Privacy Preserving Driving Style Recognition

Nicholas Rizzo , Ethan Sprissler, Yuan Hong, and Sanjay Goel
Department of Information Technology Management
University at Albany, SUNY, Albany, NY 12222
{nrizzo, esprissler, hong, goel}@albany.edu

***Abstract*—In order to better manage the premiums and encourage safe driving, many commercial insurance companies (e.g., Geico, Progressive) are providing options for their customers to install sensors on their vehicles which collect individual vehicle's traveling data. The driver's insurance is linked to his/her driving behavior. At the other end, through analyzing the historical traveling data from a large number of vehicles, the insurance company could build a classifier to predict a new driver's driving style: aggressive or defensive. However, collection of such vehicle traveling data explicitly breaches the drivers' personal privacy. To tackle such privacy concerns, this paper presents a privacy-preserving driving style recognition technique to securely predict aggressive and defensive drivers for the insurance company without compromising the privacy of all the participating parties. The insurance company cannot learn any private information from the vehicles, and vice-versa. Finally, the effectiveness and efficiency of the privacy-preserving driving style recognition technique are validated with experimental results.**

I. Introduction

One of the engineering innovations that will have a transformative impact on the society over the next few decades is the Connected Vehicles initiative [4]. The connectivity is enabled through the use of wireless communication over a dedicated spectrum to create local ad hoc networks of vehicles that are then able to communicate with other vehicles in its neighborhood, as well as with traffic infrastructure. The communication network provides opportunity for mobile devices or inertial sensors mounted in the vehicles to collect data regarding vehicle travel, driver behavior, and location data [19], [31]. This data can be processed for implementing automatic collision avoidance, optimizing traffic in real time, and planning for infrastructure needs. While, these data sets can improve safety, reduce emissions, and reduce driver wait time, they can also be used to uncover sensitive personal information. For instance, location data can be used in criminal investigations, driver behavior can be used to determine fault in accidents, and a person's lifestyle choices can be revealed by processing the data with mapping software. The privacy challenge needs to be addressed with a careful balance between the utility of the collected data and the protection of personal information as well as corporate proprietary information.

The insurance industry can leverage vehicle travel data to determine their premiums by correlating driver behavior and accidents. For instance, vehicle data can be used to create a classification system that can classify drivers as defensive or aggressive [19]. In order to better manage their premiums and also encourage safe driving, many commercial insurance companies (e.g., GEICO and Progressive) are providing options for customers to install sensors on their vehicles that can collect the vehicle's operational parameters (such as braking, acceleration, speed, etc.) [20]. Their insurance is linked to their driving behavior. At the other end, through analyzing the historical travel data from a large number of vehicles, the insurance company could build a classifier to determine the thresholds for aggressive vs. defensive driving. This would allow the companies to better understand the risks associated with each of their drivers and be able to tailor premiums based on the risks. Insurance companies could also give drivers "report cards" to help drivers better understand their own driving habits. However, collection of the vehicle traveling data to make this happen explicitly breaches the drivers' personal privacy. For instance, the accelerating & braking data, trips, turning behaviors, risky driving hours, and even the geographical locations the driver can be inferred from the data violating their personal privacy. The goal of this research is to securely build a classifier that accurately predicts any given driver's driving style (aggressive or defensive) without compromising any participating party's privacy.

There are two phases in driving style recognition: (1) building a classifier with vehicles' historical traveling data; (2) predicting the driving style for new drivers using the classifier.

1) In phase (1), some drivers' historical vehicle traveling data are analyzed to train the classifier. As part of this process, each record has a class label applied to it, the characteristics of an aggressively labeled driver would be based on the records of individuals who had received speeding tickets or been involved in an accident, while a defensive driver would be an individual who had not been in such situations. Each driver privately holds a record of his/her vehicle traveling data with the attributes such as average acceleration (m/s^2), average deceleration in the braking events (m/s^2), average turning (degrees), # of risky driving hours, and # of trips [19]. In reality, both of the driver and insurance company know the driver's class label in the training data from the traffic tickets, reported accidents, etc. Therefore, such multiparty classifier training process has a hybrid scenario of data partitions: first, all the drivers' records are horizontally partitioned – each driver privately holds a record, including the class label [27]; second, the overall data (attributes) is also vertically partitioned [27] into two shares – the insurance company holds a share (the attribute class label) while all the drivers jointly holds both shares, including the class label. Finally, the output of phase (1) is a trained classifier,

This work is supported in part by the SUNY STEM Passport Program. Corresponding Author: Yuan Hong

978-1-5090-0264-1/15/$31.00 ©2015 IEEE

DOI 10.1109/ICCVE.2015.42

privately held by the insurance company.

