PMU Data Analytics for the Resilient Electric Grid

Anurag K Srivastava

Washington State University

(anurag.k.srivastava@wsu.edu)

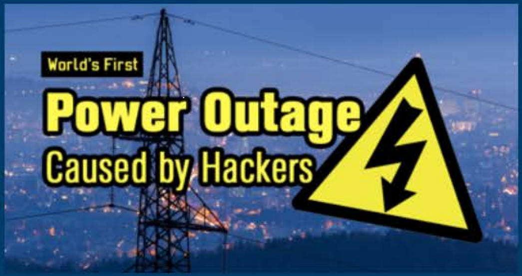
PSERC

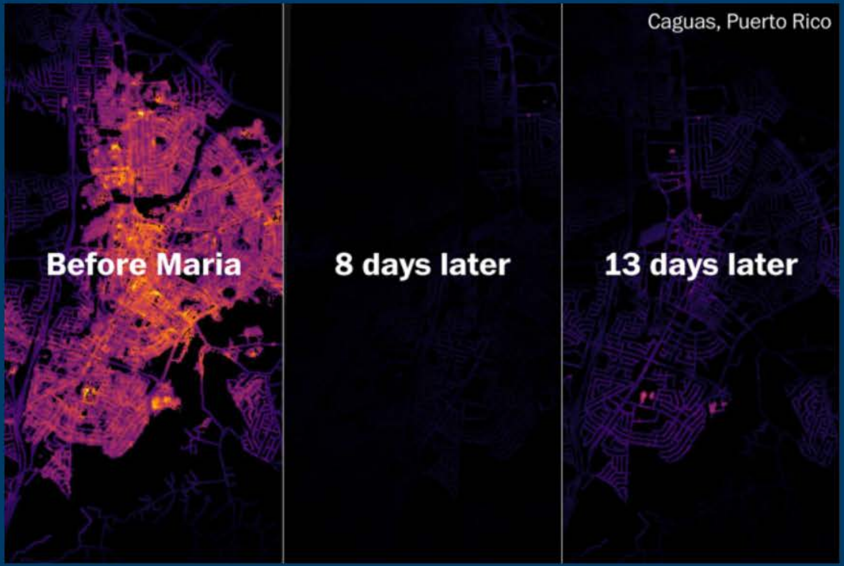
PSERC Webinar

April 16, 2019

- What is resiliency? How do we measure resiliency?
- How PMU data analytics enable resiliency?
- Use Case I: PMU based Anomaly/ Event Detection
- Use Case II: PMU based Failure Diagnosis
- Use Case III: Data-driven Resiliency Analysis
- Summary and Moving Forward

Power Grid: Reliable but Not Resilient





WRAP for Resiliency

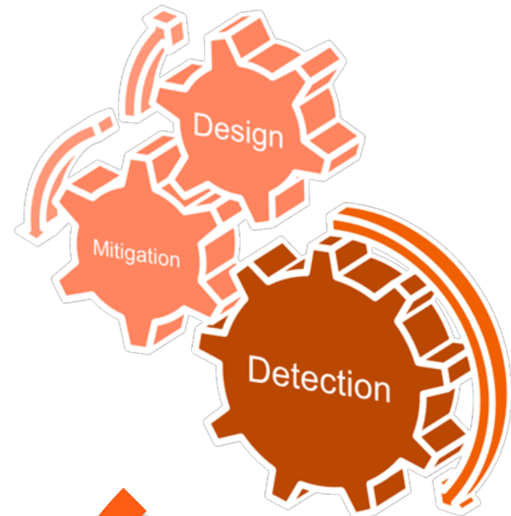
W **Withstand** any sudden inclement weather or human attack on the infrastructure.

R **Respond** quickly, to restore balance in the community as quickly as possible, after an inevitable attack.

