CERIAS

The Center for Education and Research in Information Assurance and Security

# Impact of Cyber Attacks on Traffic State Estimation for Connected and Autonomous Vehicles (CAVs) Systems


Eunhan Ka[1], Satish V. Ukkusuri[1]

[1] Lyles School of Civil Engineering, Purdue University

Email: kae@purdue.edu, sukkusur@purdue.edu

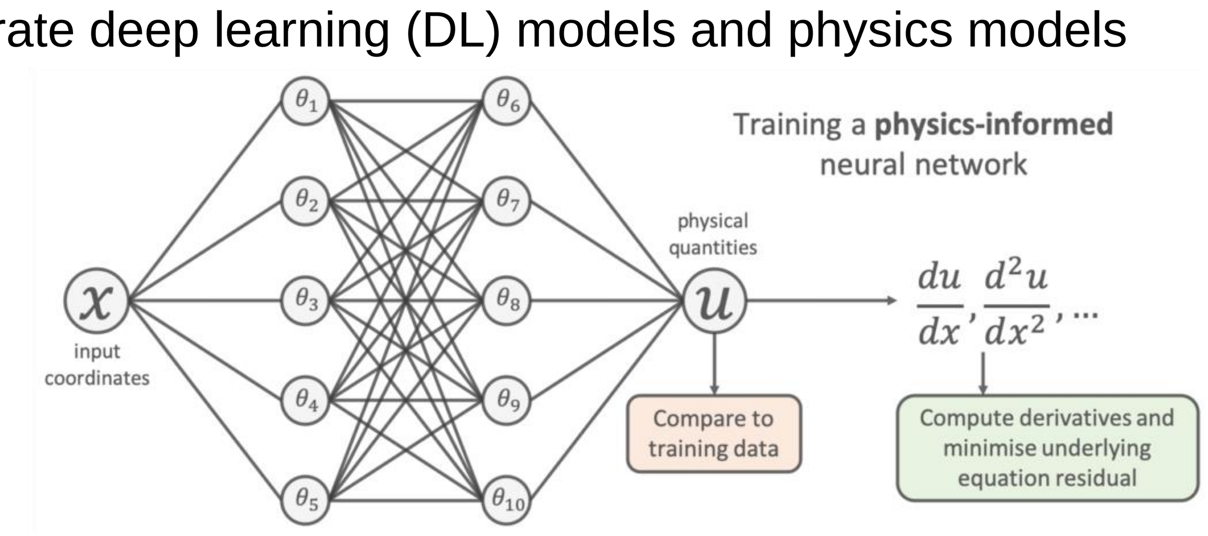

## INTRODUCTION

### Network Traffic Dynamics

- Model traffic state dynamics over time in road networks
- **Input:** Inflow rate, Trip length distribution, Initial average speed ($v$) and traffic density ($k$) equation ($v = f(k)$)
- **Output:** Traffic state variables (e.g., speed ($v$), density ($v$), flow rate ($q$))
- **Limitations: 1) Hard efforts in parameter calibrations; 2) Discretized solution algorithms**

### Physics-Informed Deep Learning (PIDL)

- Integrate deep learning (DL) models and physics models




Source: https://benmoseley.blog/my-research/so-what-is-a-physics-informed-neural-network/

- **Advantages: 1) No parameter calibrations; 2) Continuous solution algorithms**

### Cyber Attacks on Connected and Autonomous Vehicles

- Connected and Autonomous Vehicles (CAVs): Alleviate traffic congestion, enhance transportation efficiency, and reduce accidents
- Examples of cyber attacks on vehicles: 1) Remote hacking in 2015 (Chrysler) 2) Attacks on controller area network bus in 2019 (BMW)
- Potential attacks: 1) Infrastructure attacks (e.g., data theft, data poisoning); **2) Attacks on machine learning systems (e.g., data poisoning, escape attacks)**

### Research Questions

- How much do **Attacks on machine learning systems** on the PIDL model's input affect traffic state estimation?

### Objectives

- **Develop the framework for assessing the impacts of cyber attacks with PIDL models**
- Quantify the impacts of cyber attacks on traffic state estimation with PIDL models

## PRELIMINARIES

### Generalized Bathtub Model (GBM)'s Conservation Laws

**(1st Law: Conservation of trip-miles)**

$$\lambda(0)B(0) + \int_0^t f(s)\,\tilde{B}(s)ds - \int_0^t \lambda(s)\,v(s)ds = \lambda(t)B(t)$$

Initial entering trip-miles; Added trip-miles until time t; Processed trip-miles until time t; Remaining trip-miles

**(2nd Law: Conservation of total trips)**

$$G(t) = \lambda(0) + F(t) - \lambda(t)$$

Cumulative out-flux; Initial entering vehicles; Cumulative in-flux; Number of active vehicles