2) In phase (2), the insurance company privately holds its classifier whereas the new driver privately holds his/her vehicle traveling data. Since the two parties privately own different attributes, it is a vertical partition case in each individual driving style recognition. Then, they jointly predict the class label (aggressive or defensive) with their private inputs.

Since the data partitions in the previous two phases are mixed with both horizontal and vertical partitions, the prior works on privacy-preserving classification [18], [27], [28], [29] are not directly applicable to this research problem.

A. Research Contributions

In this paper, we develop privacy-preserving techniques for two phases of driving style recognition based on decision tree induction [23]. More specifically, in phase (1), each driver only knows its vehicle traveling data (a record) and its class label; the insurance company only knows all the drivers' class labels, and the final output – a decision tree. In phase (2), each new driver only knows its vehicle traveling data (a record) and learns nothing but the driving style recognition result; the insurance company only knows its input (the decision tree) and also learns nothing but the driving style recognition result. Thus, the main contributions are summarized as below:

1) We propose two secure communication protocols under secure multiparty computation (SMC) [30], [6] to implement the two phases of driving style recognition (based on decision trees) for the participating parties without private information disclosure.
2) We analyze the privacy risks for all the participating parties in the secure communication protocols for both classifier training and driving style prediction.
3) We experimentally validate the performance of our proposed approaches on the synthetic datasets generated following the format of data collected in [19].

The remainder of this paper is organized as follows. We first briefly review the related work in Section II. Then, we present the algorithms and demonstrate the experimental results in Sections III and IV respectively. Finally, we conclude this paper and discuss the future work in Section V.

II. Related Work

Vehicle traveling data has been commonly collected for analysis in Intelligent Transportation Systems (ITS). Ly et al. [19] demonstrated a methodology to collect vehicle data (e.g., accelerating, braking, turning) using inertial sensors. Hull et al. [16] collected the vehicle data with his CarTel system. There has also been expansive work into adaptive cruise control, which uses prediction algorithms to adapt to the road based on curve patterns [31]. Moreover, Rass et al. [24] formulated a system to provide feedback on the driving habits. Also, other research has focused on driver modeling & evaluation [21], [17] and maneuver recognition [25]. While there has been a variety of ventures into many ITS applications, only a surprising few of these tackled such arising privacy concerns, including: Hoh et al. [8] proposed using the virtual trip lines to monitor traffic conditions while preserving privacy. Checkoway et al. [2] examined the attack vectors for hackers to infiltrate vehicle through its Electronic Control Unit. Han et al. evaluated authentication methods for securely integrating mobile devices in vehicular networks [7]. To the best of our knowledge, privacy risks in driving style recognition have not been systematically studied.

Privacy-preserving schemes are generally developed based on data transformation and/or secure computation. The former one transforms the original data to a privacy-compliant format and minimizes the utility loss in the process of data transformation (e.g., k-anonymity [26], differential privacy [3]). The latter makes two or more parties jointly compute a function possible without revealing private data to each other (formally defined as Secure Multiparty Computation [30]). The function can be as simple as sum or as complex as big data analysis/mining [1], [18], [9]. Several researchers have addressed the privacy concerns in other contexts, such as classification [18], [29], [27], location based services [5], search engine queries [10], [11], [14], scheduling [15], transportation [13], and smart grid [12]. Following a similar line of research, we develop a privacy-preserving driving style recognition technique that can analyze the vehicle traveling data (e.g., an insurance company) without breaching participating parties' privacy.

III. Secure Communication Protocols

In this paper, we assume that the adversaries are semi-honest. The semi-honest model in Secure Multiparty Computation (SMC) [30], [6] defines that the adversaries are honest to follow the a given protocol, but are curious to infer private information from each other. Two secure communication protocols will be given for two phases of driving style recognition in Sections III-A and III-A1 respectively.

A. Phase (1): Privacy Preserving Classifier Training

In phase (1) of driving style recognition, a set of drivers and the insurance company jointly derive a decision tree based on the drivers' vehicle traveling data and class labels (aggressive or defensive). In this scenario, we have:

- Every driver's class label in the training data is known by both the driver and the insurance company. Since the "Aggressive" class label in the historical data can be derived from the "traffic tickets or accidents" which is indeed known by both parties in real world.
- All drivers and the insurance company know the name of every attribute in the dataset, such as Acceleration Events (#), Average Acceleration (m/s^2), Braking Events (#), and Average Braking (m/s^2). Insurance company initializes the attributes in the sensors which are also known to the drivers in real world.
- Every attribute has a threshold to divide the values into two categories: "less than the threshold" or "no less than the threshold". For example, the average deceleration when braking can be less than a threshold $4.9m/s^2$ or no less than a threshold. Such split is used to determine the branches of a node on the decision tree. We assume that every attribute's threshold is known to all the drivers and the insurance company

978-1-5090-0264-1/15/$31.00 ©2015 IEEE

DOI 10.1109/ICCVE.2015.42

(e.g., thresholds of speed/mileage/acceleration/braking can be available as public information).