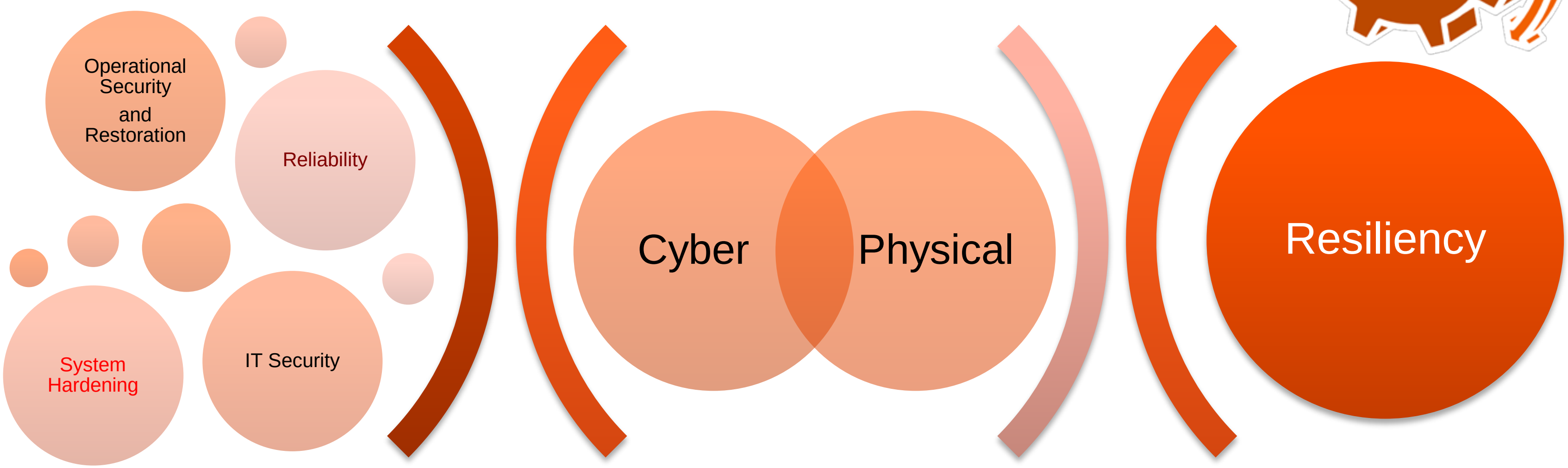
A **Adapt** to abrupt and new operating conditions, while maintaining smooth functionality, both locally and globally.

P **Predict** or **Prevent** future attacks based on patterns of past experiences, or reliable forecasts.

Electric Grid Resiliency



Resilience: The ability to supply its critical load through (and in spite of) extreme contingencies and low resource availability

