SAND2024-05444C

Sandia National Laboratories

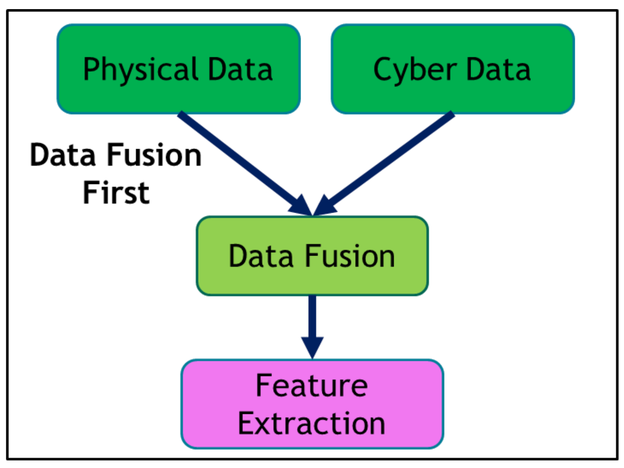
A Comparison Study of Feature Extraction and Data Fusion Techniques for Improving Cyber-Physical Situational Awareness

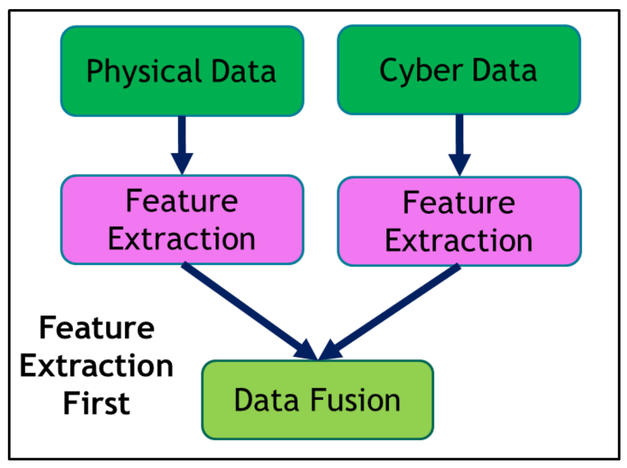
Logan Blakely, Georgios Fragkos, Shamina Hossain-McKenzie, Christopher Goes, Adam Summers
Sandia National Laboratories

Khandaker Akramul Haque, Katherine Davis
Texas A&M University

Background:

- Modern power grids are tightly integrated with communications systems and cyber networks, and this integration continues to increase as more communication-enabled devices are added
- It is no longer sufficient to independently collect and analyze physical systems data and cyber data
- Detecting modern cyber-physical attacks on the power grid requires an integrated analysis of both cyber and physical data sources





Collected Data:

- Cyber Data collected from 3 substation locations:
  - Round-trip times (RTT)
  - Packet retransmits
  - Cyber data collected at 1 sec intervals and upsampled to 33 samples per second
- Physical Data collected from 8 PMU sensors:
  - Voltage magnitude and angle
  - Current
  - Frequency
  - Physical data sampled at 33 samples per second

Scenario:

