling technique of SMART to preserve data privacy. It protects data integrity by utilizing a node disjoint between two aggregation trees rooted at the query server, where each node belongs to a single aggregation tree. When the aggregated data from both aggregation trees are compared, the query server accepts the aggregate result if the difference in the aggregated data from the two aggregation trees does not deviate from the predefined threshold value. Otherwise, it ignores the aggregated result by considering it as polluted data. However, iPDA has some shortcomings. First, it is impractical to compare aggregated values of two node-disjoint aggregation trees, because it cannot be expected that all nodes will reply to all requests, due to the unreliability of a WSN. Second, for a secure communication channel from adversaries, all sensor nodes use secret keys to encrypt their all data slices before sending it to their $2*(L - 1)$ sensor nodes. Every sensor node thus has computational overhead from decrypting all the slices before aggregating them. Because encryption/decryption is an expensive operation for resources-constrained sensor nodes, iPDA has high computation cost. Third, the technique for slicing and assembling is only operable while the collusion of sensor nodes is up to a certain threshold (i.e., the sum of out-degree and in-degree minus one). If the number of colluding sensor nodes exceeds the threshold, the sensor nodes may collaboratively reveal private information of other nodes. Although the threshold can be raised by increasing the number of slices, this will further increase communication overhead. Finally, since each sensor node has to transmit five to six messages on average, the iPDA scheme has high data propagation delay. Meanwhile, iCPDA requires three rounds of interactions. In this scheme, each node first sends a seed to other cluster members. Next, each node hides its

sensed data via the received seeds and sends the hidden sensed data to each cluster member. Each node then adds its own hidden data to the received data, and it sends the calculated results to its cluster head. To enforce data integrity, cluster members check the transmitted aggregated data of the cluster head. However, iCPDA has some disadvantages. First, its communication overhead increases significantly with respect to the cluster size. Second, its computational overhead increases rapidly with an increase of the cluster size, whereas a decrease of cluster size introduces lower privacy-preserving efficacy. Finally, iCPDA has high data propagation delay due to its three rounds of interactions.

It is thus necessary to design a new data aggregation scheme that supports both data privacy and data integrity. The new scheme should be reliable and efficient in terms of energy consumption, propagation delay, and the accuracy of the aggregated result.

3. Design Considerations

In this section, we present requirements for a data aggregation scheme to support both data privacy and data integrity. The desired data aggregation scheme should satisfy the following criteria.

(1) Data privacy: privacy concern is one of the major obstacles to civilian applications for wireless sensor networks. Curious individuals may attempt to gather more detailed information by eavesdropping on the communications of their neighbors. It is increasingly important to develop data aggregation schemes to ensure data privacy against eavesdropping.

(2) Data integrity: since data aggregate results may be used to make critical decisions, a base station needs to guarantee the integrity of the aggregated result before accepting it. Therefore, it is crucial that data aggregation schemes can protect the aggregated results from being polluted by attackers.

(3) Efficiency: data aggregation achieves bandwidth efficiency through in-network processing. In integrity-protecting private data aggregation schemes, additional communication overhead is unavoidable to achieve additional features. However, the additional overhead must be kept as small as possible.

(4) Accuracy: an accurate aggregate result of sensed data is generally desired. Therefore, we should take accuracy as a criterion to evaluate the performance of integrity-protecting private data aggregation schemes. When accurate aggregate results are needed, schemes based on randomization techniques are not applicable.

On the other hand, there exist multiple potential attacks against a data aggregation scheme. Some attacks aim at disrupting the normal operation of the sensor network, such as routing attacks and denial of service (DoS) attacks. A number of previous efforts have addressed these behavior-based attacks. In this paper, our major concern is the types of attacks that try to break the privacy and/or integrity of aggregate results, rather than worrying about those attacks. We assume that a small portion of sensor nodes can be compromised and focus on the defense of the following categories of attacks in wireless sensor networks.

(1) Eavesdropping: in an eavesdropping attack, an attacker attempts to obtain private information by overhearing transmissions over its neighboring wireless links or colluding with other nodes to uncover the private information of a certain node. Eavesdropping threatens the privacy of data held by individual nodes.

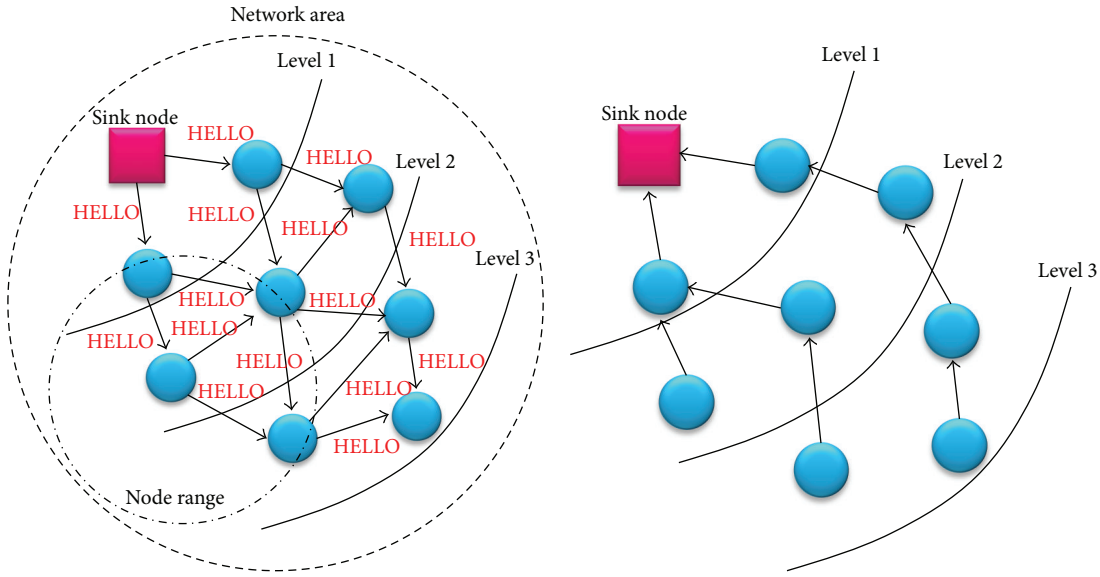
(2) Data pollution: in a data pollution attack, an attacker tampers with the intermediate aggregate result at an aggregation node. The purpose of the attack is to make the base station receive a wrong aggregate result with large deviation from the original result, which leads to improper or wrong decisions. In this paper, we do not consider the attack where a node reports a false reading value, and we assume that the impact of such an attack is usually limited. By using privacy preservation measures, individual sensory data are hidden. However, not only the sensory data but also the aggregated value of a small group of sensors must be in a reasonable range. This implies that if a malicious user pollutes the individual sensory data (at a lower level in the aggregation tree), it can be easily detected since this introduces a large deviation from the original data. Therefore, a more serious concern is the case where an aggregator close to the root of the aggregation tree is malicious or compromised.

In this paper, our goal is to design a reliable and efficient data aggregation scheme in terms of energy consumption, propagation delay, and accuracy of the aggregated result by following these design considerations.

4. Data Aggregation Scheme to Enforce Data Privacy and Data Integrity

In this section, we present a novel Hilbert-curve based data aggregation scheme that supports both privacy preservation and integrity protection for wireless sensor networks. In order to support data privacy, we first provide a data privacy preserving algorithm by using sensor nodes' seeds and Hilbert-curve values. A seed exchange algorithm is applied to reduce the number of messages during data aggregation. In order to support data integrity, we provide a private information retrieval (PIR) [16–19] based integrity checking algorithm that communicates between a child node and its parent node by exchanging a PIR message and its response message.

4.1. Privacy Preserving Algorithm. For wireless sensor networks, we provide a novel privacy preserving algorithm by using a Hilbert-curve technique [20] and seed exchanges among sensor nodes. Our privacy preserving algorithm is performed through three phases: a network construction phase, a data encryption phase, and a data transmission



(a) Message flooding (b) Network construction

Figure 1: Network construction.

phase. In the network construction phase, each node determines its sibling nodes, parent node, and child nodes by sending broadcast messages. Each node exchanges a seed to other nodes among its sibling nodes. In the data encryption phase, each node changes the sensed data into a value by using its generated seed and the received seeds. The changed value is encrypted by the Hilbert-curve algorithm. Finally, in the data transmission phase, each sensor node sends the aggregated data to a parent node where all the data from child nodes are merged with its encrypted data. A sink node aggregates all data of sensor nodes in the network. We explain each step in detail in the following.

4.1.1. Network Construction Phase. Our privacy preserving algorithm chooses a tree-based topology to perform intermediate aggregations. Note that we do not use a clustering-based topology because it is affected by the communication range between cluster heads and it suffers from a large amount of messages for constructing network. First, a sink node triggers a query by sending a HELLO message generated from a message flooding scheme [21], as shown in Figure 1(a). Upon receiving the HELLO message, a sensor node determines whether the HELLO message is from the sink node or not. If a sink node is located within its communication range, the sensor node receives the HELLO message from the sink node and sets the sink node as a parent node. Otherwise, the sensor node waits for a certain period of time to receive the HELLO message from its sibling nodes and then selects one of the sibling nodes as a parent node by broadcasting a JOIN message. The sink node forwards the HELLO message to its sibling nodes with its corresponding level (Figure 1(b)).

In this procedure, we set the maximum number of child nodes so as to avoid network imbalance. If the network has imbalance, the sensor node of the imbalanced area may consume more energy than the other areas. Therefore, we define the maximum number of child nodes as given below.

Definition 1. Let the Error Rate be the average rate of message loss from a sensor node, and let a weight (α) be a value for the density of a sensor network. The maximum number of child nodes C is defined by the following equation, where Network Area is the size of the network and Communication Range is the communication boundary reachable from a sensor node

$$\begin{aligned} &\mathrm{MIN}\Big(\#\text{ of neighbors},\\ &\quad\Big\lceil(1+\alpha)\times(1+\text{Error Rate})^2\\ &\qquad\times\frac{\pi\times(\text{Communication Range})^2}{\text{Network Area}}\\ &\qquad\times\#\text{ of nodes}\Big\rceil\Big). \end{aligned} \tag{1}$$

Figure 1 shows an example of our network construction algorithm and Algorithm 1 describes it. In algorithm 1, first, a sink node floods a HELLO message to the nearest node within its communication range (lines 1~2). The node that receives the HELLO message from the sink node sets its own level and broadcasts the HELLO message to other nodes (lines 3~12). If a node receives a JOIN message, it sets the node sending the JOIN message as a parent node. A parent node with the maximum number of child nodes sends to the child nodes a RESET message informing that they are allowed to link another node as a parent (lines 13~18).

4.1.2. Data Encryption Phase. After constructing a sensor network, each node generates random seed data for seed exchange. For this, we utilize an elliptic-curve key exchange

```
command NetworkConstruction (Message msg, MegTypemsgType)
  (1) If (A node is Sink node) {
  (2)      Flooding(initLevel, base_stationID);exit;}
  (3) Wait until receiving HELLO message;
  (4) If (a node receives message from a sensor node)
  (5)    If (msgTypeis HELLO) {
  (6)        Set parentID, recHopCnt, recLevel from message;
  (7)        NetInfo.curEntry++;
  (8)        If (curHopCnt > recHopCnt + 1)curHopCnt = recHopCnt + 1;
  (9)        else break;
  (10)       If (TOS_LOCAL_ADDRESS is not leaf node)
  (11)           Flooding(currentLevel, currentNodeID);
  (12)   If (msgTypeis JOIN)
  (13)       If (parent node does not exceed the maximum number of child node)
  (14)           NetInfo.Parent = parentID;
  (15)   }
  (16) KeySeed = GenerateSeedkey();
  (17) PairNode = chooseneighbor(random);
  (18) Send (PairNode, KeySeed) to PairNode;
  (19) Wait until response message from PairNode;
  (20) Seed = ComputeKey (Received_KeySeed from PairNode, KeySeed);
End Algorithm
```

Algorithm 1: Processing of network construction phase.

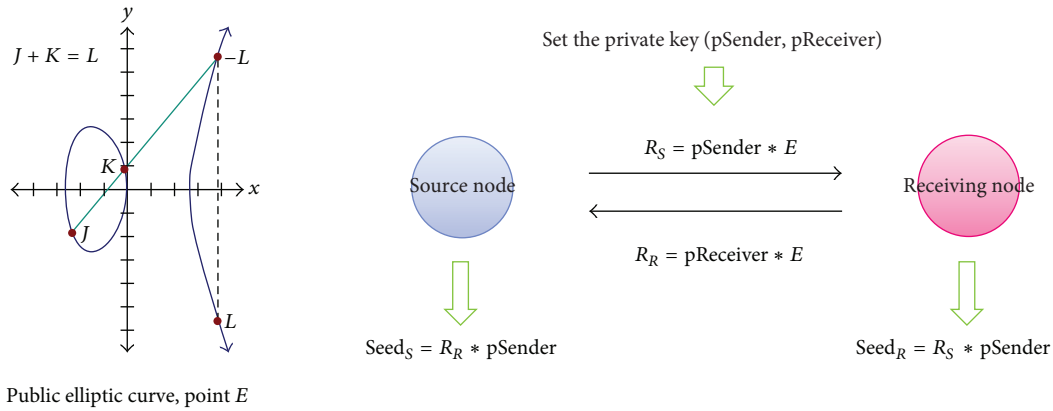


Figure 2: The overall flow of the elliptic key exchange algorithm.

algorithm that exchanges its own data by using a public elliptic curve, an arbitrary point, and its secret constant key. Figure 2 shows the flow of the elliptic key exchange algorithm. First, a source node and its neighboring node (receiving node) set a private constant key, for example, pSender and pReceiver. Second, each node makes a result R by multiplying an arbitrary point (E) and the private constant key having a public elliptic curve. Third, each node transmits the result R to the neighboring node. Finally, it calculates the seed data by multiplying R with its private constant key. The seed data are the sum of x-coordinate and y-coordinate, because the elliptic curve is a 2-dimensional equation. Because the elliptic key exchange algorithm allows each node to communicate without unnecessary messages, its own data can be protected from an attacker during communication.

The seed is used for hiding the original data from an adversary. The principle underlying our seed exchange method is as follows. The original data can be changed by extracting some part of a seed value, which is sent to other nodes. Some part of the seed value is also added from another node. As a consequence, the sensed data can be hidden among seed exchange group members. The following equation shows the final sending value from each node for data aggregation, where m is the number of seeds received

```
command EncryptData (Message msg, MegTypemsgType){
  (1)  directionValue = makeDirection()
  (2)  setCurveLevel = currentCurveLevel(Data)
  (3)  encryptedData = HilbertCurve(direction, curveLevel, NewValue)
  (4)  packing(encData)
  (5)  }
End Algorithm
```

Algorithm 2: Data encryption algorithm.

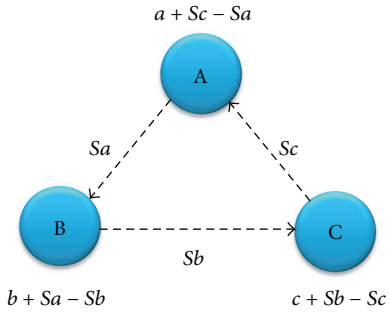


Figure 3: Original data change by seed exchange from three nodes.

from other nodes. Figure 3 shows a sensed data encryption result on each sensor node after exchanging a seed:

$$\text{processed value} = \text{original value} - \text{generated seed} + \sum_{i=1}^{m} \text{received seed}\,(i)\,. \tag{2}$$

To process a user's query, a parent node aggregates its changed data and all data received from its child nodes. Next, the parent node transforms the aggregated result into two-dimensional encrypted data by using the Hilbert curve [15]. The Hilbert curve, which was proposed by *G*. Peano, transforms *N*-dimensional data into 1-dimensional data. The Hilbert curve is a continuous fractal space-filling curve that gives a mapping between 1D and 2D space to preserve locality. The coordinates of a point (x, y), that is, projected to the unit square can be changed into a distance value from the start point to this point. To adapt the Hilbert curve to our algorithm, we assume that each sensor node transforms the one-dimensional sensed value into two-dimensional data. Here, the one-dimensional value is the aggregated value after applying the seed exchange algorithm for each node group. The two-dimensional data are the coordinate of the aggregated value along with the Hilbert curve in $2^n \times 2^n$ metrics. For this, we set as keys both the level l and the direction d of the Hilbert curve. We can encrypt the aggregated data using two-dimensional data (x, y) into a tuple of $\langle \text{key}(d, l), x, y\rangle$, where l is a level and d is the direction. For example, the aggregated value 4 of node 8 can be encrypted into $\langle \text{key}(\text{Bottom}, 2), 1, 1\rangle$ since its transformed value, level, and direction are (1, 1), 2, and Bottom, respectively in Figure 4. In a node 5, it receives encrypted data $\langle \text{key}(B, 2), 1, 1\rangle$ and $\langle \text{key}(T, 2), 3, 2\rangle$ from its child nodes 8 and 9, respectively. The encrypted data from child nodes should be changed into $\langle \text{key}(R, 2), 2, 1\rangle$ and $\langle \text{key}(R, 2), 2, 0\rangle$ by following the curve direction and the level of the node 5. Then, node 5 aggregates their data and sends aggregated data $\langle \text{key}(R, 2), 3, 2\rangle$ to the parent node. Algorithm 2 describes our data encryption algorithm. First, each node generates a Hilbert curve direction and a level based on the data (lines 1~2). Next, each node encrypts the data by the Hilbert curve (line 3). Finally, each node packs the encrypted data for sending the aggregated data to its parent node (line 4).

4.1.3. Data Transmission Phase. In the data transmission phase, each node sends the encrypted data to its parent node. The parent node then analyzes the encrypted data (e.g., key, curve direction, and curve level), that is received from the child node. If the curve direction and level of its child node are different from its own curve direction and level, the node should transform the received value based on its curve direction and level. In this way, a sink node aggregates all of the encrypted data from the hierarchy of nodes. To avoid communication loss of wireless sensor networks, we utilize a Time Division Multiple Access (TDMA) method [22] for data transmission. Definition 2 explains the principle to decide the start time of data transmission. Each child node sends the encrypted data at its own transmission time. Algorithm 3 shows our data aggregation algorithm. We start data aggregation from leafNode (lines 1~2). For aggregation, an intermediate node (InternalNode) can receive the data from its child node and reencrypt the data with its own data (lines 3~11). In this way, all encrypted data of sensor nodes reach a sink node. Finally, the sink node sends the aggregated data to the service client (lines 12~15).

Definition 2. Assume that child nodes are $N1, N2, \ldots, NC$, where the number of child nodes is C and the start time of the data transmission, that is, StartTime, for the ith sensor node Ni is determined as

$$\begin{aligned}\text{Start}_{\text{Time}}\ (Ni) &= (i-1) \times \big((\text{Life time of send section} \\ &\quad + \text{Life time of reception section}) \\ &\quad \times (\text{Life time of a user query})^{-1} \big).\end{aligned} \tag{3}$$

```
command Data Aggregation (Message msg, msgTypemsgType)
(1)   If (a node is leafNode)
(2)      Send Message(encData) to ParentNode
(3)   Else{
(4)      If (a node receive message(encData) from sensor node {
(5)         If (the node is InternalNode){
(6)            StoresencDatafrom msg;
(7)            decryptedData = decryption(encData)'
(8)            aggregatedData += decryptedData;
(9)            newEncData = HilbertCurve(direction, curveLevel, aggregatedData);
(10)           If (all data is received from childNode)
(11)              SendMessage(encData);to ParentNode;
(12)        }If (a node is SinkNode)
(13)           StoreencData from msg;
(14)           decryptedData = decryption(encData);
(15)           Send Message(decryptedData) to User;}}
End Algorithm
```

Algorithm 3: Data aggregation algorithm.

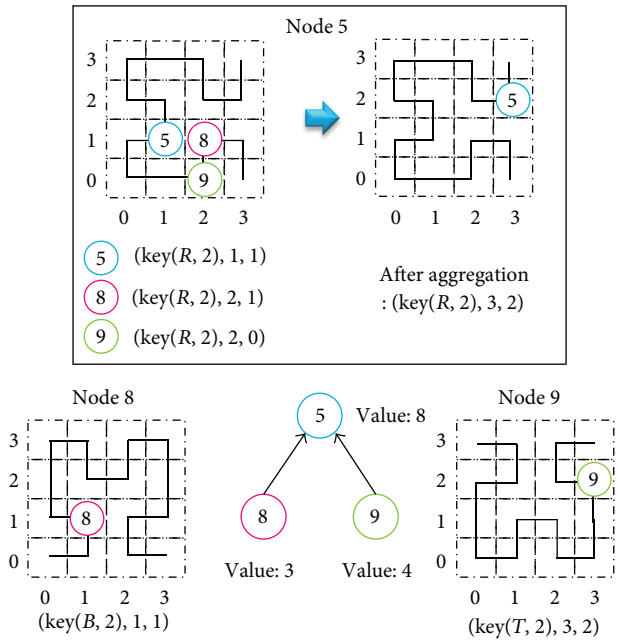


Figure 4: Example of data encryption.

4.2. Integrity Checking Algorithm. Our integrity checking algorithm is performed through three phases: a PIR message construction phase, a PIR response phase, and an integrity checking phase. In the PIR message construction phase, upon receiving the encrypted data from a child node, the parent node constructs a PIR message and sends the message to the child node to check data integrity. In the PIR response phase, a child node responds with a result message by calculating row values based on the PIR message received from its parent node. Finally, in the data integrity checking phase, the parent node checks whether the data from its child node are valid by comparing two values, that is, the first received value and the second value.

4.2.1. PIR Message Construction Phase. A parent node generates a PIR message to verify the value processed from its child node. The PIR technique was proposed to guarantee the exact result without revealing a client's desired information [18]. For this, it partitions the whole data space into a regular grid of $M \times N$ cells. Hence, a client performs modular computation where the desired cell is set to be Quadratic Nonresidues (QNR) and the other cells are set to be Quadratic

```
command PIR Message (int receieved_data, int child_ID, int PIR_p, int PIR_q)
(1)  PIR_init_data = choose_initial_value(receieved_data, k);
(2)  HCx = choose_HCx_coord(HC_dir, receieved_data, PIR_init_data);
(3)  for (i = 0; i < k; i++) {
(4)      if (i == HCx) y[i] = choose_QNR(p, q);
(5)      else y[i] = choose_QR(p, q);
(6)  }
(7)  Func_order = choose_random(func_Key_Pool[]);
(8)  PIR_init_data = Convert_Func_value(Func_order, PIR_init_data);
(9)  Msg = Construct_Msg(Func_order, PIR_init_data, HC_dir, y[]);
(10)  Send_Message(Child_ID, Msg);
End Algorithm
```

Algorithm 4: PIR message construction algorithm.

Residues (QR). A server then encrypts the dataset through a large number of computations and the user computes the result with the area of QNR. A set of QR and QNR is calculated by using Definition 3. Here, Z_N^* is a set of disjoint integers from N.

Definition 3. A set of QR and QNR.

Let $N = q1 * q2, q1$, and $q2$ large prime numbers:

$$Z_N^* = x \in Z_N \mid \gcd(N, x) = 1,$$

$$\text{QR} = y \in Z_N^* \mid \exists x \in Z_N^* : y = x^2 \bmod N, \tag{4}$$

$$\text{QNR} = y \in Z_N \mid \exists x \in Z_N^* : y \neq x^2 \bmod N.$$

However, the PIR technique is not suitable for sensor networks because its communication cost is very high while sending the whole domain partitions. To adapt the PIR technique to our algorithm, it is necessary to downsize the $M \times N$ domain index to $k \times k$ ($1 < k \leq M$ and $1 < k \leq N$) so that the PIR technique can be applicable to a wireless sensor environment. For this, we first compute k based on the available message size in a sensor network. For example, if the maximum size of one message in a sensor network is 23 bytes, candidate values for k are 2, 3, and 4 owing to $k^2 \leq 23$. Because we use the Hilbert curve technique for our privacy preserving algorithm, we can select 2 or 4 for k. If we select 4 for k, we can set the basic range of the value, that is sent from 0 to 15. Second, the parent node extracts a value x from the PIR message being processed from its child node, and thus the modified value y can range between 0 and $k^2 - 1$. The value x is transformed into $f(x)$ by using a data transformation function that is randomly selected in the given function pool. Because the ID of $f(x)$ is encrypted by using the Hilbert curve technique, it is difficult to obtain the value x. The value y is encrypted by transforming it into two-dimensional data using the Hilbert curve technique. Third, the parent node sets two large prime numbers and computes a set of QR and QNR. Finally, a cell whose Hilbert ID is the same as the modified value y is set to be QNR and the others are set to be QR. Table 2 shows our PIR message structure.

Algorithm 4 shows our PIR message construction algorithm. First, a parent node randomly selects a subtracted value x and calculates a modified value y (line 1). Second, the node converts y into two-dimensional data (line 2). Third, the parent node selects large prime numbers p and q in order to obtain the set of QR and QNR. A cell whose Hilbert ID is the same as the modified value y is set to be QNR and the others are set to be QR (lines 3–9). Finally, the node sends x and the group of the QR and QNR values (line 10).

4.2.2. PIR Response Phase. In the PIR response phase, a child node makes a response message by using both its processed data and the PIR message from its parent node. First, the child node finds a Hilbert value that is the same as the modified value y by subtracting x from the original data. A PIR response message consists of k values that represent k number of rows in $k \times k$ grid cells. Because the value of each grid cell is 0 or 1 in two-dimensional grid cells ($k \times k$), the PIR response message can be expressed by k-bit data. Definition 4 shows how to generate a response value for each column.

Definition 4. Assume that the data set of row Zi is $X_{i1}, X_{i2}, \ldots, X_{ik}$ and the data set of column j is $X_{1j}, X_{2j}, \ldots, X_{kj}$; the rule of generating the value of column j is as follows:

$$\begin{aligned} &\text{If } x' \neq j \text{ then } x_{1j} + \cdots + x_{kj} \geq 2, \\ &\text{otherwise } x_{1j} + \cdots + x_{kj} = 1. \end{aligned} \tag{5}$$

The representative value of each cell can be calculated by using (6) based on Definition 3 and Jacobi Symbol

$$\begin{aligned} Z_j = \prod_{i=1}^{m} w_{ij} \bmod N, \quad &\text{if } X_{ij} = 0 \text{ then } w_{ij} = 1, \\ &\text{otherwise } w_{ij} = y_i. \end{aligned} \tag{6}$$

Algorithm 5 shows our PIR response message construction phase. First, a child node extracts x from its processed value (line 1). Second, the child node finds the Hilbert ID of the result (lines 2-3). Third, the child node generates k^2 data based on Definition 3 (line 4). It then constructs the PIR response message by using (6) (lines 5–12). Finally, the child node sends the PIR response message to its parent node (lines 13-14).

```
command Construct PIR Response Message (intsent_data, Message PIR_Msg)
(1)  Init_data = get_initial_data(Msg.Func_order,Msg.PIR_initail_data);
(2)  HCx = Find_HCx(Msg.HC_dir, sent_data − Init_data);
(3)  HCy = Find_HCy(Msg.HC_dir, sent_data − Init_data);
(4)  rand_data[k] = generate_random_data(HCx);
(5)  for ( j = 0; j < k; i++) {
(6)       z[j] = 1;
(7)       for (i = 0; i < k; j++) {
(8)           if ( j == HCy && i == HCx)    z[j] = z[j]∗ PIR_Msg.y[i];
(9)           if ( j!= HCy && i == HCx)     z[j] = z[j] ∗ 1;
(10)          else z[j] = z[j]∗ (((rand_data[i] >> i) && 0 × 0001)?PIR_Msg.y[i]:1)
(11)       }
(12) }
(13) Message = Construct_Message(z[]);
(14) Send_Message(Parent, Message);
End Algorithm
```

Algorithm 5: PIR response message construction algorithm.

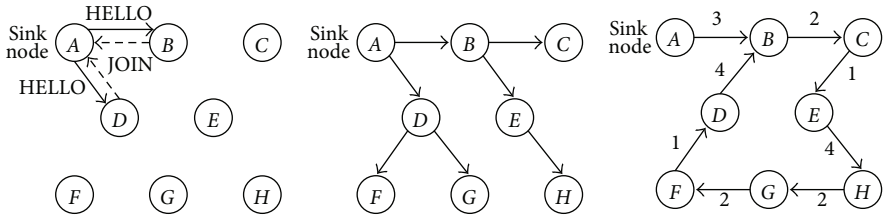
```
command Check_Data_Integrity (int HCx, int HCy, Message PIR_Response_Msg)
(1)  if (!valid_message(PIR_Response_Msg, HCx))    DBG("NOT Valid data");
(2)  for (i = 0; i < k; i++) {
(3)       if (i == HCy) {
(4)           JSvalue = JacobiSymbol(PIR_Response_Msg.z[i]);
(5)           if (IS_QNR(JSvalue) == FALSE && IS_QR(JSvalue) == FALSE)
(6)               DBG("DIFFERENT BETWEEN received_data AND PIR_data");
(7)       }
(8)  }
End Algorithm
```

Algorithm 6: Integrity checking algorithm.

4.2.3. Integrity Checking Phase. In the integrity checking phase, a parent node analyzes the PIR response message and determines whether the received data from its child node is valid. The parent node checks the QR and QNR of the received data by using the selected two prime numbers (in the second phase) and Jacobi symbol. If the received data are valid, there exist $k - 1$ QRs and one QNR. Otherwise, the received data are not valid. Algorithm 6 shows our integrity checking algorithm for the received data. First, a parent node finds the QR and QNR for all columns (lines 1-2). Second, if QNR is set to the column of the modified value, the parent node assures that the processed data from its child node are valid. Finally, the parent node also checks the validity for QR (lines 3–8).

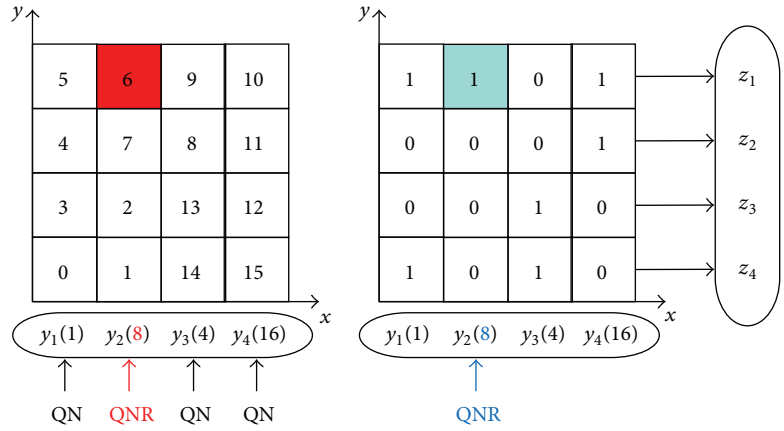
4.3. Example. To protect both data privacy and data integrity, our scheme performs six phases: network construction phase, data encryption phase, PIR construction phase, PIR response phase, data integrity checking phase, and data transmission phase. In the network construction phase, each node sets the information of its sibling nodes, parent node, and child nodes. In Figure 5(a), a sink node A triggers a query by a HELLO message. Upon receiving the HELLO message, sensor nodes B and C determine whether the HELLO message is from the sink node. When B and C receive the HELLO message from A, they set sink node A as its parent node. And other sensor nodes, that is, D, E, F, G, and H, wait for a certain period of time to receive a HELLO message from its neighbors. Upon receiving the HELLO message from any node, the node selects one of the neighboring nodes as its parent node by broadcasting a JOIN message. Figure 5(b) shows the constructed sensor network. After constructing the network, each node exchanges a seed with one node among its neighboring nodes located within its communication boundary. Figure 3 shows the process of the seed exchange. Each node changes sensed data by using the generated seed and the received seeds. All sensor nodes calculate the seed for aggregation, that is, $A = -3$, $B = 5$, $C = 1$, $D = -3$, $E = -3$, $F = 1$, $G = 0$, $H = 2$.

In the data encryption phase, the changed value is encrypted by a Hilbert curve algorithm to send the sensed data (or aggregated data) to the parent node. By selecting the direction and the level of the Hilbert curve, we can encrypt it as a tuple of $\langle \text{key}(d, l), x, y \rangle$ by using two-dimensional data (x, y), the level l, and the direction d. For example, in case of 14, we can encrypt it as $\langle \text{key (Bottom}, 2), 2, 1 \rangle$ because its transformed value, level, and direction are (2, 1), 2, and Bottom, respectively. In the PIR message construction phase, a parent node constructs a message by using Definition 4 and sends the message to a child node for checking data



(a) HELLO and JOIN messages (b) Completing of the network (c) Sending seeds to neighbor nodes

FIGURE 5: Construction phase.



(a) PIR message (1, 8, 4, 16) (b) Response message ($z_1 = 4$, $z_2 = 4$, $z_3 = 16$, $z_4 = 2$)

FIGURE 6: Encrypted data for value 14 (direction = Bottom, level = 2).

integrity. In the PIR sending phase, a parent node constructs a message with k numbers, that is, one QNR and $k - 1$ QR. For example, when the size of row, x, and y and are 4, 3, and 7, respectively, the set of QR is 1, 2, 4, 5, 6, and 8 and QNR includes others except QR. If a node B receives a value 6 from its child node C, B sets the column of the cell (1, 3) as the value of QNR while setting other columns as values of QR. By k values for one QNR and three QR, node B sends a set (1, 8, 4, 16) to its child node to check the validity of the received data. In the PIR response phase, a child node sends to its parent node k values calculated by using (6). For example, by the received data 1, 8, 4, and 16, a child node calculates four numbers, that is, $2 * ((1 * 8 * 16)\%21)$, 16, 4, and 4, as shown in Figure 6(b). For this, a cell being sent to its parent is represented by one while other cells are randomly chosen as zero or one by using Definition 2. If an adversary pollutes the original data, the PIR response phase can determine whether or not the original data are polluted during this processing. Because the adversary cannot know what column belongs to QNR, he/she cannot discover appropriate PIR values.

In the data integrity checking phase, a parent node determines whether the data received from its child node are valid by comparing the first received value with the second value. For example, for the received data z_1, z_2, z_3, z_4, we determine whether the value, that is computed by using Jacobi symbol is one or not. If the computed value is one, its cell value is zero. That is, z_4 can be calculated as $z_4(3) = 2^{(3-1)/2}/3 = 2/3 = -1^{(3^2-1)/8} = -1$, and hence the value of cell (1,3) is 1. Meanwhile, z_3 can be calculated as $z_3(3) = 16^{(3-1)/2}/3 = 4^2/3 = 1$ and $z_3(7) = 4^{(7-1)/3}/7 = 16^2/7 = 4^2/7 = 1$, and thus the value of cell (1, 2) is 0. If all values are valid, a parent node aggregates all the data received from its child nodes to process a user's query. Finally, in the data transmission phase, each sensor node sends to its parent node the encrypted data that are aggregated from its child nodes. For managing a sensor network, a sink node aggregates the data received from all the sensor nodes in the network.

5. Performance Analysis

In this section, we present performance results of both our scheme and existing schemes, in terms of communication overhead, data propagation delay, and integrity checking. For the experiment, we use a TOSSIM simulator [23] running on a TinyOS operating system [22] and a GCC compiler. We make use of 100 sensor nodes that are randomly distributed in a 100 m × 100 m area. As presented in directed diffusion, we use a receiving power dissipation of 395 mW and transmitting power dissipation of 660 mW. Table 1 shows our environment for implementation and Figure 7 shows three types of sensor node distributions for the experiment.

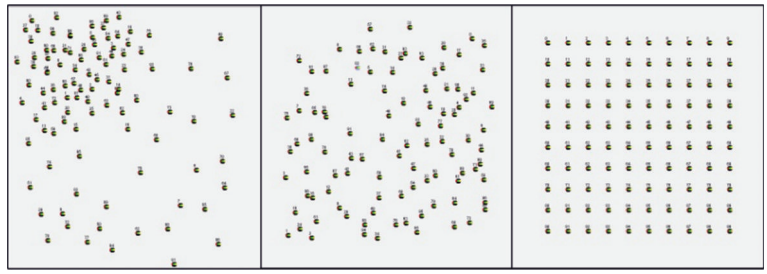


Figure 7: Three types of sensor node distributions.

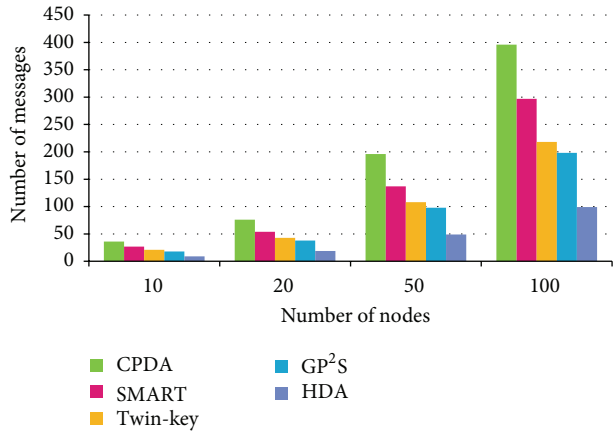


Figure 8: Number of transmission messages with varying number of nodes.

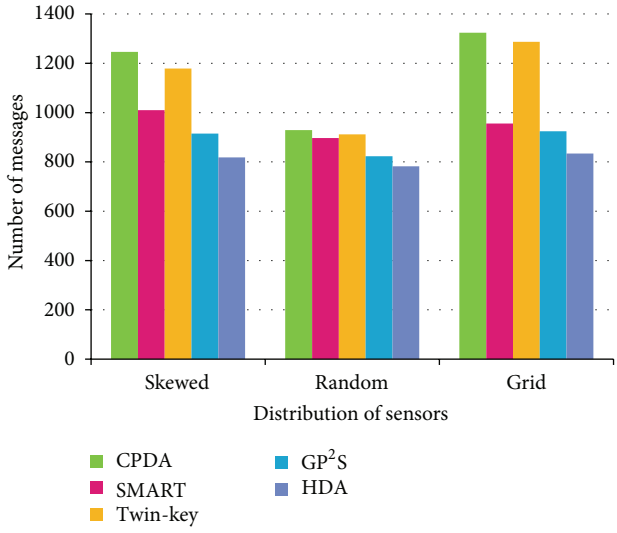


Figure 9: Number of transmission messages with respect to distributions of sensor nodes.

5.1. Experimental Results with Data Privacy Preserving Schemes. We compare our Hilbert-curve based data aggregation scheme (HDA) with CPDA, SMART, Twin-Key, and GP^{2}S, in terms of the number of transmission messages and the average lifetime of the sensor nodes. Here, the number of sensor nodes ranges from 10 to 100. Figure 8 shows the communication overhead with respect to a varying number of sensor nodes. The number of transmission messages in all schemes is increased as the number of sensor nodes increases. This is because when the number of sensor nodes is large, every sensor node in the WSN is capable of sensing data and hence a large number of messages should be transmitted. However, our scheme outperforms the existing schemes by about 10%–20%. The reason for this is that our scheme does not need to generate unnecessary messages during data aggregation since each sensor node can transform only its own data whereas the existing schemes require an additional message for privacy preservation. Figure 9 shows the number of transmission messages with respect to different distributions of sensor nodes. Figure 10 shows the number of transmission messages with a varying communication boundary when the number of sensor nodes is 100. In both figures, our scheme outperforms the existing schemes because it does not require unnecessary messages in all the cases. In particular, our scheme, SMART, and GP^{2}S show consistent performance regardless of the type of distributions and the communication boundary. This is because they are less affected by the placement of sensor nodes owing to the use of a tree topology. Meanwhile, CPDA and Twin-Key are strongly influenced by both the type of distributions and the communication boundary, because they make use of a clustering method.

Table 1: Environment for implementation.

CPU	Intel Core2 Duo CPU E4500 2.20 GHz
Memory	2 G
Language	nesC
Simulator	TOSSIM
Compiler	GCC ver. 4.0.3

Figure 11 shows the average lifetime of the sensor network with varying number of sensor nodes in the WSN. In this analysis, we measure the time until the number of sensor nodes, whose energy is completely consumed, is greater than 50% of all sensor nodes. The lifetime of all the schemes decreases as the number of sensor nodes increases. This is because the number of messages generated in the network is proportional to the number of messages required for data aggregation. However, the lifetime of our scheme becomes 100%~125% longer than those of all the existing schemes,

Table 2: Our PIR message structure.

The encrypted ID of the transformation function	Transformed value $f(x)$	The direction of the Hilbert curve	The level of the Hilbert curve	PIR data set			
				1st column	2nd column	...	k-th column

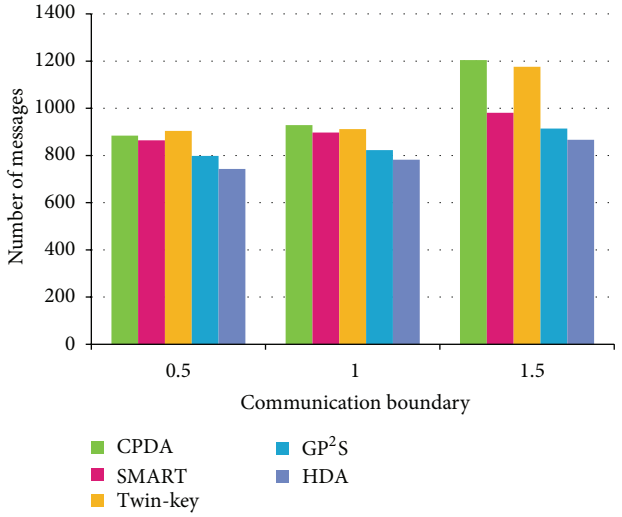


Figure 10: Number of transmission messages with respect to communication boundary.

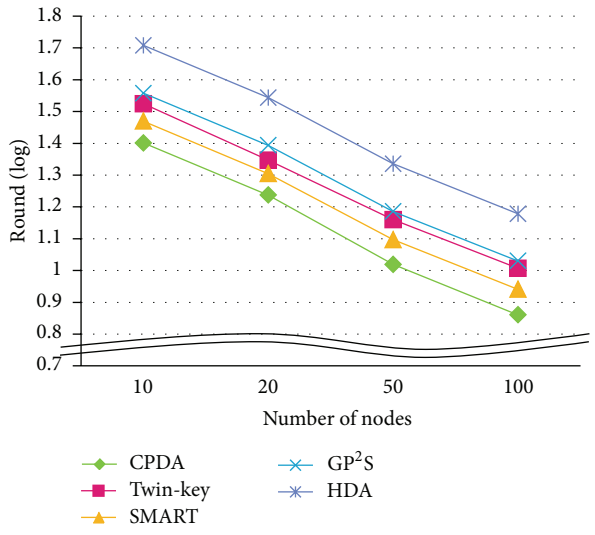


Figure 11: Lifetime of sensor network with varying number of nodes.

because our scheme can reduce unnecessary messages during data aggregation.

5.2. Experimental Results with Data Integrity Schemes. We compare our data integrity validation HDA scheme (iHDA) with iCPDA and iPDA in terms of the number of transmission messages per query round, the average lifetime of sensor nodes, and the attendance ratio of sensor nodes. Figure 12 shows the communication overhead with respect

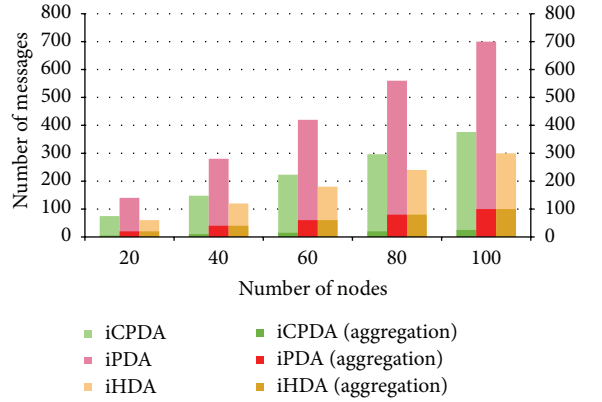


Figure 12: Number of messages generated by iPDA, iCPDA, and our scheme.

to varying number of sensor nodes in a WSN. The number of transmission messages for iPDA, iCPDA, and our scheme is increased as the number of sensor nodes increases. Our scheme outperforms the iPDA and iCPDA schemes because the existing schemes generate unnecessary messages during data aggregation in the network. That is, each sensor node generates only two additional messages for privacy preservation and integrity checking in our scheme whereas the iPDA and iCPDA schemes generate six and four messages, respectively. Due to numerous messages exchanges among sensor nodes, there is a high rate of data collisions in the existing schemes. Therefore, the iPDA and iCPDA schemes are very expensive in terms of communication overhead because the number of messages generated in the network is very large for successful data transmission. Figure 13 shows the average lifetime with respect to varying number of sensor nodes in the WSN. The dissipated energy for all three schemes is increased as the number of sensor nodes increases. This is because every message generated in the network requires energy to reach the sink node. However, in terms of lifetime, our scheme shows 35~130% better performance than iPDA and iCPDA schemes. The reason is that the iPDA and iCPDA schemes generate too many unnecessary messages for data aggregation to enforce both integrity protection and privacy preservation. In the existing schemes, every sensor node becomes active to send its messages for a longer time.

Figure 14 shows the attendance ratio of sensor nodes for data aggregation. During data aggregation, a sensor node sends the sensed data (or aggregated data) to its parent node. The attendance ratio of sensor nodes in our scheme is about 100% whereas both iPDA and iCPDA have some sensor nodes that do not take part in data aggregation. Because a given sensor node in the iPDA and iCPDA schemes has to communicate with at least six and two neighboring nodes, respectively, some sensor nodes cannot participate

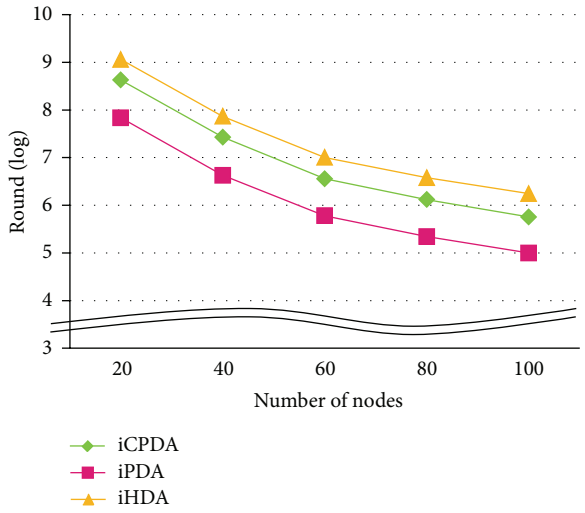


Figure 13: Average lifetime for each sensor node.

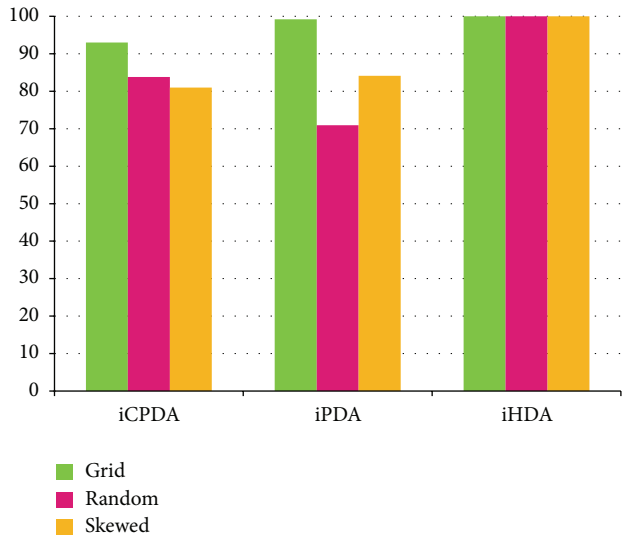


Figure 14: Attendance ratio of sensor node for data aggregation.

in data aggregation. Therefore, our scheme shows the best performance among the three schemes.

6. Conclusion and Future Work

Recently, as advanced technologies of mobile devices and wireless communication proliferate, wireless sensor networks (WSNs) have increasingly attracted interest from various applications including military and environmental monitoring. Moreover, since sensor nodes have limited resources, such as battery and memory capacity, many data aggregation techniques have been proposed for WSNs. However, the wireless communication can be easily overheard, and thus the provision of a data aggregation scheme to support data privacy is a challenging issue in WSNs. Although several data aggregation schemes have been proposed to preserve data privacy, they have the following limitations. First, the communication cost for network construction and data aggregation is considerably expensive. Second, only a part of the existing methods supports data privacy. In addition, it is necessary to assure that the aggregated data are not polluted by an unauthorized third party. For this, we propose a new data aggregation scheme for enforcing both data privacy and data integrity in WSNs. Our scheme makes use of a seed exchanging algorithm to reduce the communication cost for preserving data privacy. It also utilizes an integrity checking algorithm based on a private information retrieval (PIR) technique. From our performance analysis, we show that our HDA scheme achieves 100%–300% longer network lifetime and about a 10% better attendance rate for the aggregated data than the existing privacy preserving schemes. In addition, our iHDA scheme achieves 40%–160% better performance in terms of network lifetime and about a 16% better participation rate for the aggregated data. As future work, we plan to verify that our scheme is efficient in WSNs by applying it to a real environment.

Acknowledgments

This research was supported by Sharing and Diffusion of National R&D Outcome funded Korea Institute of Science & Technology Information (KISTI) (K-13-L02-C02-S02). And this research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0023800).

References

[1] "James reserve microclimate and video remote sensing," 2008, http://www.cens.ucla.edu/.

[2] "The firebug project," 2008, http://firebug.sourceforge.net/.

[3] "Habitat monitoring on great duck island," 2008, http://www.greatduckisland.net/.

[4] K. Du, J. Wu, and D. Zhou, "Chain-based protocols for data broadcasting and gathering in sensor networks," in *Proceedings of the International Parallel and Distributed Processing Symposium*, April 2003.

[5] W. R. Heinzelman, *Application-specific protocol architectures for wireless networks [Ph.D. thesis]*, Massachusetts Institute of Technology, 2000.

[6] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: a scalable and robust communication paradigm for sensor networks," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MOBICOM '00)*, pp. 56–67, August 2000.

[7] S. Lindsey, C. Raghavendra, and K. M. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics," *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no. 9, pp. 924–935, 2002.

[8] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG: a tiny AGgregation service for ad-hoc sensor networks," *ACM SIGOPS Operating Systems Review*, vol. 36, pp. 131–146, 2002.

[9] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks,"

IEEE Transactions on Mobile Computing, vol. 3, no. 4, pp. 366–379, 2004.

[10] N. Li, N. Zhang, S. K. Das, and B. Thuraisingham, "Privacy preservation in wireless sensor networks: a state-of-the-art survey," *Ad Hoc Networks*, vol. 7, no. 8, pp. 1501–1514, 2009.

[11] W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. Abdelzaher, "PDA: privacy-preserving data aggregation in wireless sensor networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (IEEE INFOCOM '07)*, pp. 2045–2053, Anchorage, AK, USA, May 2007.

[12] M. Conti, L. Zhang, S. Roy, R. di Pietro, S. Jajodia, and L. V. Mancini, "Privacy-preserving robust data aggregation in wireless sensor networks," *Security and Communication Networks*, vol. 2, no. 2, pp. 195–213, 2009.

[13] H. Choi, S. Zhu, and T. F. la Forta, "SET: detecting node clones in sensor networks," in *Proceedings of the 3rd IEEE International Conferenceon Security and Privacy in Communication Networks (SecureComm '07)*, pp. 341–350, Nice, France, September 2007.

[14] W. S. Zhang, C. Wang, and T. M. Feng, "GP2S: generic privacy-preservation solutions for approximate aggregation of sensor data," in *Proceedings of the 6th Annual IEEE International Conference on Pervasive Computing and Communications (PerCom '08)*, pp. 179–184, Hong Kong, March 2008.

[15] W. He, X. Liu, H. Nguyen, and K. Nahrstedt, "A cluster-based protocol to enforce integrity and preserve privacy in data aggregation," in *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops (ICDCS '09)*, pp. 14–19, Montreal, Québec, Canada, June 2009.

[16] A. Beimel, Y. Ishai, E. Kushilevitz, and J. F. Raymond, "Breaking the O(n1/(2k-1)) barrier for information-theoretic private information retrieval," in *Proceedings of the 34rd Annual IEEE Symposium on Foundations of Computer Science*, pp. 261–270, November 2002.

[17] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, "Private information retrieval," *Journal of the ACM*, vol. 45, no. 6, pp. 965–982, 1998.

[18] E. Kushilevitz and R. Ostrovsky, "Replication is NOT needed: SINGLE database, computationally-private information retrieval," in *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, pp. 364–373, 1997.

[19] S. Yekhanin, "New locally decodable codes and private information retrieval schemes," Tech. Rep. ECCC TR06-127, 2006.

[20] A. R. Butz, "Alternative algorithm for Hilbert's space filling curve," *IEEE Transactions on Computers*, vol. C-20, no. 4, pp. 424–426, 1971.

[21] Y. Panthachai and P. Keeratiwintakorn, "An energy model for transmission in Telos-based wireless sensor networks," in *Proceedings of the International Joint Conference on Computer Science and Software Engineering (JCSSE '07)*, 2007.

[22] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TinyDB: an acquisitional query processing system for sensor networks," *ACM Transactions on Database Systems*, vol. 30, no. 1, pp. 122–173, 2005.

[23] http://www.tinyos.net/tinyos-2.x/tos/lib/tossim/.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 789405, 9 pages
http://dx.doi.org/10.1155/2013/789405

Research Article

Batteryless Powering of Remote Sensors with Reversed Peltier Power Source for Ubiquitous Environments

Ondrej Krejcar and Robert Frischer

Department of Information Technologies, Faculty of Informatics and Management, University of Hradec Kralove, Rokitanskeho 62, 50003 Hradec Kralove, Czech Republic

Correspondence should be addressed to Ondrej Krejcar; ondrej.krejcar@uhk.cz

Received 7 November 2012; Revised 24 March 2013; Accepted 16 April 2013

Academic Editor: Chao Song

The need for remote monitoring of almost every place in the world is still growing, while many new technical solutions are extensively developed by research community. The knowledge of the environment and the weather conditions of human environment is, in certain cases of polluted places in heavy industrial areas, necessary to allow real-time warnings about deteriorating conditions to living community in these desired places. While huge money investments were used for real-time monitoring of every possible quantity of human body, people's living environment became increasingly poorer. In this paper we present a modern and relatively cheap solution for powering some types of intelligent sensors in which traditional battery power source is insufficient and other powering options are not applicable. Obtaining the energy to supply sensors is possible even from immediate sensors' environment. Sources (which are utilizing thermal gradient) are supplying energy from the surrounding environment and without the need for high-intensity incident light (solar energy based). In order to achieve the full working mode, only a small thermal gradient is needed in order of degrees. The developed test application is using sophisticated electronic circuit provided by Linear Technology Company, which is able to use even small energy bursts for discontinuous intelligent sensors operations. The heat energy is converted into electric energy which is stored in very high-capacity capacitor (about 1 F). This power supply type is suitable for any environment, where some permanent thermal gradient is presented.

1. Introduction

Demands on controlling heating, ventilating, and air conditioning (HVAC) processes in modern multilayer buildings or in the industry are increasing. There are many sensors, which can be used to control given process. However, if we have no sufficient cable infrastructure, these sensors are unusable. Of course, we can use battery-powered smart sensors, but their prize and the need to replace the batteries is constraining. The ideal sensor should have the capability to power itself from the surrounding free energy. This technique is called energy harvesting. Energy harvesting is by no means a new idea. It is a process, by which free ambient energy is derived, captured, and stored. Devices which make use of energy harvesting require output power that ranges from a few nanowatts to tens of milliwatts. Although nontraditional power sources such as solar cells (photovoltaic cells), thermoelectric generators (TEGs), thermopiles, and piezoelectric transducers are known sources of electrical power, harnessing power from these sources has been challenging [1]. Each of these requires some type of power conversion circuit that can efficiently collect, manage, and convert these alternative power sources into a more usable form of electrical energy to power sensors, microcontrollers, and wireless transceivers. Advances in low-power technology are making it easier to create sophisticated power energy management units, which are able to collect small pieces of energy, derive them and store them in high-capacity capacitor.

There are many types of energies, which can be utilized. We can harness vibration energy, solar energy, potential energy, mechanical energy, thermal energy, and many others. One common form of ambient energy is mechanical vibration energy, which can be caused by motors running in a factory, by airflow flowing across a fan blade, or even by a moving vehicle. A piezoelectric transducer can be used

to convert these forms of vibration energy into electrical energy, which in turn can be used to power wireless sensors. To manage the energy harvesting and the energy release to the system, the LTC3588 piezoelectric energy harvesting is used for power supply. It uses an efficient energy harvesting algorithm to collect and store energy from high impedance piezoelectric elements, which can have short-circuit currents in order of tens of microamps [2]. The LTC3588 harvests ambient vibrational energy through a piezoelectric element which is its primary application. Common piezoelectric elements are PZT (lead zirconate titanate) ceramics, PVDF (polyvinylidene fluoride) polymers, and other composites. Ceramic piezoelectric elements exhibit a piezoelectric effect when the crystal structure of the ceramic is compressed and internal dipole movement produces voltage. Polymer elements comprised of long-chain molecules produce voltage when flexed as molecules that repel each other. This type of source has one main advantage compared to the other. Output voltage is very high (in comparison to other sources), about 12–25 V. This voltage level is perfect to convert into traditional 3.3 V or 5 V (MCU core and peripherals). Conversion ratio is about 3.5 : 1, so efficiency is high. But supplied power is very low. Typically, harvestable power of this source is in order of tens microwatts. According to (1), if supplied power is low, supply voltage is high, so supplied current is even smaller. This imposes enormous requirements on power/energy management device:

$$I_{\text{supp}} = \frac{P_{\text{supp}}}{U_{\text{supp}}} \ [A, W, V]\,. \tag{1}$$

One of the most common energies is thermal environmental energy. Thermal energy is ubiquity, in a relatively large amount. However, there are only few devices which can convert this energy into electric current. One of these is called thermocouples. Thermocouples have been known for a long time. Their main advantage is that they can deliver relatively high current (hundreds mA). But often there is some drawbacks. In this case, supplied voltage is one of the lowest (order of mV). Conversion ratio is about 1 : 50, so efficiency is low. 50 years ago, Russians researchers experimented with kerosene lamps which contain many thermocouples placed on the lamps hat. Resulting power source was able to power simple transistor radio receiver. This type of source has no stabilized output voltage and requires stable heat flow. Also weight, size, and prize of such device cannot be ignored. Modern heat utilizing power sources uses TEGs. TEGs are well accessible, light weighted, and cheap. Low voltage disadvantage is compensated by relatively high supply current and good overall power delivery (hundreds mW). This type of energy has been left out, because the proper switching elements simply have not been accessible. Standard bipolar transistors are not acceptable due to their high Ucesat (Collector Emitter Voltage in saturation state).

Energy harvesting requires careful power management and sophisticated circuits with latest technology design in order to successfully capture a few microwatts of ambient power and store them in a useable energy reservoir. A capacitor with large capacity usually serves as a reservoir.

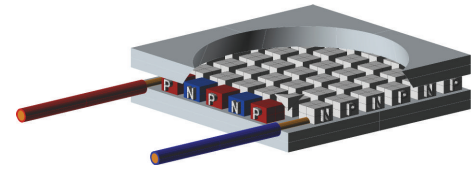


Figure 1: Interconnected semiconductor pellets in TEG module.

These electrical devices have been recently revealed. Their capacity, size, equivalent series resistance (ESR), and small leakage current make them perfect to implement into energy harvesting power sources. This capacitor is crucial, because energy is stored after small pulses and, in addition, time between energy deliveries may vary from seconds to minutes. Capacities between 0.1 F and 1 F are recommended and special care must be taken to choose low leakage (about 400 nA) version. Another important device is a switching transistor. Nowadays, special N-MOSFET transistors have parameters, which could not be achieved in the past. A standard bipolar transistor has a very long history and is practically employed on a high level; it has two semiconductor transitions, so minimum switching voltage is about 0.5 V–1 V, which makes it inapplicable in most energy harvesting sources. Modern MOSFET transistors are special electronic elements, which are capable of switching high currents (much more than 100 A) while serial resistance (RDS (on)) remains very low, in order of mOhms [3].

2. Thermoelectric Generators

Thermoelectric generators (TEGs) are simply thermoelectric modules based on Seebeck effect. The Seebeck effect is named after the German physicist Thomas Johann Seebeck (1770–1831), who, in 1826, published the results of experiments done four years earlier that opened up the new field of thermoelectricity. He observed that an electrical current is present in a series circuit of two dissimilar metals, provided that the junctions of the two metals have different temperatures. Devices working based on Seebeck effect convert heat energy into electrical energy where a temperature gradient is present. Sources of heat energy vary from body heat, which can produce tens of μW/cm^2 to a furnace exhaust stack where surface temperatures can produce hundreds of mW/cm^2. TEGs are basically Peltier cells, running in a reversed mode. The Peltier effect produces a temperature differential by applying voltage and is most commonly used in thermoelectric coolers [4]. These Peltier cells are made from many small semiconductor blocks connected together with copper tape. The most commonly used semiconductor material is bismuth-telluride (Bi2Te3). Output power is directly proportional to the cells' surface and it is directly proportional to the number of semiconductor blocks. The inner structure of Peltier cell is presented in (Figure 1).

TEGs come in a wide variety of sizes and electrical specifications. The most common modules are square shaped, ranging in size from about 10 mm to 50 mm. Their thickness

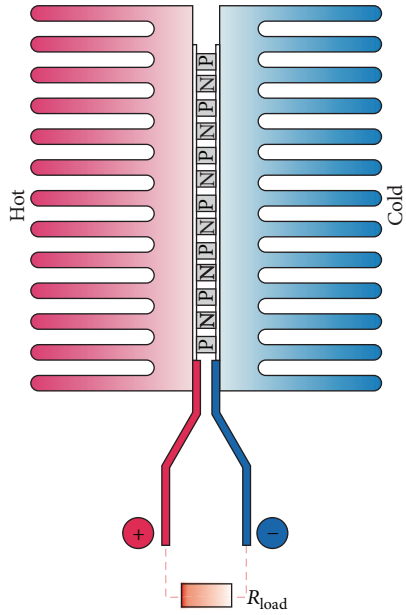


Figure 2: Measuring conditions. Heat sinks have been tightened together to ensure best heat flow.

is usually 2 mm–5 mm. We have three types of TEGs (Peltier modules) available.

Most thermoelectric module manufacturers do not provide data for output voltage or output power in relation to differential temperature, which is what the designer of a thermal energy harvester wants to see. Two parameters that are always provided are V_{max} and I_{max}, which are the maximum operating voltage and maximum operating current for a particular module (when being driven in a heating/cooling application). We want to find out if there is a difference between 8.6 V and 15.4 V TEGs. The one, which has higher working voltage should present higher output voltage, because of higher count of interconnected semiconductor pellets. Also, we want to know the available output electrical power under specific condition. One side of the TEG was cooled down to room temperature by electric ventilator and the other side was warmed up by hot air. Both sides were equipped with large heat sink to ensure good heat flow (Figure 2).

Supposed temperature layout is presented in (Figure 3). The central parts have a linear progression, because the heat perforates through a plain wall. In this area, heat flow depends on the temperature gradient and the heat passage coefficient of all materials which are involved. Individual heat flows can be easily counted according to the following:

$$Q_{\tau 1} = \frac{\lambda_1}{d_1} \cdot S \cdot (\Delta T_1)\,[W]$$
$$Q_{\tau 2} = \frac{\lambda_2}{d_2} \cdot S \cdot (\Delta T_2)\,[W] \quad (2)$$
$$\vdots,$$

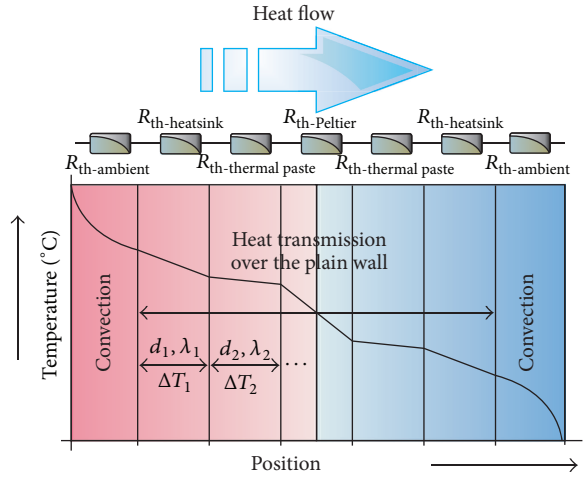


Figure 3: Temperature distribution of measured sample.

where $Q_{\tau 1}$ is heat flux through the wall $\tau 1$, ΔT is temperature gradient on both sides of the wall, λ is the heat carrying capacity of the current material, and S is the wall's surface. Total heat flow formula can be seen in the following:

$$Q_\tau = \frac{\Delta T}{(d_1/\lambda_1) + (d_2/\lambda_2) \cdots (d_n/\lambda_n)} \cdot S. \quad (3)$$

Heat flow from the heat sink's border into the air has a logarithmic run (Figure 3).

We have made several tests of some commercially available TEGs. The tests were conducted under the same conditions in order for the results to be objective. Temperatures on both sides were recorded and corresponding voltages and passing currents were measured. Supply power was calculated by formula (4). TEG supplies its power into constant load represented by power resistor with value 0R27Ohms. Results can be compared in (Figure 4):

$$P_{TEG} = U_{TEG} \cdot I_{TEG}\ \ [W]. \quad (4)$$

It can be assumed from (Figure 4) that a cell with a higher nominal voltage (meaning standard supply voltage for normal mode) reports better results. The reason for that is, most probably, the higher count of serial connected semiconductor pellets. Delivered power is only theoretical, because of the high stepup ratio. Total DC/DC efficiency is only between 10 and 50%. Delivered electric power is then between 5 and 100 mW, depending on the input voltage level.

3. Energy Harvesting Management

The basic common feature of the power sources which uses free energy of the surroundings is a discontinuous operation. Many wireless sensor systems require only low average power, which make them front candidates for powering by energy harvesting technique. In addition, many sensors monitor variables with very slow process system type (e.g., oxygen content in water, slow temperature process, and air humidity). Therefore, the measurement can proceed discontinuously,

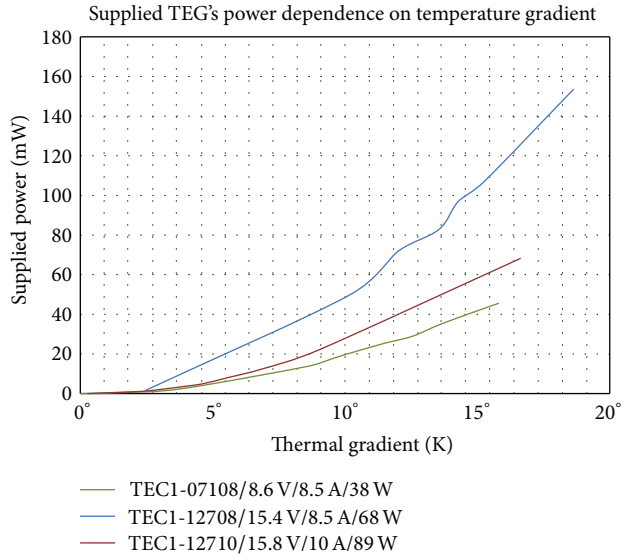


Figure 4: Theoretical electric power delivered by TEGs. Output voltage is in range 50–500 mV.

resulting in the low frequency of the cycles and correspondingly low power requirement. If some intelligent wireless sensor make a measurement sample e.g. once per a minute, the average power plummets under 250 μW. This value is significant, because most forms of energy harvesting offer very little steady-state power, so continuous operation is impossible.

As a power management component was chosen, an integrated circuit was provided by Linear Technology, LTC3108. The LTC3108 self-resonant topology steps up from input voltages as low as 20 mV. This energy harvester power source is designed for applications which are using very low average power, but requiring periodic pulses of higher load current (Figure 5). A double-layer capacitor serves as an energy reservoir with large capacity. The recommended capacity is above 0.1 F. Real output power curve can be seen in (Figure 8). The red curve represents output voltage bypassed by 470 uF capacitor, which should align any current spikes. Steady output current is only about 7 mA. The blue curve represents input voltage from TEG, green curve is voltage on supercapacitor, and finally pale blue curve represents useful energy stored in supercap expressed in Ws. As a store capacitor used 1 F sample with 5.5 V maximum voltage and very low leakage current value, maximum storage energy is about 5 Ws. It is big enough to power intelligent wireless sensor, acquire desired data, and send them to superior system to evaluate and subsequent post processing.

The accumulated energy can be consequently calculated as follows:

$$E_{\text{cap}} = \frac{1}{2} \cdot C \cdot U^2. \tag{5}$$

This energy is only theoretical and contains nonutilisable part, because we could utilize only definite voltage range, for example, 3.3–4.0 V. Voltage below the lower limit cannot be utilized (processors core). The useful energy held by capacitor can be calculated as the follows:

$$E_{\max} = E_{V_{\text{MAX}}} - E_{V_{\text{MIN}}},$$
$$E_{\max} = \frac{1}{2} \cdot C \cdot \left(U^2_{V_{\text{MAX}-}} - U^2_{V_{\text{MIN}}}\right). \tag{6}$$

The power source with advanced power management IC is presented in (Figure 6). We can see a TEG which is upheld by CIN capacitor to compensate little dropouts. A step-up transformer is designed for the given TEGs output level. The size of this transformer is not essential, because input and output currents are low. Capacitance $C1$ is the charge pump capacitor that is connected from the transformer's secondary winding to the $C1$ pin which has an effect on converter input resistance and maximum output current capability. Generally, the minimum value of 1 nF is recommended when operating from very low input voltages using a transformer with ratio of 1 : 100.

The main advantage of this IC is its ability to start oscillating, even if a very low voltage is applied. The LTC3108 utilizes an MOSFET switch to form a self-resonant stepup oscillator using an external step-up transformer and a small coupling capacitor ($C2$). This allows it to boost input voltages from 20 mV high enough to provide multiple regulated output voltages for powering other circuits. The frequency of oscillation is determined by the inductance of the transformer secondary winding and is typically in the range of 10 kHz to 100 kHz. Notice V_{out2} output. This voltage output is driven by the $V_{\text{out2\_EN}}$ logic pin from the connected microprocessor unit. It helps to minimize printed circuit board (PCB) size and reduce complexity of the whole device. This output, controlled by a host processor, can be used to power external circuits such as sensors and amplifiers that do not have a low power sleep state or shutdown capability. V_{out2} can be used to power these circuits only when they are needed. Designed power source's PCB is presented in (Figure 7).

Real output power curve can be seen in (Figure 8). The red curve represents output voltage bypassed by 470 uF capacitor, which should align any current spikes. Steady output current is only about 7 mA. The blue curve represents input voltage from TEG, green curve is voltage on super capacitor, and finally pale blue curve represents useful energy stored in super cap expressed in Ws. As a store capacitor was used 1 F sample with 5.5 V maximum voltage and a very low leakage current value. Maximum storage energy calculated by (6) is about 5 Ws. It is big enough to power intelligent wireless sensor, acquire desired data, and send them to superior system for evaluate and subsequent postprocessing.

All components were set to optimum device run. Output Schottky diode was strengthened by an external one, so the voltage drop across the diode was lowered. For that purpose was chosen the Schottky barrier rectifier in surface mount (SMD) package with average forward current of 2 A and a maximum reverse voltage of 30 V. Its transfer characteristics are presented in (Figure 9).

As can be seen, for normal temperature of 25°C is forward voltage drop for current in order of mA only 0.08 V. This can significantly improve the whole efficiency, which is very

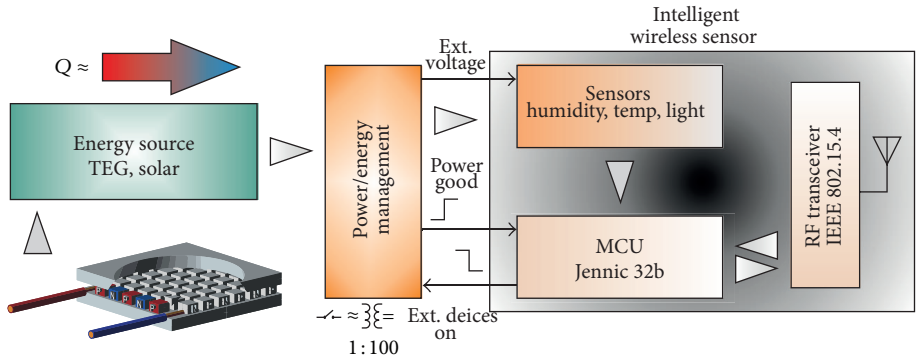


Figure 5: Basic blocks defining powering and measuring chain of independent wireless system.

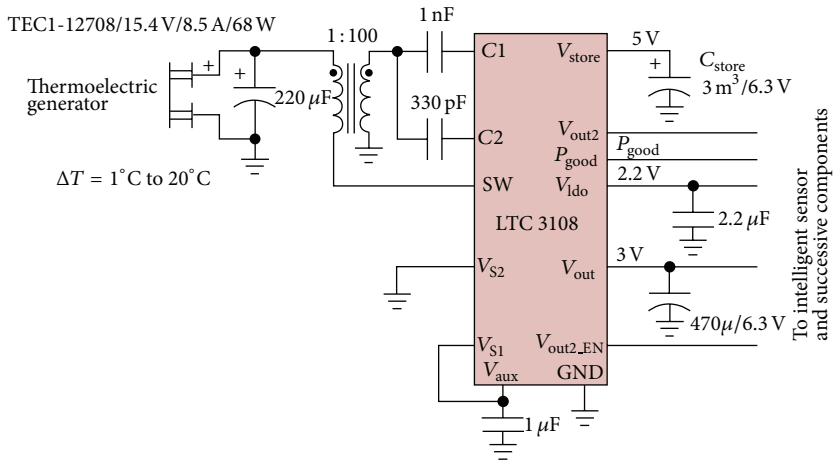


Figure 6: TEG power source for low-input TEGs.

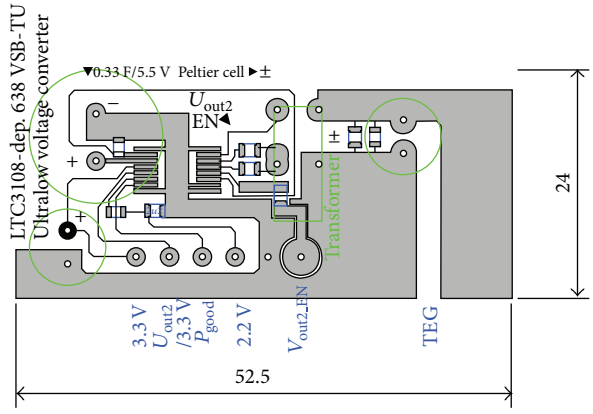


Figure 7: Developed intelligent energy harvesting power source for remote sensors powering (PCB).

low. A picture from the oscilloscope presents voltage running on this diode. Frequency of 1.898 kHz was used and the oscillogram is shown in (Figure 10).

Small glimmer in the conduction phase forms a transformer core response and is very small due to small transformed energy. These glimmers can be a serious issue if the transformed power is high.

Rectified voltage ripple is 75 mV, which is a relatively good result. Relatively, because the output supply current is very low, in other cases the ripple has tendency to decrease to 0 V. The output ripple is presented in (Figure 11).

Output ripple is substituted by yellow line and input ripple by red line. Input ripple is significantly lower. This is the sign, that the TEG is very "hard" power supply. Maximum input ripple is about 10 mV, which is partly due to 10 uF ceramic capacitors which bypass the input. The variable switching frequency is visible from the pictures. This is primarily due to the power management unit, which changes switching frequency depending on the input voltage and output power, and also due to primary winding inductance. The inductance of the primary winding is a key attribute when designing target switching frequency. In this case, big core transformer was used, to minimize switching losses (by minimizing switching frequency). The size of the transformer is not important. The goal was to find the possibilities of that type of power supply. On the other hand, the low switching frequency would not interfere any supplying wireless sensors and microprocessors.

In Figures 12 and 13 the voltage running on primary and secondary transformer side is presented.

The two pictures (Figures 12 and 13) show the voltage transformation from input to output. On the primary side, the

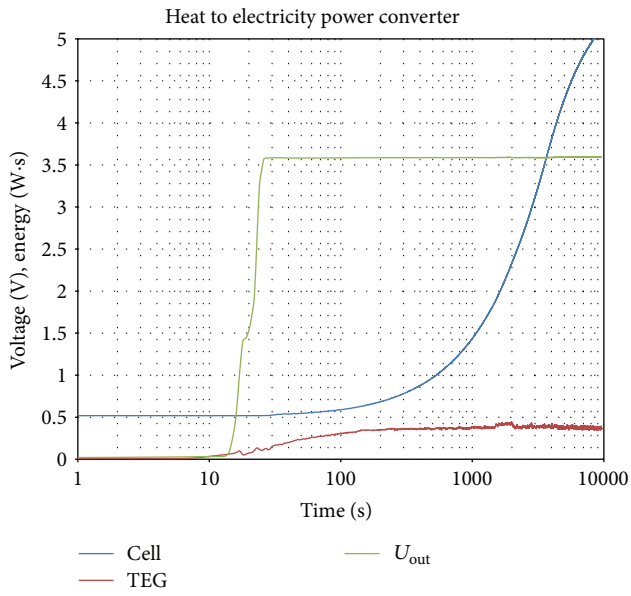


Figure 8: Output voltage running and energy stored in supercap (pale blue).

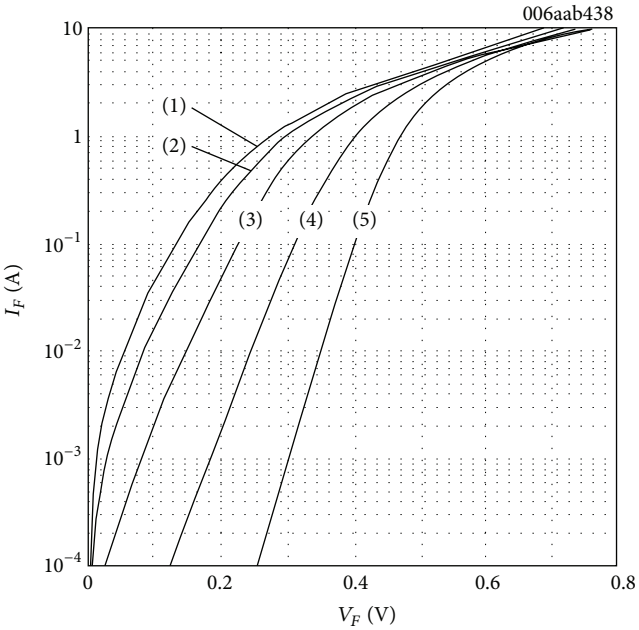


Figure 9: Forward current as a function of forward voltage for 5 temperatures from −40°C to 150°C.

relatively stable DC voltage with value of 411 mV is presented. When the internal switching transistor accumulates, voltage on the transformer suddenly decreases to almost zero level. This is a sign that internal switch is of a unipolar construction. Then the voltage is starting to increase its value, which is a sign that the switch is not able to maintain passing current and closes itself. This lasts for about 100 us. This time is sufficient to transport low-voltage energy from the input to high-voltage energy on the output. Figure 13 shows

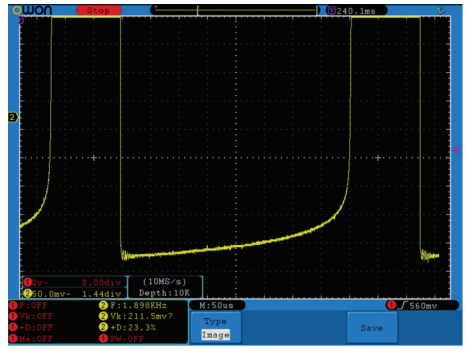


Figure 10: Oscilloscope output, when measuring rectifier diode forward running.

