Enhancing Transportation Cybersecurity: AI, Data Science, and Addressing Vulnerabilities

@ITS Texas/TexITE, Houston, 2024

Arlei Silva

Rice University

Intelligent transportation systems

Technologies:

- Navigation
- Traffic signals
- Sensors and cameras
- Incident detection
- Ride-sharing
- Forecasting
- Predictive maintenance
- Traffic control
- V2V communication
- Autonomous vehicles

Intelligent transportation systems

Technologies:

- Navigation
- Traffic signals
- Sensors and cameras
- Incident detection
- Ride-sharing
- Forecasting
- Predictive maintenance
- Traffic control
- V2V communication
- Autonomous vehicles

Benefits:

- Increased safety
- Reduced congestion
- Reduced emissions
- Reduced inequality
- Reduced maintenance cost
- Better accessibility
- Faster public transit
- Faster emergency response
- Faster deliveries
- Faster evacuation

What can go wrong?

CYBERSECURITY

St. Louis, Mo., Transit Disrupted by Weekend Cyber Attack

Trucking Grapples With Evolving Cybersecurity Threats

Technology Also Provides Opportunity for More Criminal Activity

US Transportation Department Discloses Data Breach

- No transportation safety systems affected, department says
- Cybersecurity agency joins effort to secure software systems

Road work sign hacked on I-65 near Clanton, expert discusses best cybersecurity practices

Autos & Transportation | White Collar Crime | Data Privacy

Uber investigating 'cybersecurity incident' after report of breach

Tesla hacker discovers secret 'Elon Mode' for hands-free Full Self-Driving

US DOT UTC CYBERCARE

Transportation Cybersecurity Center for Advanced Research and Education

- University of Houston (Host)
- Embry-Riddle Aeronautical University
- Rice University
- Texas A&M Corpus Christi (MSI)
- University of Cincinnati
- University of Hawai'i at Manoa

Rice's expertise: AI and ML

2023-2029, $10M (Rice: $1.5M)

$1T Infrastructure Investment and Jobs Act

CYBERCARE: research thrusts

CAV cybersecurity

E.g. how can cyberattacks be detected/avoided?

Transportation data security

E.g. which identity and privacy metadata will be shared with unknown parties in a cybersecurity incident?

ATMS cybersecurity

E.g. how to protect multiple attack surfaces in a decentralized environment?

Next generation transportation cybersecurity

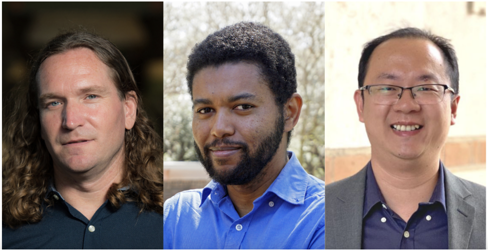
E.g. how to evaluate the resilience of transportation systems against cybersecurity attacks on one or more sub-systems?

Rice team (Chris Jermaine, Arlei Silva, Xia Hu)

RICE ENGINEERING AND COMPUTING
Department of Computer Science

Rice CS joins UH in research to improve transportation cyber security

Silva, Hu and Jermaine focus ML, AI, and graphs on DOT infrastructure



Rice team's expertise and leadership

Chris Jermaine:

- J.S. Abercrombie Professor of Eng., Professor and Chair of CS
- Research: large-scale, computationally intensive data processing, with a focus on systems for ML and AI

Arlei Silva (PI, associate director):

- Assistant Professor of Computer Science
- Research: algorithms and models for mining and learning from complex datasets, especially for graphs/networks

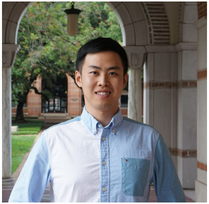
Xia Hu:

- Associate Professor of Computer Science
- Research: automated and interpretable ML algorithms and systems for large-scale, networked, dynamic, and sparse data

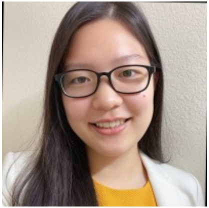
Students

Delaram Pirhayatifard

Joao Mattos



Ruixiang Tang



Xinyu Yao

Yu-Neng Chuang

Zhimin Ding

Research in Y1

CYBER-CARE has enabled the Rice team to pursue both foundational and applied research on AI and ML with either great potential or promising results towards safeguarding transportation systems against cyber-attacks

Key research thrusts:

1. Scalable, interpretable, and flexible AI and ML
2. Intrusion and misinformation detection in transportation

Eight papers published:

- Venues: ICML (3), EMNLP (2), NeurIPS, TMLR, EDS
- Multiple papers under review
- Topics: Large Language Models and Graph Neural Networks

Prompt Tuning Strikes Back: Customizing Foundation Models with Low-Rank Prompt Adaptation (NeurIPS'24, Jermaine)

Large Language Models (LLMs):

- Generative models for text able to answer complex queries
- Learn statistical relationships from large textual databases

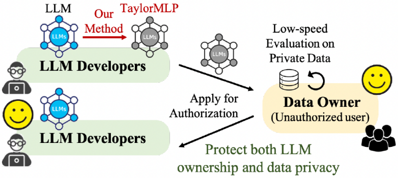
How to fine-tune an LLM without direct access to it?

