StreetSmart Traffic: Discovering and Disseminating Automobile Congestion Using VANET's

Sandor Dornbush and Anupam Joshi
Dept. of Computer Science and Electrical Engineering
University of Maryland Baltimore County
Baltimore, Maryland 21250
Email: {sandor1 joshi}@cs.umbc.edu

Abstract— **Automobile traffic is a major problem in developed societies. We collectively waste huge amounts of time and resources traveling through traffic congestion. Drivers choose the route that they believe will be the fastest; however traffic congestion can significantly change the duration of a trip. Significant savings of fuel and time could be achieved if traffic congestion patterns could be effectively discovered and disseminated to drivers. We propose a system that uses a standard GPS driving aid, augmented with peer-to-peer wireless communication. The prosed system uses a combination of clustering and epidemic communication to find and disseminate dynamic traffic patterns.**

Fig. 1. System Overview

I. INTRODUCTION

Automobile traffic is a major problem in modern societies. Throughout the world millions of hours and gallons of fuel are wasted everyday by vehicles stuck in traffic. The Texas Traffic Institute estimates that traffic congestion costs the US $68 billion dollars a year [24]. This paper details an investigation of automobile traffic monitoring using mobile peer-to-peer networks. Several companies provide live traffic information [36][28][20], and give drivers suggested routes over congested roads. The main deficiency of current systems is the lack of accurate source data.

StreetSmart's main contribution is a new method for collecting accurate real time congestion information for which there is a great need. There have been several centralized approaches to collect live traffic information. However it is cost prohibitive to implement these systems on all roads.

The goal of the project is to create a system that could be implemented with technology generally available in cars today. Specifically the system would use the Global Positioning System (GPS) to determine current location and speed, and a wireless networking communication medium such as 801.11 A, B or G, or 802.15.

As vehicles travel through congested roads the traffic device tracks the speed of the vehicle. From the traffic that each vehicle experiences it builds a local traffic map. As vehicles come close to each other they exchange their speed maps with each other. Through these interactions each peer in the system is able to build a map of expected speed on every road, even those they have not visited.

In a VANET the nodes are highly mobile, and many vehicles are not participating the network. As a result the nodes in the network are often disconnected, and any network that
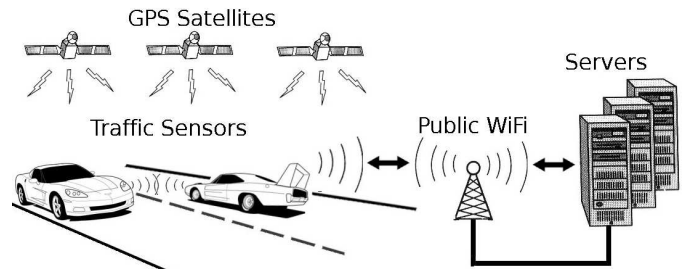
does form will be at best partially connected. We developed distributed clustering algorithms that will function in that environment using an epidemic diffusion model.

We studied our system in a simulation where cars drive random paths through a Manhattan grid and on linux laptops on Baltimore area roads.

II. RELATED WORK

StreetSmart builds upon the work of many other projects. This section explores relevant work. In each subsection we show the current state of the art and how we use and contribute to that.

A. VANETs

Many major car manufactures and leading research institutions are investigating creating Vehicle Ad-Hoc Networks(VANETs) [15][2]. Much of the work in the VANET community focuses on simulation [12][17], multi-hop routing [37][14][10] and entertainment systems such as multiplayer games [13] or streaming music [16]. The CarTel project[9] has explored stealing connections from open WiFi stations in Boston and Seattle.

A couple of researchers have studied the problem of using VANETs to discover and disseminate congestion information [18][30]. Ways of organizing such networks were discussed in [11]. Marca et al[25] discussed ways of extending an existing centralized GPS traffic monitor to use VANETs. TrafficView [27] addresses the issue of estimating road congestion using a network of vehicle based GPS systems. In essence networking car based GPS devices creates a sensor network which measures the congestion of the roads. However this

sensor network has a few key distinctions from typical sensor networks. These are the key differences:

1) **Power consumption is not an issue.** The power used by the computer systems is insignificant to the vehicle.
2) **Networks are very dynamic.** Vehicle motion can lead to a fast changing sparsely connected network.
3) **Node size and cost are not severely limited.** Vehicle born GPS systems are larger and have more resources than typical sensor network nodes.

