Randomization in Traffic Information Sharing Systems

Masaaki Tanizaki[†‡]
[†] Intelligent Media Systems Research Dept.
Hitachi, Ltd., Central Research Laboratory
1-280, Higashi-koigakubo
Kokubunji-shi, Tokyo 185-8601, Japan
masaaki.tanizaki.tj@hitachi.com

Ouri Wolfson[‡]
[‡] Department of Computer Science,
University of Illinois at Chicago
851 S. Morgan Street,
Chicago, IL, 60607, USA
wolfson@cs.uic.edu

Abstract

In this paper, we consider a traffic information sharing system based on Floating Car Data (FCD). FCD is one of the methods used to gather traffic information; it uses vehicles as sensor nodes that transmit their speed to the server. The traffic information sharing system broadcasts speed information updated by such transmission. Vehicles receiving broadcasted speed information can calculate travel time and select a minimum time route. The communication cost and the load of the server are issues, because such a traffic information sharing system can generate a lot of wireless communication between the vehicles and the server [2][3]. However, reducing the amount of communication lowers the accuracy of information provided by the server. In this paper we propose an Information Cost Model to quantify a trade-off relationship between the communication cost of the system and the accuracy of information. Additionally, we propose a randomized method to reduce the number of messages from clients to the server by avoiding redundant transmissions. We compared the performance of our proposed method with that of a conventional method, using real traffic data from Chicago highways. The result shows that our proposed method generally outperforms the conventional method.

Categories and Subject Descriptors:

D.2.8 [Metrics] : Performance metrics

General Terms

Algorithms, Measurement, Performance, Experimentation

Keywords

Information Cost Model, Randomized Policy, Traffic Information, Floating Car Data,

1. Introduction

1.1 FCD (Floating Car Data)

In order to improve the efficiency of urban transportation, Floating Car Data (FCD) techniques have been researched and developed [1]. FCD is one way to gather traffic information using floating cars as sensor nodes which have a location detecting device such as a GPS and a communication device such as a cellular phone. These traffic-information services based on FCD are already provided in Japan and the EU (see [2][3]).

In traffic information sharing systems using FCD techniques, vehicles send a measured velocity to the server. The server manages the real time traffic information of each road segment and delivers the traffic information to vehicles in the service area by continuously broadcasting the current speed on each road segment. Vehicles receiving the traffic information can calculate their travel time and also the minimum-time route to their destination based on the road network data and the broadcasted traffic information.

1.2 Problem and Proposed Solution

Generally, it is necessary to update the traffic information database (residing on the server) by a significant number of transmissions from vehicles. However, in metropolitan areas with millions of vehicles there are problems of the communication cost and load on the server to process these transmissions. These problems would be a barrier when expanding the traffic information sharing system based on FCD. Furthermore, indiscriminately reducing the number of transmissions from vehicles to the server lowers the accuracy of information provided by the server.

However, observe that an FCD traffic information sharing system has a lot of redundant transmissions; the reason is that multiple vehicles going through some road segment at around the same time measure the same speed and transmit the same information to the server. This problem increases in severity with the server delay (i.e. the time it takes from the transmission of a velocity update by a vehicle to the server, until this update is propagated to the vehicles in this service area). To address the problem, in this paper we propose a randomized update policy in which the vehicles transmit their measured speed with a certain probability smaller than 1. In other words, to determine whether or not to send the measured speed to the server each vehicle tosses a coin with probability p, and transmits the speed only if head comes up.

An important problem regarding the randomized update policy is how to determine the transmission probability p. To solve this problem, in this paper we consider how to achieve the maximum accuracy by the minimum number of data transmissions from vehicles to the server. Therefore, we quantify a trade-off relationship between the transmission cost in the whole system and the accuracy of information shared by clients. We do so by developing an Information Cost Model.

ACM GIS '07, November 7–9, 2007, Seattle, Washington, USA.
Copyright 2007 ACM ISBN 978-1-59593-914-2/07/11...$5.00

We evaluate the randomized policy and compare it with the deterministic policy by simulation. The simulation uses real traffic data from Chicago highways. The main result (Figure 6) shows that for a given accuracy (or error), the randomized policy uses fewer transmissions than the deterministic policy. More precisely, the communication savings of the randomized policy can be as high as 72%, with the actual percentage depending on the server delay and the accuracy.

1.3 Related Work

There are different ways to build a traffic information sharing system based on FCD. First, from the point of view of system architecture, one way uses a client/server architecture, where vehicles as clients send the velocity that they measure to their server individually. This informs the server of the real time traffic condition of each road segment. Then the server broadcasts updated velocities of each road segment to all vehicles.

The other way is based on a mobile peer to peer (MP2P) architecture, where vehicles send velocities of road segments to each other without server facilities. In this paper we consider the client/server architecture.

Kerner et. al. ([4]) have developed a traffic information sharing system using a client/server architecture, where the server broadcasts velocities with threshold values; each vehicle sends an update to the server when its measured speed on a road segment differs from the broadcasted speed by an amount larger than the threshold.

Goel et. al. ([5]) have considered and prototyped a system based on a mobile P2P architecture, where each vehicle sends measured velocities to other vehicles, updating each other using a wide range peer to peer communication method like SMS.