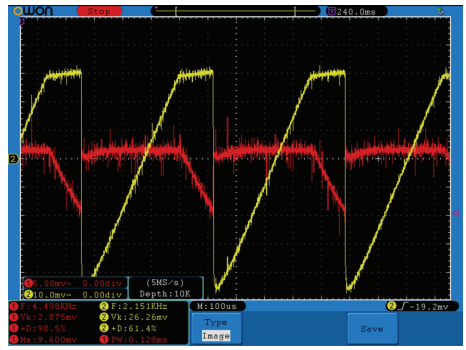


Figure 11: Output ripple (yellow line) versus input ripple (red line).

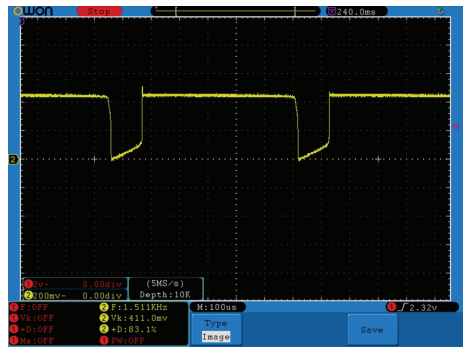


Figure 12: Switching wave form on primary transformer side.

that the secondary voltage has a maximum value of about 7.2 V which quickly falls to 5 V. This is due to the used load. Supercapacitors with several farads of capacity have very high internal resistance. Part of the delivered energy is transformed to heat on that resistance and the rest is utilized to increase overall capacitors' charge.

4. Energy Harvesting Power Management—Modern Approach

After many tests, we established that commercial products, such as Linear LTC3108, are suitable for standard, widespread applications. If we want to fit within special conditions such

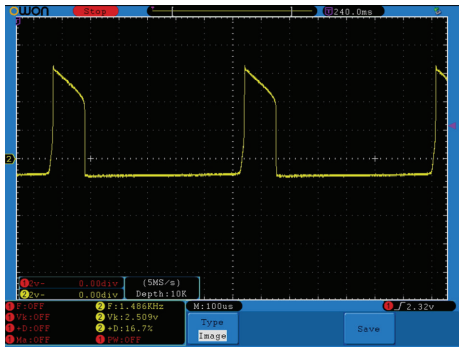


Figure 13: Switching wave forms on secondary transformer side. The output voltage is 17 times higher than the input voltage.

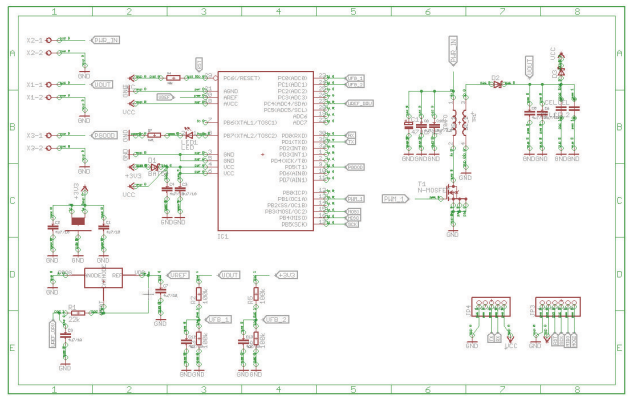


Figure 14: Electrical scheme of intelligent energy harvesting power source.

as higher output power, higher charging current, or mixed energy supplies (solar energy + thermal energy), we do not have many options. This area is relatively young and it is still being searched for applications, where this approach could be extensively used.

There are some good journal papers [5–7] that are focused on some of the mentioned problems, but none of them bring the complete solution for optimized power supply. The first one [5] is mainly focused on algorithms to achieve a better system performance as well as lower power consumption. This is certainly needed to be mentioned while developing very low power sensors, but it is not enough in certain cases. This solution is more suitable for more complex measurement cycles.

Some of the interesting solutions can be found in [6] where authors deal with two main options for power sources such as mechanical and electromagnetic energy harvesting, where they also present several circuits' designs. Unfortunately, this issue is not targeted by Peltier cell solution.

The last of the selected solutions deals with a development of a test chip using 180 nm low-threshold CMOS process which results in efficiency of 30%. Unfortunately, this chip is not available in the market now. We assume that the price of this product will be a little higher than the one affordable by our design.

Another paper [8] deals with a supercapacitor with a solar cell combination, but this solution is not possible to be easily used in case of Peltier cells. The major issue arises in a case when we want to charge power supercapacitor quickly, with greater current, which can be delivered by LTC3108. Commercial IC has limitations resulting from its size. In most cases, we have to handle it with small DFN packages, with sizes of about 3×3 mm. In this package, the high power, low RDS-ON MOS-FET switch cannot be implemented. Hand by hand with these limitations go other disadvantages. In order to deliver higher input power and larger rectifying diode, larger IC terminals, thicker PCB ways, and others must be used. That is why we design new approach, which has roots in a classic energy harvesting power supply design.

The whole idea is based on ATMEL ATMEGA8 MCU. This MCU has a very low-supply current in active (about 500 μA) and idle mode (about 50 μA). MCU is a driving power MOS-FET with very low RDS-ON (2 mΩ) and high drain current (191A ON state current). This value is excessive and will never be achieved, but other switch parameters are highly suitable for our purposes. Electrical scheme is presented in (Figure 14).

The major differences compared to commercial products are as follows.

(1) Solution is able to process much higher current than standard energy harvesting device.

(2) It can store higher energy. It can use also larger supercapacitors.

(3) Output voltage can be controlled by software.

(4) It uses CR2032 lithium cell to supply MCU when no input power is present.

(5) This prototype is relatively bulky.

(6) It can be controlled using UART or SPI interface.

Figure 15 shows a PCB of developed power source. We can focus on several major impact parts. Cell 1 and Cell 2 are high capacity multilayer capacitors supported by two small ceramic capacitors. The main energy part is stored in bulk caps, but they cannot deliver high current spikes. For that reason ceramic capacitors are good supporting choice (very low ESR). Power MOS-FET is placed on the main power path above the JP3 communication connector. Main rectifier schottky diode (D2) is of a bulky type and can carry over 3 A. This value is also unreachable, but on the other hand, due to that, this diode has a very low forward voltage (50 mV/1 A). This can help increase the whole efficiency and maximize input energy utilization.

In case that input power is not sufficient to power up MCU, installed lithium battery will ensure the right function. If there is enough input power, output stabilized voltage will pass through separation diode D3 and substitute primary lithium cell. This cell can be replaced by secondary, rechargeable cell. If the input power is too high, there is an LED1 component. Excess energy is then spent in this LED and will not harm any device in the power chain.

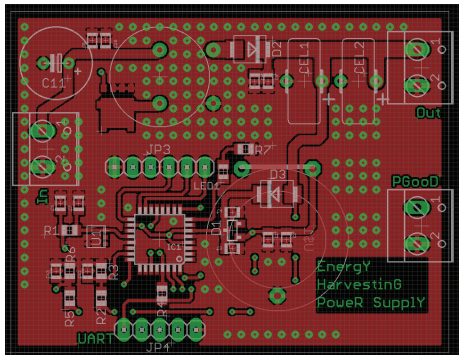


Figure 15: Intelligent energy harvesting power source (PCB layout and parts overview).

5. Discussion about the Contribution for Human Fitness

The presented solution for remote sensors powering is suitable in many cases where we do not need to measure continuous quantities several times per minute, for example, the use of remote sensor networks for monitoring of expansion of oil after sea damage (in Mexican gulf in 2010). For continental areas, a similar usage can be found in monitoring the content of oxygen in fish ponds or lakes in order to be informed about the need to increase oxygen content in water to enable fish to survive in winter periods.

For both mentioned usage areas a 24-hour monitoring cycle is needed. The use of a battery for power in such an HW solution of a remote sensor is uneconomical because there is a need of hundreds or thousands of sensors with a final price of several dollars.

Another example of the application of this solution can be found in monitoring dangerous gas concentration in living environments, monitoring of seismic activities in seismically active places, and oil soil expansiveness in cases of oil tanker accidents. When we are able to precede human contamination (by gas or oil, e.g.), we need to take care of subsequent solution for human disease [9].

It is a paradox that in every industrial manufacture, staff and their working environment are precisely monitored, but when they leave the plant, nobody takes care of them as well as of their living environment. Our solution for remote sensors powering has an ambition to change this trend around.

6. Conclusions

This paper introduced evolving segment of power sources which can utilize free environmental energy, especially thermal energy. Energy harvesting is very quickly penetrating into smart wireless sensor segment. Its advantages are obvious. There is no need to replace or charge batteries which saves time and resources. Placing diagnostics sensors onto places, where standard sensors have no chance, increases safety and can prolong life by early maintenance. Sensing some critical values in large tanks or large water areas can prevent environmental issues. Water environment is especially suitable for heat-based energy harvesters. Relatively constant thermal gradient and high heat capacity of water guarantee good conditions for running sensors array. We can say that energy harvesting is a good option for supplying low power consumption smart sensors for virtually unlimited time [10–15] also in smart cities environments [16–18].

Acknowledgments

The work and the contribution were partially supported by the Project Smart Solutions in Ubiquitous Computing Network Environments, Grant Agency of Excellence, University of Hradec Kralove, Faculty of Informatics and Management and SmartHomePoint Solutions for Ubiquitous Computing Environments, University of Hradec Kralove under the Project no. SP/2013/3. The authors also acknowledge being supported from Cautum Company (http://cautum.cz/). Last but not least, they acknowledge the technical language assistance provided by Stanislava Horakova (University of Warwick).

References

[1] A. L. Rockwood, "Relationship of thermoelectricity to electronic entropy," *Physical Review A*, vol. 30, no. 5, pp. 2843–2844, 1984.

[2] A. H. Fischer, D. Weirauch, and R. A. Schwarzer, "Peltier-cooled solid state drift-chamber detector for energy-dispersive X-ray pole figure measurement and texture mapping," *Materials Science Forum*, vol. 273–275, pp. 263–270, 1998.

[3] D. Vybiral, M. Augustynek, and M. Penhaker, "Devices for position detection," *Journal of Vibroengineering*, vol. 13, no. 3, pp. 531–535, 2011.

[4] P. Cermak, "Thermoelectrical energy and Peltier effect in transmission of solid and fluid aggregate status," *Annalen der Physik*, vol. 26, no. 8, pp. 521–531, 1908.

[5] S. B. Liu, J. Lu, Q. Wu, and Q. R. Qiu, "Harvesting-aware power management for real-time systems with renewable energy," *IEEE Transactions on Very Large Scale Integration (VLSI) System*, vol. 20, no. 8, pp. 1473–1486, 2012.

[6] P. Fiala and P. Drexler, "Power supply sources based on resonant energy harvesting," *Microsystem Technologies*, vol. 18, no. 7-8, pp. 1181–1192, 2012.

[7] A. Richelli, S. Comensoli, and Z. M. Kovacs-Vajna, "A DC/DC boosting technique and power management for ultralow-voltage energy harvesting applications," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 6, pp. 2701–2708, 2012.

[8] F. Ongaro, S. Saggini, and P. Mattavelli, "Li-ion battery-supercapacitor hybrid storage system for a long lifetime, photovoltaic-based wireless sensor network," *IEEE Transactions on Power Electronics*, vol. 27, no. 9, pp. 3944–3952, 2012.

[9] V. Kasik, M. Penhaker, V. Novak, R. Bridzik, and J. Krawiec, "User interactive biomedical data web services application," in *E-Technologies and Networks for Development*, vol. 171 of *Communications in Computer and Information Science*, part 16, pp. 223–237, 2011.

[10] J. Toldinas, V. Stuikys, R. Damasevicius, and G. Ziberkas, "Application-level energy consumption in communication models for handhelds," *Electronics and Electrical Engineering*, vol. 6, no. 94, pp. 73–76, 2009.

[11] O. Krejcar, J. Jirka, and D. Janckulik, "Use of mobile phones as intelligent sensors for sound input analysis and sleep state detection," *Sensors*, vol. 11, no. 6, pp. 6037–6055, 2011.

[12] M. Cerny and M. Penhaker, "Biotelemetry," in *Proceedings of the 14th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC '08)*, IFMBE Proceedings, pp. 405–408, June 2008.

[13] N. Chilamkurti, S. Zeadally, A. Jamalipour, and S. K. Das, "Editorial: enabling wireless technologies for green pervasive computing," *Eurasip Journal on Wireless Communications and Networking*, vol. 2009, Article ID 230912, 2 pages, 2009.

[14] N. Chilamkurti, S. Zeadally, and F. Mentiplay, "Green networking for major components of information communication technology systems," *Eurasip Journal on Wireless Communications and Networking*, vol. 2009, Article ID 656785, 7 pages, 2009.

[15] P. Gál, R. Kilík, T. Špaková et al., "He-Ne laser irradiation accelerates inflammatory phase and epithelization of skin wound healing in rats," *Biologia*, vol. 60, no. 6, pp. 691–696, 2005.

[16] J. Machaj and P. Brida, "Impact of radio fingerprints processing on localization accuracy of fingerprinting algorithms," *Elektronika Ir Elektrotechnika*, vol. 123, no. 7, pp. 129–132, 2012.

[17] W. C. Cheng, J. W. Liou, and C. Y. Liou, "Construct adaptive template array for magnetic resonance images," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, Brisbane, Australia, June 2012.

[18] G. Popek, R. Kowalczyk, and R. P. Katarzyniak, "Generating descriptions of incomplete city-traffic states with agents," in *Foundations of Intelligent Systems*, vol. 122 of *Advances in Intelligent and Soft Computing*, pp. 105–114, 2011.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 952568, 12 pages
http://dx.doi.org/10.1155/2013/952568

Research Article

Using Extension Theory to Design a Low-Cost and High-Accurate Personal Recognition System

Meng-Hui Wang and Po-Yuan Chen

Department of Electrical Engineering, National Chin-Yi University of Technology, Taiping Dist., Taichung City 41101, Taiwan

Correspondence should be addressed to Meng-Hui Wang; wangmh@ncut.edu.tw

Received 7 November 2012; Accepted 17 February 2013

Academic Editor: Chao Song

With the advancement in information technology, personal recognition systems have attracted wide attention. With more options of the recognition systems, the recognition rate and price become very important. This paper used palmprint with the extension method to design a low-cost personal recognition system. First, this paper uses a low-cost webcam to capture the image of palmprints, here the length, slop, and distance of principal line of palmprints can be captured by the image process method. Generally, the devices for capturing hand images should have higher-resolution, so their prices are higher. This paper used a low-pixel and low-cost webcam as the capturer, and it had also a high recognition rate that is equivalent in high resolution devices. The recognition algorithm of this study used extension algorithm for hand recognition and was compared with other traditional algorithms and recognition systems. Finally, the experimental results showed that the method proposed in this study has higher recognition rate than traditional algorithms and proved that low-resolution and low-cost capture tools have a high recognition rate as well.

1. Introduction

The biological recognition technology using human body as features gradually replaced traditional personal recognition technology. The traditional personal recognition systems may easily be lost, stolen, or forgotten; the "biological recognition technology" using human body as features has become a new trend [1]. In the past few years, the biological recognition technology has a lot of progress. This past studied include—face [2], iris [3], fingerprint [4], palmprint [5], and palm geometry [6] are used for identification. There are many types of biological features that can be captured, whether the biological recognition system can be accepted by the public depends on its accuracy, forgery prevention, security, convenience, and speed. The most economical and acceptable feature is hand feature. Hands have unique and stable features, and the convenience and security in recognition are obvious. Therefore, most of studies on biological recognition technology aim at hand technology [7].

Hand features can be divided into three major parts—fingerprint, palmprint, and hand geometry. This study captured the major hand feature, palmprint, for identification. The slopes, lengths, and position distances of three principal lines of palmprints were the main features. Fingerprint and fine palmprint are most likely to become vague because of external factors and aging. The uniqueness of hand geometry is lower than the former two features, and it is an unstable factor. The palmprint is unlikely to be indistinct because of age and external factors; thus, it has high stability and uniqueness. This study used a low-resolution webcam to capture and transfer images to the computer for research. Different capture instruments and different recognition methods were compared and discussed.

2. The Design of the Proposed Recognition System

Figure 1 shows the complete hand image recognition flow. First, the webcam captures the hand image, the palmar contour is determined after binarization of the original image, and the palmprint range to be captured is located. The principal line features of palmprint can be captured after image preprocessing of palmprint range, and then the feature

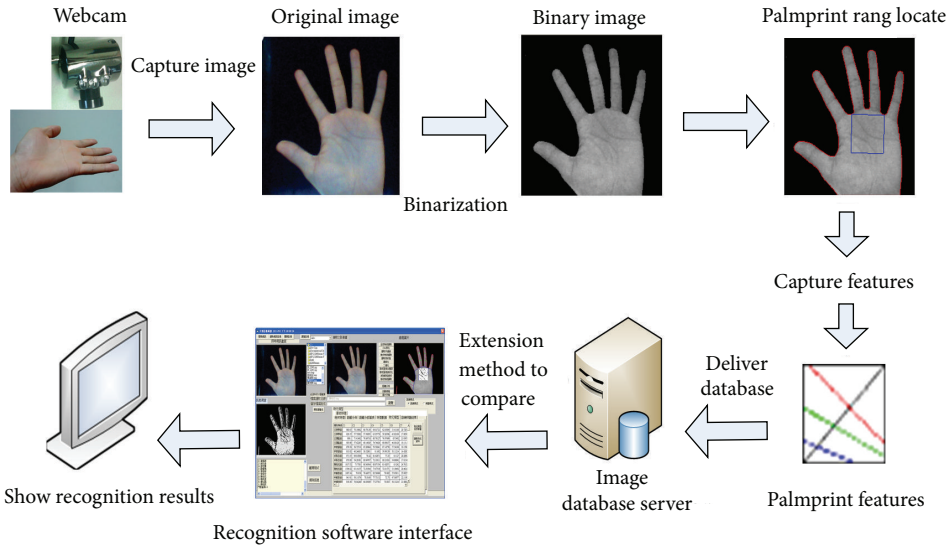


Figure 1: The structure of the proposed palmar recognition system.

Figure 2: Images of hand with contacting or too close fingers.

values are transmitted to the built database. The extension algorithm is used for identification. The results are displayed on the screen.

For capturing the hand image, the subject's fingers should be unfolded naturally, and five fingers cling to the bottom. The fingers should not contact each other or be very close to each other. If the fingers are put together or very close to each other, there will be errors or faults as the captured length, width, and contour features of fingers are influenced, as shown in Figure 2. They cannot be used as data in recognition library, and the database integrity and consistency would be impaired.

The tool for capturing hand images used in this study is a charge-coupled device. It is very sensitive to light; as long as there is slight change in the light and shadow, there will be large differences in the captured feature values, as shown in Figure 3. When the light is too dark, the hand position and size can hardly be identified, as shown in Figure 3(a). When the light is too bright, the palmar image is distorted and the palmprint cannot be seen at all, as shown in Figure 3(c). The optimal light was adjusted after multiple tests, the hand contour was obvious, and the palmprint was clear, as shown in Figure 3(b). There must be a semienclosed box in order to fix the light, and sufficient and stable light is given inside the box. The box provides proper light for capturing hand images in different places at any time, so as to avoid overly dark or bright light that would result in wrong and unsharp hand capture.

3. The Acquired Method of Palmar Images

Since the background noise of hand image is likely to cause misrecognition, the original hand image is separated from the background. The image is grayed, and then the background is expressed as gray-scale value 0 (black). The Otsu's method is used in this paper [8].

If the gray-scale value range of image is $[1, R]$, the gray-scale value is i, the obtained number of pixels is n_i, and $i \in [1, R]$. The total pixel number of various gray scales is $n_1, n_2, \ldots, n_R$, and the total pixel number is T, the probability distribution equation of gray scale i is

$$P_i = \frac{n_i}{T}, \quad P_i \geq 0, \quad \sum_{i=R}^{R} P_i = 1. \tag{1}$$

The threshold t divides the pixels of image into two clusters C_a and $C_b C_a$ denotes the pixel cluster with gray-scale value range of $[1, t]$; C_b denotes the pixel cluster with gray-scale value range of $[t+1, R]$. The probability distribution of the two clusters is W_a and W_b, respectively, and the average of the two clusters is μ_a and μ_b respectively, then

$$\begin{aligned} W_a &= \sum_{i=1}^{t} P_i = W(t), \qquad W_b = \sum_{i=t+1}^{R} P_i = 1 - W(t), \\ \mu_a &= \sum_{i=1}^{t} \frac{iP_i}{W_a} = \frac{\mu(t)}{W(t)}, \qquad u_b = \sum_{i=t+1}^{R} \frac{iP_i}{W_b} = \frac{u_T - u(t)}{1 - W(t)}. \end{aligned} \tag{2}$$

The variance of the two clusters is σ_a^2 and σ_b^2, and respectively, the total variance is σ_w^2:

$$\begin{aligned} \sigma_a^2 &= \sum_{i=1}^{t} (i - u_a)^2 \frac{p_i}{w_a}, \qquad \sigma_b^2 = \sum_{i=t+1}^{R} (i - u_b)^2 \frac{p_i}{w_b}, \\ \sigma_w^2(t) &= w_a \sigma_a^2(t) + w_b \sigma_b^2(t). \end{aligned} \tag{3}$$

(a) Light is too dark (b) Light moderate (c) Light is too bright

Figure 3: Hand image with different lights.

(a) Original image

(b) Image separated from background

Figure 4: Using Otsu's method to separate palm from background.

If there is a t value that minimizes the total variance σ_w^2, this t is the optimal gray scale threshold. If the gray-scale value of image pixels is less than t value, the gray-scale value is set as 0; if it is greater than or equal to t value, the value should not be changed. Equation (4) redistributes gray-scale value to the image to separate the hand image from the background image. Figure 4 shows the difference between

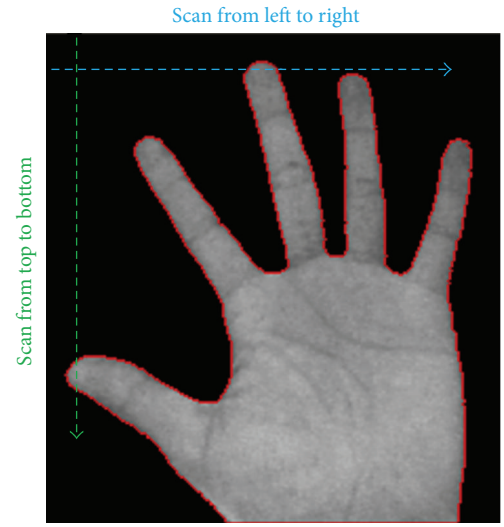


Figure 5: Scanning direction of hand image.

the original image and the Otsu's method processed image. Consider

$$Gr(i,j) = \begin{cases} 0, & \text{if } Gr(i,j) < t, \\ Gr(i,j), & \text{if } Gr(i,j) \geq t. \end{cases} \quad (4)$$

3.1. Detecting Palmar Edge. When the palm is separated from the background by using the abovementioned method, the palmar edge is detected. The pixels are scanned from top to bottom and from left to right to ensure that the full image is scanned. During scanning, if the left and right gray-scale values are greater than 0, the pixel is colored red. When the scan in two directions is finished, the complete palmar edge coordinates can be detected, as shown in Figure 5.

When the palmar contour coordinates are determined, in order to capture geometric features and palmprint range from hand, the tip points and valley points of fingers should be defined in the palmar edge coordinate set S_b. The tip points are the tip coordinates of five fingers, and the valley points are the valley coordinates between fingers, as shown

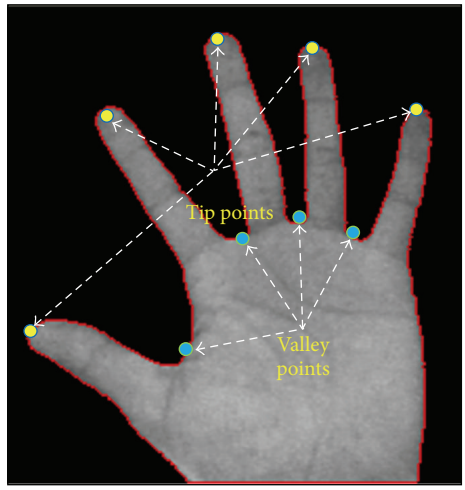


Figure 6: Coordinates of valley points and tip points.

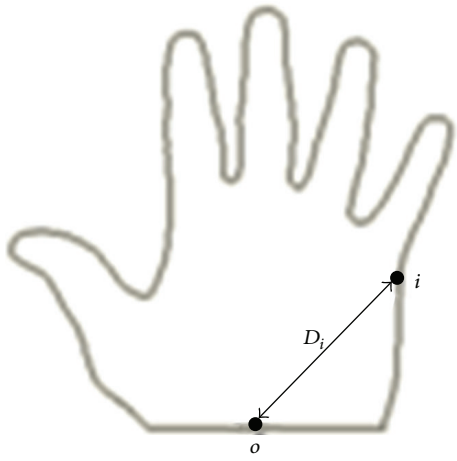


Figure 7: Distance between contour and point *o*.

in Figure 6. To determine the valley points and tip points, a point in the middle of wrist in the coordinate set is set as $o(X_o, Y_o)$, and another point $i(X_i, Y_i)$ is set. The hand contour is moved clockwise from point c and returns to point o again. The distance D_i between point i and point o is recorded in the process, and the coordinate distance between two points in the plane is calculated by using Euclid, X_i and Y_i are coordinate values in coordinate contour, as shown in (5). The distance D_i distribution is shown in Figure 7. Consider

$$D_i = \sqrt{\left(X_i - X_j\right)^2 + \left(Y_i - Y_j\right)^2}, \quad i \in S_b. \tag{5}$$

According to the distance distribution in Figure 8, there are five tip and four valley points, corresponding to five tip points and four valley points of hand contour. The tip points and valley points are determined according to the regular variance in coordinate set S_b and D_i. The principle is briefly described below.

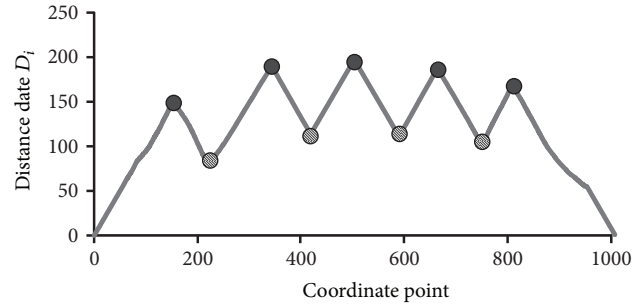


Figure 8: Distribution condition of the distance between palmar contour coordinates and point *c*.

The characteristic of tip points of palmar contour in distance D_i is that the tip point is larger than former and latter coordinate, and the former and latter coordinate distance decreases gradually. The characteristic of valley points in distance D_i is opposite to tip points. The five tip points and four valley points can be observed clearly in the above conditions.

3.2. Palmprint Rang Locate. According to the valley point coordinates (X_{v1}, Y_{v1}) of index finger and middle finger and the valley point coordinates (X_{v2}, Y_{v2}) of ring finger and little finger, two points make a straight line, the distance and slope of the line are k and m, respectively. The new coordinates (X_{u1}, Y_{u1}) are found below the valley point between index finger and middle finger in the same distance, as shown in

$$\begin{aligned} X_{u1} &= X_{v1} - \sqrt{\frac{k^2}{m^2+1}}, \\ Y_{u1} &= Y_{v1} + \sqrt{\frac{[k*m]^2}{m^2+1}}. \end{aligned} \tag{6}$$

The new coordinates (X_{u2}, Y_{u2}) are found below the valley point between ring finger and little finger in the same distance, as shown in (7):

$$\begin{aligned} X_{u2} &= X_{v2} - \sqrt{\frac{k^2}{m^2+1}}, \\ Y_{u2} &= X_{v2} + \sqrt{\frac{[k*m]^2}{m^2+1}}. \end{aligned} \tag{7}$$

The abovementioned four points are connected to form a square, this square is located as the palmprint region [9]. The palmprint feature range is shown in Figure 9. Thus, even if in different palm sizes and at different time points, the required palmprint range can be captured accurately.

3.3. Marginalized Technology. The region image processing is also known as image filter or mask image processing. The

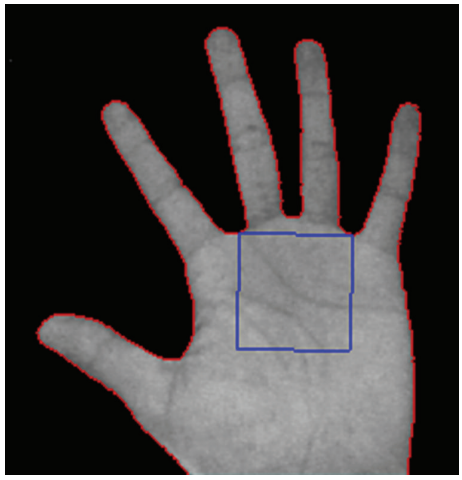


Figure 9: Image of the palmprint range.

−1	0	1
−1	1	1
−1	0	1

Figure 10: Vertical detect realized by 3 ∗ 3 mask.

image range is mostly 3 ∗ 3. The original image is $f(x, y)$, E is defined as 3 ∗ 3 image range mask as

$$E = \begin{bmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \\ w_7 & w_8 & w_9 \end{bmatrix}. \tag{8}$$

T is an operator acting on f, $g(x, y)$ is the processed image, as shown in

$$g(x, y) = E * [f(x, y)]. \tag{9}$$

3 ∗ 3 masks vertical detect to strengthen edge image, as shown in Figure 10. Large gray-scale value change occurs suddenly in the gray-scale image. Therefore, the edge with large gray-scale value change can be obtained by adjusting the threshold, that is, the required palmprint, as shown in Figure 11.

3.4. Binarization. Each pixel of gray-scale value processed image has its own gray-scale value. The image binarization means to set a threshold T first, and each pixel of the original image is detected during image processing. If it is greater than or equal to the threshold, it turns 1 (white); if less than the threshold, it turns 0 (black). The palmprint range turns black and white to show the palmprints, as shown in Figure 12. The theorem is briefly described below.

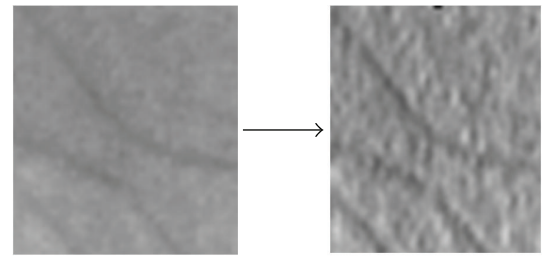


Figure 11: Marginalized palmprint range.

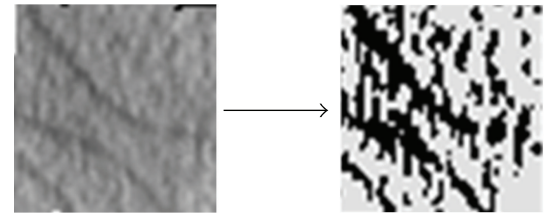


Figure 12: Marginalization of the palmprint range image.

−1	−1	−1
1	3	1
1	1	1

Figure 13: Mask for Sobel operation.

The low pass filter $L(x, y)$ and original image $I(x, y)$ are used for convolution integration, and a threshold T is set. The binary image $O(x, y)$ is obtained, as shown in

$$\begin{aligned} O(x, y) &= 0 \quad \text{if } I(x, y) * L(x, y) < t, \\ O(x, y) &= 225 \quad \text{if } I(x, y) * L(x, y) \geq t. \end{aligned} \tag{10}$$

3.5. Noise Elimination. The 3 ∗ 3 mask Sobel operation is shown in Figure 13. The Sobel operation is for edge enhancement in different directions. When the threshold is adjusted, the noise can be removed out of the binary image, and the palmprint image can be reserved, as shown in Figure 14.

3.6. Line Method. The line method aims to change three principal palmprint lines to three straight lines to reserve, and the noise is eliminated, so as to capture principal palmprint line features directly.

The block is sought for at the beginning: first, the pixel (0) is detected in the Sobel operated palmprint range, and then the pixels adjacent to the whole palmprint range are sought for, it is a block; and then each block is changed into the longest line. At this point, the palmprint range forms

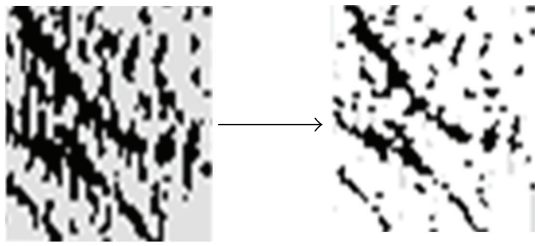


Figure 14: Palmprint Sobel operation processed.

Figure 15: Line map.

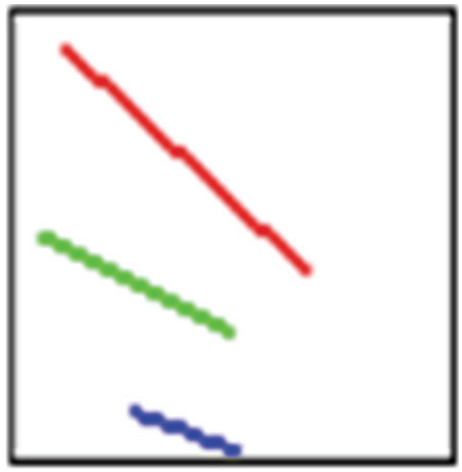


Figure 16: Principal palmprint lines.

multiple straight lines, it is the line map, as shown in Figure 15. The shorter straight lines are removed, as shown in Figure 16, leaving the principal palmprint lines.

3.7. Feature Extraction. The features can be captured directly after image preprocessing, including slope, length, and distance. The captured palmprint features are shown in Table 1 and Figure 17.

Palmprint range perimeter: the valley points of fingers form a square enclosing the range of palmprint features to be captured. The perimeter of this range is classified as a feature as shown in Figure 9.

Table 1: Palmprints features.

(1)	Circumference of the palmprint range
(2)	The slope of the heart line
(3)	The slope of the head line
(4)	The slope of the life line
(5)	The line between valley and heart line
(6)	The line between valley and head line
(7)	The line between valley and life line
(8)	Length of heart line 1
(9)	Length of heart line 2
(10)	Length of head line 1
(11)	Length of head line 2
(12)	Length of life line 1
(13)	Length of life line 2

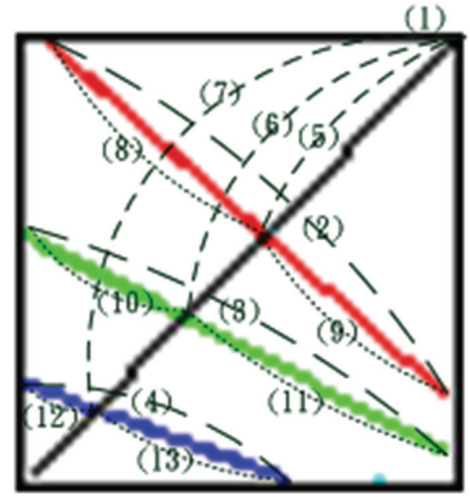


Figure 17: Features of the palmprint lines.

Principal line slope of palmprint: the image preprocessing contains line segmentation, as long as any two pixels are determined in each reserved principal line. The needed principal line slope can be obtained, as shown in Figure 16.

Distance to principal line of palmprint: in the image of principal lines of palmprint in the palmprint range, the vertex of palmprint range used in the paper. As mentioned above, the valley point between ring finger and little finger is one of vertices in palmprint range, connected to the vertex on the cross. This straight line then passes through three principal lines. The distances between the starting point of this straight line—the valley point between ring finger and little finger and three principal lines are three palmprint features.

Two lengths of principal line: the three principal lines in the palmprint range are extended and auxiliary lines are drawn. The three principal lines are cut into two, then each principal line has two length features. The different positions and angles of principal lines influence the data significantly.

4. The Proposed Recognition Method

4.1. Extension Method. The elementary theory of extension is the extension theory. The pillars of extension theory are matter-element theory and extension set theory, using matter-element transformation to solve contradiction and incompatibility problems. The extension set uses extension model to handle subjective and objective contradiction problems of traditional mathematics and fuzzy mathematics. The range of fuzzy set is extended from $\langle 0, 1\rangle$ to $\langle -\infty, \infty\rangle$ by using correlation function [10–12].

In the extension theory, elements of object consist of the name N of object R, magnitude v of feature c, called three elements of matter-element. The matter-element describing objects is shown in (11):

$$R = (N, c, v). \tag{11}$$

Generally, an object has multiple characteristics, if an object $R = (N, c, v)$ is multidimensional matter-element, and has n characteristics, the characteristic vector can be expressed as $C = [c_1 \quad c_2 \quad \cdots \quad c_n]^T$, and the magnitude vector can be expressed as $V = [v_1 \quad v_2 \quad \cdots \quad v_n]^T$. The expression of multi-dimensional matter-element is

$$R = (N, c, v) = \begin{bmatrix} N, & c_1, & v_1 \\ & c_2, & v_2 \\ & \vdots & \vdots \\ & c_n, & v_n \end{bmatrix}. \tag{12}$$

The traditional classical set uses 0 and 1 magnitudes concept to describe objects with or without some characteristic, whereas the fuzzy theory uses membership function to describe the fuzzy degree in fuzzy set range $\langle 0, 1\rangle$. The extension set uses the real number of set range extended to $\langle -\infty, \infty\rangle$ to represent the degree of characteristics.

If U is domain, and any element u in U and $u \in U$ has a corresponding real number, $K(u) \in \langle -\infty, \infty\rangle$, then the extension set is defined as

$$A = \{(u, y) \mid u \in U, y = K(u) \in (-\infty, \infty)\}, \tag{13}$$

where $y = K(u)$ is the correlation function of extension set A, and $K(u)$ is the correlation grade of u with extension set A, the range is $\langle -\infty, \infty\rangle$. The extension set A in domain U can be expressed as

$$A = A^+ \cup A^0 \cup A^-, \tag{14}$$

where A^+, A^0, and A^- are positive field, negative field, and zero boundary in extension set, respectively, expressed as

$$\begin{aligned} A^+ &= \{(u, y) \mid u \in U, y = K(u) > 0\}, \\ A^0 &= \{(u, y) \mid u \in U, y = K(u) = 0\}, \\ A^- &= \{(u, y) \mid u \in U, y = K(u) < 0\}. \end{aligned} \tag{15}$$

4.2. Recognition Procedure. The matter-element model should be built before recognition. The extension classical domain is set, and then the object R is divided into k grades of set of values, called extension classical domain of various sets. N_k $(k = 1 \sim n)$ represents matter-element names of n grades, all the characteristics of the matter-element name are c_i $(i = 1 \sim j)$. The range of characteristics is represented by x_{kj}, the characteristic size is x^U representing the upper limit of the matter-element set, and x^L is the lower limit of the matter-element set. The upper and lower limits of matter-element are called classical domain.

The matter-element model can be used for recognition, as shown in Figure 18. The recognition procedure of extension algorithm is described below.

Step 1. Read the built matter-element model as

$$R_k = \begin{bmatrix} N_k, & c_1, & \left\langle x_{k1}^L, x_{k1}^U \right\rangle \\ & c_2, & \left\langle x_{k2}^L, x_{k2}^U \right\rangle \\ & \vdots & \vdots \\ & c_j, & \left\langle x_{kj}^L, x_{kj}^U \right\rangle \end{bmatrix}, \quad k = 1, 2, \ldots, n. \tag{16}$$

Step 2. Set up neighborhood domain as

$$R = \begin{bmatrix} N, & c_1, & \left\langle x_1^M, x_1^N \right\rangle \\ & c_2, & \left\langle x_2^M, x_2^N \right\rangle \\ & \vdots & \vdots \\ & c_j, & \left\langle x_j^M, x_j^N \right\rangle \end{bmatrix}. \tag{17}$$

The neighborhood domain is the total range of whole characteristic set in all matter-element sets, the upper and lower limits of maximum classical domain are determined as (18). The upper limit and lower limit of neighborhood domain are determined, which are x^M and x^N as (19). Consider

$$\begin{aligned} x_k^U &= \max\left[x_{ki}^U\right], \quad i = 1 \sim j, \\ x_k^L &= \min\left[x_{ki}^L\right], \quad i = 1 \sim j, \end{aligned} \tag{18}$$

$$\begin{aligned} x_k^M &= x_k^U + \left(\frac{x_k^U - x_k^L}{2}\right), \\ x_k^N &= x_k^L - \left(\frac{x_k^U - x_k^L}{2}\right). \end{aligned} \tag{19}$$

Step 3. Read data to be tested as (20).

Where N_q is the name of matter-element to be tested, and the characteristic value of characteristic c_i of N_q is x_i

$$R_q = \begin{bmatrix} N_q, & c_1, & x_{q1} \\ & c_2, & x_{q2} \\ & \vdots & \vdots \\ & c_j, & x_{qj} \end{bmatrix}. \tag{20}$$

Step 4. Calculate correlation functions of matter-elements.

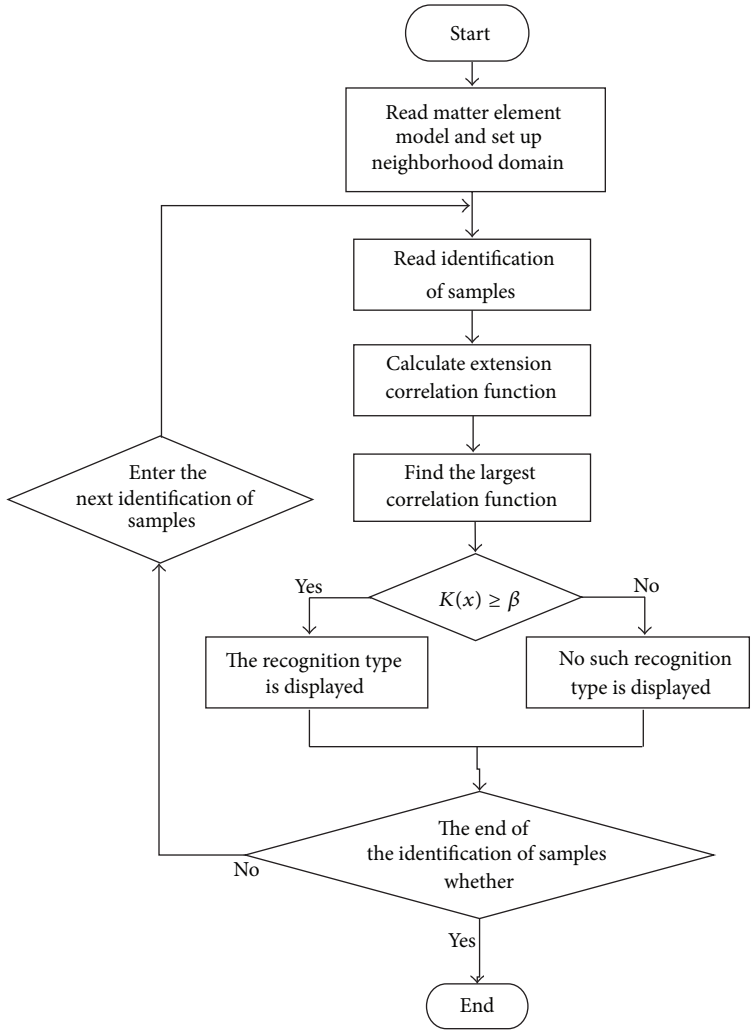


Figure 18: Flow chart of identification method.

In the extension theory, the classical domain $X_o = \langle x^L, x^U \rangle$ and neighborhood domain $X_P = \langle x^M, x^N \rangle$ are two intervals in real number field $\langle -\infty, \infty \rangle$, the interval X_o belongs to interval X_P. If x is one point in real number field, the correlation function is defined as

$$D(x, X_o, X_P) = \begin{cases} \rho(x, X_P) - \rho(x, X_o), & x \notin X_o, \\ -1, & x \in X_o, \end{cases}$$
$$K(x) = \frac{\rho(x, X_o)}{D(x, X_o, X_P)}, \tag{21}$$

where

$$\rho(x, X_o) = \left| x - \frac{x^L + x^U}{2} \right| - \frac{x^U - x^L}{2},$$
$$\rho(x, X_P) = \left| x - \frac{x^M + x^N}{2} \right| - \frac{x^N - x^M}{2}. \tag{22}$$

$D(x, X_o, X_P)$ is the correlation grade of x with X_o and X_P. The correlation function is shown in Figure 19. If $K(x) \geqq 0$, it is the degree of x belonging to X_o; if $K(x) \leqq 0$, it is the degree of x not belonging to X_o.

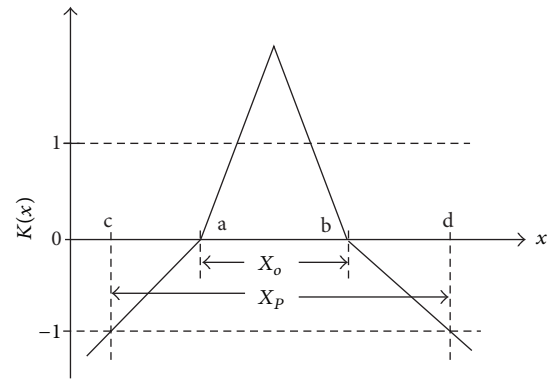


Figure 19: Extension correlation function.

Step 5. Find out the maximum correlation function, that is, maximum correlation grade.

Table 2: Experimental data.

	Database patterns	Not database patterns
Number	25	10
15 images each person	375 images	150 images
for recognition/for training	250 images for training/125 images for recognition	150 images for recognition

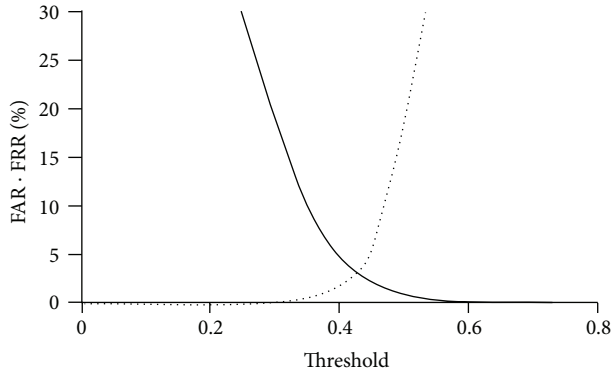


Figure 20: FAR and FRR distribution in extension theory.

Step 6. The recognition type is displayed when the maximum correlation grade is greater than or equal to threshold β; on the contrary, no such recognition type is displayed.

During recognition, the recognition system searches for the maximum correlation grade in the optimal matter-element model according to the test matter-element and takes it as the identity of the person. There should be a threshold to review the maximum correlation grade, so as to avoid any identity outside the database entering the database.

As for the setting of optimal threshold, first, an initial threshold is set, and then the false rejection rate (FRR) and the false acceptance rate (FAR) values at this threshold are calculated. FRR is the recognition rate of system misidentifying a person inside database as one outside database; FAR is the recognition rate of system misidentifying a person outside database as one inside database. The sample of false rejection is set as N_r, the sample of false acceptance is set as N_a, the total test sample is set as N_t, and the FRR and FAR are shown as

$$\begin{aligned} \text{FRR} &= \frac{N_r}{N_t}, \\ \text{FAR} &= \frac{N_a}{N_t}. \end{aligned} \tag{23}$$

The threshold is adjusted slowly to observe the variance in FRR and FAR, as shown in Figure 20. When FRR equals to FAR, this threshold is the optimal threshold. The threshold is set as 0.417 in this paper.

Step 7. If data recognition is completed, stop; otherwise return to Step 2.

Figure 21: The semienclosed box.

5. Testing Results and Discussion

5.1. Hand Image Recognition System Architecture. Since the capture tool used in this paper is a charge-coupled device-webcam, which tends to be influenced by light, the capture structure is a semienclosed space.

For hand image capture, the hand is put in the semienclosed space with palm up, as shown in Figure 21. A 35° downward light is brightest and sharpest, and the hand can lie on the bottom of the enclosed space gently. Even lying the hand for a long time does not make the subject uncomfortable, a webcam was used to shoot downward.

In order to avoid the influence of ambient light, an LED lamp was fixed inside the semienclosed space to give fixed light. The interior was pasted with light reflecting paper, as shown in Figure 22, so that the light spreads over inside of the semienclosed space uniformly, and the stability of images captured by webcam is higher. In addition, there were two fixed protrusions at the bottom of the semienclosed box, namely, in the index finger-middle finger and ring finger-little finger valley points, so as to fix the distance between palm and webcam. The entire capture device was thus finished.

Besides a webcam, the recognition system proposed in this study requires a computer with Visual Basic 6.0 program to process hand images, as shown in Figure 23. The entire image processing, from starting recognition to the end, takes 5 sec.

5.2. Experimental Data. In order to prove the recognition capability of the method proposed in this study, the experiment chose 25 males and 10 females as subjects, totaling 35

Table 3: Recognition results.

Recognition methods	Learning times	Learning accuracy	Testing accuracy
Proposed method	0	n/a	91%
K-means clustering	0	n/a	63%
Backpropagation neural network (13-20-20)	1000	89%	87%
Backpropagation neural network (13-22-20)	1000	91%	89%

Table 4: Comparison among different recognition systems.

	Price U.S.$	Accuracy	Speed (sec.)	Advantages	Disadvantages
 The proposed system	About 1000.	91%	5 s	(1) Low price. (2) Noncontact type without hygiene problems. (3) High uniqueness.	(1) Recognition rate can be further improved. (2) Difficult to identify palm with dirt.
 Thermal image system	About 7,000.	92%	5 s	(1) Free from light. (2) Infrared image capture is free from background interference.	(1) High price. (2) Influenced by temperature. (3) Low uniqueness of captured hand geometry.
 Iris system	About 11,000.	98%	15 s	(1) High accuracy. (2) Difficult to be copied. (3) Iris characteristics are unlikely to change.	(1) High price. (2) Slow recognition. (3) Infrared scan may injure eyes.
 Finger geometry system	About 4,000.	90%	3 s	(1) Quick recognition. (2) Easy technology.	(1) Low uniqueness. (2) Likely to be copied. (3) Contact type with hygiene problems.

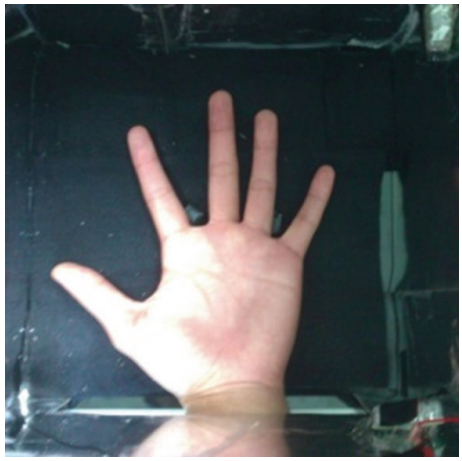


Figure 22: Inside semienclosed box.

Figure 23: Man-machine interface of the proposed system.

subjects, and 20 of them were present in the database. In addition, as the recognition algorithm of this experiment should run in two stages—training mode and recognition mode, the data of all the subjects were divided into training samples and recognition samples. The data of subjects captured in the recognition mode of Stage 1 were training samples, as the weight of each data was determined by recognition algorithm, so as to complete the optimum matter-element model. The recognition mode of Stage 2 compared the hand data outside the completed matter-element model with previously built matter-element model, to check whether the identity was inside the database, if yes, the identity of the subject could be indicated.

The images of each subject's hand were captured at three different times for objectivity and reality, 5 images each time; thus, each subject had 15 images, 10 of them were training samples, and 5 were recognition samples. There were 250 images as training samples for building the matter-element model, and the rest of 275 images were used to test the recognition accuracy rate of this recognition algorithm, as shown in Table 2.

The traditional methods were compared with the proposed method, as shown in Table 3. It is observed that the extension theory still has an accuracy rate of 91% even at low resolution, higher than other methods, and there is no need for learning, the feasibility of extension theory is proved.

In addition, different recognition systems are compared in Table 4, all of them are used for biological recognition, as their recognition rates are higher than 90%, the price becomes the key point. All the recognition systems on the market need an additional computer, and their prices are much higher than the recognition system proposed by this study. Thus, their acceptability in the market would be influenced greatly. Since the public is not willing to pay a high price for the entrance card, a low-price recognition system with high recognition rate would be accepted by the market.

6. Conclusions

The proposed recognition system is consisted of a webcam and a semienclosed box for fixing light. Visual Basic 6.0 is used to process hand images and capture features for identification. First, when the webcam captures the palmprint features, the image should be preprocessed before feature extraction. As the image is inside the semienclosed box, it is free from ambient light interference. The user only needs to lay his hand open naturally for photographing. The experimental results showed that the accuracy rate of the extension recognition algorithm in this study is 91%, which is higher than traditional neural algorithm. This new approach merits more attention, because the low-cost device deserves serious consideration as a tool in palmar recognition problems. We hope this paper will lead to further investigation for industrial applications.

References

[1] R. W. Ives, Y. Du, D. M. Etter, and T. B. Welch, "A multidisciplinary approach to biometrics," *IEEE Transactions on Education*, vol. 48, no. 3, pp. 462–471, 2005.

[2] S. Z. Li, R. F. Chu, S. C. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 627–639, 2007.

[3] C. T. Chou, S. W. Shih, W. S. Chen, V. W. Cheng, and D. Y. Chen, "Non-orthogonal view iris recognition system," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 3, pp. 417–430, 2010.

[4] D. Zhang, F. Liu, Q. Zhao, G. Lu, and N. Luo, "Selecting a reference high resolution for fingerprint recognition using minutiae and pores," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 3, pp. 863–871, 2011.

[5] D. Zhang, G. Lu, W. Li, L. Zhang, and N. Luo, "Palmprint recognition using 3-D information," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 39, no. 5, pp. 505–519, 2009.

[6] G. Zheng, C. J. Wang, and T. E. Boult, "Application of projective invariants in hand geometry biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 758–768, 2007.

[7] J. Doi and M. Yamanaka, "Discrete finger and palmar feature extraction for personal authentication," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 6, pp. 2213–2219, 2005.

[8] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on System, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[9] Y. Wang, Q. Ruan, and X. Pan, "An improved square-based palmprint segmentation method," in *Proceedings of International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS '07)*, pp. 316–319, December 2007.

[10] M. H. Wang, K. H. Chao, G. J. Huang, and H. H. Tsai, "Application of extension theory to fault diagnosis of automotive engine," *ICIC Express Letters*, vol. 5, no. 4 B, pp. 1293–1299, 2011.

[11] M. H. Wang, K. H. Chao, W. T. Sung, and G. J. Huang, "Using ENN-1 for fault recognition of automotive engine," *Expert Systems with Applications*, vol. 37, no. 4, pp. 2943–2947, 2010.

[12] M. H. Wang, Y. K. Chung, and W. T. Sung, "Emerging intelligent computing technology and applications," in *The Fault Diagnosis of Analog Circuits Based on Extension Theory*, vol. 5754 of *ICCC Lecture Notes in Computer Science*, pp. 735–744, Springer, Berlin, Germany, 2009.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 578563, 11 pages
http://dx.doi.org/10.1155/2013/578563

Research Article

A Scalable and Privacy-Aware Location-Sensing Model for Ephemeral Social Network Service

Yongqiang Lyu,[1] Dezhi Hong,[1] Ying Wang,[1] Yinghong Hou,[1] Zhengwen Yang,[1] Yu Chen,[1] Yuanchun Shi,[1] and Alvin Chin[2]

[1] *Research Institute of Information Technology, Tsinghua University, Beijing 100084, China*
[2] *Nokia, Dong Huan Zhong Lu, Daxing, Beijing 100176, China*

Correspondence should be addressed to Yongqiang Lyu; luyq@tsinghua.edu.cn

Received 10 November 2012; Revised 1 February 2013; Accepted 17 February 2013

Academic Editor: Chao Song

Social network services (SNSs) are developing at an explosive speed, which makes it easy for people to be closely connected. As a result, a new SNS type, ephemeral social network (ESN), is emerging to capture the ephemeral interactions and meetings that occur in environments such as conferences and workplaces. Most ESNs require the positioning for capturing the physical proximity between people, which impacts much the scalability, privacy protection, and the cost of the system. In this paper, we propose a scalable location-sensing model based on RFID-sensing architecture for ESN in consideration of four aspects of requirements, that is, the usability, QoS, scalability, and privacy. The model includes the perspectives of the privacy, architecture, deployment, and positioning algorithms, which can meet the four key requirements. A prototyping conference-type ESN system was also developed in this paper and was tested in the ACM UbiComp 2011 conference, which verified that the model works well with good scalability, low cost, and customized privacy protection.

1. Introduction

Social network services (SNSs) have become ubiquitous and been developing at an unprecedented speed to connect people with each other. The use of SNS also appears in many specific ephemeral scenarios and occasions, like academic conferences and workplaces, which is known as the ephemeral social network service (ESNS) and gradually show its necessity. ESNS could provide each user with the location of peer users and their activities based on locations. As many other indoor ubiquitous applications, the positioning (location-sensing) system becomes the key component to obtain position-related information of an ESNS since most ESNSs are position-sensitive applications and are mostly used indoors, where global positioning system (GPS) or assisted GPS (AGPS) fails to work effectively and accurately.

When choosing techniques to fulfill the requirements of the location-sensing system for an ESNS, we should take into account the following four perspectives of requirements as Figure 1 shows.

(i) Usability. This part involves the operability, flexibility, cost, and convenience in terms of system deployment, configuration, and use. These issues must be evaluated: how the operability of the system is, whether it is complex to set up or make adjustments, whether the system has any special requirements on the devices, what the cost is, and whether the users need to wear extra specific apparatus or devices. Usability of the system determines how much it could be accepted by users.

(ii) Quality of Service (QoS). This part consists of two aspects, that is, the power consumption and positioning accuracy. The power consumption is a key issue because ESNS would be utilized mostly in mobile scenarios, such as on mobile phones, PADs, and wearable sensors, all of which have limited battery capacities. Therefore, the power consumption of the location-sensing system must be controlled within

the accepted level. As for the accuracy, the location-sensing system should achieve certain resolutions according to the application requirements: in some cases it may be in the order of meters, while in some other cases it may be in the order of centimeters; it could be tuned automatically or manually by users according to different demands of using scenarios, for example, due to privacy policies.

(iii) Scalability. Most positioning technologies could be working well when supporting tens of users. However, when the amount of the users becomes larger, for example, hundreds or more, some kinds of techniques would fail to meet the requirement of scalability, for example, degrading in performance, usability, QoS, and so forth.

(iv) Privacy. Since a positioning system could achieve certain accuracy of the user's location information, the user's locations might be exposed and the system should involve the policy of a controllable privacy hierarchy from the perspectives of the user security levels, location-sensitive area types, and the activity-based accuracy requirements. All these are differently configured for each user because it contains users serving different roles of initiator, common user, evaluator, and analyst [1] for a specific ESNS; it covers effective areas providing services of different activities and holds several kinds of activities that can cause variation in using the same area. We should combine these concerns to standardize our model to meet the requirements of each privacy hierarchy, respectively.

In this paper, we firstly make comparisons among several state-of-the-art location-sensing technologies with respect to the four aspects of requirements and then present a scalable and privacy-aware location-sensing model dedicated for the ESNS under several typical scenarios. The reminder of the paper is organized as follows. Section 2 gives a comparative overview of the current positioning technologies and some related work that has been undertaken. Section 3 discusses the privacy issues for a standard ESNS and the corresponding privacy-aware and scalable model is shown in Section 4. In Section 5, the ESNS system we developed, the Find & Connect UbiComp at the UbiComp 2011 conference, is shown as an experimental verification to our location-sensing model.

2. Related Work

Many technologies are currently utilized to provide indoor positioning services such as image, infrared, Wi-Fi, RFID, and UWB. In this section, we make a discussion on the advantages and disadvantages of them to the four requirements respectively as listed in what follows.

2.1. Unobtrusive Methods. The positioning methods, such as image-based [2], infrared [3], ultrasonic [4], and inertial methods [5], do not need any additional accessories worn on the users and thus are classified as unobtrusive methods in this paper. Recently, Microsoft Kinect [6, 7] becomes another

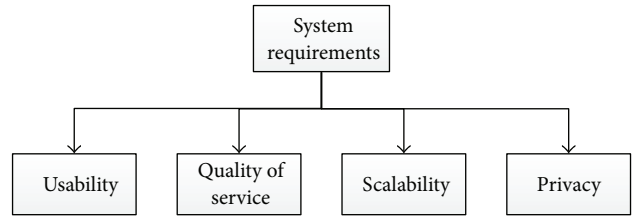


Figure 1: The system requirements of an ESNS.

option for indoor positioning by exploiting the image depth. However, the common disadvantages [8] of these methods are: (1) inaccurate positioning results, (2) inability to sense subjects due to the blockages of the walls, ceilings, and so forth, and (3) weak scalability to the large user volume due to the cost and the difficulty in user recognition.

2.2. ISM Radio-Frequency Methods. The industrial, scientific, and medical (ISM) bandwidths are reserved for daily communications in many countries and regions, for example, the Bluetooth, Wi-Fi, ZigBee, and RFID of 2.4 GHZ. These communication methods can also be employed in indoor positioning by utilizing the radio signal strength (RSS), such as Wi-Fi positioning [9], RFID positioning [10], and ZigBee positioning [11]. The most obvious advantage of using ISM communication methods is that such kind of system could share the same infrastructure as that of existing indoor wireless networks and needs little pre-setup. However, the long-term access to the ISM networks (such as Wi-Fi access points) by mobile devices or user tags may be power-consuming and requires extra ability of wireless access on user devices. The accuracy of such positioning systems is also relatively low. In addition, the costs of some communication chips are often high due to the intellectual properties (IPs) of the protocol stacks, such as Wi-Fi and Zigbee.

2.3. Wideband Radio-Frequency Methods. Ultra wideband (UWB) [12, 13] positioning utilizes the advantage of the wide band of the communication and can achieve high positioning accuracy with the error of couple centimeters [14]. However the chip of UWB is expensive and the power consumption is also high. Recently, there is another alternate method named chirp spread spectrum (CSS), which operates in narrower bandwidth than UWB but can achieve comparable positioning accuracy and ultra low power [15]. Both UWB and CSS positioning can employ time of arrival (TOA), time difference of arrival (TDoA), and radio signal strength (RSS) to sense the locations and TOA/TDoA can be more accurate than RSS. Nevertheless, the cost and power issues of UWB and CSS are still the key factors determining the usability.

2.4. Active RFID. In comparison with other related technologies, the active RFID positioning systems can gain the following advantages over the others.

Usability. With regularly distributed RFID readers collecting active tag signals, a central server is enough to process the

data and calculate the location for hundreds of tags. The system architecture is therefore simple and easy to deploy and maintain.

Although the cost of a RFID reader is relatively a little higher than a Wi-Fi access point (AP), we can regard the installation of the RFID readers as infrastructure and they can actually cover much wider area than Wi-Fi APs (up to 150-meter radius), which means we do not need to set up so many readers for the same coverage as other positioning methods especially in the scenarios of low resolution. Another important advantage of RFID is that the RFID tags are cheap (e.g., cheaper than Wi-Fi) and also easy to wear. When deploying ESNS in, for example, the academic conferences, the organizers only need to distribute the RFID tags to attendees and retrieve the tags after the activity, which provides much convenience and efficiency to both the organizers and users.

Quality of Service. For power consumption, an active RFID tag with a normal battery can theoretically last for 3~5 years at the medium data rate under the current technology node; while the Wi-Fi tags or phones suffer the remarkable power consumption which degrades the QoS. As for the accuracy, active RFID systems can reach high accuracy using algorithms with proper distribution of readers and reference tags (can be higher than Wi-Fi) and can be adjusted to lower resolutions using simple algorithms with fewer readers.

Scalability. A single RFID reader supports flexible user volume, ranging from dozens to hundreds, which is enough to cover most ephemeral social scenarios. The positioning system made up of several RFID readers could support up to 1 k users simultaneously with only average computing resources.

Privacy. The resolution of location information provided should be controlled with respect to privacy requirements from users. The strict privacy policies should be set to sensitive scenarios by default, such as rest rooms, dressing rooms, and smoking area, and the RFID-based accuracy-adaptive algorithms and deployment can also meet this requirement easily with low cost.

In summary, a RFID-based positioning system is suitable for a scalable, low-cost, QoS-controllable and privacy-protecting ESNS. The practical model is described as follows.

3. Privacy

3.1. Location Privacy in ESNS. According to Duckham and Kulik [16], the concept of location privacy can be generally defined as a special type of information privacy which concerns the claim of individuals to determine for themselves when, how, and to what extent location information about them is communicated to others.

It suggests subtle preferences of revealing location data in different forms for common positioning services [17].

(i) When. A subject may be more concerned about his current or future location being revealed than locations from the past.

(ii) How. A user may be comfortable if friends can manually request his location but may not want alerts sent automatically whenever he enters a toilet or a bar.

(iii) Extent. A user, in some cases, may rather have his location reported as an ambiguous region rather than a precise point.

As for indoor positioning, especially for ephemeral social network service, we should consolidate the guidelines above as follows.

(i) When. As ESNS explores possibility to know people in physical proximity where each encounter happens and lasts at random, users may be more concerned about their past and current location as well as those of others in certain connection with themselves, which should be provided by the system with a possible form of recommendation pushing.

(ii) How. The users should have sufficient controllability over their own position data in the sense that (1) they can view and edit private location records, and (2) they can manually decide whom to share the location data with and how much information to reveal. There should also be data-mining recommendation algorithms based on physical proximity and social connections in the system to automatically raise attention among those people who should have been connected but not yet connected in real life.

(iii) Extent. The users may still want their locations reported as ambiguous regions rather than precise points in most public places. However, in specific areas holding social gatherings such as parties and conferences, the participants would raise their tolerance for location privacy but care more about knowing people with a clear overview of "who and where" in the room. Therefore, as for an ESNS, such areas require higher accuracy in positioning, while other public places such as passageways and rest rooms need only ambiguous or even no position reports.

With the concept of location privacy refined for ESNS applications, now we can explore more about the location and privacy issues in the following section.

3.2. Privacy Concerns from the User. In previous studies, researchers have found that most users do not care about their location privacy as much as what the computer scientists have designed and advocated. According to a study of 55 interviews with subjects in Finland, Kaasinen [18] found that "The interviewees were not worried about privacy issues with location-aware services." However, he added, "It did not occur to most of the interviewees that they could be located while using the service." In another study of university students, Danezis et al. [19] asked 74 undergraduates how much they

would have to be paid to share a month's worth of their location data. The median price was £10, or £20 if the data were to be used commercially. For most indoor poisoning services, navigation combined with preference information pushing is a common purpose. Users are willing to reveal their location information to the system with the attention focusing on the services, and the services they expect to get from the system often ignore the possibility of location leak to unauthorized attackers.

Nevertheless, the privacy concern is more severe in the case of ESNS. Judging from the feedback from our own system Find & Connect UbiComp, a conference-based ESNS, some of the users had serious concerns about how the organizers would deal with all the location data. So the privacy concerns arise again under the ESNS circumstances such as the conference, where users have direct knowledge that their locations may be seen by others after permission, and some private information such as with whom they stay in proximity could be revealed. Therefore, we emphasize on "how" we should carefully define user activities in ESNS, as discussed in the previous parts, to protect privacy and compromise positioning accuracy to a reasonable extent.

3.3. Service Requirements. From the discussion above, the model of indoor positioning services in ESNS requires the following basic functions:

3.3.1. Security Access Levels. Considering the different statuses of the users, they have accordingly different levels of access to the position information of the targets. For common users, they can only access the locations of those who willingly share their information; as for authorized guardians like system administrators, they can have higher control over tracking and monitoring the targets. But their free access to all the data might also raise privacy and security issues and we will further discuss it in Section 4.

3.3.2. Controllability over Personal Data. Common users should personally decide whether to share their location information with people at the same or lower security access levels but would stick to the supervision from the higher levels to a legal extent. As for those controllable information, they can also treat different people in different ways, namely, they can specifically choose whom to share the information with.

3.3.3. Availability in History Records. Both for the system administrators to monitor and maintain the system and for common users to retrospect on places they have been to, it should be available for them to access the history data. Nevertheless, it is also constrained with respect to their security access levels.

3.3.4. Accuracy Variation for Specific Areas. Privacy limits the accuracy of the indoor positioning due to various functional zones/area in a given ephemeral social network space. On the one hand, for the highly gathering places where the social activities take place, the accuracy of location sensing could be high because users normally care more about the social activities in the willingness to be connected with other people in the same event; the position information is regarded open and ready to be achieved by others. The possible threats of position information leak are not the issues. On the other hand, for the public places which are more privacy-sensitive, such as dressing rooms, rest rooms, and smoking places, where the participants may need personal activities, the positioning accuracy should be lowered or obscured to protect user's privacy.

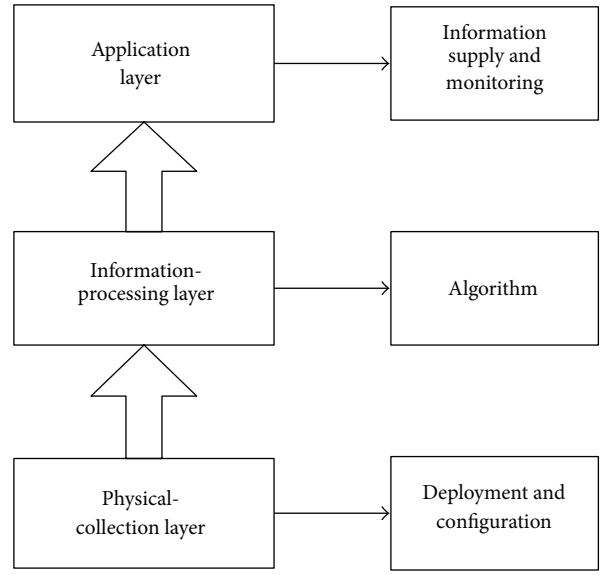


Figure 2: The architecture of the scalable privacy-protecting ESNS location-sensing system.

4. Location-Sensing Model

In this section, we firstly introduce the general architecture and the functions of each part of this model. Then we present the deployment model that could be utilized in general cases. In the following parts, we describe the details of the positioning algorithms and the privacy control, respectively.

4.1. Architecture. As Figure 2 shows, the architecture of the model from top to bottom is the application layer, information-processing layer, and physical-collection layer. The relationships between the adjacent layers and the respective content are described in Figure 2.

4.1.1. Application Layer. The application layer is the top-most layer which handles the ubiquitous ESNS services provided to users. Some features and services in this layer depend on the position-related information and the data provided by the adjacent lower layer through stipulated interfaces. In the implementation of this paper, the Find & Connect application is realized in this level based on the interfaces of the information-processing layer. For some extended applications, the developing application programming interfaces

(APIs) can also be built in this layer to support more applications.

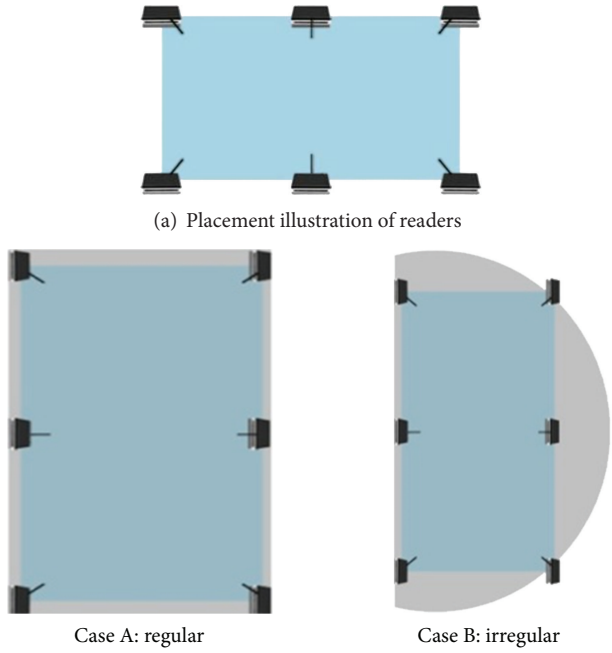
4.1.2. Information-Processing Layer. This intermediate layer implements certain algorithms to process and analyze the raw data transported from the physical layer and extracts necessary characteristics and information provided to the upper application layer. For example, the filtering algorithms can be placed in this layer to support the subject-tracking services in the application layer. Some filtering algorithms may need intensive data support from the databases, and the databases are thus also classified into this layer.

4.1.3. Physical-Collection Layer. This layer contains all the sensors and devices used in the positioning system; it handles the deployment as well as the configuration of the hardware systems. Physical-collection layer directly collects signals and related data transmitted from the devices, such as from the tags, which are processed in the upper information-processing layer. Some communication middleware and protocols are often needed in this layer to parse the physical signals from the sensors and devices into data format.

4.2. Deployment Model. We have two deployment schemes to meet the requirements of positioning accuracy, that is, the high-accuracy scheme and the low-accuracy scheme.

4.2.1. High-Accuracy Scheme. In average ESNS scenarios such as academic conferences, office buildings and public areas, the requirement of high positioning accuracy is just in the level of around one meter; the resolution of centimeter-order which can be achieved by UWB is often not necessary. LANDMARC [10] is a widely used active RFID-based positioning algorithm, which has low deployment complexity and stable performance. Therefore we adopt the positioning algorithm based on LANDMARC [10] and a hardware system consists of RFID readers and a certain number of reference tags. The LANDMARC algorithm is based on the idea that the user-tag location could be calculated by employing the known locations of the reference tags in the same space. This algorithm was also verified working well with reasonable accuracy in our experiments.

The general placement scheme of the RFID readers and reference tags is the core issue of the entire deployment model, and we propose a general scheme for it which can be employed in most common scenarios. We set 4 anchoring RFID readers with omni directional antennas in the target space with their distributive positions meeting the following requirements: (1) all four readers head towards the center of the room; (2) the four readers' positions form a rectangle-like quadrilateral; (3) the quadrilateral covers the target space to the greatest extent; and (4) each reader should minimize the blockage in its way of signal collection. We also set several assistive readers placed at the midpoint of the two longest edges of the rectangle formed by the four anchoring standard readers, as the example in Figure 3(a) shows, which can help improve the coverage of the radio frequencies in large space or places with blockages. Particularly, when the amount of the tags detected by any of the assistive readers surpasses any of the anchoring ones, the assistive reader should take over the anchoring reader to work and collect the data for positioning. Figure 3(b) shows other two placement examples to illustrate the deployment scheme for different target space shapes (vertical view).



(a) Placement illustration of readers

(b) Placement schemes in general cases

Figure 3: The deployment model.

Particularly in cases of irregular target spaces, the regular placement of the readers is recommended in this paper because the regular placement can lower down the complexity and improve the stability of the positioning algorithms. However, the interior placement, for example, the Case B in Figure 3 for the irregular space, often leaves some uncovered space outside the rectangular coverage of the readers, which may lower down the positioning accuracy within this area.

4.2.2. Low Accuracy Requirement. For low-accuracy requirements, such as the room-resolution level in which the users can only be recognized if they are in the room, we can employ the coarsening deployment model including the radio signal strength (RSS) justification scheme and the IP address-stamping scheme.

RSS Justification. The positions of the readers in each room should meet the following conditions: (1) setting readers according to the whole area or divided sections of the room; normally one reader for one room; (2) placing the reader of a room on the side far from the gate and the passageway to minimize the interference between the signals of adjacent rooms; (3) maximizing the signal range of the reader to cover as much as possible inside the room but as little as possible outside the room. With the deployment scheme, a specific RSS threshold for the reader of the room achieved from the

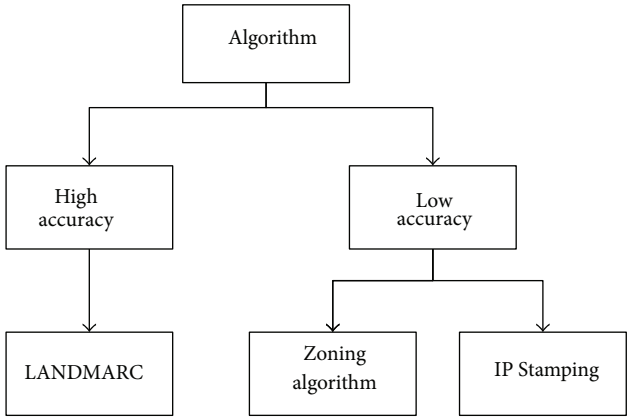


Figure 4: The algorithm framework.

manual training or machine learning can be adopted to justify whether a user is inside or outside of the room.

Given any two adjacent rooms, in the signal-overlapping areas of the two contiguous readers of them (in the case of one reader for one room), the RSS values of a user-tag to each reader can be compared to decide which room the tag belongs to especially when there is a wall between them which can often dramatically lower down the RSS values of the signal-through reader. The method is statistically verified efficient in this paper.

IP Address Stamping. Radio signals transmit in its constant speed; a reader can have its unique address of either IPV4 or IPV6 in a given networking configuration. Whenever the signal of a user tag is collected via the IP-stamped reader, the backend signal-processing server can justify which reader it is from. If there are two closely adjacent signals analyzed at the server, the earlier-coming one will be preferred to be valid and the location of the reader will be regarded as the tag position in the low-accuracy scenario.

4.3. Algorithms. According to the deployment model above, we have different algorithms for each deployment scheme. See Figure 4.

4.3.1. High-Accuracy Scenario. In scenarios which require high sensing accuracy, the positioning system of this paper just employs the LANDMARC algorithm, which is proposed by Ni et al. [10] and was verified in the tests of this paper. LANDMARC works with the idea that the unknown positions of the active user tags could be calculated with the already-known fixed positions of reference tags since they are all placed in the same space.