Note that our privacy-preserving decision tree training is extended from the ID3 Algorithm [23], [18], which iteratively finds the best attribute to split values based on its threshold (as the current node of the tree) by comparing the entropy or information gain of all the remaining attributes in the classification results. In this paper, we choose the entropy as the measure of uncertainty in the threshold based split $H = -\sum_{x \in X} p(x) log p(x)$ where X is the set of classes ("Aggressive" or "Defensive") and $p(x)$ is the proportion of the number of elements in class x to the number of elements in all the data. Therefore, in a distributed manner, each driver/vehicle owns a record and they should securely sum their shares for every $p(x)$. In this section, we first present the Secure Sum algorithm which is iteratively invoked by the protocol of privacy-preserving classifier training.

1) Secure Sum: The secure sum algorithm is developed using Homomorphic Cryptosystem (e.g., Paillier [22]). It begins by having the insurance company I generate a key pair: a public key pk and a private key sk. The insurance company (party I) then sends the public key pk to all m drivers, denoted as $D_1, \ldots, D_m$. D_1 then encrypts its share and passes along the encrypted data to the next driver (w.l.o.g, say D_2). D_2 then computes their encrypted sum to the previous number and this is passed through all the m drivers in the circuit. After this, the encrypted sum is passed back to the insurance company who decrypts it with the private key sk to get the sum. As shown in Algorithm 1, all the parties' data remains private while only allowing the insurance company to obtain the sum.

Algorithm 1 Secure Sum

```
Input: $m$ drivers' share of $p(x)$: $p(x)_1, \ldots, p(x)_m$
Output: insurance company $I$ learns $p(x) = \sum_{j=1}^{m} p(x)_j$
$I$ generates a pair of public-private key $(pk, sk)$ and sends the public key $pk$ to $P_1, \ldots, P_n$
for $i = 1, \ldots, m$ do
    $D_i$ encrypts $p(x)_i$ using $pk$ to get $E[p(x)_i]$, and computes $E[\sum_{j=1}^{i} p(x)_j] = \prod_{j=1}^{i-1} E[p(x)_j] * E[p(x)_i]$ (ensured by Homomorphic Property [22])
    $D_i$ sends $E[\sum_{j=1}^{i} p(x)_j]$ to the next party $P_{i+1}$ (if $i = m$, the next party is $I$)
end for
$I$ decrypts $E[\sum_{j=1}^{m} p(x)_j]$ with its private key $sk$ to obtain $\sum_{j=1}^{m} p(x)_j$
```

2) Secure Communication Protocol: In the secure communication protocol for classifier training, the insurance company I repeatedly finds the best attribute (that has the smallest entropy H; viz. lowest uncertainty) as the current node during the construction of the tree. Note that the key value $p(x)$ in the entropy $H = -\sum_{\forall x \in X} p(x) \log p(x)$ is split into m shares $p(x)_1, \ldots, p(x)_m$, held by m vehicles respectively. In each iteration, each of the remaining attributes' $p(x)$ is securely summed by all m vehicles and the insurance company I (Algorithm 1). This use of the secure sum ensures that I is only ever able to learn the $p(x)$ of each attribute while preserving the privacy of each of the m vehicles.

In turn, the only information that is learned by each of m vehicles is the pk that is sent by the insurance company I. The purpose of this approach is to provide a mechanism to securely sum the data needed to compute the entropy values for each attribute while obfuscating all vehicle identifying characteristics in $p(x)_1, \ldots, p(x)_m$. In addition, the use of the secure sum ensures that no other parties will be able to uncover the identity of any vehicle or the vehicle's collected data.

In the final phase of this algorithm, I locally computes the entropy H of all the remaining attributes, and selects the attribute with the smallest entropy value as the current node. This process is performed iteratively until all the leaf nodes of the decision tree have $p(x) = 1$. The details of the secure communication protocol is given in Algorithm 2

Algorithm 2 Privacy Preserving Classifier Training

```
Input: $m$ is the total number of drivers/vehicles
Output: Decision tree $T$
while existing a leaf node's $p(x)! = 1$ do
    for $i = 1, \ldots, m$ do
        Driver $D_i$ computes the share of $p(x)$ for all the remaining attributes
    end for
    Securely sum shares of $p(x)$ (Algorithm 1) for all the remaining attributes (only party $I$ knows that)
    $I$ computes entropy $H = -\sum_{i=1} p(x) log p(x)$ for all the remaining attributes
    $I$ selects the best attribute (smallest entropy) as the root or leaf node (leaf node is added along the branch of the tree with $p(x)! = 1$)
end while
```

The final product of this algorithm consists of a decision tree which contains a set of the attributes from the dataset. This tree outlines the pathways which represent determinable outcomes based on any given driver's collection of vehicle traveling data. This tree will be used to help securely predict a given new driver as either aggressive or defensive.