Shinkawa et. al. ([6]) have considered a traffic information sharing system based on mobile P2P extended by using buses running along fixed routes as "ferries" which transport the traffic information to disconnected groups of clients.

Masutani et. al ([7]) have proposed the pheromone model which is designed for mobile P2P sharing of traffic information. It expresses the congestion weight as the amount of pheromone accumulated by the short range communication among vehicles.

The above research shows the basic architecture of traffic information sharing systems, however it does not address the trade-off relationship between the communication cost and the accuracy of the information shared.

Then there are other approaches that reduce the communication cost by reducing the transmission frequency or the data volume. For example, Adachi et. al. ([8]) have developed a compression method of trajectories based on the decomposition of temporal and spatial component, and discrete wavelet transformation of the temporal component. Horiguchi et. al. ([9]) have proposed a data reduction method that is based on the extraction of portions of trajectories classified as "congestion". The classification of a sub-trajectory as congested is based on the idea that during congestion the trajectory consists of consecutive short term stop and run sections. Civilis et. al. ([10]) have developed a data reduction method based on the extraction of velocity change-points; it uses a profiled pattern of acceleration changes on a routine route. Basically, this research proposes compression methods of trajectories. However, this research does not provide a metric to indicate how much the reduction of transmission impacts the accuracy of information. Furthermore, compression methods are orthogonal to the randomized policy proposed here, and they can be combined; although, how exactly to do so remains the subject of future research.

In prior work we have researched the Information Cost Model in the context of tracking moving objects [11]. In order to estimate and compare a number of policies for predicting the location of moving objects we proposed the Information Cost Model consisting of the communication cost, the deviation cost, and the uncertainty cost. In this paper, we adapted and applied the concept of the Information Cost Model to express the trade-off relationship between the communication cost and the accuracy of the FCD information. Additionally, we propose the randomized policy to reduce redundant transmissions, and demonstrate the superiority of the policy experimentally, using real traffic data.

The rest of this paper is organized as follows. In Section 2 we introduce the deterministic and randomized update policies. In Section 3 we propose the Information Cost Model in order to derive the optimal probability of the randomized policy. In Section 4 we compare experimentally the update policies by using real traffic information. In Section 5 we conclude and discuss future work. In appendix A we provide a table of notations and terms used in this paper.

2. FCD Update Policies

In this section we describe two policies by which FCD vehicles can update the server: the deterministic policy and the randomized one.

2.1 Deterministic Policy

In order to explain a traffic information sharing system using the FCD technique, let's assume a system as follows. The server and vehicles have the same map that consists of road segments, identified individually by id. At each point in time, each road segment has a velocity v_m that represents the current speed on the segment. Each vehicle has a table storing the velocity of each road segment, and so does the server. The vehicles traversing a segment know v_m on that segment, and they inform the server so it can broadcast it to all vehicles. Clearly, this velocity is of interest to vehicles that are not on the road segment. Due to synchronization issues, the broadcasted velocity, denoted v_b, is not always identical to the current velocity v_m; sometimes it lags behind. Observe also that the velocities are given at the road-segment granularity. For all practical purposes such granularity is sufficient.

In order to update the vehicles' velocity information dynamically, the server continuously broadcasts the velocity v_b of each segment. v_b is updated by transmissions from the vehicles traversing that segment. Each vehicle receives a broadcasted velocity on each road segment, and also sends a velocity v_m to the server for each road segment it traverses. v_m is the average velocity on the road segment, as measured when the vehicle reaches the end of the segment. v_m is computed by simply dividing the length of the road segment by the time it took the vehicle to traverse the segment.

We assume that the vehicles have a location detecting device such as GPS, a transmission device such as a cellular phone, and a broadcast receiver such as a radio tuner. The location detecting device is used for determine the position on the one of the road segment in the map data. To simplify the explanation, we focus on the scenario treating only one road segment in the map.

In order to prevent small differences in velocity from being transmitted to the server, a velocity threshold is used (see [4]). We denote by T the velocity threshold, and then the *transmission rule* that permits the vehicles to send v_m to the server is expressed as follows:

$$| v_m - v_b | \geq T \quad (1)$$

In other words, the transmission rule indicates that the vehicle transmits v_m to the server if and only if the difference between the broadcasted velocity (received from the server and stored in the vehicle for the road segment) and the measured velocity exceeds T. We call the policy using a particular threshold T as the *deterministic policy*.

2.2 Randomized Policy

In order to motivate the randomized policy, let us explain the problem of redundancy in communication for the deterministic policy. Intuitively, this redundancy is a result of the fact that there is a lag between the point in time t when the first vehicle detects that the threshold T is exceeded and transmits its v_m to the server, and the point in time when the transmitted speed v_m becomes the speed v_b broadcasted by the server. This lag is called the *server delay*. During the server delay, all the vehicles that pass through the end of the road segment will have the transmission rule satisfied, and thus transmit (almost) the same v_m to the server (assuming that the average velocity on the road segment is stable during the server delay). Observe further that if the server delay is zero, then the vehicles that cross the end of the road segment after time t will have the new velocity, and thus they will not transmit anything to the server, eliminating the redundancy. However, a zero-delay server is unlikely.