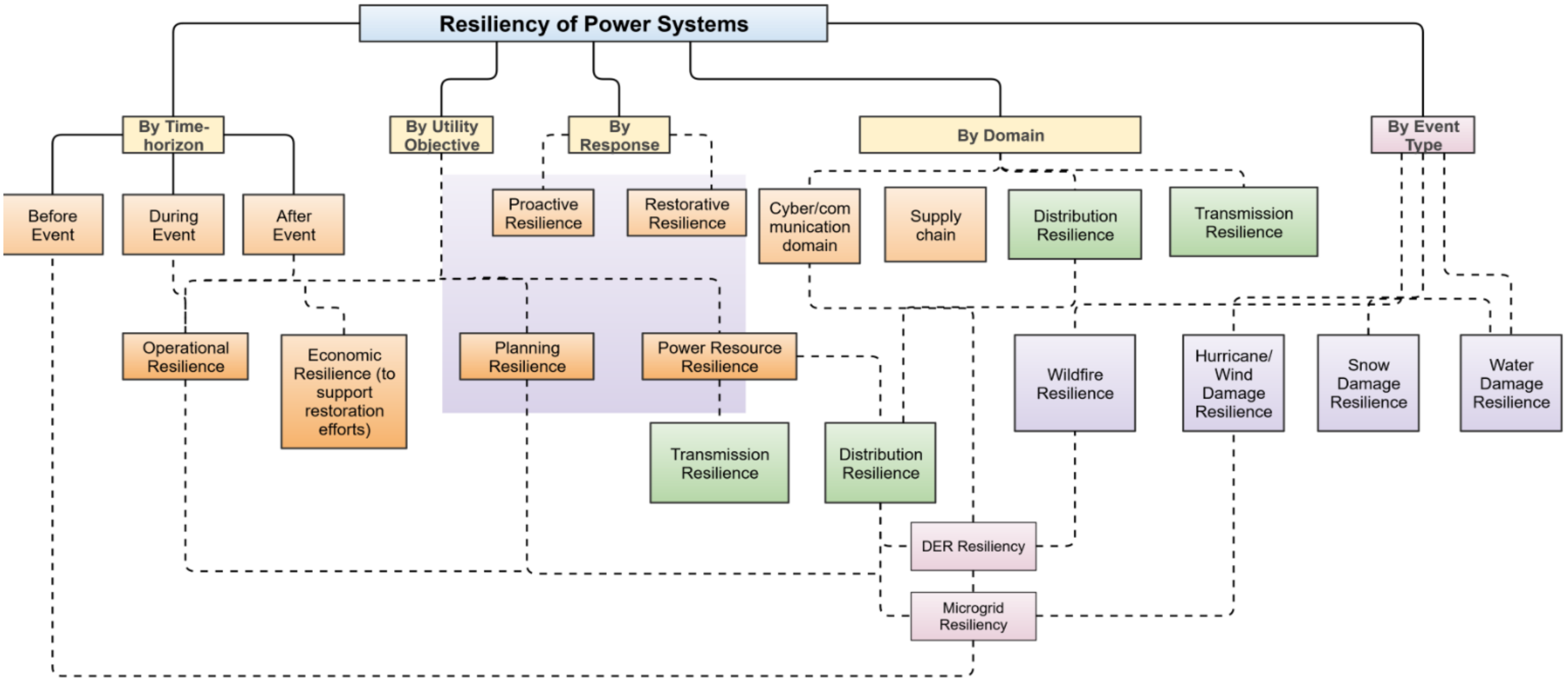


Existing Operational Practice

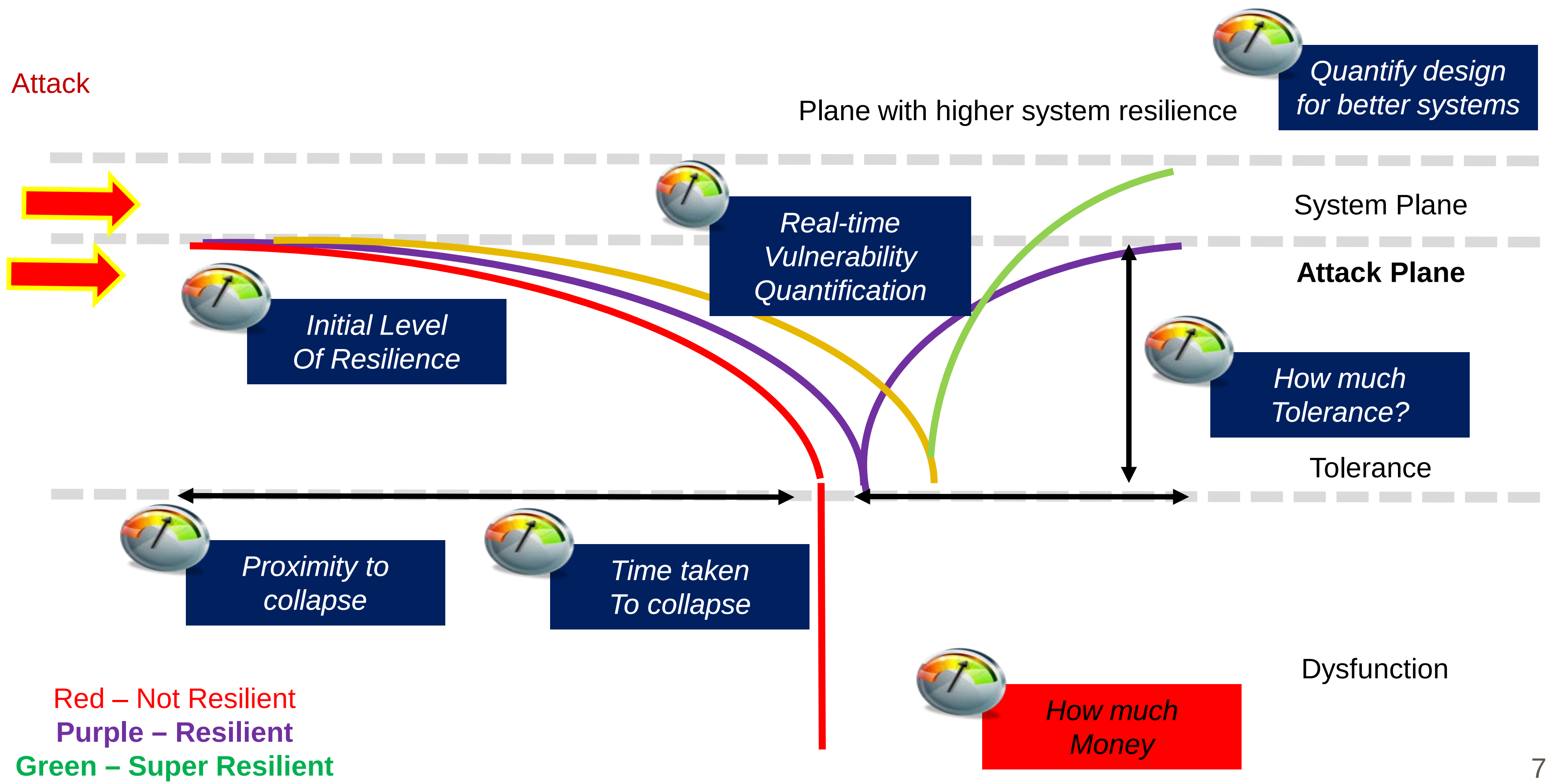
Integrated Cyber-Physical Analysis

Future Operation

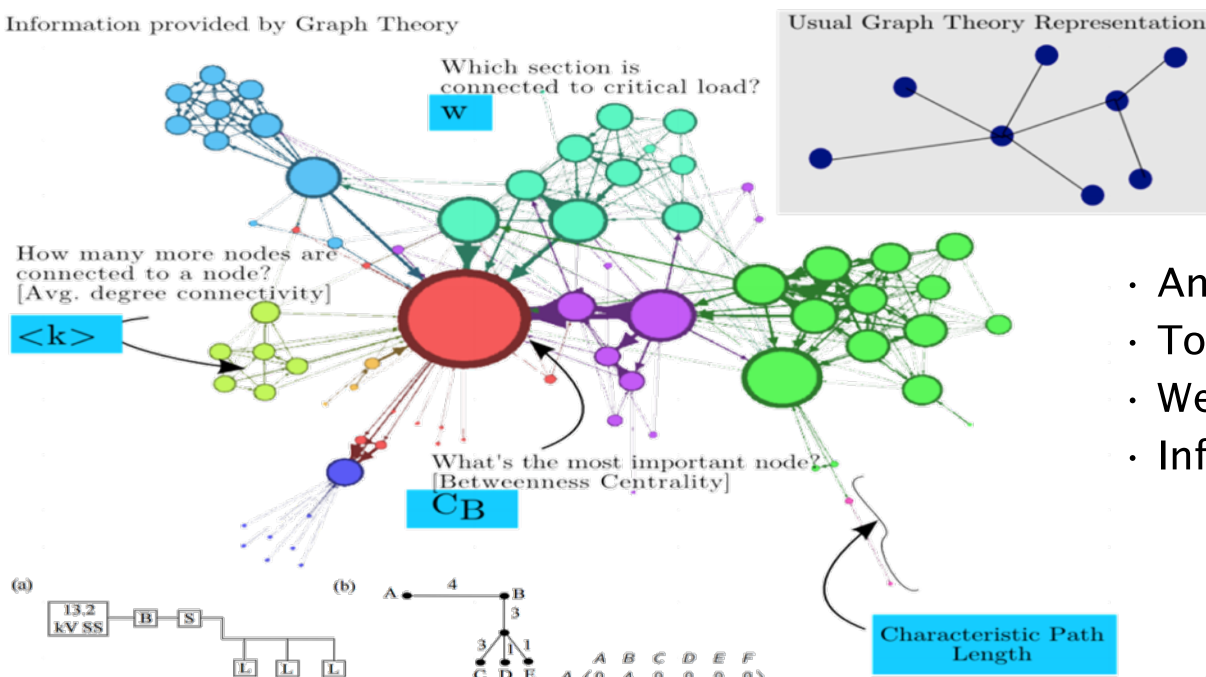
Taxonomy of Resiliency

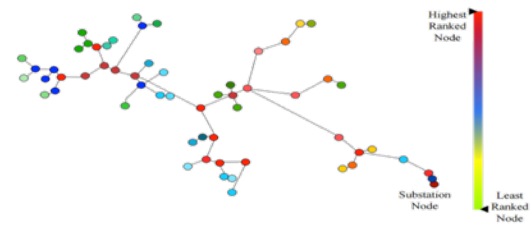


Can we measure resiliency?



Multi-criteria Decision for Physical Resiliency





$$A(G) = \begin{array}{c|cccccc} & A & B & C & D & E & F \\ \hline A & 0 & 4 & 0 & 0 & 0 & 0 \\ B & 4 & 0 & 3 & 0 & 0 & 0 \\ C & 0 & 3 & 0 & 3 & 1 & 1 \\ D & 0 & 3 & 0 & 3 & 1 & 1 \\ E & 0 & 0 & 1 & 0 & 0 & 0 \\ F & 0 & 0 & 1 & 0 & 0 & 0 \end{array}$$

