Data Validation and Correction for Resiliency in Mobile Cyber-Physical Systems

Yasmeen Mussard-Afcari, Danda B. Rawat and Moses Garuba
Data Science and Cybersecurity Center (DSC^2)
Department of Electrical Engineering and Computer Science
Howard University, Washington, DC, 20059, USA
E-mail: yasmeen.mussardafca@bison.howard.edu, {danda.rawat, mgaruba}@howard.edu

***Abstract*—Traffic congestion and accidents are increasing exponentially worldwide. More vehicles are sold every year which leads to more traffic fatalities and congestion. There have been several efforts worldwide for mobile Cyber Physical Systems (CPS) to address a range of problems including traffic congestion, accidents, unnecessary time spent in traffic jams, and overall infotainment by using onboard communicating and computing technologies. However, when we use peer-to-peer network-based communication for mobile CPS, malicious users/vehicles could mislead the mobile CPS by not reporting their true periodic status data to their neighbors on the road. In this paper, we study a data validation and correction approach for resiliency in mobile CPS that uses a diverse set of data for reducing false information. Numerical results obtained from Monte Carlo simulation are used to evaluate the proposed approach. Results show that the proposed approach minimizes the false data in the mobile CPS to enhance the resiliency.**

***Index Terms*—Mobile Cyber Physical Systems, Resilient CPS, CPS Security.**

I. INTRODUCTION

Recent advances in computing and communication technologies and the successful deployment of wireless technologies for anytime, anywhere connectivity have led to different emerging technologies such as Cyber Physical Systems (CPS), Internet of Things (IoT) and smart cities ([1]–[3]). Mobile CPS is one of the emerging areas which is regarded as a backbone for intelligent and autonomous transportation systems. Mobile CPS are expected to provide timely feedback to the vehicles so that either the vehicle can take corrective action automatically or the driver of the vehicle can take corrective action in order to enhance road safety and overall traffic efficiency. Based on the US patent and recent study ([4], [5]), "about 60% of roadway collisions could be avoided if the operator of the vehicle was provided warning at least one-half second prior to a collision". Furthermore, timely message transmission in a vehicular network is very important for emergency vehicles and incidents ([6], [7]).

It is worth noting that in mobile CPS for transportation systems, each vehicle should be able to adapt its operating parameters on the fly based on their local observations and interactions with its neighboring vehicles. Each vehicle in vehicular ad hoc networks is required to broadcast its location, speed, and other status information to its neighbors periodically, at least 8 times a second ([1], [2], [8]). This periodic reported status information can be leveraged to provide resiliency in mobile CPS through data validation and correction, and trust estimation of interacting vehicles.

Based on the report published by the U.S. National Highway Traffic Safety Administration (NHTSA), over 30 thousand traffic fatalities occur every year on US highways due to vehicle collisions. Vehicles not only driving towards traffic congested areas but also driving with different speeds in same direction need to take action instantaneously to adapt their speeds steadily to avoid any collisions.

In this paper, we study a data validation and correction approach for mobile CPS, where messages received from neighboring vehicles are used to estimate trust levels over the measuring period and adapt the operating parameters based on trust level, as well as estimated values based on received periodic status messages.

Recent related works considered data verification using conventional integrity checking approach ([9]–[11]). Traditional integrity checking approaches cannot meet the requirement of the mobile CPS, as traditional approaches rely fully on the information received from the transmitting party and an integrity check of the received message. None of the existing approaches consider the integration of estimation based on the received messages and local observation by each vehicle. In this paper, we consider that each vehicle observes different parameters such as safety distance between vehicles using ranging/radar technology and leverages the information received from neighboring vehicles. This approach bears some similarity to proposed solutions to the problem of achieving resilience from misbehaving agents in networked control systems ([12]–[14]). First, each vehicle estimates the trust level of its reachable neighbors (that are reachable by one hop vehicular communications) over the observation period. Then, for vehicles whose trust level falls above a given threshold, different parameters are estimated and updated based on the proposed algorithm (i.e., Algorithm 1).

The remainder of this paper is organized as follows. Section II presents a typical system model used in the paper. Section III presents the approach to evaluate the trust levels of interacting vehicles and the proposed algorithm to update the speed and safety separation distance between vehicles. Numerical results for performance are presented in Section IV. Conclusions are presented in Section V.

