PRIVACY WITH DSRC VEHICLE SAFETY BROADCASTS

An Undergraduate Research Scholars Thesis

by

MASON RUMULY

Submitted to the Undergraduate Research Scholars program at
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by Research Advisor: Dr. Srinivas Shakkottai

May 2017

Major: Electrical Engineering

TABLE OF CONTENTS

Page

ABSTRACT

Privacy with DSRC Vehicle Safety Broadcasts

Mason Rumuly
Department of Electrical and Computer Engineering
Texas A&M University

Research Advisor: Dr. Srinivas Shakkottai
Department of Electrical and Computer Engineering
Texas A&M University

Dedicated Short Range Communications (DSRC) is an emerging technology application allowing vehicles to communicate with each other for safety improvements. The current protocol calls for open broadcast of Basic Safety Message (BSM) reports every 100ms; these contain the time, location, speed, acceleration in three dimensions, and other attributes of the car at that point in time along with a broadcast identification (ID) randomized at random intervals. This protocol introduces privacy vulnerabilities, as anyone with a compatible device can track a vehicle's movement. The primary aim of this research is to expose the privacy shortfalls and ultimately determine how much information to omit in order to preserve users' privacy. In order to do this, I analyzed a data set of BSM messages generated from a two month deployment of the technology in an Ann Arbor, Michigan study by the US Department of Transportation (DOT). While my colleagues focused on connecting individual points, I focused on the specific question of reconnecting vehicle path segments across significant broadcast gaps and IDs changes using driver behavior. I simulated the problem by choosing an intersection in Ann Arbor, taking multiple paths through it, splitting the paths across it, and omitting any time data as well as any BSMs within 50m of the intersection. I wrote a preliminary algorithm to create statistical profiles

of the paths, from which it then attempted to match the ‘before’ and ‘after’ paths. This basic algorithm is able to match the paths with better accuracy than random guessing, but becomes less potent the more paths are introduced. As this crude method is able to stitch paths together on a harder problem than is likely in practical situations, it is likely more sophisticated methods would achieve even higher accuracy, implying that privacy is not protected sufficiently by changing ID or neglecting to broadcast during periods, as future work shall seek to determine.

ACKNOWLEDGEMENTS

Thanks to Dr. A.L. Narasimha Reddy, Dr. Alireza Talebpour, Bainan Xia, Archit Sharma, Mohammadreza Khajeh Hossein, and Vijay Subramanian for their advice and efforts on facets of this problem.

NOMENCLATURE

BSM	Basic Safety Message
DOT	Department of Transportation
DSRC	Direct Short Range Communications
ID	Identification
V2C	Vehicle to Cloud
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
VCA	Vehicle Collision Avoidance

CHAPTER I

INTRODUCTION

In order to provide for the safety and convenience of an ever-growing number of drivers, many technologies and services are being developed which fall under the umbrella of intelligent transport systems. These represent “tremendous investment from government, academia, and industry” [1], and include DSRC broadcast in addition to more central schema such as Long Term Evolution (LTE) cellular. Each has a certain set of applications to which it is suited, based mostly on scalability and latency characteristics. Some proposed applications do not require small latency, such as “reliable integration of EVs [Electric Vehicles] with the smart grid” [2], while others require short latency, such as vehicle collision avoidance (VCA) safety applications. LTE is is dependent on multiple network hops and “prior experiments have demonstrated that end-to-end communication latency increases exponentially as the number of cars in a cell increases”[3]. As such, cellular networks are infeasible for large-scale VCA applications, and developers are turning to the decentralized DSRC network.

DSRC Overview

DSRC as currently implemented uses protocol IEEE 802.11p, which uses a broadcast structure in the 5.9 GHz band allowing every device in a range extending up to one kilometer to hear from every other directly. Because of this scheme, “Simulation results reveal the immunity of DSRC PHY [physical] layer to large delay spreads”[4]. This characteristic allows DSRC the scaling necessary to operate on crowded highways. It requires only for each car to have a transmitting/receiving device, and can function without supporting infrastructure. However, this requirement also allows malicious actors with access to a DSRC unit access to anything transmitted; as such, information released through this protocol must be tightly controlled to avoid compromising the privacy of drivers.

BSM Scheme

Work is being carried out to optimize the reliability of the safety applications, which focus on broadcasting information about the vehicle once every 100ms [5]. These Basic Safety Messages (BSM) include, among other things, “vehicle size, position, speed, heading, acceleration, [and] brake system status” in addition to an ID tag which is randomized at random intervals [6]. This data is intended to allow early warning services for drivers to prevent collisions, and may in the future be used for coordinating driverless cars as well.