Assume we have n RFID readers, m reference tags, and N user tags to track, respectively, n is set to 4 in the model of this paper, and we get the signal strength vector (SSV) of a user tag $p_i, i \in [1, N]$ as

$$S_i = \left(S_{i1}, S_{i2}, S_{i3}, S_{i4}\right), \tag{1}$$

where S_i denotes the signal strength of the user tag p_i received by reader 1 to 4.

For a reference tag $r_j, j \in [1, m]$, we have similar signal strength vector as

$$\theta_j = \left(\theta_{j1}, \theta_{j2}, \theta_{j3}, \theta_{j4}\right), \tag{2}$$

where θ_j denotes the signal strength of the reference tag r_j. Then we can calculate the Euclidean distances E_{ij} from the user tag p_i to an arbitrary reference tag r_j with respect to the SSVs of them:

$$E_{ij} = \sqrt{\sum_{n=1}^{4} (S_{in} - \theta_{jn})^2}, \tag{3}$$

where $i \in [1, N]$ and $j \in [1, m]$. The nearer the user tag to track is to the reference tag, the smaller the Euclidean distance E between them will be.

Then the distance vector for each tag p_i can be obtained as E_i with respect to m reference tags:

$$E_i = \left(E_{i1}, E_{i2}, \ldots\ E_{im}\right). \tag{4}$$

For each tag p_i, we selected 4 reference tags with the smallest E values as calculation basis, and then the respective weight w_{ij} between p_i and r_j ($j \in [1, 4]$) can be calculated

$$w_{ij} = \frac{E_{ij}^2}{\sum_{j=1}^{4} E_{ij}^2}. \tag{5}$$

From the LANDMARC algorithm, the position of the user tag p_j can be then calculated as

$$\left(x_i, y_i\right) = \sum_{j=1}^{4} w_{ij} \times \left(x_j, y_j\right), \tag{6}$$

where (x_i, y_i) is the coordinate of user tag to be evaluated and (x_j, y_j) is the coordinate of one of the 4 selected reference tags. The position of user tags can thus be calculated with relatively high accuracy.

4.3.2. Low-Accuracy Scenario. In scenarios with low-accuracy requirements, we employ corresponding low-accuracy algorithms to fulfill the positioning. There are two algorithms in this category.

Hybrid Zoning Algorithm. We employ a combined zoning and RSS-justification algorithm. The transmitting power of each reader is preset to adjust the signal-detection capacity and make it cover the zone of a certain area which is preconfigured in deployment. In order to avoid any blind point in each zone, two contiguous zones would have some overlaps. We define S_{kj} as the signal strength of tag k received by reader j, whose coverage zone j has an RSS threshold S_j. We configure the reader to make sure that all S_{kj} of tags inside the zone j will satisfy $S_{kj} \geq S_j$, and satisfy $S_{kj} < S_j$ when they get outside of the zone j. S_j is obtained from testing or training before deployment, which is a relatively stable value monotone with respect to the radius of the zone. For the overlapping zone of the two contiguous zones i and j, tag k might satisfy $S_{kj} \geq$

S_i and $S_{kj} \geq S_j$ simultaneously. In this case, the position could be determined by comparing S_{ki} and S_{kj}; the tag is justified to be within the zone i if S_{ki} is larger. When making such a decision, we make a reasonable assumption that each divided zone covers almost a same area. If the areas vary a lot from one to another, the corresponding fencing schemes should be considered, which figures out each zone by the corresponding RSS thresholds S_i. The context of the user tag would be employed to break the tie in the overlapping zones of two adjacent spaces.

IP Stamping. We also employ IP stamping scheme to add location information into the data frames returned from each reader. Since the RFID readers can work with IP protocols by connecting to a LAN or a WAN, different IP addresses can be set to reflect different locations. The data frame transmitted can be defined in XML format as shown in Algorithm 1.

When receiving a data frame, the server can obtain the position of tag k by extracting tag ID and IP information from the data frame. Then the location of a certain tag could be interpreted with the tag ID and its corresponding source IP. For two very-close frames from different IPs, the earlier-coming one will be preferred to be valid since it implies that the tag may be closer to the reader of the first IP than the second due to the constant transmitting speed of the signal. Such a method can simplify the work needed for low accuracy requirement positioning.

4.4. Privacy Control. Based on the analysis of the system requirements of the privacy control, this paper employs the methodology for the privacy control as follows.

(i) Role control for users: two roles should be distinguished: administrator and common user. Administrators can access the global configurations of the system and the common users can access only their personal data. No application entries are suggested for the administrators to see all the privacy data of the common users unless they select "yes" in the corresponding options. Nevertheless, all the privacy data are tracked by the system in the backend and can be used in offline scenarios after legal authorization achieved.

(ii) Privacy policy and configuration for common users: the privacy policy is displayed by default at the first-time start-up of the application, and it can be accessed whenever the user would like to via the user interface. There are also the privacy configuration options for the common users to setup, such as which information can be shared, whom the information can be shared to, when to share and where to share.

(iii) Support for adaptable positioning accuracies: the system is not required always to offer the same-level accuracy for each scenario in ESNS. For example, in the public areas, the positioning accuracy can be the highest, while in the private areas such as rest rooms, the accuracy should be coarsened. This paper suggests that the privacy should be also guaranteed in the physical-collection layer rather than only in the software.

5. Implementation and Results

We implemented an ESNS application for the academic conference scenario—Find & Connect [20], as a validation for our model, which was actually employed in the ACM UbiComp 2011 conference. In this section we (1) describe the system setup and configuration and (2) present the outcomes and data from the system we used.

5.1. System Deployment and Setup. The physical layer of the Find & Connect UbiComp 2011 system consisted of two parts: (1) the RFID readers whose coverage ranges were controllable and were able to reach as far as 50 meters; (2) the active RFID tags which the attendees needed to wear.

The core of the system deployment and configuration was the placement of RFID readers in the selected venues. There were two types of venues in UbiComp 2011, that is, the small classrooms sized about 6 m × 8 m where the tutorials and workshops were held, and the large rooms sized about 30 × 20 m where the main conference sessions were held. We deployed the location-sensing model according to the requirements of the different scenarios.

In small venues, the organizers did not need to implement a positioning system with high accuracy because of (1) the low QoS expectation from the attendees, (2) the small space, and (3) the cost issue. Actually the real need in this conference was just to know in which room the users were located in the tutorials and workshops. So the zoning algorithm was selected to process the RFID signals. The actual deployment of the RFID readers is illustrated in Figure 5.

However, in the large venue where the main sessions took place, the attendees would ask questions, would see what was happening, and would setup connections with adjacent seats. So the high-accuracy algorithm was required in the scenario. As shown in Figure 6, the large venue with irregular shape can also be covered by the deployment model of this paper. The four readers were placed to form a rectangle which can cover most of the seats in the venue.

There were 400+ attendees in the ACM UbiComp 2011 conference. Due to the large tag amount and the limited processing capacity of the RFID readers, the refreshing interval of each RFID reader was set to 6 seconds, which can guarantee processing all the requests while keeping the response time fast enough. The LANDMARC algorithm was used to give the location information of the user tags with the support of reference tags in this scenario.

In order to figure out how the amount of the reference tags can impact the positioning accuracy, the experiments with different number of reference tags in the large venue were done. The accuracy was evaluated as the relative error to the real coordinates in the room:

$$\text{accuracy} = 1 - \frac{\sum_1^N \sqrt{(x_i - x_{i0})^2 + (y_i - y_{i0})^2}}{\sum_1^N \sqrt{x_{i0}^2 + y_{i0}^2}}, \quad (7)$$

```
<data>
    <time>time</time>
    <ID>tag ID</ID>
    <RSS>tag RSS</RSS>
    <IP>x.x.x.x</IP>
    ..........
</data>
```

Algorithm 1

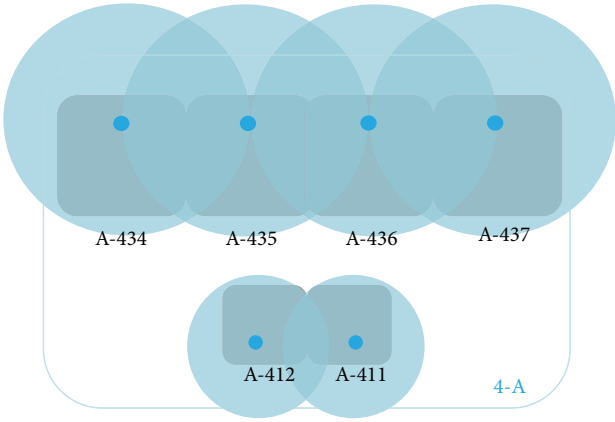


Figure 5: The RFID reader placement in classrooms. The blue points denote the positions of RFID readers; the dark rectangles denote the classrooms; the blue circles denote the radio-ranging areas.

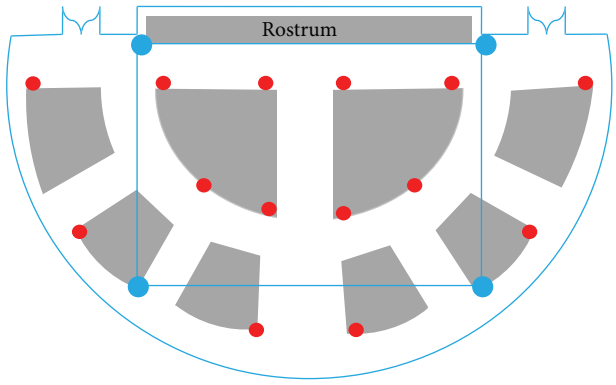


Figure 6: The RFID reader placement in the main venue. The blue points are the positions of the RFID readers and the red points are the positions of the reference tags.

where N is the tag amount, (x_i, y_i) is the calculated coordinates of a user tag, and (x_{i0}, y_{i0}) is the real coordinates of it. Figure 7 gives the testing results. It shows that the accuracy improved a lot with the amount of reference tags increasing when the amount was still little. But the system accuracy will begin to be stable after the amount of the reference tags was approaching 20 (after 15). That was also the main reason why we finally employed 14 reference tags in the implementation. The reference tags were placed evenly as Figure 6 shows to help improve the positioning accuracy of the target space.

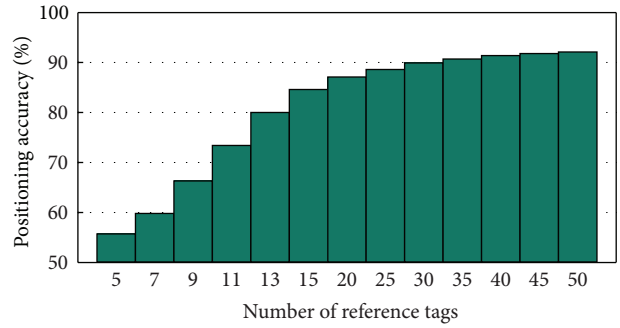


Figure 7: The positioning accuracy versus the amount of reference tags in the high-accuracy scenario.

Table 1: Positioning tests of low-accuracy model.

Tag	Time period	No. of pos.	No. of errors	Error rate
A	1 hour	1279	7	0.55%
B	2 hours	1440	9	0.63%
C	4 hours	2157	16	0.74%
D	3 hours	1648	10	0.60%
E	6 hours	3613	21	0.58%
	Average		11	0.51%

Pos.: positions.

5.2. Results and Conclusions. In this section, the corresponding results from the Find & Connect UbiComp 2011 are shown and analyzed.

5.2.1. User Configuration for Privacy. In the Find & Connect UbiComp 2011 system, the location information of each user was optionally exposed to other users and the option could be easily configured by the user via the graphic interface shown in Figure 8, where the privacy policy can also be published for users to review.

If the user checked the checkbox of *Share Location*, the real-time positions of the user could be seen by any user of the system. It was a global configuration. In personal configuration, more privacy rules could be adopted, such as sharing to specific friends, the shared regions, and the level of the positioning accuracy. All the privacy configurations were implemented based on the privacy model supported by the RFID sensing model of this paper.

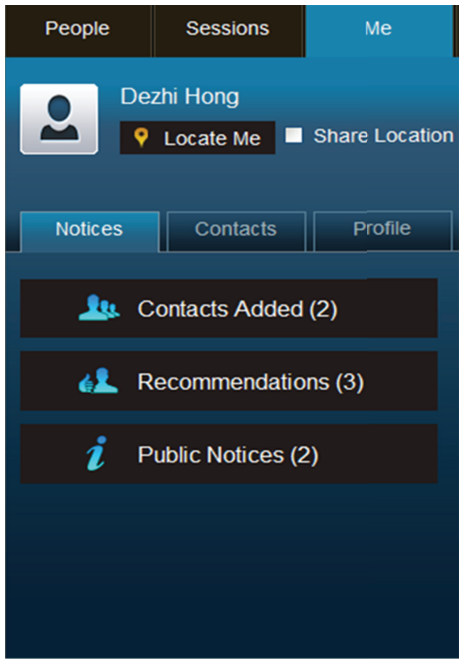


Figure 8: System configuration for the location privacy.

5.2.2. Low-Accuracy Model. In small venues, the zoning algorithm was employed to give the user locations. In order to verify its accuracy, the testing both before and on the conference days was performed. The results can be seen in Table 1.

The data in rows 1 and 2 are the pretrial data before the conference days, in which we kept the tags A and B in a selected room statically for more than two hours and compared the collected coordinates with the actual positions; the results showed the error rates of 0.55% and 0.63%, respectively. And in rows 3, 4, and 5, the data are the results obtained from the tags of three tutorial speakers on the conference days, who gave their speech in a specific room during a whole day. We compared their calculated positions from the algorithm with the actual positions recorded manually and found out the results. The error rates of 0.74%, 0.60%, and 0.58% showed that the model of this paper could work stably even with attendees filled up in the room; especially in the real conference days, the amount of the users in each small classroom was about 40, which nearly reached the upper limit of the space.

5.2.3. High-Accuracy Model. In the large venue where the high-accuracy model was employed, the testing on the 400+ attendees with 14 reference tags is depicted as Figure 9.

The positioning accuracy was evaluated by collecting arbitrary 20 trials as Table 2 shows. The worst case (marked by *) obtained 1.36 m (m for meter) of the absolute error (Euclidean distance to the real coordinate) and 84.9% relative accuracy; the best case obtained 0.39 m of the absolute error and thus 97.8% accuracy; the average accuracy among the all trials was 92.9%. All the results were a little better than the results of LANDMARC [10]. This was mainly because

Table 2: Positioning tests of high-accuracy model.

Case	Actual pos.	Calculated pos.	Absolute errors (meter)	Error %
Worst*	7.90, 4.25	8.92, 5.15	1.36	15.1%
Best	16.61, 3.86	16.07, 4.18	0.39	2.2%
Ave	\	\	0.86	7.1%

Pos.: positions.

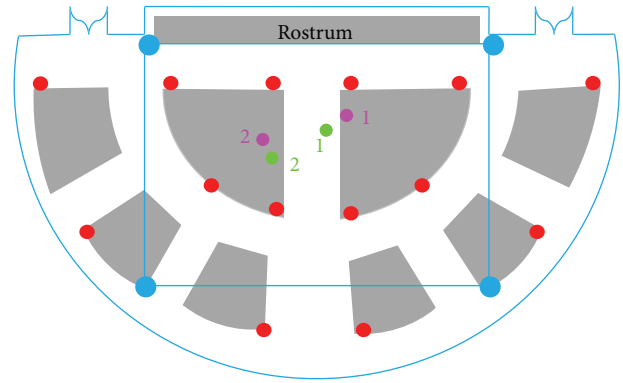


Figure 9: The accuracy test in the main venue. The magenta points are the real positions of the tracking users and the green ones are the calculated positions by the algorithm; the room size is 30 × 18 meters, and the origin is the top-left corner of the venue.

(1) the conference venue had a very clean electromagnetic environment and the space inside was also broad, and (2) the RFID tags were made in new technology and the signal quality could be maintained to a stable level; those were all good for the RSS stability and the algorithm accuracy.

It should be noticed that the results in Table 2 were tested only within the rectangular area covered by the four RFID readers. The *worst* case (marked by *) was actually not the globally worst. The testing results outside the rectangular area showed that the worst case of the positioning accuracy could be up to 50%. However, the main focus of this paper is the entire model rather than the specific positioning algorithm; the positioning accuracy may be further improved in case of necessity.

5.2.4. Analysis. Based on the implementation and the testing results shown above, the usability, QoS, scalability, and the privacy of the location-sensing model for ESNS in this paper can be analyzed as follows.

Usability. The testing system corresponding to the model of this paper is simple, flexible, and scalable to deploy and configure, which was verified from the trial in the conference with 400+ attendees. The specific deployment scheme can be easily determined in the cases with or without irregular venue shapes, and the users only need to wear the small RFID

tags. The trial application was also implemented as a browser-server application, which showed its flexible adaptability to different ubiquitous service applications.

QoS. We have models with different positioning accuracy prepared for different QoS/privacy scenarios. The hybrid zoning algorithm and the LANDMARC algorithm all worked well according to the trial results. The power consumption of the tag was also very small due to the testing, and it was estimated to last for 3 years by the data rate of one blink every thirty seconds.

Scalability. The trial system could support flexible user volume up to 500 users without any further complicated configuration and changes on the models. Based on the testing results and further theoretical analysis, the model can hopefully support up to 1000 users with the same system (the same computers, networks, and readers). The scalability of the model is promising.

Privacy. Considering the privacy requirements of the application for the conference ESNS, the users were allowed to configure their privacy levels. In the strictest case, the reader can also discard the sensing signals directly as long as the user set the location privacy to "not shared," which is considered to avoid any possible trace or leak of the position information beyond the physical-collection layer and lower down the privacy risks.

6. Conclusions

This paper presents a scalable and privacy-aware location-sensing model that could be employed by ESNS. The model consists of the architecture, the deployment model, the positioning algorithm, and the privacy control, which can meet the requirements of the usability, QoS, scalability, and privacy for a standard ESNS. We verified the model with our implemented conference-based ESNS, Find & Connect for UbiComp 2011, the analysis of the data and the trial experience show that the model is efficient and scalable. However, the positioning accuracy should also be further improved to meet more challenging requirements and the entire model should be tested much in the future work.

Acknowledgment

This work is funded by the State Key Development Program for Basic Research of China 2011CB302902 and NSFC 61201357.

References

[1] J. Burke, D. Estrin, M. Hansen et al., "Participatory sensing," in *Proceedings of the ACM Sensys, Worldwide Sensor Web Workshop*, 2006.

[2] H. Kawaji, K. Hatada, T. Yamasaki, and K. Aizawa, "Image-based indoor positioning system: fast image matching using omnidirectional panoramic images," in *Proceedings of the 1st ACM International Workshop on Multimodal Pervasive Video Analysis*, pp. 1–4, Firenze, Italy, October 2010.

[3] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The active badge location system," *ACM Transactions on Information Systems*, vol. 10, no. 1, pp. 91–102, 1992.

[4] M. McCarthy, P. Duff, H. L. Muller, and C. Randell, "Accessible ultrasonic positioning," *IEEE Pervasive Computing*, vol. 5, no. 4, pp. 86–93, 2006.

[5] R. Harle, "A survey of indoor inertial positioning systems for pedestrians," *IEEE Communications Surveys & Tutorials*, no. 99, pp. 1–13, 2013.

[6] X. Lu, C. Chia-Chih, and J. K. Aggarwal, "Human detection using depth information by Kinect," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '11)*, pp. 15–22, 2011.

[7] C. K. Schindhelm, "Evaluating SLAM approaches for microsoft kinect," in *Proceedings of the 18th International Conference on Wireless and Mobile Communications*, pp. 402–407, 2012.

[8] K. Muthukrishnan, M. Lijding, and P. Havinga, "Towards smart surroundings: enabling techniques and technologies for localization," in *Proceedings of the 1st International Workshop on Location- and Context-Awareness (LoCA '05)*, pp. 350–362, May 2005.

[9] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM '00)*, vol. 2, pp. 775–784, March 2000.

[10] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: indoor location sensing using active RFID," in *Proceedings of the 1st IEEE International Conference on Pervasive Computing and Communications (PerCom '03)*, pp. 407–415, March 2003.

[11] C. Qun, L. Hua, Y. Min, and G. Hang, "RSSI ranging model and 3D indoor positioning with ZigBee network," in *Proceedings of the IEEE/ION Position Location and Navigation Symposium (PLANS '12)*, pp. 1233–1239, 2012.

[12] J. Schroeder, S. Galler, and K. Kyamakya, "A low-cost experimental ultra-wideband positioning system," in *Proceedings of the IEEE International Conference on Ultra-Wideband (ICU '05)*, pp. 632–637, September 2005.

[13] G. Fehér, "Ultra wide band (UWB) transmitter with configurable design for indoor positioning," in *Proceedings of the 18th International Conference on Microwaves, Radar and Wireless Communications (MIKON '10)*, pp. 1–4, IEEE Press, Vilnius, Lithuania, June 2010.

[14] K. Pahlavan, X. Li, and J. P. Mäkelä, "Indoor geolocation science and technology," *IEEE Communications Magazine*, vol. 40, no. 2, pp. 112–118, 2002.

[15] K. H. Lee and S. H. Cho, "CSS based localization system using Kalman filter for multi-cell environment," in *Proceedings of the International Conference on Advanced Technologies for Communications (ATC '08)*, pp. 293–296, IEEE Press, Hanoi, Vietnam, October 2008.

[16] M. Duckham and L. Kulik, "Location privacy and location-aware computing," in *Dynamic & Mobile GIS: Investigating Change in Space and Time*, J. Drummond, Ed., pp. 34–51, CRC Press, Boca Raton, Fla, USA, 2006.

[17] J. Krumm, "A survey of computational location privacy," *Personal and Ubiquitous Computing*, vol. 13, no. 6, pp. 391–399, 2009.

[18] E. Kaasinen, "User needs for location-aware mobile services," *Personal and Ubiquitous Computing*, vol. 7, no. 1, pp. 70–79, 2003.

[19] G. Danezis, S. Lewis, and R. Anderson, "How much is location privacy worth?" in *Proceedings of the 4th Workshop on the Economics of Information Security*, Harvard University, Cambridge, Mass, USA, 2005.

[20] X. Zuo, A. Chin, X. Fan et al., "Connecting people at a conference: a study of influence between offline and online using a mobile social application," in *Proceedings of the 5th IEEE CPSCom Conference*, Besancon, France, 2012.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 272916, 11 pages
http://dx.doi.org/10.1155/2013/272916

Research Article

A Study of Mobile Sensing Using Smartphones

Ming Liu

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

Correspondence should be addressed to Ming Liu; wing.lm@gmail.com

Received 8 December 2012; Accepted 15 January 2013

Academic Editor: Chao Song

Traditional mobile sensing-based applications use extra equipments which are unrealistic for most users. Smartphones develop in a rapid speed in recent years, and they are becoming indispensable carry-on of daily life. The sensors embedded in them provide various possibilities for mobile applications, and these applications are helping and changing the way of our living. In this paper, we analyze and discuss existing mobile applications; after that, future directions are pointed out.

1. Introduction

The word sensing builds a bridge between real world and virtual world; with the help of various sensors, man-made devices are able to feel the world like God-made creatures do. Bell may be the first generation of sensors; people tie up a bell to a string so that when there is a vibration on the string, the bell will ring. Bell is a very powerful and effective sensor; it contains two parts: detection and processing. When a bell detects a vibration, it will generate a period of ringing and the volume of the ringing is proportional to the amplitude of the vibration. However, bell is the kind of sensor that connects real world to real world. With the development of electronic devices, a new man-made world has been building. This world is called virtual world; many complicated calculations are running in this world so that people in real world can enjoy their lives. Virtual world needs data to keep running, and it is far from enough to input data into the virtual world depending on human operations. Sensor is a way to sense the world and interpret the sensed information to the data form of the virtual world; therefore, sensing becomes an important part of research field and industry field.

Early sensing-based applications are mostly used for research purposes or used in some specific areas. References [1, 2] propose localization methods for finding odor sources using gas sensors and anemometric sensors. Reference [3] uses a number of sensors embedded into a cyclist's bicycle to gather quantitative data about the cyclist's rides; this information would be useful for mapping the cyclist experience. Reference [4] uses body-worn sensors to build an activity recognition system, and [5] uses body-worn sensors for healthcare monitoring. Reference [6] proposes a robotic fish carrying sensors for mobile sensing. Also, in Wireless Sensor Networks (WSN), there are a lot of sensing-based applications. References [7, 8] deploy wireless sensors to track the movement of mobile objects. References [9, 10] deploy sensors for monitoring volcano.

People-centric sensing mentioned in [11] uses smartphones for mobile sensing. Smartphones are very popular and becoming indispensable carry-on for people in recent years; they are embedded with various sensors which could be used for many interesting applications. Unlike specific sensors which are used for specific areas, sensors in smartphones could provide unlimited possibilities for applications to help and change the life of people; also, using smartphone instead of specific equipment makes an application easier to be accepted by users.

In this paper, we will discuss some existing interesting sensing-based applications using smartphones and give some possible future directions. Section 2 gives detailed descriptions of sensors embedded in modern smartphones; Section 3 introduces some sensing-based applications; Section 4 gives a conclusion and future directions.

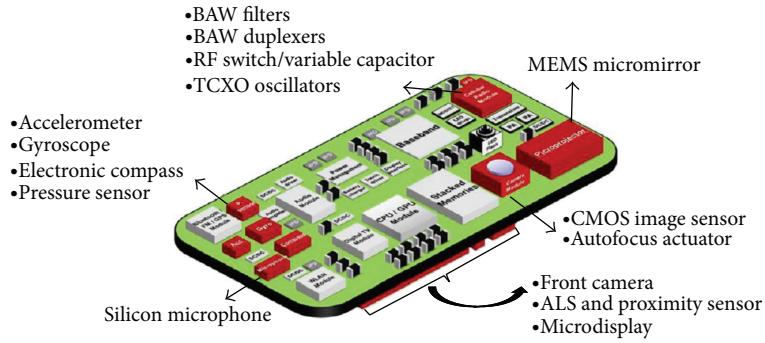


Figure 1: Sensors inside of smartphones.

2. Sensors in Smartphones

As Figure 1 shows, modern smartphones have several kinds of sensors. The most popular sensors which most smartphones have are accelerometer, gyroscope, magnetometer, microphone, and camera. In this section, we will discuss the characteristics of the sensors.

2.1. Accelerometer. An accelerometer measures proper acceleration, which is the acceleration it experiences relative to free fall and is the acceleration felt by people and objects. To put it another way, at any point in spacetime the equivalence principle guarantees the existence of a local inertial frame, and an accelerometer measures the acceleration relative to that frame. Such accelerations are popularly measured in terms of g-force [12].

The principle of accelerometer is using inertial force. Try to imagine a box with six walls, a ball is floating in the middle of the box because no force is added to the ball (e.g., the box may be in the outer space) [13]. When the box moves to the right direction, the ball will hit the left wall. The left wall is pressure sensitive that it can measure the force of hitting applied to the left wall; therefore, the acceleration can be measured. Because of gravity, when the box is placed at earth, the ball will keep pressing the bottom wall of the box and give constant ~9.8 m/s^2 acceleration. The gravity force will affect the measurement of accelerometer for measuring speed or displacement of an object in a three-dimension. The gravity force must be subtracted before any measurement. However, the gravity force can be taken as an advantage of detecting the rotation of a device. When a user rotates his smartphone, the content he/she is watching will switch between portrait and landscape. As Figure 2 shows, when the screen of smartphone is in a portrait condition, y-axis will sense the gravity; when the screen of smartphone is in a landscape condition, x-axis will sense the gravity. According to this, users can rotate their screens without affecting their reading experiences.

In theory, the displacement can be calculated as

$$d(t) = d_0 + v_0 t + \iint_0^T a(t)\, d_t d_\tau, \tag{1}$$

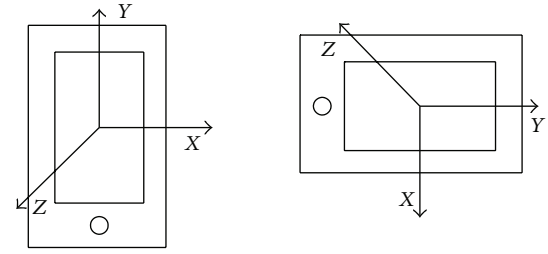


Figure 2: Screen rotation.

where $d(t)$: displacement, d_0: initial displacement, v_0: initial velocity, and $a(t)$: acceleration.

Equation (1) is a continuous function; the $a(t)$ we get in real environment is discrete due to sampling. To calculate the displacement according to discrete values, (2) has to be used as

$$\int_{t(0)}^{t(n)} a(t)\, d_t = \sum_{i=1}^{n} \left(\frac{a(i-1) + a(i)}{2} \right) \Delta t, \tag{2}$$

where $a(t)$: continuous acceleration, $a(i)$: ith sample, and Δt: time increment.

Then, the velocity and displacement can be calculated as the following [14]:

$$\begin{aligned} v(i) &= v(i-1) + \frac{a(i-1) + a(i)}{2} \Delta t, \\ d(i) &= d(i-1) + \frac{v(i-1) + v(i)}{2} \Delta t. \end{aligned} \tag{3}$$

The value the accelerometer returns is three-dimensional as Figure 2 shows; therefore, $a(t)$ will be calculated as the following shows:

$$a(t) = \vec{a_x} + \vec{a_y} + \vec{a_z}, \tag{4}$$

where $\vec{a_x}$, $\vec{a_y}$, and $\vec{a_z}$ are vectors.

2.2. Gyroscope. Accelerometer is good at measuring the displacement of an object; however, it is inaccurate to measure

the spin movement of the device, which is an easy thing for gyroscope.

A gyroscope is a device for measuring or maintaining orientation, based on the principles of angular momentum. Mechanically, a gyroscope is a spinning wheel or disk in which the axle is free to assume any orientation. Although this orientation does not remain fixed, it changes in response to an external torque much lesser and in a different direction than it would be without the large angular momentum associated with the disk's high rate of spin and moment of inertia. The device's orientation remains nearly fixed, regardless of the mounting platform's motion because mounting the device in a gimbal minimizes external torque [15].

Gyroscope is a very sensitive device; it is good at detecting the spin movement. Same as accelerometer, gyroscope returns three-dimensional values; the coordinate system is as Figure 2 shows. The value gyroscope returns is angular velocity which indicates how fast the device rotates around the axes. The gyroscope can be calculated as

$$g(t) = \vec{g_x} + \vec{g_y} + \vec{g_z}, \tag{5}$$

where $g(t)$: angular velocity $\vec{g_x}$, $\vec{g_y}$, and $\vec{g_z}$: vectors of angular velocity around x-, y-, and z-axes.

2.3. Magnetometer. A magnetometer is a measuring instrument used to measure the strength and perhaps the direction of magnetic fields [16]. Accelerometer and gyroscope are able to detect the direction of a movement; however, the direction is a relative direction; it obeys the coordinate system that a smartphone uses. Sometimes, different smartphones need to synchronize their directions; therefore, a magnetometer is needed to get an absolute direction (the direction obeys the coordinate system of earth).

The magnetometer returns three-dimensional values; if the device is placed horizontally, the orientation angle can be calculated as

$$\text{angle} = \arctan \frac{M_y}{M_x}. \tag{6}$$

Until now, we introduced three types of sensors: accelerometer, gyroscope, and magnetometer. With the help of the three types of sensors, smartphone can estimate its own all kind of movements. However, in real environment, errors of measurement happen all the time; we will talk about a way to correct the offset error of magnetometer, and the other two sensors may use the same way to correct their errors.

Firstly, place the magnetometer horizontally, rotate it in a uniform speed, measure the value of M_x and M_y, put z axis horizontally, and rotate to measure the value M_z. Calculate the offset value on the three axes as

$$X_{\text{offset}} = \frac{(\text{MAX}(M_x) + \text{MIN}(M_x))}{2},$$

$$Y_{\text{offset}} = \frac{(\text{MAX}(M_y) + \text{MIN}(M_y))}{2},$$

$$Z_{\text{offset}} = \frac{(\text{MAX}(M_z) + \text{MIN}(M_z))}{2},$$

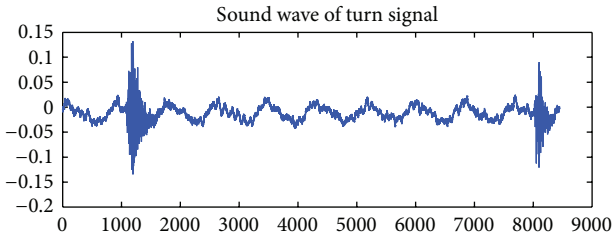


Figure 3: A sound record of turn signal of automobile.

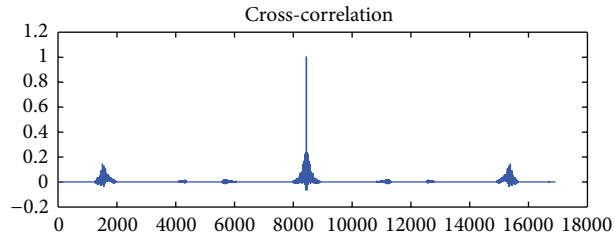


Figure 4: Cross-correlation.

$$M_x' = M_x - X_{\text{offset}},$$

$$M_y' = M_y - Y_{\text{offset}},$$

$$M_z' = M_z - Z_{\text{offset}}. \tag{7}$$

2.4. Microphones. Microphone is a very common sensor; it is usually used for recording sound. The problem is how to deal with the recorded sound. The most common way is to find a known period of sound in a sound record. Cross-correlation is a method to search a piece of signal in a mixed signal [17]. In signal processing, cross-correlation is a measure of similarity of two waveforms as a function of a time lag applied to one of them. This is also known as a sliding dot product or sliding inner product. It is commonly used for replacing a long signal for a shorter ones, which is known feature. It also has applications in pattern recognition, single particle analysis, electron tomographic averaging, cryptanalysis, and neurophysiology [18].

Cross-correlation can be calculated as (8) shows. Suppose that the known symbol pattern of sound wave of turn signal is ts of length L, $X(n)$ is the complex number representing the nth received symbol, then the cross-correlation at a shift position p is

$$C(ts, X, p) = \sum_{k=1}^{L} ts^* [k] X [k + p]. \tag{8}$$

If we use a sound wave to cross-correlate itself, for example, the sound wave as Figure 3 shows, the result will be shown as in Figure 4. The spike indicates the existence of a piece of signal.

2.5. Camera. Camera captures vision information from real world. From the human perspective, vision contains most

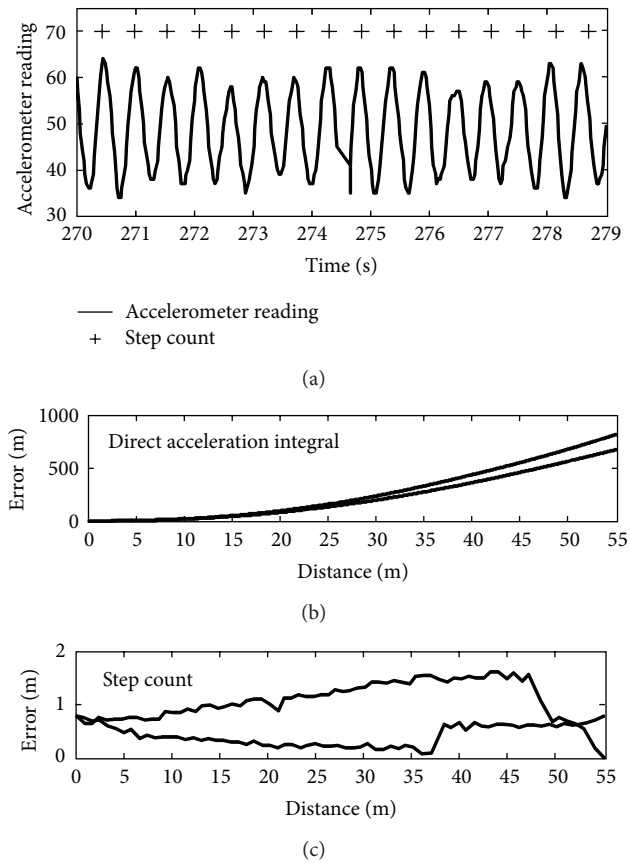


Figure 5: (a) Accelerometer readings (smoothened) from a walking user. (b) Displacement error with double integration for two users. (c) Error with step count method.

information we get. However, pattern recognition in computer area is not mature enough to work as human does. In this section, we will briefly introduce principle the pattern recognition.

A photo the camera records can be expressed as a matrix of light intensity of each pixel (here we take the grey-mode photo as an example). Suppose that the source matrixes (or as we call them dictionary) are $A_{s1}, A_{s2}, \ldots, A_{sn}$, as the matrix needed to be recognized is A_d, then the pattern recognition is proceeded as (9) shows. jth matrix in the dictionary is the result:

$$j = \arg\max A_{si} \cdot A_d. \tag{9}$$

Pattern recognition is far more complicated than (9); there are many good algorithms in pattern recognition area, like SIFT [19–23] and SVM [24–29]. But the recognition rate is still not good enough for practical applications.

3. Applications

In this section, we will introduce a few interesting sensing-based applications using smartphones. We divide the applications we are going to discuss into two categories: (1) accelerometer, gyroscope, and magnetometer; (2) microphone and camera.

3.1. Accelerometer, Gyroscope, and Magnetometer

3.1.1. Trace Track. Searching people in a public place is difficult; for example, a person is in a conference hall, a library, or a shopping mall, there are many crowds around, it is very difficult to find the target person. Even if the person tells you where he/she is, it is frustrating to find the place in an unfamiliar environment. Maps may be helpful but not always handy. Smartphones provide possibilities to develop an electronic escort service using opportunistic user intersections. Through periodically learning the walking trails of different individuals, as well as how they encounter each other in space time, a route can be computed between any pair of persons [30]. The escort system could guide a user to the vicinity of a desired person in a public place. Escort does not rely on GPS, WiFi, or war driving to locate a person—the escort user only needs to follow an arrow displayed on the phone [30].

Escort system presents an interest idea that a smartphone is an escort which tells you how many steps you walk and which direction you are heading to [30], see Figure 5(a). GPS is a way to achieve the idea; however, GPS cannot work in an indoor environment; WiFi localization is a good indoor localization method; however, it cannot guarantee, there are enough WiFi fingerprints for localization. Escort system uses accelerometer and gyroscope to achieve the idea. However, the displacement calculated based on accelerometer is inaccurate; the reasons are the jerky movement of smartphone in people's pocket and inherent error of measurement [31–33]; the displacement error may reach 100 m after 30 m walk, see Figure 5(b). Escort identifies an acceleration signature in human walking patterns to avoid the problem. This signature arises from the natural up- and down-bounces of the human body while walking and can be used to count the number of steps walked [30]. The physical displacement can then be computed by multiplying step count with the user's step size which is, a function of the user's weight and height [30], see Figure 5(c). Escort varies the step size with an error factor drawn from a Gaussian distribution centered on 0 and standard deviation 0.15 m [30]. This better accommodates the varying human step size [30].

Compass (magnetometer) is also used in the escort system. Like accelerometer has measurement errors, the compass noise is caused by several factors, including user sway, movement irregularities, magnetic fields in the surroundings, and the sensor's internal bias. Because these factors are related to the user, surroundings, and the sensor, the noise is difficult to predict and compensate. To characterize the compass noise, escort ran 100 experiments using 2 Nokia 6210 Navigator phones and has observed an average bias of 8 degrees and a standard deviation of 9 degrees [30]. In addition to this large noise range, escort made two consistent observations as follows: (1) when the user is walking in a constant direction, the compass readings stabilize quickly on a biased value and exhibit only small random oscillations, and (2) after each turn, a new bias is imposed [30]. Escort identifies two states of the user: walking in a constant direction and turning based on the two observations. Turns are identified when the compass headings change more significantly than due to

random oscillations. The turn identification algorithm uses the following condition [30]:

$$\text{Avg}\left(t_{i+1} - \text{Avg}\left(t_i\right)\right) > \frac{\text{StdDev}\left(t_i\right) + \text{StdDev}\left(t_{i+1}\right)}{2} + G, \tag{10}$$

where $\text{Avg}(t_i)$ denotes the average compass readings over a t_i time period (e.g., 1 second), $\text{StdDev}(t_i)$ is the standard deviation of compass readings during t_i, and G is a guard factor. While on a constant direction, escort compensates the stabilized compass reading with the average bias and reports the resulting value as the direction of the user. During turns, escort considers the sequence of readings reported by the compass [30].

Similar usage of accelerometer and magnetometer appears in [34]. It uses accelerometer to estimate the walking trace of people and correct the trace periodically by GPS. Accelerometer used as a supplement of GPS localization is a popular way, and many researches focused on this [35–40].

Smartphone is not only used on walking people but also on people on vehicles.

3.1.2. Dangerous Drive. When drivers are sitting in a vehicle, their smartphones are able to measure the acceleration, velocity, and turns through sensors embedded. Because the smartphone is very popular, using smartphone is an easy way to implement mobile sensing.

Driving style can be divided into two categories: nonaggressive and aggressive. To study the vehicle safety system, we need to understand and recognize the driving events. Potentially aggressive driving behavior is currently a leading cause of traffic fatalities in the United States, and drivers are unaware that they commit potentially aggressive actions daily [41]. To increase awareness and promote driver safety, a novel system that uses Dynamic Time Warping (DTW) and smartphone-based sensor fusion (accelerometer, gyroscope, magnetometer, GPS, and video) to detect, recognize, and record these actions without external processing being proposed [41].

The system collects the motion data from the the accelerometer and gyroscope in the smartphone continuously at a rate of 25 Hz in order to detect specific dangerous driving behaviors. Typical dangerous driving behaviors are hard left and right turns, swerves, and sudden braking and acceleration patterns. These dangerous driving behaviors indicate potentially aggressive driving that would cause danger to both pedestrians and other drivers. The system first determines when a dangerous behavior starts and ends using endpoint detection. Once the system has a signal representing a maneuver, it is possible to compare it to stored maneuvers (templates) to determine whether or not it matches an aggressive event [41].

References [42–44] use accelerometer and gyroscope for driving event recognition. Reference [42] includes two single-axis accelerometers, two single-axis gyroscopes, and a GPS unit (for velocity) attached to a PC for processing. While the system included gyroscopes for inertial measurements, they were not used in the project. Hidden Markov Models (HMM) were trained and used only on the acceleration data for the recognition of simple driving patterns. Reference [43] used accelerometer data from a mobile phone to detect drunk driving patterns through windowing and variation threshold. A Driver Monitor System was created in [44] to monitor the driving patterns of the elderly. This system involved three cameras, a two-axis accelerometer, and a GPS receiver attached to a PC. The authors collected large volumes of data for 647 drivers. The system had many components, one of them being detection of erratic driving using accelerometers. Braking and acceleration patterns were detected, as well as high-speed turns via thresholding. Additionally, the data could be used to determine the driving environment (freeway versus city) based on acceleration patterns.

References [45–47] use accelerometer and gyroscope to detect gesture recognition. The uWave paper [47] and the gesture control work of Kela et al. [48] explore gesture recognition using DTW and HMM algorithm,s respectively. They both used the same set of eight simple gestures which included up, down, left, right, two opposing direction circles, square, and slanted 90-degree angle movements for their final accuracy reporting. Four of these eight gestures are one-dimensional in nature. The results proved that using DTW with one training sample was just as effective as HMMs.

3.1.3. Phone Hack. Accelerometer is so sensitive that it can even detect finger press at the screen of the smartphone. Reference [49] shows that the location of screen taps on modern smartphones and tablets can be identified from accelerometer and gyroscope readings. The findings have serious implications, as we demonstrate that an attacker can launch a background process on commodity smartphones and tablets, and silently monitor the user's inputs, such as keyboard presses and icon taps [49]. While precise tap detection is nontrivial, requiring machine-learning algorithms to identify fingerprints of closely spaced keys, sensitive sensors on modern devices aid the process [49]. Reference [49] presents TapPrints, a framework for inferring the location of taps on mobile device touchscreens using motion sensor data combined with machine-learning analysis. By running tests on two different off-the-shelf smartphones and a tablet computer, the results show that identifying tap locations on the screen and inferring English letters could be done with up to 90% and 80% accuracy, respectively [49]. By optimizing the core tap detection capability with additional information, such as contextual priors, the core threat may further be magnified [49].

On both Android and iOS systems, location sensor like GPS is the only one which requires explicit user access permission, the reason probably is that people are not willing to be tracked and consider their locations information as privacy. But accelerometer and gyroscope sensors do not require explicit user permission on of the two mentioned operating systems. Because the sensors are mostly used for gaming, Android system does not restrict the access of accelerometer and gyroscope; background services with the two sensors are allowed. Moreover, there is work aimed at the standardization of JavaScript access to a device's accelerometer and gyroscope

sensors in order for any web application to perform, for example, website layout adaptation [49]. Reference [49] shows that accelerometer and gyroscope sensor data can create a serious privacy breach. In particular, it is demonstrated that it is possible to infer where people tap on the screen and what people type by applying machine-learning analysis to the stream of data from these two motion sensors. The reason the work focuses on the accelerometer and gyroscope sensors is that they are able to capture tiny device vibrations and angle rotations, respectively, with good precision [49].

Figure 6 shows a sample of the raw accelerometer data after a user has tapped on the screen—the timing of each tap marked by the digital pulses on the top line. As we can see, each tap generates perturbations in the accelerometer sensor readings on the three axes, particularly visible along the z-axis, which is perpendicular to the phone's display. Gyroscope data exhibits similar behavior and is not shown [49]. Some related works [50–52] are also about using accelerometer and gyroscope for phone hacks.

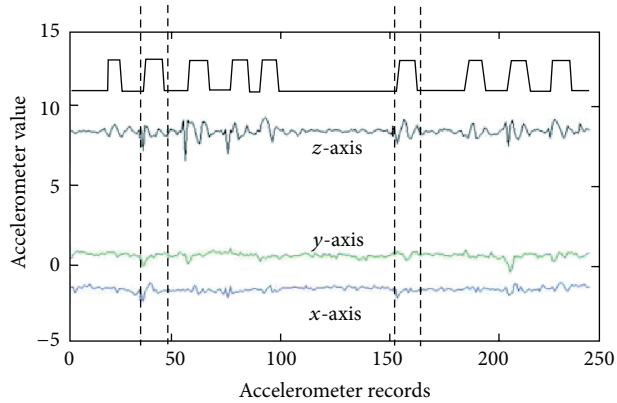


Figure 6: The square wave in the top line identifies the occurrence of a tap. Two particular taps have also been highlighted by marking their boundaries with dashed vertical lines. Notice that the accelerometer sensor readings (on the z-axis in particular) show very distinct patterns during taps. Similar patterns can also be observed in the gyroscope.

3.1.4. Phone Gesture. The first two sections are about using accelerometer, gyroscope, and magnetometer to detect a long-distance movement; actually, if the sensors are used to detect gestures, some interesting applications may appear.

The ability to note down small pieces of information, quickly and ubiquitously, can be useful. Reference [53] proposes a system called PhonePoint Pen that uses the in-built accelerometer in mobile phones to recognize human writing. By holding the phone like a pen, a user should be able to write short messages or even draw simple diagrams in the air. The acceleration due to hand gestures can be converted into an image and sent to the user's Internet email address for future reference [53].

The work is done without gyroscope, because the smartphone it used lacks gyroscope. There are two main issues to be solved: (1) coping with background vibration (noise); (2) computing displacement of phone.

Accelerometers are sensitive to small vibrations. Figure 7(a) reports acceleration readings as the user draws a rectangle using 4 strokes (around −350 units on the z-axis is due to earth's gravity). A significant amount of jitter is caused by natural hand vibrations. Furthermore, the accelerometer itself has measurement errors. It is necessary to smooth this background vibration (noise) in order to extract jitter-free pen gestures. To cope with vibrational noise, the system smoothes the accelerometer readings by applying a moving average over the last n readings ($n = 7$). The results are presented in Figure 7(b) (the relevant movements happen in the xy-plane, so we removed the z-axis from the subsequent figures for better visual representation) [53].

The phone's displacement determines the size of the character. The displacement δ is computed as a double integral of acceleration, that is, $\delta = \int(\int a d_t) d_t$, where a is the instantaneous acceleration. In other words, the algorithm first computes the velocity (the integration of acceleration) followed by the displacement (the integration of velocity). However, due to errors in the accelerometer, the cumulative acceleration and deceleration values may not sum to zero even after the phone has come to rest. This offset translates into some residual constant velocity. When this velocity is integrated, the displacement and movement direction become erroneous. In order to reduce velocity-drift errors, the velocity to zero at some identifiable points needs to be reset. The stroke mechanism described earlier is therefore used. Characters are drawn using a set of strokes separated by short pauses. Each pause is an opportunity to reset velocity to zero and thus correct displacement. Pauses are detected by using a moving window over n consecutive accelerometer readings ($n = 7$) and checking if the standard deviation in the window is smaller than some threshold. This threshold is chosen empirically based on the average vibration caused when the phone was held stationary. All acceleration values during a pause are suppressed to zero. Figure 8(a) shows the combined effect of noise smoothing and suppression. Further, velocity is set to zero at the beginning of each pause interval. Figure 8(b) shows the effect of resetting the velocity. Even if small velocity drifts are still present, they have a small impact on the displacement of the phone [53].

3.2. Microphone and Camera

3.2.1. Surround Sense. There are some research works [54–57] about using smartphones to sense the context of surroundings. A growing number of mobile computing applications are centered around the user's location. The notion of location is broad, ranging from physical coordinates (latitude/longitude) to logical labels (like Starbucks, McDonalds). While extensive research has been performed in physical localization, there have been few attempts to recognize logical locations. Reference [57] argues that the increasing number of sensors on mobile phones presents new opportunities for logical localization. Reference [57] postulates that ambient sound, light, and color in a place convey a photoacoustic signature that can be sensed by the phone's camera and microphone. In-built accelerometers in some phones may

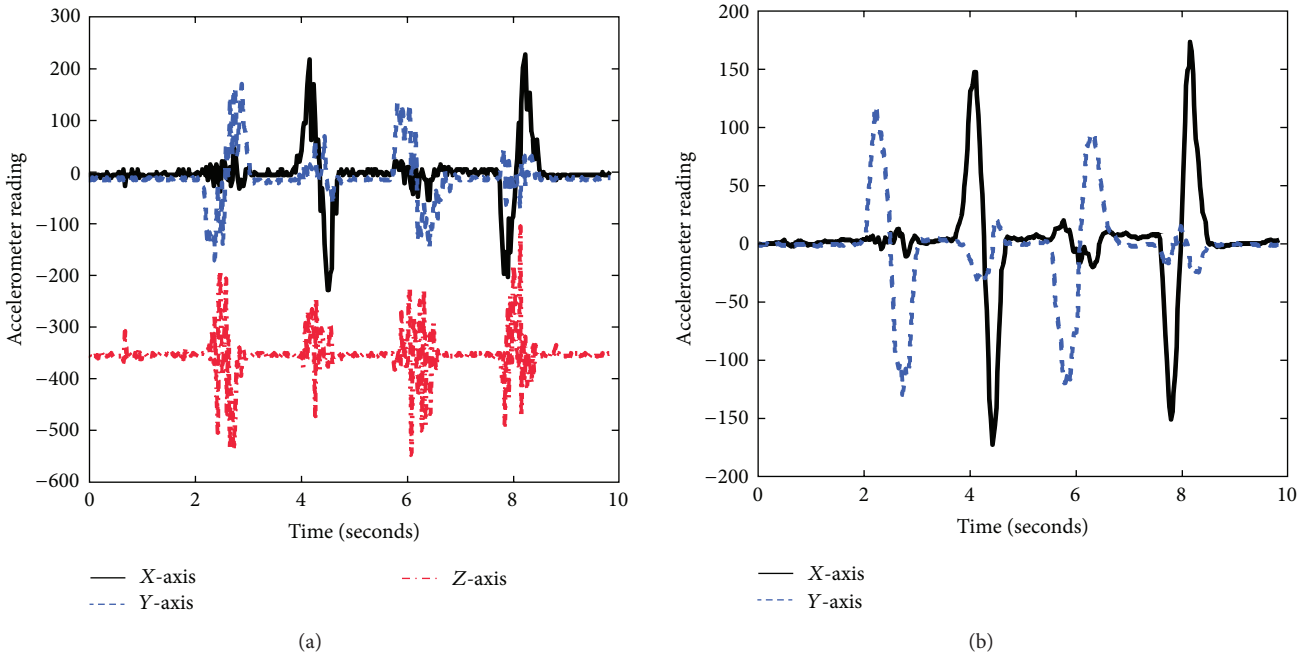


FIGURE 7: (a) Raw accelerometer data while drawing a rectangle (note gravity on the z-axis). (b) Accelerometer noise smoothing.

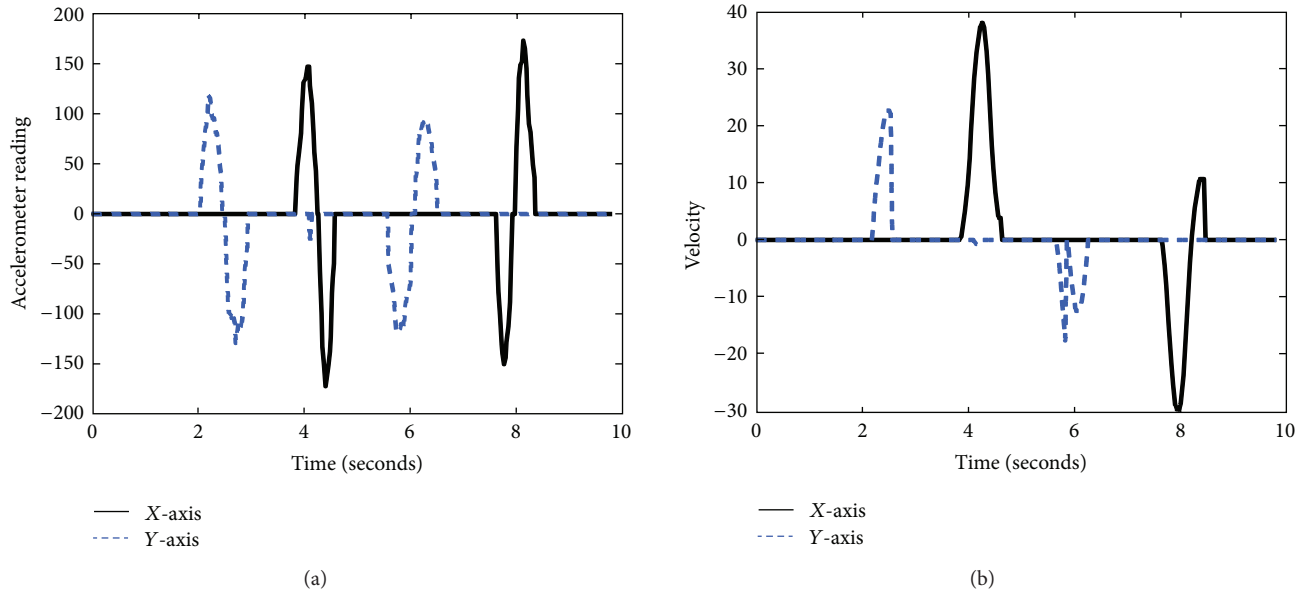


FIGURE 8: (a) Accelerometer readings after noise smoothing and suppression. (b) Resetting velocity to zero in order to avoid velocity drifts.

also be useful in inferring broad classes of user motion, often dictated by the nature of the place. By combining these optical, acoustic, and motion attributes, it may be feasible to construct an identifiable fingerprint for logical localization. Hence, users in adjacent stores can be separated logically, even when their physical positions are extremely close [57] see Figures 9 and 10.

Reference [57] takes advantage of microphone and camera to collect the surrounding fingerprint so that it can provide logical localization.

3.2.2. Localization. References [58–60] use audio for localization and the accuracy is improved to a high level. Reference [58] operates in a spontaneous, ad hoc, and device-to-device context without leveraging any preplanned infrastructure. It is a pure software-based solution and uses only the most basic set of commodity hardware—a speaker, a microphone, and some form of device-to-device communication—so that it is readily applicable to many low-cost sensor platforms and to most commercial-off-the-shelf mobile devices like cell phones and PDAs [58]. High accuracy is achieved

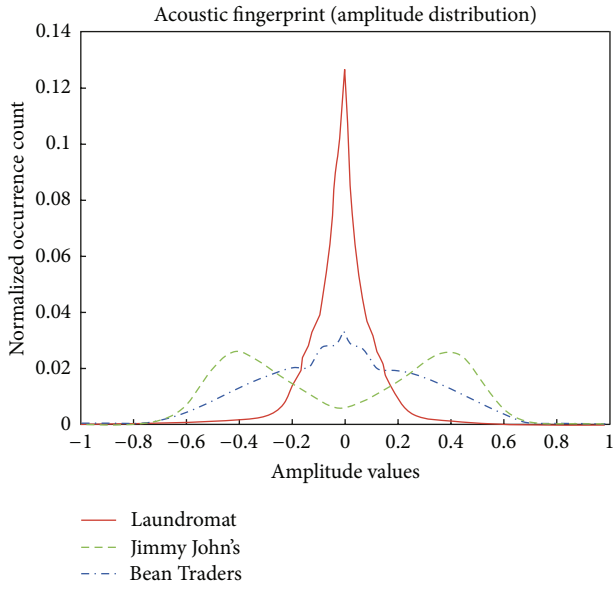


Figure 9: Sound fingerprints from 3 adjacent stores.

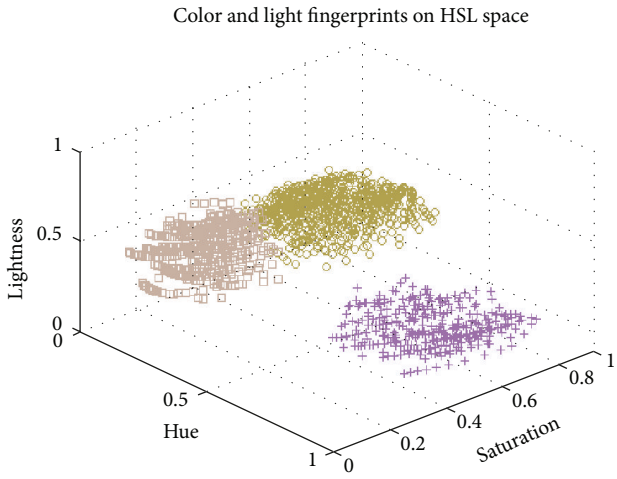


Figure 10: Color/light fingerprint in the HSL format from the Bean Traders' coffee shop. Each cluster is represented by a different symbol.

through a combination of three techniques: two-way sensing, self-recording, and sample counting. To estimate the range between two devices, each will emit a specially designed sound signal ("Beep") and collect a simultaneous recording from its microphone [58]. Each recording should contain two such beeps, one from its own speaker and the other from its peer [58]. By counting the number of samples between these two beeps and exchanging the time duration information with its peer, each device can derive the two-way time of flight of the beeps at the granularity of the sound sampling rate [58]. This technique cleverly avoids many sources of inaccuracy found in other typical time-of-arrival schemes, such as clock synchronization, non-real-time handling, and software delays [58].

Reference [58] sends beep sound to calculate the distance between two objects; the microphone will distinguish original sound and echo to get the time interval and, therefore, get a distance, see Figure 11.

4. Discussions and Conclusion

Sensors are the key factors of developing more and more interesting applications on the smartphones, and the sensors make the smartphone different from traditional computing devices like computer. Most applications used accelerometer and gyroscope because they are somehow the most accurate sensors. However, the vision contains huge information. We

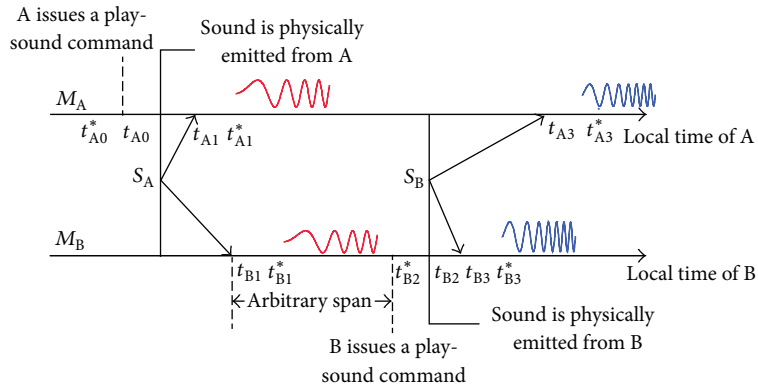


Figure 11: Illustration of event sequences in BeepBeep ranging procedure. The time points are marked for easy explanation and no timestamping is required in the proposed ranging mechanism.

believe that camera and pattern recognition will be used more and more in the future.

Acknowledgments

This work is supported by National Science Foundation under Grant nos. 61103226, 60903158, 61170256, 61173172, and 61103227 and the Fundamental Research Funds for the Central Universities under Grant no. ZYGX2010J074 and ZYGX2011J102.

References

[1] H. Ishida, K. Suetsugu, T. Nakamoto, and T. Moriizumi, "Study of autonomous mobile sensing system for localization of odor source using gas sensors and anemometric sensors," *Sensors and Actuators A*, vol. 45, no. 2, pp. 153–157, 1994.

[2] H. Ishida, T. Nakamoto, and T. Moriizumi, "Remote sensing of gas/odor source location and concentration distribution using mobile system," *Sensors and Actuators B*, vol. 49, no. 1-2, pp. 52–57, 1998.

[3] S. B. Eisenman, E. Miluzzo, N. D. Lane, R. A. Peterson, G. S. Ahn, and A. T. Campbell, "BikeNet: a mobile sensing system for cyclist experience mapping," *ACM Transactions on Sensor Networks*, vol. 6, no. 1, pp. 1–39, 2009.

[4] T. Choudhury, G. Borriello, S. Consolvo et al., "The mobile sensing platform: an embedded activity recognition system," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 32–41, 2008.

[5] B. Lo, S. Thiemjarus, R. King et al., "Body sensor network-a wireless sensor platform for pervasive healthcare monitoring," in *Proceedings of the 3rd International Conference on Pervasive Computing*, 2005.

[6] X. Tant, D. Kim, N. Usher et al., "An autonomous robotic fish for mobile sensing," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*, pp. 5424–5429, October 2006.

[7] A. Arora, P. Dutta, S. Bapat et al., "A line in the sand: a wireless sensor network for target detection, classification, and tracking," *Computer Networks*, vol. 46, no. 5, pp. 605–634, 2004.

[8] Y. C. Tseng, S. P. Kuo, and H. W. Lee, "Location tracking in a wireless sensor network by mobile agents and its data fusion strategies," *Information Processing in Sensor Networks*, pp. 554–554, 2003.

[9] G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees, and M. Welsh, "Monitoring volcanic eruptions with a wireless sensor network," in *Proceedings of the 2nd European Workshop on Wireless Sensor Networks (EWSN '05)*, pp. 108–120, February 2005.

[10] G. Werner-Allen, K. Lorincz, M. Welsh et al., "Deploying a wireless sensor network on an active volcano," *IEEE Internet Computing*, vol. 10, no. 2, pp. 18–25, 2006.

[11] A. T. Campbell, S. B. Eisenman, N. D. Lane et al., "The rise of people-centric sensing," *IEEE Internet Computing*, vol. 12, no. 4, pp. 12–21, 2008.

[12] Accelerometer, http://en.wikipedia.org/wiki/Accelerometer.

[13] A Guide To using IMU (Accelerometer and Gyroscope Devices) in Embedded Applications, http://www.starlino.com/imu_guide.html.

[14] M. Arraigada and M. Partl, "Calculation of displacements of measured accelerations, analysis of two accelerometers and application in road engineering," in *Proceedings of the Swiss Transport Research Conference*, 2006.

[15] Gyroscope, http://en.wikipedia.org/wiki/Gyroscope.

[16] Magnetometer, http://en.wikipedia.org/wiki/Magnetometer.

[17] S. Sen, R. Roy Choudhury, and S. Nelakuditi, "CSMA/CN: carrier sense multiple access with collision notification," *IEEE/ACM Transactions on Networking*, vol. 20, no. 2, pp. 544–556, 2012.

[18] Cross-correlation, http://en.wikipedia.org//wiki/Cross-correlation.

[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[20] H. Bay, T. Tuytelaars, and L. Gool, "SURF: speeded up robust features," in *Proceedings of the Computer Vision (ECCV '06)*, pp. 404–417, Springer, Berlin, Germany, 2006.

[21] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 886–893, June 2005.

[22] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[23] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[25] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*, Cambridge University Press, 2000.

[26] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.

[27] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[28] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, ACM press, New York, NY, USA, 1999.

[29] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2001.

[30] I. Constandache, X. Bao, M. Azizyan, and R. R. Choudhury, "Did you see Bob?: human localization using mobile phones," in *Proceedings of the 16th Annual Conference on Mobile Computing and Networking (MobiCom '10)*, pp. 149–160, September 2010.

[31] D. M. Boore, "Effect of baseline corrections on displacements and response spectra for several recordings of the 1999 Chi-Chi, Taiwan, earthquake," *Bulletin of the Seismological Society of America*, vol. 91, no. 5, pp. 1199–1211, 2001.

[32] D. M. Boore, C. D. Stephens, and W. B. Joyner, "Comments on baseline correction of digital strong-motion data: examples from the 1999 Hector Mine, California, earthquake," *Bulletin of the Seismological Society of America*, vol. 92, no. 4, pp. 1543–1560, 2002.

[33] D. M. Boore, "Analog-to-digital conversion as a source of drifts in displacements derived from digital recordings of ground acceleration," *Bulletin of the Seismological Society of America*, vol. 93, no. 5, pp. 2017–2024, 2003.

[34] I. Constandache, R. R. Choudhury, and I. Rhee, "Towards mobile phone localization without war-driving," in *Proceedings of the IEEE INFOCOM*, pp. 1–9, March 2010.

[35] J. Paek, J. Kim, and R. Govindan, "Energy-efficient rate-adaptive GPS-based positioning for smartphones," in *Proceedings of the 8th Annual International Conference on Mobile Systems, Applications and Services (MobiSys '10)*, pp. 299–314, June 2010.

[36] D. H. Kim, Y. Kim, D. Estrin, and M. B. Srivastava, "SensLoc: sensing everyday places and paths using less energy," in *MobiSys 8th ACM International Conference on Embedded Networked Sensor Systems (SenSys '10)*, pp. 43–56, November 2010.

[37] K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi, "Mobile data offloading: how much can WiFi deliver?" in *Proceedings of the 6th International Conference on Emerging Networking Experiments and Technologies (Co-NEXT '10)*, p. 26, December 2010.

[38] A. Thiagarajan, J. Biagioni, T. Gerlich, and J. Eriksson, "Cooperative transit tracking using smart-phones," in *Proceedings of the 8th ACM International Conference on Embedded Networked Sensor Systems (SenSys '10)*, pp. 85–98, November 2010.

[39] A. Schulman, V. Navda, R. Ramjee et al., "Bartendr: a practical approach to energy-aware cellular data scheduling," in *Proceedings of the 16th Annual Conference on Mobile Computing and Networking (MobiCom '10)*, pp. 85–96, September 2010.

[40] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Indoor localization without infrastructure using the acoustic background spectrum," in *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys '11)*, pp. 155–168, July 2011.

[41] D. A. Johnson and M. M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in *Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems (ITSC '11)*, pp. 1609–1615, 2011.

[42] D. Mitrović, "Reliable method for driving events recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 198–205, 2005.

[43] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan, "Mobile phone based drunk driving detection," in *Proceedings of the 4th International Conference on Pervasive Computing Technologies for Healthcare*, pp. 1–8, March 2010.

[44] K. C. Baldwin, D. D. Duncan, and S. K. West, "The driver monitor system: a means of assessing driver performance," *Johns Hopkins APL Technical Digest*, vol. 25, no. 3, pp. 269–277, 2004.

[45] G. Ten Holt, M. Reinders, and E. Hendriks, "Multi-dimensional dynamic time warping for gesture recognition," in *Proceedings of the Conference of the Advanced School for Computing and Imaging (ASCI '07)*, 2007.

[46] R. Muscillo, S. Conforto, M. Schmid, P. Caselli, and T. D'Alessio, "Classification of motor activities through derivative dynamic time warping applied on accelerometer data," in *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '07)*, pp. 4930–4933, 2007.

[47] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "uWave: accelerometer-based personalized gesture recognition and its applications," *Pervasive and Mobile Computing*, vol. 5, no. 6, pp. 657–675, 2009.

[48] J. Kela, P. Korpipää, and J. Mäntyjärvi, "Accelerometer-based gesture control for a design environment," *Personal and Ubiquitous Computing*, vol. 10, no. 5, pp. 285–299, 2006.

[49] E. Miluzzo, A. Varshavsky, S. Balakrishnan et al., "Tapprints: your finger taps have fingerprints," in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, pp. 323–336, Low Wood Bay, Lake District, UK, 2012.

[50] L. Cai and H. H. Chen, "TouchLogger: inferring keystrokes on touch screen from smartphone motion," in *Proceedings of the 6th USENIX Conference on Hot Topics in Security*, p. 9, San Francisco, Calif, USA, 2011.

[51] E. Owusu, J. Han, S. Das et al., "ACCessory: password inference using accelerometers on smartphones," in *Proceedings of the 12th Workshop on Mobile Computing Systems & Applications*, pp. 1–6, San Diego, Calif, USA, 2012.

[52] L. Cai, S. Machiraju, and H. Chen, "Defending against sensor-sniffing attacks on mobile phones," in *Proceedings of the 1st ACM Workshop on Networking, Systems, and Applications for Mobile Handhelds*, pp. 31–36, Barcelona, Spain, 2009.

[53] S. Agrawal, I. Constandache, and S. Gaonkar, "PhonePoint pen: using mobile phones to write in air," in *Proceedings of the 1st ACM Workshop on Networking, Systems, and Applications for Mobile Handhelds*, pp. 1–6, Barcelona, Spain, 2009.

[54] B. Clarkson, K. Mase, and A. Pentland, "Recognizing user context via wearable sensors," in *Proceedings of the 4th Intenational Symposium on Wearable Computers*, pp. 69–75, October 2000.

[55] S. Gaonkar, J. Li, and R. R. Choudhury, "Micro-Blog: sharing and querying content through mobile phones and social participation," in *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*, pp. 174–186, Breckenridge, Colo, USA, 2008.

[56] E. Miluzzo, N. D. Lane, and K. Fodor, "Sensing meets mobile social networks: the design, implementation and evaluation

of the CenceMe application," in *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, pp. 337–350, Raleigh, NC, USA, 2008.

[57] M. Azizyan and R. R. Choudhury, "SurroundSense: mobile phone localization using ambient sound and light," *SIGMOBILE Mobile Computing and Communications Review*, vol. 13, no. 1, pp. 69–72, 2009.

[58] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "BeepBeep: a high accuracy acoustic ranging system using COTS mobile devices," in *Proceedings of the 5th ACM International Conference on Embedded Networked Sensor Systems (SenSys '07)*, pp. 1–14, Sydney, Australia, November 2007.

[59] A. Mandai, C. V. Lopes, T. Givargis, A. Haghighat, R. Jurdak, and P. Baldi, "Beep: 3D indoor positioning using audible sound," in *Proceedings of the 2nd IEEE Consumer Communications and Networking Conference (CCNC '05)*, pp. 348–353, January 2005.

[60] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "BeepBeep: a high accuracy acoustic ranging system using COTS mobile devices," in *Proceedings of the 5th ACM International Conference on Embedded Networked Sensor Systems (SenSys '07)*, pp. 397–398, Sydney, Australia, November 2007.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 791981, 6 pages
http://dx.doi.org/10.1155/2013/791981

Research Article

Real-Time Localization Algorithm for Maritime Search and Rescue Wireless Sensor Network

Huafeng Wu,[1] Lei Yang,[1] Ling Liu,[1] Ming Xu,[1] and Xinping Guan[2]

[1] *Merchant Marine College, Shanghai Maritime University, Shanghai 200135, China*
[2] *School of Electronic, Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200240, China*

Correspondence should be addressed to Huafeng Wu; hfwu@shmtu.edu.cn

Received 25 October 2012; Revised 10 January 2013; Accepted 30 January 2013

Academic Editor: Yonggang Wen

Maritime search and rescue (MSR), as the last defense line for the life at sea, plays a great role in shipping industry and ocean environment protection. MSR is attracting more and more attention from technical and scientific aspects which are expected to help locate the victim goals quickly, precisely, and efficiently. So far, there have been some new techniques applied to MSR, such as computer vision. However, these techniques have a common drawback that they are too dependent on the effort of the search and rescue party, so that the victim goals can only passively wait for their search. To change this kind of passive status, wireless sensor network (WSN) is the deficiencies of the current maritime search and rescue. To make WSN localization suitable for the MSR, an improved microelectromechanical systems MEMS aided algorithm on the basis of triangle and centroid algorithm is proposed to locate and track the search targets in real time and more precisely. Finally, the simulation results show that the proposed method and algorithm are feasible and efficient.

1. Introduction

With the continuous development of the entire world shipping industry and the increasing number of ships at sea, it becomes a bigger challenge for the ships to be kept safe at sea, so the accidents often happen. When the accidents happen, the first thing is to search and rescue the floating staff to protect their lives and properties. The traditional measure is mainly using a plenty of search and rescue ships and helicopters to look for the floating staff, but the result is not ideal enough and it wastes a lot of resources [1]. Although many techniques like radar and computer vision have been applied to maritime search and rescue, it still cannot change the situation that the victim goals can only passively wait for the searching of the search and rescue party. Search and rescue consists of two parts: searching and rescuing, and searching and determining the position of the floating staff is the key to success. Wireless sensor networks WSNs are a kind of networks consisting of a large number of sensor nodes that have low cost and low power and having computing power and wireless communication capacity [2]. WSNs have the features of survival ability, accuracy and reliability, self-organization ability, scalability, and so on in harsh and dynamic environment. So the use of WSN localization can adapt to the complex environment well at sea.

For the environment at sea and the target of low cost and low power, this paper merges the advantages of range based localization and nonrange localization. We take RSSI measures to achieve the distances and propose an improved triangle and centroid algorithm, and when we get the original positions of floating staff, we can use MEMS technique to optimize the position information. Then we can get more accurate position information without cost and energy consumption and duplication iterations, and the algorithm computing process has a low complexity.

The remainder of the paper is organized as follows. In Section 2, we present the related work. In Section 3, we present an improved MEMS aided localization algorithm based on the triangle and centroid algorithm. In Section 4, we evaluate the proposed method in this paper with simulation experiments. At last Section 5 concludes the paper.

2. Related Work

Huanhuan et al. [3] introduced the researches of WSN localization in and abroad. This paper analyzed the principles and methods of mobile WSN localization and 3D localization, andit can reduce the costs of beacon nodes based on mobile WSNs. But Huanhuan did not analyze the application scenarios that the algorithms suited for. Bearing in the mind the advantages and disadvantages of range-based localization and nonrange localization, Zhao and Chen [4] proposed that it can measure the point-to-point distance by correcting RSSI ranging and choosing the preferred beacon nodes, and at last used the weighed centroid algorithm to get the position. But the weighted thinking was not related to the work status of the device itself in the harsh environment. Yun et al. [5] presented a localization algorithm using expected hop progress based on non-range localization because the accuracy of distance measurement is one of the key points in localization. Thus it can calculate the distance between any two sensor nodes and expand the range and accuracy of the localization. He et al. [6] presented APIT, a novel localization algorithm that was range-free and showed that their APIT scheme performed best when an irregular radio pattern and random node placement were considered, and low communication overhead was desired. In addition, they studied the effect of location error on routing and tracking performance. Awad et al. [7] proposed several discuss-and-analyze methods using RSSI to reduce the cost as RSSI did not need additional hardware. Also they made a comparison between two distance estimations, statistical method and artificial neural network method. But the localization accuracy of the algorithm was limited. Though the cost was low, it was not appropriate to be used in maritime search and rescue that required high localization accuracy. Li et al. [8] proposed a range-based particle filtering localization method for static WSN. Firstly, the state-space function of localization in a static WSN was proposed in the paper. Secondly, important steps in a particle filter were specifically illustrated in the paper, including initialization, prediction, sequential importance sampling, and the resampling.

Cheng and Du [9] took into account that the unbelievable GPS observations may cause positioning errors, so they gave the observations different weights using weighted least square method. The observations which are nearer can get a larger weight, and it helps to achieve more accurate position information. But Cheng and Du did not elaborate the development of WSN localization. Caballero et al. [10] proposed a new idea that combined WSNs positioning technology with robots, called network robot system (NBS). It can be divided into two parts. First, they used robots to get the original estimations of localization and then refined these estimations of localization with distributed information filtering method. But the method was established in the condition that the robot moved in accordance with the predesigned route and did not consider the random movement with wind and waves at sea. Lei and Lili [11] proposed WSNs self-positioning method based on genetic algorithm. At first they had a preliminary estimation of original position information using sampling method and then used genetic algorithm to compute for more accurate results. This algorithm can locate precisely even the anchor nodes which only account for a small proportion. But Lei and Lili did not take optimization measures in obtaining the original position of the unknown nodes. Dong et al. [12] gave a new localization algorithm based on mobile beacon nodes. The algorithm used a beacon node to traverse the entire network, and the node periodically broadcast a packet that contains the location information. At last they located using maximum likelihood estimation method instead of the trilateral localization algorithm to improve the positioning stability of the algorithm. Shang and Ruml [13] proposed a new variant of the MDS-MAP (multidimensional scaling map) method using patches of relative maps, that can be executed in a distributed fashion. The main idea was to build a local map at each node of the immediate vicinity and then merges these maps together to form a global map. Jun et al. [14] presented Non-intEractive lOcation Surveying (NEOS) to address certain deficiencies in the existing approaches. The key contribution was twofold: (i) it employed a mobile beacon to introduce mobility-differentiated time-of-arrival (MDToA) observations, a special form of time difference of arrival (TDoA), at the node side, and (ii) it involved simple computation and entailed no node-to-node communication. Tang et al. [15] studied a node localization algorithm based on moving beacon node based on the classical DV hop algorithm because of the limit energy of the node.

The previous researches have a close relationship with WSN localization algorithm at sea. Because of the unique application environment at sea, it is necessary to find a localization algorithm that is easily operated and simply calculated and has low cost and low power.

3. An Improved MEMS Aided Localization Algorithm Based on the Triangle and Centroid Algorithm

3.1. Application Scenarios at Sea for WSN Localization. As one of the supporting technologies of WSN, localization can be applied to locate the floating staff. Installing the wireless sensor nodes in the ships in distress, the survival craft, lifejackets, and other survival equipments, these nodes are artificially opened or automatically start in the case of seawater when the shipwreck occurs and self-organize into a wireless sensor network through ZigBee protocol. The nodes carried on the searching targets are defined as the unknown nodes, and the ships in distress, survival craft, rescue helicopters, and tossed sensor nodes are defined as the anchor nodes, and these anchor nodes can know their positions from GPS or other devices. The unknown node can locate itself with the actual positions of anchor nodes and send its position to the rescue organization or personnel or share with the relevant departments or personnel by connecting to the internet through the satellites. We would change the situation that the maritime search and rescue targets can only passively wait for their search, thus make the position indication of the searched targets initiative.

