Streaming Driving Behavior Data

Anas Basalamah
Department of Computer Engineering and
KACST TIC on GeoInformatics
Umm Al-Qura University
Makkah, Saudi Arabia
ambasalamah@uqu.edu.sa

Muhammad Aurangzeb Ahmad
Department of Computer Science and Engineering
Minneapolis, Minnesota 55455
mahmad@cs.umn.edu

Mohamed Elidrisi
Department of Computer Science and Engineering
Minneapolis, Minnesota 55455
elidrisi@cs.umn.edu

Saleh Basalamah
Department of Computer Engineering and
KACST TIC on GeoInformatics
Umm Al-Qura University
Makkah, Saudi Arabia
smbasalamah@uqu.edu.sa

Mohamed Mokbel *
Department of Computer Science and Engineering
Minneapolis, Minnesota 55455
mokbel@cs.umn.edu

ABSTRACT

People's driving behavior patterns have significant effects on the modern transportation systems. Public safety, traffic congestions, driving convenience are all affected by the driver's behaviors on the road. With the recent developments in data communications, streaming and mining technologies, developing driving behavior monitoring systems that can stream driving behavioral patterns, discover and predict traffic violations can be made possible. In this paper, we describe our vision towards the realization of such technologies from the view points of data steaming and analysis.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Spatial databases and GIS

General Terms

Design

Keywords

Sensor Networks, Driving Behavior, Stream Data

1. INTRODUCTION

Traffic accidents are a major cause of fatalities in many parts of the world. The Kingdom of Saudi Arabia (KSA), which has one of the highest road accident fatality rates in the world can be used as an illustrative example. In a 2009 report issued by KSA's Ministry of Interior's General Department of Traffic, the number of fatal accidents was 6,458 with a rate of 27 per 100,000 individuals, where the global fatality rate is 18 per 100,000 [20]. Over 80% of accidents occur because of bad driving habits such as excess speed and drivers disobeying traffic signals, where the total number of traffic violation was 8.97 million that year. In addition to the alarming number of lives lost, the estimated value of material loss is 13 billion Saudi Riyals a year. Moreover, the healthcare facilities are overburdened by large number of people injured in traffic accidents. Approximately, one third of all hospital beds in KSA are occupied by road accident victims [4]. In general young drivers are the largest contributor to traffic accidents mainly because of heedlessness. In the WHO's 2002 youth and road safety report, road traffic injuries are the leading cause of death globally among young people under 25 around the world. In KSA, drivers from ages 18 and 30, partake in 31.5% of all accidents [18]. Consequently there is a need to develop methods and infrastructure, which can be used to mitigate this problem not just in KSA but in the wider world as well.

In this paper we describe a vision of an Intelligent Transportation System that is capable of streaming driving behavioral patterns, discovering traffic violations and training drivers to correct reckless driving behavior. The major objectives of this vision are as follows::

1. **Sensing the traffic and the evironment.**
2. **Network connectivity and streaming data.**
3. **Mining data for intelligent decision making.**
4. **Preserving privacy of the participants.**

The most basic part of the infrastructure will be a system that will be installed on a vehicle; afterwards the onboard sensors (GPS Receiver, Cameras, etc which constitute a monitoring unit) are used to collect the information about the driver's behavioral patterns. They are then processed to identify violations and reckless behaviors. The system streams the collected data through a vehicular wireless network infrastructure to communicate with the system servers

*This work was done while the author was at sabbatical at Umm Al-Qura University

ACM SIGSPATIAL IWGSÕ12, November 6, 2012. Redondo Beach, CA, USA

for further analysis and processing. The system will employ a number of data mining and behavioral analytics techniques to analyze people's driving behavioral patterns.

To ensure privacy of drivers and preserve their identities privacy preserving systems must be deployed and extended in the current domain. This is a significant aspect to assure users that they are not being monitored which is critical for ensuring the widespread use of such a system.

Consider a scenario where a 16 year old who has just got his or her 1-year driving permit is given our monitoring system unit to log their driving patterns. When the permit-driver fails to stop at a stop sign, parks illegally, or maneuvers carelessly, the system will analyze the collected sensory data while comparing it with the predefined GIS map information, and detects this violation event and upload the streams data using the proposed network infrastructure for possible further processing and behavior analytics.

While the logistical setting is outside our research scope, this example demonstrate one instance on how our system could improve the driving behavior.

The rest of the paper raises the challenges that face the achievement of the system objectives in different domains of computer science research. We focus on the geospatial analytical and streaming aspect of the system.