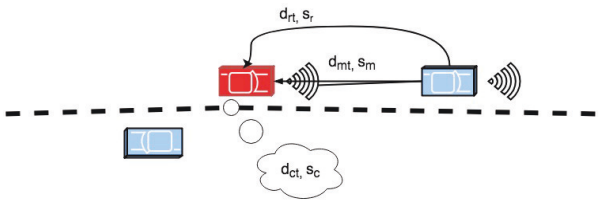

Fig. 1: A system model for mobile cyber physical systems.

II. System Model

A typical system model to study the data validation and correction for resiliency in mobile CPS is shown in Fig. 1, where each vehicle is assumed to be equipped with computing, communication, storage, camera for vision, radar technology and infrared sensors ([1], [2], [8]). Furthermore, individual vehicles are capable of measuring, transmitting, and receiving data about location, speed, and distance between vehicles. As components of a mobile CPS, the vehicles are also capable of making decisions and taking actions based on input and observed data.

From the perspective of an individual vehicle, the model contains three basic variables: d (distance to the next vehicle ahead), s (relative speed of the next vehicle ahead), and t (time).

The variables d and s are further specified in order to convey information about the source of the data, and in the case of distance, the time at which the data was recorded. Distance information measured by the vehicle's own sensors at time t is notated as d_{mt}, while distance data reported at time t by a neighboring vehicle is notated as d_{rt}. Distance information at time t that has been calculated by the vehicle using a blend of internally measured and externally reported data is notated as d_{ct}. The final assigned value for distance to the next vehicle ahead at time t, which the vehicle uses to base future decisions, on is notated d_t. Similarly, relative speed of the vehicle ahead is labeled s_m when measured by the vehicle itself, s_r when reported by the vehicle ahead, s_c when calculated by the vehicle using both internal and external data, and s when it has been finally assigned. In this model, we consider time to be a secure data point which is not vulnerable to error or attack, so a selected moment in time is notated t_i, with no reference to the source of time data. We visualize these variables for the reader in Fig. 1 in order to make clearer their relationship. The red subject vehicle on the top left drives behind the blue vehicle in front of it on the top right. Using its internal sensors, the subject vehicle measures the distance to the vehicle in front of it, d_{mt_i}, and the relative speed at which the vehicle ahead is driving, s_m. At the same time, the vehicle ahead reports to the subject vehicle its own measurements of how far away it is, d_{rt_i}, as well as its relative speed, s_r. The subject vehicle has the option of using its own measurements and those reported by the vehicle ahead in order to calculate distance d_{ct_i} and relative speed s_c, as shown in the thought bubble coming from the subject vehicle.

Furthermore, based on the interaction of a given vehicle with its neighboring vehicles, it estimates the trust levels of all interacting vehicles which is then used to take corrective actions by the vehicle.

III. Proposed Approach and The Algorithm

To estimate the trustworthiness of a given vehicle, we use an estimated suspicion level of a vehicle for a given time. A suspicion level s_n of a vehicle n among N vehicles can be expressed as

$$S_n \equiv P(C_n = \mathcal{A}|\mathcal{M}_t), \tag{1}$$

where C_n is the vehicle type: Adversarial ($\mathcal{A}$) or Honest ($\mathcal{H}$) and $\mathcal{M}_t$ is the suspicion measuring over the time t. Using Bayes' theorem, the suspicion level for a given vehicle can be written as

$$S_n(t) = \frac{P(\mathcal{M}_t|C_n = \mathcal{A})P(C_n = \mathcal{A})}{\sum_{m=1}^{N} P(\mathcal{M}_t|C_m = \mathcal{A})P(C_m = \mathcal{A})} \tag{2}$$

A given vehicle on the road could be adversarial which implies for any vehicle n and m

$$P(C_n = \mathcal{A}) = P(C_m = \mathcal{A}).$$

Then, (2) can be expressed as

$$S_n(t) = \frac{P(\mathcal{M}_t|C_n = \mathcal{A})}{\sum_{m=1}^{N} P(\mathcal{M}_t|C_m = \mathcal{A})}. \tag{3}$$

Once a given vehicle estimates the suspicion level of another vehicle using (3), we can estimate the trust level of the given vehicle as

$$T_n(t) = 1 - S_n(t). \tag{4}$$

To estimate the trust level, different data in periodic status messages can be used (e.g., [8]). For example, estimated speed of the vehicle of interest using radar technology and its reported speed through periodic status message, estimated safety separation distance using distance calculated using period message and using distance estimator such as radar technology. We propose the use of the *Algorithm* 1 to verify data about the relative speed and distance of the vehicle ahead of the subject vehicle.

