EVALUATION OF A METHODOLOGY FOR SCALABLE DYNAMIC VEHICULAR AD-HOC NETWORKS IN A WELL-CALIBRATED VEHICULAR MOBILITY TEST BED

Sandeep Mudigonda, M.Sc.
Graduate Student,
Rutgers Intelligent Transportation Systems (RITS) Laboratory
Rutgers, The State University of New Jersey,
623 Bowser Road Piscataway, NJ 08854 USA,
Tel: (732) 445-3162 Fax: (732) 445-0577
e-mail: sandeepm@rci.rutgers.edu

Junichiro Fukuyama, Ph.D.
Toyota InfoTechnology Center
601 Lexington Ave., 49th Floor, New York, NY 10022
Tel: (212)715-7466
e-mail: jfukuyama@us.toyota-itc.com

Kaan Ozbay, Ph.D
Professor,
Rutgers Intelligent Transportation Systems (RITS) Laboratory
Rutgers, The State University of New Jersey,
623 Bowser Road Piscataway, NJ 08854 USA,
Tel: (732) 445-2792 Fax: (732) 445-0577
e-mail: kaan@rci.rutgers.edu

Abstract: 270 words
Word count: Text 5, 584 + Tables 3 + Figures 7 = 8, 134 words

Paper Submitted for Presentation and Publication at the Transportation Research Board’s 92nd Annual Meeting, Washington, D.C., 2013
Submission Date: November 15th, 2012
ABSTRACT

Connected vehicles are becoming ubiquitous with each passing year. Increase in mobile computing is proliferating the possible applications of connected vehicles. Many of these applications involve a continuous need for vehicles to connect to the communication infrastructure. This could result in congestion of the communication network. In this study we evaluate a novel “dynamic grouping” methodology that combines vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication schemes to make the optimal use of the communication infrastructure. The methodology for dynamic grouping of instrumented vehicles is implemented in a realistic and well-calibrated microscopic traffic simulation test bed of the New Jersey Turnpike for the application of sensor data collection. A reduction in communication infrastructure load of 66-91% can be achieved using the dynamic grouping for systematic aggregation of vehicular information. The maximum bandwidth usage is used as a measure to show that the name-address mapping is scalable. We show that the dynamic grouping methodology is very scalable with negligible loss in data quality as compared to the scenario where each vehicle connects to the communication infrastructure independently. The scalability is shown by generating response surfaces for the load on communication channels for different market penetration and communication ranges. These response surfaces can also be useful in predicting the channel load under future scenarios with increasing market penetration and power of communication radios. The data quality is validated using reported speed and estimated travel times over the network. It is shown that on an average the error in speed is 5.5-8% albeit using far lesser bandwidth using the dynamic grouping approach. Similarly, travel time along different paths is shown to be within 5% during regular conditions and within 10% during non-recurrent congestion.
INTRODUCTION

The fast-paced growth in technology is making vehicles smarter in many ways. Vehicles are communicating with one another and the static infrastructure and exchanging information. The idea of exchanging information between cars and infrastructure, in other words, connected vehicles, has huge economic and scientific potential. This information includes: (i) sensor data for reporting traffic and weather conditions, (ii) situational awareness information for traffic safety (iii) local information for location-based services. Thus connected vehicles can provide the drivers with improved mobility, safer travel and better driver experience.

Given the widespread use of mobile computing, the possible number of applications of connected vehicles will increase as well as the bandwidth usage. Each application would require that each vehicle to connect periodically to the infrastructure. In this process, each car passes on its identity and location constantly. This phenomenon is called as name-address mapping in network communication. This repeated need for a connection for each equipped vehicle to send or accept information could result in the communication network being overwhelmed.

With various entities as stakeholders, the right amount of investment in the required infrastructure is important. The load that needs to be sustained by the infrastructure is a direct determinant of the investment to be made. Hence it is important to measure the load and, if possible, reduce it for more efficient ways for information exchange among connected vehicles.

In order to evaluate the load on the communication network, it is imperative to use a realistic test bed to measure the load. Additionally, given the scope that this load could become prohibitive due to continuous need for name-address mapping, a scalable mechanism to manage communication network load is required.

The rest of the paper is organized as follows. The second section presents the literature on VANET modeling in general along with some vehicular grouping approaches. The third section illustrates the motivation and objectives of our study. The fourth section presents a description of the methodology used in the study. Fifth section illustrates simulation results of methodology. Finally the conclusions of this study along with future research needs are discussed.