B. Phase (2): Privacy-preserving Driving Style Recognition

The insurance company has the means to classify driving style/behavior for given drivers, but has an equal interest in keeping its decision tree T private. To this end, we develop another secure communication protocol to predict the driving styles without sharing information between the insurance company and the new driver, detailed below.

Definition 1 (Aggressive Path in Decision Tree T): is defined as a path in the decision tree T that leads to the class of aggressive driving.

Letting $|T|$ be the number of aggressive paths in T, with all the thresholds of the attributes in T, each aggressive path can be represented by an n-digit binary vector:

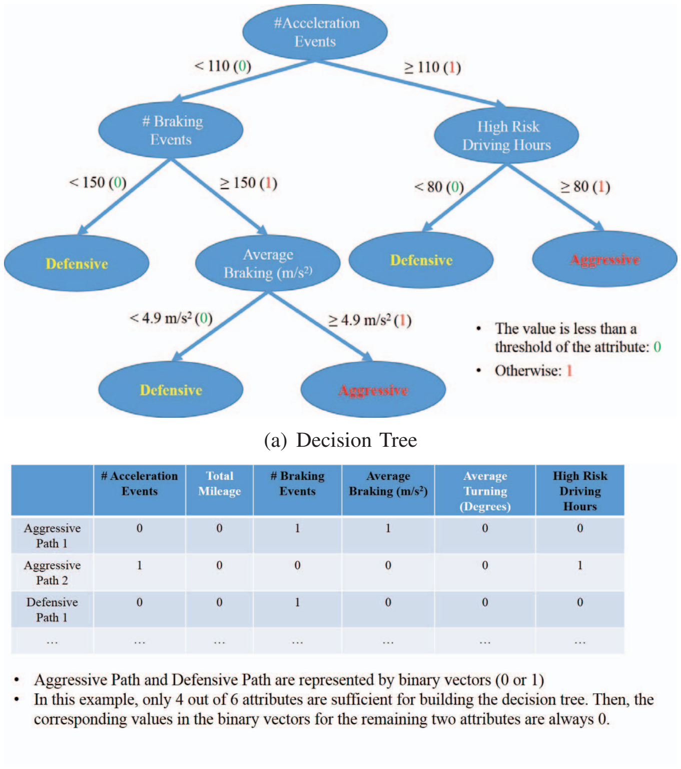
$$\forall i \in [1, |T|], \vec{c_i} = [c_{i1}, c_{i2}, \ldots, c_{in}] \quad (1)$$

where $\forall c_{ij} \in \{0, 1\}$. Note that, in the decision tree T, $c_{ij} = 1$ means the child value of the j^{th} attribute (out of all n attributes in total) along the aggressive path $\vec{c_i}$ exceeds the threshold value; otherwise, $c_{ij} = 0$. For instance, in Figure 1,

978-1-5090-0264-1/15/$31.00 ©2015 IEEE

DOI 10.1109/ICCVE.2015.42

there are six attributes in total used for training decision tree, and four of them are utilized to build the tree T (as shown in Figure 1(a)). Two paths in T: "# of acceleration events $< 110 \longrightarrow$ # of braking events $\geq 150 \longrightarrow$ average braking $(m/s^2) \geq 4.9m/s^2$" and "# of acceleration Events $\geq 110 \longrightarrow$ high risk driving hours ≥ 80" can lead to "Aggressive". Then, they can be represented as two binary vectors $(0, 0, 1, 1, 0, 0)$ and $(1, 0, 0, 0, 0, 1)$ respectively. Other paths in T are simply considered as "Defensive Paths" which can be also represented as n-digit binary vector in a similar way.



(a) Decision Tree

	# Acceleration Events	Total Mileage	# Braking Events	Average Braking (m/s²)	Average Turning (Degrees)	High Risk Driving Hours
Aggressive Path 1	0	0	1	1	0	0
Aggressive Path 2	1	0	0	0	0	1
Defensive Path 1	0	0	1	0	0	0
...	...	...	...	...	...	...

- Aggressive Path and Defensive Path are represented by binary vectors (0 or 1)
- In this example, only 4 out of 6 attributes are sufficient for building the decision tree. Then, the corresponding values in the binary vectors for the remaining two attributes are always 0.