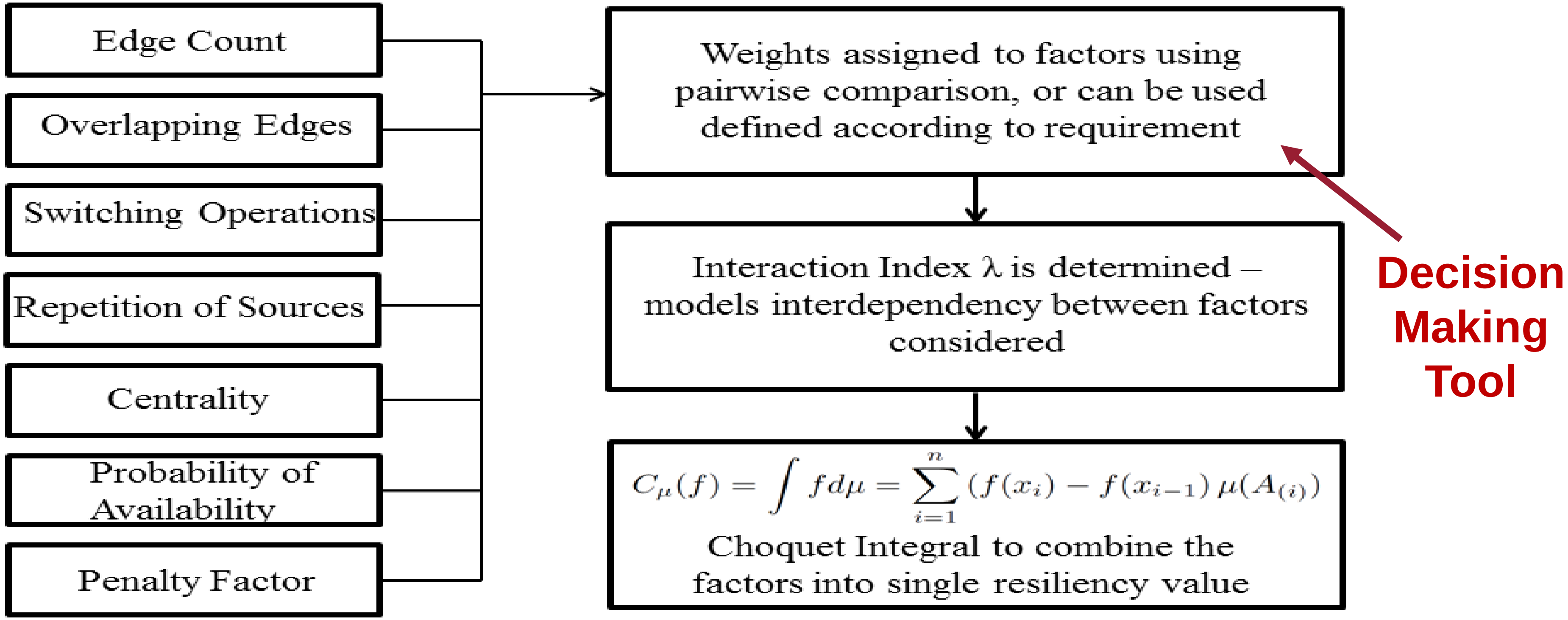
- Analytical Hierarchical Process
- Topology Parameters
- Weather Parameters
- Infrastructure Parameter

$$\rho(i,j)' = \frac{\rho(i,j) - \min_{i=1}^{\eta}(\rho(i,j))}{\max_{i=1}^{\eta}(\rho(i,j)) - \min_{i=1}^{\eta}(\rho(i,j))}$$

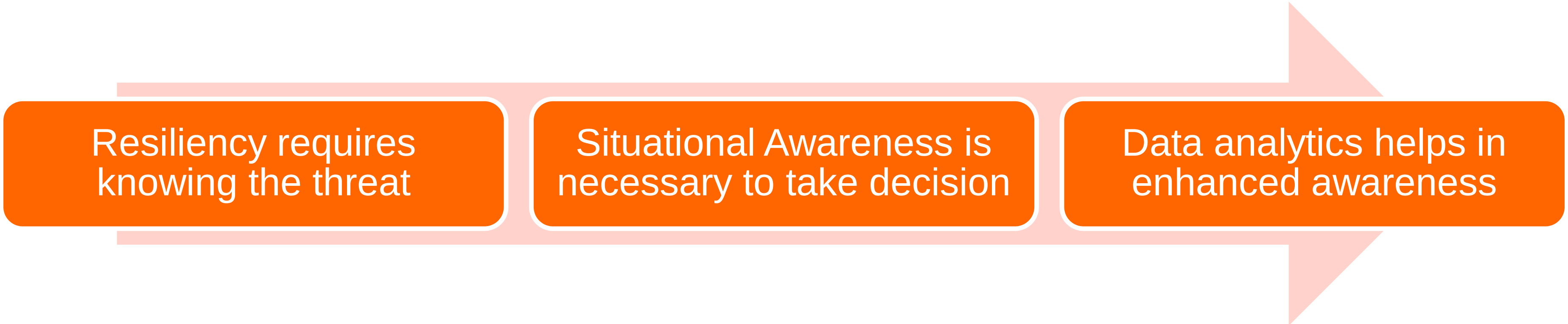
$$V = [A_{f_c}\ B_D\ C_{C_B}\ D_{l_G}\ E_{C_n}\ F_{\Delta\lambda}\ G_{\lambda_2}]^T$$