3.2. Improved Triangle and Centroid Localization Algorithm Using RSSI. RSSI is a value that indicates the quantity of the electromagnetic energy in the current media. When the wireless signal spreads in the air, the signal strength will decease with the distance increasing because of various factors from the outdoor environment. Changes in signal strength have a sort of functional model with the spreading distance, and usually the module measuring the signal strength can be easily installed in the sensor nodes. The transmission power is known, and we measure the received power in the receiving part, so we can calculate the propagation loss and use theory or experience signal propagation model to make propagation loss into distance. RSSI mainly uses RF signal, and it is a ranging way of low power and low cost. Since the error is generated from the reflection, multipath propagation, NLOS (nonline of sight), and other factors, RSSI ranging usually uses the propagation path-loss model as the following shows [16]:

$$\mathrm{PL}(d) = \mathrm{PL}(d_0) + 10n \lg\left(\frac{d}{d_0}\right) + X_\sigma. \tag{1}$$

In the formula, $d_0 = 1$ m is the reference distance; $\mathrm{PL}(d_0)$ is the loss after the distance d_0; d is the actual distance; $\mathrm{PL}(d)$ is the loss from anchor nodes to unknown nodes; X_σ is the shadowing factor whose average value is 0; n is the path-loss exponent and depends on the environment around. Change formula (1) into another form:

$$d = 10^{(\mathrm{PL}(d)-\mathrm{PL}(d_0)-X_\sigma)/10n} \times d_0. \tag{2}$$

In fact, in the real application environment of maritime search and rescue, it will produce ranging errors whatever kind of path-loss models you choose, and the distance calculated from anchor nodes to unknown nodes is always longer than their actual distance. As Figure 1 shows, A, B, and C are anchor nodes and D is an unknown node. So we can calculate the distance d_A from A to D, the distance d_B from B to D, and the distance d_C from C to D according to the principle of RSSI ranging. Then we make circles that use A, B, C as the center and d_A, d_B, and d_C as the radius and we will get the cross-area, in the figure. The basic idea of the triangle and centroid algorithm is calculating the three coordinates of the triangle in the cross-area, and the position of the unknown node is the centroid of the triangle. As Figure 2 shows, E, F, and G are vertex of the triangle, and the calculation method of E coordinate $(x_\mathrm{e}, y_\mathrm{e})$ is as the following shows [17]:

$$\begin{aligned}
\sqrt{(x_e - x_a)^2 + (y_e - y_a)^2} &\le d_\mathrm{A},\\
\sqrt{(x_e - x_b)^2 + (y_e - y_b)^2} &= d_\mathrm{B},\\
\sqrt{(x_e - x_c)^2 + (y_e - y_c)^2} &= d_\mathrm{C}.
\end{aligned} \tag{3}$$

Similarly we can have the coordinates of $\mathrm{F}(x_f, y_f)$ and $\mathrm{G}(x_g, y_g)$, so we obtain the coordinate $\mathrm{D}((x_e+x_f+x_g)/3, (y_e+y_f+y_g)/3)$ according to the formula of triangle centroid. It is worth noting that the number of anchor nodes around the unknown node and making intersection is not exactly three. When the number of anchor nodes is less than three, it will not locate. When the number of anchor nodes is three, it locates according to the method mentioned previosly. When the number of anchor nodes is more than three, we should choose three anchor nodes nearest to the unknown node, as the closer the distance, the smaller the error.

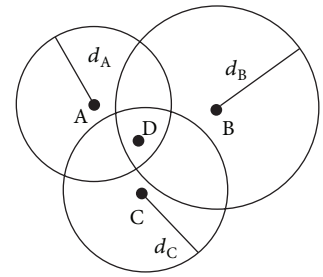


Figure 1: Cross area of three circles.

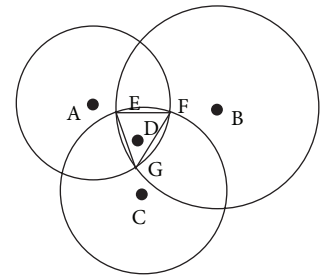


Figure 2: The schematic of triangle and centroid algorithm.

In order to get more accurate localization information, we plan to introduce the weighted thinking into the aforementioned algorithm, in other words, we give every coordinate of the triangle a weight. The anchors in this paper are mainly divided into four categories: the ships in distress, survival craft, rescue helicopters, and tossed sensor nodes. We decide to give them each different weights to distinguish their relative weighted relationship according to the status stability in the complex and bad environment at sea. The ship in distress is the mother ship close to the searching targets and its work status is stable, so we give the ships in distress the weight r_1. The survival craft is mainly well furnished and has good work status, so the weight of the survival craft is r_2. Because of the good maneuvering performance of the rescue helicopter, we give the rescue helicopter the weight r_3. The tossed sensor node will move with the wind and waves, so its work status is not stable enough and the weight of the tossed sensor node is r_4. The specific weight value which depends on the environment at sea is unique at that time. And r_1, r_2, r_3, and r_4 satisfy the following conditions:

$$\begin{aligned}
r_4 < r_3 < r_2 < r_1,\\
r_1 + r_2 + r_3 + r_4 = 1.
\end{aligned} \tag{4}$$

Imagine A is the ships in distress, B is the survival craft, and C is the rescue helicopters. For example, E is decided by the

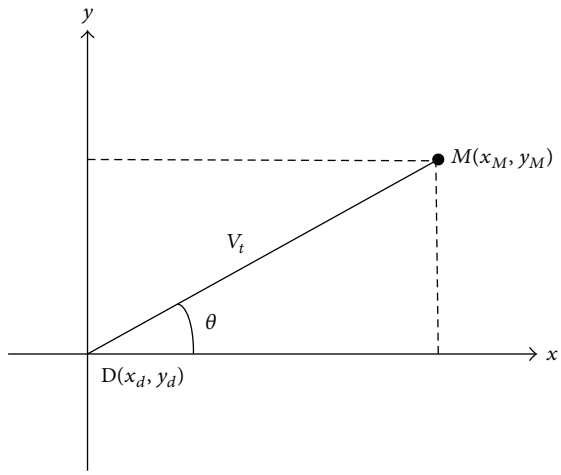


Figure 3: The schematic of MEMS localization.

circles B and C, so its weight is $(r_2 + r_3)$. Of course, if A, B, and C are other kinds of anchor nodes, it is also applicable. At last we can get the coordinate of D according to the following principle:

$$x_d = \frac{x_e(r_2 + r_3) + x_f(r_1 + r_3) + x_g(r_1 + r_2)}{6(r_1 + r_2 + r_3)},$$
$$y_d = \frac{y_e(r_2 + r_3) + y_f(r_1 + r_3) + y_g(r_1 + r_2)}{6(r_1 + r_2 + r_3)}. \quad (5)$$

After obtaining the original position information of the unknown node, it cannot avoid producing some degree of mobility taking into account the role of wind and waves. So in order to make the localization more accurate, we introduce a microelectromechanical systems (MEMS) aided method to optimize the original position of the unknown node. MEMS is a kind of microdevice or system that is available for bulk manufacture and contains miniature institutions, microsensors, microactuators and signal processing and control circuit. MEMS combines electronic technology with mechanical characteristics organically, and it can simultaneously achieve the function of physics, chemistry, biology, and so on. One of the characteristics of MEME technology is that the volume of the device is very small: the biggest is less than one centimeter and many are only several micrometers, and the thickness is even smaller. So it is suitable for the floating staff to fix in the lifejacket and carry on; the most important is that it does not affect the overall status and performance of the lifejackets.

This paper tries to integrate an acceleration sensor and a gyroscope in the unknown node with MEMS technology to monitor the state of motion of the unknown node after localization, such as speed and direction. Thus we can achieve the goal of real-time localization but only locate once and do not need to waste much cost and energy to locate again and again. As Figure 3 shows, after a period of time T, the unknown node moves to M. The speed is v, and the angle between DM and coordinate axis is θ. So the coordinate of D is

$$x_M = x_d + \int_0^T (vt\cos\theta)\,dt,$$
$$y_M = y_d + \int_0^T (vt\cos\theta)\,dt. \quad (6)$$

3.3. Design Process of the Algorithm

(1) The anchor nodes periodically transmit beacon information that contains the node ID and position information.

(2) After the unknown node receives all the beacon information, it uses formula (1) to transform the RSSI into distance.

(3) When the number of anchor nodes that the unknown node receives is not more than prescriptive n, we will make a descending order of the anchor nodes ID, positions and distance to anchor nodes to establish three sets,

(i) anchor nodes set: beacon = $\{A, B \cdots N\}$,
(ii) distance set: distance = $\{d_A, d_B \cdots d_N\}$,
(iii) positions set: position = $\{(x_a, y_a) \cdots (x_n, y_n)\}$.

(4) Choose three anchor nodes whose RSSI is the biggest and use formula (3) to obtain the coordinate of the triangle. Then get the position estimation using improved triangle and centroid algorithm in formula (5).

(5) After obtaining the original position information of the unknown node, we can observe the speed and direction of the node during the period from localization to finding the node and use formula (6) to obtain the final position information of the unknown node.

If the actual position of the unknown node is (x, y) the localization error is [18],

$$\mathrm{ER} = \sqrt{(x_M - x)^2 + (y_M - y)^2}. \quad (7)$$

The flowchart of this localization algorithm is shown as Figure 4.

4. Simulations and Analysis

The paper uses MATLAB 7.10 platform for computer simulation. Parameters are set as follows: in a square sensing area of 1000 m × 1000 m, there are 5 to 20 anchor nodes and 50 unknown nodes around. The path-loss exponent: $n = 2.4$, $X_\sigma = 7.98$ dB. The time from localization to finding the unknown node: $T = 2$ min. The weights of the anchor nodes are set like this: $r_1 = 2/5$, $r_2 = 3/10$, $r_3 = 1/5$, $r_4 = 1/10$. The simulation will analyze the algorithm (WTCA) in this paper, the trilateral localization algorithm (TLA), and the triangle-centroid algorithm (TCA), respectively [19].

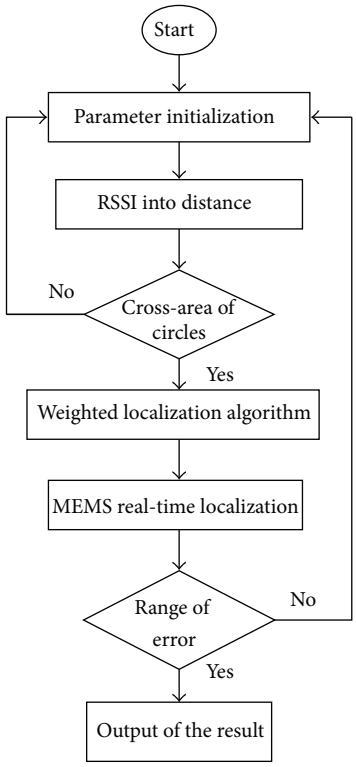


Figure 4: Algorithm flowchart.

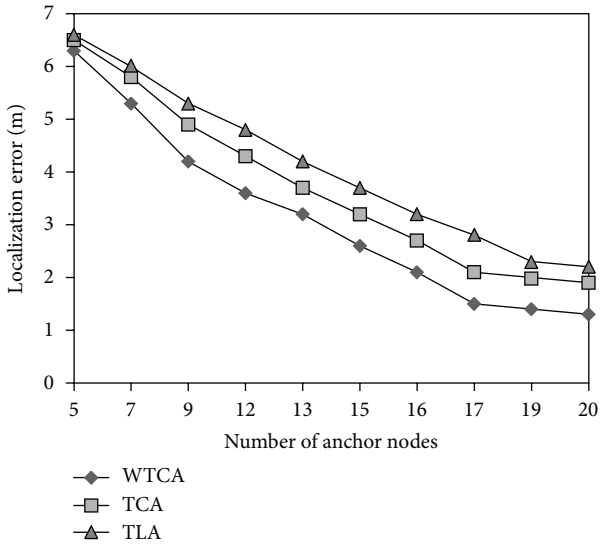


Figure 5: Localization error versus number of anchor nodes.

Figure 5 shows that in the case of the same number of anchor nodes, the error of this algorithm is smaller than the trilateral localization algorithm and the triangle-centroid algorithm. The main reason is that the trilateral localization algorithm and the triangle-centroid algorithm do not think of the work status of all kinds of anchor nodes, and different kinds of anchor nodes have different weights to distinguish their relative relationship.

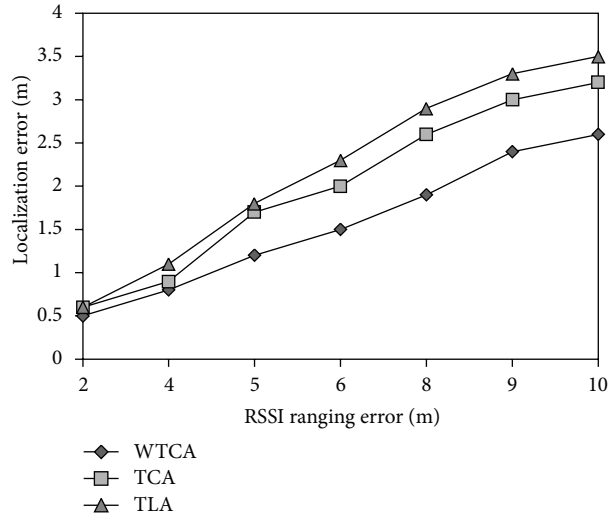


Figure 6: Localization error versus RSSI ranging error.

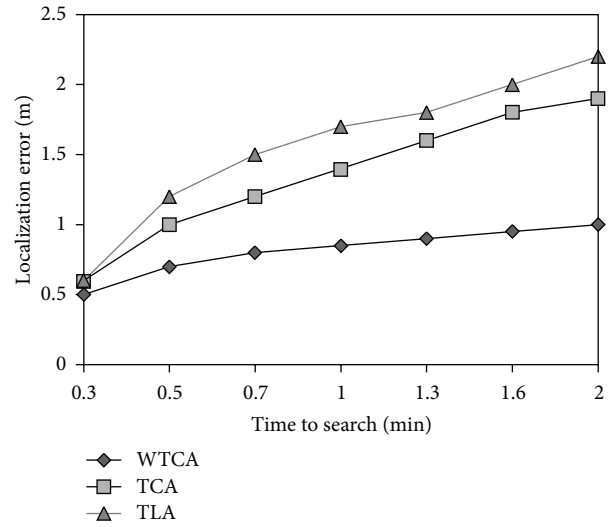


Figure 7: Localization error versus time to search.

Figure 6 shows that in the case of the same RSSI ranging error, the error of the algorithm in this paper is smaller than the trilateral localization algorithm and the triangle-centroid algorithm also because the work status of different anchor nodes is different and they have different weights.

Figure 7 indicates that if going over the same time from localization to finding the target, compared with the trilateral localization algorithm and the triangle-centroid algorithm, our algorithm has smaller localization error and is more accurate. The reason is that this algorithm thinks of the dynamic environment at sea and uses MEMS aided method to observe the speed and direction of the maritime search and rescue target, so we can achieve more accurate position information.

5. Conclusion

The paper has presented a new localization and tracking algorithm that is suitable for maritime search and rescue in wireless sensor networks based on many related researches. The algorithm is mainly divided into three stages: RSSI ranging, improved triangle and centroid algorithm and MEMS aided real-time localization method. This method combines the range-based localization algorithm with non-range localization algorithm; it can achieve the goal of real-time localization and avoid repeated computation. The simulation results have indicated that the algorithm is effective. Compared with the trilateral localization algorithm and the triangle-centroid algorithm, the new algorithm has higher accuracy and is more suitable for the mobility at sea. Also the use of wireless sensor networks can achieve an initiative position indication by the searched targets themselves. So the algorithm proposed in this paper can get more accurate position information without consuming too much energy.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51279099, 61202370), the Shanghai Natural Science Foundation (12ZR1412500), the Innovation Program of Shanghai Municipal Education Commission (13ZZ124), and the "Shu Guang" Project supported by the Shanghai Municipal Education Commission and Shanghai Education Development Foundation (12SG40).

References

[1] P. Xudong, "The application of the internet of things in searching and rescuing lives at sea," *China Water Transport*, no. 7, pp. 38–39, 2010.

[2] Q. Yan, Z. Chongchong, D. Guilan, and H. Changjun, "Rearch on localization technology for wireless sensor networks," *Computer Science*, no. 5, pp. 47–50, 2008.

[3] W. Huanhuan, Z. Xiaojia, and Y. Bin, "Rearch on localization technology for wireless sensor networks," *Journal of Shandong Polytechnic University*, no. 1, pp. 59–62, 2011.

[4] Z. Zhao and X. H. Chen, "An improved localization algorithm based on RSSI in WSN," *Chinese Journal of Sensors and Actuators*, vol. 22, no. 3, pp. 391–394, 2009.

[5] W. Yun, W. Xiaodong, W. Deming, and D. P. Agrawal, "Range-free localization using expected hop progress in wireless sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, 20, no. 10, pp. 1540–1552, 2009.

[6] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking (MobiCom '03)*, pp. 81–95, September 2003.

[7] A. Awad, T. Frunzke, and F. Dressler, "Adaptive distance estimation and localization in WSN using RSSI measures," in *Proceedings of the 10th Euromicro Conference on Digital System Design Architectures, Methods and Tools (DSD '07)*, pp. 471–478, August 2007.

[8] Y. Li, M. Q. H. Meng, S. Li, W. Chen, and H. Liang, "Particle filtering for range-based localization in wireless sensor networks," in *Proceedings of the 7th World Congress on Intelligent Control and Automation (WCICA '08)*, pp. 1629–1634, June 2008.

[9] B. Cheng and R. Du, "An accurate GPS-based localization in wireless sensor networks: a GM-WLS method," in *Proceedings of the International Conference on Parallel Processing Workshops*, pp. 33–41, 2011.

[10] F. Caballero, L. Merino, P. Gil, I. Maza, and A. Ollero, "A probabilistic framework for entire WSN localization using a mobile robot," *Robotics and Autonomous Systems*, vol. 56, no. 10, pp. 798–806, 2008.

[11] Z. Lei and D. Lili, "WSN node localization technology based on genetic algorithm," *Computer Engineering*, vol. 36, no. 10, pp. 85–87, 2010.

[12] Q. F. Dong, Y. J. Feng, and L. Yu, "Location algorithm based on the mobile beacon node in wireless sensor networks," *Chinese Journal of Sensors and Actuators*, vol. 21, no. 5, pp. 823–827, 2008.

[13] Y. Shang and W. Ruml, "Improved MDS-based localization," in *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '04)*, pp. 2640–2651, March 2004.

[14] L. Jun, H. V. Shukla, and J. P. Hubaux, "Non-interactive location surveying for sensor networks with mobility-differentiated ToA," in *Proceedings of the 25th IEEE International Conference on Computer Communications (INFOCOM '06)*, April 2006.

[15] L. Tang, W. Chai, X. Chen, and J. Tang:, "Research of WSN Localization Algorithm Based on Moving Beacon Node," in *Proceedings of the 3rd Pacific-Asia Conference on Circuits, Communications and System (PACCS '11)*, vol. 2011, pp. 1–5.

[16] C. Juan, L. Chang-Geng, and N. Xin-Xian, "Node localization of wireless sensor networks based on mobile beacon," *Chinese Journal of Sensors and Actuators*, vol. 22, no. 1, pp. 121–125, 2009.

[17] L. Wei and C. Chuanfeng, "RSSI-based triangle and centroid location in wireless sensor network," *Modern Electronics Technique*, 32, no. 2, pp. 180–182, 2009.

[18] R. Weizheng, X. Lianming, D. Zhongliang, and W. Chuan, "Distance difference localizaton algorithm based on RSSI for wireless sensor networks," *Chinese Journal of Sensors and Actuators*, no. 7, pp. 1247–1250, 2008.

[19] K. Z. Liu, J. F. Zhang, F. P. Hu, and S. Wang, "Distributed node localization scheme using moving beacons in wireless sensor networks," *Journal of Beijing University of Posts and Telecommunications*, vol. 33, no. 2, pp. 16–33, 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 615282, 9 pages
http://dx.doi.org/10.1155/2013/615282

Research Article

Mobility Crowdsourcing: Toward Zero-Effort Carpooling on Individual Smartphone

Nianbo Liu,[1] Yong Feng,[2] Feng Wang,[1] Bang Liu,[1] and Jinchuan Tang[1]

[1] *School of Computer Science and Engineering, University of Electronic Science and Technology of China, North Jianshe Road, Chengdu 611731, China*
[2] *Yunnan Key Laboratory of Computer Technology Applications, Kunming University of Science and Technology, Kunming 650500, China*

Correspondence should be addressed to Yong Feng; fybraver@163.com

Received 8 November 2012; Accepted 22 January 2013

Academic Editor: Ming Liu

In current carpooling systems, drivers and passengers offer and search for their trips through available mediums, for example, accessing carpool website by smartphone, for finding a possible match of the journey. While efforts have been made to achieve fast matching for known trips, the need for accurate mobile tracking for individual users still remains a bottleneck. For example, drivers feel impatient to input their routes before driving, or centralized systems haves difficulties to track a large number of vehicles in real time. In this paper, we present the idea of Mobility Crowdsourcing (MobiCrowd), which leverages private smartphone to collect individual trips for carpooling, without any explicit effort on the part of users. Our scheme generates daily trips and mobility models for each user, and then makes carpooling zero-effort by enabling travel data to be crowdsourced instead of tracking vehicles or asking users to input their trips. With prior mobility knowledge, one user's travel routes and positions for carpooling can be predicted according to the location of the time and other mobility context. Based on a realistic travel survey and simulation, we prove that our scheme can provide efficient and accurate position estimation for individual carpools.

1. Introduction

Nowadays, quick and easy transportation has been an essential part of modern society. Vehicles offer flexibility and mobility when it comes to our work-related and personal lives, enable rapid and timely delivery of goods, but can also cause traffic jams, carbon emissions, pollution, accidents, energy crisis issues, and other problems, inevitably. On the one hand, vehicle transportation plays a vital role in global economy; therefore, any efficiency improvement will yield great profits. On the other hand, the efficiency of vehicle transportation points out how we use vehicles and, also, the degree of expensive social costs. Although lots of institutions, resources, and research are dedicated to improving transportation efficiency, the waste of transportation capacities is still ubiquitous in current vehicle transportation.

According to NHTS [1] from the US Department of Transportation, the average occupancy rate of personal vehicle trips is 1.6 persons per vehicle mile. Since a regular vehicle carries 5 persons in full occupancy, 68% of transportation capacities are wasted during personal trips. In US alone, this involves 204 million personal vehicles and causes a great deal of loss. Similarly, such inefficiency has also been observed in business transportation. Taxies, vans, trucks, and other vehicles are often running in low occupancy or utilization or are sometimes even unoccupied. The oversupply does not just come from wasted transportation capacities, but also comes from information opacity between supply and demand.

Carpooling seems an effective method to achieve green and efficient transportation. Traditional carpools are neighbors or coworkers with similar routes, who can easily contact with each other for possible carpooling. Casual carpools, as impromptu carpools formed among strangers, can team up in public areas near HOV lanes, but it is severely limited in deployed roads [2]. Later carpool associations allow people to match their respective trips through the Internet, even if they are strangers. In the "dynamic carpooling" concept, casual carpooling is proposed, and many researchers try

to achieve this aim through special designed systems. In such carpooling, drivers and passengers offer and search for their trips through available mediums, for example, accessing carpooling website by smartphone, for finding a possible match of the journey. However, these systems failed to provide convenient and flexible carpool services for common users. While efforts have been made to achieve fast matching for known trips, the need for accurate mobile tracking for individual users still remains a bottleneck in current carpooling systems. For example, drivers feel impatient to input their routes before driving, or centralized systems have difficulties to track a large number of vehicles and to acquire their real-time positions.

Crowdsourcing has been evolving as a distributed problem solving and business production model in recent years, which was proposed to reduce production costs and make more efficient use of labor and resources with the aids of participators. An example of crowdsourcing tasks is seen in indoor localization, the system Zee [3] makes the calibration zero-effort, by enabling WiFi fingerprint training data to be crowdsourced without any explicit effort on the part of users. In this paper, we present the idea of MobiCrowd, which leverages private smartphone to collect individual trips for carpooling, without any explicit effort on the part of users. Our scheme acquires location information from smartphone, generates daily trips and mobility models for each user, and then makes carpooling zero-effort by enabling travel data to be crowdsourced instead of tracking vehicles or asking users to input their trips. With prior mobility knowledge, a user's travel routes and positions can be predicted according to the location of the time, and then possible carpooling can be arranged to fit the mobility context. Based on a realistic travel survey and simulation, we prove that our scheme can provide efficient and accurate position estimation for individual carpools.

The remainder of this paper is structured as follows: Section 2 presents a brief overview of related work. In Section 3, we explain the design of MobiCrowd step by step, including human mobility patterns, system overview, driving sensing, trajectory prediction, and position estimation. Section 4 evaluates our scheme via a survey and simulation, while Section 5 finally summarizes the paper.

2. Related Work

Based on the techniques presented in current carpool web sites, taxi service centers, and bus-tracking systems, some centralized carpooling systems are proposed. Composite traffic [4] builds hypothetical large-scale public transport for the Helsinki metropolitan area, where a centralized system collects the information on all trip demands online and then merges the trips with the same origin and destination into public vehicles with eight or four seats. Hartwig and Butchmann [5] investigate the challenges in casual carpooling and suggest a travel matching system by extending current cell phone services. Dynamic Ride Sharing Community [6] carries out a carpooling system over Traffic Information Grid so that users can order the services via Internet, in-vehicle terminal, PDA, and mobile phone. Lue and Colorni [7] develop a practical scheme for the carpooling in a university, where users are informed immediately in case of delay or changes via e-mail or short messages. In [8], Chen and Regan explore the feasibility and challenges of WiFi-based carpooling systems in metropolitan locations and indicate that WiFi connectivity works well while vehicles are traveling at slow speed. In [9], Lalos et al. describe how positioning systems can be utilized to support a dynamic network of car and taxi pool services. In [10], Agatz et al. define dynamic carpooling and outline the optimization challenges that arise when developing technology to support carpooling.

Generally, the centralized approaches have some intrinsic drawbacks. On one hand, collecting vehicle trip is an obstacle to the system, where available carpooling relies on the numbers of the drivers willing to input their trips. It often leads to a closed system running on a few registered vehicles and excludes the extensive unregistered ones. On the other hand, the centralized framework is not scalable enough and flexible enough. Vehicles are often coming late or early for traffic jams and other reasons. For a service center, tracing all vehicles and distributing updates to each passenger can be highly expensive.

Alternatively, approaches based on distributed framework to implement dynamic carpooling have also been considered. In [11], Winter and Nittel propose ad hoc shared-ride trip planning in Mobile Geosensor Networks, and look into the communication strategies among mobile agents in colocated geographical space. Intelligent Self-Organizing Transport [12] develops the idea of shared-ride trip planning and demonstrates the feasibility of an ad hoc carpooling system. Our previous work, Vehicle-to-Passenger Communication [13], combines distributed carpooling and vehicular communication and then develops a roadside vehicle calling over vehicular ad hoc networks. However, the main disadvantage of these approaches is the assumption that vehicles support ad hoc communication. It means that such schemes are not practical until the wide deployment of vehicular communication device.

Compared with previous carpooling schemes, MobiCrowd decouples carpool matching and mobile tracking, develops mobility crowdsourcing and prediction-based tracking, and results in a simple, flexible, and scalable solution to casual carpooling.

3. System Design

Before discussing the idea of MobiCrowd, we first investigate the background of human mobility and the feasibility of mobility prediction. Then, we give the global framework and explain the process of crowdsourcing step by step. Later, we explain how to detect driving with smartphone. Finally, we establish a trip history model to predict one's driving trajectory and then use a Markov model to estimate one's accurate position according to previous trajectory prediction.

3.1. Human Mobility. The research on human mobility has had rapid development in recent years. Some of them notice

the relationship between human social activities and geographic movements. The time-variant community mobility model [14] captures two properties of human mobility via empirical WLAN traces: skewed location visiting preferences and periodical reappearance of nodes at the same location. Another study [15] investigates the trajectories of 100,000 anonymous mobile phone users and finds that human trajectories show a high degree of temporal and spatial regularity: each individual can be characterized by a time-independent characteristic length scale and a significant probability to return to a few highly frequented locations. Recent research [16] has discussed the mobility patterns of 50,000 individuals from a 3-month-long record and shown a 93% potential predictability in user mobility across the whole user base, by measuring the entropy of each individual's movement.

According to NHTS, the majority of individual daily trips—87 percent—are taken by personal vehicle. The daily activities of an individual person, including "going to work," "having lunch," "shopping," and so on, often show regular features. As a kind of human activity, driving is controlled by individual drivers and follows their respective social activities, partially. Although some unexpected driving occurs, individual or household driving in a certain vehicle usually yields the same spatial and temporal features. For instance, a commuter always drives his/her car from home to office at 9:00, and from office to home at 17:00. It makes large amounts of vehicle movement predictable at the microscopic scale. Research in the field of transportation also validates the regularities, both in human mobility and vehicle mobility. The Mobidrive project [17, 18] monitored the trajectories of private cars by collecting their GPS data, which found these sorts of regular activities in the 30,000 trips performed by 320 correspondents over a six-week study. A mobility pattern is observed in Figure 1, which is constituted by the spatial distribution of those locations where a traveler has had six weeks of personal experience. The spatial regularities in vehicle mobility are marked by the gray lines of vehicle movement, in which two-to-four main locations (including home) cover more than 70% of the overall trips.

Since the daily driving activities of a user provide a high degree of temporal and spatial regularity, it is possible to establish some mobility models to describe such driving activities and to predict the related mobility features for carpooling.

3.2. System Overview. The proliferation of smartphone is motivating the research community to look at the ways for more reliable and more convenient carpooling with the support of smartphone. Today's smartphones are not only programmable, but also come with cellular and WiFi interfaces and a rich set of embedded sensors, including an accelerometer, digital compass, gyroscope, GPS, microphone, and camera, which enable great sensing and communicating abilities to play the role of carpooling terminals. Thus, smartphones carryed by carpools can automatically investigate driving behavior, record the location of the time, and establish mobility model for carpooling, without requiring any explicit effort of users. Additionally, it is very easy to identify one user

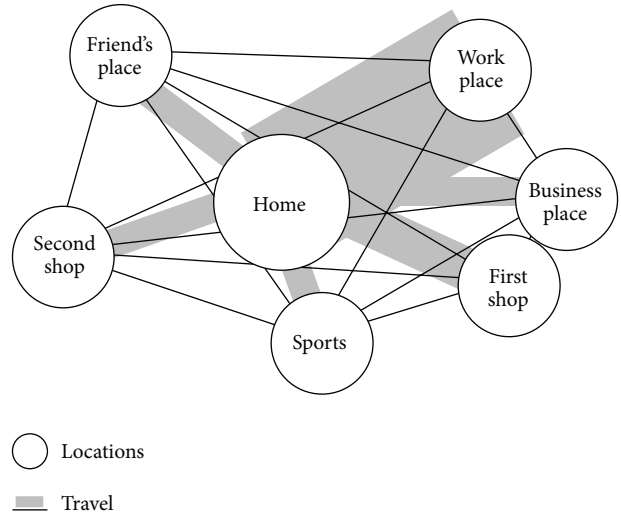


Figure 1: Mobility pattern in Mobidrive [17, 18].

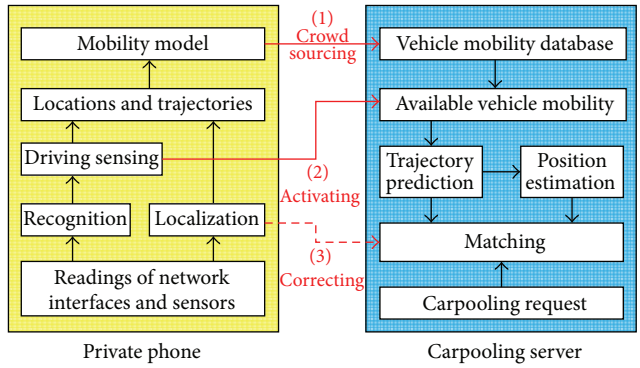


Figure 2: MobiCrowd framework.

by his/her private smartphone and to track each carpooling, which is helpful to establish insurance between unfamiliar carpools.

Figure 2 shows the envisioned phone/server architecture forcasual carpooling. We briefly describe the system components, followed by a discussion on challenges and solutions.

To one carpool participating in MobiCrowd, the smartphone will try to periodically locate itself in daily using as accurately as possible. Outdoor locations can be collected through GPS, while indoor locations could be found through WiFi and cellular localization services when the user accesses the Internet. A trail of the user's movement can be established as the ⟨*index, position, time*⟩ tuples, in the order of time sequence. At the same time, accelerometer and gyroscope readings are used to detect whether a driving event happens. During the period of driving, the trajectory is recorded and added up to local data set. After a period of learning, some daily locations, where the user often keeps unmoved for a long time, and the driving trajectories among these locations are combined into a mobility model similar to the one shown in Figure 1.

After building a mobility model for the user, the smartphone will upload it to a carpooling server to perform

the crowdsourcing of the user-specific mobility model. Thus, the server maintains many mobility models as carpooling resources. With this knowledge, it can manage mobility model, estimate driving trajectory and real-time vehicle position, response carpooling request from passengers, and then make proper vehicle-to-passenger matching. Compared to previous carpooling servers, the needs for active mobile tracking are removed. With mobility crowdsourcing, MobiCrowd cuts down a great deal of communication spending between server and each vehicle, by adapting position estimation instead of mobile tracking. However, the estimation is not always accurate. Some methods to correct estimation errors should be considered. At first, the smartphone should send a message to activate its model as soon as it detects driving, which avoids picking up false vehicle targets and provides accurate start time of each driving. Furthermore, the smartphone should estimate its position in driving according to its mobility model. Since the server and the smartphone maintain the same model, the smartphone can report correction message if the real position is far from the estimated value.

The above problems bring into question how to estimate the position of a moving vehicle accurately. Accurate estimation requires accurate mobility model, and accurate mobility model requires accurate location and trajectories and proper modeling methodology. If we want to collect location and trajectory data in the daily using of smartphone, we have to know what are the driving behavior, for example, the detailed sensor readings. At the same time, we have to predict the whole driving, not only the trajectory, but also the real-time position. The rest of this study is meant as a step towards a deeper understanding of these fundamental issues.

3.3. Driving Sensing. In our MobiCrowd scheme, the difficulty first arises from how to define driving with a smartphone, for example, with what kind of features of sensor readings we can regard a user is driving a car with his/her smartphone. Some recent works [19–21] have adopted smartphone to collect and analyze different driving behavior for providing driving safety, which prove the effectiveness of driving recognition. Human diving behavior can be regarded as a set of driving events and time, which can be detected by the sensors of a smartphone on a vehicle. However, excessive classifying and recognizing of driving behavior are not necessary to our study. In fact, we only focus on basic driving directly related to the location of the time of vehicle movement.

In this study, we adopt a simple method to detect driving by measuring the movement speed between two walking events. As shown in Figure 3(a), an acceleration signature is very easy to identify in human walking patterns [22]. This signature arises from the natural up and down bounce of the human body while walking and can be used to count the number of steps walked. For a typical driving, a user will walk to a car at one place, drive it to another place, park the car and finally walk to the destination. Thus, we can capture two successive walkings and record the end position of the first walking and the start position of the second one with timestamps, as ⟨*Position1, Time1*⟩ and ⟨*Position2, Time2*⟩, respectively. Compared to walking, driving does not have obvious characteristics in acceleration readings but shows a very high moving speed, for example, large displacement within small time interval. After calculating the Euclidean distance between the two positions, we can figure out the moving speed of the user. If the speed is not less than 30 km/h, we can regard that the user had a driving between two walking behaviors. Of course, this method cannot make very fine distinctions between driving and riding (e.g., taking a bus, taxi, subway, or flight). We simply regard that a driver proving carpooling often drives his/her own car, and errors can be removed in long-time trip history model in the next subsection.

3.4. Trajectory Prediction. Some mobility prediction schemes, Greedy Mobility Pattern agent [23], and the map matching algorithm [24] try to establish the mobility pattern in Figure 1 in an individual node by recording node-specific trajectories from GPS data. With trajectory history, a moving vehicle can compare its position to its previous trajectories and find out the most possible route.

However, such patterns also have some intrinsic drawbacks. First, the time criterion in vehicle movement is completely neglected. If a driver starts his/her car at home in a workday morning, he/she is driving to the work place most likely. If the driver does so in a nonworkday evening, the possibility that he/she is going to the work place is very slim. Since the trip history is simplified to the trajectory history, the temporal characteristics of vehicle movement are inevitably lost. Second, the overlapped trajectories can cause prediction errors. If a vehicle is moving at the only path to the driver's home, the route is clear. If the path leads to his/her home and favorite shop, the destination prediction by simple route pruning is questionable. Finally, the trajectory records bring heavy burden on data processing and storage, especially those from long distances or infrequent trips. Discarding these records may result in false predictions, whereas keeping them will exceed the capacity of the smartphone.

Thus, we need a new mobility model to predict vehicle movement, which should be spatiotemporal in mobility features, doubtless in the destination prediction, and lightweight in data size. Here we explain the idea of trip history by giving a simple example. In Figure 3, a series of trips taken by the driver can be kept as records in Table 1, in which temporal and spatial features of the trips are more concisely represented.

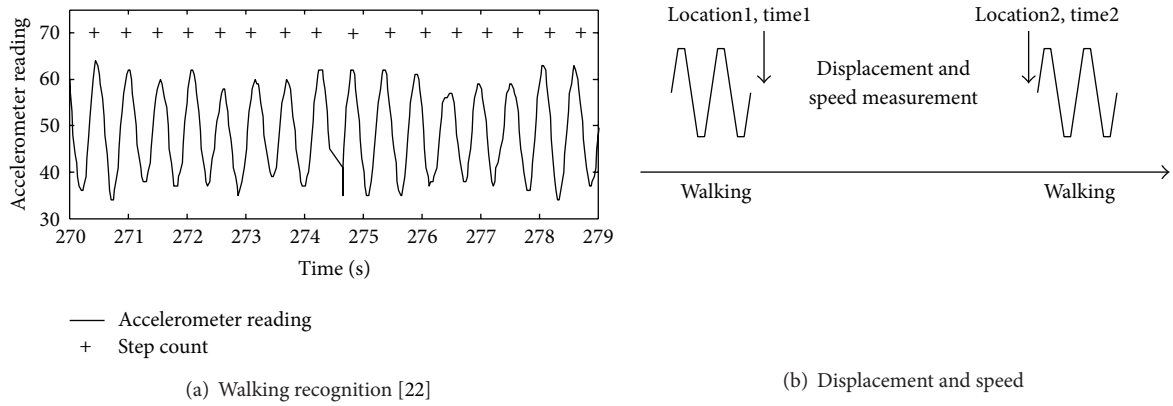
(1) To drive from home to work place at 8:40, Friday.

(2) To drive from work place to home at 17:35, Friday.

(3) To drive from home to sports at 14:05, Saturday, and send his wife to first shop on the way.

(4) To drive from sports to friend's home at 18:55, Saturday.

The locations "home," "work place," "business place," "first shop," "second shop," "sports," and "friend's place" are not accurate GPS positions, but rough regions in geography. Since the driver may have no special parking space in these



(a) Walking recognition [22]

(b) Displacement and speed

Figure 3: Driving sensing.

frequent visiting places, nearby parking positions within a certain scope, for example, within 500 m can be regarded as the same parking location of a given place. Similarly, the start time of the trip is separated into discrete time sets: Day (Mon., Tues., Wed., Thurs., Fri., Sat., and Sun.) and Time (<10, 10–12, 12–14, 14–16, 16–18, >18). Whenever a driver starts his/her car, MobiCrowd will log the start time and position as "Day," "Time," and "Source." During movement, it obtains continuous position data from the GPS every several seconds. When the car stops, the last position becomes "Destination." After finishing a trip, MobiCrowd can abbreviate the trip trajectory. If the trajectory from Source to Destination accords with the shortest path (not zero) in the electric map, it cancels all middle position records as record 1, 2, and 4. If the trajectory does not accord, it tries to find the shortest path from Source to later positions for as long as possible, cancels the middle positions from Source to the first "Midway Point," and repeats the procedure until arriving at Destination. Record 3 indicates such an abbreviation, in which the Midway Point can be the first shop. Sometimes, Source equals Destination in a trip. For example, the man drives his wife from home to second shop, and, after a live parking, he drives back to home finally. In this case, the trip is abbreviated similarly with a Midway Point as second shop. Since most daily driving trajectories take the shortest paths in geography, the trip records can be largely shortened in data size.

With the trip history, we develop a heuristic and context-dependent induction method based on decision trees, to predict vehicle moving trajectories. In data mining and machine learning, decision trees are widely used as the predictive tool mapping from observations about an item to conclusions about its target value. The related theories and algorithms can be found in [25]. When a car starts, MobiCrowd constructs a decision tree, where previous trip records in Table 1 are expressed as branches and leaves in Figure 4. In each leaf node, the probability of selecting a destination is given by

$$p_{kq} = \frac{f_{kq}}{N_k}, \tag{1}$$

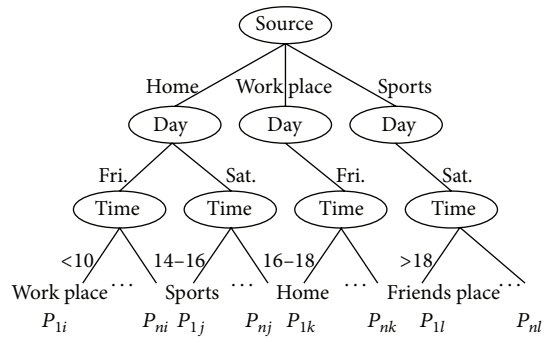


Figure 4: Decision tree structure.

where N_k is the total number of trips starting at root node k, and f_{kq} is the number of trips according with specific time, midway point, and destination choices at leaf node kq. Since category kq gives a final destination choice, p_{kq} represents the probability of moving toward the destination in history. Since the car knows current Source, Day, and Time values, it can find a Destination choice with a maximum p_{kq} value in the tree as an initiate prediction.

While driving, the smartphone periodically checks whether its position is on the way to the predicted destination or midway point. If the position disagrees with the predicted route, the car needs to calculate a new destination probability by

$$p_{kq} = \begin{cases} 0 & \text{if the route is impossible,} \\ \dfrac{f_{kq}}{\sum f_k} & \text{otherwise,} \end{cases} \tag{2}$$

where $\sum f_k$ is the total number of the rest of the trips after removing the infeasible ones. Then, it can find a new destination prediction with maximum probability p_{kq}. Consequently, our model can predict one's driving trajectory according to the trip history data of the user.

3.5. Position Estimation. Knowing a driving trajectory is still far from casual carpooling, for the accurate position

Table 1: Trip history record.

	Day	Time	Source	Midway point	Destination
1	Fri.	<10	Home		Work place
2	Fri.	16–18	Work place		Home
3	Sat.	14–16	Home	First shop	Sports
4	Sat.	>18	Sports		Friend's home

Table 2: Collected trips in travel survey.

Car	1–8 week	9 week	Total	Average
A	155	17	172	2.73
B	214	29	243	3.86
C	169	18	187	2.97
D	180	25	205	3.25
E	120	15	135	2.14

of the vehicle is necessary in real-time vehicle-to-passenger matching.

To predict the position of one vehicle based on driving trajectory, we can easily eliminate the influence of ambiguous past route. Thus, we model the sequence of vehicle's positions in one specific trajectory as $X(t)$, with t representing the time stamp which records the length of time from vehicle's driving trajectory to initially be calculated. That is to say, the ideal model is we just have to get the time stamp to find out the position of vehicle on the trajectory. And some positions on the trajectory are very different and cause the vehicle to make a turn at these places, and then we denote these place as landmarks. We simply denote the vehicle's positions as $\ldots, X(-3), X(-1), X(3), X(7), \ldots$, where $X(-3)$ is a landmark through which vehicle was passed within three time units. $X(3)$ and $X(7)$ are the landmarks into which vehicle is going to swerve within three and seven time units. The time between $X(t)$ and $X(t+1)$ is defined as a time unit. This means that the time unit is a constant and is not affected by road situation or time. Accordingly, the length between each two landmarks is different, and it takes different numbers of time units to go through them. At any time, when the vehicle swerves, this position becomes $X(-1)$. As knowing the driving trajectory, the chain of landmarks is clearly laid on it and predictable to be reached at certain time.

Even if we can predict the time when the vehicle will arrive at destination. It is still not easy to calculate vehicle's exact position just by time. So it naturally turns into the problem of probability theory field. We redefine the sequence of vehicle's positions as $\ldots, X(t_0), X(t_1), \ldots X(t_n)$, where n is the number of landmarks and $X(t_0)$ means that driver should make a first turn at t_0. By this way, we can eliminate the influence of different time units between every two landmarks, and there may exist several time units between t_i and t_j, where $0 \le i, j \le n$. Then, we denote $\Delta t_0 = t_1 - t_0, \ldots, \Delta t_{n-1} = t_n - t_{n-1}$ and separate the trajectory into certain segments. The function $Y[\Delta t_0]$ means the probability that vehicle is in the road segment between first and second landmark during Δt_0. The probability distribution of which segments that vehicle should be there is given by $P[Y(\Delta t_i)]$.

The Markov model leads one way to do position estimation task based on known trajectory. The first Markov model is $P[Y(\Delta t_i)]$ for the trajectory segments only depend on $Y(\Delta t_{i-1})$:

$$P\left[Y\left(\Delta t_i\right) \mid Y\left(\Delta t_{i-1}, Y\left(\Delta t_{i-2}\right)\cdots\right)\right] = P\left[Y\left(\Delta t_i\right) \mid Y\left(\Delta t_{i-1}\right)\right]. \quad (3)$$

This means that the position of vehicle can be estimated by knowing its last segment, and the last segment is determined by last two landmarks. The probability distribution for every segments is correlated with the past segments' probability distribution.

As usual, if we count two or more trajectory segments, higher accuracy will be got. The most apparent superiority compared with first Markov model is the sense of direction. We can build this model to be similar to the one mentioned above. However, in this paper, we have already known the direction at any time. In the result, there is only a few increments of accuracy by large number of segments to be counted.

Although the Markov model can be transformed to estimate the position, for instance, $P[Y(\Delta t_i) \mid Y(\Delta t_{i-2})]$, the distribution of the trajectory segment is based on the segment before the prior one. In our result, the nearest landmark plays the most important role in the process of estimation. The estimation accuracy is higher as the vehicle is closer to the last landmark. Markov model does not let us precisely get the vehicle's location. However, we can still estimate the trajectory segment, and more accurate position can be calculated just by knowing vehicle's speed. It is not hard to get because we have already known the trajectory and time.

4. Performance Evaluation

In order to evaluate MobilCrowd accurately, we first construct a realistic travel survey to demonstrate the trip history model and then examine the performance of position estimation in real driving.

4.1. Travel Survey. To evaluate the driving regularities for individual vehicles and the accuracy of our trip history model, a travel survey, taken over 9 weeks, has been performed. Five volunteers from the academic staff were invited to attend the survey, not involving any authors or contributors of this study. Each one processes a private car with GPS device. During the survey, they and their family members driving the cars were asked to note down the start time, source position, and destination position of each car, as well as the midway point positions, if there were any. Then, we collected the records and calculated trips for each car in Table 2. We found partial travel numbers from 135 to 243, due to the different household driving habits. And the average trips in each car, respectively, range from 2.14 per day to 3.86 per day.

After the survey, the trip records were translated into discrete results in Table 2 and input into the electric map of Chengdu City, China. All trips outside of the city scope were discarded. In order to evaluate the accuracy of predictions,

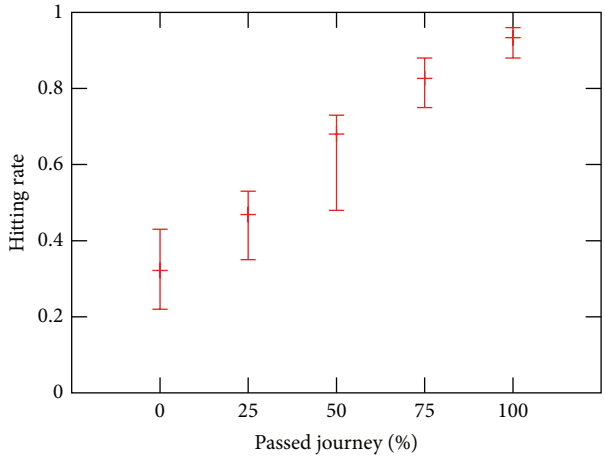


Figure 5: The accuracy of trajectory prediction.

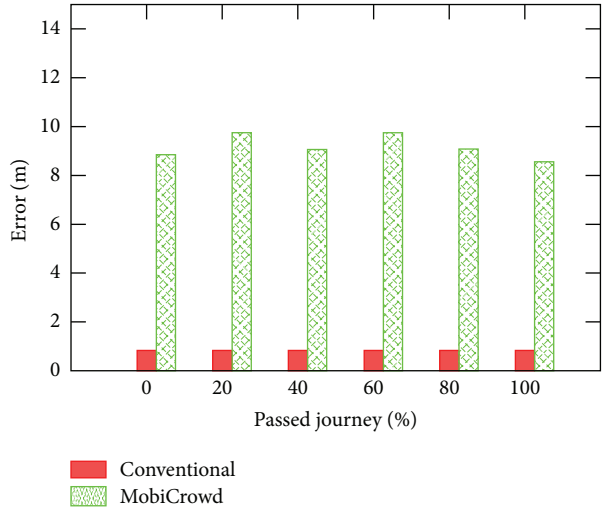


Figure 6: The accuracy of position.

we established a trip history model with the trips in the first eight weeks as a training set and validated the prediction with the trips in the last week as a target set. We examined the generated up-to-date predictions at different stages in the vehicle moving process. The hit rate, shown in Figure 5, demonstrates that the trip history model accords with the actual vehicle moving from 65.83% at the beginning to 97.04% at the end of the journey. It proves that the trip history model can provide an accurate prediction of vehicle movement, even when the vehicle first starts moving.

4.2. Simulation Results. In this study, we compare MobiCrowd with the conventional method in previous carpooling systems. Fortunately, GPS localization often shows consistent errors nearly 10 m, making it able to provide reference position. Thus, the conventional method, mobile tracking, enables smartphone to receive GPS signals and report the server periodically during the whole driving. Here we assume the smartphone reports its GPS position every 5 seconds.

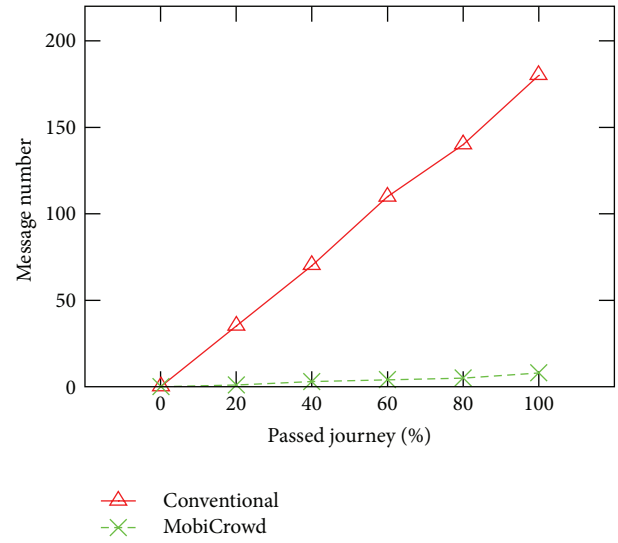


Figure 7: The numbers of generated messages.

As shown in Figure 2, MobiCrowd first sends an activating message to the server, estimates vehicle position, and sends correcting messages if necessary. Since correcting message is used, errors in trajectory prediction can be removed, for the smartphone can report its time and position. Here we set the max position error as 20 m, which is an acceptable distance to ordinary carpools. If the estimation error is beyond this limit, the smartphone will send a correcting message. In the simulations, all trip data are collected from the realistic survey in the last subsection.

As shown in Figure 6, the accuracy of position is different when carpooling server uses two methods to acquire the position of a moving vehicle. Using the conventional method, the accuracy equals the accuracy of GPS localization and keeps constant during the whole trip. With an interval of 5 second, the average errors are less than one meter at most of time. Using MobiCrowd with auxiliary correcting messages, the average errors are often nearly 10 meters. Even for casual carpooling, the accuracy of MobiCrowd is enough to provide real-time matching, without misleading drivers and passengers.

In Figure 7, we represent the numbers of generated messages of two method. Although both the methods produce messages linear to the driving time, MobiCrowd shows overwhelming advantages in the spending of communication. During the whole driving, MobiCrowd generates about ten messages averagely, while the conventional method exceeds 200 messages in average. Considering that a carpooling server often simultaneously tracks tens of thousand of vehicles or more, mobile tracking becomes a huge burden to the server and network inevitable. That is to say, the complexity and cost of carpooling system will increase sharply, which leads a bad performance and terrible user experience. On the contrary, MobiCrowd relies on position estimation based on mobility crowdsourcing, which is greatly reducing

the communication, simplifies the system architecture, and results in efficient and accurate carpooling service.

Generally, our MobiCrowd scheme shows great performance both in accuracy and spending. It also proves the value of mobility prediction in human daily activities.

5. Conclusion

Motivated the needs of mobile tracking in current carpooling systems, we propose MobiCrowd to achieve simple and flexible mobility crowdsourcing for carpools. The basic idea of MobiCrowd is simple: if private smartphone knows daily movement of the owner, why not let it predict the owner's driving trajectory and real-time position for carpooling?

In this paper we introduce human mobility patterns to investigate daily movement of carpools. Since carpools drive cars with their smartphones, we use smartphones to build mobility models to acquire accurate position of moving vehicles. First, we explain how to detect driving with smartphone. Then, we establish a trip history model to predict one's driving trajectory. According to trajectory prediction, we use a Markov model to estimate the accurate position of the moving vehicle. Finally, we prove that our scheme can provide efficient and accurate position estimation through realistic travel survey and simulation.

We believe that MobiCrowd has a bright future in next-generation traffic networks. In such networks, mobile tracking is no longer a bottleneck for casual carpooling, carpools are characterized as their mobility models, carpooling matching is made without any explicit effort, vehicles are running in higher occupancy and less traffic, and all of us benefit from green and efficient vehicle transportation.

Acknowledgment

This paper is supported in part by the China NSF Grants (61103226, 60903158, 61170256, 61173172, and 61272526), and the Fundamental Research Funds for the Central Universities Grants (ZYGX2010J074, ZYGX2011J102, and ZYGX2012J083).

References

[1] "National household travel survey," 2001, http://nhts.ornl.gov/introduction.shtml#2001.

[2] M. Burris, J. Winn, and W. Associates, "Slugging in Houston: casual carpool passenger characteristics," *Journal of Public Transportation*, vol. 9, no. 5, p. 23, 2006.

[3] A. Rai, K. Chintalapudi, V. Padmanabhan, and R. Sen, "Zee: zeroeffort crowdsourcing for indoor localization," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, pp. 293–304, ACM, 2012.

[4] J. T. Tuomisto and M. Tainio, "An economic way of reducing health, environmental, and other pressures of urban traffic: a decision analysis on trip aggregation," *BMC Public Health*, vol. 5, no. 1, article 123, 2005.

[5] S. Hartwig and M. Butchmann, "Empty seats travelling: next generation ridesharing and its potential to mitigate traffic and emission problems in the 21st century," Tech. Rep., Nokia Research Center, 2007.

[6] Y. Fu, Y. Fang, C. Jiang, and J. Cheng, "Dynamic ride sharing community service on traffic information grid," in *Proceedings of the International Conference on Intelligent Computation Technology and Automation (ICICTA'08)*, vol. 2, pp. 348–352, Hunan, China, October 2008.

[7] A. Lue and A. Colorni, "A software tool for commute carpooling: a case study on university students in Milan," *International Journal of Services Sciences*, vol. 2, no. 3-4, pp. 222–241, 2009.

[8] R. Chen and A. Regan, "Wireless networks for car-and ridesharing systems: assessment of 802. 11 Wi-Fi," *Transportation Research*, vol. 9, p. 56, 2009.

[9] P. Lalos, A. Korres, C. K. Datsikas, G. S. Tombras, and K. Peppas, "A framework for dynamic car and taxi pools with the use of positioning systems," in *Proceedings of the Computation World: Future Computing, Service Computation, Adaptive, Content, Cognitive, Patterns (ComputationWorld'09)*, pp. 385–391, IEEE, Athens, Greece, November 2009.

[10] N. Agatz, A. Erera, M. Savelsbergh, and X. Wang, "Sustainable passenger transportation: dynamic ride-sharing," Research Paper, 2010.

[11] S. Winter and S. Nittel, "Ad hoc shared-ride trip planning by mobile geosensor networks," *International Journal of Geographical Information Science*, vol. 20, no. 8, pp. 899–916, 2006.

[12] S. Winter, "Intelligent self-organizing transport," *Kunstliche Intelligenz*, vol. 22, no. 3, pp. 25–28, 2008.

[13] N. Liu, M. Liu, J. Cao, G. Chen, and W. Lou, "When transportation meets communication: V2P over VANETs," in *Proceedings of the 30th IEEE International Conference on Distributed Computing Systems (ICDCS'10)*, pp. 567–576, Genova, Italy, June 2010.

[14] W. J. Hsu, T. Spyropoulost, K. Psounis, and A. Helmy, "Modeling time-variant user mobility in wireless mobile networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (IEEE INFOCOM'07)*, pp. 758–766, Anchorage, Alaska, USA, May 2007.

[15] M. González, C. Hidalgo, and A. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[16] C. Song, Z. Qu, N. Blumm, and A. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[17] S. Schönfelder and S. Raumplanung, "Some notes on space, location and travel behaviour," in *Proceedings of the Swiss Transport Research Conference*, Ascona, Switzerland, 2001.

[18] S. Schönfelder and U. Samaga, "Where do you want to go today? More observations on daily mobility," in *Proceedings of the 3rd Swiss Transport Research Conference*, Ascona, Switzerland, 2003.

[19] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan, "Mobile phone based drunk driving detection," in *Proceedings of the 4th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth'10)*, pp. 1–8, IEEE, March 2010.

[20] D. Johnson and M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in *Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems (ITSC'11)*, pp. 1609–1615, IEEE, Washington, DC, USA, 2011.

[21] H. Eren, S. Makinist, E. Akin, and A. Yilmaz, "Estimating driving behavior by a smartphone," in *Proceedings of the 4th*

IEEE Intelligent Vehicles Symposium, pp. 234–239, IEEE, Alcala de Henares, Spain, 2012.

[22] I. Constandache, X. Bao, M. Azizyan, and R. R. Choudhury, "Did you see Bob? Human localization using mobile phones," in *Proceedings of the 16th Annual Conference on Mobile Computing and Networking (MobiCom'10)*, pp. 149–160, ACM, September 2010.

[23] J. Kurhinen and J. Janatuinen, "Geographical routing for delay tolerant encounter networks," in *Proceedings of the 12th IEEE International Symposium on Computers and Communications (ISCC'07)*, pp. 463–467, Aveiro, Portugal, July 2007.

[24] K. Miyashita, T. Terada, and S. Nishio, "A map matching algorithm for car navigation systems that predict user destination," in *Proceedings of the 22nd International Conference on Advanced Information Networking and Applications Workshops (AINA'08)*, pp. 1551–1556, Okinawa, Japan, March 2008.

[25] I. H. Whitten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Boston, Mass, USA, 2005.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 678214, 8 pages
http://dx.doi.org/10.1155/2013/678214

Research Article

Energy-Efficient Prediction Clustering Algorithm for Multilevel Heterogeneous Wireless Sensor Networks

Jian Peng,[1,2] Tang Liu,[1,2,3] Hongyou Li,[1] and Bing Guo[1]

[1] *College of Computer Science, Sichuan University, Chengdu 610065, China*
[2] *State key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[3] *College of Fundamental Education, Sichuan Normal University, Chengdu 610068, China*

Correspondence should be addressed to Tang Liu; crikey@163.com

Received 6 December 2012; Accepted 18 January 2013

Academic Editor: Ming Liu

In designing wireless sensor networks, it is important to reduce energy dissipation and prolong network lifetime. This paper presents research on the existing clustering algorithm applied in heterogeneous sensor networks and then puts forward an energy-efficient prediction clustering algorithm, which is adaptive to sensor networks with energy and objects heterogeneous. This algorithm enables the nodes to select the cluster head according to factors such as energy and communication cost, thus the nodes with higher residual energy have higher probability to become a cluster head than those with lower residual energy, so that the network energy can be dissipated uniformly. In order to reduce energy consumption when broadcasting in clustering phase and prolong network lifetime, an energy consumption prediction model is established for regular data acquisition nodes. Simulation results show that compared with current clustering algorithms, this algorithm can achieve longer sensor network lifetime, higher energy efficiency, and superior network monitoring quality.

1. Introduction

Over recent years, wireless sensor networks (WSNs) [1] have become a hot research and have a wide range of potential applications including environmental monitoring, military detection, industrial control, and home networks [2–5]. But in practical applications, in order to meet the demands of various applications for the technologies of sensor networks, increasing attentions have been attracted to the researches on heterogeneous wireless sensor networks HWSN [6].

HWSN is composed of different types of sensor nodes, which are in a wide range of applications [7–9]. For HWSN, it should be given priority to reduce energy dissipation in network operation, improve network load and stability, and prolong the network lifetime.

Energy consumption in networks can be effectively reduced by organizing clustering sensor nodes. Due to the dynamic and complex nature of energy configuration and network evolution, it is very difficult to design a clustering protocol which can save energy and provide reliable data transmission in heterogeneous networks.

In this paper, a new heterogeneous sensor networks model with heterogeneity of monitored objects and energy heterogeneity of all nodes is proposed. For the heterogeneous networks with such properties, in order to make more rational use of network energy and prolong the lifetime of the networks, this paper presents an energy-efficient prediction clustering algorithm (EEPCA).

EEPCA determines node energy factor by comparing the energy of a node with the average energy of other nodes within the communication range and determines communication cost factor according to the ratio of the average energy consumed in one communication within all nodes and the ideal average energy consumption after the node becomes the cluster head. The probability for nodes to become cluster heads is directly related to energy factor and communication cost factor. In order to save energy consumed by broadcasting energy information in each round of nodes clustering, an

energy predication model is established for nodes whose data collection (such as temperature, humidity) is of regularity in time interval and message length. Considering the changes in networks environment and errors between calculated and actual node energy consumption, set the nodes that do not need to broadcast their energy information if the difference between the node residual energy in the initial stage at the current round and the predicted value at the last round is within a certain range. Simulation results show that EEPCA can achieve longer lifetime, higher energy efficiency, and superior network monitoring quality compared with other clustering protocols such as LEACH, SEP, and EDFCM.

2. Related Work

LEACH [10] is one of the most popular distributed cluster-based routing protocols in wireless sensor networks. In the initialization phase, LEACH carries out cluster head selection. In order to balance the load of all network nodes, LEACH elects cluster head nodes in each round, where is the proportion of optimal cluster heads. If the probability of a node i is less than the following probability threshold, it becomes the cluster head as follows:

$$T(i) = \begin{cases} \dfrac{p_{\text{opt}}}{1 - p_{\text{opt}}\left(r \bmod \left(1/p_{\text{opt}}\right)\right)}, & \text{if } i \in G \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where r is the current number of rounds, G is a set of cluster head nodes which fail to make cluster heads in the latest round ($r \bmod (1/p_{\text{opt}})$).

However, LEACH has some constraints, including: (1) it does not take into account the optimization of the number of cluster heads; (2) LEACH therefore must base itself on two assumptions so as to achieve uniform energy consumption at per node: (1) the initial energy of each node is equal; (2) the energy consumed at each node when acting as the cluster head is equal.

Many researchers have done profound work probing into HWSN.

In [10], authors improved LEACH algorithm and put forward LEACH-C algorithm of electing cluster heads according to the residual energy. However, each node needs to know the total energy of the current network to determine whether it can become the cluster head, which requires support of routing protocols, and therefore distributed implementation is difficult to achieve. SEP [11] is designed for two-level heterogeneous networks. But in the multilevel heterogeneous networks, SEP does not suitable.

For further researches, a heterogeneous network model in term of different initial energies is discussed in [6, 12–14]. In [15], the authors introduced a cluster head election method using fuzzy logic to overcome the defects of LEACH. They investigated that the network lifetime can be prolonged by using fuzzy variables.

In [14], authors proposed EEHC protocol. This protocol selects cluster heads based on the weighted probability of each node related to the initial energy, the more initial energy, and the higher probability; the node will be selected as a cluster head. However, this protocol cannot predict energy consumption, so its performance is limited in heterogeneous networks in which parts of nodes are regular data acquisition nodes.

In [16], authors proposed EDFCM protocol, which applies to networks with three different kinds of heterogeneous nodes. Nodes in the networks model of this protocol fall into two ordinary types: one performing the function of managing information and the other collecting different data (type_0, type_1). type_1 has more complex hardware and software architectures, so it has more initial energy and greater data transfer capability. But the application of this protocol is limited to the networks with only two types of ordinary nodes.

In [17], authors proposed ERP clustering routing protocol. In this paper, an evolutionary algorithm with an appropriate fitness function is proposed with the intrinsic properties of clustering in mind.

3. System Model and Problem Description

3.1. Heterogeneous Model for Wireless Sensor Networks. To meet the demands of efficient monitoring, we describe our HWSN model with both different initial energies and monitored objects. The basic assumptions of networks model are as follows: the networks is located in a $M \times M$ square area (Figure 1), N sensor nodes are randomly distributed within the networks, nodes are stationary, and base station is located in the middle of the area. Sensor nodes monitor a variety of objects. Define some nodes to be regular data acquisition (RDA) nodes: these nodes send back messages of fixed length at a fixed interval; some nodes are not regular in acquiring data; and the messages sent back are not regular.

Therefore, nodes are heterogeneous in two ways: (1) heterogeneous data-acquisition regularity: some nodes are regular in acquiring data, and some are not. All regular nodes send $n_1 \sim n_2$ times messages in a rotation cycle times, and the message sizes are between $[l_1, l_2]$ bits; (2) the initial energy of all nodes are heterogeneous.

Nodes do not have any location information, but they can calculate the distance between nodes according to signal strength received. Nodes in the networks are organized in the form of clusters. Cluster heads perform the function of data fusion and are responsible for the resultant data transmission to the BS. There is only one BS in the networks.

Node initial energy is randomly distributed in $(E_{\min}, E_{\max})$. For any node i, its initial energy is E_i.

3.2. Energy Consumption Model. This paper applies a simple energy consumption model [10] to calculate energy consumption in communication, ignoring energy consumption of nodes in the process of computing, storage, and so forth.

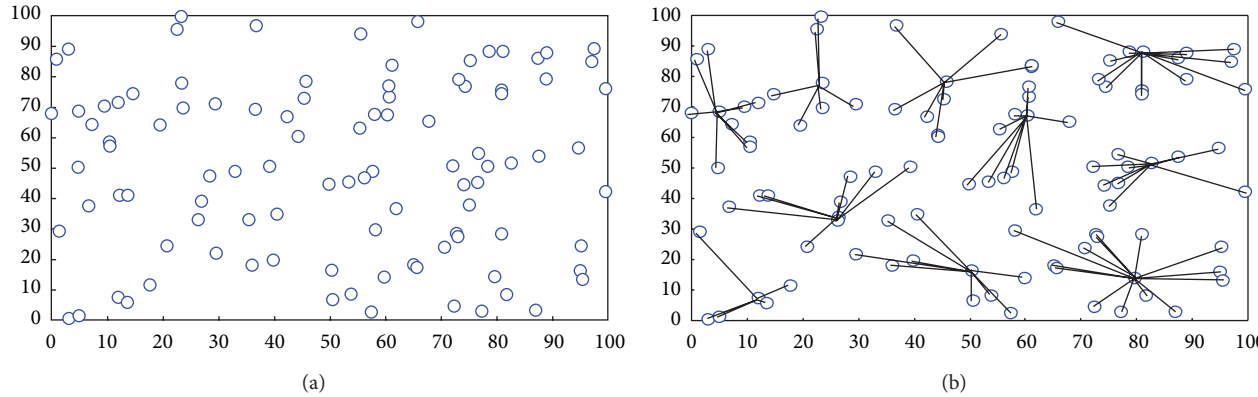


Figure 1: (a) 100-node random heterogeneous network. (b) Dynamic clustering structure by EEPCA.

In the process of transmitting l bits message through distance d, the energy consumption of the transmitter is

$$E_{Tx}(l,d) = E_{Tx\_elec}(k) + ET_{Tx\_amp}(l,d) = \begin{cases} lE_{elec} + l\varepsilon_{fs}d^2, & d < d_0, \\ lE_{elec} + l\varepsilon_{mp}d^4, & d \geq d_0. \end{cases} \quad (2)$$

Receiver's energy consumption is

$$E_{Rx}(l) = E_{Rx\_elec}(l) = lE_{elec}, \quad (3)$$

where E_{elec} is the energy dissipated per bit to run the transmitter or the receiver circuit, and $\varepsilon_{fs}d^2$ and $\varepsilon_{mp}d^4$ are the amplifier energy that depend on the transmitter amplifier model.

3.3. Problem Description. EEPCA must take full account of the following:

(1) algorithm should be fully distributive and self-organized. Each node must decide whether to become a cluster head or a member belonging to a cluster in the clustering phase [10];

(2) nodes with more residual energy must have higher probability to become cluster head, and it must be ensured that the cluster has a smaller communication cost, but energy is not the only factor for cluster head selection;

(3) cluster load balancing must be ensured;

(4) in order to save energy consumption when nodes broadcast in initial clustering phase of each round, an energy prediction model of RDA nodes is established.

4. EEPCA Clustering Algorithm

4.1. Calculation of Distance between Nodes. Nodes can perceive their mutual distance according to attenuation of signal strength in the process of transmission. In clustering phase, all nodes use certain transmission energy for broadcast. For instance, with energy E_i^{tran}, node i broadcasts information to other nodes, including its message sending cycle t_i, message length l_i, and its energy information E_i. Node j detects the received signal strength $E_{j,i}^{rec}$ while receiving messages. The relationship between transmission energy and reception energy is as follows [18]:

$$E_{j,i}^{rec} = \frac{K}{d_{i,j}^{\alpha}} \times E_i^{tran}, \quad (4)$$

where K is a constant, $d_{i,j}^{\alpha}$ is the relative distance between node i and node j. α is distance-energy gradient, and its value varies from 1 to 6 according to the physical environment in which the sensor networks operate. Thus, the distance between i and j is

$$d_{i,j} = \sqrt[\alpha]{\frac{K \times E_i^{tran}}{E_{j,i}^{rec}}}. \quad (5)$$

The nodes establish a routing table of neighboring nodes based on received data and save all relevant information of all nodes within its communication range. All nodes in the networks are marked by the only integer value, which is each node's ID. The information stored in the routing table includes the distance between the node and its neighboring nodes, cluster head node's ID, the distance to the cluster head, the current energy, and predicted energy consumption.

4.2. Cluster Head Selection. The cluster head node has to perform extrafunctions such as data fusion and relaying messages. In order to prevent some nodes from dying too soon due to excessive energy cost, the nodes with more residual energy should be given greater opportunity to become cluster heads, and all nodes take their turns to be cluster head nodes.

Set p_{opt} to be the proportion of optimal cluster heads and p_i to be the probability for node i to be selected as the cluster head. In energy-heterogeneous WSN, p_i calculation is complicated. Currently, many clustering algorithms in HWSN determine p_i by using the ratio of nodes' current residual energy and the average energy of the entire networks, but the latter is very difficult to obtain [13], especially for networks in which different nodes are monitoring different

objects. Consequently, major error is likely to happen to the estimated average energy.

Ideally, nodes are distributed uniformly and send back data at identical frequency and length. Set d_{toBS} to be the average distance between the head node and the BS and d_{toCH} to be the average distance between member nodes in a cluster and the head node, it can be concluded that [10, 19]

$$d_{\text{toBS}} = 0.765\frac{M}{2}, \qquad d_{\text{toCH}} = \frac{M}{\sqrt{2\pi k}}. \tag{6}$$

The number of optimal cluster heads is [13]

$$k_{\text{opt}} = \frac{\sqrt{N}}{\sqrt{2\pi}}\sqrt{\frac{\varepsilon_{\text{fs}}}{\varepsilon_{\text{mp}}}}\frac{M}{d_{\text{toBS}}^2}. \tag{7}$$

Therefore, the proportion of optimal cluster heads is

$$p_{\text{opt}} = \frac{k_{\text{opt}}}{N}. \tag{8}$$

In the initial stage of clustering, for any node i, there are a total of n nodes within its communication range, of which the distance between n_1 nodes and i is $< d_0$, and the distance between n_2 nodes and i is $> d_0$. So considering the ratio of the energy of i and the average energy of all nodes within its communication range, $\varpi(E)_i$, the energy factor influencing the probability of cluster heads can be obtained as follows:

$$\varpi(E)_i = \frac{E_i}{\sum_{j=1}^{n}\left(E_j/n\right)}. \tag{9}$$

Consider the nodes distribution in the networks. If after nodes have been clustered, the average distance between nodes within the cluster and cluster head is far, a high communication cost is inevitable for one communication within the cluster. Set $\overline{E_{i-\text{round}}}$ to be the average energy consumed in one communication between each node in the cluster and node i after i has been selected as the cluster head as follows:

$$\overline{E_{i-\text{round}}} = \frac{\sum_{j=1}^{n_1}\left(l_j E_{\text{elec}} + l_j\varepsilon_{\text{fs}}d_{i,j}^2\right) + \sum_{k=1}^{n_2}\left(l_k E_{\text{elec}} + l_k\varepsilon_{\text{mp}}d_{i,k}^4\right)}{n}. \tag{10}$$

On an ideal occasion that nodes in the networks are uniformly distributed and every data transmission sends data identical in length l, the number of nodes in each cluster is N/k_{opt}. If the distance from m_1 nodes to the cluster head is $< d_0$, m_2 nodes $> d_0$, the ratio of these two types of nodes is

$$\lambda = \frac{\pi d_0^2}{\pi d^2 - \pi d_0^2}. \tag{11}$$

Therefore, the number of these two types of nodes is

$$m_1 = \frac{\lambda \times \left(N/k_{\text{opt}}\right)}{\lambda + 1}, \tag{12}$$

$$m_2 = \frac{\left(N/k_{\text{opt}}\right)}{\lambda + 1}. \tag{13}$$

The random distribution of nodes can be viewed as a Poisson point process [19]. Ideally, there are n points in circle A, and their locations which are uniformly distributed in A are mutually independent random variables. d_i is a random variable, presenting the distance from a point (x_i, y_i) to the circle centre point. The expectation of all the points in the circle to the center point is

$$E\left(d_i\right) = \iint_A \sqrt{x_i^2 + y_i^2}\,\rho\left(x_i, y_i\right)dx\,dy. \tag{14}$$

A circle can be obtain after any radius revolves around the center, so consider the distribution of points on a random radius. Points are distributed uniformly in the circle, and accordingly, the density of points is proportional to radius squared. Therefore, the probability density of points on a random radius is

$$f\left(x\right) = \frac{2x}{R^2}, \tag{15}$$

where R is radius length.

Therefore, the calculation of $E(d_i)$ can be simplified to

$$E\left(d_i\right) = \int_0^R xf\left(x\right)dx. \tag{16}$$

By formula (15) and (16), the average distance expectation of nodes whose distance to the cluster head is less than d_0 is

$$E(d_i)_1 = \int_0^{d_0} x\frac{2x}{d_0^2}dx = \frac{2}{3}d_0. \tag{17}$$

The average distance expectation of nodes whose distance to the cluster head is more than d_0 is

$$E(d_i)_2 = \int_{d_0}^{d} x\frac{2x}{\left(d - d_0\right)^2}dx + d_0 = \frac{2}{3}\left(d - d_0\right) + d_0. \tag{18}$$

Therefore, ideally the average energy consumption within one data transmission in the cluster is

$$\overline{E_{\text{consume}}} = \frac{l\left(m_1\left(E_{\text{elec}} + \varepsilon_{\text{fs}}E(d_i)_1\right) + m_2\left(E_{\text{elec}} + \varepsilon_{\text{mp}}E(d_i)_2\right)\right)}{N/k_{\text{opt}}}. \tag{19}$$

By formulae (10) and (19), communication cost factor $\varpi(C)_i$ which has influence on probability of cluster head election is

$$\varpi(C)_i = \frac{\overline{E_{\text{consume}}}}{\overline{E_{i-\text{round}}}}. \tag{20}$$

Integrating node energy factor and communication cost factor, the probability for node i to become the cluster head node is

$$p_i = p_{\text{opt}} \times \left(\alpha\varpi(E)_i + \beta\varpi(C)_i\right), \tag{21}$$

where α and β are the calculation factors regulating the proportion of energy factor and communication cost factor in calculation p_i, $\alpha + \beta = 1$.

The constraints of LEACH threshold formula $T(i)$ should be improved in two steps: (1) to promote $T(i)$ into multilevel heterogeneous networks; (2) in EEPCA, to take energy factor and the communication cost factors into account and to improve calculation method of $T(i)$, as is shown in:

$$T(i) = \begin{cases} \dfrac{p_i}{1 - p_i\left(r \bmod \left(1/p_i\right)\right)} \\ \times \left[\left(\alpha\varpi(E)_i + \beta\varpi(C)_i\right) + \left(r_s \operatorname{div}\left(1/p_i\right)\right)\right. \\ \left.\quad \times \left(1 - \left(\alpha\varpi(E)_i + \beta\varpi(C)_i\right)\right)\right] & \text{if } i \in G \\ 0, & \text{otherwise,} \end{cases} \tag{22}$$

where r_s is the number of rounds when a node fails to be selected as the cluster head. Once the node elected, r_s is reset to 0.

4.3. Energy Consumption Prediction Mechanism. Obviously, after the networks complete a round, a new node need to be selected as the cluster head. Because it is necessary to re-evaluate the energy factor and the communication cost factor so as to determine the probability for the node to become the cluster head, the current node residual energy must be obtained. The easiest way is that all nodes in the networks carry out a broadcast through the method utilized in the first round of clustering. However, considerable energy will be consumed when broadcasting in each round of clustering, so this paper establishes an energy consumption prediction mechanism for RDA nodes.