2. OPEN QUESTIONS

The long-term objective of this research is to design and develop a system that monitors individual driving behavioral patterns, and provides data to preform data mining on the collective driving behavior. A system that helps the driver but at the same time preserve their privacy

A number of techniques and methods have been proposed for detecting detrimental traffic behaviors (e.g., excessive speeding, hard turning, reckless driving etc), using GPS receivers [5] , accelerometers [5] [19] [14], gyroscopes [19] and video cameras [5][14]. One main limitation of the previous techniques is that none of these techniques stream driving behaviors for further analysis as in our objectives.

2.1 Designing a Vehicular Sensor Network

A vehicular sensor network (VSN) platform provides a means of collecting/processing/streaming sensor data. Vehicles continuously collect sensor data from urban streets (e.g., images, accelerometer data, etc), which are then processed to search for information of interest (e.g., recognizing license plates, or inferring traffic patterns). The architecture of information access in a vehicular sensor network is mainly dependent on the underlying wireless access methods in vehicular environments. The infrastructure architecture integrates Access Points with the road network, such that they are distributed to provide vehicles with network connectivity. If vehicles are only equipped with inter-vehicular communications devices, an infra-structure-free network is used [17] [9]. Vehicles can also be equipped with a broadband wireless access method such as 2/3G and WiMax [13] [3]. The network infrastructure design highly depends on the streaming requirements. What access methods should be designed to provide efficient connectivity to vehicles under different road requirements. does the existing 2/3G network serve the design requirements?

In designing protocols for the vehicular networks, nodes have significantly different characteristics and demands from traditional ad hoc/mobile device networks [17]. Vehicles have much higher power reserves, and orders of magnitude larger in size and weight, and can support significantly heavier computing (and sensorial) components with higher data connectivity rates. Vehicles travel at speeds up to one hundred miles per hour, making sustained, consistent vehicle-to-vehicle communication difficult to maintain. However, "existing statistics" of vehicular motion, such as tendencies to travel together or traffic patterns during commute hours, help in maintaining connectivity across mobile vehicular groups. There are many wireless ad hoc routing protocols: proactive routing , reactive routing, geographic routing, and hybrid geo-graphic routing, and yet they cannot directly be used due to high mobility and non-uniform distribution of vehicles, which causes intermittent connectivity. However, in vehicular networks, geographic or hybrid geographic routing protocols are often preferred. Also, the carry-and-forward strategy is used to overcome intermittent connectivity; when disruption happens, a node stores a packet in its buffer and waits until connectivity is available [8]. What routing protocols are most suitable and how can we maintain them and contain abrupt disconnections? Would intermitted connectivity suffice? how tolerant is the system to network delay?

One feature of vehicular networks that was rarely discussed in the literature is the road network specific design considerations. The carry-and-forward strategy, for example, can use the road network specific knowledge to set better locations of access points and provide better coverage. In road networks, there are some specific locations where a high percentage of vehicles pass through in their daily routes. Allocating access points at those location can cover the majority of city vehicles depending on the application requirements. How many access points are required to cover vehicles? what is the optimal access point placement mechanism?

Additionally, vehicles follow a certain mobility pattern that is a function of the underlying roads, the traffic lights, the speed limit, traffic condition, and drivers' driving behaviors. Because of the particular mobility pattern, evaluation of vehicular network protocols only makes sense from traces obtained from the pattern. How can we develop country specific mobility patterns, where vehicles follow different city specific patterns?

2.2 Mining Driving Sensor Data

Mining vehicular sensor data is a unique problem that combines existing challenges in distributed and data stream mining. There are existing systems that have been successfully designed and/or deployed for mining vehicle sensor data. MineFleet was recently deployed for fleet management by Agnik Inc [15]. For more detailed review, Gaber et al. [11] provided an analysis of the major issues on data mining streams that apply to mining vehicle data.

Mining the vehicle data can provide answers to either the individual drivers or the policy makers at transportation agencies. In both cases, the problem could be either real-time or post collection mining depending on the task at hand. The spatial and temporal aspect of the data will require more innovation in applying existing data mining techniques.

The data streamed will create a unique resource for analysis of driving trends, socials behaviors, and traffic dynamics. By utilizing existing data mining algorithms and extending them in new domains, a wide variety of tools and applications can be explored. The open questions that relate to the collected data span a wide area of questions that include identifying reckless behavioral patterns, dealing with congestion, accidents and speed analysis.

2.3 Protecting Users's Privacy

Privacy is a major concern for telematic applications in the consumer market. Most people do not want to be tracked everywhere they go. Existing work on location privacy focuses mainly on hiding user locations either through many method like access control [6], location perturbation [10] to mention a few.