Previous VANET congestion systems had limited scope. The TrafficView [27] project focused the congestion of the road directly ahead. The TrafficView project was able to demonstrate that it is possible to monitor vehicle congestion using a real VANET. The idea was extended to both sides of the road by SOTIS [38]. We are the first to address the problem of discovering traffic on a road network.

B. Clustering

Clustering is an important data mining tool. Clustering is the process of combining data points that are similar to each other by some measure. An optimal cluster assignment will provide the maximum compression with minimum distortion. StreetSmart uses clustering as a data aggregation technique to combine related recordings of unusual speed.

Several algorithms have been presented to find clusters in a distributed network such as what we are proposing. Several papers present ways to perform a bit of data analysis at each node then send the results of that analysis to a central node [21] [23][7]. We use a similar technique, however each node tries to do the analysis of the collected statistics, instead of just the central node.

There are several techniques for calculating clusters in a P2P fashion. However all of them have relied on a fully connected network[3][39] which is not possible with VANETs. A group of algorithms such as [32] or [34] calculate distributed clustering using centralized ensembles.

C. Epidemic Communication

This paper deploys a form of communication that is alternately referred to as rumor [31], gossip [22] or epidemic [5] based. This class of communication protocols is based on the practice of saving messages from previous nodes and relaying those messages to other nodes. While these communication protocols do not provide any guarantee on the timeliness of the data, they are often the best possible protocols in disconnected settings. Performing data mining using gossip protocols was used in the Astrolabe system [6].

III. SYSTEM DESIGN

StreetSmart is designed to address the needs of drivers. This system does not exchange information on every section of road instead it only exchanges summary information on areas of unexpected traffic. Unlike many similar works we do not expect that all drivers adopt StreetSmart Traffic, it is designed to perform well with only a small fraction of total drivers participating in the network. This is a marked

difference between most other proposed VANETs. Due to the rapid movement found int VANETs we cannot rely on a fully connected network. We present a distributed clustering algorithm that does not require constant connectivity.

The motivation of the algorithm is for each node to keep a compressed version of every other nodes traffic information. This information is expressed as summary statistics about clusters. Nodes exchange summary statistics using epidemic communication. Every participating node calculate higher level clusters from the summary statics gathered locally and from the network.

A. Need To Say

When designing the communications we followed a "Need to Say" principle. The "Need to Say" principle seeks to reduce communication load by only transmitting useful information. Other researchers in the VANET field have focused on data fusion as a way to reduce communication load. We have not ignored that tool, but rather augmented it. A similar approach has been used for perform association rule mining (ARM) [33].

The first application of this principle to this project is observing that one need not communicate any information if the vehicle is traveling at, or above the posted speed limit. This is the motivation for representing the traffic as a collection of clusters of slow traffic. It is assumed that the number of clusters of congestion is significantly smaller than the total number of road segments. This simple insight compresses the set of all road segments down to just the deviations from the speed limit.

It is possible to extend the idea of expected speed beyond the posted speed limit. Traffic congestion has very predictable trends which can be exploited. For example, it is common to have significant congestion on commuter routes at peak travel times. If this information is available to all of the nodes of the network then each node only needs to communicate when the recorded speed is outside the variance of expected speed. The JamBayes[19] project explored predicting traffic patterns using bayes networks. This work focused on the unexpectedness of traffic to drivers, however the same techinque could be used to reduce the communication on a sensor network.

B. Data Representation

The naive choice of data representation for geostatics data is latitude and longitude coordinates. Our early experiments used this data model. The distance can be calculated using the L2Norm of the latitude and longitude. However that model presented significant problems. If one simply uses latitude and longitude coordinates it is possible that clusters of data appear to be inside city blocks or on the wrong side of the road. To overcome this deficiency congestion data is represented by a road identifier and an offset on that road, similar to a mile marker. A digital map such as those provided by NAVTEQ[28] or TeleAtlas[35] can be used to find a road id and offset from a latitude and longitude location. The similarity of two data points can be expressed
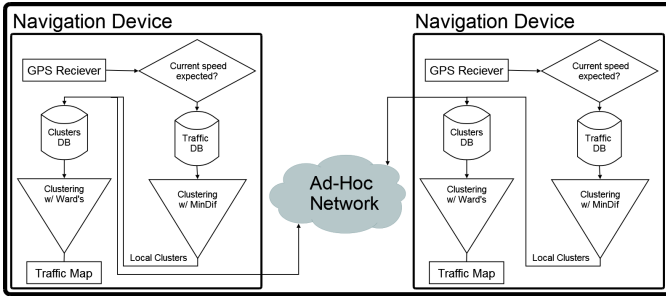
Fig. 2.   Hierarchical Agglomerative Clustering Overview