LITERATURE REVIEW

V2V communication is unique and different than the standard modeling of ad-hoc networks. The aspect of uniqueness can be characterized in terms of the following:

- Time-varying network topology
- Constantly changing number of vehicles with different speeds

VANETs have been modeled by modeling the communication layer and mobility layer using different tools. Tools such as NS-2 (1,2), TraCI (3), NCTUns (4), Jist/SWANS (5), etc. have been widely used to simulate the communication network layer. Vehicular
mobility simulation has been undertaken using simulation packages such as PARAMICS (1), VISSIM (2,3,4,6), VATSIM (7), SUMO (8), or other customized packages (5,9).

These two layers can be coupled in the end as shown in (1-9).

Sometimes the objective of the study is not to model the intricacies of the communication network but to find some average metrics of the application layer and model the impact on the vehicular layer. To achieve this goal, the high-fidelity simulation of the communication layer may not be necessary. For such studies, especially on a larger scale of application, if the signal transmission is fairly constant, the vehicular mobility is more dominant in evaluating the VANET performance. Examples of studies that used vehicular simulation without high-fidelity network simulators but with approximation for communication channel modeling via assuming a fixed communication range:

1. Ozbay et al. (2007) studied the propagation of information using vehicular communication on a simulation model in PARAMICS for various levels of market penetration. (10)
2. Du et al., (2008) used a simulation model of a straight section of a freeway in MITSIM to evaluate the effect of interference on data exchange on VANETs. (11)
3. Jordan et al., 2012 built a microscopic model of a straight freeway section in VISSIM to investigate warning time to be given to drivers for the presence of emergency vehicles. (12)
4. Yang and Jin (2012) investigate the possible reduction in vehicular emission via inter-vehicle communication using a simulation model based on a car-following algorithm. (13)
5. Kim (2012) tested the effectiveness of traveler information using V2V communication for a hypothetical urban grid (6-by-6 blocks) network in a microsimulation network. (14)
6. CORSIM was used by Elefteriadou and Martin (2012) to test advanced cruise control (ACC) and lane change assist (LCA) technologies by changing some default parameters in CORSIM. (15)
7. Johnson et al. (2012) evaluated the effectiveness of location and conditions of charging station information for electric vehicles by simulating a network in MATLAB. (16)

Grouping of vehicles in V2V communication has been proposed in some studies including,

1. Briesemeister (2001) proposed a methodology for grouping of equipped vehicles based on the identity of the nearest neighbors. The purpose of this methodology is the propagation of information relevant to a group. The methodology was implemented for a simplistic uncalibrated traffic network. (17)
2. Chennikara-Varghese et al. (2007) theorized the possible use of integration of V2V to V2I with local peer group architecture for propagation of safety information. (18)
3. Rybicki et al. (2009) used an integrated approach to V2V and V2I on a mobile internet platform and evaluated for safety information propagation. (19)
4. Korkmaz et al. (2006) proposed a method of grouping vehicles based on fixed spatial and temporal segments to integrate V2V information with communication
infrastructure. (Error! Reference source not found.)

Li et al. (2012) proposed a dynamic vehicle grouping algorithm to form VANETs to aggregate probe vehicle data. The methodology has been tested on a linear highway in VISSIM and shown to have reduced communication load. (21)

MOTIVATION AND OBJECTIVES

Most studies presented above (11-Error! Reference source not found.) implemented their own vehicular mobility tools or alternatively used un-calibrated traffic simulation models. The disadvantage of this approach is that by constructing new mobility tools, the powerful traffic simulation algorithms in the existing packages are not used. Also the mobility modeling may not be accurate.

Microscopic traffic simulation is a very detailed representation of vehicular flow using many complex mathematical equations. When it is used to represent traffic flow on a certain roadway, it is not only important to represent features such as geometry, number of lanes, etc. accurately, but also the outputs such as speed, traffic flow, travel time, etc. Adjusting the complex model parameters in the model to obtain the outputs accurately forms the process of calibration. The advantages of using a calibrated vehicular mobility model are the following:

1. Accurate representation of traffic metrics such as flow, speeds, travel time (current/future)
2. Accurate prediction of spatial and temporal distribution of congestion
3. Estimate impact on network layer performance under real-world conditions

Using an uncalibrated model could result in inaccurate findings. Overestimation of congestion due to use of an uncalibrated model results in, (a) unrealistic overload of network bandwidth and, (b) unrealistic excessive route changing by drivers. On the other hand, underestimation of congestion due to use of an uncalibrated model results in, (a) unrealistic extenuating congestion and unsafe conditions (b) incorrect prediction of (lack of) benefits VANETs.