In $r-1$ round, it takes n_j times for any node j to send messages with a length l_j to cluster head node i and the distance between i and j is $d_{i,j}$. Since each node keeps relevant information of all nodes within communication range and their mutual distance, any node within node j' communication range can calculate the energy consumption of node j in $r-1$ round as follows:

$$E_{j_{r-1}\text{-comsume}} = \begin{cases} n_j\left(l_j E_{\text{elec}} + l_j\varepsilon_{\text{fs}}d_{i,j}^2\right), & d_{i,j} < d_0, \\ n_j\left(l_j E_{\text{elec}} + l_j\varepsilon_{\text{mp}}d_{i,j}^4\right), & d_{i,j} \geq d_0. \end{cases} \tag{23}$$

The residual energy of node j can be predicted at the beginning of r round when $r-1$ round starts as follows:

$$E_{j_r\text{-prediction}} = E_{j_{r-1}} - E_{j_{r-1}\text{-comsume}}. \tag{24}$$

Due to reasons such as networks environment changes, when r round starts, all nodes need to be reclustered, and new cluster head nodes need to be elected. Node j determines whether its current residual energy is close to the residual energy predicted in the last round or not as follows:

$$\gamma = \left|1 - \frac{E_{j_r\text{-prediction}}}{E_{j_r}}\right|. \tag{25}$$

If γ is less than constant ε, the energy predication error can be tolerated. In the initial phase of r round, node j does not broadcast its energy information and the remaining nodes update node j' energy information in the routing table according to calculation results.

Table 1: Parameters used in simulations.

Parameter	Value
Network grid	(0, 0)~(100, 100)
Node numbers	100
Coverage radius (m)	12
Threshold distanced_0 (m)	75
Initial energy (J)	1–3
E_{elec}	5 nJ/bit
ε_{fs}	10 pJ/bit/m^2
ε_{mp}	0.0013 pJ/bit/m^4
Message size	2000–6000 bits
Broadcast packet size	2500 bits
Round	5 TDMA frames

5. Simulation Experiment

5.1. Establishment of Simulation Environment. To evaluate performance of the algorithm proposed in this paper, we have made a simulation experiment on it with the help of MATLAB simulation software. The experiment simulates a sensor network randomly formed within a 100 m × 100 m area. After the formation, nodes become static. And 100 sensor nodes are randomly distributed in this area. Assuming the BS is located in the center of the area. Parameters used in this experiment can be seen in Table 1. This paper will compare the performance of EEPCA and that of LEACH, SEP, and EDFCM. All results, unless otherwise stated, are average values of 100 times independent experiments.

5.2. Experiment Results and Analysis. In EEPCA, α and β are calculation factors regulating the proportion of energy factor and communication cost factor in calculation p_i, satisfying $\alpha + \beta = 1$. Change the values of α and β, and observe the performance of EEPCA. This experiment sets all nodes that are energy heterogeneous, and the initial energy is 1–3 J. All monitored objects in the network are homogeneous, excluding RDA nodes. All nodes send 4000 bits messages to the cluster head at TDMA timeslot.

Figure 2 shows the death time of the first node, 10% nodes and 50% nodes when the values of α and β vary in the above circumstances. It can be seen when the values of α are in the vicinity of 0.74, death time of the first node and 10% nodes appears the latest; while when the values of α are within the range of 0.66–0.68, death time of 50% nodes appears the

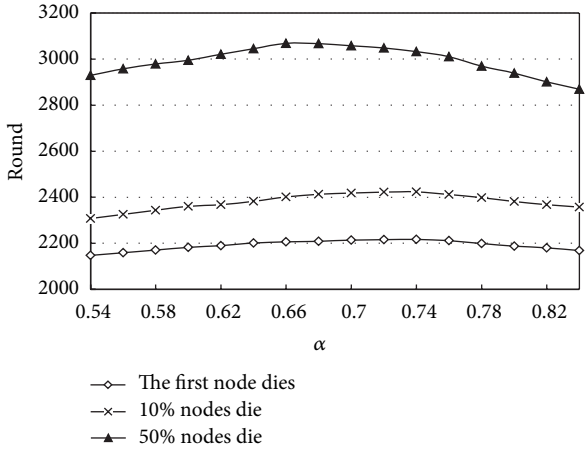


Figure 2: Influence of α and β' values on performance.

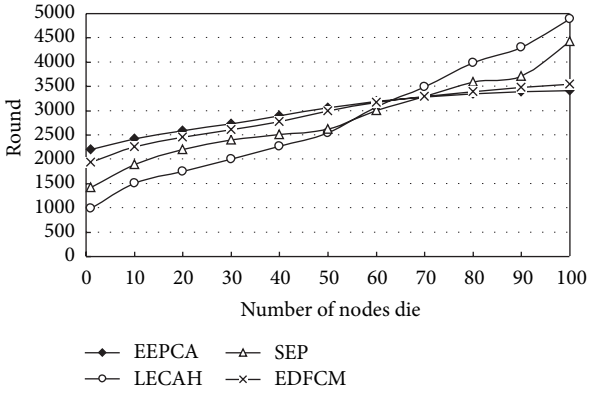


Figure 3: Death time of nodes.

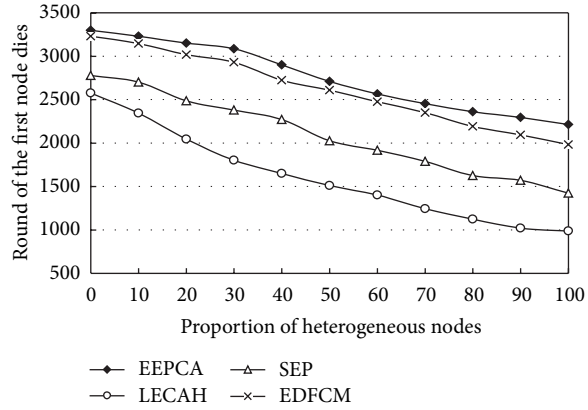


Figure 4: Death time of the first node when the number of heterogeneous nodes changes.

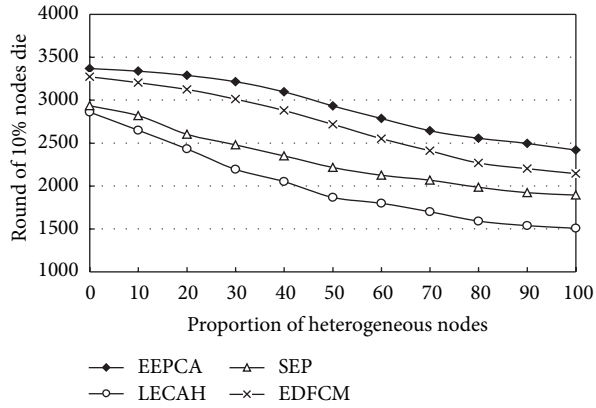


Figure 5: Network stable period when the number of energy-heterogeneous nodes changes.

latest. In subsequent experiments, the values of α and β are unified as 0.7 and 0.3.

In the previous experimental environment, EEPCA and LEACH, SEP, and EDFCM will be compared and tested to analyze EEPCA cluster head selection mechanism's impact on the algorithm performance when all nodes are heterogeneous.

The simulation results in Figure 3 show the variation of the number of dead nodes over time in the previous experimental environment in different algorithms. It can be seen in Figure 3, LEACH cannot make good use of the additional energy of heterogeneous nodes, the stable period is very short, and nodes die at a fixed speed rate. Compared with LEACH, SEP has longer stable periods. EEPCA and EDFCM curves are lines with smaller slope versus X-axis. Because EEPCA distributes energy consumption uniformly on each node in the heterogeneous network, the death time of the first and the last node is relatively closer.

In the previous experimental environment, change the proportion of heterogeneous nodes in the total number of nodes, and observe the performance of each algorithm. Figure 4 presents the number of rounds from the beginning to the death of the first node when the proportion of heterogeneous nodes varies from 0 to 100%. In this experiment, the initial energy of all nonenergy-heterogeneous nodes is 2 J.

Before the death of 10% nodes, the network can send back to the BS data of high quality and reliability [13]. So Figure 5 presents the number of rounds from the beginning to the death of 10% nodes, namely, the stable period.

It can be seen that as LEACH is not a clustering algorithm for heterogeneous networks, with the increase of the proportion of heterogeneous nodes, attainable network stable period quickly reduces. SEP can obtain 25% more stable period than LEACH, which is basically consistent with the experimental results presented by [11]. As EDFCM takes into account heterogeneous energy of different nodes, it gets longer stable period than SEP. EEPCA takes into account the energy consumption of nodes in the communication process in addition to residual energy, so the decline rate of stable period is significantly less than other algorithms in the process of increasing proportion of heterogeneous nodes. Therefore, with greater proportion of heterogeneous nodes, a more stable period is obtained.

To go further, RDA nodes are introduced into the experiment. Set all nodes energy in the networks to be

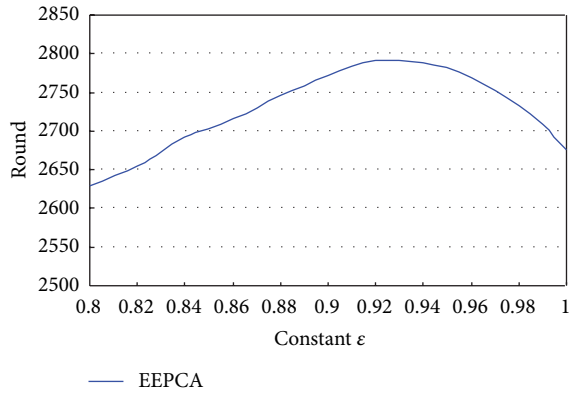


Figure 6: Impact of constant ε on network stable period.

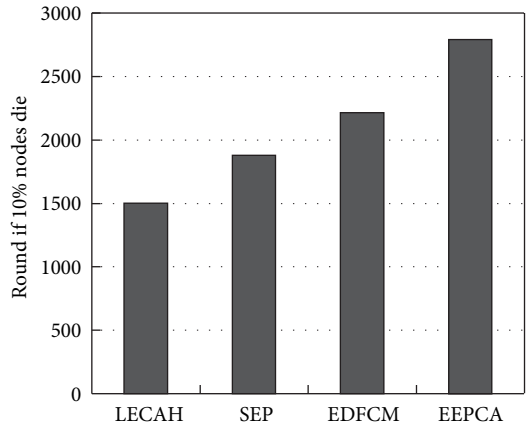


Figure 7: Network stable period.

heterogeneous, 50% nodes to be RDA nodes, and 10% nodes to be malfunctioning. All RDA nodes send messages 3–7 times in a round, and the sizes of messages are valued randomly between 2000 and 6000 bits. Examine the impact of the constant ε on networks stable period.

Figure 6 shows when the value of ε is near 0.92–0.93, the network achieves maximum stable period.

RDA nodes are introduced, and the stable periods of all the algorithms are examined. This experiment sets all nodes that are energy heterogeneous, 50% of which are RDA nodes, constant $\varepsilon = 0.93$, and 10% nodes in the network are malfunctioning. The results are shown in Figure 7.

Obviously, due to the introduction of energy consumption prediction mechanism, broadcast frequency in the clustering phase in each round is effectively reduced. Therefore, in a network heterogeneous in two ways, initial energies and monitored objects, EEPCA makes significant improvement in network stable period compared with the other three algorithms.

Figure 8 shows that all the nodes are energy heterogeneous, 50% are RDA nodes, and 10% of the nodes are malfunctioning. In EEPCA, the number of messages received by BS is on linear rise for a long period of time, while in other algorithms, the growth rate of the number of messages

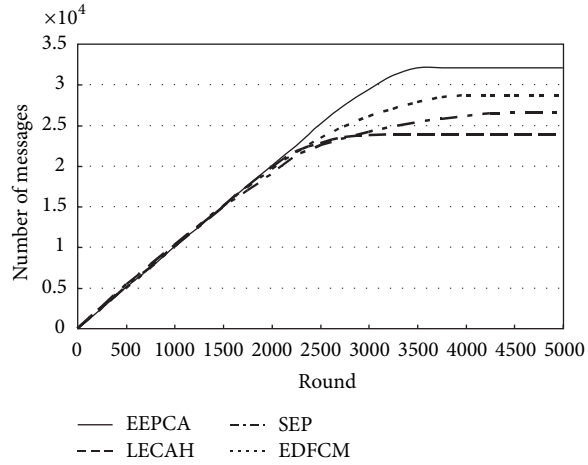


Figure 8: The number of messages received by BS.

received by BS begins to decline earlier. To sum up the total number of messages sent back to BS by all nodes in these four algorithms when the network fails, the amount of data collected by EEPCA is much larger than that by the other three algorithms. Therefore, EEPCA has better network monitoring quality.

6. Conclusions

In this paper, we describe the HWSN model with both different initial energies and monitored objects. We present an effective energy prediction clustering algorithm EEPCA for multilevel heterogeneous sensor networks. In EEPCA, each node independently selects itself as the cluster head node based on energy factor and communication cost factor, which leads to the probability of cluster head election related to nodes' current residual energy and average communication cost after being selected. At the same time, with the consideration that the WSNs are frequently used to monitor objects such as temperature and humidity which need to report data regularly, and the length of reported data is usually fixed, an energy consumption prediction mechanism is established for RDA nodes. Simulation results show that compared with LEACH, SEP, and EDFCM, and EEPCA can achieve longer lifetime, higher energy efficiency, and better network monitoring quality. Its performance is superior to other protocols.

Acknowledgments

This work was supported by the opening Project of State key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications) (SKLNST-2010-1-03), the Technology Fund of support in Sichuan Province (2009GZ0153, 2011GZ0188).

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirici, "Wireless sensor network: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[2] M. Haenggi, *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*, CRC Press, 2005.

[3] C. Y. Chong and S. P. Kumar, "Sensor networks: evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, 2003.

[4] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Proceedings of the IEEE Interntional Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, pp. 2033–2036, May 2001.

[5] C. Y. Chang and H. R. Chang, "Energy-aware node placement, topology control and MAC scheduling for wireless sensor networks," *Computer Networks*, vol. 52, no. 11, pp. 2189–2204, 2008.

[6] E. J. Duarte-Melo and M. Liu, "Analysis of energy consumption and lifetime of heterogeneous wireless sensor networks," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '02)*, pp. 21–25, IEEE Press, Taipei, Taiwan, November 2002.

[7] E. P. de Freitas, T. Heimfarth, C. E. Pereira, A. M. Ferreira, F. R. Wagner, and T. Larsson, "Evaluation of coordination strategies for heterogeneous sensor networks aiming at surveillance applications," in *Proceedings of the IEEE Sensors Conference (SENSORS '09)*, pp. 591–596, Christchurch, New Zealand, October 2009.

[8] J. M. Corchado, J. Bajo, D. I. Tapia, and A. Abraham, "Using heterogeneous wireless sensor networks in a telemonitoring system for healthcare," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 234–240, 2010.

[9] I. Dietrich and F. Dressler, "On the lifetime of wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 5, no. 1, pp. 5:1–5:39, 2009.

[10] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

[11] G. Smaragdakis, I. Matta, and A. Bestavros, "SEP: a stable election protocol for clustered heterogeneous wireless sensor networks," in *Proceedings of the International Workshop on SANPA*, pp. 251–261, 2004.

[12] V. P. Mhatre, C. Rosenberg, D. Kofman, R. Mazumdar, and N. Shroff, "A minimum cost heterogeneous sensor network with a lifetime constraint," *IEEE Transactions on Mobile Computing*, vol. 4, no. 1, pp. 4–14, 2005.

[13] L. Qing, Q. X. Zhu, and M. W. Wang, "Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks," *Computer Communications*, vol. 29, no. 12, pp. 2230–2237, 2006.

[14] K. Dilip, C. A. Trilok, and R. B. Patel, "EEHC: energy efficient heterogeneous clustered scheme for wireless sensor networks," *Computer Communications*, vol. 32, no. 4, pp. 662–667, 2009.

[15] J. M. Kim, S. H. Park, Y. J. Han, and T. M. Chung, "CHEF: cluster head election mechanism using fuzzy logic in wireless sensor networks," in *Proceedings of the 10th International Conference on Advanced Communication Technology (ICACT '08)*, pp. 654–659, 2008.

[16] H. Zhou, Y. Wu, Y. Hu, and G. Xie, "A novel stable selection and reliable transmission protocol for clustered heterogeneous wireless sensor networks," *Computer Communications*, vol. 33, no. 15, pp. 1843–1849, 2010.

[17] B. A. Attea and E. A. Khalil, "A new evolutionary based routing protocol for clustered heterogeneous wireless sensor networks," *Applied Soft Computing Journal*, vol. 12, no. 7, pp. 1950–1957, 2012.

[18] S. Doshi, S. Bhandare, and T. Brownl, "An on-demand minimum energy routing protocol for a wireless ad hoc network," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 6, no. 3, pp. 50–66, 2002.

[19] V. Mhatre and C. Rosenberg, "Design guidelines for wireless sensor networks: communication, clustering and aggregation," *Ad Hoc Networks*, vol. 2, no. 1, pp. 45–63, 2004.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 837381, 10 pages
http://dx.doi.org/10.1155/2013/837381

Research Article

Driving Path Predication Based Routing Protocol in Vehicular Ad hoc Networks

Yong Feng,[1] Feng Wang,[1] Jingjing Liao,[2] and Qian Qian[1]

[1] *Yunnan Key Laboratory of Computer Technology Applications, Kunming University of Science and Technology, Kunming, Yunan 650500, China*
[2] *Faculty of Civil Engineering and Architecture, Kunming University of Science and Technology, Kunming, Yunan 650500, China*

Correspondence should be addressed to Yong Feng; fybraver@163.com

Received 9 November 2012; Accepted 16 January 2013

Academic Editor: Chao Song

The vehicular mobility is reflection and extension of the human social activity. Since human trajectories show a high degree of temporal and spatial regularity, thus vehicular driving paths are predictable to a large extent. In this paper, we firstly analyze the predictabilities of different types of vehicles and then propose a new driving path predication based routing protocol (DPPR). With hello messages to broadcast vehicles' driving path predication information to neighbor vehicles, DPPR can observably increase the successful ratio to find the proper next hop vehicles that move toward the optimal expected road in intersection areas. In roads with sparse vehicle density, DPPR utilizes vehicles to carry messages to roads with high vehicle density while the messages' forward paths partially coincide with the vehicles' driving paths. Moreover, as to messages that can tolerate long delay, they can be carried to destinations by vehicles whose driving paths will pass the messages' destination in order to optimize bandwidth utilization. Simulation results demonstrate the effectiveness of the proposed DPPR protocol.

1. Introduction

In recent years, vehicular networks have gradually become an important research field in wireless communication and computer network and received broad attention from both industry and academy [1, 2]. As vehicle nodes' high speed mobility and uneven distribution, vehicular networks have dynamic and changing network topology, which makes it difficult to maintain persistent connection among vehicle nodes. To improve data delivery performance, vehicular networks widely adopt ad hoc networking and delay tolerant networking technology. Thus vehicular ad hoc networks (VANETs), which evolved from mobile ad hoc networks (MANETs) and delay tolerant networks (DTNs) and are formed by cars and any supporting fixed nodes, emerges and are regarded as an important component of intelligent transportation systems (ITS) [3].

VANETs promise a wide range of valuable applications including real-time traffic estimation for trip planning, mobile access to Internet, and in-time dissemination of emergency information such as accidents and pavement collapses. To realize the applications above, one of the key research topics is to design effective and efficient data delivery schemes. Therefore, many schemes have been presented to solve the problem in recent years. Among the existing schemes, some works mainly take advantage of geographic position information, such as GPSR [4] and CAR [5]. The performances of these protocols mainly depend on the network connectivity, and they are sensitive to the vehicle node density. Many works are based on the traffic statistics and network layout, such as SADV [6] and VADD [7]. A few protocols such as TBD [8], TSF [9], and STDFS [10] are designed to utilize available vehicle trajectories to improve the data delivery performance. To disseminate vehicle trajectory information, these protocols assume that numerous wireless access points (APs) need to be deployed along roads. That will undoubtedly request a large amount of investment. Some researchers have made explorations on the prediction of vehicle driving paths and take use of the character of anticipating vehicle routes to develop data delivery schemes. In the literature [11–14], several prediction models and schemes are proposed. However, these works are based on either historic driving records or local and current vehicle running status, and thus there are prediction accurate problems.

Many research works [15–17] show that human trajectories show a high degree of temporal and spatial regularity: each individual can be characterized by a time-independent characteristic length scale and a significant probability to return to a few highly frequented locations. Sincerely the vehicular mobility is reflection and extension of the human social activity; thus vehicular driving paths are predictable to a large extent. In general, vehicles can be categorized to three types: (i) bus, tramway, and light rail, which have stable trajectories and schedules; (ii) private car, which has remarkable regular trajectories and obviously temporal and spatial regularity; for example, a private car generally travels among limited several places such as home, workplace, supermarket, and park; (iii) taxi, which has flexible and variable running paths. In this paper, we analyze the driving path predictability for different sorts of vehicles and discuss the corresponding prediction methods. Doing this makes ordinary vehicles foresee their driving paths. Based on the character, we propose a new driving path prediction routing protocol (DPPR), which can increase data delivery ratio through (i) carrying messages by vehicles whose driving paths will pass through the messages' destination; (ii) increasing chance to find the vehicles that move toward the optimal expected road in intersection areas. The major contributions of this work may be listed as follows.

(a) We analyze the driving path predictability of various types of vehicles with a microscopic perspective and divide vehicles into three categories in which the prediction method of each category vehicle is discussed. Moreover, we also discuss the feasibility and practicability of utilizing the vehicle driving path prediction method to improve the routing performance in VANETs.

(b) We propose a new routing protocol called DPPR to improve the data forwarding performance of VANETs, which effectively takes advantage of the characteristic of foregone driving path for average vehicles. Through extensive simulations, the effectiveness of our proposed DPPR scheme is evaluated.

The rest of the paper is organized as follows: Section 2 summarizes the related works of routing protocols in vehicular ad hoc networks. In Section 3, we analyze the driving path predictability of various types of vehicles and discuss the feasibility and practicability of DPPR scheme. In Section 4, we describe the design of our proposed DPPR routing protocol in detail. Section 5 shows the effectiveness of DPPR via simulation experiments. Section 6 concludes the paper.

2. Related Works

In recent years, data delivery and forwarding issues about vehicle-to-vehicle and vehicle-to-infrastructure in VANET have gained lots of attention [1–6]. For the frequent network partition and merging due to the high mobility of vehicles, the physically constrained nodal mobility resulted from the fixed roadways, and for the constrained nodal moving speed limited by the roadway conditions, the data forwarding in VANET is different from that in the traditional mobile ad hoc networks (MANETs). These unique characteristics of the road networks make the MANET routing protocols ineffective in the VANET settings [18]. Thus, many routing protocols based on carry-and-forward thinking have been proposed in order to reach efficient and effective data forward performance in VANETs.

Among these works, Epidemic [19] is an early approach to deal with the data forward issue in frequent network partition and merging settings. It allows the random pair wise exchange of data packets among mobile nodes in order to maximize the possibility that data packets can be delivered to their destination node. Thus, a great number of copies of data packets are generated during the delivery, which weakens its performance to a great extent, especially when the resources of bandwidth and buffer are limited.

Some works mainly take advantage of geographic position information, such as GPSR [4], CAR [5], MMR [18], and VVR [20]. Similar to GPSR, both MMR and VVR use greedy forwarding strategy to find the next packet carrier based on the geographical proximity toward the packet destination. Through using the approach of "guard node," CAR forwards data packets through the connected path from the packet source to the packet destination. The performances of these protocols mainly depend on the network connectivity, and they are sensitive to the vehicle node density. Thus the geographic position based routing schemes cannot work well when the vehicular traffic is sparse and of noneuniform-distribution.

Some works are based on the traffic statistics and network layout, such as SADV [6], VADD [7], and DBR [3]. SADV routing leverages on the stationary nodes to improve the network connectivity and the data forward performance. Through using a stochastic model based on vehicular traffic statistics, VADD tries to achieve as high delivery successful ratio as possible with low delivery delay from mobile vehicles to stationary packet destinations. DBR scheme focuses on satisfying the user-defined delay bound rather than the lowest delivery delay, so that it can economize resource such as channel utilization and buffer space.

A few protocols such as TBD [8], TSF [9], and STDFS [10] are designed to utilize available vehicle trajectories to improve the data delivery performance. TBD utilizes the vehicle trajectory information along with vehicular traffic statistics in order to compute the accurate expected delivery delay for better forwarding decision making. However, to disseminate vehicle trajectory information, these protocols assume that numerous wireless access points (APs) need to be deployed along roads. That will undoubtedly request a large amount of investment.

Some researchers have made explorations on the prediction of vehicle driving paths and take use of the character of anticipating vehicle routes to develop data delivery schemes. Several prediction models and schemes are proposed such as PBR [11], MOPR [12], and PLR [13]. PBR exploits the location and velocity information of vehicles to predict route lifetime and takes preemptive action to minimize route failure. MOPR improves the routing process by selecting the most stable route in terms of lifetime. PLR uses a location predictor to

solve the problem of location inaccuracy and vehicle mobility. In [14], Jeung et al. propose a network mobility model to predict the driving paths of vehicles. However, these works are based on either historic driving records or local and current vehicle running status, and thus there are prediction accurate problems. Moreover, the approaches mainly focus on short-term and short-distance prediction and do not take into account how long-distance driving path prediction information is used to improve the data forwarding in VANETs.

3. Vehicle Driving Path Prediction

In this section, we firstly review the mobility trajectory features of various types of vehicles and analyze the predictability of their driving paths with a microscopic perspective. According to the differences of predictability, vehicles are classified as three categories in which the prediction method of each category vehicle is discussed. Moreover, we also discuss the feasibility and practicability of utilizing the vehicle driving path prediction method to improve the routing performance in VANETs.

3.1. Driving Trajectory Predictability for Vehicles. In the real world, there exist various kinds of vehicles. As to the stability of driving route, the vehicles can be divided into three categories: (i) the first sort of vehicles that have certain driving trajectories, such as bus, tramway, and light rail, which have stable trajectories and schedules; (ii) the second sort of vehicles that have remarkable regular trajectories, such as private car whose trajectory has obviously temporal and spatial regularity; for example, a private car generally travels among limited several places such as home, workplace, supermarket, and park; (iii) the third sort of vehicles that have changing trajectories, such as taxi, which has flexible and variable running paths. Based on the above classification, we will respectively discuss the driving path predictability for each sort of vehicles as follows.

(a) It is clear that the driving paths of the first sort of vehicles are fully predictable; as a result of this feature they have stable trajectories and schedules.

(b) The driving paths of the second sort are predictable to a large degree because the vehicles have relative regular trajectories and their driving destination places are generally limited and relatively fixed. Some works have focused on the research topic of vehicle driving path prediction, and a few prediction models and methods have been proposed in recent years. For example, in [21], Liu et al. develop a heuristic and context-dependent induction method based on decision trees, to predict vehicle moving trajectories. PLR [13] uses a location predictor to solve the problem of location inaccuracy and vehicle mobility. In [14], Jeung et al. propose a network mobility model to predict driving paths of vehicles.

(c) As to the third sort of vehicles, it seems difficult to predict their driving paths because the vehicles have flexible and variable running paths. Taking a deeper look at the trajectory feature of the third sort of vehicles, however, we find that the prediction difficulty results from the fact that the vehicles' destination locations are uncertain, that is, their destinations may be any place. But if their destination information can be collected in the initial stage of every trip, then the sort vehicles' trajectories will be predictable in a large part. In our previous work [22], we discuss the issue of destination information gathering and propose a driving path prediction method for taxis based on destination information gathering. For the third sort of vehicles, they will always choose the paths with the shortest driving time to reach their destinations as soon as possible. Based on the gathered destination information, GPS device, electronic map, and traffic statistic information at different times, a practicable prediction method is to utilize the Dijkstra algorithm to look for the lowest cost path from the current position to the destination, where the cost means the average travel time for each road segment. To further improve prediction accuracy, we will take advantage of the feature that drivers are relatively fixed and familiar with the road and traffic conditions. Through recording the drives' historical route and personal preferences at different times on the on-board unit of each vehicle itself, the prediction path can be adjusted and more consistent with the real situation. In general, it is reasonable of presume that the driving paths third sort of the vehicles' are predictable while corresponding prediction methods are assisted by destination information.

3.2. Feasibility. From resource requirement aspect, necessary hardware equipments and software systems may include on-board unit, GPS device, electronic map, traffic statistic, input device such as phonetic recognition device, handwriting board, keyboard, and touch panel. Because the above equipments and systems are already popular and becoming common in vehicles, there is hardly any hardware, software, and cost problem to predict vehicles' driving paths.

From prediction scheme aspect, for the first sort of vehicles that have stable trajectories and schedules, undoubtedly, driving path prediction is fully feasible. As to the second sort, as they have relative regular trajectories and stable destinations, a few meritorious prediction methods and models have already been proposed and discussed, and thus it is feasible to predict their driving trajectories with a high reliability. To predict the trajectories of the third vehicles, the first step is to collect vehicle destination information. As previously discussed, there is no technology and cost problem because whether phonetic recognition or handwriting board, keyboard, and touch panel they are all very mature and reliable technology. The key problem lies to create an incentive measure that can prompt drivers to carry out the information gathering activities. A proper solution can be that senders or receivers should pay a fee to drivers for messages' success delivery.

3.3. Availability. From the analyses above, we can get such a useful feature from most vehicles, that is, each vehicle can know the driving path beforehand or in the beginning of its current trip. Through broadcasting driving path information to neighbor vehicles in hello messages, each vehicle can know the trajectories of the vehicles that it will meet on road, which will bring such benefits for improving routing performance in VANETs as follows.

(a) To increase data delivery success ratio through carrying messages to their destinations by vehicles whose driving paths will pass the messages' destination. The carrying message method is very effective in regions where vehicle density is sparse and especially useful for applications that can tolerate long delivery delay due to high delivery successful ratio and negligible low transmission cost.

(b) To improve multiple-hop forwarding success ratio. For time-sensitive applications, multiple-hop forwarding is essential, but the difficulty lies in how to find vehicles that move toward the optimal expected road in intersection areas. Fortunately, driving path information beforehand will effectively solve the problem and increase the chance to find proper next hop.

(c) To reduce wireless channel resources occupancy. For it is envisioned that a wide variety of applications can be running on VANETs in the near future, ranging from road safety, cooperative driver, to entertainment and Internet access, the coexistence of a large spectrum of vehicular applications means that they will inevitably compete with each other for the use of finite wireless network resources [23]. That will easily lead to severe congestion at these areas with high vehicle density and likely result in the appearance of "hotspot" in the rush hour. Therefore, the carrying method can lessen wireless collision possibility and improve the quality of channel, which in turn can improve the delivery success rate.

(d) Effective vehicle trajectory forecast can provide new opportunity to forward data from static locations to mobile vehicles, even between two or more mobile vehicles. In VANETs, existing routing protocols usually only consider how mobile vehicles send data message to location-fixed destinations, but seldom involve how to send messages to specific mobile vehicles, much less two-way data transmission between two moving vehicles. Obtaining driving path information, however, will make it possible or easier to realize message forwarding from fixed positions to mobile vehicles and two-way data transmission between moving vehicles.

4. Driving Path Predication Based Routing Framework

Through utilizing the proposed driving path prediction method in Section 3, each vehicle can know its driving path in advance. Therefore, a new valuable character is introduced into VANETs, that is, the driving trajectories of vehicle nodes are of foreknowledge. Based on this characteristic, we propose a driving path predication based routing protocol called DPPR to improve the data delivery performance in VANETs. In the rest of this section, our proposed DPPR will be described in detail. In Figure 1, the sketch map of DPPR thinking is given out.

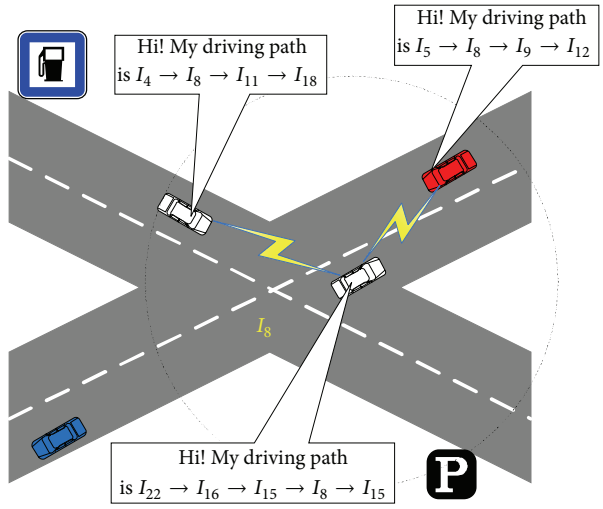


Figure 1: A sketch map of DPPR thinking.

4.1. Assumption. Every vehicle can obtain its current location through GPS device and can be equipped with a preloaded street-level digital map, which not only describes road topology and traffic light period but also provides traffic statistics such as traffic density and average vehicle speed on roads at different times of the day. Such kind of digital map has already been commercialized [24], and more detailed traffic statistics will be integrated into digital map in the near future. Vehicles communicate with each other through short range wireless channel and can find their neighbors through beacon messages. Each beacon provides vehicle's information such as its unique ID, location, velocity, and direction. What more, and each vehicle will carry out the prediction method proposed in Section 3 and announce its driving paths in beacon message.

More formally, the street-level digital map is abstracted as a directed graph $G(V, E)$. For any two intersections I_i and I_j, $r_{ij} \in E$ if and only if there is a road segment connecting I_i and I_j, and vehicles can travel from I_i towards I_j on that segment. And the notation used in the paper is listed in Table 1.

4.2. Path Selection Algorithm. In VANETs, a vehicle needs to send messages in three cases as follows: (a) the vehicle itself produces messages; (b) it forwards the message which it received from other vehicles; (c) it periodically takes from its routing buffer to send. When vehicle v_i has a message M to send, it needs to leverage the knowledge of global traffic statistics such as average vehicle density and speed on the

Table 1: Notation.

Notation	Definition
I_i	Intersection marked i, $I_i \in V$
r_{ij}	Road segment from I_i to I_j, $r_{ij} \in E$
l_{ij}	Euclidean distance of r_{ij}
ρ_{ij}	Vehicle density on r_{ij}
v_{ij}	Average vehicle velocity on r_{ij}
d_{ij}	Expected packet forwarding delay from I_i to I_j
R	Wireless communication range of vehicles
α, β	Correction factor

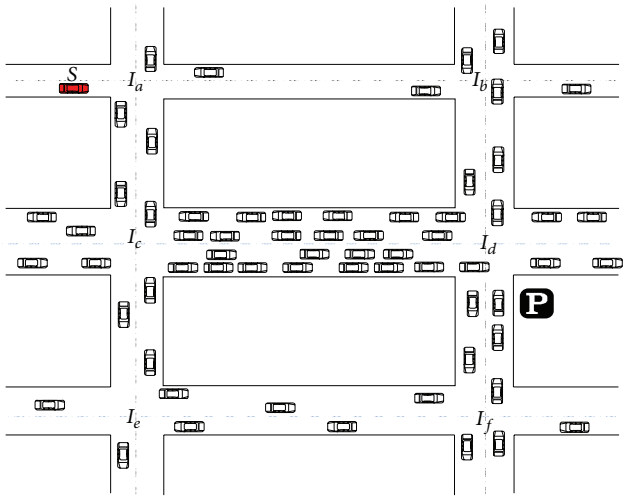


Figure 2: Find a path to the park.

road segments of the directed graph G. Based on the information, DPPR works out the path with optimized performance for M. As shown in Figure 2, vehicle v_i near intersection I_a wants to send a message to the park near intersection I_d for reserving a parking space. The first important issue is to select an optimal forwarding path to deliver the message M among the three paths: $I_a \rightarrow I_b \rightarrow I_d$, $I_a \rightarrow I_c \rightarrow I_d$, and $I_a \rightarrow I_c \rightarrow I_e \rightarrow I_f$.

For each edge in G, we will evaluate it with transmission delay metric, that is, the time cost used to transmit a message through the edge. To estimate the transmission delay of a road segment (e.g., r_{ij}), we utilize the delay model proposed in [7], which is shown in (1) as follows:

$$d_{ij} = \begin{cases} \alpha \cdot l_{ij}, & \text{if } \dfrac{1}{\rho_{ij}} \leq R, \\ \dfrac{l_{ij}}{v_{ij}} - \beta \cdot \rho_{ij}, & \text{if } \dfrac{1}{\rho_{ij}} > R. \end{cases} \quad (1)$$

Since such information as road segment distance, vehicle density, and average velocity is available, we can get the transmission delay of each road segment through (1), so that each edge has a weight, that is, its transmission delay, in the directed graph G. The proposed DPPR protocol assumes that the best path to the destination is the shortest delay one, that is, the path that minimizes the sum of transmission delay of the edges on the directed graph G that abstracts the street map. By running a single invocation of Dijkstra on G, the shortest delay path can be worked out.

4.2.1. Straightway Mode. As previously mentioned, each vehicle maintains a neighbor list by periodically broadcasting beacons. According to the content of the beacon messages, each vehicle can get the driving path information of its neighbors. In DPPR, as to a vehicle, its driving path is defined as the intersection sequence along which it will arrive at the destination of its current trip. According to changing geographical position, DPPR protocol will switch between Straightway and Intersection modes. In this section, we present the protocol used in the Straightway mode.

Since the traffic is at most bidirectional, data forwarding in the Straightway mode is simple. As vehicle v_i has a message M to send, it firstly operates the path selection algorithm to determine the nearest target intersection then chooses the optimal next hop vehicle from its neighbor list, and the chosen one is indicated as *Nexthop*. In the Straightway mode, we can simply apply the geographically greedy forwarding to choose *Nexthop*. At the same time, the vehicle also looks for the vehicles whose driving paths will pass the destination of the message. If there is more than a matching vehicle, then the one with the fastest average running speed is selected out, and its vehicle identification is denoted as *Dest_accord*. As shown in Algorithm 1, after completing the above query steps, vehicle v_i can make a forwarding decision as follows.

Case 1. Driving path of v_i itself will pass *M's* destination: if *Dest_accord* is found and has faster average running speed than v_i, then M is sent to *Dest_accord*. Otherwise, v_i checks whether there exists *Nexthop*; if so, it generates a copy of M, denoted as M', increments flag_copy of M', and sends it to *Nexthop*.

Case 2. Driving path of v_i itself will not pass *M's* destination: if *Dest_accord* is found and the distance between v_i and message *M's* destination is within the range of two intersections, then M is sent to *Dest_accord*. Otherwise, v_i checks whether there exists *Nexthop*; if so, then M is sent to *Nexthop*.

When no proper *Nexthop* or *Dest_accord* is found, vehicle v_i only puts M into its routing queue. While receiving a message from another vehicle, the current firstly checks whether it has to keep a copy of the message in its routing queue. If so, it directly drops the message, else or puts the message into its queue.

4.2.2. Intersection Mode. When a vehicle enters an intersection, it switches to Intersection mode. For each packet in its buffers, the vehicle checks whether there is a contact available to forward the packet to the next road segment along the selected optimal path. If so, the vehicle transmits the packet to the neighboring vehicle, otherwise it continues to carry it. For example, when vehicle v_i has a message M to send, it firstly operates the path selection algorithm to sort all the outgoing directions. Secondly, v_i looks for the neighbor vehicles whose running directions are in accordance with the optimal direction based on their driving path information.

```
Input: Nexthop, Dest_accord, neighbor list of v_i, that is, Σ
Output: Straightway forwarding decision
BEGIN
(1)   if(v_i itself will pass M's destination)
(2)     if(Dest_accord exists and speed(Dest_accord) > speed(v_i))
(3)       send(M, Dest_accord); //send message M to Dest_accord
(4)     else
(5)       if(Nexthop exists)
(1)         M' = copy(M); //generate a replica of message M
(2)         flag_M' ++; send(M', Nexthop);
(6)       endif
(7)       enqueue(M);
(8)     endif
(9)   else
(10)    if(Dest_accord exists and dist(v_i, M' destination) > 2)
(11)      send(M, Dest_accord); //send message M to Dest_accord
(12)    else
(13)      if(Nexthop exists)
(3)         send(M, Nexthop);
(14)      else
(15)        enqueue(M);
(16)      endif
(17)    endif
(18)  endif
END
```

Algorithm 1: Pseudocode of straightway forwarding algorithm.

```
Input: Nexthop, Direc_accord, optimal direction, denoted as direc
Output: Intersection forwarding decision
BEGIN
(1)   if(v_i itself runs toward direc)
(2)     if(Direc_accord exists and closer direc than v_i)
(3)       send(M, Dest_accord); //send message M to Direc_accord
(4)     else
(5)       enqueue(M);
(6)     endif
(7)   else
(8)     if(Direc_accord exists)
(9)       send(M, Direc_accord);
(10)    else
(11)      if(Nexthop exists)
(4)         send(M, Nexthop);
(12)      else
(13)        enqueue(M);
(14)      endif
(15)    endif
(16)  endif
END
```

Algorithm 2: Pseudocode of intersection forwarding algorithm.

If there is more than a matching vehicle, then the one which is geographically closest towards the optimal direction is selected out, and its vehicle identification is denoted as *Direction_accord*. But if such *Direction_accord* cannot be found, v_i will apply the geographically greedy forwarding to choose *Nexthop*, that is, the vehicle which is the nearest to the optimal direction among all its neighbors. As shown in Algorithm 2, after completing the above query steps, vehicle v_i can make a forwarding decision as follows.

Case 1. Running direction of v_i itself is the same as the optimal: if *Direction_accord* is found and is geographically

closer towards the optimal direction than v_i, then M is sent to *Direction_accord*. Otherwise, v_i puts M into its routing queue and goes on carrying it.

Case 2. Running direction of v_i itself is not the same as the optimal: if *Direction_accord* is found, then M is sent to *Direction_accord*. Otherwise, v_i checks whether there exists *Nexthop*; if so then M is sent to *Nexthop*. When neither *Direction_accord* or *Nexthop* is found, this means that v_i has no proper forwarding contact at present; it only puts M into its routing queue.

4.3. Queue Management Algorithm. For each mobile vehicle node in VANETs, the size of its routing buffer queue is limited. Therefore, the queue management algorithm would greatly influence the data delivery performance. In the proposed DPPR, the flag_copy of a message indicates how many copies have been propagated, and the survival time shows how long a certain message has existed in the network. Therefore the flag_copy and survival time together denote the importance of a message, and the queue management is just based on the two parameters.

Messages are sorted in the routing queue based on an increasing order of their flag_copy. For those messages with the same flag_copy value, they are further sorted according to an increasing order of survival time. Thus messages with smaller tickets and shorter survival time are closer to the top of the queue, and can be transmitted with higher priorities. Moreover, messages will be dropped in the following two occasions: (a) when a message arrives and the queue is full, it is compared with that message at the end of the queue and the one with bigger flag_copy is dropped among them. If the flag_copy values of the two messages are equal, then the one with a longer survival time is dropped; (b) whenever a message's survival time is longer than the delay tolerance of the network, it is dropped to avoid unnecessary resource occupation.

5. Performance Evaluations

In this section, we will evaluate the impact of the proposed DPPR scheme on the data transmission performance in VANETs. We choose the classic wireless ad hoc network routing algorithm GPSR [4] and Epidemic as the referential. Since pure GPSR has no carry-and-forward ability, which will result in poor performance in intermittently connected VANETs, we extend it by adding buffers and thus make it have basic carry-and-forward ability. In the following simulation experiments, the data delivery performances of GPSR only with simple carry-and-forward ability, denoted as GPSR (with buffer), Epidemic, and the proposed DPPR will be analyzed and evaluated.

5.1. Simulation Setting. The experiment is based on a 4000 m × 3000 m approximate rectangle street area, which is derived and normalized from a real street map of Beijing city in China. The rectangle contains 25 intersections and 40 bidirectional roads. We utilize NS-2.34 as the simulation tools. Since modeling of complex vehicle movement is important for accurately evaluating protocols, the open source software VanetMobiSim-1.1 [25] is used to generate the movement of vehicles. Detailed simulation parameters are shown in Table 2.

Table 2: Simulation parameters.

Parameter	Value
Simulation area	4000 m × 3000 m
Number of intersections	25
Number of mobile vehicles	100, 300
Proportion of taxis	0–30%
Number of packet senders	20
Communication range	250 m
Vehicle velocity	40–80 km/h
CBR rate	0.1–1 packet per second
Data packet size	512 B
Buffer size	200 packets
Vehicle beacon interval	1 second

To evaluate the performance of the protocols in different traffic density environments, we deploy two kinds of different number of vehicles, that is, 100 and 300, respectively, into the network to imitate different vehicle node density, which means different network connectivity. Vehicles are with the average speed ranges from 40 to 80 kilometers per hour. Among all vehicles, 20 of them are randomly picked out to send CBR data packet to fixed spots. To evaluate the performance on different load status, we change the CBR rate from 0.1 to 1 packet per second.

5.2. The Packet Delivery Ratio. In this section, we compare the performance of DPPR protocols with GPSR (with buffer) and Epidemic routing in the aspect of packet delivery ratio. Here, the packet delivery ratio is defined as the fraction of packets that have reached their destination without exceeding the delay tolerant threshold of certain application. For multiple-copy routing algorithm such as Epidemic and DPPR, a message is regarded as one successfully delivered only if any one copy of all its copies is received by the destination. We vary the data generation rate, that is, from 0.01 to 0.1 message/s, in order to evaluate the data delivery performance of the three protocols under different transmission loads.

Figures 3 and 4 show the data delivery ratio as a function of the CBR sending rate under two different vehicle densities; that is, the total number of vehicles is 100 and 300, respectively. From the two figures, we can find that the data delivery ratios of all three protocols go down with the increasing data sending rate, but the decreasing tendencies are much different. Epidemic routing has best performance and almost reaches the upper bound of the data delivery ratio when the data sending rate is very low; however its performance quickly deteriorates with the increase of sending rate. This is due to MAC layer collision and rapid exhaustion of the limited resources (i.e., network bandwidth, routing buffer, etc.) resulted from forwarding a tremendous amount of copies in epidemic routing. Though GPSR is extended and

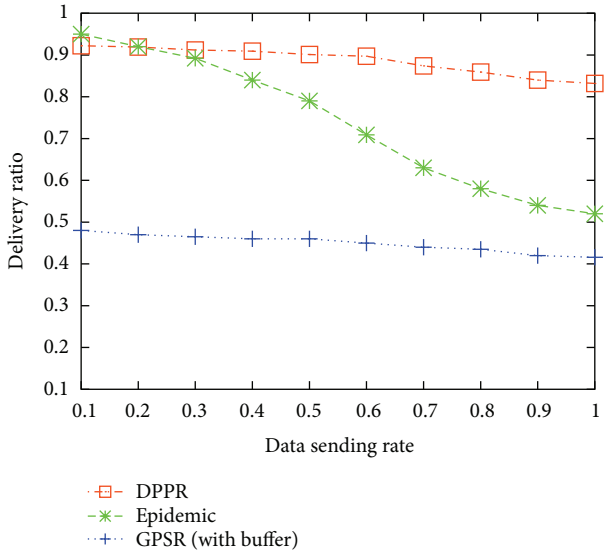


Figure 3: Packet deliver ratio under different CBR sending rate (100 vehicles).

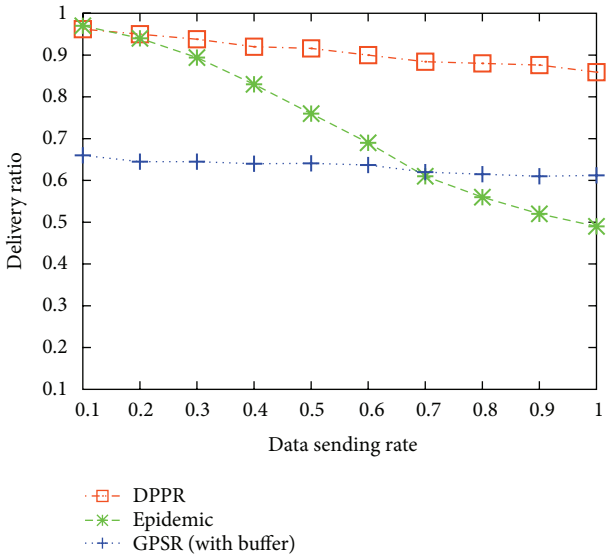


Figure 4: Packet deliver ratio under different CBR sending rate (300 vehicles).

thus has simple carry-and-forward ability, its performance is still the worst in the case of sparsely connected networks, as shown in Figure 3. However, its delivery performance is very steady and almost insusceptible to the change of data sending rate. From the figure, we also can find that our proposed DPPR algorithm outperforms the others as the data sending rate increases, especially when the data sending rate is normal or high.

Figure 4 shows the data delivery ratio of the three protocols under high vehicle density environment that is, when the total vehicle number is 300. We can find that the change tendencies of data delivery ratio among these protocols remain unchanged as the data sending rate increases.

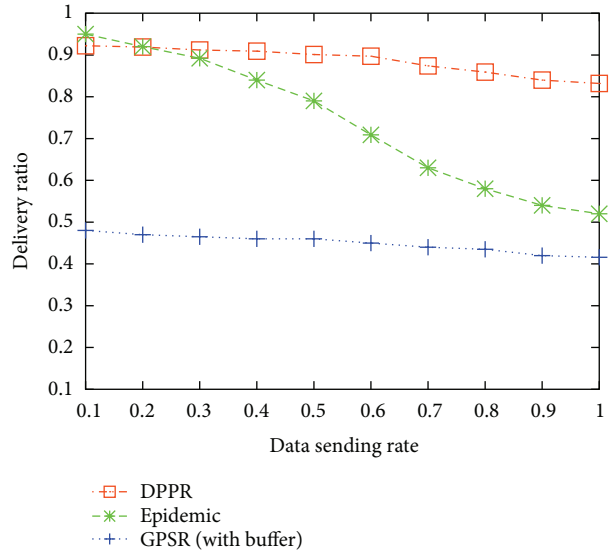


Figure 5: Delivery delay under different CBR sending rate (100 vehicles).

As the increment of vehicle density improves the network connectivity, the performance of DPPR and GPSR routing is obviously enhanced. particularly for GPSR, it achieves relatively good delivery ratio when vehicle number is 300 because the network connectivity is improved with the increasing vehicle density. On the other hand, the high vehicle density is not good news for Epidemic protocol. The reason is that the number of message copies increases dramatically in epidemic routing as the node density increases, which results in an increasing number of collisions and dropped data packets.

5.3. The Packet Delivery Delay. In this section, we compare the packet delivery delay among DPPR, Epidemic routing, and GPSR protocols when transmitting data from moving vehicles to fixed destination spots. Figures 5 and 6 show the packet delivery delay as a function of the CBR sending rate under two vehicle densities; that is, the total number of vehicles is 100 and 300, respectively. From Figure 5, we can see Epidemic routing has the least delivery delay when the data sending is very low, but its performance drastically deteriorates; that is, its delivery delay rapidly goes up, as the data sending rate increases. GPSR has relatively low data delivery delay at low node density, but it is not meaningful simply because of its low delivery ratio. We also see the delivery delay of DPPR algorithm is generally lower than that of GPSR and Epidemic, and it has much lower delay than Epidemic especially while data sending ratio is high.

Figure 6 shows data delivery delay of the three protocols when the total vehicle number is 300. We can find that the delay of epidemic routing dramatically increases with the increasing data sending rate because it generates a great number of redundant packets. As the traffic load increases, many packets may be dropped, which is similar to the case in Figure 5. The delivery delay of DPPR and GPSR shows slowly decreasing tendency. As the increment of vehicle density

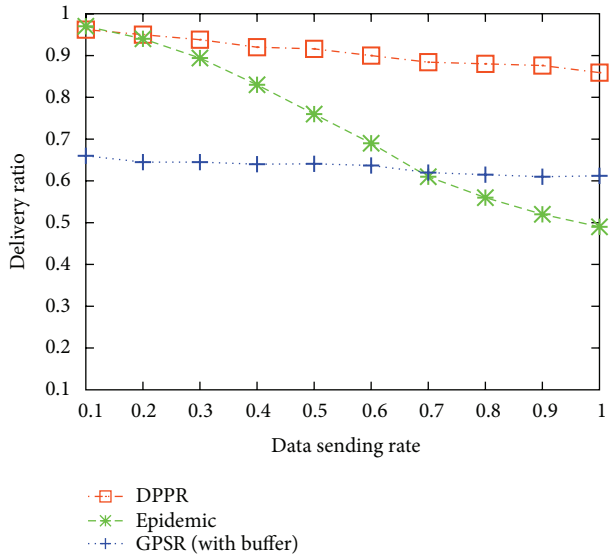


Figure 6: Delivery delay under different CBR sending rate (300 vehicles).

improves the network connectivity, we also see that the delay of DPPR and GPSR is clearly lower than that in Figure 5, and the proposed DPPR routing shows optimal performance in general.

6. Conclusion

In near future, traffic safety and many other commercial applications will be running on VANETs. For realizing these promising applications, it is an important research topic to develop effective routing protocols that can reach high data delivery performance. In this paper, we analyze the driving path predictability for different sorts of vehicles and discuss the corresponding prediction methods. Doing this makes ordinary vehicles foresee their driving paths. Through utilizing the features, we propose a new routing protocol called DPPR to improve the performance of data delivery in VANETs. Through simulation experiments, we analyze and evaluate the performance of GPSR (with buffer), Epidemic, and DPPR protocol. The experiment result shows that our proposed DPPR protocol can reach high data delivery performance in VANETs.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant no. 61262081; the Yunnan Provincial Applied Fundamental Research Project under Grant no. KKSY201203027.

References

[1] J. Ott and D. Kutscher, "Drive-thru internet: IEEE 802.lib for "automobile" users," in *Proceedings of the IEEE INFOCOM Conference on Computer Communications*, pp. 362–373, March 2004.

[2] D. Reichardt, M. Miglietta, L. Moretti, P. Morsink, and W. Schulz, "Cartalk 2000—safe and comfortable driving based upon inter-vehicle communication," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '02)*, pp. 545–550, Jun 2002, http://www.cartalk2000.net/.

[3] Research and I. T. A. (RITA), "IntelliDrive: Safer, Smarter and Greener," http://connection.ebscohost.com/c/articles/52707131/intellidrive-safer-smarter-greener.

[4] B. Karp and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom '00)*, pp. 243–254, ACM, August 2000.

[5] V. Naumov and T. R. Gross, "Connectivity-aware routing (CAR) in vehicular ad hoc networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 1919–1927, May 2007.

[6] Y. Ding, C. Wang, and L. Xiao, "A static-node assisted adaptive routing protocol in vehicular networks," in *Proceedings of the 4th ACM International Workshop on Vehicular Ad Hoc Networks (VANET '07)*, pp. 59–68, September 2007.

[7] J. Zhao and G. Cao, "VADD: vehicle-assisted data delivery in vehicular Ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1910–1922, 2008.

[8] J. Jeong, S. Guo, Y. Gu, T. He, and D. Du, "TBD: trajectory-based data forwarding for light-traffic vehicular networks," in *Proceedings of the 29th International Conference on Distributed Computing Systems (ICDCS '09)*, Montreal, Quebec, Canada, June 2009.

[9] J. Jeong, S. Guo, Y. Gu, T. He, and D. H. C. Du, "TSF: trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks," in *Proceedings of the 30th IEEE International Conference on Distributed Computing Systems (ICDCS '10)*, pp. 557–566, June 2010.

[10] F. Xu, S. Guo, J. Jeong et al., "Utilizing shared vehicle trajectories for data forwarding in vehicular networks," in *Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM '11)*, pp. 441–445, April 2011.

[11] V. Namboodiri and L. Gaw, "Prediction based routing for Vehicular ad hoc networks," in *Proceedings of the Vehicular Ad Hoc Networks (VANET '04)*, 2004.

[12] H. Menouarand, M. Lenardi, and F. Filali, "Movement Prediction-based Routing (MOPR) concept for position-based routing in vehicular networks," in *Proceedings of the IEEE 66th Vehicular Technology Conference (VTC '07)*, pp. 2101–2105, October 2007.

[13] Z. Li, Y. Zhu, and M. Li, "Practical location-based routing in vehicular ad hoc networks," in *Proceedings of the IEEE 6th International Conference on Mobile Adhoc and Sensor Systems (MASS '09)*, October 2009.

[14] H. Jeung, M. L. Yiu, X. Zhou, and C. S. Jensen, "Path prediction and predictive range querying in road network databases," *VLDB Journal*, vol. 19, no. 4, pp. 585–602, 2010.

[15] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[16] S. Schönfelder and U. Samaga, "Where do you want to go today?-More observations on daily mobility," in *Proceedings of the 3rd Swiss Transport Research Conference*, Ascona, Switzerland, 2003.

[17] K. Kelley, "Casual carpooling-enhanced," *Journal of Public Transportation*, vol. 10, no. 4, p. 119, 2007.

[18] Y. Lee, H. Lee, N. Choi, Y. Choi, and T. Kwon, "Macro-level and Micro-level Routing (MMR) for urban vehicular ad hoc networks," in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 715–719, November 2007.

[19] A. Vahdat and D. Becker, "Epidemic routing for partially-connected Ad Hoc networks," Technical Report, 2000, http://issg.cs.duke.edu/epidemic/epidemic.pdf.

[20] H. Lee, Y. Lee, T. Kwon, and Y. Choi, "Virtual Vertex Routing (VVR) for course-based vehicular ad hoc networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '07)*, pp. 4408–4413, March 2007.

[21] N. Liu, M. Liu, J. Cao, G. Chen, and W. Lou, "When transportation meets communication: V2P over VANETs," in *Proceedings of the IEEE 30th International Conference on Distributed Computing Systems (ICDCS '10)*, 2010.

[22] Y. Feng, Q. Qian, F. Wang, and X. Fu, "Public transportation assisted data delivery scheme in vehicular delay tolerant networks," in *Proceedings of the 8th Mobile Ad hoc and Sensor Networks*, Chengdu, China, 2012.

[23] A. Skordylis and N. Trigoni, "Delay-bounded routing in vehicular ad-hoc networks," in *Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '08)*, pp. 341–350, May 2008.

[24] "Digital Map Data," 2012, http://www.mapmechanics.com/digital-map-data/.

[25] J. Härri, F. Filali, C. Bonnet, and M. Fiore, "VanetMobiSim: generating realistic mobility patterns for VANETs," in *Proceedings of the 3rd ACM International Workshop on Vehicular Ad Hoc Networks (VANET '6)*, pp. 96–97, ACM, September 2006.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 809742, 7 pages
http://dx.doi.org/10.1155/2013/809742

Research Article

A Study on the Tracking Problem in Vehicular Ad Hoc Networks

Xing Zhang, Bang Liu, and Jinchuan Tang

Department of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Correspondence should be addressed to Bang Liu; bangliuox1989@gmail.com

Received 3 December 2012; Accepted 12 January 2013

Academic Editor: Ming Liu

Vehicles have the characteristic of high mobility which makes vehicular ad hoc networks (VANETs) different from other mobile ad hoc networks (MANETs), it is more difficult to establish an end-to-end route in VANETs, and the source and destination nodes keep moving fast from their original locations. To guarantee a data packet will finally be received by the destination node in VANETs, and the location of the destination node must be tracked constantly. Advanced Greedy Forwarding (AGF (Naumov et al. (2006))) is a good way to estimate the location of the destination node by taking advantage of the velocity vectors; however, a vehicle changes its velocity constantly, it would be very inaccurate to estimate the location of a vehicle via its initial velocity. In this paper, we study the tracking problem in VANETs and propose two solutions: area-based tracking (ABT) and parked vehicle-assisted tracking (PVAT). ABT works well when the delays in data transmission from source to destination are small, and PVAT works as a supplement of ABT and deals with the situation of which the delays are large. PVAT takes advantage of vehicular sensors to detect whether a vehicle is parking. Simulations show better results compared to AGF.

1. Introduction

Routing is a great challenge in VANETs because of the high mobility of vehicles. Unlike routing in some static networks or other MANETs that the relay nodes in the routing path are static or moving slowly, the relay nodes in VANETs change their locations constantly and rapidly, and it is hard to find a relatively stable routing path in VANETs. However, the worst thing of routing in VANETs is not the location change of relay nodes, but the location change of the destination node. It is acceptable to find another relay node as long as it can deliver the data packet to the destination node finally, but it is not possible to change the destination node. Therefore, how to track the location of moving destination node and make sure the data packet can be finally delivered to it is an important topic in routing research of VANETs.

Traditional routing algorithm in ad hoc network is based on node topology, each node will maintain a routing table on a global view [1], and any changes on the neighbor node will cause updates on all routing tables. Because the routing table is global, the tracking of the destination node is not necessary, and data packets from the source node will arrive at the destination node in a very low latency. However, the overhead of maintaining the routing table is very big; when the network scale increases, the overhead will become unacceptable. Although some improvements have been made, like some reactive routing protocols [2, 3], the overhead is still proportional to the network scale.

Geographic routing is a routing that each node knows its own and neighbor node geographic position by position determining services like GPS [4]. It does not maintain any routing table or exchange any link state information with neighbor nodes [4]. Taking Greedy Perimeter Stateless Routing (GPSR), for example, in GPSR [5], the source node knows the location of the destination node, and the source node checks its neighbor table whether the destination node is in it; if the destination node is one of the source node's neighbor, the source node will deliver the data packets to the destination node, otherwise the source node will choose a neighbor which is closest to the destination node geographically. Same process happens on the relay nodes repeatedly until the destination node received the data packets, or delivery time expired. GPSR only needs partial information from neighbor nodes, therefore it has a very low overhead. However, GPSR has the tracking problem. The source node obtains the location of the destination node in the start of data

transmission; when the data packets arrive at the location, the destination node has moved away already from the location in a big possibility. To address this problem, AGF is proposed in [6].

AGF adds a velocity vector to the context of the HELLO beacon packet which is used to inform the existence of a node to its neighbors. The velocity vector contains the moving direction and the speed of a node. The source node knows the velocity vector and the location of the destination node in advance. When a relay node receives a data packet, it will add the elapse time of the data packet sending from the source node to here to the processing time of itself; it will multiply the time result by the speed in the velocity vector of the destination node to get an estimation of the location of the destination node with the help of the moving direction also. It will check whether the estimated location is in the neighbor range, and the destination node is in the neighbor table; if condition satisfied, it will deliver the data packet to the destination, otherwise it will choose a neighbor which is closest to the estimation location as a next hop. AGF considers the movement of each node, therefore it obtains much higher data delivery rate compared to GPSR. However, velocity varies constantly, using a static value of velocity to estimate the location of the destination node is very inaccurate. As Figure 1 shows, AGF performs well in an origin case that no speed and direction varied. When speed begins to jitter, the performance starts to degrade. Because AGF has a searching process when the destination node is supposed to be in a neighbor table but it is not, AGF can handle some speed-variation cases when the whole packet delivery time is short, and the destination node is not far from its estimated location. However, when there is a delay in the whole packet delivery process and the destination node drives far away from its estimated location, the packet delivery will fail. Direction change is much worse than speed change to the performance of AGF, because when a vehicle makes turns, it is hard to track it according to its original direction. AGF uses broadcast to search for the destination node; however, wireless signal will be blocked by the obstacles by the road side as Figure 2 shows, like trees and buildings; if a vehicle wants to send a packet to a road segment on the orthogonal direction, unless it is in the center of the crossroad, the packet will not be able to penetrate the obstacles.

To remedy the disadvantages of AGF, we propose two tracking methods: ABT and PVAT. The basic idea of ABT is that instead of maintaining velocity vectors, ABT uses statistical results. The source node will choose an optimum route to the destination node based on traffic density according to VADD [7], see in Figure 3, the time of which the data packet travels on the selected route will be estimated. Therefore, we can calculate a distance of which how far the destination node moves away during the packet delivery by multiplying the time with the average speed of vehicles on the selected route. Instead of estimating the location of the destination node, ABT estimates an area of which the destination node may appear by considering all directions the destination node may head to. As Figure 4 shows, the area is rectangular and centers at the original location of the destination node. When a data packet arrives at the rectangular area, it will

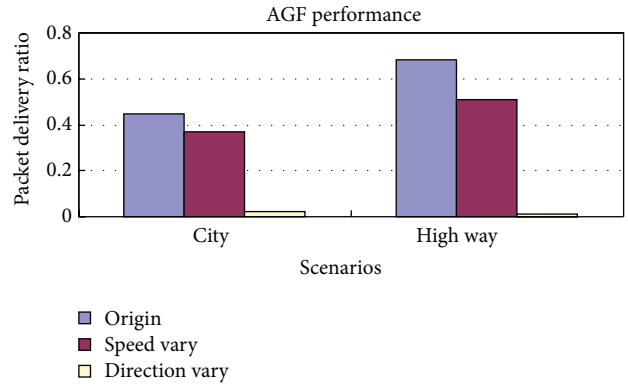


Figure 1: We use a Gaussian distributed random variable v_x with v_i mean and standard deviation σ. In city scenario, v_i equals to 40 kilometers per hour (km/h), and σ is 30 km/h; in high way scenario, v_i equals to 80 km/h, and σ is 10 km/h. The density is settled to a medium one. Detailed simulation setup is in Section 3.

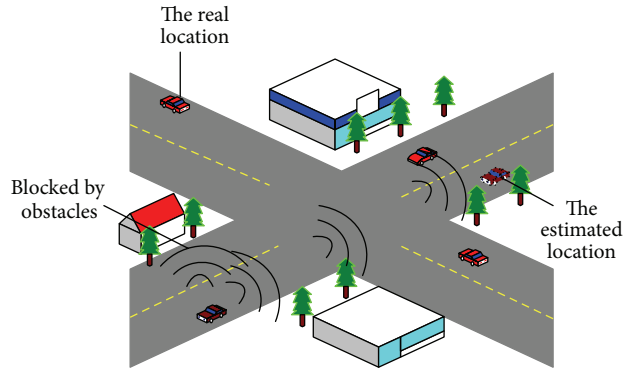


Figure 2: A vehicle is supposed to appear at the estimated location, and it turns left at the crossroad in fact. Another vehicle initiates a broadcasting to search for the vehicle, but the wireless signal is blocked by the obstacles by the roadside.

propagate inside the area until it meets the destination node, or delivery time expired. ABT does not need the information of moving speed and direction, therefore the performance will not suffer from speed and direction variation. However, ABT is at the expense of increasing packet redundancy, if the distance traveled by the destination node is long, the area will be large and the redundancy will be large, and therefore we propose another tracking method PVAT to deal with this problem. Vehicles' movement is restricted by road topology, and direction change of vehicle mostly happens at the intersection except for u turn. If every time the destination node passes by an intersection, the intersection will record the moving direction, and the destination node can easily be tracked. However, it is not easy to keep the information to a certain place in VANETs, because vehicles are highly mobile. [8] uses moving vehicles to cache the information at a certain location; once a vehicle with the information is moving out of the location area, it will pass the information to another vehicle which is in the location area. Vehicle density is changing constantly; when a vehicle leaves its location, it

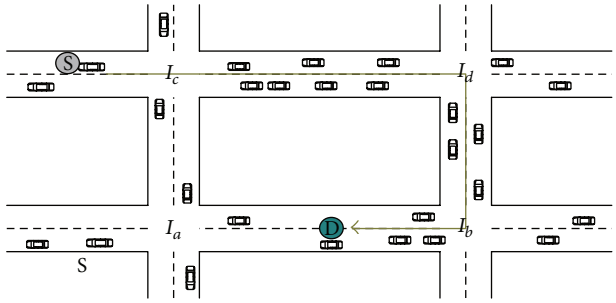


Figure 3: Traffic density-based route selecting. The shortest path from S to D is $I_c \longrightarrow I_a \longrightarrow I_b$, but the density on the path is low, so $I_c \longrightarrow I_d \longrightarrow I_b \longrightarrow I_a$ is chose.

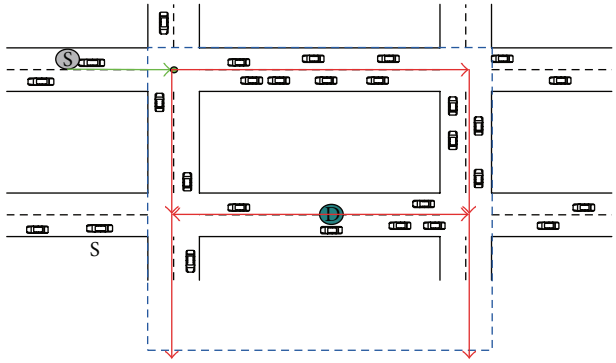


Figure 4: Area-base routing. The data packet will be delivered along the route selected as Figure 3 (green line); once it enters the area (blue rectangle), the data packet will spread over the whole area (red lines).

is not certain that it can find a substitution; also it is a big communication cost that the information keeps transmitting and receiving. Parked vehicles are wide spread in the city, and they can provide a good cache [9]. Parked vehicles are static and wide spread, therefore they can guarantee a more stable and costless cache than [8].

The rest of the paper is organized as follows: detailed descriptions of ABT and PVAT are in Section 2; performance evaluation is in Section 3; related works and conclusion is in Section 4.

2. Area-Based Tracking and Parked Vehicle-Assisted Tracking

ABT and PVAT aim to solve the tracking problem in VANETs, therefore they are more like patches to the existing routing algorithm. Before discussing about the detail of ABT and PVAT, we will describe the routing algorithm we use.

2.1. Traffic Density-Based Routing. Traditional geography-based routing like GPSR finds a neighbor node which is closest to the destination node, as a next hop. In theory, if the traffic density is satisfied, a data packet will finally reach the destination in GPSR. However, it is not optimum; because vehicles travel along road, a neighbor node which is closest to the destination in Euclidean distance may not be the closest node in the road topology. VADD is a routing algorithm based on road topology, and it achieves better performance than traditional geographic routing, therefore we use VADD as the main reference in routing.

However, VADD is a dynamic routing which decides the next hop of data delivery according to real situation, ABT needs to estimate the time for data delivery before real data sending, and therefore before real data routing using VADD, we select an optimum route based on statistical traffic density and calculate the elapse time according to (1).

Suppose Φ is the set of all connected path from the source node to the destination node; $W(x)$ gives the sum of each subpath's weight of path x; $N(x)$ is the number of subpath of x; $t(x)$ is the time elapse of data delivery on path x; l_i is the Euclidean distance of subpath i; ρ_i is the traffic density of subpath i; v_i is the velocity of vehicles on subpath i; R is the communication range of wireless signal, then

$$x = \arg\max W(i), \quad i \in \Phi,$$

$$t_j(x) = \begin{cases} \alpha \cdot l_j & \text{if } \dfrac{1}{\rho_j} \le R, \\ \dfrac{l_j}{v_j} - \beta \cdot \rho_j & \text{if } \dfrac{1}{\rho_j} > R, \end{cases} \tag{1}$$

$$t(x) = \sum_{k=1}^{N(x)} t_k(x).$$

The equation chooses an optimum path x which has the maximum sum of traffic density, because the higher the traffic density, the better for data relay. If the average distance between vehicles is smaller than R, wireless transmission is used to forward the packet; otherwise, vehicles are used to carry the data; even in this case, it is still possible to occasionally have wireless transmissions; hence β is used as a correction factor [7]. The whole time elapse of data delivery is the sum of the time of each sub-path.

2.2. Area-Based Tracking. The basic idea of ABT is that when a data packet enters into the area of ABT, the data packet will be delivered to all roads in the area; as long as the destination node is in the area, it will finally receive the data packet. ABT needs to settle two challenges: how to set the area and how to propagate a data packet in the area.

2.2.1. Area Setting. In Section 2.1, an estimated time elapse of data delivery is given as $t(x)$, we can get an average vehicle speed in urban traffic $\overline{v}$, and therefore an estimated distance that the destination node may move away can be calculated as $d = t(x) \cdot \overline{v}$. The area can simply be a rectangle that it centers at the original location of the destination node and uses $2\lambda d$ as the length and width of the rectangle, and λ is an elastic coefficient which can be adjusted according to real situation.

However, the original rectangle may have to adjust to fit the road topology. As Figure 5 shows, sometimes the area

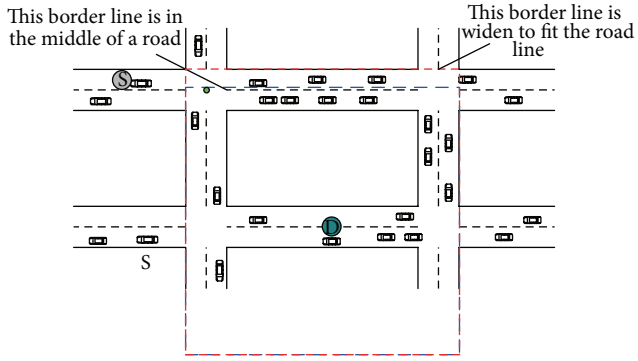


Figure 5: Border line is in the middle of a road.

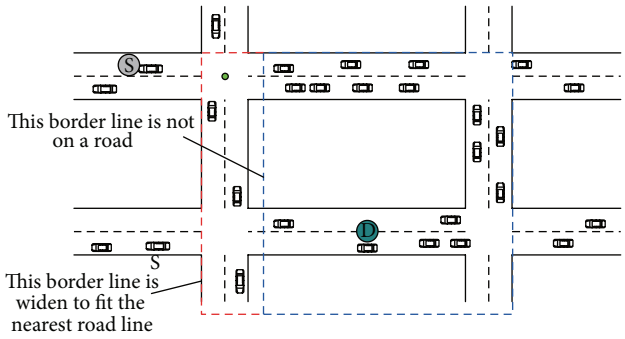


Figure 6: Border line is not on a road.

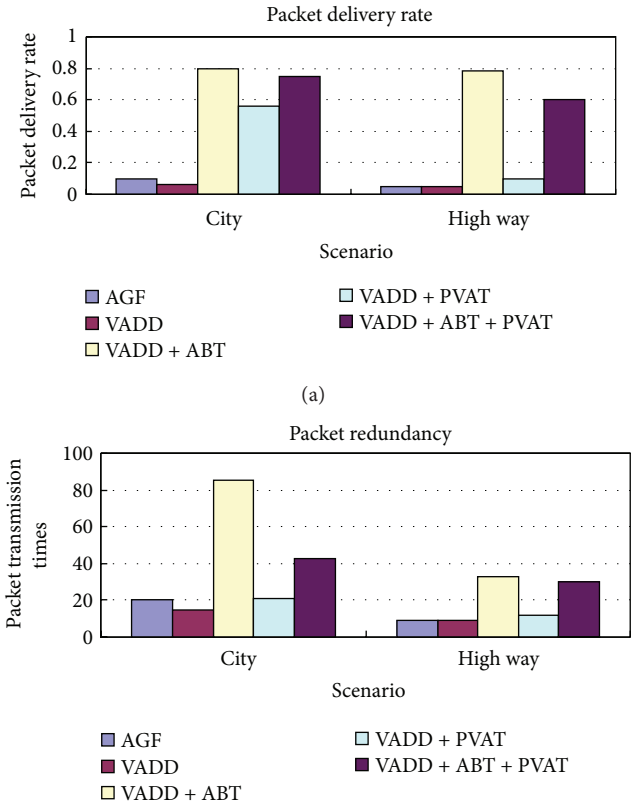


Figure 7: Simulation results.

border line is in the middle of a road. In this case, the border line will be widen to fit the road line. As Figure 6 shows, sometimes the area border line is not on any road. In this case, the border line will be widened or narrowed down to its nearest road line.

2.2.2. Data Packet Propagation. Once a data packet enters into the area, ABT will spread the packet all over the area to guarantee that the destination node will receive the packet. Epidemic routing [10] is suitable to achieve the goal. To reduce unnecessary packet redundancy, ABT makes sure that if a node already has the packet, it will not receive the packet again; if a node is not in the area, it will not receive the packet; if a node already delivered the packet to the destination node, it will not send the packet to another node. DBT algorithm is given in Algorithms 1 and 2.

If a node needs to deliver a packet, it will traverse its neighbor table; if a neighbor is not in the area scale, it will not deliver the packet; if a neighbor is in the area scale, it will send a request to the neighbor; the neighbor will check whether the packet has already been received, if true, reply ALLOW, else reply DISALLOW; if the neighbor allow to send packet, the sender will send the packet; if the neighbor is the destination node, the sending process will be end. After looptime ends, if the packet still has not been delivered to the destination node, and the sending process will be end.

2.3. Parked Vehicle-Assisted Tracking. Parked vehicles by the roadside are very common in city environment, [9] mentioned to use these parked vehicles as data delivery relay nodes. Parked vehicles are static nodes which can provide stable communication and data cache. If a data packet needs to be cached to a location, using moving vehicles will cause the data packet been delivered over and over again when the data-carry vehicle moves out of the location area. Parked vehicles do not leave an area as frequent as the moving vehicles do, therefore using parked vehicles will greatly reduce the overhead of caching data. Moreover, parked vehicles are wide spread in the city, especially in the night that traffic density is very low, the density of parked vehicles are very high, and they can provide good data delivery which cannot be achieved by using moving vehicles.

To detect whether a vehicle is parking or not, PVAT takes advantage of vehicular sensors. OBD scanner can read revolutions per minute (RPM) of vehicle engine which can be used to indicate the parking state of a vehicle.

PVAT records the direction information of the destination node at the intersections when the destination node passes by, it is similar to [8], and the difference is that PVAT uses parked vehicles which can provide low overhead data cache and high reliability of vehicle tracking, because moving vehicles suffer from nonuniform traffic density and traffic holes [11] which will cause data caching fail. The PVAT algorithms are given in Algorithms 3 and 4.

In PVAT, when a node receives a packet, it will firstly check whether the destination node is in its communication

```
Require:
  data packet, area scale, destination node ID, period, looptime.
Ensure:
  true, false.
  repeat
    for each node in the neighbor table
      if node is not in area scale
        continue;
      end if
      send request for data packet delivery;
      if wait for response time out
        continue;
      end if
      if response == ALLOW
        send the data packet;
        if node ID == destination node ID
          return true;
        end if
      end if
    end for
    wait for period time;
  until looptime-- == 0;
  return false;
```

Algorithm 1: Sender of area-based tracking.

```
Require:
Ensure:
  while 1
    if received a request
      if the data packet according to the request
      has already been received
        send DISALLOW response;
        continue;
      end if
      send ALLOW response;
    wait to receive the data packet
    end if
  end while
```

Algorithm 2: Receiver of area-based tracking.

range and send data packet to it, otherwise it will check if it is in an intersection, because that is where the destination node leaves its marks. If marks are found, the data packet will be delivered to the nodes holding the marks and continue data delivery using VADD to the other end of the road segment until no marks are found, or the data packet is delivered to the destination node.