by this function:
DIFFERENCE(d1,d2) =

$$\begin{cases} \text{L2NORM}((d1.dist, d1.speed), \\ \qquad (d2.dist, d2.speed)) & d1.roadId = d2.roadId \\ 0 & Otherwise \end{cases}$$

C. Traffic Monitoring

After sampling the data, nodes exchanging messages then each node merges their snap shot of other nodes centroid data with its local centroids. We experimented with two different methods of calculating the MetaClusters.

The first method of calculating MetaClusters we experimented with was an adaptation of K-Means to a peer-to-peer environment. Each cluster is built as a weighted combination of the source data. This performed well, however we had difficulty specifying the number of clusters apriori.

To avoid some of the difficulties of K-Means we created a distributed hierarchical agglomerative clustering system that does not need to know the number of clusters a-priori and can use the custom similarity measure defined in III-B. To adapt agglomerative clustering to a distributed setting we used a combination of two proximity measures. When clustering local traffic information we used the MIN proximity measure. This measure defines the proximity of two clusters and the proximity of their two closest elements. With just the summary statistics we used Ward's method of proximity to measure proximity of distributed clusters. Under Ward's method the proximity of two clusters is defined to be the proximity of the centroids of the two clusters.

D. Message Size

The size of the message can be constrained to a fixed constant and still monitor the road conditions. If every a node tried to exchange every cluster in the system message size would be bound by $O(nk)$ where n is the number of nodes and k is the number of clusters. This linear growth in the number of nodes is not desirable. We used a heuristic to fix the size of the messages to a constant c. Each node chooses to send only the nodes with the highest heuristic value that can fit within a fixed message size. We used the heuristic: $h(centroid) = \sum_j memberShipCount_j age_j$. This will send the most significant and the newest centroids.

Each cluster is quite small, just 48 bytes. Assuming that the expected number of traffic clusters any one car encounters is 10. Then the expected size of one local traffic map is only 480 bytes. If there are 1000 participating nodes all of the local traffic maps would have a combined size of 469 KB. In this simple example the size of the largest messages is 469 KB. These messages are not tiny, however they are very reasonable size for transmission over 802.11 networks. We have shown that typically much fewer than 250 clusters are needed for good results.

IV. SIMULATION

StreetSmart Traffic was evaluated in simulation. Our simulation provides a sufficient level of detail to experiment with our application. Many of the assumptions, such as a random way point model, are shared by similar work. It would be desirable to use a more sophisticated simulation, however no one suitable simulation was found. Different simulators model different aspects of automobile traffic and communication. Some simulators such as STRAW[12] have combined traffic simulation and MANET simulation to create a VANET simulator. However even this does not model traffic congestion. At a future point it would desirable to combine a system like STRAW with a congestion simulator such as DYNAMIT[1].

The StreetSmart simulator attempts to model basic digital broadcasting. The broadcast distance and allowable message size are designed to approximate the performance of 802.11 G. For most simulations the broadcast distance is set to 500 meters. [29] The simulation is a Manhattan grid of highways. The vehicles travel along paths generated using random way point mobility. [4].

V. SIMULATION RESULTS

We defined the accuracy of the StreetSmart Traffic system as the difference at every point along every road between the actual speed of vehicles on that road and the estimated speed of vehicles on that road. At uniform intervals the speed is sampled from the true speed map and the estimated speed map. The difference between those two values is used to calculate the mean squared error across all road segments. The results are averaged over ten or more simulations.

The figures 3(a), 3(b), 3(c) and 3(d) illustrate the various traffic maps from a simulation of 500 cars in an 80km x 80km manhattan grid. The first figure 3(a) shows the actual speed on all roads, as imposed by the simulation. The server that connects to all of the base stations calculated the speed map shown in 3(b). Figure 3(c) and 3(d) show the smallest error and largest error, respectively, of any vehicle in the simulation.