In order to solve this problem, we propose the *randomized policy* which uses a randomization function for avoiding redundant transmissions. basically, the randomized policy is the same as the deterministic policy, except that when the transmission rule is satisfied, instead of transmitting v_m to the server with probability 1, the randomized policy does so with some given *transmission probability p*.

How does this help? Let's say that ten vehicles run through the end of a segment during the server delay τ, and the transmission rule is satisfied (with the previous v_b) for each one of these vehicles. If p=1/10, then the expected number of vehicles that will transmit v_m to the server is one; so there are no redundant transmissions.

Observe that in the above discussion we implicitly make the assumption that the time interval between two consecutive times when the threshold T is exceeded is not higher than the server delay; otherwise the system thrashes, indicating that the threshold T is too low.

Observe that if p is too small, then there will not be enough updates and the accuracy of the system will decrease; on the other hand if p is too large there will be redundant transmissions. The question now is what should p be and how it is determined. Intuitively it is clear that the optimal probability p depends of factors such as: 1) K, the number of vehicles running through the end of the segment during server delay τ, 2) the number of time units between two consecutive times the threshold is exceeded, 3) the threshold T, and so on. Furthermore, some of these parameters change dynamically, thus the optimal probability changes dynamically. In the randomized policy the following information is continuously broadcast by the server for each road segment:

- the optimal probability p for each road segment (computed by the server)
- the current speed v_b and
- u: The *uncertainty of* v_b for the segment. This value may be the velocity threshold T, or the free flow speed V. In Section 3.2 we will explain how the server chooses one of these two values.

3. Information Cost Model

In this section we introduce, analyze, and demonstrate a model that enables the computation of the optimal p for the randomized policy. In subsection 3.1 we define the model, in subsection 3.2 we apply it to compute of the information cost of the randomized policy, and in subsection 3.3 we derive the optimal probability.

3.1 Definition of Information Cost Model

Compared to the deterministic policy, the randomized policy increases the uncertainty in the FCD information system, and decreases the communication cost. Actually, the randomized policy is a set of policies, one for each transmission probability p.

In order to estimate the optimal probability p for the randomized policy, we need to quantify the trade-off between the communication cost and the uncertainty cost. The following discussion pertains to a particular road segment.

We define the total information cost $COST_I$ of an update policy to be the sum of the communication cost $COST_C$ (expressing the total amount of transmitted data) and the uncertainty cost $COST_u$ (expressing the inaccuracy (or uncertainty) of the information about the road segment).

$$COST_I = COST_C + COST_u \quad (2)$$

The communication cost is defined as the number of transmissions from a vehicle to the server per time unit. We set the unit cost of one transmission, as 1 to simplify the problem.

The intuition behind the uncertainty cost is as follows. Let's say that the server continuously broadcasts the current speed v_b on each road segment along with a threshold T; it indicates to vehicles currently on the road segment that if they experience a velocity v_m that differs from v_b by a value higher than T, they should update the server. T indicates to the vehicles currently outside the road segment, i.e. the ones interested in the velocity information for the road segment, that the velocity may change within T without the server broadcasting such change. Thus T is the *uncertainty* for the road segment.

It is clear that as the uncertainty increases, the communication cost decreases since the threshold will be reached less frequently. It is also clear that the uncertainty of the segment may change over time, if, for example, the threshold T changes.

Let U be the cost of a unit of uncertainty per time unit. If the uncertainty on the road segment is T over a time interval of length I, then the uncertainty cost for the segment over the interval is $U*T*I$. U is given in messages. So, for example, if U=0.01 this means that the system designer is willing to pay 0.01 transmissions from a vehicle to the server in order to reduce the segment-uncertainty by one unit (e.g. from T to T-1), during a single time unit.

3.2 Information Cost of Randomized Policy

In this subsection we analyze the information cost of the randomized policy. In order to do so, we consider the period of time between two consecutive times, t1 and t2, when the transmission threshold T is exceeded (see Figure 1).

The first exceeding of the threshold T on a vehicle occurs at t1. However, the transmission does not occur immediately, because the result of the randomization function may be false. Furthermore, some of the subsequent vehicles running after the first vehicle don't send their measured velocity because the result of the randomization function is not true until t1'. We denote by α the period of time in which the vehicles run through the end of a road segment without sending the measured velocity even though the transmission rule is satisfied. In other words, the vehicle that runs through the end of the segment at t1' is the first after t1 to transmit the updated v_m to the server.

Subsequently, the updated v_b is broadcasted to the all vehicles after the server delay τ, i.e. at time t1'+ τ. Between this time and t2 vehicles do not send their velocity to the server, because the transmission rule is not satisfied.