Secondly, as mentioned in the earlier section, integrating V2V with V2I will ensure that the load on the communication infrastructure will be lower. V2V and V2I have been integrated in some of the previous studies; however they have been done for the purpose of situational awareness, safety applications or network security, which are applications on a smaller scale. For vehicular sensor applications there is a constant occurrence of name-address mapping. For larger areas, due to the increase in data needs and vehicles periodically seeking a connection, the communication network would get inundated.

Integration of V2V and V2I has been theorized (18) and implemented on a smaller scale and less sophisticated vehicular mobility model (Error! Reference source not found.). A peer-to-peer architecture on a mobile internet platform was evaluated by (19). (Error! Reference source not found.) proposed a method of grouping vehicles based on fixed roadway and time segments, which may not be ideal since the vehicular density could be highly variable based on time of day and unexpected conditions (weather, accidents, etc.).

Thus the motivation for this study is:
i. To develop a systematic aggregation approach of vehicle information to reduce the number of vehicles that access the communication network,

ii. evaluate the scalability of the systematic data aggregation and name-address mapping using the developed simulation test bed without loss of data quality.

The means to achieve the objective of aggregating vehicle information is by forming groups of equipped vehicles. The vehicles within a group exchange information with one another and only one chosen vehicle (called the leader) exchanges the information with the communication network.

In this study our major contributions are,

i. Implementation of the methodology for formation of a disjoint collection or group of VANET nodes dynamically proposed by (21,22) in a realistic well-calibrated microscopic traffic simulation of the New Jersey Turnpike (NJTPK).

ii. Demonstration of benefits of dynamic grouping of vehicles for sensor data collection in terms of 70-90% reduction in communication load using the realistic microscopic vehicular simulation test bed.

iii. Scalability of the methodology for dynamic grouping of vehicles when compared with vehicles connecting independently to the communication network. This result suggests that the name-address mapping is also scalable.

iv. The scalability and reduction in bandwidth are obtained without loss in data quality in speed (5.5-8%) and travel time (5-10%)

MODELING METHODOLOGY

Mobility Tracking Load Modeling and Measurement

An important and basic step involved in modeling vehicular communication and its application to sensor data collection is modeling the infrastructure host or servers, also called as road-side units (RSU). Note that these servers could also be cell phone towers, if the communication medium is a mobile phone. However, we use RSU as a generic term to denote a communication infrastructure server. The RSUs are used to collect and consolidate the data accumulated by various instrumented vehicles. This is also called as vehicle-to-infrastructure (V2I) communication. Hence analysis of vehicular communication would entail the measurement of the load on the RSUs. Unlike, the instrumented vehicles, topology of RSUs is fixed. The road network can be divided into various regions based on the coverage area of the RSUs. If cellular communication towers are used as RSUs the coverage area can be large.

When a vehicle is in the area of coverage of an RSU, then the instrumented vehicle is assumed to be able to communicate with the RSU without interruption. The instrumented vehicles within the coverage area of an RSU will upload the probe data to the RSU. The assumptions and parameters used in the mobility load measurement are shown below. These are only simplifying assumptions and the findings of this study will not be affected much with these assumptions (similar parameters have been used in (23)):

1. Vehicles are assigned a first upload time from a uniform distribution so that the uploading events can be distributed (in time) as uniformly as possible
2. Upload frequency = 1 s.
3. Upload time of each packet is one simulation time step (0.1s)
4. Packet size = 1kB

Simulation Test Bed
A 25-mile section of a much larger calibrated microscopic simulation model of the New Jersey Turnpike (NJTPK) in PARAMICS has been used during the PM peak period.

(24,25) The model has been calibrated using parameters such as travel time, section volumes. It has been shown that the average of relative error in volumes is between 8-11% of the observed data for AM and PM peaks. 97% of travel times were within 10% of observed data. This model has been applied for various other applications such as estimating travel times and delays (24) and changes in emission levels due to introduction of time-of-day pricing (25). A snapshot of the network can be seen in Figure 1.

In this study, a powerful microscopic simulation tool, PARAlell Microscopic Simulation (PARAMICS) is used to model the vehicular mobility. There are many features in PARAMICS that are useful in this study such as,

- Model individual vehicle movement at the most disaggregate level
- Provide dynamic network information such as level of congestion, travel time, etc.
- PARAMICS with Application Programming Interface (API)
  - Override or extend the default models such as car following, lane changing, route choice, etc.
  - Incorporate customized functionalities
- Test and validate new models before field testing
Figure 1 Screenshot of the study area

Dynamic Grouping of Instrumented Vehicles

The first sub-section describes the modeling and analysis of V2I communication. Depending on the type of data exchanged between the RSU and the vehicles, the usage and load on the RSU can vary. In order to make optimal use of vehicular communication and to ensure a better spread of information, V2I has to be combined with V2V communication. Also, integrating V2V with V2I will ensure that the load on the RSU infrastructure will be less, at least for the type of communication, such as vehicular sensor data application, which is evaluated in this study. Also, unlike the case of other type of communication such as location-based services and advertising, the frequency of vehicles accessing the RSU infrastructure is much higher in the case vehicular sensor data.