The number of connections between nodes has a strong effect on the accuracy of the StreetSmart system. Several experiments were performed to test this relationship. The first set of these experiments were conducted with different numbers of participating vehicles from 50 to 500, the results are presented in figure 4(a). As more vehicles participate in the
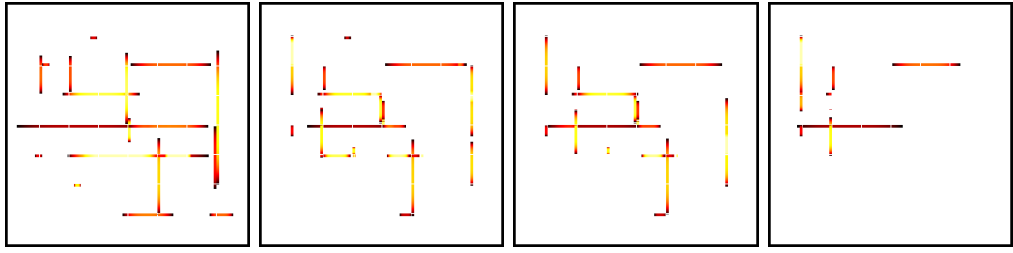
(a) The speed imposed on the vehicles by the simula- tion.

(b) The speed estimated by the base station.

(c) The best speed estimated by any vehicle.

(d) The worst speed esti- mated by any vehicle.

Fig. 3.   Various results and visualizations of a typical simulation of 500 vehicles.
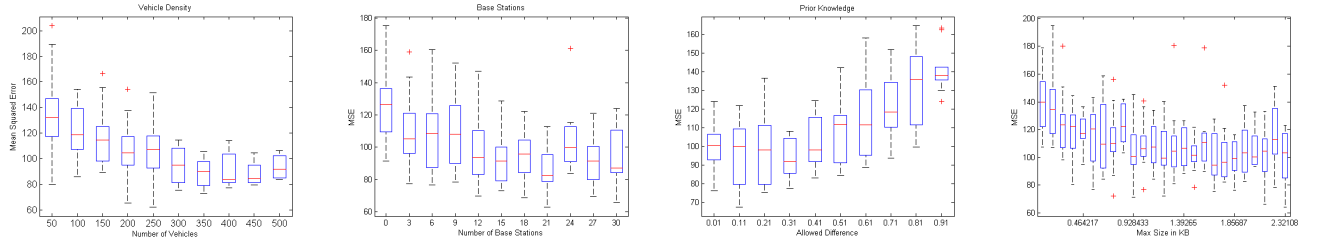
(a) Accuracy as a function of the number of vehicles

(b) The affect on the MSE of the number of wired sinks

(c) The affect on the MSE of the use of known priors.

(d) The affect on the system accuracy of the maximum number of clusters allowed in each message.

Fig. 4.   Various results and visualizations of a typical simulation of 500 vehicles.

network the error at each node decreases. Experiments with even fewer vehicles showed poor performance.

Wireless base stations connected to a wired network can significantly increase the performance of the overall network. These experiments show in figure 4(b) that the reliable fast nature of wired network improve the performance of the over- all system. Linear regression showed that each base stations has a more powerful effect on system accuracy than each vehicle. Linear regression found that each vehicle has a - 0.1178 influence on the mean squared error, however each base station has a larger influence of -0.9408. In these simulations we used a relatively small number of open base stations, 0- 30, typical large US cities have thousands of open 802.11 base stations.

We performed a series of experiments to see the effect prior knowledge has on the accuracy of the systems. In these experiments traffic was only recorded if the speed was different than some percent less than the expected speed the results are presented in figure 4(c). These experiments show that the accuracy decreases if too much of the speed deviations are ignored. It is safe to ignore low intensity traffic and still accurately identify the worst areas of traffic.

To constrain the size of messages, only the newest and most significant clusters are exchanged. We experimented with several different maximum message sizes, the results of these experiments are show in figure 4(d). To our surprise we were able to achieve high levels of accuracy while only communicating a small number of clusters. It appears that our simple heuristic does a good job a choosing the most relevant clusters to exchange.