In this situation, the expected communication cost during the time period Δ+α is the following: (the number of vehicles K that run through the end of the segment during the server delay τ) x (the transmission probability p). Therefore, $COST_c$, the communication cost per time unit is:

$$COST_c = \frac{pK}{\Delta + \alpha} \qquad (3)$$

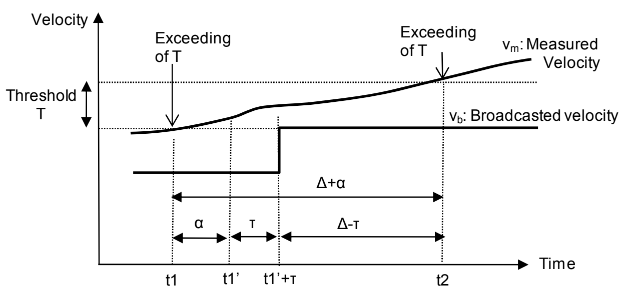

Figure 1. Sample Situation of the Randomized Policy

The uncertainty cost is computed as follows. Consider the period of time of length α+τ, starting from t1 and ending at t1'+ τ. Due to randomization and server delay, during this period of time the server does not know the measured speed. Therefore it assumes the maximum uncertainty, V, which is the free flow speed; for example, V can be taken to be 60 mph. Thus, for each time unit during this period, the cost of the uncertainty is UV.

Now consider the period of time of length Δ−τ, starting from t1'+ τ and ending at t2. During this period the vehicles know that the uncertainty is T (because this is the value broadcasted by the server) and this knowledge is accurate. Thus the uncertainty cost during this period is 2TU for each time unit. Thus $COST_u$ is as follows:

$$COST_u = \frac{UV(\alpha + \tau) + U * 2T * (\Delta - \tau)}{\Delta + \alpha} \qquad (4)$$

The above discussion of the uncertainty cost assumes that the server knows the values of α, τ, and Δ at time t1. τ is estimated by the load on the server. Since Δ changes dynamically, the server predicts its next value based on a sliding window of the last values for these parameters. For example, the server saves the last five values for each one of these two parameters (τ and Δ), and takes the next one to be the average of these. The parameter α is calculated as explained in the next subsection. Thus, the server estimates the uncertainty of v_b , denoted u, for each time unit between t1 and t2. The estimate is V between t1 and t1'+τ, and T*2 between t1'+τ and t2.

The information cost is expressed by the addition of (3) and (4) as follows:

$$COST_I = \frac{pK + UV(\alpha + \tau) + 2UT(\Delta - \tau)}{\Delta + \alpha} \qquad (5)$$

3.3 Optimum Probability

In order to determine the optimum probability of the randomized policy, we find the minimum of the cost equation (5) as a function of the probability p.

In order to do so, we need to first express α as a function of p. α is the expected length of the period of time starting when a vehicle exceeds the threshold for the first time, and ending when a vehicle transmits the revised velocity to the server. Therefore, α is calculated as the sum of the lengths (in time units) s between two consecutive vehicles, multiplied by the probability $p*(1-p)^{i-1}$. It models the situation where (i-1) consecutive vehicles get "false" as a result of the coin toss, and the i-th vehicle gets "true". Specifically,

$$\alpha = s*(p*0 + p(1-p)*1 + ... + p(1-p)^i * i + ...) = s * p \sum_{i=0}^{\infty} (1-p)^i i$$

It can be shown that the following formula holds for each real number q, where 1>|q|:

$$\sum_{i=0}^{\infty} q^i i = \frac{q}{(1-q)^2}$$

Hence, by substitution of (1-p) for q:

$$\alpha = sp * \frac{1-p}{p^2} = \frac{s(1-p)}{p}$$

Then, by substituting the above expression for α in equation (5) we obtain the cost as a function of probability p, as follows:

$$f(p) = \frac{Kp^2 + \{VU(\tau - s) + 2TU(\Delta - \tau)\}p + VsU}{(\Delta - s)p + s}$$

Next, we find p where f'(p) = 0 during (0,1], and show that f''(p) > 0. Thus we find p for which f(p) is minimum.

$$f'(p) = \frac{(\Delta - s)Kp^2 + 2Ksp + (V - 2T)(\tau - \Delta)sU}{\{(\Delta - s)p + s\}^2}$$

We can solve for the case where f'(p)=0 using the quadratic formula as follows:

$$p = \frac{-Ks \pm \sqrt{K^2 s^2 + (\Delta - s)(\Delta - \tau)(V - 2T)KsU}}{(\Delta - s)K}$$

Because p is between 0 and 1, p which satisfies f'(p) = 0 is:

$$p = \frac{-Ks + \sqrt{K^2 s^2 + (\Delta - s)(\Delta - \tau)(V - 2T)KsU}}{(\Delta - s)K} \qquad (6)$$

And there are restrictions where $0 < p < 1$ as follows:

$$U < \frac{(\Delta + s)K}{(\Delta - \tau)(V - 2T)s}, \ \Delta > s, \ \Delta > \tau, \text{ and } T < V/2 \qquad (7)$$

If these restrictions do not hold, then it is possible that the optimal p comes out to be a number that is bigger than 1 or smaller than 0. If so, then such number is turned into a probability of 1 (in case the number is bigger than1), or a small probability close to 0 (if the number is negative). Overall, the probability p shown in equation (6) is the value which minimizes the cost of information expressed in equation (5).