**(3rd Law: Conservation of the number of trips with remaining distances)**

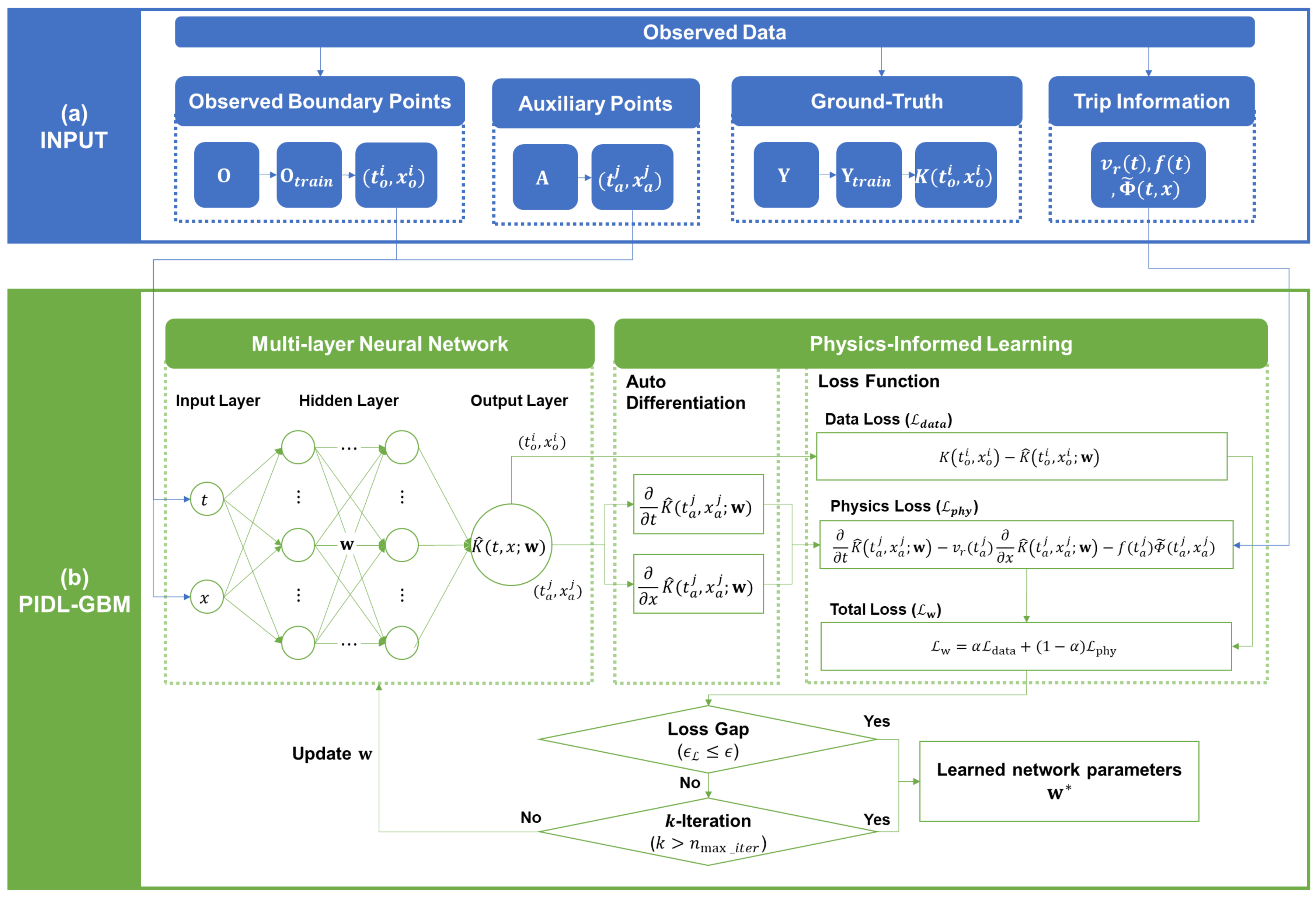
$$\frac{\partial}{\partial t}K(t,x) - v(t)\frac{\partial}{\partial x}K(t,x) = f(t)\tilde{\Phi}(t,x)$$

# of trips with a remaining distance not smaller than x; # of trips with a remaining distance not smaller than x+v(t)dt; # of entering trips with a remaining distance not smaller than x

## METHODOLOGY

### Framework of PIDL-GBM

- **Input:** Observation ($O$), $(\boldsymbol{t_o}, \boldsymbol{x_o})$; Auxiliary points ($A$), $(\boldsymbol{t_a}, \boldsymbol{x_a})$; Ground-Truth ($Y$), $\boldsymbol{K}(\boldsymbol{t_o}, \boldsymbol{x_o})$; Trip information $(v_r(t), f(t), \tilde{\varphi}(t,x))$
- PIDL-GBM: Multi-layer Neural Network, Auto Differentiation, Loss Function
- **Output:** Learned network weights ($\mathbf{w}^*$) ➔ Estimation of $K(t,x)$; $\boldsymbol{\hat{K}(t,x)}$