$$\Re_{\mathcal{T}} = \sum_{j=1}^{\eta} V_j \rho(i,j)'$$

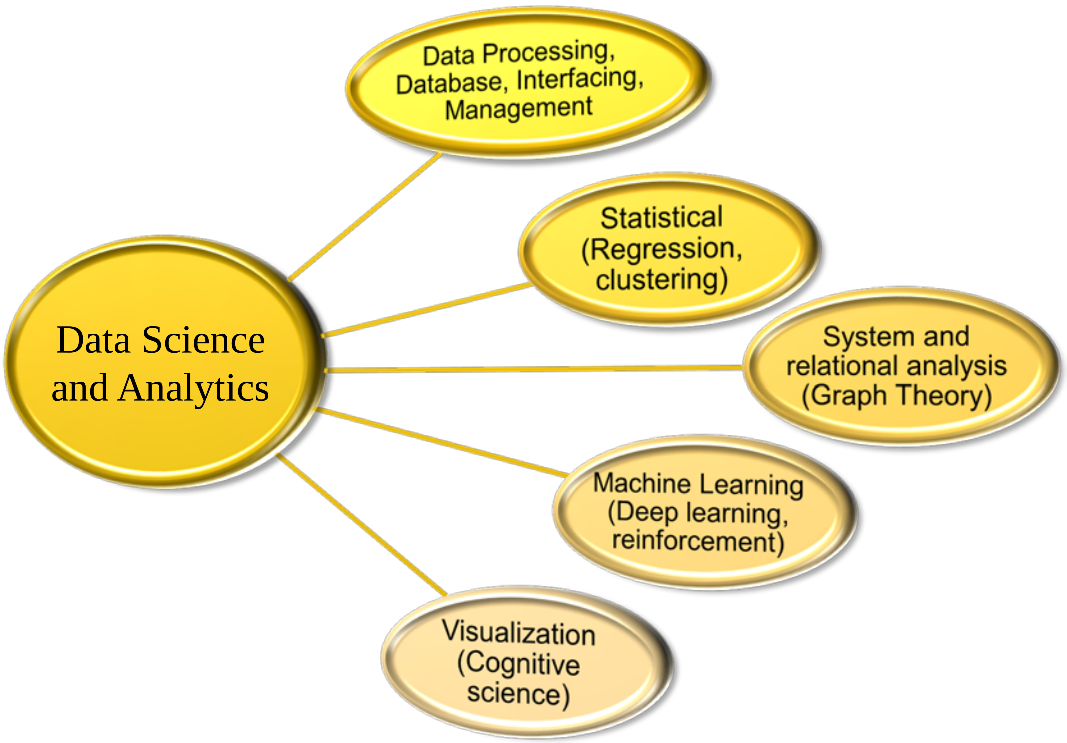
Overview of Resiliency Quantification Process



- What is resiliency? How do we measure resiliency?
- **How PMU data analytics enable resiliency?**
- Use Case I: PMU based Anomaly/ Event Detection
- Use Case II: PMU based Failure Diagnosis
- Use Case III: Data-driven Resiliency Analysis
- Summary and Moving Forward



- Predicts the future based on past patterns.
- Explores and examines data from multiple disconnected sources.
- Develop new analytical methods and machine learning models.
- Leverage data for relevant applications.
- Deliver actionable insights from the data.
- Store and process the data for insights.
- Design and create data reports using various reporting tools.
- Query database and package data for insights.



Data Collection by PMUs: Example of Operational Data

- PMU sampling rates: 30 per second
- Assume 100 values per second

If we assume all 100 points in a sub are PM

- Average data rate per sub is 10K/sec
- Average data rate for the total of 100 subs in a BA is 1M/sec
- Average data rate for the RC is then 10M/sec

Data Analytics Needed for Making Sense of this Steaming Operational Data for Cyber or Physical Events !!!!