2.4. VADD Plus Tracking Solutions. In this Subsection, we give a whole routing solution combined VADD, ABT, and PVAT, see Algorithm 5. Each packet has a *flag* bit which indicates whether PVAT should be activated. ABT has more packet redundancy than PVAT, which increases communication overhead. It seems that PVAT is better than ABT, but neither parked vehicle nor moving vehicle may fail to cache the marks of the destination node because there are no such vehicles there for caching sometimes. Although ABT uses more packet redundancies, it has much better data delivery rate than PVAT, and therefore we use ABT as main tracking method and PVAT as a supplement way.

3. Performance Evaluation

In this paper, we use ns-2 simulator for simulations. We test packet delivery rate and packet transmission times of AGF, VADD, VADD plus ABT, VADD plus PVAT, and VADD plus ABT and PVAT. We write the AGF and VADD codes according to [6, 7].

```
Require:
Ensure:
  while 1
    mapping its location from GPS to the road segment in the map;
    compare the road segment with the previous road segment;
    if it is in a different road segment
      send destination ID to parked vehicles;
      if no response from parked vehicles
        send destination ID to moving neighbors;
        if no response from moving neighbors
          return;
        end if
      end if
    end if
    wait for a period;
  end while
```

Algorithm 3: Marks leaving of the destination node.

```
Require:
  data packet, location
Ensure:
  newlocation
  if the destination node is in the communication range
    send the data packet to the destination node;
    return 0;
  end if
  if in an intersection
    request neighbor vehicles including parked and
    moving vehicles which have the marks of the
    destination node to reply their locations;
    map the locations to the road segment in the map;
    //the replied locations should be one end of the road segment.
    send the data packet to the replied nodes;
    return the location of the other end of the road segment;
  end if
  send data packet to the node of next hop according to VADD;
  return location
```

Algorithm 4: Parked vehicle-assisted tracking.

We use a similar simulation setup of AGF. The radio propagation model that is using $[P_r(d)/P_r(d_0)]_{db} = -10\beta \log(d/d_0) + X_{db}$, X_{db} is a Gaussian distributed random variable with zero mean and standard deviation σ_{db}; β is the path loss exponent, and $P_r(d_0)$ is the reference received power for a close-in distance d_0 [6]. σ_{db} is set to 8, and β is set to 2 in city scenario and 4 in high scenario. The communication range is set to 550 meters, and 802.11b protocol is used for communication. Sources start generating data packets within the first 50 s of the simulation time and stop generating data packets 50 s before the simulation ends to avoid data packet drops due to the end of the simulation; all source/sink nodes stay inside the simulated area for the duration of their communications [6]. Speed variation and direction variation are same as Figure 1.

We can see from Figure 7 that VADD plus ABT plus PVAT achieve high packet delivery ratio, meanwhile it costs much small packet redundancy comparing to VADD plus ABT.

4. Conclusions

In this paper, we propose two solutions of the tracking problems existing in VANETs routing. ABT that has good packet delivery rate yet costs high packet redundancy, and PVAT costs less redundancy, but the packet delivery rate is not as good as ABT. We combine VADD, ABT, and PVAT to achieve both high packet delivery rate and low packet redundancy.

```
Require:
  data packet, destination location, destination node ID, location
Ensure:
  generate an area A according to (1);
  if size of A is smaller than B
    if destination location not belongs to A
    use VADD to find the next hop;
      return;
    end if
    use Algorithm 1 for data delivery; //call ABT(data packet, A, destination node ID)
    return;
  else
    if flag == PVAT
      use Algorithm 4 for data delivery; //call PVAT(data packet, location)
      return;
    else
      if location is in the communication range of destination location
        set flag to PVAT;
        use Algorithm 4 for data delivery; //call PVAT(data packet, location)
        return;
      else
        use VADD for data delivery;
        return;
      end if
    end if
  end if
```

Algorithm 5: VADD plus ABT plus PVAT.

Acknowledgment

This work is supported by National Science Foundation under Grant nos. 60903158, 61003229, 61170256, 61103226, and 61103227 and the Fundamental Research Funds for the Central Universities under Grant nos. ZYGX2010J074, ZYGX2011J073, ZYGX2011J060, and ZYGX2011J102.

References

[1] G. Pei, M. Gerla, and T. W. Chen, "Fisheye state routing: a routing scheme for ad hoc wireless networks," in *Proceedings of the IEEE International Conference on Communications*, pp. 70–74, June 2000.

[2] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'99)*, pp. 90–100, February 1999.

[3] D. Johnson and D. Maltz, "Dynamic source routing in Ad Hoc wireless networks," in *Mobile Computing*, pp. 153–181, Springer, New York, NY, USA, 1996.

[4] S. Jayachandran, J. Jothi, and S. Krishnan, "A case study on various routing strategies of VANETs," in *Global Trends in Computing and Communication Systems*, pp. 353–362, Springer, Berlin, Germany, 2012.

[5] B. Karp and H. T. Kung, "GPSR: Greedy Perimeter Stateless Routing for wireless networks," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MOBICOM '00)*, pp. 243–254, Boston, Mass, USA, August 2000.

[6] V. Naumov, R. Baumann, and T. Gross, "An evaluation of inter-vehicle ad hoc networks based on realistic vehicular traces," in *Proceedings of the 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC '06)*, pp. 108–119, May 2006.

[7] J. Zhao and G. Cao, "VADD: vehicle-assisted data delivery in vehicular Ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1910–1922, 2008.

[8] K. Shafiee and V. C. M. Leung, "Connectivity-aware minimum-delay geographic routing with vehicle tracking in VANETs," *Ad Hoc Networks*, vol. 9, no. 2, pp. 131–141, 2011.

[9] N. Liu, M. Liu, W. Lou, G. Chen, and J. Cao, "PVA in VANETs: stopped cars are not silent," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM '11)*, pp. 431–435, April 2011.

[10] A. Vahdat and D. Becker, "CS-200006," Tech. Rep., Duke University, 2000.

[11] C. Song, M. Liu, Y. Wen et al., "Towards the traffic hole problem in VANETs," in *Proceedings of the 9th ACM International Workshop on Vehicular Inter-Networking, Systems, and Applications, Low Wood Bay*, pp. 139–140, Lake District, UK, 2012.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 491756, 9 pages
http://dx.doi.org/10.1155/2013/491756

Research Article

PASS: Parking-Lot-Assisted Carpool over Vehicular Ad Hoc Networks

Jinqi Zhu,[1] Yong Feng,[2] and Bang Liu[1]

[1] *Department of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China*
[2] *Yunnan Key Laboratory of Computer Technology Applications, Kunming University of Science and Technology, Kunming 650500, China*

Correspondence should be addressed to Jinqi Zhu; cszjq@uestc.edu.cn

Received 9 November 2012; Accepted 18 December 2012

Academic Editor: Ming Liu

Information interaction is a crucial part of modern transportation activities. In this paper, we propose the idea of PASS: a parking-lot-assisted carpool method over vehicular ad hoc networks (VANETs). PASS aims at optimizing transport utilization by the carpooling among car drivers who cover a part of the same traveling route. With wireless device enabled in the vehicle, a user can easily get matched vehicles information and then express his travel demands via radio queries over VANETs to the corresponding driver. The driver can decide whether to provide carpooling services or not. We investigate the main challenges in PASS design, via the parking lot to collect vehicle trajectories, via accelerator sensor to sense vehicle's movement, establish a routing tree to delivery vehicle trajectory information to nearby parking lots, and design a suitable matching scheme to match the target vehicle in VANETs. Finally, simulation results prove that PASS is effective and efficient in carpooling among vehicle drivers.

1. Introduction

Vehicular ad hoc networks (VANETs) are network of vehicles which communicate with each other via wireless communications. They are emerging as a new technology, integrating the capabilities of new generation peer-to-peer (P2P) wireless networks with vehicles. In VANETs, vehicles can exchange messages with each other as vehicle-to-vehicle communication (V2V) or exchange messages with roadside infrastructures as vehicle-to-infrastructure (V2I) as well. Nowadays, with the technology development, various intelligent sensor nodes are equipped in vehicles. These sensor devices can sense various physical quantities with very low cost and high accuracy. VANETs have many appealing applications such as driving safety [1], Internet access [2], and intelligent transport and infrastructure monitoring [3]. However, such applications are all aimed at providing services for drivers in moving cars, while transport utilization of vehicles is ignored. In other words, the traditional V2V and V2I communication mode cannot offer any additional advantages to the waste of transportation abilities.

In recent years, although government and more and more researchers have paid more attention to improve transportation efficiency, the waste of transportation abilities is still general in current vehicle transportation. For example, according to NHTS [4] from the US Department of Transportation, the mean occupancy rate of personal vehicle trips is 1.6 persons per vehicle mile. Since a regular vehicle carries 5 persons in full occupancy, 68% of transportation abilities are wasted as a result. In USA alone, more than 87% of people travel in private vehicles every day [5], which result in a great deal of loss. On the other side, the large travel demand for personal car with low occupancies also leads to traffic congestion problem which has become an increasingly important issue in many countries. In 2007, congestion caused urban Americans to travel 4.2 billion hours more and to purchase an extra 2.8 billion gallons of fuel [6]. As more and more private vehicles are purchased and the degree of expensive transportation fuel cost in nowadays, the traffic congestion problem will be worsen. Moreover, personal cars also cause environment pollution inevitably. Thus, wise usage

of personal automobiles to reduce the consumption of energy and the emission of harmful gases is gaining more and more attraction.

Many transportation enterprises, carpool associations, and other institutions try to provide effective ride sharing for drivers, but the centralized methods they introduced are not scalable enough for real-time, dynamic, convenient, and rapid-changing carpooling service. Several carpooling methods have been proposed to increase the occupancy of vehicles, but these unfortunately, interaction methods are all made between moving vehicles and roadside passengers, while ride sharing among vehicle drivers is not considered. Some administrations have introduced high occupancy vehicle (HOV) lanes to encourage ride sharing. Casual carpools, as impromptu carpools formed among strangers, can team up in public areas near HOV lanes [7], but it is severely limited in deployed roads.

In this paper, we propose that a parking-lot-assisted carpooling method (PASS) over VANETs, which enable direct, instant, and flexible communication between car drivers. By taking the parking lot as a central database, the PASS method aims at optimizing transport utilization by ride sharing among car drivers who cover a part of the same traveling route. With wireless device enabled in the vehicle, a user can easily get matched vehicles information and then expresses his travel demands via radio queries over VANETs to the corresponding matched vehicle driver. The driver can decide whether to provide carpooling services or not. We investigate the main challenges in PASS design, via the parking lot to collect vehicle trajectories knowledge, via the accelerator sensor to sense vehicle movement, establish a routing tree to delivery vehicle trajectory information to nearby parking lots, and design a suitable matching scheme to match the target vehicle in VANETs. Since office worker and staff made up 68% user of current carpool service [8] people who work together usually prefer to live near each other in China. For example, most of teachers of our university live in the *bali* residential. We believe that PASS will be very useful for people who live neighboring and share common or adjacent destinations. Based on a realistic simulation, we prove that the PASS method is effective and efficient in ride sharing.

Our major contributions in this paper are summarized as follows.

(1) To the best of our knowledge, we are the first to consider the using of parking lot to assist ride sharing. Compared to previous carpooling researches, by encouraging car drivers to share their personal cars, PASS succeeds in improving transportation efficiency, reducing the price of travel cost and improving the traffic congestion problem.

(2) After clustering parked vehicles in a parking lot, we properly select a cluster head for vehicular trajectory knowledge gathering and matching. The key insight is to guarantee the effective and accuracy in trip matching.

(3) We design a suitable vehicle-matching scheme to match the target vehicles. As a result, a target user can inquiry matched vehicles according to his wishes.

The remainder of this paper is structured as follows: Section 2 discusses related works. Section 3 discusses PASS and its implementation. Section 4 evaluates PASS via extensive simulation and Section 5 summarizes the paper.

2. Related Works

In recent years, carpool system became a hot point for the VANETs researchers, and many studies on carpool have been proposed for VANETs. Traditional carpools usually take place among neighbors or coworkers who have similar routes, and they can easily contact with each other for possible ride sharing. Nowadays, with the development of the Internet, many carpool associations appear [9]. Thus, people are allowed to match their trips over the Internet, even though they are strangers. However, such carpooling systems have their own problems. Firstly, whether we can find a matched driver or not by using these systems entirely depends on the number of registered users. Since people are not widely encouraged to practice carpooling, for example, by governments or employers, the number of registered users is usually limited and the process of trip matching is usually failed as a result. Secondly, the centralized framework which relies on a central server to collect the information on all trip demands online is not scalable enough and flexible enough.

Alternatively, other endeavors aim to implement dynamic ride sharing have been considered in recent years. Paper [10] investigates the challenges in casual carpooling and suggests a travel matching system by extending current cell phone services. In paper [11], the authors undertake a review of technology-assisted carpooling in order to understand the challenge of designing participation and consider how mobile social software and interface design can be brought to bear. But this study mainly focuses on examining the possibility of designing participation, while a practical ride sharing system is not considered. In paper [12], Ghoseiri et al. present a dynamic rideshare matching Optimization model (DRMO) to match between the passenger and car drivers. DRMO receives passengers and drivers information and preferences continuously over time and assigns passengers to drivers with respect to proximity in time and space. The ride sharing preferences such as age, gender, and smoke are considered in the model. In [13], the authors define dynamic ride sharing and outline the optimization challenges that arise when developing technology to support ride sharing. However, not only the vehicle trip information collection process in this paper is obstacle to the system but also available ride sharing relies on the numbers of the drivers who are willing to input their trips. Moreover, the use of smartphones while driving is illegal in many countries and is also a major cause of accidents. Thus, this method is not so appropriate to support dynamic ride sharing. Plan et al. [14] explore the feasibility and challenges of WiFi-based ride sharing systems in metropolitan locations and indicate that WiFi connectivity works well while vehicles are traveling at slow speed. In [15], Liu et al. propose the idea of vehicle-to-passenger communication (V2P), which allows direct communication between moving vehicles and

roadside passengers. However, since passengers have to call moving vehicles via extra wireless devices, this method is inconvenient for passenger. Furthermore, the method or systems mentioned above all concentrate on ride sharing between passengers and car drivers but have not taken ride sharing among vehicle drivers into consideration. Compared with the initial ride sharing work, PASS aims at developing a flexible vehicle calling and travel-matching solution among car drivers.

Besides, http://www.carpoolworld.com/ [16] uses the commuter's precise latitude and longitude coordinates to find the best matches among the other commuters in the database; based on exactly how close together they live and exactly how close together they work. In [17–19], dynamic ride sharing approaches based on distributed framework have also been proposed, but these work lack the concrete realization of implementation of travel matching, which are far from a practical ride sharing system.

3. PASS and Its Implementation

In this section, we present the PASS method for dynamic ride sharing. We first describe the vehicle-matching idea and then focus on the implementation of PASS.

3.1. The Vehicle Matching Idea. For the idea of PASS, we first discuss how to get a user to communicate with a matched vehicle driver through a vehicular communication. As shown in Figure 1, a typical vehicle carpool process among car drivers has the following steps

(1) information acquisition: once a user activates the vehicle carpool application, he starts an onboard device such as the trip computer equipped in his car in order to acquire all the matched vehicles information. Since GPS, electric map, trip computer as well as wireless devices are widely deployed in today's vehicles, we believe that drivers are allowed to get useful information and communicate with each other through these devices.

(2) Send query message: the user chooses a matched vehicle according to his/her requirements and then sends a special query message over VANETs to the selected vehicle.

(3) Echo: when getting the query message via the trip computer, the matched vehicle driver can then decide to accept or ignore the message calling. Once the driver accepts the query message, an ACK message is sent to the user. They can further communicate with each other via QQ application or smartphones.

(4) Send query message again: if the query message is reject by the vehicle driver in step 3, the user will continue to send query messages until a right car driver is found.

In the implementation stage of PASS, some social and economic problem should be considered, which will be discussed later. Moreover, we assume query message can be received successfully by its receiver.

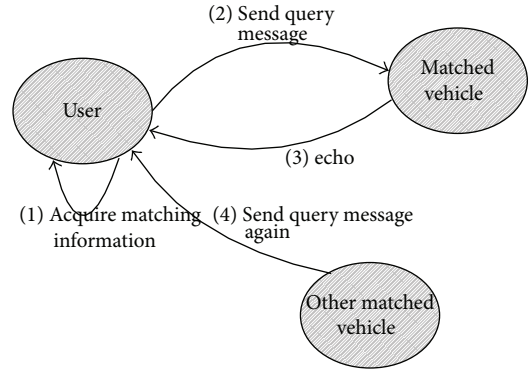


Figure 1: A typical vehicle carpool process.

3.2. Vehicle Trajectory Collection Assisted by Parking Lot. This section concentrates on parking-lot-assisted vehicle trajectory information collection and realization. Generally, from the viewpoint of vehicle drivers, the relationship of vehicles in driven on the urban street can be divided into four categories: vehicles with the same origin and the same destination, vehicles which have the same origin but different destinations, vehicles with different origins but the same destination and vehicles whose origins and destinations are distinct with each other. Since vehicles in the first three categories enter the same parking lot when parked, we explain the idea of parking-lot-assisted vehicle trajectory collection process from two cases consequently.

3.2.1. Vehicles Parked in the Same Parking Lot. For the vehicles have the same origins or the same destinations, they enter the same parking lot when parked. Thus, we try to collect vehicle trips assisted by the parking lot. Here we assume that some vehicle drivers would like to share their devices during parking. In other words, vehicle users can leave their wireless devices alive in parking, for downloading requested contents or continuing unfinished transmission.

As shown in Figure 2, at first all parked vehicles in the same parking lot are grouped into a cluster. Since the radio range of a vehicle can achieve to 250 m, the parked vehicles in the cluster can coordinate with each other from one hop or multihop communication. Besides, we assume each vehicle in the parking lot has a unique ID number (can be identified by vehiclelicense plate). Secondly, a vehicle located in the parking lot is randomly elected main cluster head, as H1 in Figure 2. Other vehicles are taken as cluster members. Thirdly, since the cluster head may fail in the complex environment of parking lot, we appoint a spare head as H2. A spare head is a cluster member nearest to the cluster head, which always keep a copy of recent cluster status and a copy of gathered vehicle trajectories knowledge from the main cluster head. Thus, a spare head becomes a special cluster member, working as the "warm backup" for the management of the parking lot cluster. The introduced spare head enhances the robustness of the parking lot cluster. In our study, the main cluster head needs to handle the following tasks: (a) cluster management, including membership management

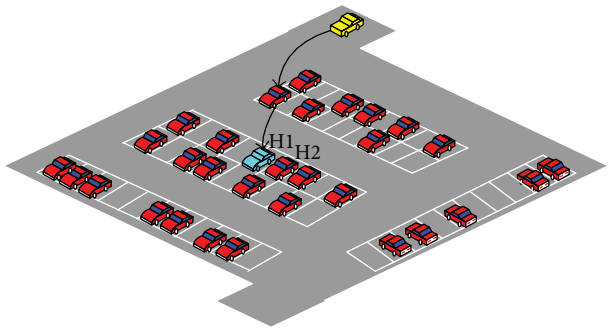


Figure 2: A typical parking lot cluster.

Figure 3: Trips for two drivers.

and vehicle trajectories knowledge management; (b) trip matching, that is, to find the matched trips from various vehicle trajectories. (c) Send matched vehicle information to demanders. Finally, once the vehicle acts as cluster head stops parking and drives away, the information it owes should be transmitted at once to another vehicle, which confirms the absence of the old cluster head and then becomes the new cluster head instead.

Each parked vehicle operates the following operations.

(1) Parking: once parked, a vehicle periodically senses its parked status and sends beacon messages to the main cluster head to report its existence.

(2) Joining a cluster: a vehicle receives the echo from the cluster head and then becomes a parking lot cluster member.

(3) Sending message: after operation 2, a vehicle begins to route its history vehicle trajectory records to the cluster head, whose main task is to maintain the trajectory records of all parked vehicles. As a kind of human activity, driving is controlled by individual drivers. As depicted in paper [20], human trajectories show a high degree of temporal and spatial regularity: each individual can be characterized by a time-independent characteristic length scale, and a significant probability to return to a few highly frequented locations. Paper [20] also discovers that two-to-four main locations (including home) cover more than 70% of the overall trips of a vehicle. Consequently, we can predict a vehicle's movement via its trip history for transparent travel matching.

(4) Send query message: once a vehicle driver enters his/her car and actives the vehicle carpool application through the on-board device, he chooses a promising driver on demand and sends a query message to the matched vehicle over VANETs.

(5) Leaving: once a vehicle senses its moving tendency through the accelerator sensor, it sends a message to inform the cluster head that it will stop parking and drive away from the parking lot.

3.2.2. Vehicles with Different Origins and Destinations. In regards to vehicles with different origins and destinations, dynamic carpool can be realized among car drivers contenting the following two cases: (a) for two drivers such as i and j, when both their origins and destinations near each other and (b) when their origins are adjacent to each other and the point of destination of driver i is within the acceptable walking distance, from the vehicle route of driver j. Since the average human walking speed is 3 m/s, if a person tolerates a walking time of ten minutes; for instance, the acceptable walking distance for a person is 1800 meters as a result.

Here we explain our idea by giving a simple example. In Figure 3, trips taken by driver i and j are concisely presented.

(1) Driver i: trip is from home (the origin is A) to work place (denote as B) at 8:40.

(2) Driver j: trip is from home (the origin is C) to work place (denoted as D) at 8:30.

We can see from Figure 3 that though the two trips differ in sources and destinations, since their sources and destinations are all not so far away from each other, the two divers can make a ride sharing once needed.

For vehicle drivers such as i and j mentioned above, since their vehicle trajectories are maintained by different cluster head, in order to find matched vehicles, we need a proper strategy to spread their history trip trajectories out to neighboring parking lot clusters over VANETs. To achieve this aim, when a vehicle stops parking and starts moving out, it gets the exact location of its local parking lot exit by GPS. This location will be taken as the location of the local parking lot in geography in our study. Later, this location information is passed to other vehicles and broadcasted on urban streets within a certain scope limit. Whenever a vehicle keeps this location information access a destined nearby parking lot, the location information message will be delivered to the destined cluster head via VANETs communication. In addition, in order to pass location information to all nearby parking lots as quickly and efficiently as possible, we consider vehicles parked at roadside near the underground parking lot. Figure 4 describes when a vehicle drives on a road, it delivers parking lot location messages it maintains to a roadside-parked vehicle, which keeps these messages and transmits them to a vehicle ready to enter the underground parking lot. These messages are finally passed to that underground parking lot cluster head. After a certain time of message

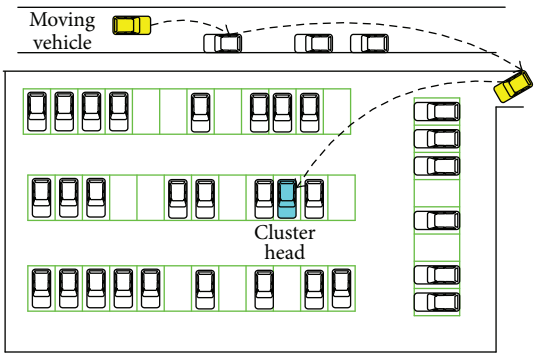


Figure 4: Position information dissemination.

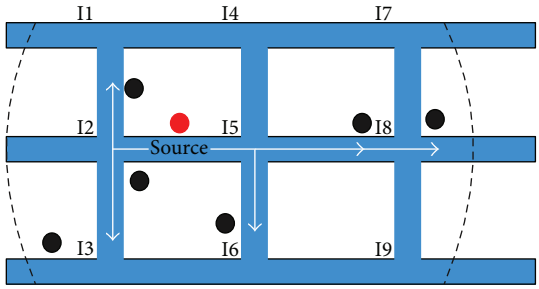


Figure 5: A routing tree for message transmission.

broadcasting, every parking lot cluster head masters the exact location information of all nearby parking lots.

In order to deliver vehicle trajectories mastered by a parking lot cluster head to all its nearby parking lots with low overhead, here a shortest path routing tree is constructed, as depicted in Figure 5. Messages are then routed along the routing tree to all its nearby parking lot.

Vehicle trajectories information forwarding process is described as follows. This process will be executed if and only if the vehicle trajectories knowledge mastered by the cluster head updated.

(1) At first, vehicle trajectory messages maintained by the underground parking lot cluster head (denoted as H1) are sent to a car which ready to drive away from this parking lot. These messages are then replicated to neighboring roadside-parked vehicles, which keep these messages and send them to moving vehicles in the local street directly, as shown in Figure 6.

(2) Secondly, in the city scenario shown in Figure 5, vehicle trajectory messages should be transmitted by moving vehicles along the routing tree to all parking lots within an acceptable detour scope (the size of this scope is determined by a special user). To achieve this aim, the messages from source parking lot are attached to a scope limit and destined for neighbor intersections as I2 and I8. We assume all intersection positions from I1 to I11 are marked in the electric map. Whenever a message carried and forwarded by vehicles in driven reaches the destined intersection, the message is discarded definitely. More specifically, we assume each data message has a field to record its survival time, whenever the message is timeout, it is discarded.

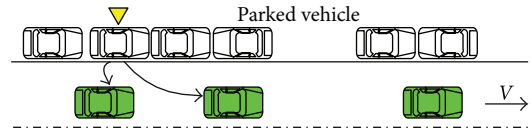


Figure 6: Data transmission from parked vehicle to moving vehicle.

(3) Finally, we let vehicle trajectory messages transmit to roadside parked vehicles near the destined parking lot. These messages are passed to a vehicle ready to enter the parking lot and finally delivered to the underground parking lot cluster head accordingly.

3.3. Vehicle Matching Scheme. Another task for the main cluster head of a parking lot is to match target vehicles for a user. For each driver i, a trip trajectory consisting of successive points, $v_1^i, v_2^i, v_3^i \cdots v_n^i$ with v_1^i as source point and v_n^i as destination point, is specified at the time that he/she starts this trip.

Here we take two vehicles A and B, for example. The start time and trajectories of A and B are described as T^A, v_1^A, $v_2^A, v_3^A \cdots v_n^A$ and T^B, v_1^B, $v_2^B, v_3^B \cdots v_n^B$, respectively. A matches B if and only if

$$\left|v_1^A - v_1^B\right| < d, \qquad \left|v_n^A - v_n^B\right| < d\,, \qquad \left|T^A - T^B\right| < \tau \quad (1)$$

or

$$\left|v_1^A - v_1^B\right| < d, \quad \text{distance}\left(v_n^A, v_1^B, v_2^B, v_3^B \cdots v_n^B\right) < d, \quad \left|T^A - T^B\right| < \tau, \quad (2)$$

where d is the acceptable walking distance assigned by a user, τ is the acceptable waiting time the user spends to wait a matched car, and distance $(v_n^A, v_1^B, v_2^B, v_3^B \cdots v_n^B)$ denotes the expected shortest distance from the destination of vehicle A to trip trajectory of vehicle B.

In our study, we let matching degree indicate the likelihood that two vehicles match each other. The higher the matching degree, the more possible two vehicle matched. Here function $f(x)$ is defined to calculate the matching degree value. For two matched vehicle A and B, if they satisfy formula (1), we have

$$f(x) = \alpha\left|v_1^A - v_1^B\right| + \beta\left|v_n^A - v_n^B\right| + \gamma\left|T^A - T^B\right|. \quad (3)$$

Otherwise, when formula (2) is content, we set

$$f(x) = \alpha\left|v_1^A - v_1^B\right| + \beta\ \text{distance}\left(v_n^A, v_1^B, v_2^B, v_3^B \cdots v_n^B\right) + \gamma\left|T^A - T^B\right| \quad (4)$$

here

$$\alpha + \beta + \gamma = 1. \quad (5)$$

Matching degree	Geography position	ID number
2	(30.6, 104.1)	905YV
1.4	(50.2, 104.1)	30328
1	(58.8, 150)	25687

Figure 7: Matched vehicles on on-board device.

In the above formulas, $0 \le \alpha \le 1, 0 \le \beta \le 1, 0 \le \gamma \le 1$ are constants decided on users demand. We can see from (8) and (9) that the definition of $f(x)$ is simple yet flexible, and it is entirely based on the user's will.

After trip matching, if one or more matched vehicles are founded, the local main cluster head sends matched vehicles information, including matching degrees, geography positions these matched vehicles parked and, their ID numbers to the target user once the user activates the vehicle carpool application. For user convenient, we believe that the vehicle with higher matching degree is more important and should appear with a higher priority on the target user's on-board device. This is done by arranging the matched vehicle on the target user's on-board device with a decreasing order of their matching degree. That is to say, vehicle with the largest matching degree value should appear at the most front, as seen in Figure 7. If no vehicle matches, a null message will be sent to the target user instead. The target user can then select a matched vehicle on demand and send query message via VANETs to that matched vehicle for carpooling service.

3.4. Implementation Discussion. Even though vehicle matching between car drivers is feasible in technique, the implementation still confronts lots of social and economic problems. We list them as follows.

Privacy is a matter that causes a lot of worry. Some drivers may have a problem with exposing their travel destinations. In this case, accurate travel destinations can be replaced by some common transit points, such as intersection or street name. For further anonymity, some encryption and decryption mechanisms should be considered. Furthermore, security and comfort are other important factors that deeply concern users. Strangers in the same vehicle need to have a basic level of trust for comfortable carpooling. We think that a reputation system is helpful in regulating individual actions in a trustworthy manner. Once a matched driver and the user make a deal and start a ride sharing, two sides can, respectively, send the deal note to the reputation system via VANETs. After finishing a carry, they can send feedback to one another, similarly.

Moreover, since people going to the same or adjacent working places usually prefer to live close to each other in China, the matched vehicle drivers found by using our method may know each other with a high probability. Thus, people using PASS are easier to trust each other and make a comfortable ride sharing than using other ride sharing methods mentioned before. Although involving many problems, the PASS method not only represents an opportunity to decrease the cost and undesirable impacts of traffic congestion, fuel consumption, and pollution, but also brings convenient to vehicle drivers.

Figure 8: Road topology in simulation.

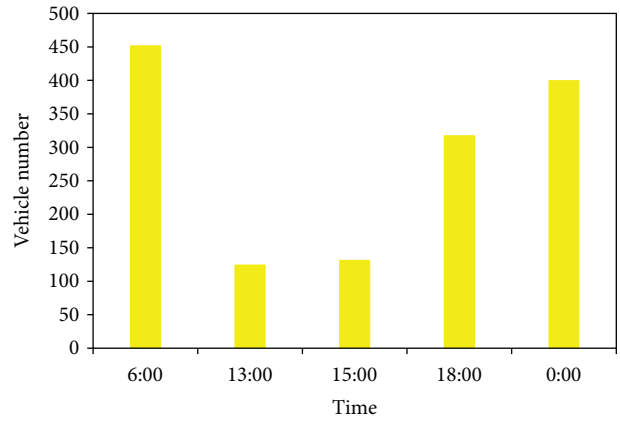


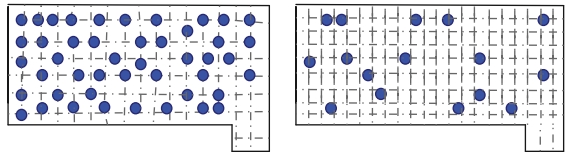
Figure 9: Parked vehicle number at different time.

4. Simulation

In this section, in order to evaluate PASS accurately, we examine the performance of PASS in vehicle carpooling.

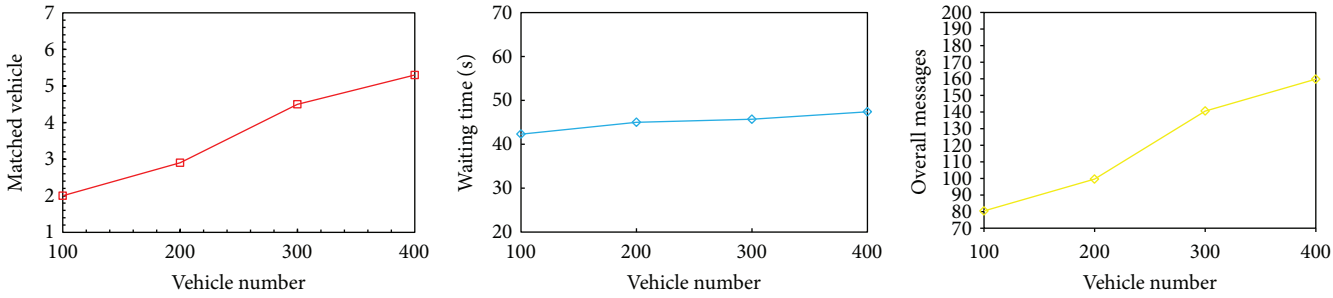
4.1. Simulation Environment. As shown in Figure 8, we extract a regional urban area with the range of 2400 m × 1600 m from a real street map of Chengdu, which is a city in China and contains 12 intersections and 20 bidirectional roads. Each intersection is marked by a number from 0 to 11.

To evaluate the vehicle carpooling in different traffic density environments, we deploy different vehicle numbers, that is, 100, 200, 300, and 400, to the map, with the average speed ranges of from 40 to 80 kilometers per hour. The radio range is set at 250 m, and the MAC protocol is 2 Mbps 802.11. The beacon interval is 1s, and all messages have a uniform size of 1 kb. In the simulation, we assume that each vehicle has enough buffer space to prevent message elimination from occurring. Besides, we find 14 parking lots are deployed in the



(a) 6 AM (b) 13 PM

Figure 10: Vehicle distribution at 6 AM and 13 PM.



(a) Matched vehicle (b) Waiting time (c) Overhead

Figure 11: Vehicle matching in carpooling.

urban region. Since accurately modeling vehicles' movement is very important for performance evaluation, we use the open source software, VanetMobiSim-1.1 [21], to generate the sources and destinations for all vehicles. That is, we let all vehicles drive from one randomly chosen parking lot to another parking lot in our simulation area. The movement of vehicles from source to destination on the road is also generated by VanetMobiSim-1.1 and can be directly utilized by NS-2.33. Since vehicles are more likely to enter or drive out of their local parking lot on the morning or from work due to human moving characteristics, inspired from long-term observation, we define 65% of the total vehicles moving out or driving in a parking lot at time periods 7:00–8:30 and 18:00–19:30. In other time periods, the overall vehicles moving in or out of a parking lot are provided to follow uniform distribution with a mean value of 10%. The average parking time is 41.40 minutes with a standard deviation of 27.17, which is provided in [22]. Moreover, in the simulation, we assume vehicles that moving on the road enter a randomly chosen parking lot with a probability of 10%. Some parked vehicle nodes are located on random positions of each street. Other simulation parameters and their default values are summarized in Table 1.

4.2. The Number of Vehicles in a Parking Lot. We investigate parked vehicle statistics in a residential parking lot at 6:00, 13:00 15:00, 18:00, and 24:00 of every Monday, Wednesday, and Friday, respectively. We counted the vehicle numbers parked in the parking lot and find that, generally, the total amount vehicles at 6:00 and 24:00 are much larger than vehicles numbers at 13:00 and 15:00 as shown in Figure 9. Besides, we also find that the parked vehicle numbers are stable in working hours of a day.

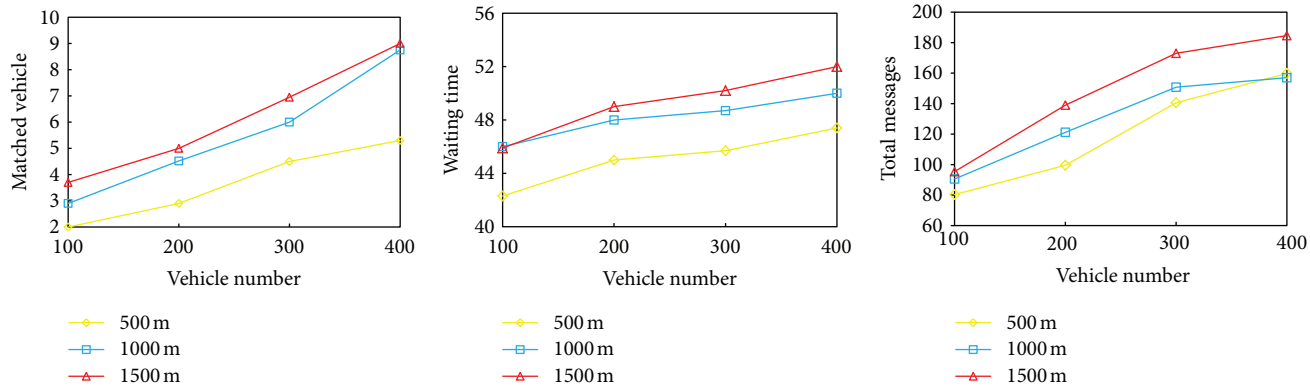
Table 1: Simulation parameters.

Parameter	Default value
Network size	3200 × 2200
Transmission radii R (m)	250 m
d	500 m
τ	30 minutes
α	0.4
β	0.4
γ	0.2

Figure 10 shows the number of vehicles at 6 AM and 13 PM in a parking lot, respectively, in our simulation. We can see that the number of parked vehicles changes, with the amount of vehicles parked at 6 AM is much larger than the vehicle number at 13 PM. This is consisted with our investigation since a large number of vehicles drive out at day time, which fits vehicle distribution in the parking lot.

4.3. Performance Analysis. In parking-lot-assisted carpool, we assume a vehicle driver who needs ride sharing has randomly chosen source parking lot and destination parking lot in the whole simulation area. Other vehicles are personal vehicles having randomly predefined movement sources and destinations. These vehicles record their history trajectories and deliver them to their local parking lot once parked. We assume that parking lot clusters are established at the beginning of simulation and are maintained at a cycle of 60 seconds.

Figure 11(a) demonstrates that the target vehicle can find an average number of 3–7 matched vehicles when the total vehicle number changes from 100 to 400. This is



(a) Matched vehicle (b) Waiting time (c) Overhead

FIGURE 12: d to vehicle.

reasonable because the number of vehicles parked in nearby parking lot increases then, and thus more matched vehicles would be found as a result. This result also means that our method is effective in matched vehicle finding. The impact of vehicle density to the effect of target users waiting time is presented in Figure 11(b). As we can see, the target user waiting time which is the time a user spends to wait for the matched vehicles information varies slightly with the increment of vehicle numbers, which denote that vehicle density does not have significant impact on the waiting time. Since PASS introduces roadside parked vehicles to speed up the dissemination of message, it also shortens the waiting time of users. Figure 11(c) illustrates the overhead in PASS in order to find matched vehicles. The overhead is defined as the total number of messages distributed in the simulation area in order to find matched vehicles. As we can see, the amount of messages is no more than 30. The main reason is that, by establishing a routing tree to forward messages, unnecessary message transfers are greatly reduced. With higher vehicle density, the overhead in PASS increases. This mainly stems from the vehicle trajectories collected by main cluster updated more quickly, and thus more messages are delivered.

In our study, the target user can gradually expand the acceptable walking scope to match nearby target vehicles. We exam the performance of PASS with different acceptable walking distance d, including 500 m, 1000 m, and 1500 m, in target vehicle matching, with results described in Figure 12. Figure 12(a) describes that the number of matched vehicles increases rapidly with the increase of acceptable walking distance. This is because more vehicles are provided in vehicle matching, and thus the target user can find target vehicle in a large area more easily. In Figure 12(b), users waiting time increase slightly with acceptable walking distance, while, in Figure 12(c), acceptable walking distance is positive to the network overhead, mainly because more nearby parking lots are founded with larger matching scope. Generally, simulation results show that the proposed PASS method for ride sharing is a simple and efficient vehicle-to-vehicle interaction. Although the carpooling response is not real-time, the average waiting time spent to get matched vehicles information is tolerable for users.

5. Conclusion

We presented a dynamic rideshare matching method assisted by parking lot (PASS) in this paper. PASS aims at identifying suitable matches between vehicle drivers requesting ride sharing services. The basic idea of PASS is simple since every vehicle in driven will be parked at a parking lot, why not we group these parked vehicles into a connective cluster and select a parked vehicle to support vehicle trajectories knowledge collection and trip matching? In VANETs, through the accelerator sensor equipped in a vehicle, the vehicle could sense the movement of itself. The cluster head then delivers all vehicle trajectories knowledge messages it mastered to this vehicle. These messages are transferred along urban streets to all nearby vehicles for trip matching. At last, the simulation results show the effectiveness of PASS. PASS contributes revolutionary vehicle ride sharing and promotes the vehicle carpool research in depth.

In the future, ride sharing preferences and characteristics such as age, gender, smoke, and pet restrictions as well as the maximum number of people sharing a ride should be considered in carpool matching algorithm design. Furthermore, how to use sensors equipped in today's vehicles to sense urban surroundings, such as traffic congestion, in order to reduce environmental effect on message delivery and trip matching will be considered.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grants No. 61103227, 61003229. The Fundamental Research Funds for the Central Universities No. ZYGX2011J060.

References

[1] L. Chisalita and N. Shahmehri, "A peer-to-peer approach to vehicular communication for the support of traffic safety applications," in *Proceedings of the 5th IEEE Conference on Intelligent Transportation Systems*, pp. 336–341, Singapore, 2002.

[2] H. Hartenstein, B. Bochow, A. Ebner, M. Lott, M. Radimirsch, and D. Vollmer, "Position-aware ad hoc wireless networks for inter-vehicle communications: the fleetnet project," in *Proceedings of the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '11)*, pp. 259–262, October 2001.

[3] P. B. Gibbons, B. Karp, Y. Ke, S. Nath, and S. Seshan, "IrisNet: an architecture for a worldwide sensor web," *IEEE Pervasive Computing*, vol. 2, no. 4, pp. 22–33, 2003.

[4] "National household 6 travel survey," http://nhts.ornl.gov/download.shtml.

[5] J. Saranow, "Carpooling for grown-ups high gas prices, new services give ride-sharing a boost, rating your fellow rider," Wall Street Journal, 2006.

[6] D. Schrank and T. Lomax, *2009 Urban Mobility Report*, Texas Transportation Institute, The Texas A&M University System, 2009.

[7] M. Burris, J. Winn, and W. Associates, "Slugging in houston-casual carpool passenger characteristics," *Journal of Public Transportation*, vol. 9, no. 5, p. 23, 2006.

[8] N. Chan and S. Shaheen, "Ridesharing in North America: past, present, and future," in *Transportation Research Board Annual Meeting*, Washington, DC, USA, January 2011.

[9] http://www.aapinche.cn/.

[10] F. K. Ozenc, L. F. Cranor, J. H. Morris et al., "Adapt-a-ride: understanding the dynamics of commuting preferences through an experience design framework," in *Proceedings of the Conference on Designing Pleasurable Products and Interfaces (DPPI '11)*, 2011.

[11] M. Brereton, P. Roe, M. Foth, J. M. Bunker, and L. Buys, "Designing participation in agile ridesharing with mobile social software," in *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group*, pp. 257–260, November 2009.

[12] K. Ghoseiri A Haghani M Hamedi, "Real-Time Rideshare Matching Problem," Tech. Rep., 2011.

[13] N. Agatz, A. Erera, M. Savelsbergh, and X. Wang, "Sustainable passenger transportation: dynamic ride-sharing," ERIM Report ERS-2010-010-LIS, 2009.

[14] S. Plan, S. Committees, P. Committees et al., "Wireless networks for car-and ridesharing systems: assessment of 802.11 Wi-Fi," in *Proceedings of the Transportation Research Board 88th Annual Meeting*, 2009.

[15] N. Liu, M. Liu, J. Cao, G. Chen, and W. Lou, "When transportation meets communication: V2P over VANETs," in *Proceedings of the 30th IEEE International Conference on Distributed Computing Systems (ICDCS '10)*, pp. 567–576, June 2010.

[16] http://www.carpoolworld.com/.

[17] Y. Fu, Y. Fang, C. Jiang, and J. Cheng, "Dynamic ride sharing community service on traffic information grid," in *International Conference on Intelligent Computation Technology and Automation (ICICTA '08)*, vol. 2, pp. 348–352, October 2008.

[18] A. Lue and A. Colorni, "A software tool for commute carpooling: a case study on university students in Milan," *International Journal of Services Sciences*, vol. 2, no. 3, pp. 222–241, 2009.

[19] K. Kelley, "Casual Carpooling—enhanced," *Journal of Public Transportation*, vol. 10, no. 4, p. 119, 2007.

[20] W. J. Hsu, T. Spyropoulost, K. Psounis, and A. Helmy, "Modeling time-variant user mobility in wireless mobile networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 758–766, May 2007.

[21] J. Härri, F. Filali, C. Bonnet, and M. Fiore, "VanetMobiSim: generating realistic mobility patterns for VANETs," in *Proceedings of the 3rd International Workshop on Vehicular ad hoc Networks (VANET '06)*, pp. 96–97, New York, NY, USA, 2006.

[22] A. Adiv and W. Wang, "On-street parking meter behavior," *Transportation Quarterly*, vol. 41, pp. 281–307, 1987.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 101878, 11 pages
http://dx.doi.org/10.1155/2013/101878

Research Article

Traffic-Aware Data Delivery Scheme for Urban Vehicular Sensor Networks

Chunmei Ma and Nianbo Liu

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

Correspondence should be addressed to Nianbo Liu; liunb@uestc.edu.cn

Received 9 November 2012; Accepted 18 December 2012

Academic Editor: Ming Liu

Vehicular sensor network (VSN) is a promising technology which could be widely applied to monitor the physical world in urban areas. In such a scenario, the efficient data delivery plays a central role. Existing schemes, however, cannot choose an optimal route, since they either ignore the impact of vehicular distribution on connectivity, or make some unreasonable assumptions on vehicular distribution. In this paper, we propose a traffic-aware data delivery scheme (TADS). The basic idea of TADS is to choose intersections to forward packets dynamically as the route from a source to destination based on link quality and remaining Euclidean distance to destination. Specifically, we first present an optimal utility function as the criteria of intersection selection. Besides the packet forwarding through intersections, we also propose an improved geographically greedy routing algorithm for packet forwarding in straightway mode. Moreover, in order to decrease the routing overhead brought by the traffic information gathering, we build a traffic condition prediction model to estimate the link quality. The simulation results show that our TADS outperforms existing works on packet delivery ratio, end-to-end delay, and routing overhead.

1. Introduction

With the advance of technology, most of the vehicles are equipped with on-board sensors; thus, a new network is the emerging-vehicular sensor network (VSN). Different from traditional sensor networks, VSN is not limited in energy consumption, owns much powerful processing units and wireless communication, and can determine the position of the nodes through GPS. Because of these properties, VSN has been envisioned to be useful in monitoring the physical world, especially in urban areas in which more vehicles equipped with sensors are available. For example, a vehicular network can be used to monitor traffic [1, 2] (e.g., jams, traffic accident, etc.), for improving driving convenience and efficiency. It can also be used in environmental surveillance [3], since urban areas are full of vehicles especially the taxis. To realize this vision, efficiently data transmission mainly relies on intervehicle communication in VSN. However, the special property of vehicle in mobility brings many challenges to VSN in throughput and latency of data transmission.

For the past decade or so, quite a few number of researchers and institutions were dedicated to improving transportation efficiency in urban VSN. Usually, these protocols mainly focus on selecting an intersection candidate to forward data according to the density of vehicles on the road [4, 5], such as VADD based on historical traffic condition to select a road with high density to forward packets. However, not only the density but also the distribution of vehicles on the road can affect the node connectivity. As shown in Figure 1, although the density is high (Figure 1(b)), unbalance traffic, caused by signal light and car-following driving, will lead to disconnection. Consequently, both the density and distribution of vehicles are needed to be taken into consideration for data transmission. In order to consider the vehicular distribution in routing protocol, some of protocols give unreasonable assumptions. For example, in [6, 7], they assume that the speed of the vehicles in the traffic flow is uniform, so the distribution of the space headway is exponential, which does not hold in reality scenario.

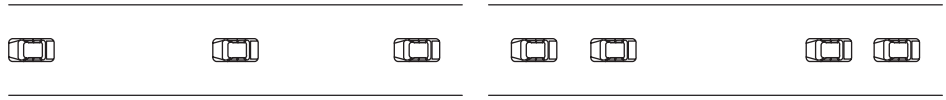


Figure 1: Connectivity in VSN.

In this paper, we solve the problem of efficient data delivery in vehicular sensor network when a vehicle sends a delay-tolerant data to some fixed site and how to efficiently route the packet to the destination. We propose a traffic-aware data delivery scheme (TADS), which emphasizes the intersection selection, that is, how to choose a robust route to forward packet. Different from existing routing protocols, TADS takes vehicular real-time spatial distribution together with vehicular density and the remaining Euclidean distance to the destination into consideration to choose an intersection for efficiently forwarding packet in VSN. For data delivery between intersections, we employ improved geographically greedy routing algorithm. We also build a vehicular density and vehicular spatial distribution prediction model based on the traffic pattern and road layout to estimate the link quality with time. Extensive simulations are conducted to evaluate the proposed protocol; the results show that TADS outperforms previous protocols in terms of packer delivery ratio, data packet delay, and routing overhead. TADS paves the way for delay-tolerant VSN applications like traffic surveillance and event reporting.

The major contributions of this paper are as follows.

(1) Uneven distribution of vehicles has great impact on routing performance; we present a method to denote the vehicular spatial distribution. Once the vehicular distribution can be obtained, the connectivity can be computed more accurately.

(2) Due to the traffic pattern and road layout, we establish a vehicular density and vehicular spatial distribution prediction model, through which TADS is efficient and resource efficient.

(3) We take vehicular density, distribution, and the remaining distance to the destination into account to design a routing protocol for VSN. The simulation results show that TADS outperforms most of the existing protocols.

The rest of the paper is organized as follows. In Section 2, we present a brief overview of related work on data delivery. In Section 3, we introduce the model of TADS. The detailed traffic-aware data delivery scheme will be introduced in Section 4. Section 5 evaluates the performance of TADS, and Section 6 summarizes the paper.

2. Related Work

In the last few years, a number of researchers have made many contributions to the routing protocol of ad hoc network. These routing algorithms are divided into two categories: topology-based and position-based routings. Topology-based protocols [8–10] should establish the route from source to the destination and maintain it in a table before the packet is sent, which are suitable for dense networks. Position-based protocols [11–13] leverage geographic positioning information to select the next forwarding hop; there is no need to create and maintain the global route between source and destination. As more and more vehicles are equipped with GPS and digital maps, position-based routing strategy is more convenient for urban vehicular sensor network. Owing to that the connectivity in urban vehicle networks heavily relies on the traffic condition, we classify position-based routing algorithms into two categories. One category exploits the historical traffic information to assist packet forwarding, and the other exploits the real-time traffic information to forward packet.

For the first category algorithms, there is no need to collect traffic information or make prediction. With the help of the navigation system, the traffic conditions are priori known before the packet to be sent. Vehicle-assisted data delivery (VADD) [4] is a kind of geographic routing protocol. It assumes that vehicles can obtain traffic statistics such as traffic density and vehicular speed on the road at different times of the day through preloaded digital map. The shortest expected delivery delay to the destination of the data depends on each road vehicular density. A drawback is that the historical information will not confirm to the realist traffic conditions that may cause routes to be incorrectly computed.

For the second category, the traffic conditions are gathered through an on-the-fly collection process. Greedy traffic-aware routing protocol (GyTAR) [6] is an intersection-based geographical routing protocol. For the intersection selection process, an estimation score is given to each junction by combining the road density and the curve metric distance to the destination. The junction with the highest value will be chosen for data delivery. The work in [14] selects an optimal route with the best network transmission quality model that takes into account vehicle real-time density and traffic light periods to estimate the probability of network connectivity and data delivery ratio for transmitting packets. Both of these two protocols divide the road into segments, and the vehicle headers in each of the segments, which are highly dynamic, are needed to exchange information. Intersection-based routing protocol (IBR) [15] finds a minimum delay routing path in various vehicular densities. Moreover, vehicles reroute each packet according to real-time road conditions in each intersection, and the packets sent to a node at the intersections depend on the moving

direction of the next vehicle. All of these protocols are needed to periodically collect road information, which significantly increases the routing overhead.

Most of these routing protocols do not take into account the vehicular distribution on the road where there may exist traffic hole [16], which means that it may be possible to lose some good candidate intersections when they try to forward a packet. Moreover, some of these protocols divide roads into segments, which are too complicated to be realized in the networking protocol design. To provide a solution to the aforementioned problems, in this paper, we take the vehicular density and distribution into account to design a new intersection-based routing protocol. The proposed protocol is easy to operate. In the following section, we will give a detailed description of our approach and present its added value compared with other existing vehicular routing protocols.

3. The TADS Model

3.1. Assumptions. We assume that each vehicle in the network is equipped with GPS, which enables them to acquire their own position and speed, and vehicles maintain a neighbor table that is built through beacon messages. The neighbor table records each neighboring vehicle's position, velocity, and moving direction. Furthermore, vehicles are equipped with preloaded with digital map, with the help of which we can determine the current geographical position of the destination. Finally, we assume that the traffic condition will not change significantly for a period of time.

3.2. Problem Statement. In city scenarios, data delivery has two modes: intersection mode and straightway mode. The most important issue is to select a robust route that can reduce delivery latency and improve delivery ratio. Although some of the routing protocols, such as ACAR [14] which always selects route with high density, are sometimes very inefficient for data delivery, as they ignore the distribution of vehicle on the road. As shown in Figure 2, the source node S wants to send a message to the destination D at the corner of intersection I_d. To forward the message through $I_a \rightarrow I_c \rightarrow I_d$ would be better than through $I_a \rightarrow I_b \rightarrow I_d$, even though the route $I_a \rightarrow I_b$ owns higher density. The reason is that the uneven distribution caused by traffic accidents or signal lights may lead to disconnection of the message that has to be carried by the vehicle. Since wireless communication is far faster than vehicle moving, the routing protocols should do their effort to leverage wireless communication.

Therefore, a proper routing protocol is important to forward the packet in VSN, and it is determined by several factors such as vehicular density, vehicular distribution, and road length. In this paper, we first model the link quality and then propose an approach to select the optimal route that can achieve the lowest forwarding delay.

3.3. TADS Model. TADS is an intersection-based multihop routing protocol that is capable of finding optimal route

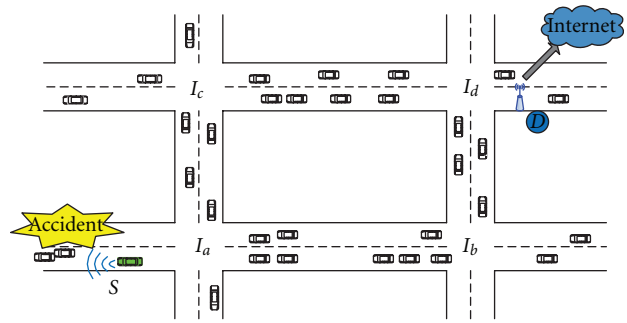


Figure 2: The problem of routing in VSN.

according to the real-time traffic condition. Owing to that the traffic flow will not change rapidly within a short time in urban scenario, we can build the link quality prediction model to increase the beacon cycle. To reach these objectives, TADS is organized into two schemes: (1) a scheme for the selection of the optimal route through which packets at each junction make decision to forward the packet to the destination; (2) a mechanism for the estimation of vehicular density and vehicular distribution to estimate link quality with time, which can significantly reduce the overhead of the routing protocol. Using TADS, packets will ultimately reach to the destination.

3.3.1. Intersection Selection. Similar to GPSR, TADS is a partial routing protocol and dynamically chooses the next intersection. When the packet carrier is approaching an intersection, it will calculate the utility function which depends on the link quality and the Euclidean distance from the candidate to the destination. Usually, the optimal route is the geographically closer candidate intersection to the destination and the route having higher link quality.

To better understand how the next intersection is chosen, it is illustrated by Figure 3 as an example. Since the source node S is approaching the intersection, it computes link quality of each neighboring road based on the traffic density and distribution. In order to guarantee the package forward to destination, the shorter distances from the candidate intersection to the destination will make a contribution to selecting the intersection. The source node knows the Euclidean distance from each candidate intersection to the destination through map. Thus, Swill not select the intersection far away from destination. Intersection I_a has the optimal utility function and is chosen as the next one.

To formally calculate the utility function, we define the following notations.

(i) r_{ij} : The road from intersection I to intersection J;

(ii) L_{ij} : The Euclidean distance on r_{ij};

(iii) p_{ji_k} : The position of vehicle k on r_{ij};

(iv) N_{ij} : The number of vehicles on r_{ij};

(v) σ : The vehicular distributions on r_{ij};

(vi) D_J : The Euclidean distance from the candidate intersection J to the destination.

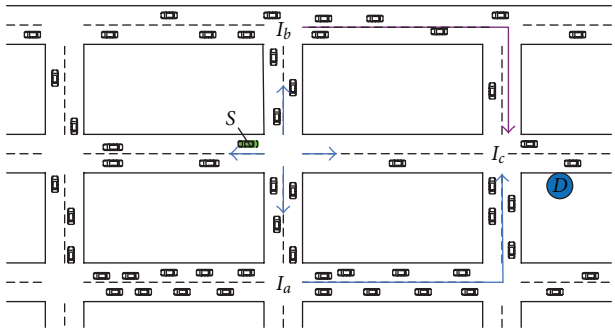


Figure 3: An example of selecting an intersection.

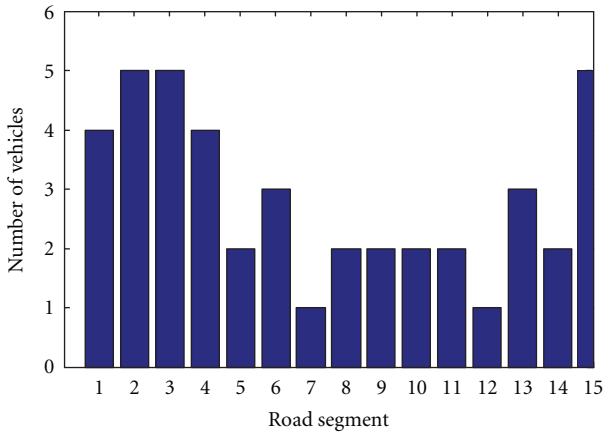


Figure 4: Vehicle spatial distribution on road.

Vehicular density depicts the status of the road segment that owns the ability to forward packets, it can be formulated on the road r_{ij} as

$$\rho_{ij} = \frac{N_{ij}}{L_{ij}}. \tag{1}$$

Before giving the vehicular distribution on road r_{ij}, we propose a definition called relative distance degree, which indicates the relative distance between the intersection and the current position of the vehicle.

Definition 1. For any node k and its position is p_{ij_k} on the road r_{ij}, the relative distance degree is

$$l_{ij_k} = \frac{p_{ij_k}}{L_{ij}}. \tag{2}$$

We can apply (2) to all vehicles on road r_{ij} and then acquire the vehicular spatial distribution. For each road, we employ a histogram to obtain the statistics results. Since the radio range is 200 m–250 m, we set the interval of histogram to 100 m; x-axis represents the road segments, the y-axis indicates the number of vehicles per segment, and the fluctuation of the graph indicates the vehicular unbalance distribution. Figure 4 is an example of a single road of length 1500 m. In the graph, the first 100 m has five vehicles, and there is only one vehicle between 600 m–700 m, which will significantly affect the routing performance.

Histograms of all the road segments are calculated based on the collected beacons from vehicles on the road. Suppose that n_m denotes the number of vehicles in the ith interval zone, the deviation (it indicates the balance or unbalance vehicular distribution) of the road r_{ij} can be expressed as

$$\begin{aligned}\sigma_{ij} &= \frac{\sum_{m=1}^{N}\left(n_m - \sum_{m=1}^{N}\left(n_m/N\right)\right)^2}{N} \\ &= \frac{\sum_{m=1}^{N} n_m^2 - 2\sum_{m=1}^{N} n_m + \left(N_{ij}^2/N\right)}{N} \\ &= \frac{N\sum_{m=1}^{N} n_m^2 - N_{ij}^2}{N^2}.\end{aligned} \tag{3}$$

Hence, the utility function can be formulated as

$$\Theta_{\rho\sigma D_J} = \left(1 - \alpha\frac{D_J}{D_K + D_H}\right)\frac{\rho_{ij}}{1+\sigma_{ij}}. \tag{4}$$

α is used as a weighting factor for the distance. $D_K = \max(D_1, D_2, \ldots, D_m)$ and $D_H = \min(D_1, D_2, \ldots, D_m)$.

As we can see, the utility function depends on three parameters (ρ, σ, D_J). For a given street, (1) provides the traffic density ρ and through the statistics results of the relative position of vehicles, the deviation σ (indicates if the road is balanced or not) is calculated using (3), whereas $D_J/(D_K + D_H)$ represents distance degree of different route, which guarantees that the packet is forwarded to the destination. $\rho_{ij}/(1+\sigma_{ij})$ determines how high the whole street link quality is. Thus, by multiplying $\rho_{ij}/(1+\sigma_{ij})$ with $\alpha(D_J/(D_K + D_H))$, we provide the streets with a correct value since this corresponds to scenarios where the candidate intersections are far away destination.

3.3.2. The Traffic Condition Prediction Model. In order to achieve the intersection selecting, we need to periodically collect the road information (e.g., the vehicle moving direction, speed, and position) to calculate the link quality, which consumes large resources, especially the bandwidth, and has bad impact on routing performance, for example, increasing routing overhead. However, owning to that the traffic flow may keep steady within a short time period, the traffic condition can be predicted based on the intersection historical traffic flow and moving vehicles on the road.

Link quality depends on the vehicular density and vehicular spatial distribution on the road; thus, in this section, we propose a vehicular density and vehicular spatial distribution prediction model to estimate the link quality. For prediction, each packet carrier maintains a road table where three tuples, $s_i(p, v, a)$, are recorded; here the three elements represent the position, velocity, and moving direction of vehicles on the road, and the table is divided into two sets $T1$ and $T2$ according to the vehicle moving direction. Each road periodically broadcasts its state to update the road table.

The problem of prediction of the link stability at time t is to compute the future density $\rho(t)$ and the deviation

$\sigma(t)$. The deviation is determined by the statistics results of relative distance degree; therefore, we just need to predict the vehicular number and their positions on the road. The vehicular number is affected by two parts: (1) the new coming vehicles and (2) the vehicles that have left. Suppose the vehicle arrivals at each intersection follow Poisson distribution. Thus, the probability of i of vehicles coming at time t is

$$P(A(t) = i) = e^{(-\lambda t)} \frac{(\lambda t)^i}{i!}. \quad (5)$$

When $P(A(t) = w) < h_c$, we ignore the possibility of coming vehicles; thus, the average number of arriving vehicles $A(t)$ is

$$A(t) = \sum_{i=1}^{w-1} e^{(-\lambda t)} \frac{i(\lambda t)^i}{i!} = \sum_{i=1}^{w-1} e^{(-\lambda t)} \frac{(\lambda t)^i}{(i-1)!}. \quad (6)$$

At time τ, the packet carrier receives the information from each road, then, it can compute the vehicle arrival rate of each intersection and applies (6) to estimate the number of the new coming vehicles. For the vehicles that have left, it is much simpler. The vehicles moving in the packet expectation direction belong to $T1$, and the others belong to $T2$. In this paper, we use D to denote the number of vehicles that have left. In $T1$, when $P_k + V_k t > L$, we can determine the vehicle has left the road, that is, $D = D + 1$ and in the same way, in $T2$, when $P_K - V_k t < 0$, $D = D + 1$. Thus, the number of vehicles on the road at time t is

$$N'(t) = N(t) + A(t) - D. \quad (7)$$

And then, we can apply (1) to compute the vehicular density on the road. For the vehicular distribution, it is essential to determine the relative distance of the new coming vehicles. We first consider the situation for moving in the packet expectation direction.

For Poisson distribution, N vehicles arrive at the intersection at time t, w_i denotes the ith arrival waiting time, and the conditional probability of waiting time is

$$f(t) = \begin{cases} \frac{n!}{t^n}, & \text{if } 0 < w_1 < w_2 < \cdots < w_n < t, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Suppose the new coming vehicles move at the average speed of road, and y_n denotes the nth vehicular position; thus, its relative position is

$$y_n = (t - w_n) V. \quad (9)$$

Hence, the conditional probability density function of the new arrival vehicles relative position is

$$f(y) = \begin{cases} -\frac{n!}{(t - y/V)^n}, & \text{if } 0 < y_1 < y_2 < \cdots < y_n < L, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

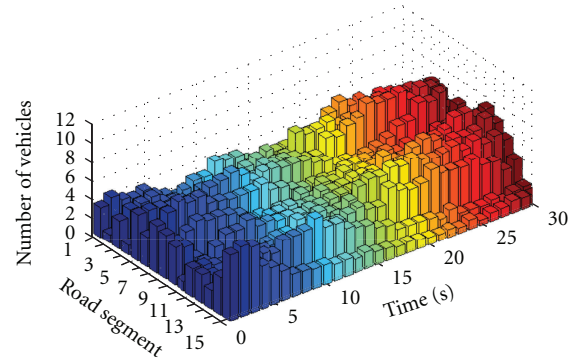


Figure 5: Prediction model.

The estimation position of nth vehicle is

$$E(y_n) = \int_0^L -\frac{yn!}{(t - y/v)^n} dy. \quad (11)$$

If the first arrival vehicle belongs to the first zone, the following ones do not need to be calculated, since they will not pass it; else, the second one is also computed until the one falling into the first zone emerges. For the vehicles on the opposite of the expectation direction, the relative position is

$$Y_n = L - y_n. \quad (12)$$

For the vehicles on the road, the relative distance is $P_K + v_k t$, v is a vector. When obtaining all the relative distances, we apply (2) to compute the relative degree and then apply (3) to compute vehicular spatial distribution.

Figure 5 shows an example of the change of vehicular spatial distribution over time. In this scenario, the vehicle arrivals rates of two intersections on the road are 0.6 and 0.3, respectively.

4. Traffic-Aware-Based Data Delivery Protocol

TADS is designed for city scenarios VSN routing, which only has two modes: intersection and straightway. In this section, we orderly present the protocol for the two modes.

4.1. TADS Used in the Intersection Mode. When approaching the intersection, the packet carrier can determine the best forward path deriving from (4) and then check if there is an available relay node to forward packets toward that intersection. As Figure 6(a) shows, vehicle A at intersection has a packet to forward to certain destination. Assume that north is the optimal direction for this packet. Both B and C are available relay nodes, for difference, B moving north and C moving south. TADS will select B as the next hop instead of vehicle C, since it can guarantee that the packet is forwarded to the optimal road. If there is only vehicle C available which is geographically closer to north, A will select C as the relay node, since C has the possibility to forward the packet to D immediately. Due to that the moving direction of C is on the opposite of the expectation direction, A always

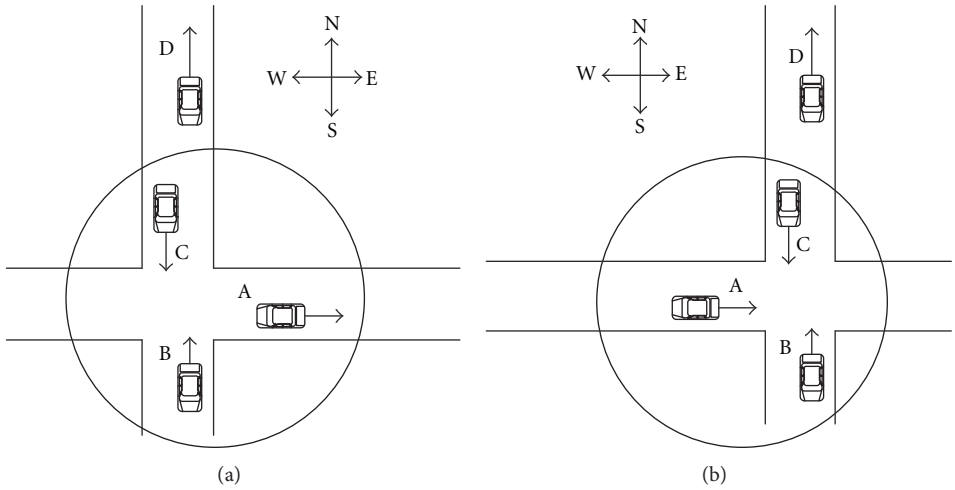


Figure 6: Packet forwarding in the intersection mode.

restores the packet until it receives the ACK message from C. If A failed to meet any relay node at current intersection, it will keep holding the packet and move ahead, it still has the opportunity to forward the packet to the expectation direction, as Figure 6(b) shown. When A passes through current intersection, nodes B and C are likely to emerge into the communication scope of A, and it can make a routing planning.

4.2. TADS Used in the Straightway Mode. In the straightway mode TADS applies improved geographically greedy routing algorithm to forward the packets to the destination intersection (the intersection ahead). To implement the scheme, all packets are marked by the location of the next intersection, and the packet carrier needs to record the velocity vector information of each neighbor vehicle. When a packet is received, the forwarding vehicle uses the corresponding recorded information to select the next neighbor that is closest to the destination among the nodes moving in the packet expectation direction. For instance, as shown in Figure 7(a), A has a packet to forward to the next intersection, both B and C are its neighbors, A will select C which is geographically closer to destination to forward the packet. In Figure 7(b), even though B is further from the destination, A will select B to forward the packet, since this seems like better than selecting C. This is because if there is no relay node for C, it will pass the packet to B shortly, which increases the delay, and B can ensure the packet to be sent to destination.

5. Performance Evaluation

In this section, we analyze the weighting factor α and evaluate the performance of the prediction model and TADS. We use VanetMobiSim-1.1 [17], a flexible framework for vehicles mobility modeling, to generate the real traces that can be used by NS2. Since the simulation should be offered a network environment as close as possible to the real world one, we make an effort to define a realistic scenario where VSN may be deployed.

5.1. The Weighting Factor Analysis. In this section, we analyze the weighting factor of the utility function to determine the good balance between distance and link quality. We simulated the packet delivery ratio of TADS for different values of α. The simulation scenario is shown in Figure 10, and we set the number of vehicles to be 250. As Figure 8 shows, in most of the cases, $\alpha = 0.4$ achieves the highest packet delivery ratio. This is mainly because the vehicle moving speed is much slower than the wireless communication, and it is better to favor a higher link quality to forward the packet.

5.2. Evaluation of the Prediction Model. We simulated a 1500 m long straight road with two bidirectional lanes and set the period of vehicle light to 60 s. The traffic light can cause unbalanced vehicle distribution in various degrees, which has great impact on the performance of prediction model. In order to analyze the performance in various situations, we must set the simulation time to be greater than 60 s (in this section we set the simulation time to be 70 s). Since the different vehicles arrival rate of each intersection has impact on the performance of prediction accuracy, we extract different arrival rate to display our algorithm. In order to give the analysis result of the prediction model, firstly we define the error rate.

Definition 2. $\sigma(t)$ and $\sigma * (t)$ are the estimated and real deviation of road L at time t, respectively, and the error rate is given as

$$\text{err} = \frac{|\sigma(t) - \sigma * (t)|}{\sigma * (t)}. \tag{13}$$

Figure 9 plots the error rate of the prediction of vehicular spatial distribution over time. The figure shows that the error rate is lower than 10% for the low arrival rate and lower than 13% for the higher arrival rate when the estimation time is under 30 s, which means that the estimated deviation

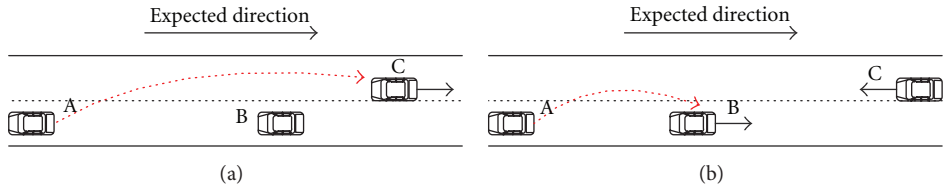


Figure 7: Packet forwarding in straightway mode.

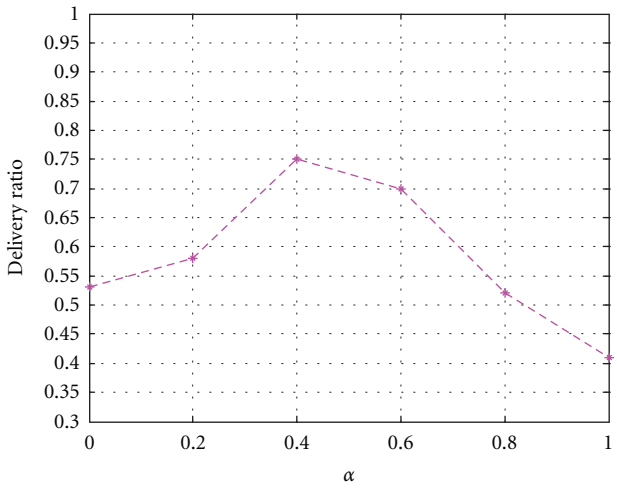


Figure 8: The delivery ratio versus weighting factor.

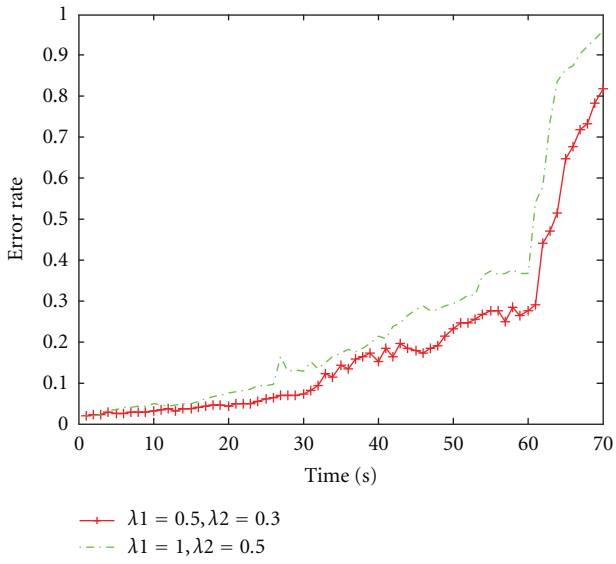


Figure 9: The error rate of road over time.

is close to the actual deviation, and it is observed that longer estimation time causes more errors rate. This is mostly explained by the fact that the vehicles are highly dynamic; by contrast, in our model we set a constant arrival rate during the estimation process which causes larger error rate with time increases. Furthermore, it is observed that when the time exceeds 60 s, the error rate sharply increases. This is because,

Figure 10: Road topology of simulation area.

in reality scenario, the traffic red light is working beyond 60 s, and at the current intersection the vehicles stop moving into the road, which is consistent with the reality scenarios.

5.3. Evaluation of TADS. To evaluate the performance of the TADS, we conduct a more complex simulation and compare our algorithm with two existing protocols: GSR [18] and ACAR [14]. Since GSR always selects the forwarding node that is closest in geographical distance to the destination and drops packets when the network is disconnected, to have a fair comparison, we extend GSR by adding carry-and-forward schemes in it.

5.3.1. Simulation Setup. The simulation was based on a real street map with the range of 1600 m × 1400 m, which is extracted from an urban area of Chengdu, a city of China. As shown in Figure 10, it consists of 14 intersections and 21 bidirectional roads. At the beginning of the simulation, a number of vehicles are located on the position to the map following the predefined value, and each of them chooses one of the intersections as its destination. Then, they start moving along the road on both directions with an average speed range from 40 to 80 km/h that depends on the speed limit of a specific road. To produce traffic change, the total numbers of nodes are deployed in the region ranging from 100 to 300, and the traffic light period is set to 60 s to produce unbalanced vehicular distribution. Among all vehicles, 10 of them are source nodes to send CBR data packets at 0.1 to 1 per second with a packet size of 512 B. All the key simulation parameters are listed in Table 1.

Table 1: Simulation parameters.

Parameter	Value
Simulation time (s)	400
Simulation area (m)	1600 × 1400
Number of intersections	14
Period of traffic lights (s)	60
Number of vehicles	100150200250300
Communication range (m)	250
Vehicle velocity (km/h)	40–80
CBR (packet per second)	0.1–1
Packet size (B)	512
Vehicle beacon interval (s)	1.0, 30

5.4. Simulation Results. We mainly evaluate our algorithm on packet delivery ratio, end-to-end delay, and routing overhead as a function of the data transmission rate and the vehicular density in simulation. As a static metric, packet delivery ratio is the number of successfully received packets at the destination divided by the total number of packets in the networks, which is significantly affected by the simulation of time. End-to-end delay reflects transmission duration between two nodes, which indicates how long it takes for a packet to be forwarded across the network from the source to the destination. Routing overhead denotes the ratio between the total number of the control packets and the number of data packets sent into the networks and the control packets.

The routing protocols are compared under various data transmission rates that set the number of vehicles as 150 and various vehicular densities that set the packet sending rate as 0.3 pkt/s.

Packet Delivery Ratio. Figure 11(a) shows that GSR has the lowest packet delivery ratio, even if it is implemented in a carry-and-forward way, as it always chooses the geographically shortest path to destination without considering the vehicular traffic. Consequently, some data packets cannot reach their destination due to that the wireless transmission quality is low on some sections of the road. On the other hand, for almost all packet sending rates, TADS gives the highest packet delivery ratio, since it forwards packets along the route on the road following the road traffic density, vehicular spatial distribution, and the Euclidean distance to the destination. Hence, a packet will successively arrive at the destination along the streets with the highest transmission quality. ACAR has a lower delivery ratio than TADS, and this is because it just considers vehicular density to estimate the probability of network for transmitting packets. Consequently, some data packets cannot reach their destination due to the problem of traffic hole.

In Figure 11(b), it is observed that as the vehicular density increases, GSR achieves very good delivery ratio, and since the connectivity is much better than the previous scenario, there are more nodes that can help carry and forward the packets to the destination. For ACAR, its packet delivery ratio will increase when the network density is low, as it will forward packets along the path with higher connectivity. However, when network density is larger than 150 nodes, its packet delivery ratio slightly increases. When the network density becomes larger, ACAR may choose the highest density road to forward the packet, which causes MAC layer collisions, so the delivery ratio cannot drastically increase and sometimes may decrease. Due to full consideration of connectivity in TADS, the optimal utility function may be different; thus, there are few collisions, and the packet delivery ratio of TADS increases when network density increases.

End-to-End Delay. As Figure 12(a) shows, the end-to-end delay is a function of packet sending rates. When the packet sending rates increase, more opportunities are obtained to forward the packet to the destination; thus, the forwarding delay will decrease. GSR has the largest end-to-end delay compared to ACAR and TADS, which is mainly due to the long time vehicles carry packets as there is no next hop available. ACAR has relatively lower end-to-end delay at most of the packet sending rates. An interesting observation is that when the packet sending rate is closer to 0.9 prt/s or when the vehicular density increases (in Figure 12(b)) to 200 nodes, GSR shows lower delay. This is because of the forwarding rules in ACAR; when more packets are injected into the same route, there are more packet collision and longer queuing time. In this case, the end-to-end delay in ACAR will increase.