This approach can be summarized as validating internally measured and externally reported data against each other, within some thresholds of trust and similarity bounded by λ_T and λ_x^T, respectively. The algorithm uses relative speed and time data to calculate distance when there is too much disagreement between internally measured and externally reported distance data but relative speed data is in agreement. It uses distance and time data to calculate relative speed when the reverse is the case. The extent to which externally reported data is taken into account in these calculations is modulated by the trust level of the reporting vehicle. When neither distance nor relative speed data from internally measured and externally reported data is in agreement, the cyber-physical system must determine the reality of the situation using other means, such as human intervention or computer vision, because the data is too compromised to draw meaning from.

Algorithm 1 Data Verification for Resiliency in Mobile CPS

Input: Periodic messages from neighboring/front vehicles, readings of local sensors and λ_T, λ_s, and λ_d threshold values

Output: Trust level, corrected safety separation distance, and steadily updated speed for the vehicle

For each vehicle n, to verify data at time t **do**

$S_n(t) \leftarrow \frac{P(\mathcal{M}_t|C_n=\mathcal{A})}{\sum_{m=1}^{N} P(\mathcal{M}_t|C_m=\mathcal{A})}$

$T_n(t) \leftarrow 1 - S_n(t)$

if $T_n(t) < \lambda_T$ **then**

Trust level is too low to take neighboring vehicle's report into account

$s \leftarrow s_m$

$d_{t_i} \leftarrow d_{mt_i}$

else

$\lambda_s^T \leftarrow \frac{\lambda_s}{T_n(t)}$

$\lambda_d^T \leftarrow \frac{\lambda_d}{T_n(t)}$

if $\frac{|s_m - s_r|}{s_m} \leq \lambda_s^T$ AND $\frac{|d_{mt_i} - d_{rt_i}|}{d_{mt_i}} \leq \lambda_d^T$ **then**

Measured data and data reported by neighboring vehicle are in close enough agreement that no alterations are needed

$s \leftarrow s_m$

$d_{t_i} \leftarrow d_{mt_i}$

else if $\frac{|s_m - s_r|}{s_m} \leq \lambda_s^T$ AND $\frac{|d_{mt_i} - d_{rt_i}|}{d_{mt_i}} > \lambda_d^T$ **then**

Measured speed data and speed data reported by neighboring vehicle are in agreement but measured and reported distance data are significantly different– speed data must be used to calculate distance

$s \leftarrow s_m$

$d_{t_i} \leftarrow d_{ct_i} = s(t_i - t_{i-1}) + d_{t_{i-1}}$

else if $\frac{|s_m - s_r|}{s_m} > \lambda_s^T$ AND $\frac{|d_{mt_i} - d_{rt_i}|}{d_{mt_i}} \leq \lambda_d^T$ **then**

Measured distance data and distance data reported by neighboring vehicle are in agreement but measured and reported distance data are significantly different– distance data must be used to calculate speed

$d_{t_i} \leftarrow d_{mt_i}$

$s \leftarrow s_c = \frac{d_{t_i} - d_{t_{i-1}}}{t_i - t_{i-1}}$

else

Data is not verifiable- trigger alternate/backup systems (human intervention, computer vision, etc.)

end if

end if

IV. Performance Evaluation

In this section, we evaluate the performance of the proposed algorithm using Monte Carlo simulation for mobile CPS. We consider that each vehicle is equipped with computing, communication, infrared, camera and radar devices. As discussed in the previous section, each vehicle in mobile CPS reports its periodic status to its neighbors and each vehicle estimates its surrounding by using its sensors and computing capabilities.

First, we randomly generated trust levels for 1000 neighboring vehicles with a uniform distribution. We also randomly generated two sets each of speed and distance data for 1000 vehicles with a normal distribution centered around means of 37.5 mph and 125 meters, respectively. One of these pairs of generated speed and distance data represented measured data, while the other represented reported data.