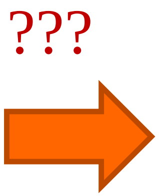
Credit: Prof Anjan Bose, WSU

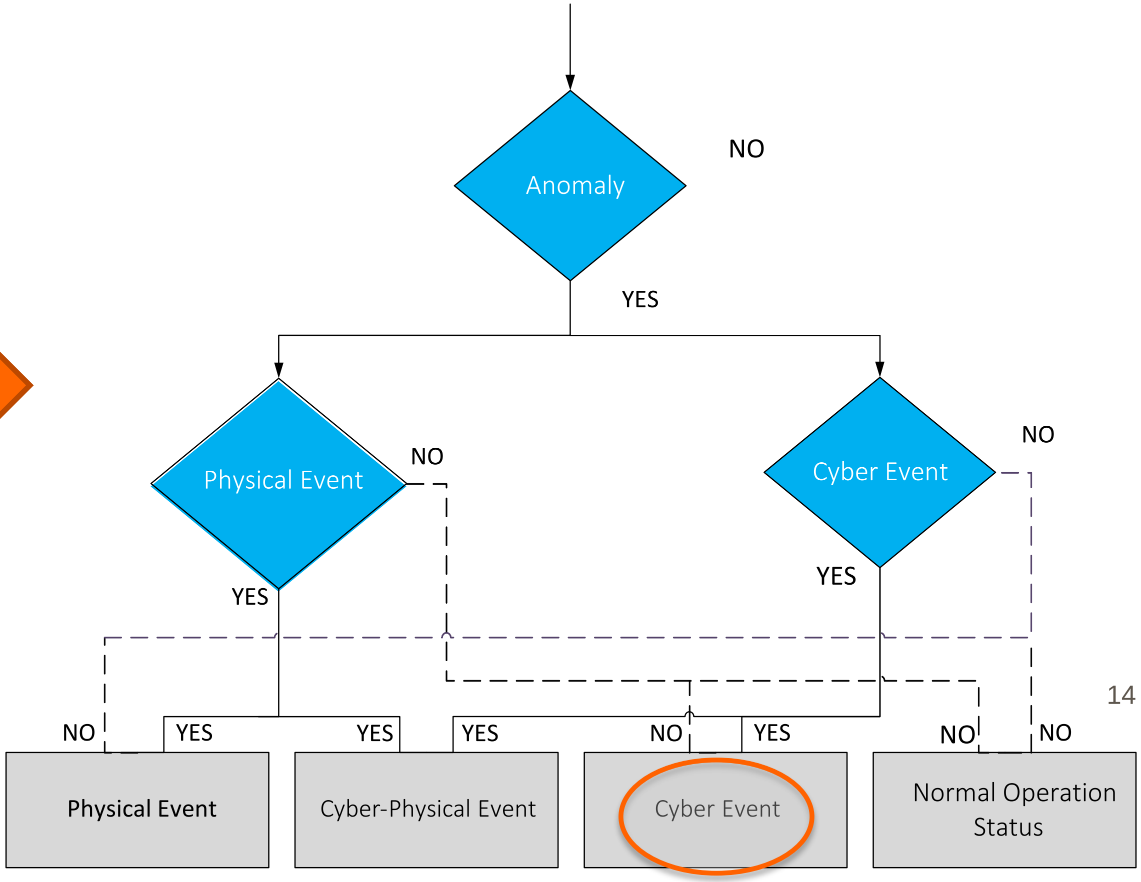
- What is resiliency? How do we measure resiliency?
- How PMU data analytics enable resiliency?
- **Use Case I: PMU based Anomaly/ Event Detection**
- Use Case II: PMU based Failure Diagnosis
- Use Case III: Data-driven Resiliency Analysis
- Summary and Moving Forward

Use Case I: Anomaly Detection and Classification: Processing lots of data in real time

Data

- Physical
  - PMU measurements
  - CT/PT measurements
  - Breaker status
  - Relay operations
- Cyber
  - Network data
    - Pcaps, netflows, Ids alerts
  - Hosts
    - Event logs, Ids alerts

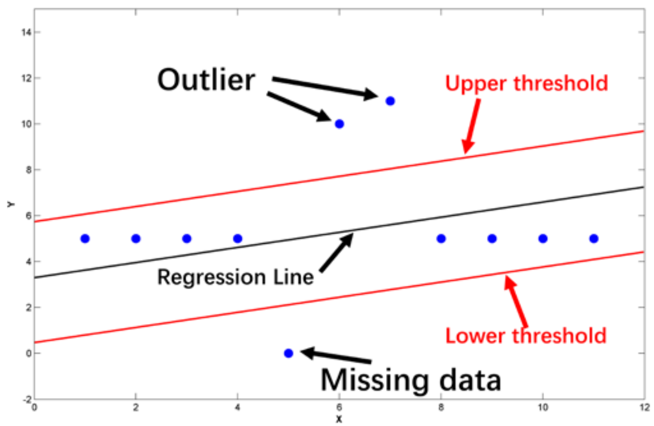




Options?