In Figure 12(b), TADS achieves the lowest end-to-end delay; this is because in TADS, it has an efficient scheme of selecting intersection that guarantees that the packet is sent to the destination with the least delay, and in straightway, the chosen sending node is the one closest to the candidate intersection which reduces the number of hops involved in delivering packets, which to some degree reduces the delivery delay.

Routing Overhead. In this section, we evaluate the routing overhead of these protocols. For certain packet sending rates, the total number of packets sent into the networks is similar for all protocols; thus, the routing overhead is determined by the control message. As Figure 11 shows, although the packet delivery ratio of ACAR is higher than GSR, when the packet sending rate increases, GSR outperforms ACAR in routing overhead (Figure 13(a)), and the major reason is the on-the-fly density collection scheme in ACAR. Due to the prediction scheme, the periodic beacon interval of TADS is much longer (30 s in the simulation) than the other two protocols; thus, TADS has the lowest routing overhead.

As shown in Figure 13(b), the routing overhead will increase along with the increase in vehicular density, since the size of control messages is proportional to the number of vehicles in the networks. Although GSR needs to send hello message to maintain its neighbor table, it achieves lower routing overhead than ACAR for different vehicular density. This is because the cost is lower compared with the real-time density collection for a whole street. In TADS, even if the size of control message increases similarly, the long periodic beacon interval decreases the ratio between the control packets and the total packets sent into the networks.

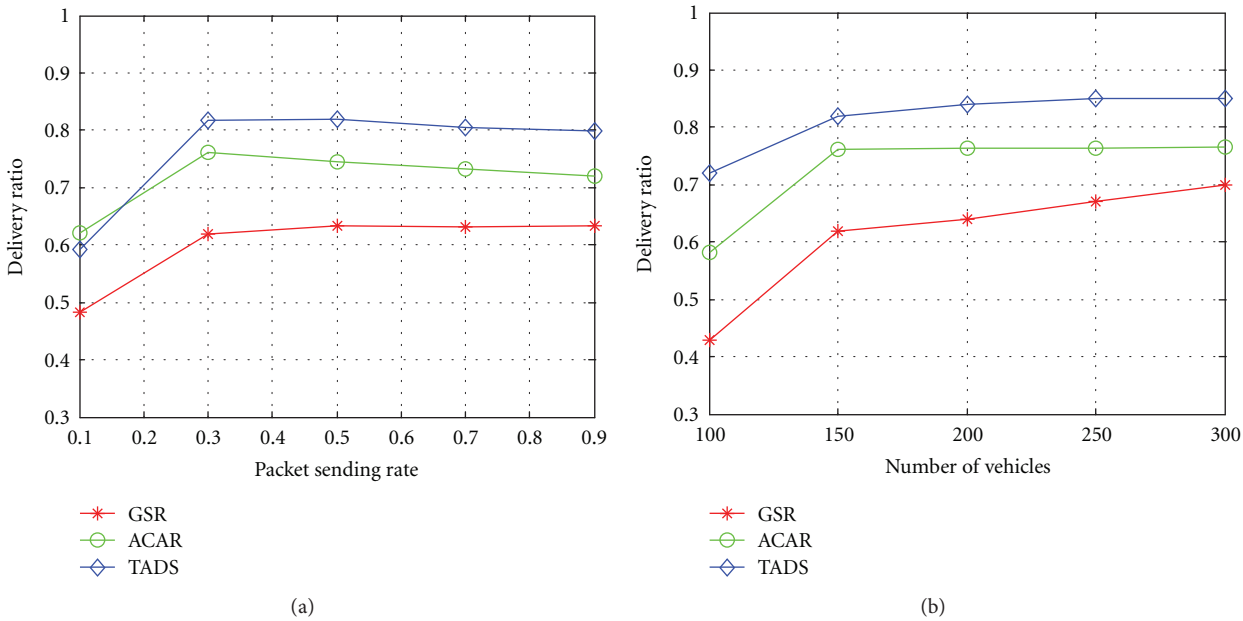


Figure 11: Data delivery ratio versus (a) packet sending rate and (b) number of vehicles.

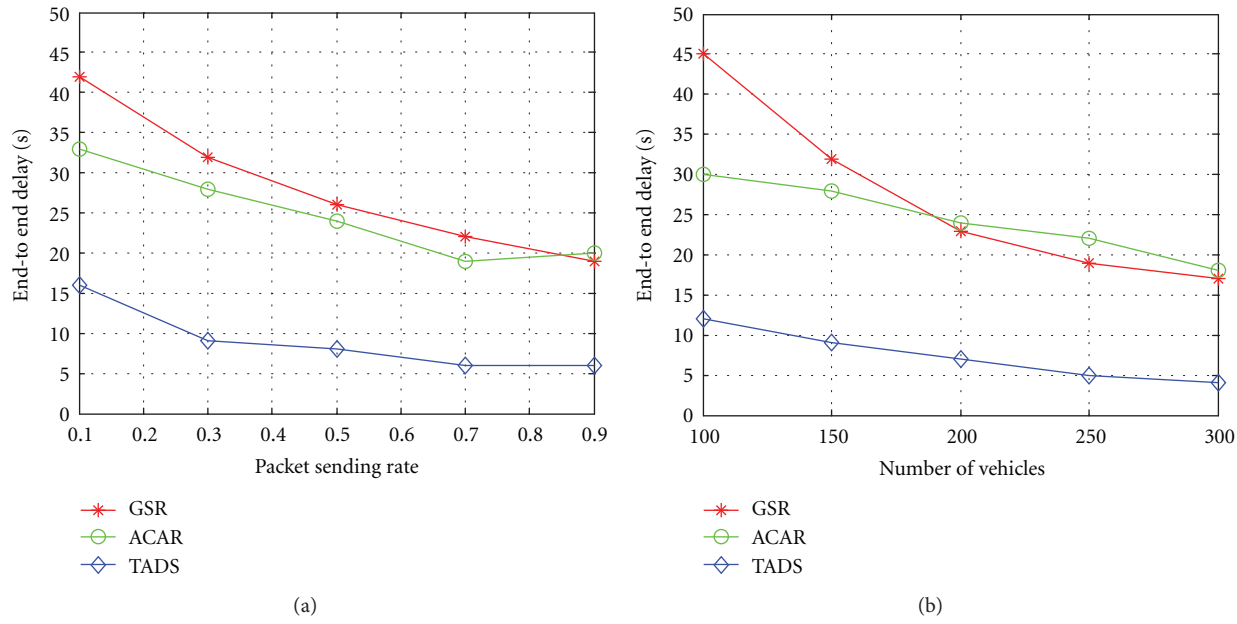


Figure 12: Data end-to-end delay versus (a) packet sending rate and (b) number of vehicles.

6. Conclusions

Motivated by the great impact of uneven distribution on the performance of routing protocol, we propose an efficient and a lightweight traffic aware-based data delivery scheme (TADS) to achieve data delivery, which will benefit many applications for urban VSN. The selection of intermediate intersection is based on the optimal utility function to each road. The utility function is determined by the dynamic traffic density, the Euclidean distance to the destination, and the vehicular distribution. Different from most of the existing protocols, we make use of real realistic vehicular distribution. In this paper, we first give a definition of relative distance degree, which can reflect the real distribution of vehicles. Then, we make use of the mathematical statistics method to calculate the deviation of road per 100 m. The result reflects the different distribution among road segments. Due to the character of vehicle mobility, which is constrained by

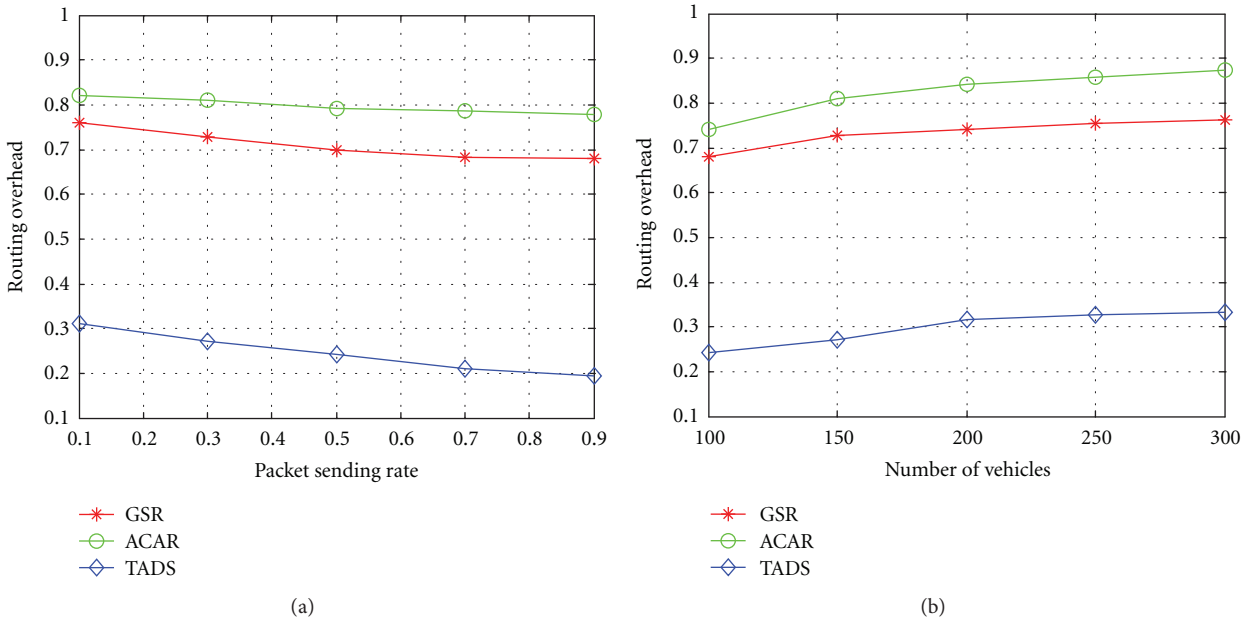


Figure 13: Data routing overhead versus (a) packet sending rate and (b) number of vehicles.

the road topology and traffic pattern, we build a prediction model. Our simulation results show that TADS achieves a higher successful throughput and lower delay with low cost compared to GSR and ACAR.

Acknowledgments

The authors sincerely thank their shepherd and the anonymous reviewers for their invaluable feedback which helped improve this paper. This work was supported in part by the National Natural Science Foundation of China under Grants (nos. 61003226, 61170256, 61173171, 61173172, 61103226, 61103227, and 61272526), the Fundamental Research Funds for the Central Universities under Grants (nos. ZYGX2011J060, ZYGX2011J073, and ZYGX2011J074).

References

[1] T. ElBatt, S. K. Goel, G. Holland, H. Krishnan, and J. Parikh, "Cooperative collision warning using dedicated short range wireless communications," in *Proceedings of the 3rd ACM International Workshop on Vehicular Ad Hoc Networks (VANET '06)*, pp. 1–9, September 2006.

[2] C. L. Robinson, L. Caminiti, D. Caveney, and K. Laberteaux, "Efficient coordination and transmission of data for cooperative vehicular safety applications," in *Proceedings of the 3rd ACM International Workshop on Vehicular Ad Hoc Networks (VANET '06)*, pp. 10–19, September 2006.

[3] L. E. Cordova-Lopez, A. Mason, J. D. Cullen, A. Shaw, and A. I. Al-Shamma'A, "Online vehicle and atmospheric pollution monitoring using GIS and wireless sensor networks," *Journal of Physics: Conference Series*, vol. 76, no. 1, Article ID 012019, 2007.

[4] J. Zhao, "VADD: vehicle-assisted data delivery in vehicular adhoc networks," in *Proceedings of the 25th IEEE International Conference on Computer Communications*, pp. 1–12, 2006.

[5] Q. Yang, "Connectivity aware routing in vehicular networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '08)*, pp. 2218–2223, Piscataway, NJ, USA, 2008.

[6] M. Jerbi, S. M. Senouci, T. Rasheed, and Y. Ghamri-Doudane, "Towards efficient geographic routing in urban vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 5048–5059, 2009.

[7] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, and O. Tonguz, "Routing in sparse vehicular ad hoc wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 8, pp. 1538–1556, 2007.

[8] G. Pei, M. Gerla, and T.-W. Chen, "Fisheye state routing: a routing scheme for ad hoc wireless networks, Communications," in *Proceedings of the IEEE International Conference on Communications (ICC '00)*, vol. 1, pp. 70–74, New Orleans, La, USA, 2000.

[9] C. Perkins, E. Belding-Royer, and S. Das, "Ad Hoc on-demand distance vector (AODV) routing," in *RFC 3561, Network Working Group*, 2003.

[10] Z. J. Haas, "The zone routing protocol (ZRP) for ad hoc networks," in *Internet Draft*, 1997.

[11] B. Karp and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MOBICOM '00)*, pp. 243–254, August 2000.

[12] C. Lochert, M. Mauve, H. Fler, and H. Hartenstein, "A routing strategy for vehicular ad hoc networks in city environments," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 156–161, June 2003.

[13] B. C. Seet, G. Liu, B.-S. Lee, C.-H. Foh, K.-J. Wong, and K. K. Lee, "A-STAR: a mobile Ad Hoc routing strategy for

metropolis vehicular communications," in *Mobile and Wireless Communications*, vol. 3042, pp. 989–999, 2004.

[14] L. D. Chou, J. Y. Yang, Y. C. Hsieh, D. C. Chang, and C. F. Tung, "ACAR: adaptive connectivity aware routing for vehicular Ad Hoc networks in city scenarios," *Mobile Networks and Applications*, vol. 15, no. 1, pp. 36–60, 2010.

[15] L.-D. Chou, J.-Y. Yang, Y.-C. Hsieh, D.-C. Chang, and C.-F. Tung, "Intersection-based routing protocol for VANETs," in *Proceedings of the 2nd International Conference on Ubiquitous and Future Networks (ICUFN '10)*, pp. 105–124, June 2010.

[16] C. Song, M. Liu, Y. Wen, G. Chen, and J. Cao, "Towards the trafic hole problem in VANETs," in *Proceedings of the 9th ACM International Workshop on Vehicular Inter-Networking, Systems, and Applications*, pp. 139–140, 2012.

[17] J. Harri, F. Filali, C. Bonnet, and M. Fiore, "VanetMobiSim: generating realistic mobility patterns for VANETs," in *Proceedings of the 3rd International Workshop on Vehicular ad hoc Networks (VANET '06)*, pp. 96–97, 2006.

[18] C. Lochert, M. Mauve, H. Fubler, and H. Hartenstein, "Geographic routing in city scenarios," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 9, no. 1, pp. 69–72, 2005.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 389874, 7 pages
http://dx.doi.org/10.1155/2013/389874

Research Article

A Hot-Area-Based Selfish Routing Protocol for Mobile Social Networks

Haigang Gong and Xiaomin Wang

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Correspondence should be addressed to Haigang Gong; hggong@uestc.edu.cn

Received 8 November 2012; Accepted 11 December 2012

Academic Editor: Ming Liu

Data delivery in mobile social network is a challenging task due to the nodal mobility and intermittent connectivity. It is natural to utilize the inherent social properties to assist in making forwarding decisions. However, existing routing schemes seldom consider selfishness of nodes and assume that nodes are willing to forward messages for others. In the real world, most people are selfish and nodes attached to people behave selfishly too. Based on the regularity of human behavior, we propose a hot-area-based selfish routing protocol (HASR) tailored for mobile social networks in this paper. Data transmission is based on the active degree of the node, which is calculated by the weight of hot areas that nodes will visit, when there are no selfish nodes. When nodes behave selfishly, routing decision is made by the contribution index that indicates the contributions to data transmissions of the network made by nodes. Simulation results show that HASR performs better when nodes behave selfishly.

1. Introduction

The portal devices such as smart phone, laptop, and tablet computer have been very popular in the world with the rapid development of the technologies of wireless communication and integrated circuit. These devices with the wireless capabilities such as Bluetooth, WiFi, or 3G are often carried by people and cooperate with each other to form an ad hoc network for exchanging and sharing their data. Social behavior analysis has been introduced to resolve the routing issues when nodes are attached to the human and could achieve better performance by using social relationship or human behavior in real life environment. Hui et al. [1] named the network Pocket Switch Networks (PSNs), a type of Delay Tolerant Networks (DTNs) [2]. Because there have been some inherent social features in the network, the network is also called Mobile Social Networks (MSNs).

Since mobile social networks have the great potentials of collaborative data exchanging, opportunistic routing for MSNs has attracted a great interest. Unfortunately, it is hard to find an end-to-end path between the source node and the destination node in the networks and the network is usually intermittently connected due to the nodal mobility and the spare distribution of nodes, which pose great challenges for data delivering in mobile social networks. Different from traditional delay tolerant networks, nodes in mobile social networks are often controlled by people so that nodes have some social features due to the social relationships or social ties among people. It is natural to utilize the inherent social properties to assist in making forwarding decisions.

Some institutes try to find social properties of the mobile social networks in the real world based on the data set collected from the portable devices attached to human, for example, Reality Mining [3], Topology Discovery [4], and Haggle [5]. And there have been lots of routing strategies for mobile social networks, such as LABEL [6], BUBBLE [7], SimBet [8], Peoplerank [9], and PLBR [10], which all employ the social network properties to help message forwarding. However, in the previous routing techniques, there is a common assumption that all nodes in the network are unselfish and coordinated. That is to say, each node is willing to receive and relay the messages sent by other nodes. In the real world, most people are selfish and there would be some selfish nodes dominated by human. For example, the resources (energy, buffer, bandwidth, etc.) of nodes are usually limited and nodes try to preserve their own resources

while just consuming the resources of other nodes. Besides, most people are willing to forward messages for nodes with whom they have social relationship. Social selfishness will affect node behaviors.

In this paper, we propose a Hot-Area-based Selfish Routing protocol (HASR) tailored for mobile social networks. Based on the regularity of human behavior, which people may visit some locations regularly and would spend more time at a few specific locations than other locations, HASR proposes a scheme to resolve routing issue for the network with selfish nodes.

The rest of the paper is organized as follows. Section 2 discusses related works. Section 3 describes the system model. Section 4 introduces the details of HASR. Section 5 shows the simulation results of HARP. Section 6 concludes the paper and presents the future work.

2. Related Work

In the recent years, social structures have been used to help forwarding in intermittently connected networks and there have been a lot of studies on data gathering for social networks []. Routing techniques in mobile social networks could be classified into two categories: social relationship based approach and human behavior-based approach.

2.1. Social Relationship Based Approach. In the real world, there are inherent social relationships between people such as relatives, friends, colleagues, schoolmates, and so forth. The relationships usually maintain stable in a long period of time. Based on the social relationships, message could be forwarded efficiently.

Hui and Crowcroft have proposed a routing algorithm called LABEL which takes advantage of communities for routing messages [6]. LABEL partitions nodes into communities based on only affiliation information. Then ach node in the network has a label telling others about its affiliation. A node only chooses to forward messages to destinations, or to the next-hop nodes belonging to the same group (same label) as the destinations. LABEL significantly improves forwarding efficiency over oblivious forwarding using their dataset, but it lacks a mechanism to move messages away from the source when the destinations are socially far away.

BUBBLE combines knowledge of the community structure with knowledge of node centrality to make forwarding decisions [7]. Centrality in BUBBLE is equivalent to popularity in real life, which is defined as how frequently a node interacts with other nodes. People have different popularities in the real life so that nodes have different centralities in the network. Moreover, people belong to small communities like in LABEL. When two nodes encounter, the node forward the message up to the node with higher centrality (more popular node) in the community until it reaches the same level of centrality as the destination node. Then, the message can be forwarded to the destination community at the same ranking (centrality) level. BUBBLE reduces the resource consumption compared to Epidemic and PRoPHET. However, this reduction may not be large since the ranking process creates significant communication overhead. In addition, this protocol still uses multicopy forwarding which means that it is not efficient in terms of resource consumption.

SimBet presented in [8] makes routing decisions by centrality (betweenness) and similarity of nodes. Centrality means popularity as in BUBBLE. More specifically, the centrality value captures how often a node connects nodes that are themselves not directly connected. Similarity is calculated based on the number of common neighbors of each node. SimBet routing exchanges the preestimated centrality and locally determined similarity of each node in order to make a forwarding decision. The forwarding decision is taken based on the similarity Utility Function (SimUtil) and Betweenness Utility unction (BetUtil). When nodes contact with each other, the node selects the relay node with higher SimBet utility for a given destination.

LocalCom proposed by Li and Wu [11] is a community-based epidemic forwarding scheme in disruption-tolerant network. LocalCom detects the community structure using limited local information and improves the forwarding efficiency based on the community structure. It defines similarity metrics according to nodes' encounter history to depict the neighboring relationship between each pair of nodes. A distributed algorithm, which only utilizes local information, is then applied to detect communities and the formed communities have strong intracommunity connections.

In social-greedy [12], forwarding decision is made by the closeness and social distance. Closeness is calculated by the common attributes (address, affiliation, school, major, city, country, etc.) of the two nodes. The more common attributes, the closer the two nodes. Social-greedy forwards a message to the next node if it is socially closer to the destination. Social-greedy outperforms the LABEL protocol. However, the delivery ratio of Epidemic and BUBBLE is better than social-greedy.

2.2. Social Relationship Based Approach. Another social-based routing strategy employs the regularity of human behavior to aid in routing decision.

Liu presents a cyclic MobiSpace [13], which is a MobiSapce where the mobility of the node exhibits a regular cyclic pattern there exists a common motion cycle for all nodes. In a cyclic MobiSpace, if two nodes were often in contact at a particular time in previous cycles, then the probability that they will be in contact around the same time in the next cycle is high. Based on this phenomenon, Routing in Cyclic MobiSpace (RCM) scheme is proposed. Routing decision is made by the Expected Minimum Delay (EMD), which is the expected time that an optimal forwarding scheme takes to deliver a message at a specific time from a source to a destination, in a network with cyclic and uncertain connectivity. When nodes contact, messages would be relayed to the next-hop with minimum EMD.

Liu et al. consider that there are preference locations that people visit frequently and they propose preference location-based routing strategy (PLBR) [10]. Firstly, PLBR provides the approach of acquiring one's preference locations and then calculates the closeness metric which is used to measure

the degree of proximity of any two nodes is proposed. On the basis of that the data forwarding algorithm is presented. The closeness is defined to indicate the similarity of the preference locations that the two nodes visit. The higher the closeness of the two nodes is, the more the common preference locations they are. If the closeness of the two nodes is high, the probability of the two nodes contact is high. The messages would be forwarded to the next hop with the highest closeness. However, the calculation of the closeness requires the preference locations of the destination node, introducing large network overheads.

An expected shortest path routing (ESPR) [14] scheme improves PLBR by utilizing the stable property of humans that they have preference locations in their mobility traces, and the direct distance between node pairs can be calculated according to the similarity of their location visiting preferences. Then an expected shortest path length (ESPL) can be achieved by Dijkstra's algorithm. Messages are forwarded to nodes which are closer to the destination than the previous nodes in the message delivery history. In addition, ESPR also employs the priority of message in the queue management.

CSI [15] is a behavior-oriented service as a new paradigm of communication in mobile human networks, which is motivated by the tight user-network coupling in future mobile societies. In such a scenario, messages are sent to inferred behavioral profiles, instead of explicit IDs. At first, user behavioral profiles are constructed based on traces collected from two large wireless networks, and their spatio-temporal stability is analyzed. The implicit relationship discovered between mobile users could be utilized to provide a service for message delivery and discovery in various network environments. CSI shows that user behavioral profiles are surprisingly stable. Leveraging such stability in user behaviors, the CSI service achieves delivery rate very close to the delay-optimal strategy with minimal overhead.

3. System Model

3.1. Network Model. Assuming that in an $M \times M$ city scenario, the city is divided into several areas by the road. Each area has a unique ID. As shown in Figure 1, there are 10 areas such as $A1, A2, \ldots, A10$.

Definition 1 (hot area). A hot area is the zone in which nodes stay for a long period of time and visit frequently. Each area has its own weight call hot degree and a hot area has the higher hot degree. In city, the hot areas include the downtown, scenic spot, and so forth. In school, the dormitory, canteen, and library are hot areas for students.

Definition 2 (hot degree). Hot degree denotes the popularity of a hot area or the attraction for the mobile nodes. The more the hot degree is, the more frequently the area is visited.

3.2. Mobility Model. Based on the regularity of human behavior, the mobility of all nodes is assumed to follow the schedule-based mobility model described in [16], where each node carries a unique schedule that describes its whole day journey. Each item of the schedule indicates when and where the node will be. As shown in Table 1, the node arrives at $A1$ on 8:00 and stays in $A1$ for 100 minutes. Then the node leaves for $A2$ and will reach $A2$ on 10:00. A node moves only according to its schedule and it moves from the current location towards the next one in its schedule.

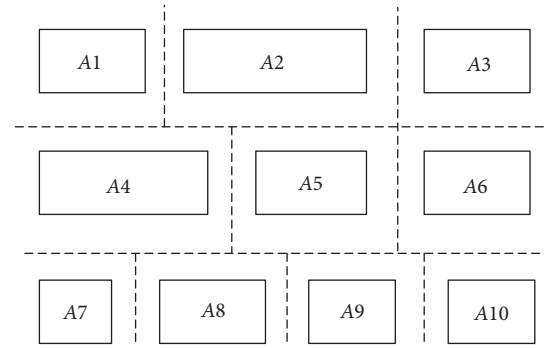


Figure 1: Network model.

Table 1: Agenda of the node.

ID of area	Arrival time	Residence time (min.)
$A1$	8:00	100
$A2$	10:00	90
$A3$	12:00	60
⋮	⋮	⋮
Ak	14:30	90
⋮	⋮	⋮

Whether a node is selfish or not could be determined by different metrics. For example, a node behaves selfishly when most of its resources have been depleted. When the resources of the node are consumed much, the node only delivers message for the node that relay message for itself before. The selfishness of the node could also be determined by the social relationship between nodes. A node only relays the message from nodes with strong relationships (e.g., friends, classmates, and colleagues) among them and is not willing to transfer the message for strangers. Without loss of generality, we assume that the node is selfish only when its energy is lower than a predefined threshold in this paper.

4. HASR Design

HASR modifies and extends our previous work HARP [17]. Data forwarding in HASR is based on the active degree and the contribution index of the node. When the energy of the node is enough, data forwarding is determined only by the active degree of the node. When the energy of the node is consumed too much, the node behaves selfishly and routing decision is based on the contribution index.

4.1. Hot Degree. Hot degree of an area reflects the popularity of the area and the attraction for the mobile nodes. There are more nodes in the area with higher hot degree, which means there are more chances to meet the destination node in the

area. Obviously, the hot degree of an area varies with time but it fluctuates a little for a long period of time. Assuming that there are N nodes entering into an area i per unit time, the residence time of node k in the area i is t_i. Then the hot degree could be expressed by

$$H_i = \frac{N}{\sum_{k=1}^{N} t_i/N} = \frac{N^2}{\sum_{k=1}^{N} t_i}, \qquad h_i = \frac{H_i}{\sum_{i=0}^{n} H_i}, \tag{1}$$

where h_i is the normalization hot degree of the area i and n is the number of the areas. As in (1), the hot degree is proportional to the number of nodes visiting the area and inversely proportional to the mean time staying in the area of these nodes. The more the number of nodes visiting the area, the higher probability to forwarding message is. By contrast, the mean time staying in the area indicates the activity of the area. The higher the mean residence time is, the less probability to leave for another area is.

How to get the hot degree of the areas is a fundamental problem for HASR. Hot degree of the area could be acquired from the data set collected from the real world for a long period of time such as Reality Mining project. Based on the data set, we can learn the number of nodes visiting each area and the residence time that they stay in that area. Then hot degree of the area can be calculated by (1). Table 2 shows an example of the hot degree of different locations in school.

4.2. Active Degree. In HASR, data forwarding decision is decided by the active degree of the node, which indicates the visiting frequency of the areas. As mentioned before, each node follows an agenda schedule like in Table 1. As in Table 1, the first column A denotes that the areas the node will visit today; the second column T is the time to arrive at the area; and the third column S indicates the residence time that the node stays in the area. Then the active degree of the node x is defined as the sum of the hot degree of the areas that the node will visit later today,

$$AD_x = \sum_{i}^{k} h_{A(i)} \quad \text{if } T_c < T(i) + S(i), \tag{2}$$

where $A(i)$, $T(i)$, and $S(i)$ are the values of the ith item of the agenda schedule table; k is the number of items of the table; T_c is the current time. The condition $T_c < T(i) + S(i)$ means that the node stays in the area $A(i)$ and the previous $i - 1$ items in the table is invalid. For instance, as in Table 1, if the current time is 11:30, the first item is invalid and the second time will be invalid 30 minutes later. When computing the active degree of nodes, the invalid items should be removed.

The active degree of the node reflects the capability to relay messages. Higher the active degree means the node is more popular and there is more chance to meet the destination node. When two nodes encounter with each other, messages will be transferred to the node with the higher active degree.

Table 2: The hot degree of different locations.

	Dormitory	Lab	Playground	Canteen	Library
Hot degree	0.233	0.182	0.132	0.216	0.177

Table 3: Transmission statistics of the node.

The ID of node	The number of relays, N_R	The number of transmissions, N_T
y	15	18
z	31	80
⋮	⋮	⋮
Network	577	890

4.3. Contribution Index. When the energy of the node is enough, forwarding decision is only decided by the active degree of the node. However, if the energy of the node is lower than a preset value, the node should behave selfishly to save its energy and routing decision is made by the contribution index of the node. That is to say, messages from the node with the more contribution to the network should be delivered when the next hop behaves selfishly.

The contribution index (CI) indicates the contributions to the data transmissions of the network and consists of two parts: network contribution and contribution for specific node. The network contribution can be determined by the number of messages generated by itself and the number of messages that it receives from other nodes and relays to the next hop. Similarly, the contribution for specific node could be decided by the number of messages that it sends to a specific node and the number of messages that the specific node sends to it.

Contribution index can be calculated based on the transmission statistics in the network. As shown in Table 3, each node maintains the table recording the data transmissions and updates the table once data transmission occurs.

Let N_R be the number of messages received from a specific node and N_T be the number of messages that send to the specific node. SN_R and SN_T are the number of messages received from other nodes and the number of messages generated by itself. Define the balance of contribution B for two specific nodes as

$$B = \frac{|N_R - N_T|}{N_R + N_T}. \tag{3}$$

From (3), if B is 0, the two specific nodes relay equal messages for each other. If B equals 1, it means that one node relays all messages from the other node that only sends message and does not relay messages. For selfish routing, the balance of contribution of the two nodes should be as small as possible.

```
Node x initializes its agenda
Calculates AD_x
Get the set of its neighbors, SET_N
Node y = the neighbor with the maximal AD_y
    if (the energy of node is enough)        // normal routing
        if (AD_y > AD_x)
            forwarding message to node y
    else if (AD_y == AD_x)
        computing H_x and H_y
        if (H_y > H_x)
            forwarding message to node y
        end if
    end if
else                    // selfish routing
    for node y in SET_N
        if (AD_y > AD_x and |CI_x − CI_y| < γ)    /
            forwarding message to node y
        end if
    end for
end if
```

Pseudocode 1: HASR pseudocode.

The contribution index of the node can be expressed by

$$\mathrm{CI} = \begin{cases} \alpha \cdot \dfrac{SN_R}{ST_R + SN_T} + (1-\alpha) & \\ \quad \cdot \dfrac{N_R}{N_R + N_T}, & \text{if } \left(SN_R > \lambda,\ B < \Delta\right), \\ 0, & \text{else,} \end{cases} \tag{4}$$

where α is a weight indicating the importance of the network contribution to CI; λ is a predefined constant that only when the number of messages the node relay exceeds λ, the network contribution is; Δ is a threshold and satisfies $0 < \Delta < 1$. Similarly, only when the balance of contribution of the two specific nodes is greater than Δ, the contribution for specific node could be computed. Clearly, Δ should not be set to a large number because a node is not willing to relay messages from the node who contributes little for its data transmission. For selfish node, it should receive and relay the messages from the node with the approximately equal contribution index.

4.4. Data Transmission. Data transmission in HASR is based on the active degree of the node when the energy is enough and the contribution index when routing selfishly. The pseudocode of HASR is shown in Pseudocode 1. The node calculates its active degree according to the initial agenda schedule at first. When it meets other nodes, they exchange the data transmission statistics. If the energy of the node is greater than the threshold, the node behaves unselfish and relays all messages without consideration of the contribution index. Then the node will send its message to its neighbor with the maximal active degree. The active degree of the node is different in different time for that some items in the schedule table would expire. If the active degrees of the two nodes are equal, it needs another metric to make routing decision. Let SET_x and SET_y be the set of areas in the valid items of the node x and node y; N_x and N_y is the number of elements of the two sets, respectively. N_{xy} is the number of the same elements of the two set and $N_{xy} = |\mathrm{SET}_x \cap \mathrm{SET}_y|$. Let SET'_x and SET'_y be the set of different areas that the two nodes will visit and there are $\mathrm{SET}'_x = \mathrm{SET}_x - \mathrm{SET}_x \cap \mathrm{SET}_y$ and $\mathrm{SET}'_y = \mathrm{SET}_y - \mathrm{SET}_x \cap \mathrm{SET}_y$. Define H_x and H_y as the sum of hot degrees in SET'_x and SET'_y. When node x contacts node y, if AD_x equals to AD_y and H_y is greater than H_x, messages would be delivered to the node y because node y is more active than node x. For example, if $\mathrm{SET}_x = \{A3(0.1), A4(0.15), A5(0.18)\}$ and $\mathrm{SET}_y = \{A1(0.2), A3(0.1), A4(0.15), A6(0.12)\}$, which the values in brackets is the hot degree of the area, the different areas of the two set are $A1$, $A5$, and $A6$. Then there are $H_x = 0.18$ and $H_y = 0.2 + 0.12 = 0.32$, and messages will be transmitted to node y because H_y is larger than H_x.

When the energy of the node is consumed too much, the node would behave selfishly and routing decision is made by the contribution index. The pseudo code of HASR is shown in Pseudocode 1. As in Pseudocode 1, node x calculates its AD_x according to its agenda set by people who hold it. When node x meets with node y, if the energy of node x is enough, message will be delivered to the node y with higher AD_y. If the AD_x equals AD_y, message will be delivered to the node y with higher H_y. Once the energy of node x is lower than a preset threshold, node x will send messages to node y based on the contribution index. If AD_y is greater than AD_x and they have almost the same contribution index, message will be transferred to node y. γ denotes the difference between the contribution indexes of the two nodes.

Table 4: Simulation parameters.

Parameter	Default value
Number of sensor node	100
Initial energy of each sensor node (J)	10
Size of each messages (bit)	250
Communication radius (m)	100
λ	50
Δ	0.1
γ	0.1
α	0.6
E_{elec}	50 nJ/bit
ε_{fs}	10 pJ/bit/m^2
ε_{mp}	0.0013 pJ/bit/m^4

5. Simulation

5.1. Simulation Setup. We simulate three protocols: the proposed HASR, HARP, and RSD [18] and evaluate their performance on data delivery ratio. We assume that message generation of each node follows a Poisson process and the destination node is randomly selected. To calculate the energy consumption, we use the same radio energy dissipation model as in [19]. The initial energy of each node is 10 J and the node behaves selfishly when half of the energy is consumed. Other simulation parameters and their default values are summarized in Table 4.

5.2. Simulation Results. Figure 2 shows that the impacts of weight factor α on data delivery ratio of HASR. α determines the contribution index is calculated mainly by the network contribution of the node or not. As in Figure 2, the data delivery ratio increases with the rising of α. When α is small, the contribution index is mainly determined by the contribution for the specific node. Even if the node contributes more to data transmission for the network, its message might not be relayed by other node. For example, node x encounters node y. Node x did not relay any message from node y before which means there is little contribution for node y, but it contributes more for the whole network. However, node y will not relay the message from node x because node x did not serve it before. When α increases, the network contribution is more important than the contribution for the specific node. The node that contributes more for data forwarding in the network has more chance to deliver its messages. When α is greater than 0.6, there is little increase on the data delivery ratio. In the simulation, α is set to 0.6.

The influence of the proportion of selfish nodes denoted by ρ on data delivery ratio is shown in Figure 3. Obviously, the more the number of selfish nodes is, the lower the data delivery ratios of HASR and HARP are. When ρ is greater than 0.8, the data delivery ratio is just about 10%. As seen from Figure 2, the data deliver ration of HARP decreases drastically than that of HASR with the increase of the ratio of selfish nodes. The reason is that when a portion of node behaves selfishly, they still delivery messages according to the contribution index.

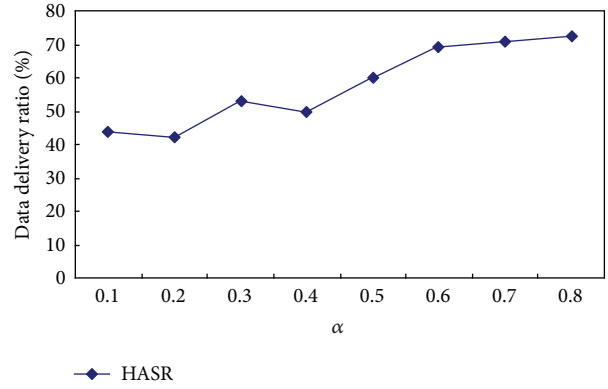


Figure 2: Impact of weight factor.

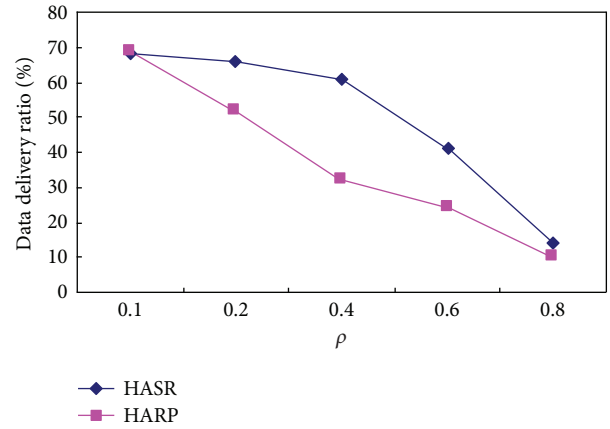


Figure 3: Impact of the proportion of selfish nodes.

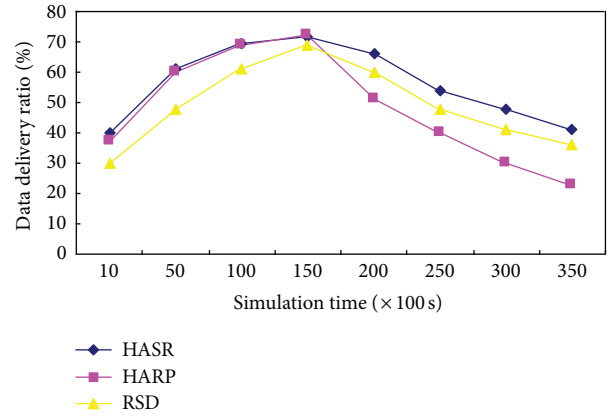


Figure 4: Performance of the three schemes.

The following simulations measure the data delivery ratio of HASR, HARP, and RSD. As shown in Figure 4, the data delivery ratio of the three schemes increases after simulations begin, but it decreases instead at a specific time. This is because the energy of the node is enough at the beginning of the simulation, the node transmits messages unselfishly. When the energy of the node is lower than the threshold (5 J), the node in HARP does not forward any messages in order

to save its energy, so that the data delivery ratio decreases quickly. For HASR and RSD, the node would deliver messages selfishly. HASR forwards messages based on the contribution index and routing decision of RSD is decided by a reputation-based scheme. That is to say, the selfish nodes still relay messages for some other nodes. So, the data delivery ratio of the two schemes decrease slower than that of HARP.

6. Conclusions and Future Works

In this paper, we propose a hot-area-based selfish routing (HASR) protocol for mobile social network. In HASR, routing decision is made by the active degree and the contribution index of the node. When the energy of the node is enough, data forwarding is determined only by the active degree of the node. When the energy of the node is consumed too much, the node behaves selfishly and routing decision is based on the contribution index. Simulation results show that HASR performs better than other schemes when nodes behave selfishly.

In the future works, we will study the selfishness of the node based on the social relationship between nodes. A node only relays the message from nodes with strong relationships among them and is not willing to transfer the message for strangers.

Acknowledgments

This work is supported by National Science Foundation of China under Grants nos. 60903158, 61003229, 61103226, 61103227, 61170256, and 61173172 and the Fundamental Research Funds for the Central Universities under Grants nos. ZYGX2010J074, ZYGX2011J060, and ZYGX2011J073.

References

[1] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *Proceedings of ACM SIGCOMM Workshops: Conference on Computer Communications*, pp. 244–251, ACM, Philadelphia, Pa, USA, August 2005.

[2] K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets," in *Proceedings of ACM SIGCOMM Conference on Computer Communications*, pp. 27–34, August 2003.

[3] N. Eagle and A. Pentland, "Reality mining: sensing complex social systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255–268, 2006.

[4] UCSD, "Wireless topology discovery project," 2007, http://sysnet.ucsd.edu/wtd/wtd.html.

[5] C. Diot, M. Martin, and N. Erik, "Haggle project," 2004, http://www.haggleproject.org.

[6] P. Hui and J. Crowcroft, "How small labels create big improvements," in *Proceeding of the 5th Annual IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2007*, pp. 65–70, March 2007.

[7] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE rap: social-based forwarding in delay tolerant networks," in *Proceeding of The 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '08)*, pp. 241–250, May 2008.

[8] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant MANETs," in *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '07)*, pp. 32–40, September 2007.

[9] A. Mtibaa, M. May, C. Diot, and M. Ammar, "PeopleRank: social opportunistic forwarding," in *Proceeding of IEEE 29th conference on Information communications (INFOCOM '10)*, pp. 111–115, San Diego, Calif, USA, March 2010.

[10] J. Liu, H. Gong, and J. Zeng, "Preferance location based routing in delay tolerant netork," *Journal of Digital Content Technology and its Applications*, vol. 5, no. 12, pp. 468–474, 2011.

[11] F. Li and J. Wu, "LocalCom: a community-based epidemic forwarding scheme in disruption-tolerant networks," in *Proceedings of the 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '09)*, June 2009.

[12] K. Jahanbakhsh, G. C. Shoja, and V. King, "Social-greedy: a socially-based greedy routing algorithm for delay tolerant networks," in *Proceedings of the 2nd International Workshop on Mobile Opportunistic Networking (MobiOpp '10)*, pp. 159–162, February 2010.

[13] C. Liu and J. Wu, "Routing in a cyclic mobispace," in *Proceeding of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '08)*, pp. 351–360, May 2008.

[14] J. Liu, M. Liu, and H. Gong, "Expected shortest path routing for social-oriented intermittently connected mobile network," *Journal of Convergence Information Technology*, vol. 7, no. 1, pp. 94–101, 2012.

[15] W. J. Hsu, D. Dutta, and A. Helmy, "CSI: a paradigm for behavior-oriented profile-cast services in mobile networks," *Ad Hoc Networks*, vol. 6, no. 4, pp. 13–24, 2011.

[16] Q. Zheng, X. Hong, and J. Liu, "An agenda based mobility model," in *Proceedings of the 39th Annual Simulation Symposium*, pp. 188–195, April 2006.

[17] H. Gong, L. Yu, X. Wang, and C. Song, "Hot area based routing protocol for delay tolerant mobile sensor network," *Journal of Convergence Information Technology*, vol. 8, 2012.

[18] M. K. Denko, "A reputation-based service differentiation scheme for mobile ad hoc networks," in *Proceedings of IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob '05)*, pp. 197–204, August 2005.

[19] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 921208, 9 pages
doi:10.1155/2012/921208

Research Article

Jam Eyes: A Traffic Jam Awareness and Observation System Using Mobile Phones

Xing Zhang, Haigang Gong, Zongyi Xu, Jinchuan Tang, and Bang Liu

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

Correspondence should be addressed to Haigang Gong, hggong@uestc.edu.cn

Received 31 October 2012; Accepted 3 December 2012

Academic Editor: Ming Liu

Traffic jam is a very common and very annoying thing in urban traffic. The most annoying part in traffic jams is not that you have to wait for a long time but that you do not even know how long you have to wait and what causes the traffic jam. However, the pain of being trapped in traffic jams seems to be neglected by existing research works; they put their focuses on either mathematical modeling or optimal routing for those not trapped in traffic jams. In this paper, we propose a traffic jam awareness and observation system using mobile phones. It can tell a driver how many vehicles ahead are trapped in traffic jam and how much time the driver would probably wait. Moreover, it can provide real-time video streams from the head vehicles of the traffic queue so that the driver can see what causes the traffic jam and the progress of handling the traffic jam. The system is environment independen; it can even work when the traffic jam happens in a tunnel. Experiments show that our system can find the head vehicles of the traffic queue and give the queue length accurately, and the video streams coming from the head vehicles reflect the actual situation of the traffic jam basically.

1. Introduction

Traffic jam is already a daily routine of modern urban traffic. The sources of traffic jam can be categorized into three ways: (1) a temporary obstruction, (2) a permanent capacity constraint in the network itself, and (3) a stochastic fluctuation in the demand within a particular sector of the network [1]. Obviously, the second way is the fundamental reason why traffic jam happens so frequently. Researchers have been trying their best to reduce the frequency of traffic jam; however, their works are basically a kind of optimization, as long as the network capacity is far from handling the actual increasing traffic flows, traffic jams will be inevitable and be getting worse. Now that traffic jam is inevitable, we should at least pay some attention to relieving sufferings of people from trapping in traffic jams.

Almost everyone living in the city has experienced traffic jam; the most annoying thing in traffic jam is not that people have to wait for a long time but that people even do not know how long they have to wait. When people are trapped in a traffic jam, unless they are the head of traffic queue, they hardly know what causes the traffic jam, how long is the traffic queue, and how is the progress of handling the traffic jam. In psychology, lines of evidence shows that people have strong fear of unknown [2]. Although the mentioned information cannot handle the traffic jam, it can handle the fear of people trapped in traffic jam.

Traffic density is a way to distinguish traffic jam from free flow. Traffic jam will cause a much higher traffic density than the density in free flow, and if we can obtain the locations of all vehicles in an area and depict a density map, traffic jam can be recognized. However, this method will cause big communication and computation cost; every vehicle will have to broadcast its location constantly and analyze the gathered locations to judge whether there is a traffic jam. Moreover, existing localization methods including GPS and WiFi Fingerprint [3] are environment dependent, the satellite signal and WiFi signal are easily blocked by buildings, trees, and so forth, and if traffic jam happens in a tunnel, GPS and wifi will fail to work. In fact, traffic density is a kind of relationship among each vehicle, and relative locations are enough to represent the relationship.

When a vehicle is trapped in a traffic jam, it has no speed or drive in a very low speed for at least a period of time, also it has neighbor vehicles surrounding it, and the neighbor vehicle also has no speed or low speed. The observations can help a vehicle to aware whether it is trapped in a traffic jam. The problem is how to distinguish parking from trapping in a traffic jam, because when a vehicle is parking at somewhere, it has neighbors surrounding it and it has no speed. Actually, we cannot distinguish the two cases, but with the help of a reasonable assumption, we can guarantee that the parking case will not interfere the judgment of traffic jam. In this research, we use mobile phones to implement our system; mobile phones have various sensors which can provide enough information of vehicle conditions, and they have networking ability to construct networks and communicate with each other which is not available on modern vehicles (although vehicular ad hoc network (VANET) has been researched for years, it has not equipped on modern vehicles). The assumption is that mobile phone is kind of a belonging to people; drivers will carry their mobile phones with them once they parked and left their vehicles. When a vehicle parks at a place, although it will sense a low speed of itself, it cannot sense any neighbors surround it because there are no surrounding mobile phones available for traffic jam judgment, and the vehicle therefore will not consider itself as trapping in a traffic jam.

Although vehicles are able to aware whether they are trapped in a traffic jam, they do not know their surrounding vehicles' relative locations: whether a surrounding vehicle is before or behind itself. Because the detecting signal is omitted omnidirectionally, a vehicle can know whether there is any surrounding vehicle but it cannot know the relative location of the surrounding vehicle. Without the information of relative location, a vehicle will not be able to count how many vehicles trapped in the traffic jam are ahead and estimate how long it has to wait. Fortunately, the time sequence of traffic jam can be mapped to the location sequence; if vehicle A joins a traffic queue earlier than vehicle B, vehicle A is ahead of vehicle B. If each vehicle records a time stamp when it is aware that it is trapping in a traffic jam, we can get the relative locations of each vehicle according to the time sequence. Therefore, the head vehicles of traffic queue are easily identified. If the head vehicles are willing to share their cameras, the drivers in the behind will be able to see what causes the traffic jam and how is the progress of handling the traffic jam.

The rest of the paper is organized as follows: Section 2 explains the system in detail; Section 3 shows the evaluation results. Related works and conclusions will be given in Sections 4 and 5.

2. System Overview

In this research, we do not need any extra hardware; the system is implemented on carry-on mobile phones. As mentioned before, to be aware of the traffic-jam state for a vehicle, two conditions must be satisfied: (1) the vehicle drives in no speed or low speed for at least a period of time; (2) the vehicle has neighbors which are also trapping in the traffic jam. We can use a triple to describe the traffic jam state: $\text{Jam}_i\,(Neighbors, Speed, Elapse)$, i is vehicle's ID, $Neighbors$ is the set of neighbor vehicles' IDs, $Speed$ is the driving speed of vehicle, and $Elapse$ is the elapse time of staying in the $Speed$. As (1) shows

$$\text{Jam}_i = \begin{cases} \text{level1}: & 0 \le Speed \le 20\,\text{km/h}, \\ \text{level2}: & \text{level1 and } Elapse \ge 5\ \text{min}, \\ \text{level3}: & \left(\text{level2 and } \exists j \in Neighbors \right. \\ & \left. \text{that Jam}_j = \text{level2 or level3}\right) \text{ or} \\ : & \left(\text{level1 and } \exists j \in Neighbors \right. \\ & \left. \text{that Jam}_j = \text{level3}\right), \\ \text{normal}: & \text{otherwise}, \end{cases} \tag{1}$$

there are four different states of Jam_i: level1 is set when vehicle i drives in a speed which is lower than 20 km/h, and it has some possibility that vehicle i is trapping in a traffic jam; level2 is set when vehicle i stays in level1 for more than 5 minutes, and there is a big possibility that vehicle i is trapping in a traffic jam; however, it has to check whether there is any neighbor which is also trapping in the traffic jam, otherwise it is hard to distinguish it from mechanical failure; level3 is set when vehicle i is already in level2 and it at least has one neighbor which is also in level2, or if vehicle i is in level1 and it has at least one neighbor which is in level3. If vehicle i is in level3, it is trapping in a traffic jam. Therefore, if vehicle i is in level1 and it has at least one neighbor in level3, traffic jam is the most possible reason which slows it down, and Jam_i is set to level3 directly.

Once Jam_i is set to level1, the system will record the time t_i, and t_i will be used to judge the position of vehicle i in the traffic queue. Vehicle i or the driver can initiate a request for acquiring information of vehicles in which $t_i > t_j$, $j \ne i$, through mobile ad hoc network (MANET) constructed by mobile phones. When the information is returned, the head vehicle of traffic queue is the one that has the smallest time stamp; the length of traffic queue is the number of qualified vehicles; video stream will send from the head vehicle through multihops in MANET.

2.1. Traffic Jam Awareness. The triple $\text{Jam}_i\,(Neighbors, Speed, Elapse)$ indicates that the system needs three dimensional information. $Elapse$ simply uses timer, and $Speed$ and $Neighbors$ use both sensors and wireless signals.

2.1.1. Speed. The readings of sensors are three dimensional, and they all obey the coordinate system of mobile phones. Take android system for example; if the mobile phone is placed as Figure 1 shows, the x axis is horizontal and points to the right, the y axis is vertical and points up, and the z axis points towards the outside of the front face of the screen [5].

Accelerometer gives three dimensional acceleration values, we use A_z to represent acceleration at z axis, A_x to

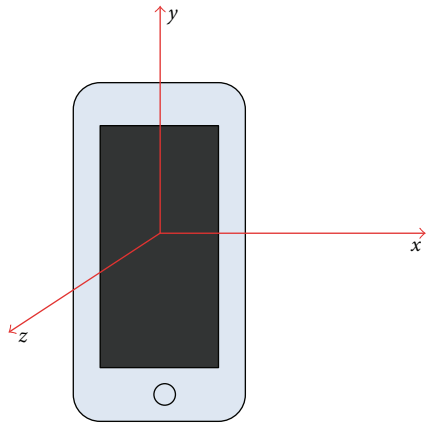


Figure 1: Coordinate system of mobile phones.

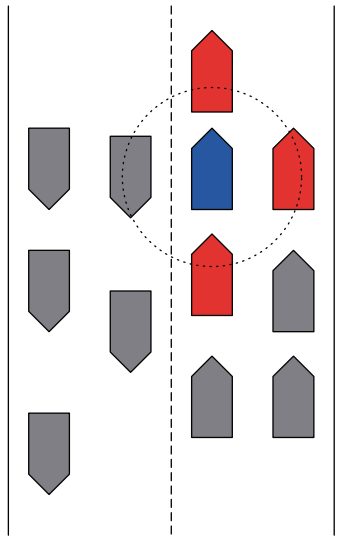


Figure 2: Neighbors detection. Blue block is a vehicle which initiates a detection of neighbors, red blocks are qualified neighbors of the blue block, and grey blocks are not the neighbors of the blue block.

represent acceleration at x axis, A_y to represent acceleration at y axis. If we fix the mobile phone as Figure 5(a) shows, A_z and A_y will be used to calculate vehicle speed as (2) shows, and A_x will never be used unless vehicles can fly. The sample rate of accelerometer we set is 20 ms; therefore $\Delta t = 20$ ms:

$$a_i = \sqrt{A_y^2(i) + A_z^2(i)}, \tag{2}$$

$$v_{i+1} = v_i + a_{i+1}\Delta t. \tag{3}$$

2.1.2. Neighbors. Normally, vehicles are closed to each other in traffic jam. In this research, direct neighbors will be detected as *Neighbors*. As Figure 2 shows, the blue block is the vehicle detecting its neighbors, the red blocks are the neighbors of the blue block, and the grey blocks are vehicles not qualified as neighbors of the blue block. We can see that there is a grey block which is in the neighbor range of the blue block, the reason why this block is not the neighbor of the blue block is that it is heading a different direction from the blue block, and traffic jam happens among vehicles in the same direction. Therefore, distance and direction are two factors for detecting neighbors.

Mobile phones have orientation sensor which returns three-dimensional values, Azimuth (degrees of rotation around the z axis), Pitch (degrees of rotation around the x axis), and Roll (degrees of rotation around the y axis) [6]. If the mobile phone is fixed as Figure 5(a) shows, then Pitch is the degrees we need. Suppose a vehicle i is detecting its neighbors, vehicle j is closed to vehicle i, and Pitch_i is the Pitch value of vehicle i, then vehicles i and j have the same direction when $|\text{Pitch}_i - \text{Pitch}_j| \leq 30°$.

In this research, each vehicle needs to find its direct neighbors, and direct neighbors means they are very close to the detecting vehicle that there is no other vehicles between direct neighbors and detecting vehicle. If a neighbor is not a direct neighbor, for example, it is 100 meters away from the detecting vehicle, even the neighbor does be trapped in a traffic jam, it cannot be used as a reference to indicate a traffic jam because it is hard to tell whether the traffic jam is the same as the traffic jam the detecting vehicle is trapped in. Modern mobile phones are equipped with WiFi network cards, and the network cards can be used to construct MANET. It is easy to communicate in MANET, but it's hard to detect direct neighbors. If we use distance to represent the relationship of the direct neighbors, as [7] says the average car length is around 4 meters, the average distance between two vehicles is around 2 meters, then 6 meters will be a proper distance. RSSI-based localization [8] is a good way to measure distance, but it is difficult to acquire each node's RSSI value in a MANET with existing tools. In this research, we propose a new and simple detection method called ECHO. Common WiFi signal with 20 dbm initial power can cover an area with at least 100 meters radius, and the area is too large for the neighbor detection. If the initial transmit power is reduced to a much lower level that the signal can only reach up to 6 meters distance, when the detecting vehicle sends one detecting packet, only vehicles within 6 meters from the detecting vehicle can receive the packet and echo back their IDs. *Neighbors* will contain the vehicles which are within 6 meters from detecting vehicle and they have the same direction as the detecting vehicle.

Now that the three parameters of $\text{Jam}_i(Neighbors, Speed, Elapse)$ are all collected, vehicles are able to judge whether they are trapped in a traffic jam according to (1). Figure 3 shows the states transition of (1). When a vehicle's state changes to level1, it will record a time stamp, and the time stamp will be used to decide the relative location of each vehicle in the same traffic jam.

2.2. Basic Information of Traffic Jam. It is easy to calculate the length of traffic queue and find the head of the traffic queue as long as we have all the time stamps of vehicles in level3. To obtain the time stamps, the detecting vehicle needs to broadcast a detection packet, and due to the limitation of

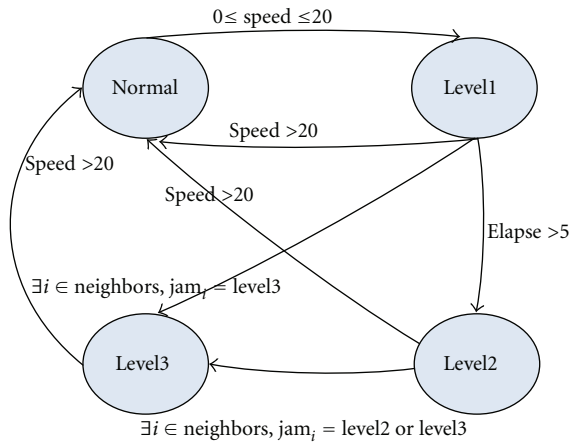


Figure 3: State transition of traffic jam. Vehicle starts in normal state; if it drives in a speed lower that 20 km/h, it will change to level1; if level1 stays for at least 5 minutes, it will change to level2; if level1 has one neighbor in level3, it will change to level3; if level2 has one neighbor in level2 or level3, it will change to level3. All levels will change to normal when speed is higher than 20 km/h. We consider a vehicle is trapping in a traffic jam when it is in level3.

```
while 1 do
  if receive one detection packet then
    if already has received the detection packet with same
    source vehicle ID in one minute then
      continue;
    end if
    if Jam_i = level3 then
      reply a packet with its vehicle ID and time stamp in
      it;
      broadcast the received detection packet;
    end if
  end if
end while
```

Algorithm 1: Reply and relay actions for received detection packet on vehicle i.

communication range of WiFi signal, multihop is needed. Algorithm 1 shows the actions of a vehicle when it receives a detection packet.

When detection finished, the system will sort all vehicle IDs according to time stamps. The vehicle with smallest time stamp is the head of the traffic queue, and we use V_{head} to represent it; the number of vehicles which have smaller time stamps than the detecting vehicle is the length of the traffic queue, and we use Q_i to represent the number of vehicles ahead of vehicle i in the traffic jam.

To estimate the time vehicle has to wait, we have to monitor V_{head}'s movement. When V_{head} changes its state to *normal*, and the time interval between the latest zero speed and state change to *normal* is recorded as T_{head}, the time between a vehicle moves and its following vehicle moves

```
while 1 do
  if it is V_head then
    broadcast the T_head;
  else
    if received the broadcast packet then
      if hasn't received the broadcast packet within a
      minute then
        get the T_head from the packet;
        broadcast the packet;
      end if
    end if
  end if
end while
```

Algorithm 2: T_{head} broadcast.

is Δt, the time vehicle i will wait for will be calculated as $Q_i \times \Delta t - T_{\text{head}}$, then the calculation result could be negative because when V_{head} changes its state to *normal*, some following vehicles already moves. If V_{head} has not changed its state to *normal*, T_{head} will be ∞. The problem is how to get the T_{head} from V_{head}. One way is that each vehicle periodically requests the information from V_{head}, and the other way is that V_{head} broadcasts the T_{head} which it changes its state to *normal*. Obviously, the second way costs less, but it requires a vehicle that can be aware whether it is the V_{head}. With the help of the information of *Neighbors*, a vehicle that can be aware it is the V_{head} through comparing its time stamp with the time stamps of *Neighbors*; if its time stamp is the smallest, the vehicle is the V_{head}. Once a vehicle is aware that it is the V_{head}, it will broadcast its T_{head} using Algorithm 2.

2.3. Traffic Jam Observation. The basic information can tell drivers how long the traffic queue is and how long they probably have to wait; however, there are still a lot of useful information which sensors cannot sense and process; people trapped in the traffic jam usually want to know what causes the traffic jam and how is the progress of handling the traffic jam. It is better to let people see the traffic scene from the position of V_{head} and make their own judgment. However, camera is a privacy-sensitive device on mobile phone; people usually would not be willing to share their cameras, and therefore incentive schemes are needed.

Our system is crowdsourcing based, and there are some incentive researches on crowdsourcing systems, like [9–11]. Moreover, there are some researches on leveraging phone cameras to capture traffic lights [1], some researches on taking advantage of phone sensors to monitor driving behaviors [4, 12, 13], and some applications like driving navigation; if our system can be combined with these works, it is possible to make people wish to share their cameras.

Figure 4 shows two figures from [1, 4]; we can see the way the mobile phone fixed is convenient for cameras to capture outside scene. Therefore we use the same way to fix the mobile phone and, as Figure 5(b) shows, the camera

(a) (b)

Figure 4: Existing mobile phone placement. (a) is from [1], and (b) is from [4].

(a) (b)

Figure 5: Mobile phone placement of our system. (a) is the placement of the mobile phone, and (b) is the camera view.

captures wide enough scene which can give remote drivers the ability to know what causes the traffic jam.

Video stream is large, if V_{head} sends each vehicle the video stream using a separate link, such as http, it will cause high overhead on V_{head} and network congestion when there are many vehicle requests for the stream. Multicast is a good way to solve the problem. In an ad hoc network, like MANET, every node will act like a router, once a node requests to receive a multicast channel, one other node in the range of this node will act as a multicast router to deliver the multicast packets to the node. As long as we can find a route from the node of request to V_{head}, and each node on the route will act as a multicast router, the node of request will receive the multicast packets. Algorithm 3 shows how to achieve a video stream multicast in MANET. In this algorithm, we assume that the relative location of each vehicle is stable and they are closed to each other; therefore the multi hop packet forward process is simply to choose the vehicle with the smallest time stamp.

3. Implementation and Evaluation

We rent ten automobiles to simulate a traffic jam in an open space of campus, each automobile is equipped with an android mobile phone fixed as Figure 5 shows. We set up three scenes of traffic jam as Figure 6 shows: (a) a single lane with an obstacle in the middle, and automobile has to slow down and drive around the obstacle; (b) two lanes change to one lane, and it causes automobiles distributed on two lanes to compete for one lane; (c) automobiles are stopped due to traffic control. We test the accuracy of V_{head} judgment, the accuracy of waiting time estimation, the accuracy of traffic queue length, and the frames dropped rate of video stream multicast in the three scenes, and we repeat the tests for 20 times to get the statistics.

Before scene testing, we test the mapping relationship between RSSI and distance, we find that 2 mw is a good initial power to detect direct neighbors.

3.1. Scene 1. In scene 1, ten automobiles cannot cause more than 5-minute traffic jam; therefore we parked the ten automobiles in front of the obstacle for 5 minute, and the first automobile starts to drive around the obstacle in a speed of 5 km/h; once the first automobile moves, the second automobile will move in a delay time around one to two seconds due to human response time and driving operation time. The first automobile will spend 3 seconds driving around the obstacle and 1 second speeding up to 20 km/h. The detection of V_{head} is doing in a period time of 0.5 second.

Figure 7 shows that the accuracy of V_{head} is averagely 63%; the reason is that when a vehicle starts to speed up after driving around the obstacle, the vehicle is no longer the V_{head}, but our system will wait until its speed reaches

```
while 1 do
  if received a request packet for video stream then
    if it is V_head then
      if it has not initiate the video stream multicast then
        initiate the video stream multicast;
      end if
      continue;
    end if
    record the vehicle ID of the packet sender as
    a multicast client;
    collect the time stamp and ID information from its
    neighbors;
    choose a neighbor with the smallest time stamp and
    send it the request;
    continue;
  end if
  if receive a reply of request then
    collect the time stamp and ID information from its
    neighbors;
    if the request vehicle in among its neighbor then
      send the reply to the request vehicle;
    else
      choose a neighbor with the largest time stamp and
      send it the reply;
    end if
    continue;
  end if
  if received a video stream packet then
    if it has one or more multicast client records then
      deliver the packet;
    end if
    continue;
  end if
  if receive a quit packet of multicast then
    if it has one or more multicast client records and the
    quitter is one of these client then
      delete the record accordingly;
    end if
    continue;
  end if
end while
```

Algorithm 3: Video stream multicast.

20 km/h; the accuracy of waiting time estimation is within 1 second averagely, −1 means the timer says there is still one second to wait while the vehicle is already moving, +1 means the timer already changes to zero while the vehicle moves after one second; the misjudgment of traffic queue length is 37% according to the accuracy of V_{head}. Frames dropped rate is somehow a little high and unstable, and the reason is probably that the V_{head} is moving all the time and it may affect the signal propagation.

3.2. Scene 2. Scene 2 is more complicated than scene 1 because vehicles on difference lanes will compete for the only one lane; vehicles may change their lanes to occupy a better position, and it will cause that a vehicle with a bigger time stamp is in front of a vehicle with a smaller time stamp; therefore the accuracy of V_{head} judgment may be affected. Figure 8 shows the accuracy of V_{head} judgement is dropped to 48% compared to scene 1, and the view from wrong V_{head} somehow does not reflect the progress of handling the traffic jam; the accuracy of waiting time estimation is bigger than scene 1 because the the competition will interrupt the moves of vehicle and prolong the Δt; the misjudgment of traffic queues length is 52% according to the accuracy of V_{head}; frames dropped rate is still high due to the same reason.

3.3. Scene 3. Scene 3 is another simple case, because vehicles will not compete for better positions and vehicles are not moving until the traffic control released; therefore the frames dropped rate is relatively lower than the other two cases. Figure 9 shows the result, and the accuracy of V_{head} judgment is 100%, because no vehicle moves in this scene; therefore the time estimation and queue length have perfect results which are not necessary to present here.

4. Related Works

Research works on traffic jam mainly focus on analyzing traffic jam through mathematical modeling. Reference [14] shows a new perspective of the origin of traffic jam, and it presents experimental evidence that the emergence of a traffic jam is a collective phenomenon like "dynamical" phase transitions and pattern formation in a nonequilibrium system. Reference [15] gives modeling and simulation of traffic jam in a macroscopic level, and [16, 17] give modeling of traffic jam in a microscopic level; they both give some interesting perspectives. Mathematic modeling can reveal some facts of traffic jam and help reduce the frequency of traffic jam, but we believe that the principal contradiction of traffic jam is the capacity of network that is unable to handle the increasing popularity of vehicles; therefore math can do some optimizations to traffic jam, it can never solve the problem. Now that traffic jam is unbeatable, we should put more energy on relieving suffering of people trapped in the traffic jam.

Sensors on vehicles and mobile phones are able to reflect the behaviors of vehicles: how fast they are, which direction they are heading to. Some researches leveraged mobile and vehicular sensors to monitor and correct drivers' behaviors to enhance driving safety [4, 18], and some works capture the surrounding information to enhance driving comfort [1]. However, they all neglect the pain of trapping in the traffic jam.

5. Conclusions

In this paper, we propose a novel traffic jam awareness and observation system; it gives some basic information of traffic jam like the length of traffic queue, the time waiting in the traffic jam, and so forth, and also it provides video streams from the head of traffic queue to present the cause of the traffic jam and the progress of handling the traffic jam to people who are not in the front of the traffic queue. The system is environment independent, and it need not

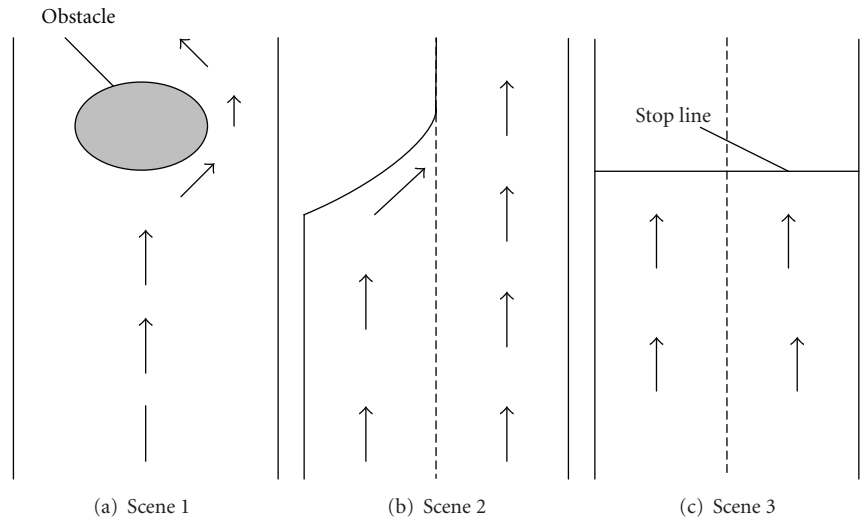


Figure 6: Three scenes of traffic jam.

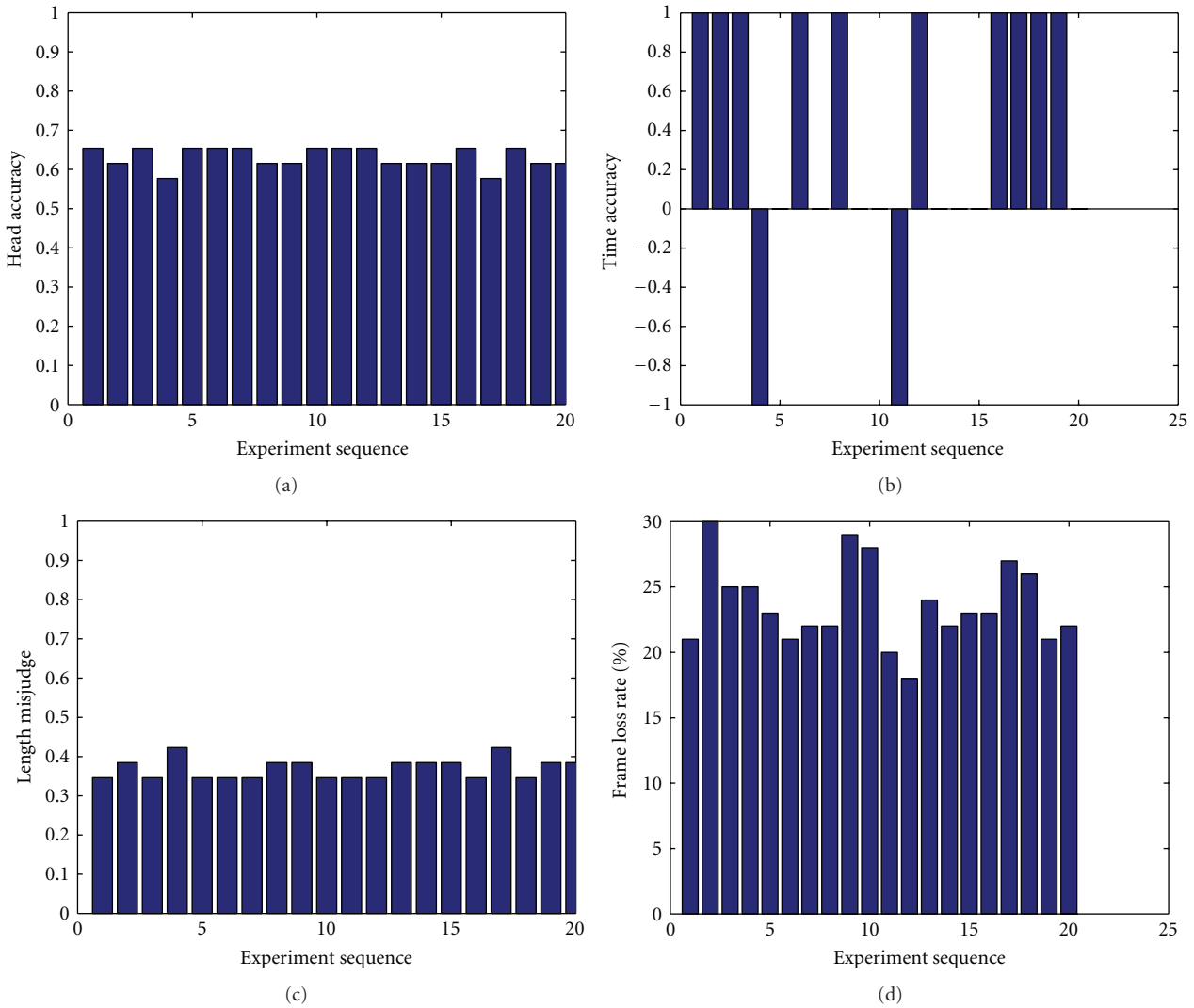


Figure 7: Experiment results of scene 1. (a) is the accuracy of V_{head} judgment, (b) is the accuracy of time estimation, (c) is the queue length misjudgment rate, and (d) is the frame loss rate.

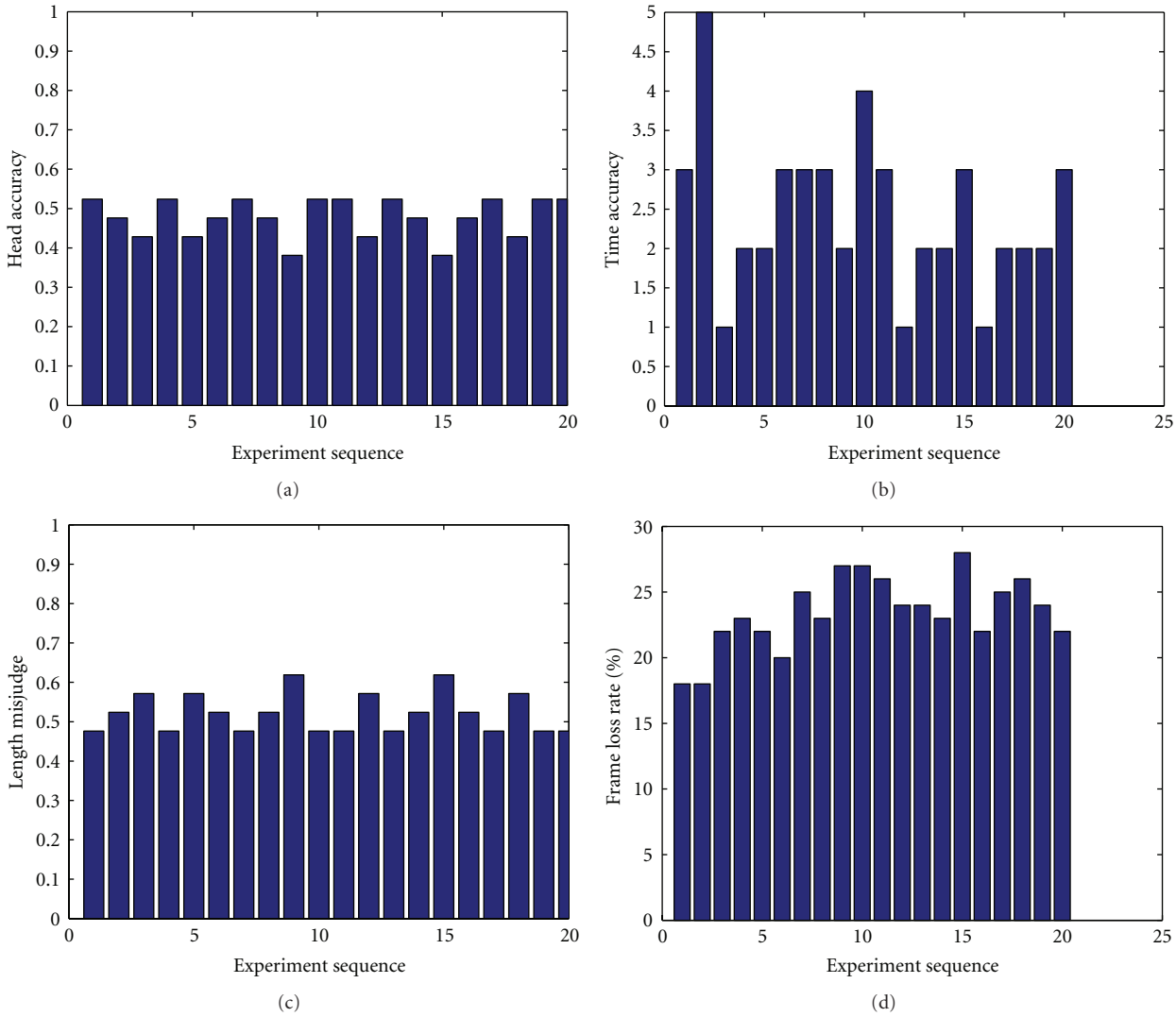


Figure 8: Experiment results of scene 2. (a) is the accuracy of V_{head} judgment, (b) is the accuracy of time estimation, (c) is the queue length misjudgment rate, and (d) is the frame loss rate.

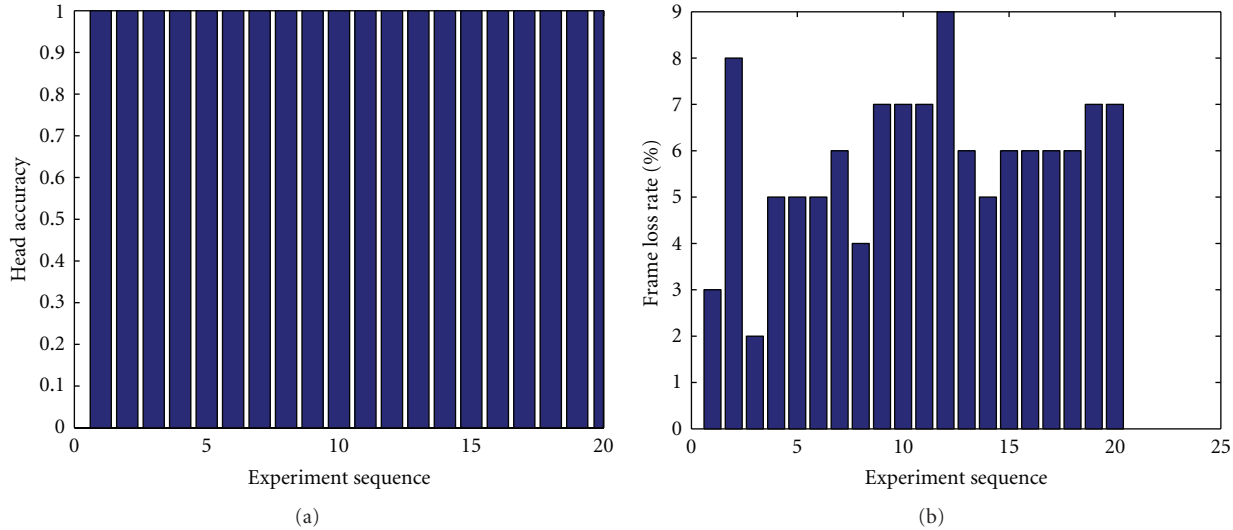


Figure 9: Experiment results of scene 3. (a) is the accuracy of V_{head} judgment, and (b) is the frame loss rate.