However, previous work is limited as: (1) the focus is only on snapshot user locations with no direct extension to continuous user movement over a trajectory, (2) it lacks adversary attack models, (3) it is not clear how data mining services will be delivered after hiding the user location. Very few work has focused on the trajectory movement, e.g., [12], however, that was done for very specific functionality, and not for general data mining techniques like what we aim to do in our monitoring system.

3. DIRECTIONS

In the remainder of the paper we plan to demonstrate our vision for this research. The 3 main directions are: 1) building a streaming network architecture that allow us to share the data collected in real-time and effective matter, 2) mine the stream collected data with the knowledge that the data will be geospatial and temporal in nature. 3) build a system that assures privacy of the users without a loss of context.

3.1 Streaming Network Architectures:

Architecture: In order to support streaming driving behavior data, the network infrastructure should be equipped with the various wireless technologies that deliver seamless connectivity and wide coverage. Hence, heterogeneous wireless communication (for both vehicle-to-vehicle and vehicle-to-infrastructure) and well as broad-band access will be considered. Fortunately, vehicular communication can be realized using various wireless technologies. Each wireless technology has its own advantages and disadvantages. For vehicle-to-vehicle communication, 802.11 WiFi, 802.15.4 Zigbee and 802.15 Bluetooth technologies are suitable candidates (in addition to DSRC and WAVE). On the other hand, WMN, 802.16 WiMAX and cellular networks can be considered for vehicle-to-infrastructure communication. The optimal selection of the wireless technology (or load distribution) should be performed based on the service coverage, resource utilization, type of service, loading conditions, load balancing and quality-of-service (QoS).

Routing Protocols: The behavior monitoring application needs sets of services such as unicasting, broadcasting, geo-casting, and multicasting routing in order to facilitate the data collection and content dissemination. Location-based routing can be used to minimize the delay and the overhead of route discovery and setup time. Also, connectivity in sparse vehicular networks is another challenge for data routing. Utilizing designed wireless infrastructure in routing packets can significantly mitigate the connectivity problem. High efficient routing protocols for vehicular communication networks with low overhead should be investigated along with their scalability. Finally support for carry and forward strategies should be investigated to provide seamless services and to overcome intermittent connectivity.

Road network specific design: The behavior monitoring application has different network connectivity requirements. For example, the frequency and size of uploads depends on the real-timeliness and processing requirement of the violation recognition algorithm. When the application requirements allow for intermittent connectivity, designing a vehicle to infrastructure only network can be made possible if careful understanding of road network characteristics and driver behavior are considered.

Developing simulation mobility patterns: A critical aspect when studying VANETs, is the need for a mobility model which reflects, as close as possible, the real behavior of vehicular traffic. It would be desirable for a trustworthy simulations to consider motion-constraints (road topology, street characterization, vehicle class dependent constraints, traffic signs, etc.) and also traffic generator (vehicle-to-vehicle interactions, vehicle-to-road interactions, acceleration and deceleration, overtaking, etc.) in modeling vehicular movements. The properties of both motion constraints and traffic generators vary significantly depending on the driving behavior patterns, local congestions and local road network.

3.2 Mining of Vehicular Sensor Data:

While pattern mining of driver behavior and classification of such behaviors as normal vs. violation behaviors has been the main tasks that have been addressed before, we note that this domain has not really benefited from the burgeoning field of Social Behavioral Analytics which seeks to analyze not just the behavioral data of individuals (in this domain drivers) but also include social factors like what are the contexts in which they exhibit such behaviors.

There is ample literature in the behavioral and social sciences which attests that the behaviors of people is context dependent and is usually contingent upon their social networks [7]. In our previous work we have addressed similar problem in the context of online massive datasets [2] [16] In a pevious work we have demonstrated that algorithms inspired from social science theories can be used for predicting for the formation of social relationships [1]. Similarly techniques in social computing can be extended in the traffic domain where the driving patterns of habitual offenders, thus exhibiting deviance, can be automatically extracted and predicted in advance. Such an analysis can yield both micro and macro level insights. At the macro level one can determine the general driving behaviors of people which can be used ot adjust traffic policies and protocols. At the microlevel one can make recommendations to the individual drivers based on his or her history as well as in the presence of other drivers.

3.3 Protecting Users's Privacy

Trajectory-based location anonymizion: Most of existing location privacy techniques focus on hiding single location information. It is more challenging to hide the location in our case as we have trajectory information for each user, i.e., a sequence of consecutive locations for each user. Thus, we need to develop new trajectory-based location privacy techniques that focus on hiding the information of the trajectory as a whole. In the mean time, the developed location anonymize r should take into account the adversary attack models developed above.