(b) Aggressive and Defensive Paths

Fig. 1. An Example of Decision Tree and Aggressive/Defensive Paths

The insurance company I now owns the decision tree T, the number of aggressive paths $|T|$, as well as the path(s) that identify aggressive driving behavior $\vec{c_{|T|}}$. Since the insurance company privately possesses such information, the aggressive paths must be encrypted for computation. Specifically, I generates a public/private key pair (pk, sk) based on the Paillier cryptosystem [22]. I encrypts the aggressive paths from the decision tree and the total number of aggressive paths $|T|$ as well as the inner products of all the aggressive paths (which is the total number of "1" in each binary vector) using the public key pk such that: $E(\vec{c_1})$, ..., $E(\vec{c_{|T|}})$ and $E(\vec{c_1} \cdot \vec{c_1})$, ..., $E(\vec{c_{|T|}} \cdot \vec{c_{|T|}})$ are both then transmitted along with pk to the new driver/vehicle D.

At the other end, similar to the aggressive/defensive paths, the new driver D privately holds a:

Definition 2 (Vehicle Traveling Vector $\vec{v}$): is an n-digit binary vector: $\vec{v} = [v_1, v_2, \ldots, v_n]$ with 0 representing the attributes with values below the threshold, and otherwise 1.

After receiving pk and the ciphertexts from the insurance company I, the driver/vehicle D then securely computes the following scalar products with the ciphertexts and its vector $\vec{v}$:

$$\forall i \in [1, |T|], E(\vec{c_i} \cdot \vec{v}) = E(c_{i1})^{v_1} * E(c_{i2})^{v_2} * \cdots * E(c_{in})^{v_n} \tag{2}$$

Then, driver/vehicle D encrypts $\vec{v}$ and computes:

$$\forall i \in [1, |T|], E(\vec{c_i} \cdot \vec{v} - \vec{c_i} \cdot \vec{c_i}) = \frac{E(\vec{c_i} \cdot \vec{v})}{E(\vec{c_i} \cdot \vec{c_i})} \tag{3}$$

If any of $\forall i \in [1, |T|], \vec{c_i} \cdot \vec{v} - \vec{c_i} \cdot \vec{c_i}$ equals 0, the vehicle traveling vector $\vec{v}$ would match the corresponding aggressive path, and the driver D is predicted as an aggressive driver. If all of $\forall i \in [1, |T|], \vec{c_i} \cdot \vec{v} - \vec{c_i} \cdot \vec{c_i}$ are non zero, the vehicle traveling vector $\vec{v}$ would not match any aggressive path, and the driver D can be predicted as a defensive driver. To minimize information disclosure, the driver permutes all the ciphertexts $\forall i \in [1, |T|], E(\vec{c_i} \cdot \vec{v} - \vec{c_i} \cdot \vec{c_i})$ and send them to the insurance company I one by one, and I decrypts a ciphertext immediately. As long as a 0 is found, conclude D as an aggressive driver and terminate the protocol (no more ciphertext will be sent). If no 0 is found after examining all $|T|$ results, then conclude D as an defensive driver. I can share the classification result to D if necessary.

```
Algorithm 3 Privacy Preserving Driving Style Recognition
Input: Insurance company $I$ and a new driver $D$;
       $D$'s vehicle traveling vector $\vec{v}$;
       $T$ represents the complete decision tree;
       The number of aggressive paths $|T|$;
       All the aggressive paths $\vec{c_1}, \ldots, \vec{c_{|T|}}$
Output: The new driver is aggressive or not
    {A random nonce is generated for every single encryption}
Party $I$ generates a public/private key pair based on Paillier Cryptosystem $(pk, sk)$
$I$ encrypts $\vec{c_1}, \ldots, \vec{c_{|T|}}$, $|T|$, and inner products $\vec{c_1} \cdot \vec{c_1}$, ..., $\vec{c_{|T|}} \cdot \vec{c_{|T|}}$ using $pk$ to obtain $E(\vec{c_1}), \ldots, E(\vec{c_{|T|}})$, $E(\vec{c_1} \cdot \vec{c_1})$, ..., $E(\vec{c_{|T|}} \cdot \vec{c_{|T|}})$, and sends the ciphertexts and $pk$ to the driver $D$
$D$ encrypts $\vec{v}$ and computes the ciphertexts of $|T|$ scalar products $E(\vec{c_1} \cdot \vec{v})$, ..., $E(\vec{c_{|T|}} \cdot \vec{v})$ using Equation 2
$V$ computes $E(\vec{c_1} \cdot \vec{v} - \vec{c_1} \cdot \vec{c_1})$, ..., $E(\vec{c_{|T|}} \cdot \vec{v} - \vec{c_{|T|}} \cdot \vec{c_{|T|}})$ using Equation 3 and permutes them
for $i = 1, \ldots, |T|$ do
    $D$ sends the permuted ciphertext $E(\vec{c_i} \cdot \vec{v} - \vec{c_i} \cdot \vec{c_i})$ to $I$
    $I$ decrypts the current $E(\vec{c_i} \cdot \vec{v} - \vec{c_i} \cdot \vec{c_i})$ using its private key $sk$ to get $\vec{c_i} \cdot \vec{v} - \vec{c_i} \cdot \vec{c_i}$
    if $\vec{c_i} \cdot \vec{v} - \vec{c_i} \cdot \vec{c_i} = 0$ then
        $D$ is an aggressive driver and terminate the algorithm
    end if
end for
if $\forall i \in [1, |T|], \vec{c_i} \cdot \vec{v} - \vec{c_i} \cdot \vec{c_i} \neq 0$ then
    $D$ is a defensive driver
end if
```

