

JUNO2: Collaborative Research: STEAM: Secure and Trustworthy Framework for Integrated Energy and Mobility in Smart Connected Communities (US-Japan Collaboration)



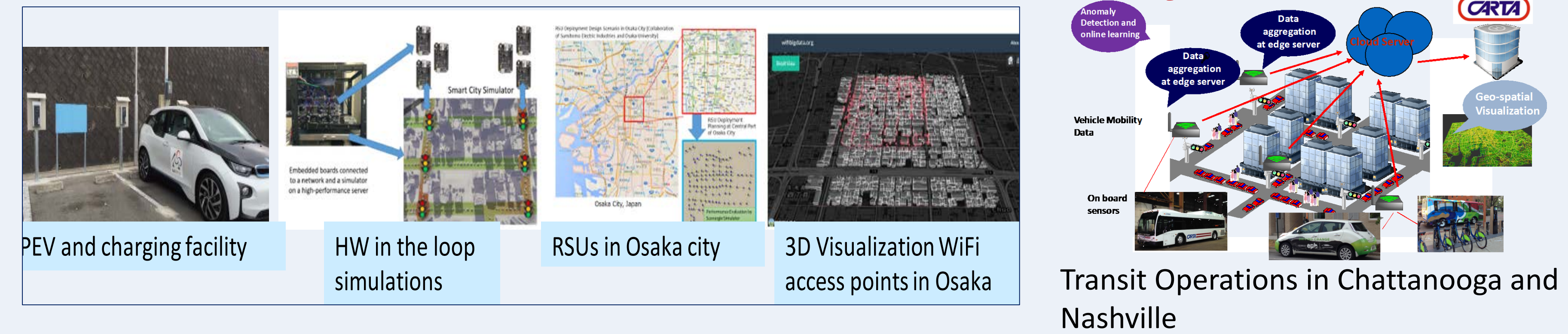
Investigators: Sajal K. Das (Missouri S&T); Abhishek Dubey (Vanderbilt University); Shameek Bhattacharjee (WMich), Hayato Yamana (Waseda Univ); Keiichi Yasumoto, (NAIST); Hirozumi Yamaguchi (Osaka University)

Students: Michael Wilbur (Vanderbilt University), Geoffrey Pettet (Vanderbilt University), Venkat Praveen Madhavarapu (MST), Prithwiraj Roy (MST), Yu Ishimaki (Waseda Univ), Jose Talusan (NAIST)

STEAM Project: Objective

- Develop integrated frameworks, algorithms and models to address security and trustworthiness challenges in mobility and energy under various threats.
- Design lightweight resilient anomaly detection and privacy preserving encryption schemes; a middleware architecture.
- Mechanisms to handle conflicting goals of identifying anomalies; preserving privacy and integrity at scale.
- Efficient co-design and calibration of encryption and robust anomaly detection schemes.

Validation with Real Datasets and Systems



Efficient Anomaly Detection

Threat Model: Orchestrated attacks on a collection of sensors to maximize the effect of attack on the global transportation system

Tried on real data from Nashville, TN

Optimal RSU Placement: designed for optimal anomaly detection using ROC curves

Macro Model: designed for large scale decentralized anomaly detection in real time

Micro Model: highly accurate and fine grained-anomaly detection. Computationally intensive

Congestion Progression: how the effects of anomalies propagate through the network

Optimized (for detection) RSU Deployment

Macro-Level Detection: Built efficient streaming statistical algorithm

Micro-Level Detection: Efficient long short-term memory (LSTM) based traffic predictor by modeling each road segment in large scale traffic network as a function of neighboring roads.

Anomaly Detection over Encrypted Data

Adopted Approximate Homomorphic Encryption scheme (HEAAN) to leverage floating-point arithmetic (log computation) over encrypted data

Anomaly detection algorithm over encrypted data w/o using non-colluding servers (more secure than Approach 1)

Pre-computations of logarithm and its inverse at each household

Optimized for FHE-friendly anomaly detection

Homomorphic evaluation up to daily statistics that hide individual power consumption

Almost same accuracy over both encrypted and unencrypted methods (possible to mitigate small accuracy error via post-processing)

Server-side computation is feasible: 3.303 s/hour (each hourly time-slot)

Next Challenges

- Enhance table lookup method to adopt multiple values, and propose less-than comparison for input values to handle wide range of inputs
- Balance Privacy-Performance trade-off with quantification (e.g., FHE with Differential Privacy)

Utility Company (Data Analyst)

Perform anomaly detection using the statistics

Daily Statistics

Hourly Data

Pre-computation (Inverse of Log) at each household

Normal Range

Unencrypted

Encrypted

Ciphertext Size:

Household → Server: 2,270KB/hour

Server → Utility: 224KB/date

All experiments are with 136 households. Experiment is done with Intel Xeon CPU E5-1620 v4 @ 3.50GHz in single-threaded mode.

Distributed Processing and Data Store

Challenges

- How to build multi-domain architecture for smart mobility and smart energy?
- How and where to implement computations related to privacy, security, and trust?
- What are computational/resource challenges for scalability?

Distributed Data Store

Designed a middleware architecture to assign tasks over IoT devices (e.g., RSUs, smart meters) taking into account required QoS level

Implemented/evaluated a prototype middleware

Used Docker technology for easy development

Developed Smart Transportation Service Emulation Testbed

Flow Consumers: Application/Actuators

Service Developers: Datasets

Recipe Program for Context Flow Generation

IFoT

Flow Provider: Sensors

Computational Resource Provider: Edge IoT devices, Cloud

Transaction Log [Apache Pulsar]

Spatial representation of data using a graph structure.

Material View [Not Only SQL]

Analysis

Goal: All buses in a region, Identify bottlenecks in this region

Data Sources [Various Forms]: Vehicle (bus), Traffic, Real-time Bus Position, Automated Passenger Count, Weather

Data Sources Needed: bus locations, bus mode, bus speed, surrounding traffic, road network infrastructure

Queries Needed: buses within path, aggregation across time, road network shape along route, traffic per road segment, bus route ID, bus speed, weather

Optimal Data Structures: spatial indexed data store, graph data store, spatial indexed data store

The architecture can support different kinds of spatio-temporal queries

Robust Decisions Under Uncertainty

Problem and Challenges: Improve publish decision accuracy in vehicular social sensing under attacks and observation uncertainty

Two Level Decision Tree Formulation:

- Which event type?
- What event confidence?

Key Theory for Tree Design:

- Modified Prospect Theory (CPT)
- Tversky Kahneman Function
- Dual Prob. Weighing Function

Compare with classical decision tree with expected utility maximization

Classification (Detection)

util(P) = g1 * v(Cj) * π+(pj) + l1 * v(Cj) * π-(pj)

Modified Prospect Theory Decision Scheme

Modified Dual Prob. Weighing Function

Modified Tversky-Kahneman Utility Function

Classification (Detection)

util(P) = g1 * v(Cj) * π+(pj) + l1 * v(Cj) * π-(pj)

Classification (Detection)

util(P) = g1 * v(Cj) * π+(pj) + l1 * v(Cj) * π-(pj)

Award #s: CNS-1818942 (Missouri Univ. of Science and Technology); and CNS-1818901 (Vanderbilt University)