CAREER: Towards Secured and Efficient Energy-based Critical Infrastructure

A Gaussian-Mixture Model Based Detection against Data Integrity Attacks in Smart Grid

Dr. Wei Yu Associate Professor

Director of Cyber-Physical Networked System and Security Research Laboratory

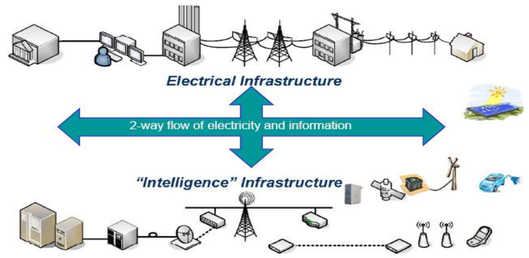
Department of Computer and Information Sciences, Towson University

Web: http://wp.towson.edu/wyu Email: wyu@towson.edu

Acknowledgement: NSF CAREER Award - CNS-1350145

TOWSON UNIVERSITY™

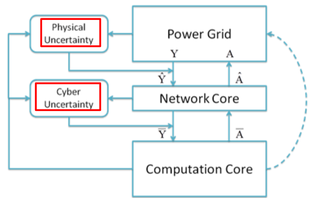
Overview



Smart Grid (Source: NIST)

Project Goals and Research Focus

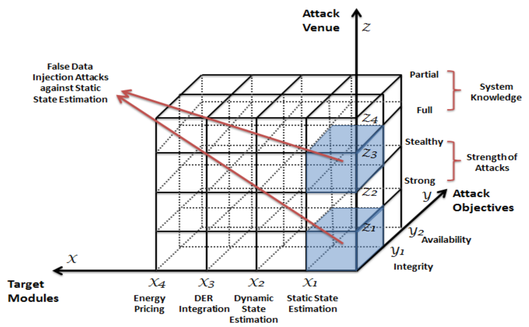
- Goals
  - Establish a theoretical and empirical basis for securing energy-based infrastructure
- Our research focuses on
  - Developing modeling and co-simulation frameworks for designing efficient energy CPS
  - Conduct a systematical study of exploring attack space and designing countermeasures



Uncertainty in Coordinating Cyber & Physical Components

Cyber Attacks on Power Grids

- Real-world Examples
  - In 2003, computers infected by Slammer worm shut down safety display systems at power plant in Ohio
  - In 2010, Stuxnet worm provides a blueprint for aggressive attacks on control systems
  - In 2011, malware BlackEnergy disrupts processes controlled HMIs products from vendors, e.g., General Electric, Siemens, Advantech
  - In 2013 and 2014, there were 224 hacking incidents at energy companies investigated by the Computer Emergency Readiness Team, a division of the Department of Homeland Security (DHS)
  - Between April 2013 and 2014, hackers managed to break into 37% of energy companies, according to a survey by ThreatTrack Security
  - …



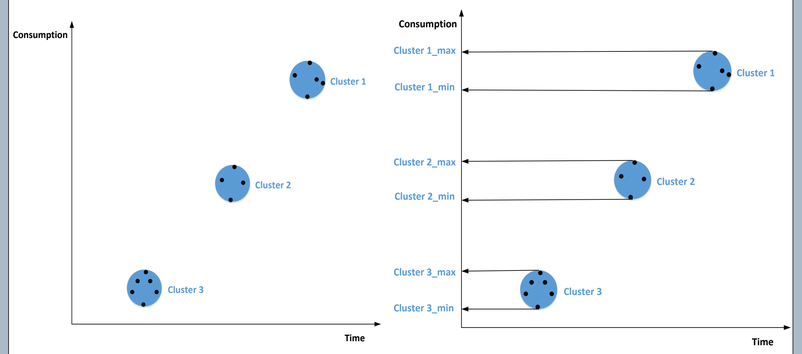
A Framework to Explore Attacks against Smart Grid Operations

Our Contributions

- Propose a new Gaussian-Mixture Model-based Detection (GMMD) scheme against data integrity attacks
  - Our scheme does not need to predefine a threshold
  - Our scheme does not need to obtain external knowledge from domain experts to label historical data
  - Our scheme could narrow the range of normal data through clustering the historical data set
- Carry out the theoretical analysis, demonstrating that our scheme can achieve a high detection accuracy
- Conduct extensive simulation to validate our findings