We then simulated running our algorithm in order to plot the frequency of data agreement type for each of the 1000 driving scenarios represented by a generated set of five data points: trust level, measured speed, measured distance, reported speed, and reported data. We set λ_s and λ_d, the thresholds which allow us to scale trust level to maximum percent difference, to 0.1 on each of the twenty runs of the simulation. The minimum trust level λ_T began at 0 on the first run and increased by 0.05 on each simulation run, ending at 1. We now present a representative selection of simulation results.

When minimum trust is at the low end of the possible range, such as 0.15 as shown in Fig. 2, the scenarios are divided between having trust levels that are too low to take reported data into account and having all data agree. The percentage of scenarios which have trust levels which are too low to take reported data into account ("no trust") is approximately the minimum trust level, 15.8% compared to a minimum trust level of 0.15 in Fig. 2. This makes sense as our generated trust levels were uniformly distributed across scenarios, however, it is noteworthy that every scenario in which the reporting vehicle was trusted had close enough data agreement across both speed and distance data.

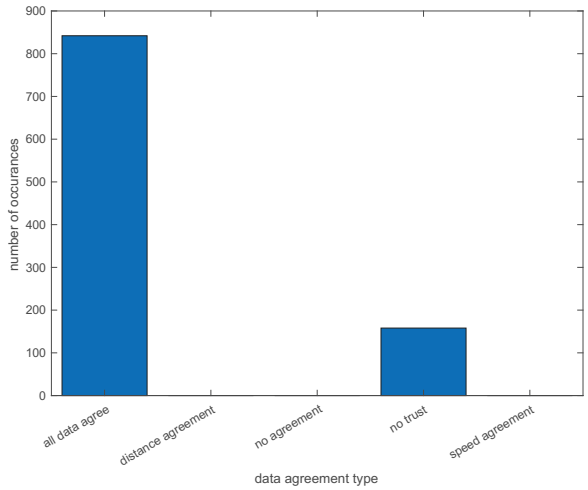

Fig. 2: Frequency of data agreement type when minimum trust level is 0.15.

When minimum trust is in the middle of the possible range, such as 0.5 as shown in Fig. 3, we begin to see a small number of scenarios (0.3% and 0.2%, respectively, when trust level is 0.5) which only agree on one type of data, speed or distance, rather than both. As to be expected, the scenarios which have only one type of agreement seem to only reduce the number of scenarios in which all data agree, rather than those in which there is no trust.

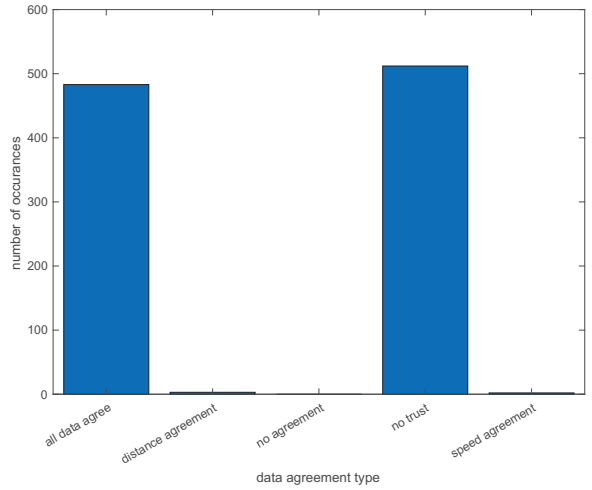

Fig. 3: Frequency of data agreement type when minimum trust level is 0.5.

Finally, as minimum trust level starts to climb into the upper end of the possible range, such as trust level 0.9 as shown in Fig. 4, we begin to see the balance of the small number of scenarios which have only distance or speed agreement shift towards distance agreement. In Fig. 4, where trust level is 0.9, 0.6% of all scenarios had only distance agreement, while only 0.2% off all scenarios had only speed agreement. It is also notable that there is no point at which there is no agreement at all between data types in any of the simulations, suggestion that it would be very rare to need to trigger alternate/backup systems. Further research will be needed to determine both why it is extremely unlikely to achieve one type of data agreement without the other, but also why the balance of this small number of one data agreement type scenarios shifts towards distance agreement as minimum trust level increases.