Privacy-preserving data mining techniques: Hiding the trajectory location information is not the end of the story. We still need to get the desired service in terms of monitoring driving patterns and mining the collected data. Existing data mining techniques rely on the data availability. With our developed trajectory-based location anonymization techniques, we need to develop new data mining techniques that can still be applicable for the case where the data is not completely available.

Quality vs. Privacy trade-off: The objective is to find a suitable trade-off between quality and privacy where the system can have access to part of the data (e.g., an anonymized data), and hence can give data analysis with reasonable quality. We would like to achieve something like giving the system an access to 90% of the data to produce 90% quality of data analysis. The system will have a chart for this trade-off, then, an incentive model would be developed to encourage users to reveal more of their data.

4. CONCLUSIONS

In this paper, we have presented the challenges that face the development of a driving behavior monitoring systems from the aspects of communications, streaming and mining of sensed data. We also presented our vision towards the realization of the technology.

5. ACKNOWLEDGMENTS

The work was funded by the KACST National Science and Technology Plan under grant #11-INF2062-10

6. REFERENCES

[1] M. A. Ahmad, Z. Borbora, J. Srivastava, and N. S. Contractor. Link prediction across multiple social networks. In *ICDM Workshops'10*, pages 911–918, 2010.

[2] M. A. Ahmad, B. Keegan, D. Williams, J. Srivastava, and N. S. Contractor. Trust amongst rogues? a hypergraph approach for comparing clandestine trust networks in mmogs. In *ICWSM'11*, pages –1–1, 2011.

[3] J. Ahnn, U. Lee, H. Moon, and M. Gerlag. Senster: scalable smartphone based vehicular sensor networking systems.

[4] S. Ansari, F. Akhdar, M. Mandoorah, and K. Moutaery. Causes and effects of road traffic accidents in saudi arabia. *Public Health -London- Society of Public Health*, 114(1):37–40, 2000.

[5] K. Baldwin, D. Duncan, and S. West. The driver monitor system: A means of assessing driver performance. Technical Report 25, Johns Hopkins, APL Technical Digest.

[6] C. Bettini, X. Wang, and S. Jajodia. Protecting privacy against location-based personal identification. *Secure Data Management*, pages 185–199, 2005.

[7] A. Campbell. The rise of people-centric sensing. In V. Garg, R. Wattenhofer, and K. Kothapalli, editors, *Distributed Computing and Networking*, volume 5408 of *Lecture Notes in Computer Science*, page 9. 2009.

[8] Z. D. Chen, H. Kung, and D. Vlah. Ad hoc relay wireless networks over moving vehicles on highways. In *MobiHoc '01: Proceedings of the 2nd ACM MobiHoc*, pages 247–250, New York, NY, USA, 2001. ACM.

[9] M. Dikaiakos, S. Iqbal, T. Nadeem, and L. Iftode. Vitp: an information transfer protocol for vehicular computing. In *Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks*, pages 30–39. ACM, 2005.

[10] M. Duckham and L. Kulik. A formal model of obfuscation and negotiation for location privacy. *Pervasive Computing*, pages 243–251, 2005.

[11] M. M. Gaber, A. Zaslavsky, and S. Krishnaswamy. Mining data streams: a review. *SIGMOD Rec.*, 34(2):18–26, June 2005.

[12] A. Gkoulalas-Divanis, V. Verykios, and M. Mokbel. Identifying unsafe routes for network-based trajectory privacy. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*. Citeseer, 2009.

[13] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, and S. Madden. Cartel: a distributed mobile sensor computing system. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, pages 125–138. ACM, 2006.

[14] T. Imkamon, P. Saensom, P. Tangamchit, and P. Pongpaibool. Detection of hazardous driving behavior using fuzzy logic. In *5th International Conference ECTI-CON 2008.*, volume 2, pages 657 –660, may 2008.

[15] H. Kargupta, K. Sarkar, and M. Gilligan. Minefleet®: an overview of a widely adopted distributed vehicle performance data mining system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 37–46. ACM, 2010.

[16] B. Keegan, M. A. Ahmad, D. Williams, J. Srivastava, and N. S. Contractor. Dark gold: Statistical properties of clandestine networks in massively multiplayer online games. In *SocialCom/PASSAT'10*, pages 201–208, 2010.

[17] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, and A. Corradi. Dissemination and harvesting of urban data using vehicular sensing platforms. *Vehicular Technology, IEEE Transactions on*, 58(2):882–901, 2009.

[18] Ministry of Interior, Department of Traffic. 2009 statistical report, 2009.

[19] D. Mitrovic. Reliable method for driving events recognition. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):198–205, 2005.

[20] W. H. Organization. Global status report on road safety: time for action. Technical report, WHO, 2009.