Privacy Concerns

Because these BSM packets must be standardized and receivable by other automobiles regardless of make, model, or owner, one must assume that any malicious actor within range would know the contents of every BSM; encryption of the data or limiting recipients would defeat the purpose entirely. Vehicles using this scheme would be traceable along their entire journey. The current proposed method to baffle such attempts is to change broadcast IDs after random intervals, breaking up the paths into disjoint segments. However, it is not clear whether this method is sufficient to prevent the paths being reconnected algorithmically due to both physical realities of driving and other identifiers such as driver behavior. This project seeks to determine the effectiveness of this method as a privacy safeguard, as well as propose and test other possible safeguards.

CHAPTER II

METHODS

Below I describe the methods used to carry out this research.

Sample Data

In order to investigate the problem, I used a dataset of BSM logs collected from 3,000 vehicles over the course of two months during the Safety Pilot Model Deployment in Ann Arbor, Michigan [7]. Because this data was from a real deployment of DSRC technology, operations on it should produce practical results. The data and algorithms were hosted on the ADA Cluster at Texas A&M and processed and analyzed using programs written in C++ and Python. We can apply certain privacy safeguards to the data, such as changing ID, total removal of explicit ID, removal of messages, and introduction of noise to the data. By then attacking the modified data, we expect to find the relationship between safety applicability and privacy preservation, hopefully uncovering an optimal solution to increase safety significantly over degrading privacy.

Kinematics vs. Behaviors

Prior and concurrent work by my colleagues focused on the kinematic realities to join segments which begin and end near each other. Their current algorithm analyzes this case by choosing a set of nearby paths, normalizing time data to produce concurrency, and chopping the path into ten second segments, then attempting to reconnect these mini-segments by looking for the closest message within 200ms of the last point in the segment. My focus is on the case where broadcasts are omitted in a location, and attempting to rejoin the segments on either side of the dead-zone using driver behaviors evident in the path segments.

Chosen Metric

In order to test the above problem, I first hypothesized that drivers can be profiled based on behavior near the break, and so can be reconnected with a significantly large probability. In order to test this, a set of paths through a common location were chosen and then split at that location; in this case, intersections were chosen as the locations, and when the paths were split all BSM data within 50m of the intersection were excluded from the disjoint paths. I then created a program to profile each path segment statistically, detailing the mean and variance of each attribute in the BSM packets while discarding time data to focus on behavior, both over the entire segment and over the minute of transmitting closest to the dead-zone. These profiles were then used to create a simple matching algorithm.

Algorithm

My latest algorithm creates a matrix of match costs, then uses the Hungarian Algorithm to find the least-cost 1-to-1 matching. The costs are evaluated as follows:

Take two paths, one in-coming to the dead-zone and one outgoing, calling them A and B respectively. Let them have statistical attributes $\{A_0, A_1, \dots, A_{N-1}\}$ and $\{B_0, B_1, \dots, B_{N-1}\}$ respectively, as produced by the characterization program. The similarity s_n for an attribute is determined by equation (1).

$$s_n = \begin{cases} \left|\frac{A_n}{B_n}\right|, & 0 \leq \left|\frac{A_n}{B_n}\right| < 1 \\ \left|\frac{B_n}{A_n}\right|, & 0 \leq \left|\frac{B_n}{A_n}\right| \leq 1 \end{cases} \tag{1}$$

Next, define a series of weights for each attribute position, $\{w_0, w_1, \dots, w_{N-1}\}$. This will be used in accordance with how important each attribute is found to be; in practice at this time, each weight array used has one $w_i = 1$ while the other $w_j\ \forall\ j \neq i$ are set to 0. The total similarity S between the paths is then set per equation (2).

$$S = \sum_{n=0}^{N-1} s_n w_n \tag{2}$$

The entries to the cost matrix are then computed as $C = 1 - S$, and the cost matrix solved using the Hungarian Algorithm.

Once the algorithm is run, the program compares the results to ground truth and determines the accuracy of the matching. The entire evaluation program was run over three intersections with paths binned by month and time of day in order to eliminate effects of season and time of day, for a total of 36 problems over which any weighting could be averaged. On average, an intersection had 73 paths (146 split segments) available to evaluate.

CHAPTER III

RESULTS

The current results are as follow.

Kinematic

The simple kinematic algorithm achieves approximately 90.5% total path recovery for paths over five minutes long, achieving about 99.7% success on a per-split basis.