We performed a number of experiments exploring the effect of having more traffic congestion. In those experiments we found little relationship between the amount of congestion and the accuracy of the system. In our view the mobility of the nodes has little effect on the system, most of the data mobility comes from the wired network.

VI. Experiments

We have developed a prototype system using Linux based laptops. Each laptop runs the RoadMap[26] system. RoadMap is an open source portable navigation system. We used maps built from TigerLine[8] data. Each device has an 802.11 G connection with has an external unidirectional antenna mounted on the vehicle roof. Each vehicle used a static IP and broadcast UDP packets.

In our experiments we were able to create a small network of 2 laptops on Baltimore area highways. In these experiments we were able to communicate traffic disruptions between vehicles. We are currently resolving problems we found with the RoadMap system.

VII. Conclusion and Future Work

The StreetSmart Traffic system has been shown in simu- lation to effectively find and disseminate automobile traffic congestion using ad hoc wireless networking. We built proto- types of the system on commodity hardware. We introduced a

technique to find clusters in a disconnected peer-to-peer network. Our system performed well in challenging environments with non-stationary data and a sparsely connected dynamic network, converging to a good answer in the constrained setting of an automobile network.

This system has been designed to provide high value information to drivers in a timely fashion. While the promise of networking together vehicles to broadcast music or play games is intriguing, it has limited value to drivers. Instead of making the time spent in cars more pleasurable, this system will allow people to spend as little time as possible in cars. If people can avoid traffic they can arrive at their destination quicker and then use traditional networks to enjoy music or video games.

Recent advances in wireless networks such as WiMax or EV-DO raise the possibility of having cheap ubiquitous connections in most locations where traffic is likely. In view of that we believe the best method for finding the global state of a road network will rely primarily on centralized communication. In view of that we believe the most important questions involve optimizing the available bandwidth and protecting users privacy.

REFERENCES

[1] http://mit.edu/its/dynamit.html.

[2] Calling all cars. *Daimler Chrysler Hightech Report*, 2001.

[3] S. Bandyopadhyay, C. Gianella, U. Maulik, H. Kargupta, K. Liu, and S. Datta. Clustering distributed data streams in peer-to-peer environments. *Information Science Journal*, 2004. In Press.

[4] Christian Bettstetter and Christian Wagner. The spatial node distribution of the random waypoint mobility model. In *Mobile Ad-Hoc Netzwerke, 1. deutscher Workshop Uber Mobile Ad-Hoc Netzwerke WMAN 2002*, pages 41–58. GI, 2002.

[5] K. P. Birman. The surprising power of epidemic communication. *LECTURE NOTES IN COMPUTER SCIENCE*, pages 97–102, 2003.

[6] Ken Birman, Robbert van Renesse, and Werner Vogels. Navigating in the storm: Using astrolabe for distributed self-configuration, monitoring and adaptation. 5th Annual International Active Middleware Workshop (AMS 2003), June 2003.

[7] C. Boulis and M. Ostendorf. Combining multiple clustering systems. In *8th European conference on Principles and Practice of Knowledge Discovery in Databases(PKDD), LNAI 3202*, pages 63–74, 2004.

[8] U.S. Census Bureau. Topologically integrated geographic encoding and referencing system, 2006.

[9] Vladimir Bychkovsky, Bret Hull, Allen K. Miu, Hari Balakrishnan, and Samuel Madden. A Measurement Study of Vehicular Internet Access Using In Situ Wi-Fi Networks. In *12th ACM MOBICOM Conf.*, Los Angeles, CA, September 2006.

[10] Z. Chen, H. Kung, and D. Vlah. Ad hoc relay wireless networks over moving vehicles on highways. MobiHoc, 2001.

[11] I. Chisalita and N. Shahmehri. A peer-to-peer approach to vehicular communication for the support of traffic safety applications. pages 336–341. 5th IEEE International Conference on Intelligent Transportation Systems, Singapore, Sept 2002.

[12] David Choffnes and Fabin E. Bustamante. Straw - an integrated mobility and traffic model for vanets. Number 10. International Command and Control Research and Technology Symposium (CCRTS), June 2005.

[13] M. Delio. Wireless caravan: Geeks on parade. *Wired Magazine*, 2003.

[14] M. Dikaiakos, S. Iqbal, T. Nadeem, and L. Iftode. Vitp: An information transfer protocol for vehicular computing. 2nd ACM International Workshop on Vehicular Ad-Hoc Networks (VANET), September 2005.