LOPA: Low-Rank Prompt Adaptation

- Soft prompts appended to the input query
- Combines task-specific and instance-specific information

Relevance to CYBER-CARE: transportation-specific queries can be answered by fine-tuning a general purpose LLM

Taylor Unswift: Secured Weight Release for Large Language Models via Taylor Expansion (EMNLP'24, Hu)



How to release LLMs without compromising the privacy/security of the model?

TaylorMLP: shares a Taylor approximation of the model

- Model can be used but cannot be reconstructed
- Similar accuracy compared with controllable latency

Relevance to CYBER-CARE: LLMs for transportation can be shared with partners without the risk of misuse

Cross-Domain Graph Anomaly Detection via Test-time Training (under review, Silva)

Graph Anomaly Detection (GAD)

- Identifying unusual patterns in graph data
- Often require labeled anomalies for training the model

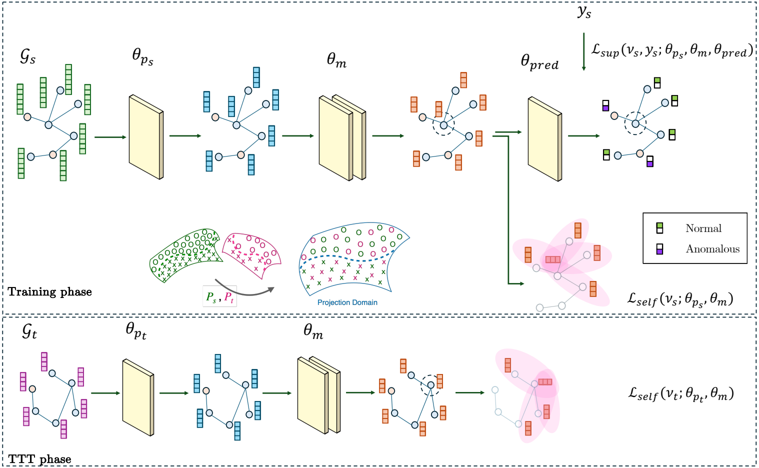
How to leverage labeled anomalies across different domains?

From computer network (source) to IoT (target)

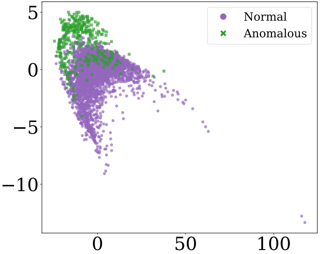
GADT3: domain adaptation for GAD

- Source and target-specific data encoders
- Test-time training scheme

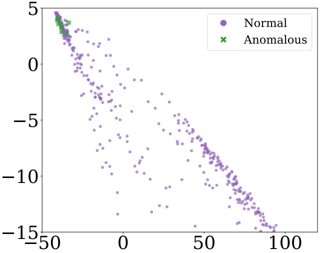
Cross-Domain Graph Anomaly Detection via Test-time Training



Cross-Domain Graph Anomaly Detection via Test-time Training



Source



Target

Using source model to identify anomalies in the target dataset.

Relevance to CYBER-CARE: intrusion detection for cybersecurity in transportation

Education in Y1

PhD students trained on topics relevant to CYBER-CARE

- Ruixiang Tang is now an assistant professor of CS@Rutgers

Data to Knowledge project on traffic misinformation

- Semester-long project involving six students
- Data: Twitter, Caltrans sensors and incident reports
- Tools: data science, LLMs, geoprocessing

Sanjay Rajasekha, Yifan Wu, Frank Ran,
Bryant Cassady, Ningzhi Xu, Anthony Yan

Challenges

Lack of realistic large-scale labeled benchmarks

- Synthetic intrusion detection benchmarks are too easy for ML
- Evaluation often does not reflect real-world settings

Tradeoff: identifying new attacks vs. false positives

- Identifying vulnerabilities in real systems is expensive
- Lack of realistic testbeds, honeypots, etc.

Research scattered across communities

AI/ML, cybersecurity, transportation

Is the challenge "between the computer and the chair"?

Enhancing Transportation Cybersecurity: AI, Data Science, and Addressing Vulnerabilities

@ITS Texas/TexITE, Houston, 2024

Arlei Silva

Rice University