For the rest of this section we consider the two parameters K and s in equation (6) above. We cannot obtain these values from the vehicles directly. To determine these parameters we use a model giving the relationship between the velocity and the traffic flow (Greenshield's model [12]). This model is old but widely cited (see [13]). The relationship is given by equation (8). In this equation, r is the traffic flow, i.e., how many vehicles pass a point per time unit; thus r is the inverse of s. d is the traffic-jam density, i.e., how many vehicles fit per unit-length. v_m is the velocity. V is free flow speed of vehicles.

$$r = d * v_m (1 - v_m / V) \qquad (8)$$

Intuitively, equation (8) indicates that the flow (i.e. 1/s) is determined based on d, v_m and V. These parameters are obtained as follows. We know v_m from the vehicle. d and V are determined a priori by the least square method, using equation (8); the determination is based on real traffic data, which includes (velocity, flow) pairs for each minute. Now, given s and the server delay, it is easy to compute K.

4. Evaluation by Simulation

In order to compare the efficiency of the randomized policy with the deterministic one, we developed simulation system using real traffic data. In this section, we describe the simulation environment, and the results of the comparison.

4.1 The Traffic Data for Simulation Input

We have utilized highway traffic speed information provided by GCM travel web site [15] for evaluation of our proposed policies. This web site has operated by Illinois, Indiana and Wisconsin Department of Transportations and provides real time traffic information: e.g. travel times, congestion information, incidents, and so on as text, image and video data format.

We have used a "detector record" that is gathered from loop sensors installed in highways. The detector records are provided by this site as XML data, and they contain a data set of records, each of which consists of update time of the record, sensor device unique id, velocity, number of vehicles per lane per hour and etc.

In order to generate evaluation data for one road segment, we downloaded the records generated by the sensor at the end of a road segment on highway I90 in Chicago, for the period 6AM to 8AM on Tuesday May 22nd, 2007. The data in this period involves congestion and non congestion situations and the total number of vehicles that ran over the sensor was 6495. The sequence is $\{(t_i, v_{mi}, r_i), (t_{i+1}, v_{mi+1}, r_{i+1}), \ldots\}$, where the format of each record is as follows:

- t_i [sec] is the update-time of the record. A record is generated every minute.
- v_{mi} [m/s] is the average velocity of the vehicles running over the loop sensor between t_i and t_{i+1}.
- r_i [vehicles/lane/sec] is the number of vehicles per lane passing over the loop detector at t_i

Based on this set of records we compute the following information: how many vehicles went through the end of the road segment, at what time each one went through, and what was the speed of the vehicle at that time. In other words, we generate a *vehicles-sequence*; each vehicle in the sequence is a pair consisting of the vehicle's sensor (i.e. end-segment) crossing time, and its velocity at that time.

4.2 Procedure of the Simulation

In this subsection we describe how a simulation run was conducted. Remember that the symbols used are described in Table 1 in Appendix A.

The input to each simulation run is as follows:

- the vehicles-sequence s (see last subsection),
- a velocity threshold T used in the transmission rule,
- a server delay τ which is fixed for the duration of the simulation.
- an update policy (deterministic or randomized)

The simulation is executed by two threads, the server thread, and the client thread. The procedure of the simulation is shown in Figure 2. The server thread keeps receiving the velocities v_{mi} sent by the client thread and updates v_b by v_{mi} after the server delay. Then the updated v_b is used for continuous broadcasting to the client thread until the next update.

In the case of the deterministic policy, the client thread is executed as follows. Each vehicle sends its velocity v_{mi} to the server at the time it crosses the end of the road segment (the timing is: every $t_i + j/r_i$ during t_i and t_{i+1}), if the transmission rule $|v_{mi} - v_b| \geq T$ is satisfied; where v_b is the speed broadcasted by the server at the crossing time. T is the speed threshold that is used in the transmission rule.

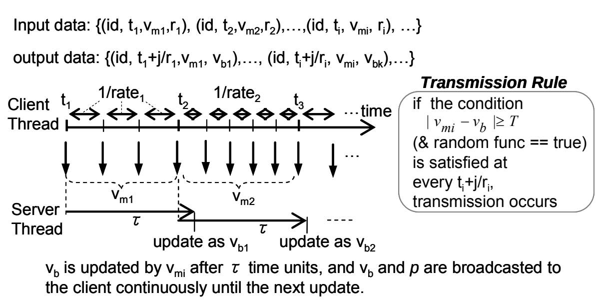

Figure 2. Simulation Procedure

The differences of the randomized policy from the deterministic policy are as follows. First, the condition [the randomization function is true (i.e. the result of the toss of a coin with probability p is heads)] is added to the transmission rule. Second, the server broadcasts to the client thread not only v_b, but also the optimized probability p. The calculation of the probability p is executed as explained in subsections 3.2 and 3.3.

Finally, the simulation output data is as follows. The vehicles thread outputs a record that consists of (time-of-sensor-crossing, v_m, v_b, and transmission-flag) when each vehicle runs through the end of the segment. The flag indicates whether or not a transmission from the vehicle to the server occurs at the time-of-sensor-crossing. The number of times a transmission occurs is counted using the transmission-flag. The simulation was executed for these ranges of parameters: T = {0.1, 0.5, 1, 2, 3, 4, 5, 6, 7}[m/s], τ = {60, 120, 180, 240, 300}[sec], V=34.8[m/s], d=0.0217[vehicles/m], U= 0.05. The size of the sliding window used to determine the size of the next Δ is five.