[15] Holger Fler, Marc Torrent-Moreno, Hannes Hartenstein, Matthias Transier, Roland Krger, and Wolfgang Effelsberg. Studying vehicle movements on highways and their impact on ad-hoc connectivity. *ACM SIGMOBILE Mobile Computing and Communications Review (MC2R)*, 2006.

[16] J. Garretson, W. Hess, J. Kanarek, M. Pignol, and M. Shai, 2005. http://roadcasting.org.

[17] S. Ghandeharizadeh and B. Krishnamachari. C2p2: A peer-to-peer network for on-demand automobile information services. First International Workshop on Grid and Peer-to-Peer Computing Impacts on Large Scale Heterogeneous Distributed Database Systems (GLOBE'04), August 2004.

[18] Mark Hallenbeck. Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation.

[19] E. Horvitz, J. Apacible, R. Sarin, and L. Liao. Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service. *Twenty-First Conference on Uncertainty in Artificial Intelligence*, July 2005.

[20] Inrix, 2006.

[21] E. Januzaj, H.-P. Kriegel, and M. Pfeifle. Dbdc: Density based distributed clustering. pages 88–105. EDBT in Lecture Notes on Computer Science, 2004.

[22] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. Number 44. Annual IEEE Symposium on Foundations of Computer Science, 2003.

[23] A. Lazarevic, D. Pokrajac, and Z. Obradovic. Distributed clustering and local regression for knowledge discovery in multiple spatial databases. Number 8, pages 88–105. European Symposium on Artificial Neural Networks, 2000.

[24] Tim Lomax and David Schrank. Urban mobility study. Technical report, Texas Transportation Institute, 2005.

[25] James E. Marca, Craig R. Rindt, and Michael G. McNally. Towards distributed data collection and peer-to-peer data sharing. Technical Report UCI-ITS-AS-WP-02-4, August 2002.

[26] Pascal Martin, Latchesar Ionkov, Stephen Woodbridge, and Ehud Shabtai. A car navigation system for linux and unix (and pocketpc too), 2006.

[27] T. Nadeem, S. Dashtinezhad, C. Liao, and L. Iftode. Trafficview: Traffic data dissemination using car-to-car communication. *ACM Sigmobile Mobile Computing and Communications Review, Special Issue on Mobile Data Management*, 8(3):6–19, July 2004.

[28] NAVTEQ. www.navteq.com, 2006.

[29] Ramano P. The range vs. rate dilemma of wlans. *Wireless Net DesignLine*, January 2004.

[30] Filip Perich, Anupam Joshi, Tim Finin, and Yelena Yesha. On Data Management in Pervasive Computing Environments. *IEEE Transactions on Knowledge and Data Engineering*, May 2004.

[31] P. Reiher, G. Popek, M. Gunter, and J. Salomone. Peer-to-peer reconciliation based replication for mobile computers. European Conference on Object Oriented Programming, Second Workshop on Mobility and Replication, June 1996.

[32] N. F. Samatova, G. Ostrouchov, A. Geist, and A. Melechko. RACHET: An Efficient Cover-Based Merging of Clustering Hierarchies from Distributed Datasets. *Distributed and Parallel Databases*, 11(2):157–180, 2002.

[33] A. Schuster and R. Wolff. Communication-Efficient Distributed Mining of Association Rules. *Data Mining and Knowledge Discovery*, 8(2), March 2004.

[34] A. Strehl and J. Gosh. Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3:583–617, 2002.

[35] TeleAtlas, 2006.

[36] Traffic.com, 2006.

[37] Jrg Widmer, Martin Mauve, Hannes Hartenstein, and Holger Fler. *The Handbook of Ad Hoc Wireless Networks*, chapter Position-Based Routing in Ad-Hoc Wireless Networks. CRC Press, 2002.

[38] Lars Wischhof, Andr Ebner, and Hermann Rohling. Self-organizing traffic information system based on car-to-car communication: Prototype implementation. International Workshop on Intelligent Transportation (WIT), MARCH 2004.

[39] R. Wolff, K. Bhaduri, and H. Kargupta. Local l2 thresholding based data mining in peer-to-peer systems. Technical Report TR-CS-05-11, UMBC, 2005.