Upon completion, each driver has the ability to access his or her computed rating of either aggressive or defensive. I has sole possession of the decision tree T developed from training data and is the only party which is able to view all of the pathways which lead to an aggressive classification. The driver

978-1-5090-0264-1/15/$31.00 ©2015 IEEE

DOI 10.1109/ICCVE.2015.42

D, on the other hand, is the only party able to access $\vec{v}$, keeping specific driving behavioral data private from I. Ultimately, both parties will have access to the computed classification result. However, the insurance company can only infer some trivial information from D such as how many values in $\vec{v}$ has met or exceeded the corresponding attributes' threshold.

IV. Experimental Results

We implemented application of privacy-preserving driving style recognition in Java on a PC with AMD FX-4350 4.55 GHZ CPU and 16G RAM. Synthetic datasets are generated falling into a similar value range as [19]. 7500 drivers' vehicle traveling data are generated for training classifier (with class labels in the training data) while 2500 drivers' traveling data are generated for predicting the driving style by the classifier (without class labels in the dataset). These records simulated driving activity over a 6-month period and featured 9 attributes:

- total number of trips taken
- total mileage driven
- the number of acceleration events
- the average amount of acceleration
- the number of braking events
- the average deceleration when braking
- the average number of degrees turned
- the total number of hard braking events
- the hours driven in the highest risk time (0 to 4 AM)

The cryptographic keys are generated with lengths of 512 and 1024-bit using Paillier Homomorphic Cryptosystem [22].

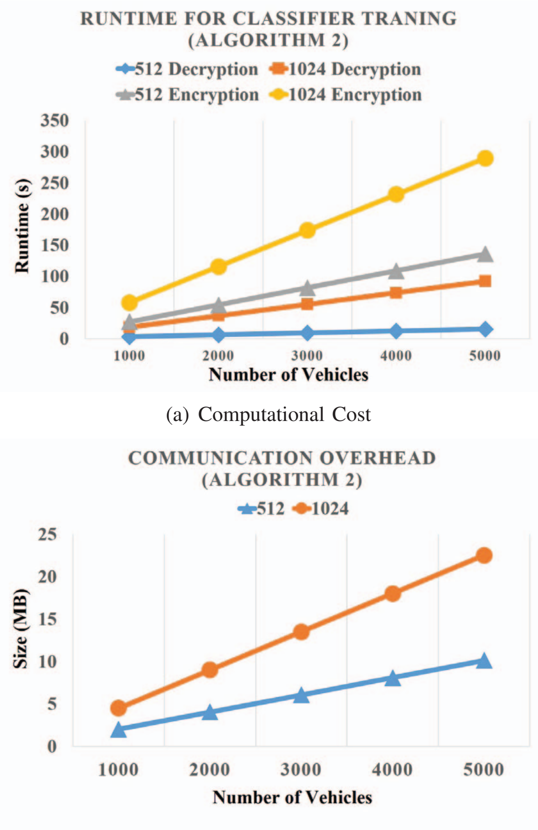
For examining the computational costs, we tested the overall runtime of the protocols including encryption, computation and decryption. For examining the communication overheads, we tested the overall bandwidth consumption, which is equivalent to the overall size of the ciphertexts and plaintexts to be transmitted among all the distributed parties in the protocols. Due to the novelty of the data partition scenario and protocols devised for driving style prediction, there are no available benchmarks to compare against.

A. Classifier Training

Algorithm 2 securely trains a decision tree out of the privately held distributed records. We conducted a group of experiments for classifier training (Algorithm 2) using 1000, 2000, 3000, 4000 and 5000 drivers' traveling data featuring 9 attributes. In each group of experiments, we tested the runtime for the encryption and decryption as well as the size of all the ciphertexts. As shown in Figure 2(a) and 2(b), both the computational cost and the communication overheads present a linear increase trend as the number of vehicles increases.