Behavioral

By limiting the number of paths per problem and taking the optimal matches, my program produced the relation between accuracy and path density shown in Figure 1. Note that the relation is better than random guessing would be expected to achieve, for example nearly 75% accuracy at two paths per problem instead of the expected 50%.

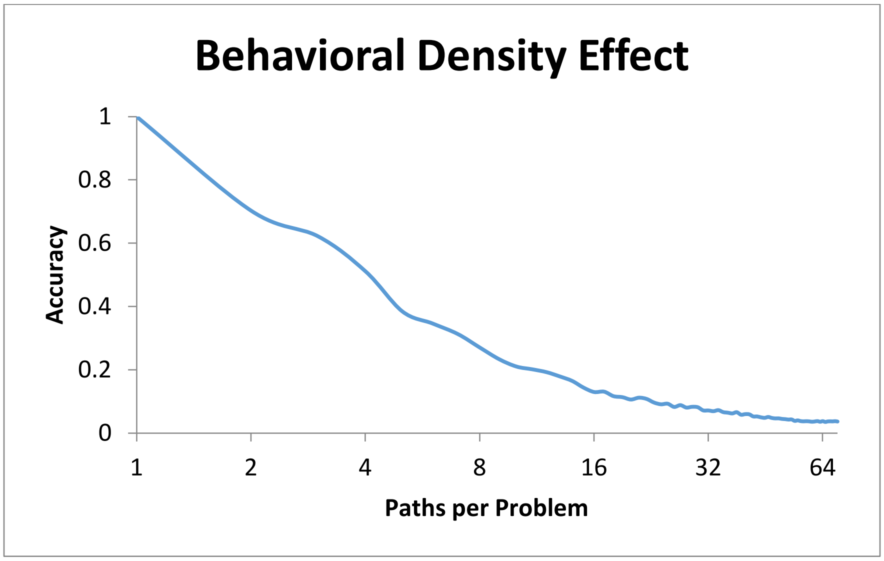

Figure 1: Accuracy of Behavioral Algorithm

CHAPTER IV

CONCLUSION

Current Indications

The results from the kinematic portion indicate that a vehicle can probably be tracked over any period of continuous 10Hz BSM transmission. The current behavioral results do not rule out that there may be some privacy preserved in high-density traffic situations where transmission is omitted for periods of time.

Future Direction

Future work is planned to refine both the kinematic and behavioral algorithms, ultimately linking the two together and attempting to determine a transmission scheme which preserves privacy while allowing the safety application of BSMs to still be useful. I will pursue machine-learning techniques such as clustering to determine whether more sophisticated algorithms can match behaviors across disconnects more accurately.

REFERENCES

[1] Y. Li: “An Overview of the DSRC/WAVE Technology”. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering Quality, Reliability, Security and Robustness in Heterogeneous Networks*, pp544-558, 2012.

[2] B. Cronin: “Vehicle Based Data and Availability”. United States Department of Transportation Intelligent Transportation Systems Joint Program Office Research and Innovative Technology Administration, October 2012. Accessed 30 August 2016. http://www.its.dot.gov/itspac/october2012/pdf/data_availability.pdf.

[3] I. Al-Anbagi, H. T. Mouftah: “WAVE 4 V2G: Wireless Access in vehicular environments for Vehicle-to-Grid applications”. *Vehicular Communications*, vol. 3, pp31-42, 2016.

[4] K. A. Hafeez, A. Anpalagan and L. Zhao, "Optimizing the Control Channel Interval of the DSRC for Vehicular Safety Applications," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3377-3388, May 2016. doi: 10.1109/TVT.2015.2440994

[5] S. Kato, M. Hiltunen, K. Joshi and R. Schlichting, "Enabling vehicular safety applications over LTE networks," *2013 International Conference on Connected Vehicles and Expo (ICCVE)*, Las Vegas, NV, 2013, pp. 747-752. doi:0.1109/ICCVE.2013.6799889

[6] J. Yin et al., “Performance Evaluation of Safety Applications over DSRC Vehicular ad hoc Networks”. *Proceedings of the 1st ACM International Workshop on Vehicular ad hoc Networks*, 2004, doi:10.1145/1023875.102387

[7] Safety Pilot Model Deployment and the members of the test conductor team. “Safety Pilot Model Deployment – Sample Data, from Ann Arbor, Michigan”. U.S. Department of Transportation’s (USDOT) Intelligent Transportation Systems (ITS) Joint Program Office (JPO), 28 October 2014.