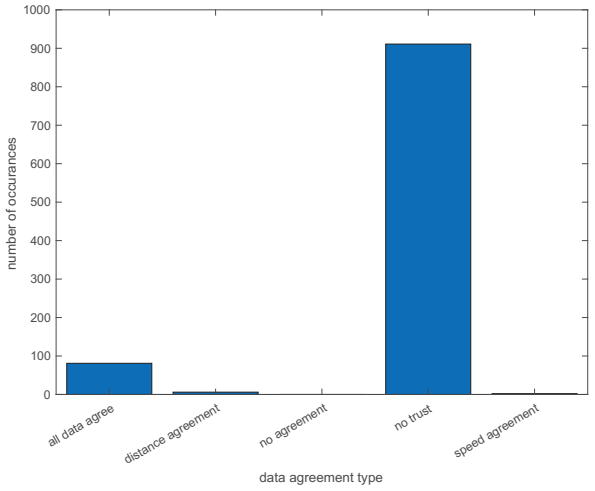

Fig. 4: Frequency of data agreement type when minimum trust level is 0.9.

V. Conclusion

In this paper, we have presented an algorithm for calculating the trust level of surrounding vehicles as well as data validation and correction in mobile CPS. The goal of the proposed approach is to correct the speed and safety separation distance of vehicles on the road on the fly based on the observations of the following vehicle and data reported by vehicles traveling in the front of the given vehicle in a mobile CPS. We have used simulation results to evaluate the proposed approach. The results have shown that the proposed approach can correct the speed and safety separation distance (in case of untrustworthy vehicle) for reducing vehicle collisions in the mobile CPS.

Our future work includes an extensive study using in depth formal analysis, model based data validation and correction, consideration of the impact of errors in on-board sensors, and validation of the proposed approach using real test data.

Acknowledgment

This work is supported by the U.S. National Science Foundation (NSF) under grants CNS 1650831 and HRD 1828811. However, any opinion, finding, and conclusions or recommendations expressed in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the NSF.

References

[1] M. C. Weigle and S. Olariu, *Vehicular networks: from theory to practice*. Chapman and Hall/CRC, 2009.

[2] D. B. Rawat, J. J. Rodrigues, and I. Stojmenovic, *Cyber-Physical Systems: From Theory to Practice*. CRC Press, 2015.

[3] D. B. Rawat and K. Z. Ghafoor, *Smart Cities Cybersecurity and Privacy*. Elsevier Press, 2018.

[4] C. D. Wang and J. P. Thompson, "Apparatus and method for motion detection and tracking of objects in a region for collision avoidance utilizing a real-time adaptive probabilistic neural network," Mar. 18 1997, uS Patent 5,613,039.

[5] D. S. Breed, W. E. Duvall, and W. C. Johnson, "Accident avoidance system," Apr. 9 2002, uS Patent 6,370,475.

[6] D. B. Rawat, D. C. Popescu, G. Yan, and S. Olariu, "Enhancing VANET Performance by Joint Adaptation of Transmission Power and Contention Window Size," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 9, pp. 1528–1535, Sept 2011.

[7] H. Hartenstein and K. Laberteaux, *VANET: vehicular applications and inter-networking technologies*. Wiley Online Library, 2010, vol. 1.

[8] D. B. Rawat and C. Bajracharya, *Vehicular Cyber Physical Systems: Adaptive Connectivity and Security*. Springer, 2016.

[9] E. A. Lee, "Cyber physical systems: Design challenges," in *11th IEEE Symposium on Object Oriented Real-Time Distributed Computing (ISORC)*, 2008, pp. 363–369.

[10] L. Deka and M. Chowdhury, *Transportation Cyber-Physical Systems*. Elsevier, 2018.

[11] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[12] W. Zeng and M. Chow, "Resilient distributed control in the presence of misbehaving agents in networked control systems," *IEEE Transactions on Cybernetics*, vol. 44, no. 11, pp. 2038–2049, Nov 2014.

[13] A. Khazraei, H. Kebriaei, and F. R. Salmasi, "Replay attack detection in a multi agent system using stability analysis and loss effective watermarking," in *2017 American Control Conference (ACC)*, May 2017, pp. 4778–4783.

[14] D. B. Rawat, T. White, M. S. Parwez, C. Bajracharya, and M. Song, "Evaluating secrecy outage of physical layer security in large-scale MIMO wireless communications for cyber-physical systems," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1987–1993, 2017.