B. Driving Style Recognition

Algorithm 3 securely predicts the class for new individual drivers. Since the algorithm runs driving style recognition for all the drivers individually, we tested its computational costs and communication overheads again for a predicting a single driver's class. A group of experiments is conducted with a varying number of aggressive paths in the decision tree: $|T| = 1, \ldots, 10$. As shown in Figure 3(a) and 3(b), the costs increase slowly as the number of aggressive paths $|T|$ increases.



(a) Computational Cost

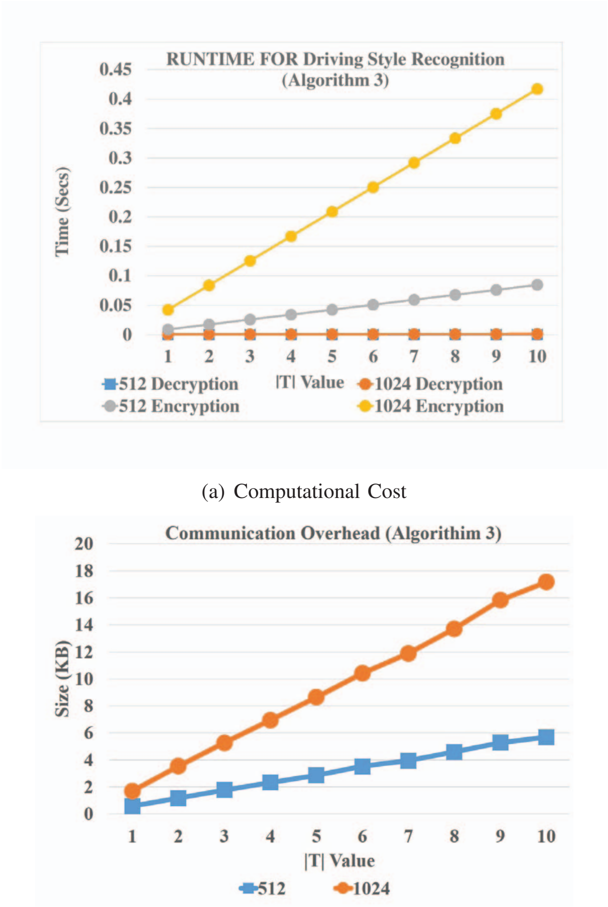
(b) Communication Overheads

Fig. 2. Privacy Preserving Driving Style Classifier Training (Algorithm 2)

V. Conclusion and Future Work

In this paper, we have developed two secure communication protocols to tackle the privacy concerns in the two phases of driving style recognition among the insurance company and various vehicles. Participating parties can jointly train a decision tree based on the vehicles' historical traveling data without compromising their privacy. The insurance company can also use its decision tree to securely predict the driving style (aggressive or defensive) of any given driver with limited disclosure. We have also experimentally validated the performance of our proposed secure communication protocols.

In the future, we have several directions to extend this work. First, we assume a semi-honest adversarial model in this paper such that every participant will follow the outlined secure communication protocols. In the real world, one or multiple parties may become more malicious to corrupt the protocol for additional payoff, or even collude with each other to breach privacy or jeopardize the utility of the protocol. We will explore efficient solutions to address the security and privacy concerns for multiparty driving style recognition in malicious adversarial model. Second, maybe more than one entities (e.g., multiple insurance companies and police department) would like to collaboratively predict the driving style of the drivers with their private inputs. Introducing more parties into this problem will influence the data partition scenario, and then the required secure communication protocols might be thoroughly different from the current ones. Third, in the future, vehicles' traveling data used for driving style recognition might be in



(a) Computational Cost

(b) Communication Overheads

Fig. 3. Privacy Preserving Driving Style Recognition (Algorithm 3)

real-time format rather than the historical aggregated format. In such scenario, the challenges on efficiency and bandwidth should be resolved.

REFERENCES

[1] R. Agrawal and R. Srikant. Privacy-preserving data mining. In *Proceedings of SIGMOD Conference*, pages 439–450, 2000.

[2] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, and T. Kohno. Comprehensive experimental analyses of automotive attack surfaces. In *Proceedings of the 20th USENIX Security Symposium*, 2011.

[3] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284, 2006.

[4] A. Elliott. The future of the connected car, February 25 2011. Mashable.

[5] B. Gedik and L. Liu. Location privacy in mobile systems: A personalized anonymization model. In *Proceedings of IEEE International Conference on Distributed Computing Systems (ICDCS)*, pages 620–629, 2005.

[6] O. Goldreich. *The Foundations of Cryptography*, volume 2, chapter General Cryptographic Protocols, pages 599–764. Cambridge University Press, Cambridge, UK, 2004.