### Escape attacks

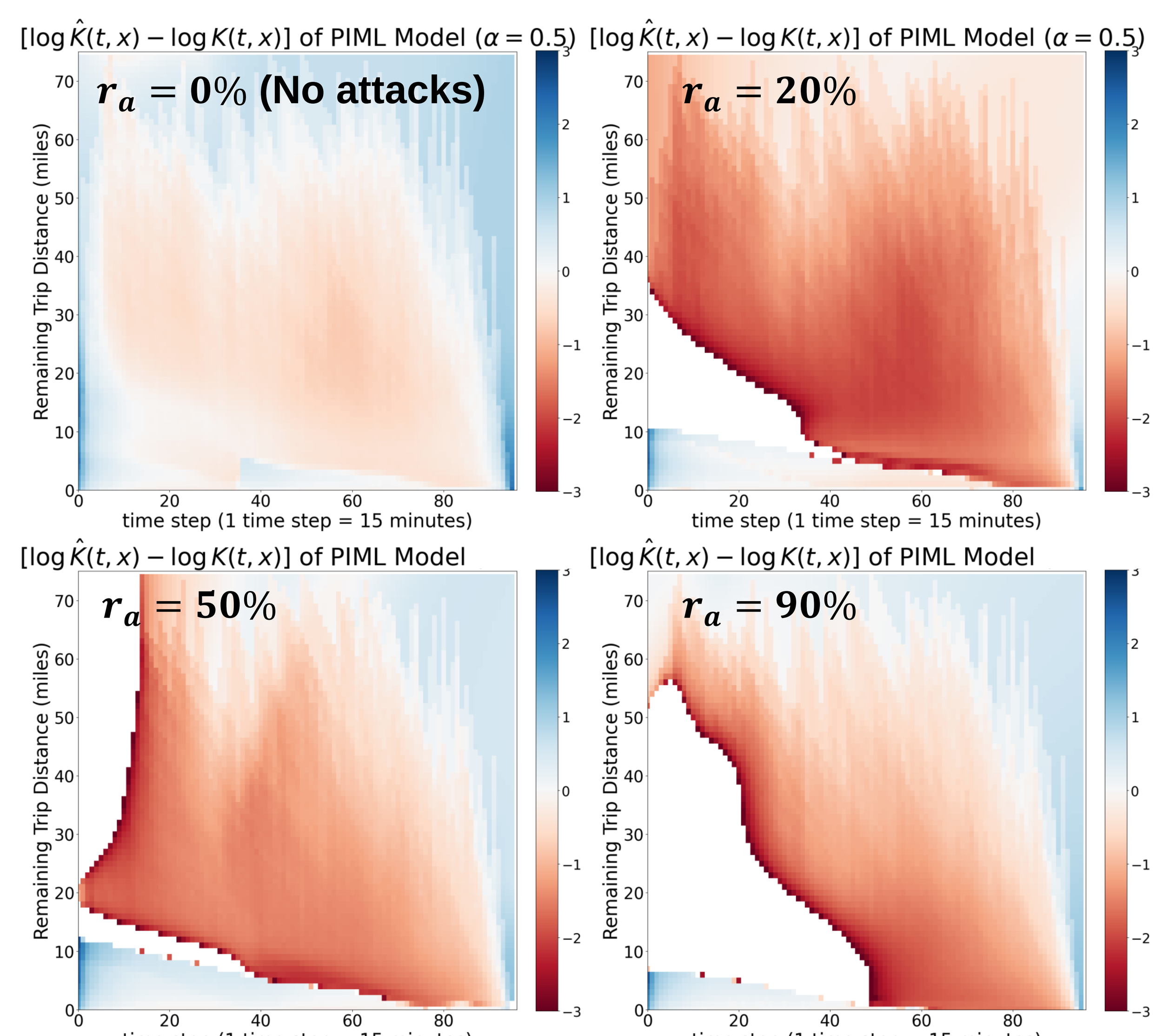
- Assume that escape attacks randomly remove input data in PIDL-GBM
- Escape attacks hinder traffic state estimation by manipulating input data

## EXPERIMENTS

- Study Area: Indianapolis road network (35,742 nodes and 49,455 links)
- Data Collection: Mobile data (14.4 M unique devices and 4.8 B records)
- Ratio of attacks ($r_a$) = {0, 10, 20, …, 90%}; Performance Metrics: RMSE

| $r_a$ | 0 % | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
|---|---|---|---|---|---|---|---|---|---|---|
| RMSE | 0.0603 | 0.0647 | 0.0853 | 0.0859 | 0.0860 | 0.1074 | 0.0977 | 0.0967 | 0.1035 | 0.1211 |

(No attack)




**Red: Underestimation**
**Blue: Overestimation**

PURDUE UNIVERSITY

CERIAS