GPS device and any extra hardware, but the mobile phones carried by drivers are good enough.

References

[1] E. Koukoumidis, L. S. Peh, and M. R. Martonosi, "SignalGuru: leveraging mobile phones for collaborative traffic signal schedule advisory," in *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys '11)*, pp. 127–140, ACM, Bethesda, Md, USA, July 2011.

[2] H. H. Cao, B. Han, D. Hirshleifer, and H. H. Zhang, "Fear of the unknown: familiarity and economic decisions," *Review of Finance*, vol. 15, no. 1, pp. 173–206, 2011.

[3] Y. C. Cheng, Y. Chawathe, A. Lamarca, and J. Krumm, "Accuracy characterization for metropolitan-scale Wi-Fi localization," in *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys '05)*, pp. 233–245, ACM, June 2005.

[4] D. A. Johnson and M. M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in *Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems*, pp. 1609–1615, Washington, DC, USA, 2011.

[5] Android sensor coordinate system, http://developer.android.com/reference/android/hardware/SensorEvent.htmll.

[6] Orientate sensor, http://developer.android.com/guide/topics/sensors/sensors_position.html.

[7] "Vehicles keep inching up and putting on pounds," http://usatoday30.usatoday.com/money/autos/2007-07-15-big-cars_N.htm.

[8] J. Arias, A. Zuloaga, J. Lázaro, J. Andreu, and A. Astarloa, "Malguki: an RSSI based ad hoc location algorithm," *Microprocessors and Microsystems*, vol. 28, no. 8, pp. 403–409, 2004.

[9] P. Trompette, V. Chanal, and C. Pelissier, "Crowdsourcing as a way to access external knowledge for innovation," in *Proceedings of the 24th EGOS Colloquium*, Amsterdam, The Netherlands, July 2008.

[10] Y. Zhang and M. van der Schaar, "Reputation-based incentive protocols in crowdsourcing applications," in *Proceedings of the IEEE INFOCOM*, pp. 2140–2148, IEEE, Orlando, Fla, USA, March 2012.

[11] D. Yang, G. Xue, and X. Fang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking (MobiCom '12)*, pp. 173–184, Istanbul, Turkey, August 2012.

[12] G. A. ten Holt, M. J. Reinders, and E. A. Hendriks, "Multi-dimensional dynamic time warping for gesture recognition," in *Proceedings of the 13th Annual Conference of the Advanced School for Computing and Imaging*, June 2007.

[13] R. Muscillo, S. Conforto, M. Schmid, P. Caselli, and T. D'Alessio, "Classification of motor activities through derivative dynamic time warping applied on accelerometer data," in *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4930–4933, 2007.

[14] Y. Sugiyamal, M. Fukui, M. Kikuchi et al., "Traffic jams without bottlenecks-experimental evidence for the physical mechanism of the formation of a jam," *New Journal of Physics*, vol. 10, no. 3, Article ID 033001, 2008.

[15] P. Degond and M. Delitala, "Modelling and simulation of vehicular traffic jam formation," *Kinetic and Related Models*, vol. 1, no. 2, pp. 279–293, 2008.

[16] G. Orosz, R. E. Wilson, R. Szalai, and G. Stépán, "Exciting traffic jams: nonlinear phenomena behind traffic jam formation on highways," *Physical Review E*, vol. 80, no. 4, Article ID 046205, 2009.

[17] J. Long, Z. Gao, X. Zhao, A. Lian, and P. Orenstein, "Urban traffic jam simulation based on the cell transmission model," *Networks and Spatial Economics*, vol. 11, no. 1, pp. 43–64, 2011.

[18] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan, "Mobile phone based drunk driving detection," in *Proceedings of the 4th International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health '10)*, March 2010.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 151795, 11 pages
doi:10.1155/2012/151795

Research Article

Efficient Data Dissemination in Urban VANETs: Parked Vehicles Are Natural Infrastructures

Hui Zhao and Jinqi Zhu

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

Correspondence should be addressed to Hui Zhao, jenniferzhao09@gmail.com

Received 9 November 2012; Accepted 3 December 2012

Academic Editor: Ming Liu

Data dissemination is the fundamental operation in vehicular ad hoc networks (VANETs); for example, after an accident or congestion is detected by the corresponding sensors mounted on the vehicles, an alert message should be swiftly disseminated to the vehicles moving towards the affected areas. However, the unique characteristics of VANETs, such as high mobility of vehicle nodes, intermittent connectivity, and rapidly dynamic topology, make data dissemination over them extremely challenging. Motivated by the fact that there are large amounts of roadside parked vehicles in urban areas, this paper proposes a parking-based data dissemination scheme for VANETs. Data to be disseminated are buffered at the roadside parked vehicle, which continuously provides data dissemination services for the vehicles passing by. We analyze the challenging issues in achieving parking-based data dissemination and provide possible solution for each issue. Theoretical results illustrate the effectiveness of our approach, and simulation results based on a real city map and realistic traffic situations show that the proposed data dissemination paradigm achieves a higher delivery ratio with lower network load and reasonable delivery delay.

1. Introduction

Nowadays, to facilitate better road safety and comfort driving, more and more vehicles are equipped with wireless devices and different types of sensors. Consequently, large-scale vehicular networks are expected to be available in the near future. With its popularity, VANETs are envisioned to provide us with numerous useful applications. One typical application is intelligent transportation system; for example, after an accident or congestion is detected by the corresponding sensors mounted on the vehicles, an alert message would be swiftly disseminated to the vehicles moving towards the affected areas via vehicular communication. Taking advantage of this application, incoming vehicles will be informed in advance of these accidents/congestions and the drivers may take another route/appropriate actions. Other applications also include available parking spaces notification and commercial ads dissemination. Undoubtedly, these applications would improve our driving experience greatly.

The basic operation in the aforementioned applications is data dissemination. Unfortunately, VANETs are characterized by rapidly dynamic topology, intermittent connectivity, and high mobility of vehicle nodes, which make data dissemination over it a challenging issue. Most of the existing works take advantage of the inter-vehicle communication to achieve data dissemination [1–4]. The weakness of the inter-vehicle scheme is that data to be disseminated can hardly be kept within a target area in highly mobile environments. Towards solving the problem, two abiding geocast techniques [5] could be adopted. One is periodically broadcasting each data at the deployed server. Another is maintaining each data at selected moving vehicles within the target area. For the first approach, when tens of thousands of messages are routed over a long distance to the target areas, excessive transmissions and severe congestion are inevitable. For the second approach, continuous node selection and message handover are required due to the high mobility of vehicle nodes, which incurs great overheads.

In view of the insufficiency of inter-vehicle data dissemination, some researches put forward the infrastructure-based data dissemination. In [6], Zhao et al. propose to deploy roadside units to assist data dissemination. Data to be disseminated are stored temporarily at roadside units in the target area and broadcasted periodically to the vehicles passing by. This scheme is proved to be effective. However, the deployment of roadside units at the city scale also requires a large amount of investment.

In this paper, we propose a parking-based data dissemination scheme, which harnesses the free resource offered by roadside parking for data dissemination in urban areas. Our proposal is inspired by a real world urban parking report [7], which provides the parking statistics of two surveys in a central area of Montreal city in Canada. It investigated the 61,000 daily parking events in an area of 5,500 square kilometers. According to the report, street parking accounts for 69.2% of total parking, and the average duration of street parking lasts 6.64 hours. It generates many roadside vehicle nodes easy to communicate and enables them to support long-time communication. The basic idea of our parking-based data dissemination scheme is simple: if a vehicle often drives through extensive vehicles parked at roadside, why not let these parked vehicles support data dissemination as roadside infrastructure?

We organize the parked vehicles into different clusters, propose an effective routing scheme to distribute each data message to appropriate roadside parking, and adopt the pub/sub scheme to perform data dissemination. Moreover, we investigate our scheme through theoretic analysis, realistic survey, and simulation. The results prove that our scheme achieves a higher delivery ratio with lower network load and reasonable delivery delay.

The original contributions that we have made in the paper are highlighted as follows.

(i) We exploit the roadside parked vehicles to achieve data dissemination in urban VANETs. Our scheme aims at reducing the overhead brought by inter-vehicle scheme and avoiding the costs brought by constructing roadside infrastructure.

(ii) We tackle the main challenges in realizing parking-based data dissemination, for example, how to manage the roadside parked vehicles and how to route a data message to the targeted parking efficiently.

(iii) We evaluate our parking-based data dissemination scheme through theoretical analysis, realistic survey, and simulation. The numeric results show that our scheme is effective.

The remainder of this paper is structured as follows. Section 2 makes a brief overview of related work. Section 3 presents the system model. In Section 4, we explain our parking-based data dissemination scheme in detail. Section 5 proves the effectiveness of our scheme through theoretical analysis, while Section 6 evaluates our scheme through realistic survey and simulation. Finally, Section 7 summarizes the paper.

2. Related Work

Data dissemination over VANETs is extremely challenging due to the unique characteristics of VANETs. In the last decade, many research effects have been devoted to addressing the data dissemination issues in VANETs. Xu et al. propose an opportunistic dissemination (OD) scheme [8]. In this scheme, the data center periodically broadcasts some data, which will be received and stored by passing vehicles. Whenever two vehicles move into the transmission range of each other, they exchange data. This scheme does not rely on any infrastructure. However, the performance of the OD scheme is poor in areas with high vehicle density due to media access control (MAC) layer collisions. This can easily lead to severe congestion and significantly reduce the data delivery ratio. To mitigate the excessive transmissions and congestion, Korkmaz et al. [9] propose a link-layer broadcast protocol to help disseminate the data. The protocol relies on link-layer acknowledge mechanisms to improve the reliability of the multihop broadcast. However, in the case of network congestion, the link-layer solution is not enough. Furthermore, since many information sources may exist in a given urban area, the amount of broadcasted data from these sources can easily consume the limited bandwidth. In [1], Nekovee M proposes an improved Epidemic scheme, which takes advantage of the clustering characteristics of vehicle flow and broadcasts message at the edge of each cluster. This scheme reduces the communication overhead at some extent. In [2], the notification area is divided into several subareas, and message is disseminated based on each subarea, which effectively limits the broadcast range of each message. In [3], the authors put forward MDDV scheme, which exploits the vehicles called message holder to carry the message to the notification area and broadcast it in this area. In [4], Wu et al. propose a mobile distribution-aware data dissemination scheme MDA for VANETS. In MDA, the subscribers' distribution is predicted, and the forwarding of the notification token is controlled to achieve effective distribution of notification brokers (notification-token holder). Although [1–4] cut down the network overhead to some extent. The data to be disseminated can hardly be kept in the target area owing to the intermittent connectivity of VANETs.

Recently, many approaches have been proposed to realize persistent data availability in VANETs. A basic approach is the server approach in [5], in which the server periodically delivers the message to the destination region using a geocast routing protocol. The deficiency of this approach is that frequent broadcasting at the server would consume a large amount of bandwidth. An alternative approach is the Election approach in [5]. It stores the messages at elected mobile nodes inside the geocast destination region. Due to the high mobility of vehicle nodes, continuous node selection and message handover are required in this case.

To reduce the amount of data poured from the server, Zhao et al. [6] propose the idea of intersection buffering, in which the relay and broadcast station (IBer) is used to buffer data copies at the intersection. The IBer broadcasts each message periodically. As a result, the server does not have to frequently broadcast data to guarantee that each

vehicle receives the data. In [10], the authors also propose to use stationary roadside units to improve data dissemination performance. In [11], they further discuss the strategic placement of roadside units. Although the deployment of roadside units could improve the dissemination performance dramatically, the widely deployed roadside units will lead to great investments.

3. System Model

3.1. Assumptions. First, we assume that vehicles are equipped with various types of sensors, GPS, and preloaded electric maps, which are already popular in new cars and will be common in the future. Second, we assume that some vehicle users will share their devices during parking. This could be motivated by effective incentives, as indicated in [12, 13]. Finally, we assume that each data message is attached with the following two attributes: (1) target areas, which are the areas where the data is most likely to be interested, and (2) survival time, which indicates the survival time of the data. This assumption is based on the following observation: the disseminated data are often spatial or/and temporal sensitive; for example, for an accident notification message, it is most likely to be the interest of drivers moving towards the affected area, and this message will be invalid after the traffic accident is properly treated.

3.2. Scenario. As shown in Figure 1, the parked vehicles are widely distributed at the roadside in urban area. At a certain moment, a traffic accident happens in one road segment. Assume that the vehicles are equipped with accident detection sensors and the sensor output is monitored and processed by a microcontroller. After the microcontroller detects this traffic accident based on the input from the sensors, it would broadcast an emergency notification message. To lower the impact of this accident on the traffic condition, the emergency notification message should be forwarded to the vehicles moving towards the affected area, so that the drivers could choose to take another route. Similar applications also include parking statistics dissemination. In [14], it is reported that cruising for parking wastes 47,000 gallons of gasoline and produces 730 tons of CO_2 emissions per year in a small business district of Los Angeles. If drivers are provided with parking data dissemination services, the parking space searching costs would be greatly reduced. With the popularity of VANETs, more and more applications would be emerging in VANETs. While tens of thousands of data messages are flooded into the VANETs, an efficient data dissemination scheme is indispensable. Therefore, it is of great significance to develop highly efficient data dissemination scheme for urban VANETs.

In our parking-based data dissemination scheme, data to be disseminated are buffered at the roadside parked vehicle, which continuously provides services for the vehicles passing by. Overall, our parking-based scheme involves the following four components.

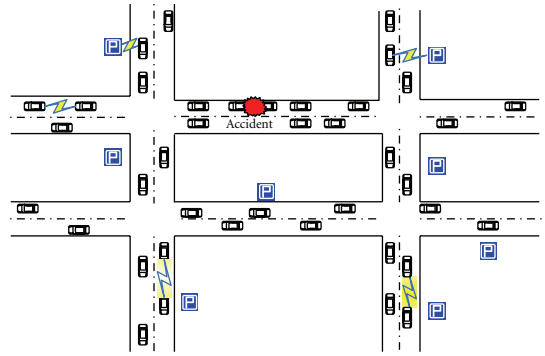


Figure 1: A sample scenario.

(1) Data source could be a computer with a wireless interface, a wireless access point, or an infostation [15].

(2) Data forwarders are the vehicles which help to forward a data item from the data source to the targeted parking clusters.

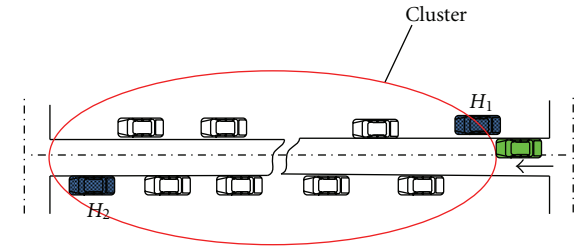
(3) Roadside parking cluster is composed of a group of vehicles which are parked along the same road segment and belong to the same partially connected network.

(4) End users are the vehicle users who have interests in a certain set of data messages while driving.

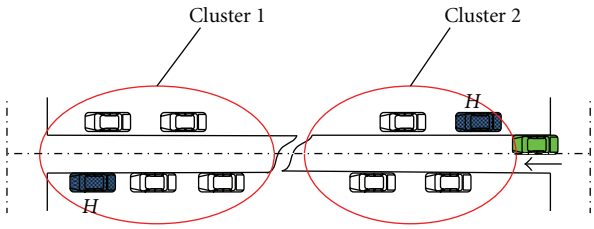
4. The Proposed Parking-Based Scheme

To facilitate data dissemination, we organize the roadside parked vehicles into clusters. Generally, our proposed parking-based data dissemination scheme is divided into two phases: data forwarding from the data source to appropriate parking clusters within the target area and data dissemination from the parking cluster to vehicles passing by.

4.1. Parking Cluster. A realistic survey [16] provides a quantitative understanding of roadside parking in cities, in which the on-street parking meters in the Ann Arbor city are continuously monitored during six midweek days. It shows that the parking time is 41.40 minutes in average, with a standard deviation of 27.17. The occupancy ratio, defined as occupied space-hour/available space-hour, averages 93.0% throughout one day. Even the occupancy ratio during off-peak time reaches almost 80%. Due to the high stability and utilization of roadside parking, clustering parked vehicles is feasible in urban areas. In our parking-based scheme, we group the vehicles which are parked along the same road segment and are mutually reachable into a cluster and take it as data buffering unit at street level. Considering the fact that vehicle mobility is strictly constrained by traffic rules and street layout, buffering each data at some clusters in the target area is enough. Therefore, we will first introduce how to elect data buffering units from the existing clusters and then give our cluster management scheme.



(a) One cluster



(b) Multicluster

Figure 2: Data buffering unit.

In some road segments, the parked vehicles form one cluster, as shown in Figure 2(a). In other road segments, the parked vehicles are isolated from each other and form different partially distributed groups, as shown in Figure 2(b). To determine whether it should act as data buffering unit, we let each cluster periodically report its distribution to other clusters along the same road (with the help of vehicles traveling across the road). After obtaining the distribution of other clusters along the same road, a cluster decides whether it would work as buffering unit according to following rule: if there is only one cluster along the road, this cluster is undoubtedly elected as data buffering unit; if there are two or more than two clusters along the road, the two clusters located at the two ends of the road are elected as data buffering units. After elected as data buffering unit, a cluster needs to be responsible for the cluster management, including head election and membership management.

In our scheme, we specify the following head selection mechanism. In a scenario in Figure 2(a), the two vehicles located at the two ends of the cluster are elected as cluster head. In a two-way road, the two cluster heads, respectively, provide services for the vehicles coming from the nearest intersection. In a scenario in Figure 2(b), the vehicle which locates at the end of the road segment is elected as cluster head in each cluster; this is also to ensure that a vehicle moving into the road could encounter the cluster head in a short time. After the cluster head is determined, the cluster members periodically report their position to the cluster head. Thus, the cluster head is able to manage all parked vehicles, act as local service access points, and perform the data dissemination operation. Considering the fact that the vehicle works as cluster head might leave at any time, we specify the following rule: while the cluster head is leaving (the engine is started), a new round of head selection is triggered, and the data to be disseminated as well as the cluster state are transferred from the old cluster head to the new one.

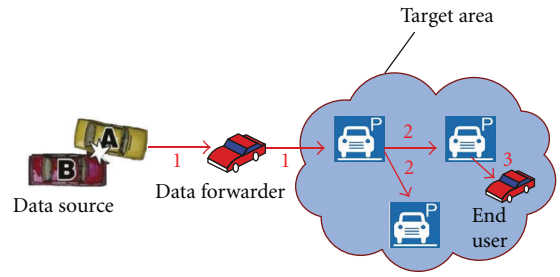


Figure 3: Data forwarding process.

4.2. Data Forwarding from Data Source to Roadside Parking. In our parking-based scheme, the parking clusters help to buffer the data messages in their target area and provide data dissemination service for the vehicles passing by. To realize this one-hop data dissemination, the data source should first distribute each data to the selected parking clusters within the target area. According to the strategy used, this process could be further divided into two phases: routing from data source to one parking cluster (step 1 in Figure 3) and routing from one parking cluster to other parking clusters (step 2 in Figure 3). We will describe them in detail in the following part.

4.2.1. Routing from Data Source to One Parking Cluster. While investigating the routing from the data source to one parking cluster in the target area, we first focus on the most common scenario, in which the location of the data source is out of the target area of the data message. In our scheme, apart from taking advantage of the mobile vehicles, we also exploit the parked vehicles for data forwarding. To be specific, in the straightway mode, the geographically greedy forwarding is used to forward the data message to the intersection ahead. Here, specially, the parked vehicles are deemed as special mobile vehicles (velocity = 0) and involved in the process of geographically greedy forwarding. In the intersection mode, a vehicle finds the next road to forward the packet according to the utility function of each available road, which is determined by the vehicle density (including both the moving vehicles and parked vehicles) in this road and the distance from the next intersection to the target area.

The utility function of a road segment is defined as follows:

$$U = \frac{\rho_m + \rho_p}{d}, \tag{1}$$

where ρ_m represents the density of mobile vehicles, ρ_p represents the density of parked vehicles, and d is the shortest distance form the next intersection to the target area. If we assume N_m to be the number of parked vehicles in a road segment, N_p to be the number of parked vehicles, L to be the length of this road segment, and r_{pva} to be the

ratio of parked vehicles which are willing to provide parking assistance service, we have

$$\rho_m = \frac{N_m}{L}, \tag{2}$$

$$\rho_p = \frac{r_{\text{pva}} N_p}{L}. \tag{3}$$

For N_p, it could be easily obtained, and for N_m, it could be estimated as follows: the cluster head first estimates the driving time within this segment based on the average velocity as $T = L/v$ and then counts the number of vehicles passing by within time period of T.

Using the above data forwarding strategy, a message could be routed to its target area efficiently. After arriving at a road in the target area, the data is propagated along this road. While its carrier encounters the first parking cluster, it forwards the data to this parking cluster. This parking cluster is then responsible for sending the data to the other parking clusters in the same target area. To indicate whether a parking cluster is the first one obtaining the data in the target area, we adopt an additional bit in the head of each message, where 0 represents it has not traversed any parking cluster until now, while 1 represents it has traversed at least one parking cluster.

If the data source is within the target area of a data, the routing process becomes much simpler. Data is sent to a vehicle that moves into its communication range, which works as mobile helper and forward this data along this road, until the carrier encounters a parking cluster.

4.2.2. Routing from One Parking Cluster to Other Parking Clusters. To effectively route a data to all parking clusters in the target area, we propose a tree-based data forwarding scheme, which forwards each data message from one parking cluster to the other parking clusters in the same target area over a tree structure. We assume that one parking cluster knows the location of other parking clusters within the same target area. This could be realized through a simple mechanism with the help of moving vehicles. For example, each parking cluster periodically broadcasts its location (the location of cluster head) to the parking clusters within two hops (the TTL is set as 2), and adjacent parking clusters exchange the information (similar like <*cluster ID*, *location*>) they obtain with each other. This process is similar to Link-State Broadcast [17]. Due to the high occupancy of parking lots, a long broadcast cycle is enough. As some vehicles may move away while others may move in, the location reported from the same cluster at different time might be slightly different. We abstract the parking clusters and the roads in a target area as a weighted connected graph $G(V, E)$, where V is the set of parking clusters and E is the set of roads between two adjacent parking clusters (might be more than one segment). Weight d_{ij} on E is the estimated transmission delay between adjacent parking clusters. Figure 4 shows one such weighted connected graph. We let adjacent parking lot clusters periodically send a delay probe packet to each other and estimate the transmission delay according to the history record. As the transmission delay between two parking lot clusters is affected by their mutual distance, the traffic density, and other factors that change slowly, this approximation is reasonable.

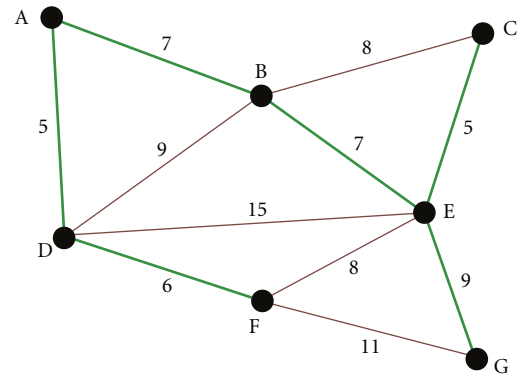


Figure 4: One minimum spanning tree.

The transmission delay between each pair of parking clusters forms a delay matrix, which is updated periodically. With this delay matrix, each parking vehicle could derive a minimum spanning tree, such that the total estimated transmission delay is minimized while routing over this tree. The minimum spanning tree could be easily acquired at each parking cluster through the classic Kruskal's algorithm or Prim's algorithm, both of which are of polynomial complexity. As these two algorithms are all very simple, we will not elaborate them here. If the minimum spanning tree is not unique, the one covers that the shortest road length is chosen as data forwarding tree. Through this way, we could make sure that each parking cluster in a target area maintains the same MST at the same time point. With this tree obtained in each cluster, each data message records its previous hop and is forwarded along this tree. Here, the data forwarding from one cluster to a next-hop cluster uses the routing approach presented in the previous section.

Although routing a packet along any one spanning tree could make sure that the packet could be received by every parking cluster, routing along the minimum spanning tree could realize the same goal in a shorter time. With this routing scheme, each packet only needs to be replicated while new tree branch appears, which greatly decreases the transmission overhead. Moreover, the consistency among packets buffered at different parking clusters could also be maintained.

4.3. Demand-Driven Data Dissemination. VANETs are characterized by limited bandwidth. To make full use of this scarce resource, blind data dissemination should be avoided. We observe that the vehicle users usually only have interests in certain types of data items. Thus, we adopt a demand-driven data dissemination scheme. The vehicles users express their interests in certain types of data messages, while the parking cluster delivers the matched ones to them. In this sense, our system is a pub/sub system. The data source acts as publisher, the mobile vehicle acts as subscriber, while the parking cluster acts as a broker, which is used to ensure

that the data from the data source could be delivered to the subscribers.

To achieve the demand-driven data dissemination, the format of a data message is defined as <*MsgID, AOI, topic, TTL*>, among which the *MsgID* represents the *ID* of this data message, *AOI* represents the target areas, *topic* indicates the type, and *TTL* is the survival time of this data message. For the *topic*, it is represented by a tree as follows in Figure 5.

The data dissemination at the parking cluster includes the following three phases.

(1) Subscribe: an end user customizes a subscription according to his/her requirements, and this subscription is periodically broadcasted in the control channel.

(2) Match: once receiving a subscription, the parking cluster compares it against the stored data messages. This could be realized using the existing matching algorithm [18, 19].

(3) Data dissemination: if there is any data messages which match the subscription, the parking clusters broadcast it in the service channel, which is then received by the subscribers.

Due to the fact that each data item is buffered at multiple parking clusters in the same target area, a vehicle may receive replicas of the same data message while driving in this area. To avoid this problem, we let the subscriber piggyback the IDs of the last n data messages received while broadcasting the subscription. Through this way, we could guarantee that the vehicle users will not be disturbed by replicas of the same data message.

5. Theoretical Analysis

We consider a road segment S with length L. Assume that the number of vehicles moving on this road is K_m, among which the number of vehicles that carry the desired message is K_c. The communication range of each vehicle is R, and there are K_P vehicles parked uniformly along one side of this road. Imagine that a vehicle moves into road S at time 0. We will investigate the probability of getting the desired message through the inter-vehicle-based scheme and the parking-based scheme, respectively, on this road segment.

5.1. Parking-Based Scheme. As the vehicles are uniformly parked along road S, we have the number of vehicles parked within a distance of R of the intersection is

$$N_e = K_p \cdot \frac{R}{L}. \tag{4}$$

Here, the width of the road is neglected. Among the N_e vehicles, the probability of at least one vehicle willing to provide PVA services is

$$p = 1 - \left(1 - \text{pva}_{\text{ratio}}\right)^{N_e}. \tag{5}$$

Substituting N_e with (3), we have the probability for a vehicle getting a data from the parking cluster at the intersection is:

$$p = 1 - \left(1 - \text{pva}_{\text{ratio}}\right)^{K_p R/L}. \tag{6}$$

Now we assume $L = 1000\,\text{m}$, $R = 200\,\text{m}$, $\text{pva}_{\text{ratio}} = 30\%$, and study how the probability p varies with the number of parked vehicles, with the results shown in Figure 6.

We observe that with 40 vehicles parked along a road with a length of 1 km, the probability for the vehicle getting the data at the intersection is higher than 94%. From the parking report [16], we learn that the average number of parked vehicles along a road (in one side) with 1 km is much higher than 40 in urban areas. Thus, while taking advantage of the parking cluster, the probability of getting the desired data at the intersection is greater than 94%.

5.2. Inter-Vehicle Scheme. We assume $N(t)$; $t \geq 0$ denotes the number of encountered mobile vehicles in the time of $(0, t]$. Notice that the $N(t)$, $t \geq 0$ satisfies the conditions of the Poisson process [20]. Therefore, $N(t)$, $t \geq 0$ is a Poisson process. We define W_n as a random variable and have the sequence of $W_0 = 0, \ldots$, $W_i = t_i, \ldots$, where t_i stands for the beginning until encountering the number i mobile vehicle. According to the properties of Poisson process, we can derive that $W_n, n = 1, 2, \ldots$ is an Erlang distribution, with the probability density function expressed as

$$f_{w_n}(t) = \begin{cases} \dfrac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t}, & \text{if } t \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Then, we have the probability of encountering the number n mobile vehicle in the time of $(0, t]$ is

$$F(t) = \int_0^t \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t} dt = \sum_{k=n}^{\infty} \frac{\lambda t^k}{k!} e^{-\lambda t}. \tag{8}$$

As the possibility for a moving vehicle carrying the desired data item, represented by P, is K_c/K_m, the possibility of obtaining the desired data item from the number n encountered vehicle is

$$p_n = \sum_{k=n}^{\infty} \frac{\lambda t^k}{k!} e^{-\lambda t} (1 - P)^{n-1} P. \tag{9}$$

Considering the fact that a moving vehicle might obtain the desired data item from the number $1, 2, \ldots, N(t)$ encountered vehicle, we have

$$p = \sum_{n=1}^{N(t)} \sum_{k=n}^{\infty} \frac{\lambda t^k}{k!} e^{-\lambda t} (1 - P)^{n-1} P. \tag{10}$$

This can be further represented by

$$p = \sum_{n=1}^{N(t)} \left(1 - \sum_{k=0}^{n-1} \frac{\lambda t^k}{k!} e^{-\lambda t}\right) (1 - P)^{n-1} P. \tag{11}$$

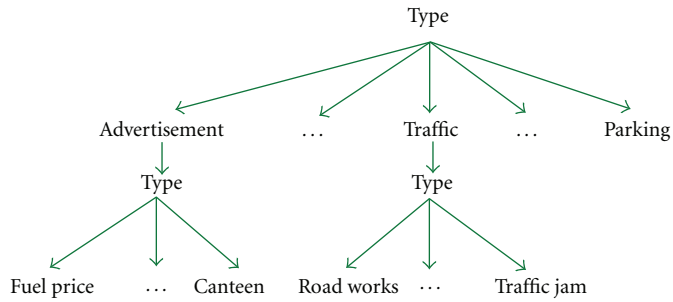


Figure 5: Topic representation.

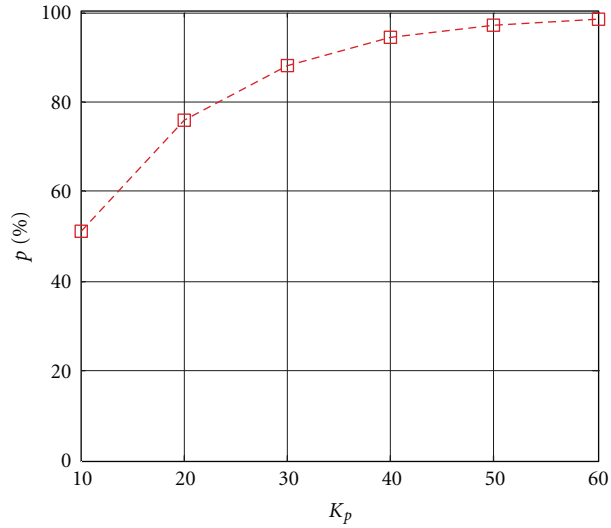


Figure 6: Impact of K_p on the probability p.

Now we assume $L = 1000$ m, set $\lambda = 2$, $K_m = 100$, $t = 20$, $N(t) = 60$ (as the average value obtained in our survey), and study the probability p. According to formula (11), if $K_c = 2$, p equals 69%. That is to say, if there are only 2 copies of the same message kept within a road segment, the possibility for a moving vehicle getting the desired message within 20 s is only 69%. Obviously, the parking-based data dissemination scheme outperforms the inter-vehicle-based scheme.

Table 1: Roadside parking in survey.

Street	Policy	Density	Average
R_{04}, R_{15}, R_{26}	No limits	280–320 veh/km	308 veh/km
R_{37}, R_{79}	Strict limits	15–25 veh/km	21 veh/km
R_{01}, R_{12}, R_{23}, R_{45}, R_{56}, R_{67}, R_{48}, R_{68}, R_{89}	Moderate limits	72–180 veh/km	95 veh/km

6. Performance Evaluation

In this section, we investigate realistic parking and traffic profile in real urban environments and evaluate the performance of parking-based scheme and other two alternative data dissemination schemes in NS-2.33.

6.1. Survey. We performed a six-week survey on an urban area of Chengdu, a city in China, for collecting realistic parking and traffic profile. Since choosing target area is crucial in performance evaluation, we prefer ordinary urban region with typical parking distribution to downtown areas where the parking is above average. As shown in Figure 7, we extract a real street map with the range of 1600 m × 1400 m, which contains 10 intersections and 14 bidirectional roads totaled up to 7,860 meters. Each intersection is marked by a number from 0 to 9.

During the survey, we investigated the traffic and roadside parking statistics at 16:00, 18:00, and 22:00 of every Tuesday, Thursday, and Saturday. We counted the vehicles parked along each street within 5 meters and skipped those parked in the middle of obstacles or too far from the roads. To on-street parking lots, only fringed vehicles along road direction were calculated. As shown in Table 1, there are three classes of streets with different parking limits. The first class permits free parking at roadside, as R_{04}, R_{15}, and R_{26}, which results in a very high node density. The second one, as R_{37} and R_{79}, lacks public parking spaces. These

Figure 7: Road topology in survey and simulations.

Table 2: Performance under default parameters.

Parameter	Default value
Number of vehicle	200
Vehicle velocity	40~80 kph
Size of data message	10 kb
Interval of beacon message	1 second
Data generation rate	0.1/second
Data survival time	30 minutes

Table 3: Performance under default parameters.

Scheme	Parking-based	Inter-vehicle	OD
Average delivery ratio (%)	93.2	85.4	80.6
Average delivery delay (s)	3.5	18.2	7.6
Network traffic overhead	1.53×10^4	2.4×10^5	0.7×10^6

streets have a very low vehicle density that comes from some reserved parking spaces and illegal parking. The rest of the streets belong to the third one, which has a moderate vehicle density. Generally, the parked vehicle numbers are stable in different hours of a day. During the survey, we also calculated daily traffic by counting the passing vehicles within fifteen minutes at random positions and found traffic fluctuating from 300 veh/h (vehicle per hour) to 2200 veh/h at different time of one day. If the road width is 20 m, the corresponding moving vehicles within the area range from 60 to 400, with the average speed ranges of 40 km/h to 80 km/h.

6.2. Simulations. Since accurately modeling node movement is very important for simulation, we use the open source software, VanetMobiSim-1.1 [21], to generate realistic urban mobility traces. The generated traffic file can be directly utilized by NS-2.33. To produce sparse traffic and traffic changes, we deploy different vehicle numbers, that is, 50, 100, 150, 200, 250, and 300, to the map. The radio range is set at 250 m, and the MAC protocol is 2 Mbps 802.11. In the simulation, parked vehicle nodes are located on random positions of each street, following the density collected in Table 1. The average parking time is 41.40 minutes with a standard deviation of 27.17, which is provided in [18]. Since not all parked vehicles are willing to share their wireless devices, a participating ratio of 30% is deployed in default. We assume that the parking clusters are established at the beginning of simulation and are maintained at a cycle of 60 seconds.

To simulate data dissemination, a data source is deployed at the center area of the simulated area, which generates new message with a given time interval. For each message, its target area is specified as a rectangle area which includes four intersections and the roads among them (e.g., the area composed of R_{01}, R_{04}, R_{45}, and R_{15} in Figure 5.), and we assume that 20% of vehicles moving in the target area are interested in it. The default parameters are shown in Table 2.

We mainly discuss three data dissemination mechanisms: our parking-based data dissemination, inter-vehicle-based data dissemination, and OD [8]. For the inter-vehicle based scheme, data messages to be disseminated are routed to the target area using GPSR [22] routing protocol and are maintained within each road segment by the mobile vehicles. While the carrier is leaving a road segment, the maintained data would be transmitted to the furthest vehicle that located within its communication range and drives on the same road segment. Here, similar to our parking-based scheme, we let the message carrier respond to the message subscription within one hop.

The performance of the three mechanisms is measured by the following three metrics.

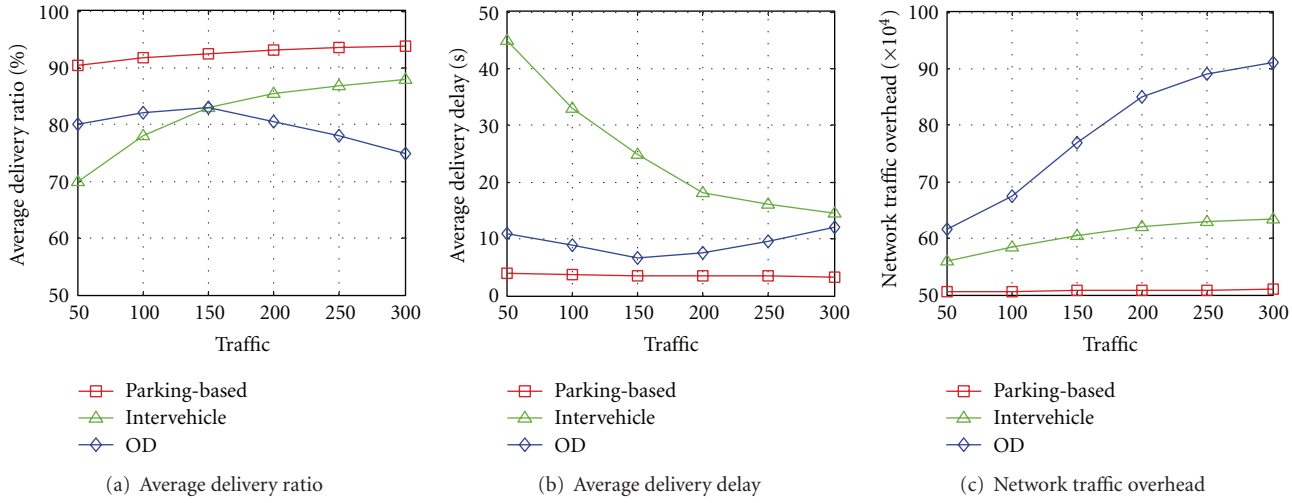
Data Delivery Ratio. For each message, the delivery ratio is defined as the fraction of subscribers that successfully received this message.

Data Delivery Delay. For each message, the delivery delay is defined as the time spent for a subscriber obtaining this message after entering the target area of this message.

Network Traffic Overhead. The network traffic overhead is defined as the total amount of data generated during the simulation.

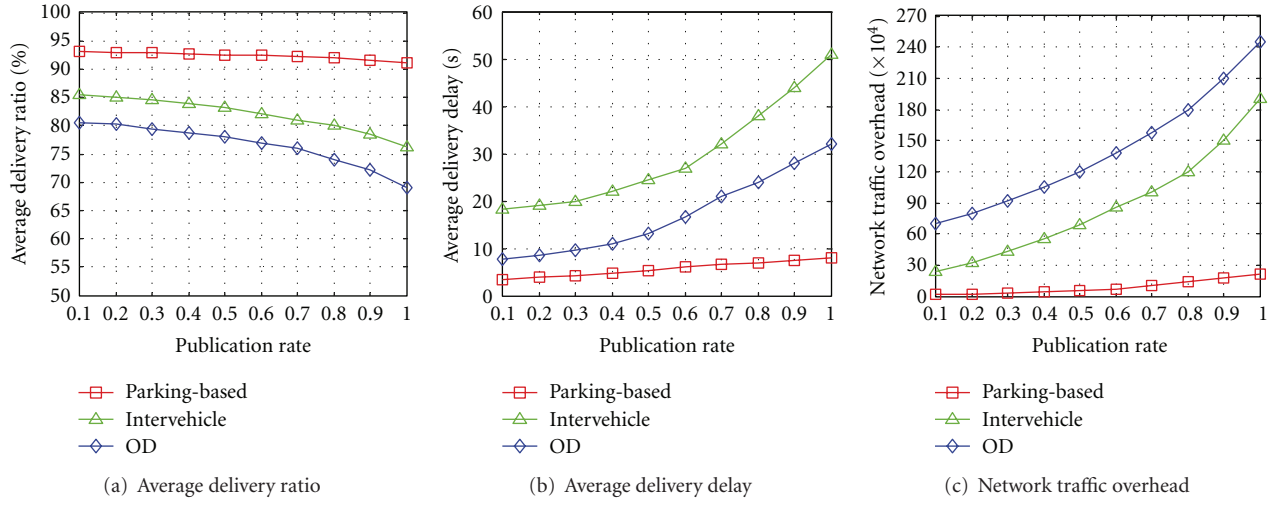
The average delivery ratio is the mean value of delivery ratio of all the disseminated messages, and the average delivery delay is the mean value of the delivery delay of all the disseminated messages. For each measurement, 30 simulation runs are used, and each simulation lasts for 60 minutes.

We first test the performance of the above three schemes under the default parameters. The results are shown as Table 3. We notice that compared to the inter-vehicle scheme and OD, parking-based scheme shows better performance. It achieves a higher delivery ratio with less delivery delay at lower overhead. For parking-based scheme, replicas of the same message are maintained at many parking clusters in the target area. Once a vehicle comes to a road with parking cluster, it will get the desired message in short time. Thus, the average delivery ratio is higher and the average delivery delay is lower. In addition, as each message only needs to



(a) Average delivery ratio (b) Average delivery delay (c) Network traffic overhead

Figure 8: Impact of vehicle density.



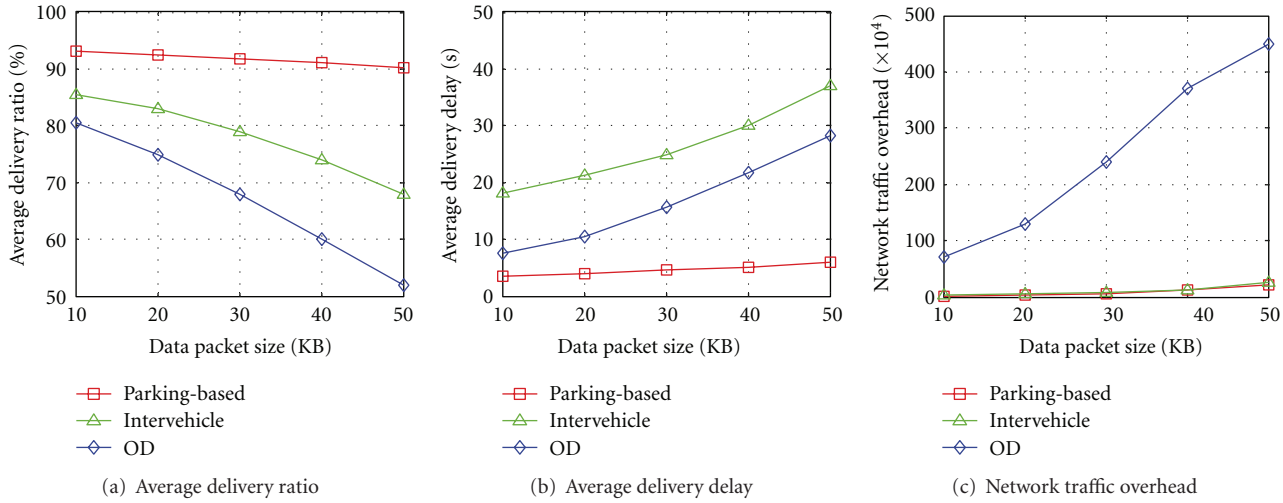
(a) Average delivery ratio (b) Average delivery delay (c) Network traffic overhead

Figure 9: Impact of data publication rate.

be broadcasted within one hop of the parking cluster, the overhead is very low. For the inter-vehicle scheme, data to be disseminated are maintained at the mobile vehicles. Owing to the high mobility of vehicles, frequent handovers among the mobile vehicles are needed to maintain a data message within a road segment. Hence, the network overhead is high. Moreover, there might exist some special cases, in which the vehicle which carries a message leaves a road segment, and it has no chance to perform data handover (there are not any other vehicles within its communication range), and the new coming vehicles have no way to acquire this data. Thus, the delivery ratio is lower and the data delivery delay is higher. For OD, the overhead is much higher than the other two schemes. However, the data delivery ratio is not as high as it should be. In OD, whenever two vehicles move into the transmission range of each other, they will exchange data, which leads to severe congestion and significantly reduces the data delivery ratio.

6.2.1. Impact of Vehicle Density. This group of experiments illustrates the impact of vehicle density on the performance of three data dissemination schemes. Form Figure 8, we observe that the parking-based scheme works well under different road traffic, while the inter-vehicle scheme shows bad performance under sparse traffic. The parking based scheme relies on the roadside parking. As long as there are a certain amount of parked vehicles, the message availability within the target area could be guaranteed. However, the inter-vehicle scheme relies on the moving vehicle, which can hardly ensure the message availability in sparse traffic and thus lead to low delivery ratio. For OD, while the vehicles density increases, the possibility of collisions in media access control (MAC) layer is increased. Thus, the delivery ratio decreases while the delivery delay reduces.

6.2.2. Impact of Data Publication Rate. The data publication rate determines the number of messages to be disseminated



(a) Average delivery ratio (b) Average delivery delay (c) Network traffic overhead

Figure 10: Impact of data packet size.

over VANETs. Higher data publication rate means larger network load. Through this group of experiments, we will see how the data publication rate affects the performance of the three data dissemination schemes. As shown in Figure 9, while the data publication rate varies from 1 message/10 s to 1 message/1 s, the delivery ratio of parking based scheme decreases slightly, while the delivery ratio of inter-vehicle scheme drops obviously. This is because parking-based scheme buffers data at roadside parking and performs data dissemination within one hop, which greatly reduces the possibility of transmission collision. The inter-vehicle scheme maintains data at mobile vehicles, which causes frequent handover and excessive transmission while the publication rate is high. With the increase of the publication rate, the overheads of the three schemes are all increased, and that of OD scheme is more obvious. Here, we also observe that the parking based scheme outperforms the other two schemes.

6.2.3. Impact of Data Packet Size. This group of experiments investigates the impact of data packet size on the performance of three data dissemination schemes. As shown in Figures 10(a) and 10(b), the data delivery ratio and delivery delay of parking-based scheme are superior to that of the other two schemes under different data packet size. For the parking-based scheme, data messages are maintained at the roadside parked vehicles, which thus could provide stable data dissemination services for the vehicles passing by. However, for the inter-vehicle scheme, data messages need to be frequently handed over to the vehicles still moving within the road segment. With the increasing data packet size, the handover suffers from more losses; thus, the data delivery ratio is decreased and the data delivery delay is increased. For OD, larger data packet size means much more serious collision. Hence, the performance becomes worse.

7. Conclusion

Data dissemination over VANETs is challenging due to the fact that data messages can hardly be kept in a specified target area. In this paper, we propose the idea of parking-based data dissemination, which leverages the roadside parking to buffer the data to be disseminated and performs data dissemination. We organize the parked vehicles into clusters, offer a routing scheme to distribute each data message to appropriate roadside parking, and introduce the pub/sub scheme into the last stage of data dissemination. Our parking-based data dissemination scheme exhibits a low capital overhead by exploiting the free resources offered by parked vehicles and a low operational overhead via efficient operations. The theoretical analysis demonstrates the superiority of our scheme. At last, the numerical results also show that our scheme achieves a higher data delivery ratio at lower network traffic overhead and reasonable delay.

Acknowledgments

This work is supported by National Science Foundation of China under Grant nos. 61170256, 61103226, 61103227, 61173172, and 61272526 and the Fundamental Research Funds for the Central Universities under Grant nos. ZYGX2011J060, ZYGX2010J074, and ZYGX2011J073.

References

[1] M. Nekovee, "Epidemic algorithms for reliable and efficient information dissemination in vehicular Ad Hoc networks," *IET Intelligent Transport Systems*, vol. 3, no. 2, pp. 104–110, 2009.

[2] M. Caliskan, D. Graupner, and M. Mauve, "Decentralized discovery of free parking places," in *Proceedings of the 3rd ACM International Workshop on Vehicular Ad Hoc Networks (VANET'06)*, pp. 30–39, New York, NY, USA, September 2006.

[3] H. Wu, R. Fujimoto, R. Guensler, and M. Hunter, "MDDV: a mobility-centric data dissemination algorithm for vehicular networks," in *Proceedings of the 1st ACM International Workshop on Vehicular Ad Hoc Networks (VANET'04)*, pp. 47–56, New York, NY, USA, October 2004.

[4] L. Wu, M. Liu, X. Wang, G. Chen, and H. Gong, "Mobile distributionaware data dissemination for vehicular Ad Hoc Networks," *Software Journal*, vol. 22, no. 7, pp. 1580–1596, 2011.

[5] C. Maihöfer, T. Leinmüller, and E. Schoch, "Abiding geocast: time-stable geocast for Ad Hoc networks," in *Proceedings of the 2nd ACM International Workshop on Vehicular Ad Hoc Networks (VANET'05)*, pp. 20–29, ACM, 2005.

[6] J. Zhao, Y. Zhang, and G. Cao, "Data pouring and buffering on the road: a new data dissemination paradigm for vehicular Ad Hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 6 I, pp. 3266–3277, 2007.

[7] C. Morency and M. Trépanier, "Characterizing parking spaces using travel survey data," CIRRELT, 2006.

[8] B. Xu, A. Ouksel, and O. Wolfson, "Opportunistic resource exchange in inter-vehicle Ad Hoc networks," in *Proceedings of the IEEE International Conference on Mobile Data Management (MDM)*, pp. 4–12, 2004.

[9] G. Korkmaz, E. Ekici, F. Ozguner, and U. Ozguner, "Urban multi-hop broadcast protocol for inter-vehicle communication systems," in *Proceedings of the 1st ACM International Workshop on Vehicular Ad Hoc Networks (VANET'04)*, pp. 76–85, 2004.

[10] Y. Ding, C. Wang, and L. Xiao, "A static-node assisted adaptive routing protocol in vehicular networks," in *Proceedings of the 4th ACM International Workshop on Vehicular Ad Hoc Networks (VANET'07)*, pp. 59–68, September 2007.

[11] C. Lochert, B. Scheuermann, C. Wewetzer, A. Luebke, and M. Mauve, "Data aggregation and roadside unit placement for a vanet traffic information system," in *Proceedings of the 5th ACM International Workshop on VehiculAr Inter-NETworking (VANET'08)*, pp. 58–65, New York, NY, USA, September 2008.

[12] M. Eltoweissy, S. Olariu, and M. Younis, "Towards autonomous vehicular clouds," in *Ad Hoc Networks*, vol. 49 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 1–16, 2010.

[13] S. Olariu, I. Khalil, and M. Abuelela, "Taking VANET to the clouds," *International Journal of Pervasive Computing and Communications*, vol. 7, no. 1, pp. 7–21, 2011.

[14] D. Shoup, "Cruising for parking," *Access*, vol. 30, pp. 16–22, 2007.

[15] R. H. Frenkiel, B. R. Badrinath, J. Borràs, and R. D. Yates, "Infostations challenge: balancing cost and ubiquity in delivering wireless data," *IEEE Personal Communications*, vol. 7, no. 2, pp. 66–71, 2000.

[16] A. Adiv and W. Wang, "On-street parking meter behavior," *Transportation Quarterly*, vol. 41, pp. 281–307, 1987.

[17] J. F. Kurose and K. W. Ross, "A top-down approach featuring the internet".

[18] R. Meier and V. Cahill, "STEAM: event-based middleware for wireless Ad Hoc networks," in *Proceedings of the 1st International Workshop on Distributed Event-Based Systems (DEBS'02)*, pp. 639–644, 2002.

[19] L. Fiege, F. Gärtner, O. Kasten, and A. Zeidler, "Supporting mobility in content-based publish/subscribe middleware," in *Proceedings of the ACM/IFIP/USENIX International Conference on Middleware (Middleware'03)*, pp. 103–122, 2003.

[20] W. Feller, *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons, New York, NY, USA, 2008.

[21] J. Härri, F. Filali, C. Bonnet, and M. Fiore, "VanetMobiSim: generating realistic mobility patterns for VANETs," in *Proceedings of the 3rd International Workshop on Vehicular Ad Hoc Networks (VANET'06)*, pp. 96–97, ACM, New York, NY, USA, 2006.

[22] B. Karp and H. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom'00)*, pp. 243–254, ACM, 2000.

Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2012, Article ID 527097, 6 pages
doi:10.1155/2012/527097

Research Article

Latency Estimation-Based Data Delivery Scheme for Vehicular Ad Hoc Networks

Haigang Gong,[1] Lingfei Yu,[2] Ke Liu,[3] Fulong Xu,[1] and Xiaomin Wang[1]

[1] *School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China*
[2] *College of Computer Science and Technology, Southwest University for Nationalities, Chengdu 610043, China*
[3] *School of Foreign Language, Southwest University for Nationalities, Chengdu 610043, China*

Correspondence should be addressed to Lingfei Yu, linphie@163.com

Received 2 November 2012; Accepted 5 December 2012

Academic Editor: Ming Liu

Data delivery is the most compelling issue for the applications in vehicular ad hoc networks. The characteristics of VAENTs such as highly dynamic topology and the road-constrained mobility pose great challenges for data delivery in vehicular ad hoc networks. In this paper, a latency estimation-based data delivery (LED) scheme is presented. According to the current location and the moving direction of the vehicles, LED calculates the expected delivery latency (EDL) of each vehicle and chooses the vehicle with the shortest EDL as the next hop to forward data. Simulation results show that LED performs better than VADD.

1. Introduction

Vehicular ad hoc networks have been envisioned to be promising in road safety and many other commercial applications, such as the intelligent transportation system (ITS), one of the most important applications of VANET, which have been deployed in USA, Europe, and Asia. For example, with the ITS, an emergency warning can help drivers behind a crashed vehicle (or incident) to avoid multicar collisions [1]. Besides, a vehicular network can be used to alert drivers to potential traffic jams, providing more convenience and efficiency.

Vehicular ad hoc networks have some characteristics as follows: (1) the topology of VANETs changes rapidly due to the high mobility of the vehicles. Different from the traditional node-to-node pattern of routing which is based on the mobile node topology in mobile ad hoc networks (MANETs) [2–6], VANETs usually cannot establish a stable end-to-end path to transmit data packets due to the intermittent connectivity [7–9]. (2) The mobility of vehicle nodes in VANETs is constrained by predefined roads. Due to the buildings and other obstacles, the data delivery based on the V2V communications over VANETs is also constrained by the roads.

Efficient data delivery based on infrastructure-less V2V communications is the most compelling issue for the mentioned applications in VANETs. However, the characteristics of VAENTs such as highly dynamic topology and the road-constrained mobility pose great challenges for data delivery in vehicular ad hoc networks. Some researches have been proposed to resolve data forwarding for vehicular networks. GPSR [4] is a geographical forwarding approach which always chooses the next hop closer to the destination. It is a classic and very efficient data delivery scheme in ad hoc networks, but it may not be suitable for sparsely connected vehicular networks and it cannot select a path with lower latency. Zhao and Cao present VADD [7], which utilizes the predictable mobility in VANET for data delivery. VADD regards each road as a link whose delay is the time consumed to deliver a packet through the corresponding road by the multihop communication and the carrying of moving vehicles on this road. Therefore, packets will be delivered along the shortest-delay trajectory. In cases when no vehicles are available in the next road for data delivery along the optimal trajectory, VADD improves packet delivery reliability by making a routing decision at each intersection to select the best currently available trajectory. Though VADD performs better than GPSR, it neglects the performance difference of individual vehicle while forwarding data.

In this paper, a latency estimation-based data delivery (LED) scheme is presented. According to the current location and the moving direction of the vehicles, LED calculates

the expected latency of the individual vehicles and chooses the best vehicle as the next hop to forward data.

The paper is organized as follows: Section 2 discusses some related works. Section 3 states the problems. Section 4 presents the detailed design of LED, followed by the evaluation of LED in Section 5. Section 6 concludes the paper.

2. Related Works

There have been a lot of research works on vehicular ad hoc networks. In this section, we just discuss some related works about the data delivery scheme.

Ding et al. [10] introduce that the sensing data from a vehicle may need to be sent to a sink that is deployed miles away, or a vehicle may want to send queries to a remote site such as gas station, restaurant, or parking area. Thus, a multihop routing algorithm is needed in a large vehicular network for these applications. Then authors propose a static node-assisted adaptive data dissemination protocol (SADV) for vehicular networks. They suggest employing mechanisms that enable the packet to wait at an intersection until the best path is available. To achieve this, they add static nodes at intersections that store and forward the packet when appropriate. Sormani et al. [11] take an accident happening on the highway as an example. In order to prevent dangerous situations, the crashed cars and possibly other cars in the vicinity immediately generate a message that is delivered to the nearest ambulance and police stations, the vehicles within one-mile approaching the accident area and the closest highway entrances.

Most protocols assume that intermediate nodes can be found to set up an end-to-end connection; otherwise, the packet will be dropped [3, 5]. Naumov and Gross [2] present a position-based routing scheme called CAR designed specifically for intervehicle communication in a city and highway environment. CAR integrates locating destinations with finding connected paths between source and destination. Once a path is found, its scheme is called CAR designed specifically for intervehicle communication in a city and highway environment. CAR integrates locating destinations with finding connected paths between source and destination. Once a path is found, it is auto-adjusted on the fly to account for changes, without another discovery process.

Wisitpongphan et al. [12] indicate that although the average recovery time for I-80 type of freeways is, on average, less than 30 seconds, such long network disconnection time could be a major problem for conventional ad hoc routing protocols [3, 5], which just tolerates a network disconnection time of up to 2-3 seconds. In addition, some time-critical applications may not be able to function properly in disconnected VANET as the end-to-end delay could be on the order of several minutes. Vahdat and Becker in [13] use epidemic routing. Whenever two nodes meet, they exchange the data that they do not possess. The random pairwise exchanges of messages among mobile hosts ensure eventual message delivery. In [7], Zhao and Cao make use of the predicable vehicle mobility, which is limited by the road traffic pattern and road layout. Based on the existing road traffic pattern, a vehicle can find the next road to forward the packet to reduce the delay. The estimation of packet forwarding delay through each road is based on some statistical data such as the average vehicle density. As the vehicle density on each road may vary with time, which greatly influences the packet forwarding delay, the shortest-delay path calculated based on the statistical data may not reflect the real optimal one.

Some protocols add controlled nodes to help message forwarding. Li and Rus [14] proposed to have mobile nodes proactively modifing their trajectories to transmit messages. Zhao et al. [15] proposed to utilize special mobile nodes called message ferries, which move nonrandomly in the deployed area, to help data delivery in a sparse network. However, in vehicular networks, it is impossible to modify the trajectories of the moving vehicles or finding such ferries.

3. Problem Statement

3.1. VADD Analysis. VADD is based on the idea of carry and forward. The most important issue is to select a forwarding path with the smallest packet delivery delay. As shown in Figure 1, VADD has three packet modes: Intersection, StraightWay, and Destination based on the location of the vehicle that carries the packet. In StraightWay mode, the geographically greedy forwarding is applied to deliver the packet to the intersection, entering into Intersection mode. In Intersection mode, packet is determined to be delivered to one of the road segments connected with the intersection. Among the three modes, the Intersection mode is the most critical and complicated one, since vehicles have more choices at the intersection.

Before making a forwarding decision, VADD estimates the data delivery delay of different paths to the destination.

As shown in Figure 2, for a packet at the intersection I_m, the expected delay of delivering the packet through road segment r_{mn} is as follows:

$$D_{mn} = d_{mn} + \sum_{j \in N(n)} \left(P_{nj} \times D_{nj} \right), \tag{1}$$

where D_{mn} is the expected data delivery delay from I_m to the destination if the packet carrier at I_m chooses to deliver the packet following road segment r_{mn}; d_{mn} is the delivery delay between I_m and I_n; P_{nj} is the probability that the packet is forwarded through road r_{nj} at the intersection I_n; $N(n)$ is the set of neighboring intersections of I_n.

To shorten the delivery delay, VADD sorts D_{mn} for each neighboring intersections of I_m and chooses the road with the smallest expected delay.

3.2. Problem Description. VADD has some disadvantages as follows. (1) In StraightWay mode, packet could be determined to deliver to which direction according to the expected delivery delay. However, it takes some time to deliver the packet to the intersection, which was influenced by the location of the packet carrier, the moving speed, and the traffic flow of the vehicles. (2) Though the geographically greedy forwarding method achieves shorter delivery delay in the StraightWay mode, the packet may be delivered to a road with high latency because the moving direction of the vehicle

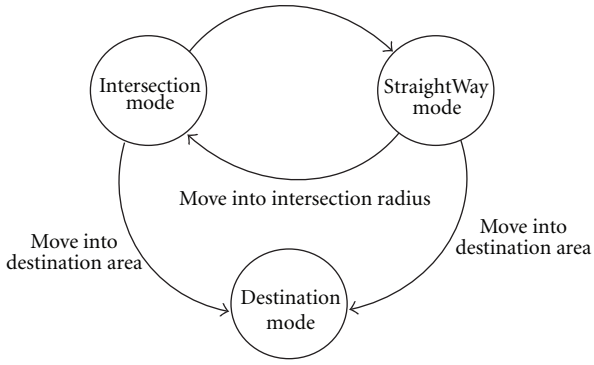


Figure 1: The transition mode in VADD.

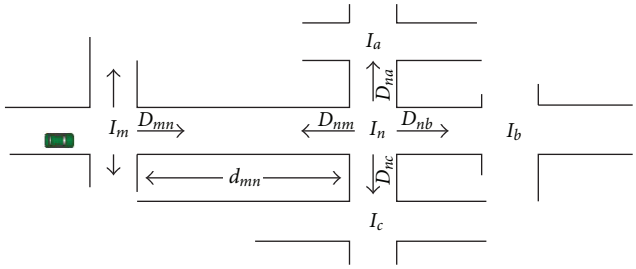


Figure 2: Delay model in VADD.

ahead is uncertain and it may not contact the neighboring vehicle in the optimal road. The wrong choice of the road with high latency leads to more end-to-end delivery delay. In fact, when encountering a vehicle driving to the road with lower delivery delay, packet can be delivered to the vehicle and carried by it to the road with shorter delay, which may achieve shorter data delivery delay than the geographically greedy forwarding approach. Unlike VADD that estimates the expected data delivery delay based on the road, our proposed LED calculates the latency of the individual vehicle and chooses the optimal vehicle as the packet carrier according to the current location and turning direction of the vehicle to achieve the lowest transmission latency.

4. The Design of LED

4.1. The Expected Delivery Latency. LED scheme calculated the expected delivery latency (EDL) of the vehicles based on their current location and moving direction. Assuming that the vehicle V_1 moves in the road r_{ij} at the speed of v and the distance to the intersection I_j is l, the EDL of V_1 includes two parts. (1) The time carried by V_1 to the intersection I_j, denoted by t_{carry}. (2) The expected end to end delay to the destination from the intersection I_j, denoted by t_{e2e}. t_{carry} can be calculated by $t_{\text{carry}} = l/v$.

As in VADD, LED selects the next road with the shortest expected delivery delay in the intersection. Assuming that the intersection I_j connects n road such as $r_{j1}, r_{j2}, \ldots, r_{jn}$ sorted by the nondecreasing order of $D_{j1}, D_{j2}, \ldots, D_{jn}$ calculated by (1). If the vehicle V_1 carries the data to the intersection I_j and will move into road r_{ja} $(1 \le a \le n)$, the data will be carried by V_1 and delivered to road r_{ja} in the worst case. Obviously, the probability that data delivered to the road whose expected delivery delay is higher than road r_{ja} is zero. The probability that data delivered to road $r_{jm}(1 \le m \le a)$ is defined as (2)

$$p_{jm} = \begin{cases} CP_{j1} & m = 1, \\ \left[\prod_{s=1}^{m-1}\left(1 - CP_{js}\right)\right] CP_{jm} & 2 \le m \le a-1, \\ 1 - \sum_{s=1}^{a-1} p_{js} & m = a, \end{cases} \tag{2}$$

where CP_{jm} is the probability of meeting at least one neighboring vehicle towards road r_{jm}.

So, if the vehicle V_1 carries the data to the intersection I_j, the end-to-end delivery delay t_{e2e} from the intersection I_j to the destination is as follows:

$$t_{\text{e2e}} = \sum_{m=1}^{a} p_{jm} D_{jm}. \tag{3}$$

And then, the expected delivery latency of the vehicle V_1 could be calculated by the equation (4) at the current moment:

$$\text{EDL} = t_{\text{carry}} + t_{\text{e2e}} = \frac{l}{v} + \sum_{m=1}^{a} p_{jm} D_{jm}. \tag{4}$$

As seen from (4), LED computes EDL of the individual vehicle more accurate than VADD while it considers the moving direction of the vehicle in the next intersection and the transmission direction of the data in the worst case.

4.2. Data Delivery. According to (4), each vehicle calculates its EDL and broadcasts a beacon message containing EDL periodically. While receiving beacon message, the vehicle could find its neighboring vehicles and the EDL of the neighbors. When making a forwarding decision, data is always delivered to the vehicle with the shortest EDL until it arrives at the destination.

LED can be used to select an optimal driving route because (4) considers four factors: the current location of the vehicle, the moving direction of the vehicle, the expected delivery latency of the road connected with the intersection ahead, and the probability that data will be delivered to the road connected with the intersection ahead. For example, as in Figure 3, the vehicles V_1 and V_2 move to the intersection I_j and the vehicle V_3 moves to the intersection I_i. The transmission direction of data is determined by the EDL of the three vehicles. If V_3 has the shortest EDL, data will be delivered to the intersection I_i. Otherwise, data will be transmitted to the intersection I_j. Compared with LED, VADD cannot choose the transmission direction of data according to the current traffic, which may incur longer delivery delay. In addition, if V_3 has the largest EDL, data will be delivered to the intersection I_j. Assume that the intersection I_j connects the road r_{j1} and r_{j2}, and the end-to-end delay of r_{j1} is far less than r_{j2}. In the intersection I_j, the vehicle V_1 will enter into the road r_{j1} and V_2 will enter into the road r_{j2}. If the probability that V_2 meets a vehicle moves to the road r_{j1} in the intersection I_j is too small, data

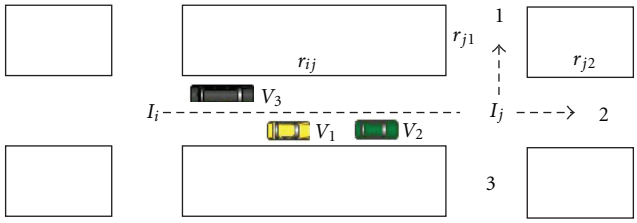


Figure 3: Data delivery of LED.

will be carried by V_1 and delivered to the road r_{j1} in spite of the fact that V_2 is closer to the intersection I_j than V_1. On the contrary, if V_2 has a high probability to contact a vehicle that moves to the road r_{j1} in the intersection I_j, data will be transmitted to V_2 and delivered to the vehicle that moves to the road r_{j1} by V_2 quickly.

5. Performance Evaluation

We choose VanetMobiSim+NS2 to evaluate the performance of the LED, in which VanetMobiSim is used to produce the motion model of the vehicles. VADD is also simulated to compare with the LED.

The experiment is based on a 5400 m × 4800 m map, including 36 intersections and the destination is set to the central point of the map. The speed of vehicle is differentiated from 40 km/h~80 km/h and is set before the vehicle enters into the road segment, which is restricted by the speed limit of the road segment. MAC protocol of the vehicles uses 802.11 and the transmission radius is set to 200 m. TTL is an upper bound on the time that a packet is valid or not. If the TTL of packet expires, packet will be dropped. The simulation parameters are listed in Table 1.

We evaluate the performance of the two schemes on data delivery ratio and average delivery delay. Firstly, we observe the performance of the two schemes when the packet is without TTL constrains. Then TTL is set to 600 seconds, the performance of the two schemes under different scenarios is discussed.

5.1. Without TTL Constrains. In this section, we set TTL to infinite. That is to say, packet will not be dropped forever and all packets will be delivered to the destination. The average delivery delay of the two schemes is listed in Table 2.

As seen from Table 2, LED has less delivery delay than VADD. The reason is that VADD computes the expected delivery delay based on the statistics of traffic flow, which is used to assist in packet forwarding, without any consideration of the current delivery capability of the individual vehicle. In contrast, LED takes the current state of the individual vehicle and calculates the EDL of the individual vehicle more accurately based on the current location and moving direction of the vehicle, which packet is delivered to the vehicle with the lowest EDL.

5.2. Impact of Density. Figure 4 shows the impact of the number of vehicles on the performance of the two schemes in terms of the data delivery ratio and the average delivery delay, respectively.

Table 1: Default simulation parameters.

Parameters	Value
Simulation area	5400 m × 4800 m
Intersections	36
Number of vehicles	100
Transmission radius	200 m
Speed of vehicles	40 km/h–80 km/h
MAC protocol	802.11
Time-to-live (TTL)	600 seconds
Packet generation rate	1 packet/second

Table 2: Performance of the two schemes without TTL constrains.

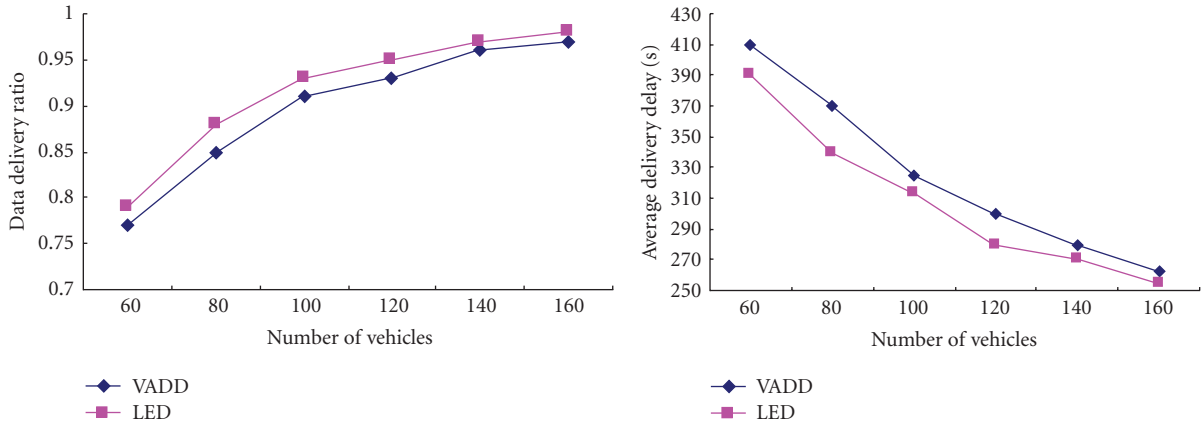
	LED	VADD
Average delivery delay(s)	323.6	347.1

Figure 4(a) depicts the impact of the number of vehicles on the data delivery ratio. As seen from Figure 4(a), the data delivery ratio increases quickly with the increase of the number of vehicles. This is because the growth of the density of the vehicles improves the connectivity of the networks and there are more chances to deliver the packet to meet an appropriate next hop. But when the number of vehicles is high, the data delivery ratio increases slowly and there is no much difference between the two schemes. The reason is that when the density of the vehicles is high, there are more chances to meet an optimal vehicle in the intersection and the packet will be delivered to the vehicle closer to the intersection, in which LED performs like VADD. Figure 4(b) shows the average delivery delay of the two schemes. The higher the density of the vehicles, the higher probability the packet is delivered to the road with the shorter expected delivery delay to the destination, which reduces the average delivery delay.

5.3. Impact of Communication Radius. Figure 5 shows the performance of the two schemes under different communication radii. Figure 5(a) compares the two schemes on data delivery ratio and Figure 5(b) describes the average delivery delay of the two schemes. When the communication radius is small, the network connectivity is poor and there are little chances to communicate with each other, leading to poor performance. With the enlargement of the communication radius, the data delivery ratio of the two schemes increases and the average delivery delay decreases. The enlargement of the communication radius improves the network connectivity, introducing more chances to meet the next hop. Moreover, there is less time to carry the packet if the communication radius is large and the packet will be delivered to the destination rapidly by multihop.

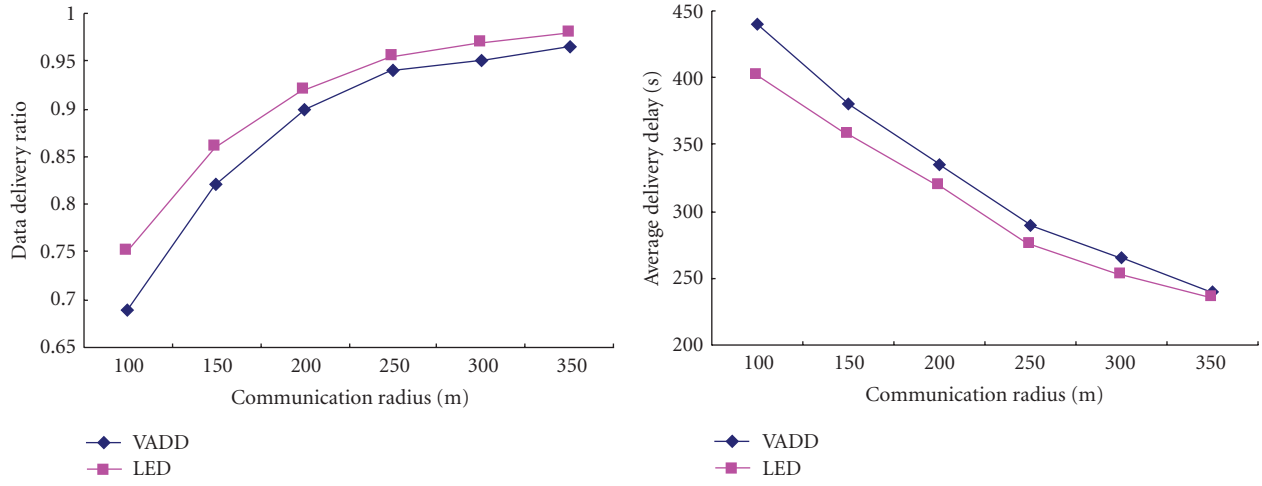
6. Conclusions

In this paper, we proposed a latency estimation-based data delivery scheme. Based on the current location and the moving direction of the vehicles, LED calculates the expected delivery latency of the individual vehicles and chooses the



(a) Data delivery ratio (b) Average delivery delay

Figure 4: Performance under different densities of the vehicles.



(a) Data delivery ratio (b) Average delivery delay

Figure 5: Performance under different communication radii.

vehicle with the shortest EDL as the next hop to forward data. Simulation results show that LED performs better than VADD.

Acknowledgments

This work is supported by the National Science Foundation under Grants no. 61003229, 61170256, 61173172, and 61103226 and the Fundamental Research Funds for the Central Universities under Grants no. ZYGX2010J074, ZYGX2011J073, and ZYGX2011J060.

References

[1] R. Fracchia, M. Meo, and D. Rossi, "Knowing vehicles location HELPS avoiding broadcast packets storm," in *Proceedings of the 4th Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops'06)*, pp. 118–123, Pisa, Italy, March 2006.

[2] V. Naumov and T. R. Gross, "Connectivity-aware routing (CAR) in vehicular Ad Hoc networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (IEEE INFOCOM'07)*, pp. 1919–1927, Anchorage, Alaska, USA, May 2007.

[3] D. Johnson and D. Maltz, "Dynamic source routing in Ad Hoc wireless networks," in *Mobile Computing*, pp. 153–181, Kluwer Academic, New York, NY, USA, 1996.

[4] B. Karp and H. T. Kung, "GPSR: greedy perimeter stateless routing for wireless networks," in *Proceedings of the 6th ACM Annual International Conference on Mobile Computing and Networking (MOBICOM'00)*, pp. 243–254, August 2000.

[5] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'99)*, pp. 90–100, New Orleans, La, USA, February 1999.

[6] E. M. Royer and C. K. Toh, "A review of current routing protocols for Ad Hoc mobile wireless networks," *IEEE Personal Communications*, vol. 6, no. 2, pp. 46–55, 1999.

[7] J. Zhao and G. Cao, "VADD: vehicle-assisted data delivery in vehicular Ad Hoc networks," in *Proceedings of the 25th IEEE International Conference on Computer Communications (INFOCOM'06)*, pp. 1–12, Barcelona, Spain, April 2006.

[8] K. C. Lee, M. Le, J. Härri, and M. Gerla, "LOUVRE: landmark overlays for urban vehicular routing environments," in *Proceedings of the 68th Semi-Annual IEEE Vehicular Technology Conference (VTC'08-Fall)*, pp. 1–5, Calgary, Canada, September 2008.

[9] N. Wisitpongphan, F. Bai, P. Mudalige, and O. K. Tonguz, "On the routing problem in disconnected vehicular Ad Hoc networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (IEEE INFOCOM'07)*, pp. 2291–2295, Anchorage, Alaska, USA, May 2007.

[10] Y. Ding, C. Wang, and L. Xiao, "A static-node assisted adaptive routing protocol in vehicular networks," in *Proceedings of the 4th ACM International Workshop on Vehicular Ad Hoc Networks (VANET'07)*, pp. 59–68, September 2007.

[11] D. Sormani, G. Turconi, P. Costa, D. Frey, M. Migliavacca, and L. Mottola, "Towards lightweight information dissemination in inter-vehicular networks," in *Proceedings of the 3rd ACM International Workshop on Vehicular Ad Hoc Networks (VANET)*, pp. 20–29, New York, NY, USA, September 2006.

[12] N. Wisitpongphan, F. Bai, P. Mudalige, and O. K. Tonguz, "On the routing problem in disconnected vehicular Ad Hoc networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (IEEE INFOCOM'07)*, pp. 2291–2295, Anchorage, Alaska, USA, May 2007.

[13] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Tech. Rep. CS-200006, 2000.

[14] Q. Li and D. Rus, "Sending messages to mobile users in disconnected ad-hoc wireless networks," in *Proceedings of the 6th ACM Annual International Conference on Mobile Computing and Networking (MOBICOM'00)*, pp. 44–55, August 2000.

[15] W. Zhao, M. Ammar, and E. Zegura, "A message ferrying approach for data delivery in sparse mobile Ad Hoc Networks," in *Proceedings of the 5th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MoBiHoc'04)*, pp. 187–198, United States Department of Transportation, May 2004.