[7] K. Han, S. Potluri, and K. Shin. On authentication in a connected vehicle: secure integration of mobile devices with vehicular networks. In *Proceedings of ACM/IEEE 4th International Conference on Cyber-Physical Systems*, pages 160–169, 2013.

[8] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J. Herrera, A. Bayen, M. Annavaram, and Q. Jacobson. Virtual trip lines for distributed privacy-preserving traffic monitoring. In *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services (MobiSys 2008)*, pages 15–28, 2008.

[9] Y. Hong. *Privacy-preserving Collaborative Optimization*. PhD thesis, Rutgers University, Newark, NJ, 2013.

[10] Y. Hong, X. He, J. Vaidya, N. R. Adam, and V. Atluri. Effective anonymization of query logs. In *CIKM*, pages 1465–1468, 2009.

[11] Y. Hong, J. Vaidya, H. Lu, and M. Wu. Differentially private search log sanitization with optimal output utility. In *EDBT*, pages 50–61, 2012.

[12] Y. Hong, S. Goel, and W. M. Liu. An efficient and privacy-preserving scheme for p2p energy exchange among smart microgrids. *International Journal of Energy Research, to Appear*.

[13] Y. Hong, J. Vaidya, and H. Lu. Secure and efficient distributed linear programming. *Journal of Computer Security*, 20(5):583–634, 2012.

[14] Y. Hong, J. Vaidya, H. Lu, P. Karras, and S. Goel. Collaborative search log sanitization: Toward differential privacy and boosted utility. *IEEE Trans. Dependable Sec. Comput.*, 12(5):504–518, 2015.

[15] Y. Hong, J. Vaidya, H. Lu, and B. Shafiq. Privacy-preserving tabu search for distributed graph coloring. In *Proceedings of SocialCom/PASSAT*, pages 951–958, 2011.

[16] 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 International Conference on Embedded Networked Sensor Systems*, pages 125–138, 2006.

[17] Y. Kishimoto and K. Oguri. A modeling method for predicting driving behavior concerning with driver's past movements. In *Proceedings of Vehicular Electronics and Safety, 2008*, pages 132–136, Sept 2008.

[18] Y. Lindell and B. Pinkas. Privacy preserving data mining. In *Proceedings of Advances in Cryptology – CRYPTO 2000*, pages 36–54, 2000, Springer-Verlag.

[19] M. V. Ly, S. Martin, and M. M. Trivedi. Driver classification and driving style recognition using inertial sensors. In *Proceedings of 2013 IEEE Intelligent Vehicles Symposium (IV)*, pages 1040–1045, 2013.

[20] T. Meek. In-car sensors put insurers in the driver's seat, June 27 2014. Forbes Business.

[21] C. Miyajima, H. Ukai, A. Naito, H. Amata, N. Kitaoka, and K. Takeda. Driver risk evaluation based on acceleration, deceleration, and steering behavior. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1829–1832, 2011.

[22] P. Paillier. Public key cryptosystems based on composite degree residuosity classes. In *Proceedings of Advances in Cryptology - Eurocrypt '99, LNCS 1592*, pages 223–238, 1999.

[23] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

[24] S. Rass, S. Fuchs, and K. Kyamakya. A game-theoretic approach to co-operative context-aware driving with partially random behavior. In *Proceedings of Smart Sensing and Context, Third European Conference*, pages 154–167, 2008.

[25] A. Sathyanarayana, P. Boyraz, Z. Purohit, R. Lubag, and J. H. L. Hansen. Driver adaptive and context aware active safety systems using can-bus signals. In *Proceedings of IEEE Intelligent Vehicles Symposium (IV)*, pages 1236–1241, 2010.

[26] L. Sweeney. k-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570, 2002.

[27] J. Vaidya, M. Kantarcioglu, and C. Clifton. Privacy preserving naive bayes classification. *International Journal on Very Large Data Bases*, 17(4):879–898, July 2008.

[28] J. Vaidya, H. Yu, and X. Jiang. Privacy preserving svm classification. *Knowledge and Information Systems*, 14(2):161–178, February 2008.

[29] M. Xiao, L. Huang, Y. Luo, and H. Shen. Privacy preserving ID3 algorithm over horizontally partitioned data. In *Proceedings of Sixth International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT 2005)*, pages 239–243, 2005.

[30] A. C. Yao. How to generate and exchange secrets. In *Proceedings of the 27th IEEE Symposium on Foundations of Computer Science*, pages 162–167, IEEE Computer Society.

[31] D. Zhang, Q. Xiao, J. Wang, and K. Li. Driver curve speed model and its application to acc speed control in curved roads. *International Journal of Automotive Technology*, 14(2):241–247, 2013.