Since the objective is to reduce the number of vehicles that access the RSUs, groups of equipped vehicles are to be formed dynamically. The vehicles within a group exchange information with one another and only one chosen vehicle (called the leader) exchanges the information with the communication network. There are various dynamics involved in the formation of these groups. The relative location of the vehicles is constantly changing; vehicles can enter and exit the highway where the groups are being formed dynamically. Hence the procedure involved in grouping equipped vehicles is far from trivial. For this study the forest forming algorithm proposed by (21,22) is implemented for grouping. A pictorial illustration of the grouping algorithm is shown in Figure 2.
The pseudo code for the grouping algorithm can be seen in Figure 3. Through periodic beacon messages including the sender’s GPS location, every car knows the location of other cars in its proximity. Initially, Step 2-1 finds $L$ as the set of all the other cars $c_l$ within $R$-radius. Step 2-5-1 chooses either back or front with probability 1/2. If two adjacent cars agree, i.e., their chosen directions direct toward each other, then they merge into a group of level 1 and size 2 whose leader is the front car. The back car sets $\text{IsLeader} = \text{false}$ to terminate its process. Among $c_l \in L$ such that $\text{IsLeader} = \text{true}$, suppose as in the figure that two adjacent leaders of level 1 agree to merge. Then the leader in the back sends its aggregate probe data to the front leader, terminating its process. This creates a new group of level 2 and the link $(c; c_l)$ in the spanning forest. (21,22)

In order to simplify the communication modeling, we assumed that once a singleton instrumented vehicle is within the communication range of an instrumented vehicle which is a ‘leader’, then the exchange of information takes one simulation time step to complete. We recognize that the communication is influenced by aspects of channel propagation such as path loss, signal fading, etc. (26). However, since the test bed is that of a fairly linear highway, we assumed that a line-of-sight radio link is sufficient for communication. Also the signal can be augmented by increase in power and adding multiple frequencies. Since the application is that of probe vehicle data transfer, once the vehicle connects to a leader and transfers its location data, the need for a communication link is not necessary. Hence the scope for multi-path interference can be assumed to be minimal.
Algorithm FormForest

Key Variables
- $c$: the current vehicle.
- $IsLeader$: true if and only if $c$ is a group leader.
- $dir(car) \in \{ \text{front, back} \}$: merging direction of the argument car.
- $GroupSize$: the number of vehicular nodes in the car group.
- $level(car)$: recursion level of the group the argument car belongs to.
- $ElapsedTime$, $MaxElapsedTime$: elapsed time and its upper bound.
- $MaxInitialDistance$: maximum distance to another leader to join initially.
- $L$: the set of all the group leaders in the radio transmission range of $c$.
- $c_l \in L$: a group leader in $L$.

begin
/* Every car $c$ in the network runs the following algorithm. A global clock starts it at the same time. Assume $c$ knows its GPS location. */
/* Every time a car sends information, it is attached with its GPS location. When $c$ hears from another car, it knows its location; thus the closest neighbors $c$ both in the back and front. */
1. $IsLeader = true; GroupSize = 1; level(c) = 0; ElapsedTime = 0; /* Initialize the basic variables when $c$ starts the algorithm. */
2. while $IsLeader = true$ and $ElapsedTime \leq MaxElapsedTime$
do
2-1. $L$: the set of all group leaders $c_l$ in the radio transmission range of $c$, i.e., nearby vehicular nodes $c_l$ such that $IsLeader = true$;
2-2. if $L = \emptyset$ then send the aggregate probe vehicle data owned by $c$ to the infrastructure and set $IsLeader = false$ (to terminate the algorithm);
2-3. else /* The current car $c$ merges to its front or back. */
2-3-1. $dir(c) =$ value randomly chosen from $\{ \text{front, back} \}$;
2-3-2. $c_l =$ the leader in $L$ closest to $c$ in the merging direction $dir(c)$ if exists; otherwise $c_l =$ the leader in $L$ closest to $c$;
2-3-3. if $(dir(c) = \text{back})$ and $(dir(c_l) = \text{front})$ then /* $c$ is ahead of $c_l$. */
2-3-3-1. Receive the probe vehicle data of $c_l$ and update those of $c$;
2-3-3-2. $c$ has become the parent of $c_l$ in the forest remaining as a group leader. It aggregates the probe data of $c$ with those of $c_l$.
    Also, update the aggregation level $level(c)$ as follows. /*
    2-3-3-4. else if $(dir(c) = \text{front})$ and $(dir(c_l) = \text{back})$ then /* $c$ is behind $c_l$. */
    2-3-4-1. Send the probe vehicle data of $c$ to $c_l$;
    2-3-4-2. $IsLeader = false$;
    /* $c$ has become a child of $c_l$ terminating the process. */
2-4. end if
2-5. Update $ElapsedTime$ correctly by the global clock;
3. end while