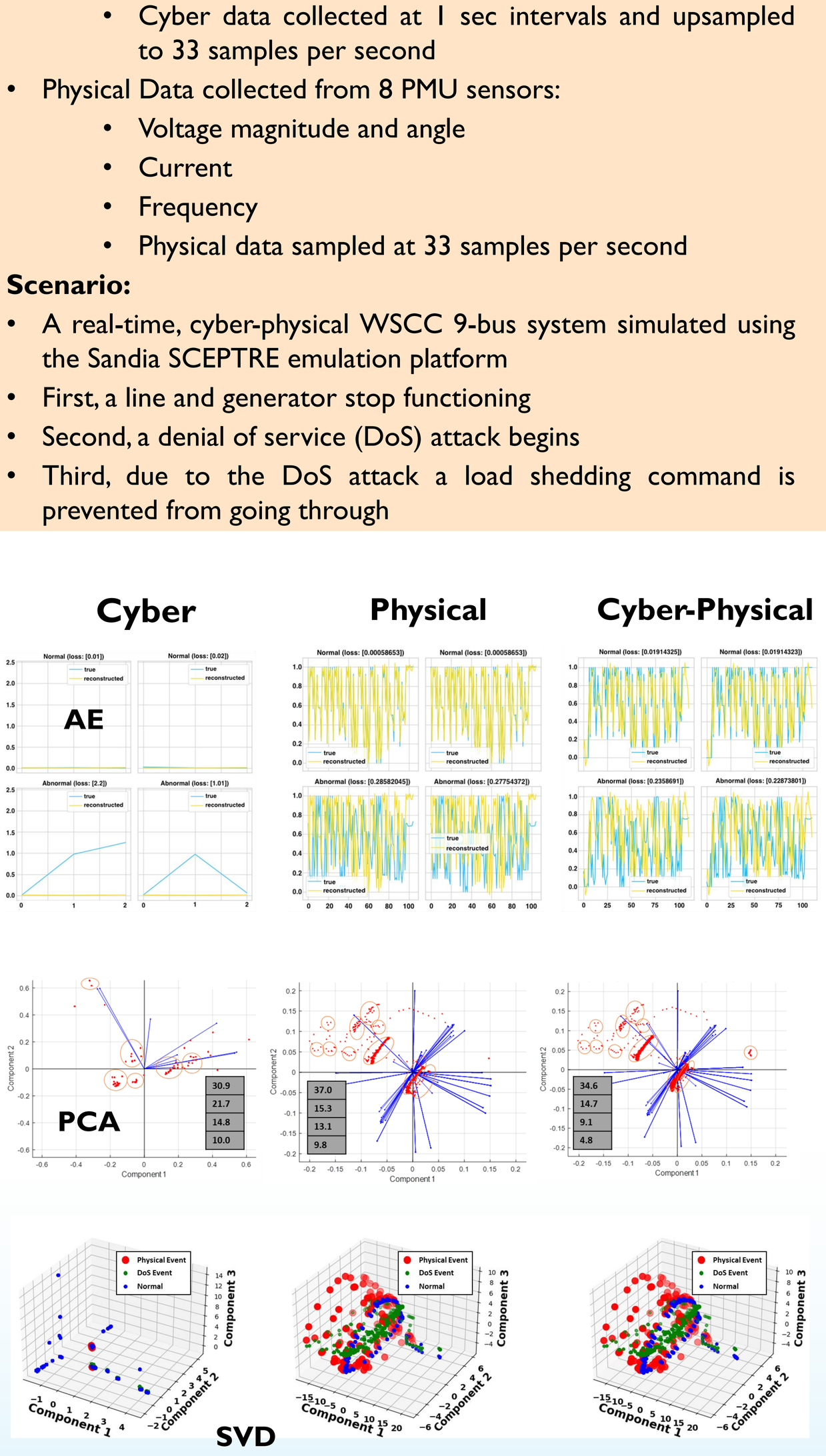
- A real-time, cyber-physical WSCC 9-bus system simulated using the Sandia SCEPTRE emulation platform
- First, a line and generator stop functioning
- Second, a denial of service (DoS) attack begins
- Third, due to the DoS attack a load shedding command is prevented from going through

Methodology:

- Tested three feature extraction methodologies:
  - Principal Component Analysis (PCA) with Singular Value Decomposition (SVD)
  - Stochastic Neighbor Embedding (t-SNE)
  - Autoencoder (AE)
- Each method was run on the cyber only, physical only, and combined cyber/physical datasets

Key Takeaways:

- Each feature extraction technique is a feasible choice for cyber-physical system analysis
- AE provides a fully unsupervised feature extraction plus anomaly detection solution
- The result plots for the physical and cyber-physical cases are nearly identical. We hypothesize that this is due to the larger number of physical features as well as the upsampling of the cyber data



This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan https://www.energy.gov/downloads/doe-public-access-plan.

Sandia National Laboratories