4.3 Comparison of the Two Update Policies

We compare the deterministic policy and the randomized policy by using two values that are measured during the 2-hour execution of our simulations. One is the communication cost, namely, the total number of transmissions from vehicles to the server which occurred during the simulation. The other is the average error, which is calculated from the difference between the velocity of road segment and the broadcasted velocity from the server. Specifically, the *average error* for a simulation run is calculated as follows. For each second of the simulation we calculate the absolute value of the difference between the broadcast velocity and the measured velocity. Then we sum these values over all seconds of the simulation (7200), and divide the sum by 7200. Figure 3 shows the measured velocity and broadcasted velocities of the randomized and deterministic policies. The average error of the deterministic policy is simply the difference between the integral of the measured velocity and the integral of the broadcasted velocity of the deterministic policy; similarly for the randomized policy average error.

Observe that in determining the optimal probability p we used the uncertainty cost, whereas in the evaluation we are using the error. The reason is that the error is a better metric of the imprecision, however, the error is unknown to the server in real-time. But we can compute the error given the vehicles sequence generated retroactively using the sensor (loop detector) records.

We compare the policies by considering their communication cost for the same error. More precisely, for each value of the error we subtract the communication cost of the randomized policy from the deterministic one (see Figure 6). If the difference of the total communication cost between the two policies is greater than zero, then the randomization policy is more efficient than the deterministic policy, because it means that the randomized policy achieves the same accuracy as the deterministic policy, but with less communication. However, the simulation cannot get as a parameter an error value (to produce a value of the communication cost). The parameter of the simulation that induces the error is the transmission threshold T. Thus we first obtain the error (Figure 4) and the communication cost (Figure 5) as a function of T, and then we compute the difference in communication cost between the two policies, for each given error (see Figure 6). Furthermore, observe that the results of the comparison may be different for various values of the server delay. Thus we perform the comparison for 5 different values of the server delay, 1 minute, 2 minutes, …, 5 minutes.

4.4 Simulation Result

Figure 3 shows the measured velocity (i.e. speed) as a function of time for T=1.0, τ=180, U=0.05. This graph indicates that the measured speed v_m is changing every minute. Then, the two velocities v_{br}, v_{bd} representing the randomized and deterministic broadcasted velocities, are also shown as a function of time. This graph shows that the updates of v_{br} occur at different times from the updates of v_{bd}. The reason is that in the randomized policy updates are delayed by the randomization function. The broadcasted velocities are also different for the two policies.

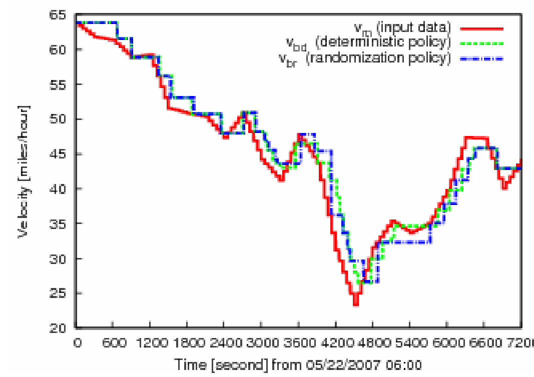

Figure. 3 Speed Transition by Simulation **(T=1.0, τ=180, U=0.05)**

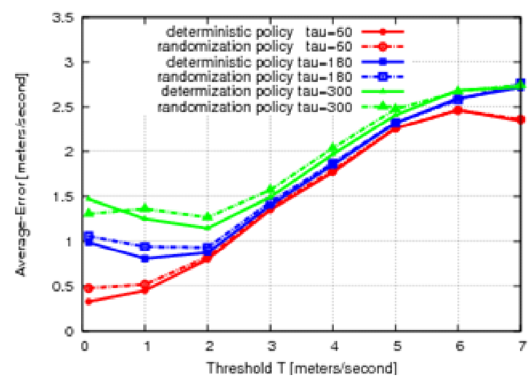

Figure 4. Threshold T and Average Error **(U=0.05)**

Next, we show the relationships among the threshold T, the server delay τ, and the error, during 2-hour execution of the simulation. These relationships are shown in Figure 4 for each one of the policies. In order to better demonstrate the trends, the curves are shown only for τ = 60, 180, and 300 where U=0.05.

According to Figure 4, the error for a given delay τ varies with the threshold T. In general, the error, expectedly, increases with the threshold T. However, there are exceptions. For example, for τ = 300, the minimum error is obtained when the T=2, and further decrease of the threshold increases the error. In general, the error also increases with the server delay τ.

The relationships among the threshold T, the server delay τ, and the total communication cost is shown in Figure 5. Expectedly, for each policy and for each server delay, the total communication cost decreases as the transmission threshold T increases. The reason for this, clearly, is that the smaller the threshold, the more frequently it is reached. The figure demonstrates another expected result, namely that the communication cost of the randomized policy is lower than that of the deterministic policy for the same threshold *T* and server delay τ, especially when *T*<3. Additionally, when the server delay τ is increases, the communication cost also tends to increase, especially when *T* is small. The reason is that as the server delay increases, the broadcasted velocity is less accurate, i.e. more frequently different than the measured velocity.