Figure 3 Pseudo code for the grouping algorithm (21,22)
MICROSCOPIC SIMULATION RESULTS

The dynamic grouping methodology is implemented in microscopic traffic simulation model using PARAMICS for the application of sensor data collection for two scenarios,

i. A hypothetical linear highway of length 5 km and,

ii. A 25-mile section of a well-calibrated realistic model for NJTPK for the PM peak period

Dynamic Grouping Implementation for Hypothetical Linear Highway

A two-lane linear highway of length 5 km with an RSU every 500 m is the configuration of the hypothetical scenario. The grouping algorithm is implemented as mentioned in (21,22). According to this algorithm the leader of the group to which an equipped vehicle connects is chosen only from the lanes left of the equipped vehicle. It should be noted that the larger the size of the equipped vehicle group the lesser is the demand on the RSU infrastructure. A modification in the implementation is made by choosing the leader of the group to which an equipped vehicle connects from all possible lanes in order to increase the group size under lower density of equipped vehicles. The set of parameters used in the implementation are shown in Table 1. The metrics used to evaluate the performance of this algorithm are number of times the RSU infrastructure is accessed and size of the equipped vehicle group.

With a demand on the highway of 1750 veh/hr/lane, the density of instrumented vehicles is 0.015/meter. The performance of the methodology is estimated by the comparison of the number of accesses between a scenario where there is no grouping and other scenario with the grouping performed using forest forming algorithm (Figure 2). We found that when there is no grouping of equipped vehicles the communication load to be 0.06 megabits per second (Mbps), whereas with grouping the same was 0.01 Mbps. Also average group size when using grouping algorithm is 6.95 vehicles/group.

Table 1 Relevant Parameters for the Grouping Algorithm used in each scenario

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Hypothetical Linear Highway</th>
<th>NJTPK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section Length</td>
<td>3.13 mi (5 km)</td>
<td>25 mi</td>
</tr>
<tr>
<td>Area of Coverage of RSU</td>
<td>0.31 mi (500 m)</td>
<td>5 mi</td>
</tr>
<tr>
<td>Market Penetration</td>
<td>25%</td>
<td>5%, 15%</td>
</tr>
<tr>
<td>Upload Frequency</td>
<td>1 Hz.</td>
<td>1 Hz.</td>
</tr>
<tr>
<td>Upload Time</td>
<td>0.1 s</td>
<td>0.1 s</td>
</tr>
<tr>
<td>Packet Size</td>
<td>1 KB</td>
<td>1 KB</td>
</tr>
</tbody>
</table>

Dynamic Grouping Implementation for NJTPK Network (Exit 9-14)

The dynamic grouping methodology explained in the previous section is implemented for
the 25-mile section of a calibrated microscopic simulation model of NJTPK during the
PM peak period (shown in Figure 1). In addition to the modification mentioned above,
we use a three-hop communication approach to increase the equipped vehicle group size.
The set of parameters used in the implementation are shown in Table 1.

The channel usage for the RSUs is aggregated for each minute. Since the coverage area
of each RSU is assumed to be 5 miles, the study area can be divided into six regions,
namely, interchange 9, 10, 11-12, 12-13, 13A-13 and 13A-14. The simulation is run for
5% and 15% market penetration and the results are based on average of seven simulation
runs.

Average and maximum channel usage per minute in the peak period for 5% and 15%
market share is shown in Table 2(a) and Table 2(b) respectively. Note that these values
are a result of data upload by instrumented vehicles on NJTPK network only and does not
include the traffic from other roads in the area.