Workflow

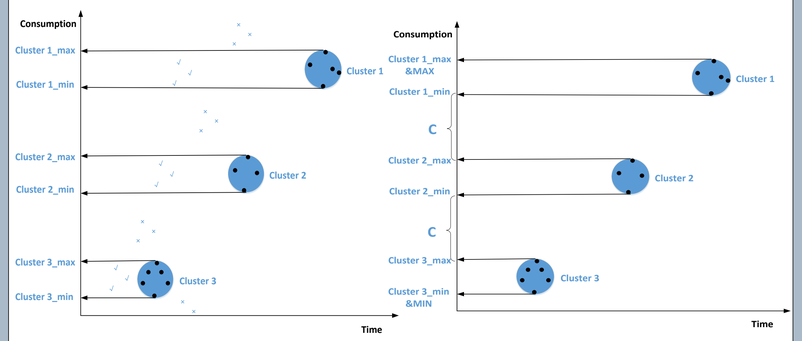
- Feature Extraction: customer ID, day, time slot, consumption data
- Historical Data Clustering & Normal Range Determination



- False Data Decision
  - If the measurement value is in the range of S, it is true

$$clu_k\_\min \le c_i^t \le clu_k\_\max, 1 \le k \le K$$

  - Else, it is false



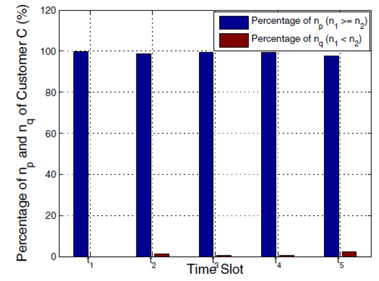
Evaluation

P_d The ratio of the number of false data of measurement values being detection versus the total number of false data

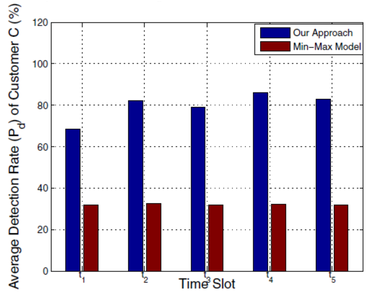
P_e The ratio of the number of true data of measurement values being determined as false and the number of false data of measurement values being determined as true versus the total number of data

Training data (80 days) & Testing data (40 days)

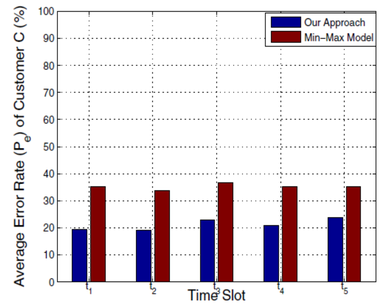
$$t_1 = 7:45, t_2 = 10:45, t_3 = 12:45, t_4 = 14:45, t_5 = 17:45$$



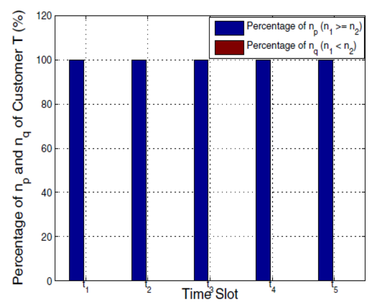
Ratio of n_p and n_q of Customer C



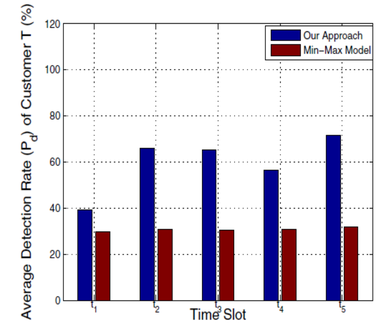
Detection Rate of Customer C



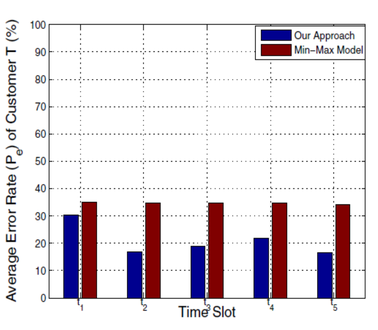
Error Rate of Customer C



Ratio of n_p and n_q of Customer T



Detection Rate of Customer T



Error Rate of Customer C