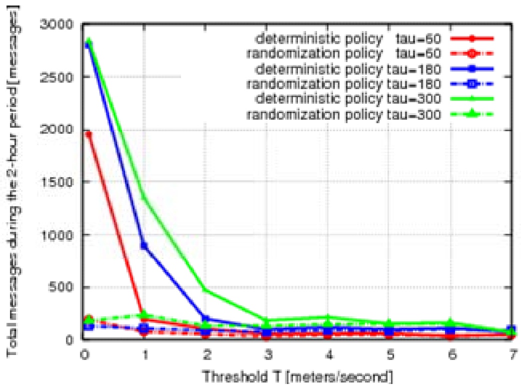

Figure 5. Threshold *T* and Total Messages During the 2-hour Period (U=0.05)

Finally, in Figure 6 we show the difference in communication cost between the randomized policy and the deterministic one, for a given average error (regardless *T*). More specifically, in Figure 6(a) the y-axis is the difference in communication cost (deterministic - randomized), and thus a positive value indicates superiority of the randomized policy, and vice versa, a negative value indicates superiority of the deterministic policy. Each line indicates the communication cost difference for a different server delay.

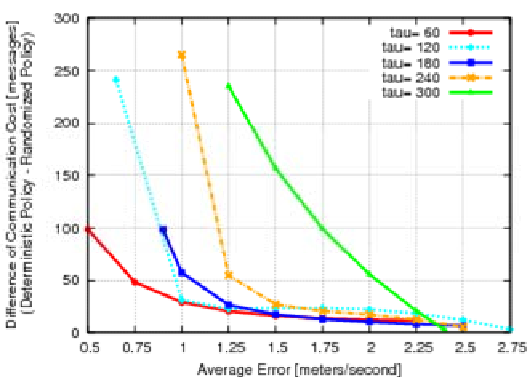

Figure 6(a): **Difference in Communication Cost between Deterministic Policy and Randomized Policy as a Function of the Average Error.**

Observe that each line has its own average-error range. The reason is that for larger server delays a lower average error is not achievable. For example, for τ=300 the minimum error is achieved when T=3.0, and lowering T further does not decrease the error (see also Figure 4).

According to Figure 6(a), each line is greater than zero. This means that in order to achieve the same level of accuracy, the deterministic policy needs more transmissions than the than the randomized policy. Observe that this advantage of the randomization policy is independent of the threshold *T*. Without this observation, one may be tempted to suggest that the effect of the randomized policy can be achieved in the deterministic policy simply by increasing the threshold *T*. However, Fig. 6 (a) refutes this suggestion by indicating that the advantage of the randomized policy is for a given error, not for a given threshold.

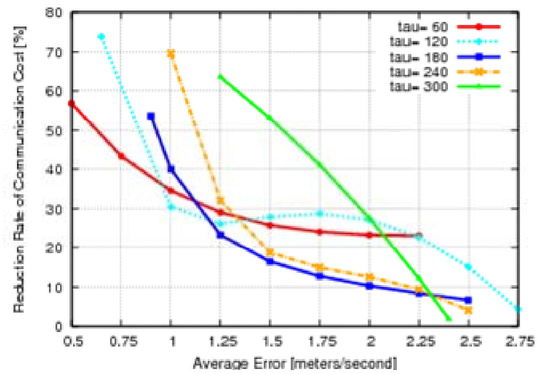

Figure 6(b): **Communication Cost Savings Afforded by the Randomized Policy as a % of the total cost.**

Finally, how significant is this advantage of the randomized policy? In other words, what percentage of the communication cost does the difference in Figure 6(a) represent? In Figure 6(b) we present the percentage:

(DeterministicMessageCost – RandomizedMessageCost)*100/ DeterministicMessageCost

as a function of the Average Error. The figure indicates the communication cost savings of the randomized policy is indeed significant, representing up to 72% of the communication cost.

5. Conclusion and Future Work

In this paper we addressed the problem of reducing the communication cost in traffic information sharing systems. Such a system is based on Floating Car Data, in which vehicles measure the average velocity on road segments, and then transmit this information to a server. In turn, the server broadcasts this information throughout the road network, to be used by vehicles in computing optimal travel routes.

We pointed out that the traditional deterministic policy incurs redundant transmissions if the server delay (i.e. the time it takes from the transmission of a velocity update by a vehicle to the server, until this update is propagated to the vehicles in the road network) is significant. We introduced the randomized policy to address this problem. In this policy each vehicle transmits its speed to the server only if the result of the toss of a coin with probability *p* is "true".

Then we addressed the problem of determining the optimal probability p. In order to do so we introduced the Information Cost Model to quantify the trade-off between the communication cost and the accuracy of the traffic information sharing system. Then we evaluated our proposed policy by comparing it with the conventional deterministic policy. The evaluation and comparison was carried out by using real traffic data collected in one of the Chicago highways. The evaluation shows that randomized policy is superior to the deterministic one, with an advantage to up to 72% reduction in communication cost (for the same accuracy).