**Table 2 (a) Average channel usage (Mbps) with and without grouping**

<table>
<thead>
<tr>
<th>RSU</th>
<th>5% Grouping</th>
<th>5% Reduction</th>
<th>15% Grouping</th>
<th>15% Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int. 9</td>
<td>3.08</td>
<td>0.67</td>
<td>78%</td>
<td>8.94</td>
</tr>
<tr>
<td>Int. 10</td>
<td>2.69</td>
<td>0.75</td>
<td>72%</td>
<td>7.79</td>
</tr>
<tr>
<td>Int. 11-12</td>
<td>2.65</td>
<td>0.39</td>
<td>85%</td>
<td>7.97</td>
</tr>
<tr>
<td>Int. 12-13</td>
<td>3.12</td>
<td>1.01</td>
<td>68%</td>
<td>9.25</td>
</tr>
<tr>
<td>Int. 13A</td>
<td>2.01</td>
<td>0.46</td>
<td>77%</td>
<td>6.08</td>
</tr>
<tr>
<td>Int. 14</td>
<td>6.18</td>
<td>2.11</td>
<td>66%</td>
<td>18.17</td>
</tr>
</tbody>
</table>

**Table 2 (b) Maximum channel usage (Mbps) with and without grouping**

<table>
<thead>
<tr>
<th>RSU</th>
<th>5% Grouping</th>
<th>5%</th>
<th>15% Grouping</th>
<th>15%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int. 9</td>
<td>4.31</td>
<td>1.05</td>
<td>10.75</td>
<td>1.30</td>
</tr>
<tr>
<td>Int. 10</td>
<td>3.49</td>
<td>1.22</td>
<td>9.10</td>
<td>1.97</td>
</tr>
<tr>
<td>Int. 11-12</td>
<td>3.46</td>
<td>0.62</td>
<td>9.73</td>
<td>1.06</td>
</tr>
<tr>
<td>Int. 12-13</td>
<td>4.13</td>
<td>1.42</td>
<td>10.95</td>
<td>3.58</td>
</tr>
<tr>
<td>Int. 13A</td>
<td>2.89</td>
<td>0.70</td>
<td>7.50</td>
<td>1.26</td>
</tr>
<tr>
<td>Int. 14</td>
<td>7.47</td>
<td>2.66</td>
<td>20.74</td>
<td>5.19</td>
</tr>
</tbody>
</table>

From the results (Table 2(a) and Table 2(b)) it can be seen that there is a significant
reduction in the average channel usage with grouping. The reduction is between 68-85%
for 5% market penetration and 72%-91% for 15% market penetration.

Another inference from Table 2(a) and Table 2(b) is that the name-address mapping
involved in the continuous location reporting is well within the range of cell tower
channel capacity. The maximum channel usage with 5% market penetration is 7.47 Mbps
without grouping and 2.66 Mbps with grouping. The same for 15% market penetration
are 20.74 Mbps and 5.19 Mbps. Hence it can be said that the name-address mapping is
Scalability of Forest Forming Algorithm

With increasing use of technologies in vehicles, the market penetration of equipped vehicles will only increase. Similarly, the improvement in DSRC communication is set to increase the power of the DSRC radios. Hence, in order to investigate the scalability of the algorithm for different market penetration of technology and communication ranges, a sensitivity analysis of the communication infrastructure load is performed. The communication infrastructure load is estimated for market penetration of 2.5%, 5%, 7.5%, 10%, 12.5% and 15% and communication range of 200m, 400m, 600m, and 800m. Additionally, the communication infrastructure load in a scenario where there is no grouping and each vehicle independently connects is also included.

Figure 4(a) shows the response surface for communication channel load for RSU located between interchanges 13A and 14. The communication channel load when there is no grouping can be seen towards the extreme left of the surface shown. The increase in communication channel load with no grouping with increasing market penetration is indicated by the slope of the red dotted line on the left. The rate of increase in load can be quantified as 1.22 Mbps/% increase in market penetration.

The communication channel load with grouping is shown with communication range increasing from left to right in Figure 4(a). The increase in communication channel load with grouping with increasing market penetration is indicated by the slope of the blue solid line on the right. The rate of increase in load can be quantified as 0.23 Mbps/% increase in market penetration, which is an 80% reduction when compared to the scenario with no grouping. Thus the difference in slopes of the red and blue lines distinctly shows the scalability of the forest forming algorithm.
Figure 4 (a) Response Surface for Communication Channel load for RSU located between Interchanges 13A and 14 (b) Response Surface for Communication Channel load for RSU located between Interchanges 13A and 14 with only vehicular grouping

The same response surface with channel load under only vehicular grouping is shown in Figure 4(b). The communication channel load decreases with increasing market penetration. Additionally, it can be seen from Figure 4(b) that increase in communication range reduces the communication channel load.

A similar response surface for communication channel load for RSU located between interchanges 12 and 13 is shown in Figure 5. The communication channel load when there is no grouping can be seen towards the extreme left in the surface shown. The increase in communication channel load with no grouping with increasing market penetration is indicated by the slope of the red dotted line on the left. The rate of increase in load can be quantified as 0.62 Mbps/% increase in market penetration.

The communication channel load with grouping is shown with communication range increasing from left to right in Figure 5. The increase in communication channel load with grouping with increasing market penetration is indicated by the slope of the blue
solid line on the right. The rate of increase in load can be quantified as 0.15 Mbps/%
increase in market penetration, which is 75% reduction when compared to the scenario
with no grouping.

Figure 5 Response Surface for Communication Channel load for RSU located
between Interchanges 12 and 13
The group size increases with increase in market penetration and communication range.
This is the reason for significant reduction in RSU channel load with increasing market
penetration and communication range.

Validation of Quality of Reported Data
The clear benefit of dynamic grouping of vehicles is a much lower bandwidth usage.
However, the quality of reported data should also be maintained in order to conclude that
the dynamic grouping methodology has a clear advantage over individual probe data
collection. Quality of data represents how closely the aggregated vehicular sensor data
represents the actual traffic conditions.

For this purpose the data that is reported under the two scenarios with 10% market
penetration are compared: (a) speed data reported to RSU with individual probe vehicles,
and (b) speed data reported to the group leader with dynamic grouping methodology (400
m communication range). The average speed reported between 3:00 PM – 5:00 PM is
estimated using each approach is compared and mean absolute percent error (MAPE) is
estimated. This MAPE averaged for each link using the dynamic grouping methodology
for each 15-minute time interval is between 5.5-8.0%. Additionally, we also computed
the standard deviation of difference in mean link speeds estimated using individual probe
data and dynamic grouping approach and found it to be between 1.76 and 2.25 mph.

Note: The probe data from individual vehicles (without grouping) is averaged using the
data uploaded by each equipped vehicles every second. With the grouping methodology,
however, each ungrouped vehicle uploads data every second. Each grouped vehicle
transfers the data collected until before grouping and transfers the aggregated data packet
to the group leader, which in turn uploads an aggregate of its own and new group
members’ data every second. Since there are a high number of data points with probe
data individual vehicles, it is treated as baseline for comparison.

An important application of probe vehicle data is to estimate travel time between origins
and destinations, which may have more than one path. The travel time on each path
should be accurate in order to validate the benefit of the dynamic grouping methodology.
Based on the speed and location data reported the speed profile can be estimated for all
the links along the path. Travel time is calculated by the summation of link travel times.

For estimating travel times of multiple paths, the extent of the network is expanded to
incorporate an urban network of Jersey City, adjacent to the NJTPK. The three most
commonly used paths (Figure 6) between Holland Tunnel and Newark are used to
compare the speed profiles and the travel times.

Figure 6 Three Paths used for Travel time Evaluation

The speed profiles (Figure 7) using data from dynamic grouping methodology closely
follow that from individual probe data. Travel time estimates and bandwidth used for the
estimate for each path averaged for the six 15-minute time intervals is shown in Table 3.
It can be seen that the travel time using dynamic grouping approach is less than 4.5%
within the travel time from individual vehicles’ probe data, however, using 80-90% less
bandwidth.
Figure 7 Speed Profile along Different Paths

Table 3 Comparison of the Average Travel Time and Bandwidth Usage

<table>
<thead>
<tr>
<th>Path</th>
<th>No Grouping</th>
<th>Dynamic Grouping Methodology</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Travel Time (min.)</td>
<td>Average Bandwidth (Mbps)</td>
<td>Average Travel Time (min.)</td>
</tr>
<tr>
<td>1</td>
<td>16.04</td>
<td>3.70</td>
<td>16.39</td>
</tr>
<tr>
<td>2</td>
<td>16.27</td>
<td>3.67</td>
<td>16.98</td>
</tr>
<tr>
<td>3</td>
<td>18.94</td>
<td>3.53</td>
<td>19.31</td>
</tr>
</tbody>
</table>

Probe data is also useful in estimating travel times in the event of non-recurrent congestion such as inclement weather or an accident. To replicate a non-recurrent congestion scenario, an incident of 20 minute duration at 3:45 PM along path 1 (please see Figure 6) is simulated. Estimated travel times on each path from dynamic grouping are almost always less than 10% and less than 5% on an average within the individual...
probe vehicle travel times. (Unfortunately, due to space constraint we are not able to present all the details of travel time comparison during an incident here)

CONCLUSIONS

In this study, a well-calibrated microscopic traffic simulation test bed of NJTPK in PARAMICS is used to simulate the vehicle mobility to measure the communication channel load in receiving location information from instrumented vehicles. PARAMICS allows for modifying existing algorithms, collect very detailed data and test numerous alternatives. We implement the methodology proposed by Li et al. (2012) for formation of disjoint groups of VANET nodes dynamically in the test bed. It is important to implement the methodology on a well-calibrated model, since it captures flows and congestion very accurately.