Furthermore, it showed that the effect of the randomized policy cannot be achieved by simply increasing the threshold of the deterministic policy.

Future work includes extensions in several directions. First, we suspect that the randomized policy can be further optimized, for example by varying other parameters such as U, the ratio between the uncertainty cost and the communication cost. Second, there are other variations of the randomized policy which are worth examining. For example, instead of deciding in a probabilistic fashion whether or not a vehicle should transmit the velocity, it possible to pick the transmission threshold probabilistically. Third, it will be interesting to adapt the randomized policy to a P2P architecture such as the one proposed in [5]. Fourth, it will be interesting to compare the proposed randomized policy to a variant in which the server does not revise the broadcasted speed based on every update, but broadcasts the average of speeds from multiple vehicles. We have modified randomized policy runs as is, except that the server divides the time into periods during which it averages the received updates. However, the presented policy has a better performance than this variant, because the former matches the mathematical model. So the challenge is to develop the model for the averaging randomized policy.

6. Acknowledgments

We thank IGERT members, especially Prof. Prasad Sistla for their constructive comments and discussions. This research was supported in part by the National Science Foundation under Grants DGE-0549489, OII-0611017, 0513736, 0326284. The data is courtesy of the Illinois Department of Transportation.

7. References

[1] Turksma, S.: The various uses of floating car data, Road Transport Information and Control, 2000. Tenth Intl. Conference on (Conf. Publ. No. 472), April 2000 pp.51-55

[2] http://www.premium-club.jp

[3] http://www.vmz-online.de/vmz/

[4] Kerner, B.S., et al: Traffic state detection with floating car data in road networks, Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE, Sept. 2005, pp.44- 49

[5] S. Goel, et al: Grassroots: A Scalable and Robust Information Architecture. Technical Report DCS-TR-523, Department of Computer Science, Rutgers University, June 2003. http://paul.rutgers.edu/~gsamir

[6] T. Shinkawa, et al: A Technique for Information Sharing using Inter-Vehicle Communication with Message Ferrying , Proceedings of the 2006 International Workshop on Future Mobile and Ubiquitous Information Technologies, (FMUIT'06), pp.221-225, (May 2006)

[7] Masutani, et, al: Traffic Prediction using Pheromone Model, 12th World Congress on ITS, 2005,

[8] Adachi, et al: Compression Method for Probe Data, Inter-Vehicle P2P Communication Experiment System, ITS World Congress, Nagoya, 2004

[9] R. Horiguchi, et. al: Effective Probe Data Transmittal with Detection of Congestion Pattern, Proc. of 11th World Congress on Intelligent Transport Systems, Nagoya, Oct., 2004

[10] A Civilis, C. S. Jensen, Efficient tracking of moving objects with precision guarantees, Proc. MobiQuitous 2004

[11] Ouri Wolfson, et al: Updating and Querying Databases that Track Mobile Units, Distributed and Parallel Databases, v.7 n.3, p.257-387, July 1999

[12] B.D. Greenshields, "A Study of Traffic Capacity," Highway Research Board Proc., vol. 14, pp. 448-477 1935

[13] L. Henry, Traffic Flow Theory - A State-of-the-Art Report: Revised Monograph on Traffic Flow Theory, Turner Fairbank Highway Research Center (TFHRC), 2002, http://www.tfhrc.gov/its/tft/tft.htm

[14] B. Xu, and O. Wolfson, "Opportunistic Resource Exchange in Inter-vehicle Ad-hoc Networks", Proc. of 2004 IEEE Int'l Conf. on Mobile Data Management (MDM 2004), pp.4–12, Jan. 2004.

[15] GCM Travel http://www.gcmtravel.com/

Appendix A. Definition of Terms (Table 1)

The notations and definitions used in this paper are as follows.

v_m: Measured velocity of a road segment; it is measured when a vehicle goes through the end of a road segment.

v_b: Broadcasted velocity from the server to the vehicles.

T: Threshold of velocity, used in transmission rule to reduce the number of transmissions.

p: The probability used by the randomized policy in combination with the transmission rule.

Δ: In the deterministic policy, the number of time units between two consecutive times the threshold T is exceeded. In the randomized policy, it is the length of the time period from the time a vehicle succeeds in transmitting v_m to the server, until the threshold T is exceeded the next time (see Figure 1.).

τ: The number of time units from the update of the server until the vehicles receive the updated velocity. In the simulation, it is the time-length of a server cycle.

K: The number of vehicles running through the end of a road segment during the server delay τ.

U: The cost of a unit of uncertainty per time unit, in messages.

V: The free flow speed, (i.e. uncongested speed) on a road segment.

α: The expected number of time-units from the instance when the measured speed first exceeds the threshold T until a vehicle first transmits the new speed to the server. It depends on the probability p used by the randomized policy.

s: The number of time-unit between two consecutive vehicles.

r: Traffic flow: The number of vehicles per lane per second passing on the road. It is the inverse of s.

d: Jam density: The number of vehicles per unit-length when the traffic is so heavy that it is at a complete standstill.

u: The *uncertainty of* v_b. This value is broadcast by the server in the randomized policy, and it may be the velocity threshold T, or the free flow speed.