For the purpose of vehicular sensor data collection, the load on the RSU is an important factor in determining the investment in V2I infrastructure. The dynamic grouping methodology proposed integrates V2V with V2I to significantly reduce the load on communication infrastructure. We show that this aggregation of vehicle information will result in reduction in communication infrastructure load. Reduction of 66-85% for 5% market penetration and 72%-91% for 15% market penetration was observed for NJTPK. These reductions are only possible because of the large groups formed using the dynamic grouping methodology.

We also show that the name-address mapping involved in the continuous location reporting is well within the range of cell tower channel capacity. The name-address mapping has been tested on a realistic test bed and maximum bandwidth usage is used as a measure to show that it is scalable. However, due to the large number of vehicular nodes, the communication side of the name-address mapping is not accounted for in this study. The maximum channel usage with 5% market penetration is 7.47 Mbps without grouping and 2.66 Mbps with grouping. The same for 15% market penetration are 20.74 Mbps and 5.19 Mbps.

We show that with the dynamic grouping methodology the communication load is much more scalable when compared with vehicles connecting independently to the communication network. For this purpose we developed response surfaces for channel load on each RSU for different combinations of market penetration and communication range. The scalability can be measured as the rate of increase in load with percent increase in market penetration. This measure for the RSU channel located between interchange 13A and 14 is found to be 1.22 Mbps/% increase in market penetration without any grouping of vehicles. When the vehicles are grouped the scalability measure is as low as 0.23 Mbps/% increase in market penetration, which is an improvement of 80%. Similarly, the improvement in scalability measure for the RSU channel located between interchange 12 and 13 is found to be 75%. Additionally, these response surfaces will also be useful in predicting the infrastructure load under future scenarios of increasing market penetration and power of DSRC radios.

Despite using much lesser bandwidth, the dynamic grouping methodology maintains the
data quality. We validate this by using average reported speed from over the NJTPK simulation network. The percent error in reported speed for dynamic grouping approach when compared to the individual probe data is only 5.5-8%. This accuracy is obtained with much lesser bandwidth used for the dynamic grouping methodology. Similarly travel time estimated using speed data from probe data is within 5% of the travel time from individual probe data during normal traffic conditions and 10% during non-recurrent congested conditions.

In this study we show the benefits of using the dynamic grouping of equipped vehicles for the application of sensor data collection. However, the benefits of reduction in communication bandwidth usage are applicable to other V2V/V2I applications. The evaluation of dynamic grouping for other applications forms a part of our future work. As a part of improving the methodology, we modified the grouping methodology in order to sustain larger group sizes. (27) Also we evaluated the algorithm, as a proof-of-concept, for the application of the dynamic grouping methodology for efficient broadcast of vehicular situational awareness data such as safety data on a highway section. (27) As part of future work, we are working on periodic refreshing of the grouping methodology to create a framework for data broadcast to vehicular nodes constantly and accurately. Also the simulation of communication channel for V2V communication will be improved.

ACKNOWLEDGEMENTS
This study was partially conducted with support from Toyota InfoTechnology Center to whom the authors are grateful.

REFERENCES


3 Spanning Forest in a Vehicular Network - with Application in Probe Data
4 Aggregation and Upload", Presented at the 91st Annual Meeting of
8 23. Zhao, J., and G. Cao, "VADD: Vehicle-Assisted Data Delivery in Vehicular Ad
9 Hoc Networks", Proceedings of 25th IEEE International Conference on Computer
10 Communications, INFOCOM 2006
12 of an Integrated Freeway and Toll Plaza Model." Presented at the 85th TRB
13 Annual Conference, 2006, Washington, D.C.
15 Toll Collection on Air Pollution Levels using Microscopic Simulation Model of a
16 Large-Scale Transportation Network.” In Transportation Research Record:
20 27. Mudigonda, S., J.Fukuyama, K.Ozbay, "Efficient Data Broadcast for Generic
21 Applications using a Scalable Dynamic VANET Algorithm", Accepted for the
22 International Conference on Connected Vehicles and Expo, Beijing, China, Dec.