Linear regression

find straight line $y = \alpha + \beta x$ to provide a "best" fit for the data points w.r.t least-squares



Chebyshev method

Determine a lower bound of the percentage of data that exists within k standard deviations from t

$$P(|X - \mu| \leq k\sigma) \geq (1 - \frac{1}{k^2})$$

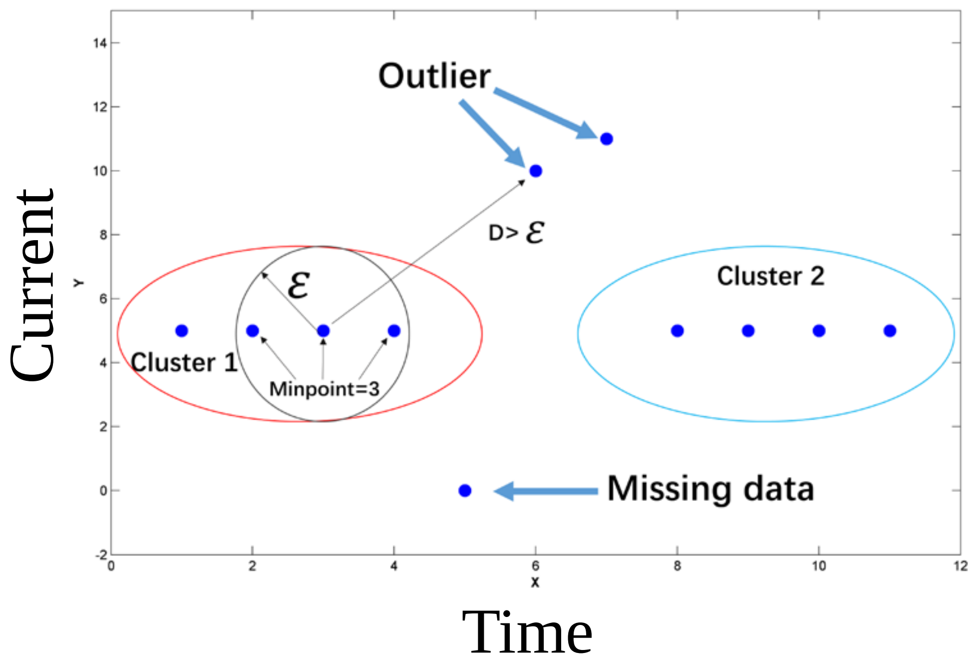
μ: mean, σ: standard deviation, k: number of standard deviations from the mean.

Amidan, Brett G., Thomas A. Ferryman, and Scott K. Cooley. "Data outlier detection using the Chebyshev theorem." *Aerospace Conference, 2005 IEEE*. IEEE, 2005.

DBSCAN

- DBSCAN uses two thresholds radius ε and *min*.
- A data point is a center node if it has more than *min* ε-neighbors (points within distance ε);
- Two centers are reachable if they are in ε-neighbor of each other; a cluster is a sequence of reachable centers and their ε-neighbors

- New clusters is formed after the event ends. Points far away from any cluster are outliers.



Does standalone method suffice?

LSTM Auto-encoder Model

- The model consists of two RNNs – the encoder LSTM and the decoder LSTM as shown in Figure
- The input to the model is a sequence of vectors (PMU data)
- The encoder LSTM reads in this sequence
- Once input vector is read, the decoder LSTM takes over and outputs a prediction for the target sequence
- The encoder can be seen as ‘creating a list’ of new inputs and previously constructed list (learned weights).
- The decoder essentially unrolls this list, with the hidden to output weights extracting the element at the top of the list and the hidden to hidden weights extracting the rest of the list.
- Thus the LSTM weights are learned using the auto encoder method.

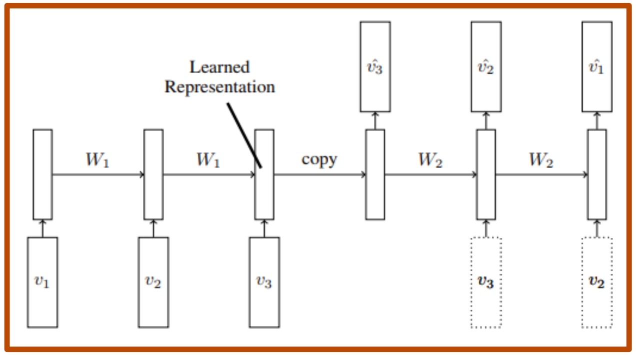
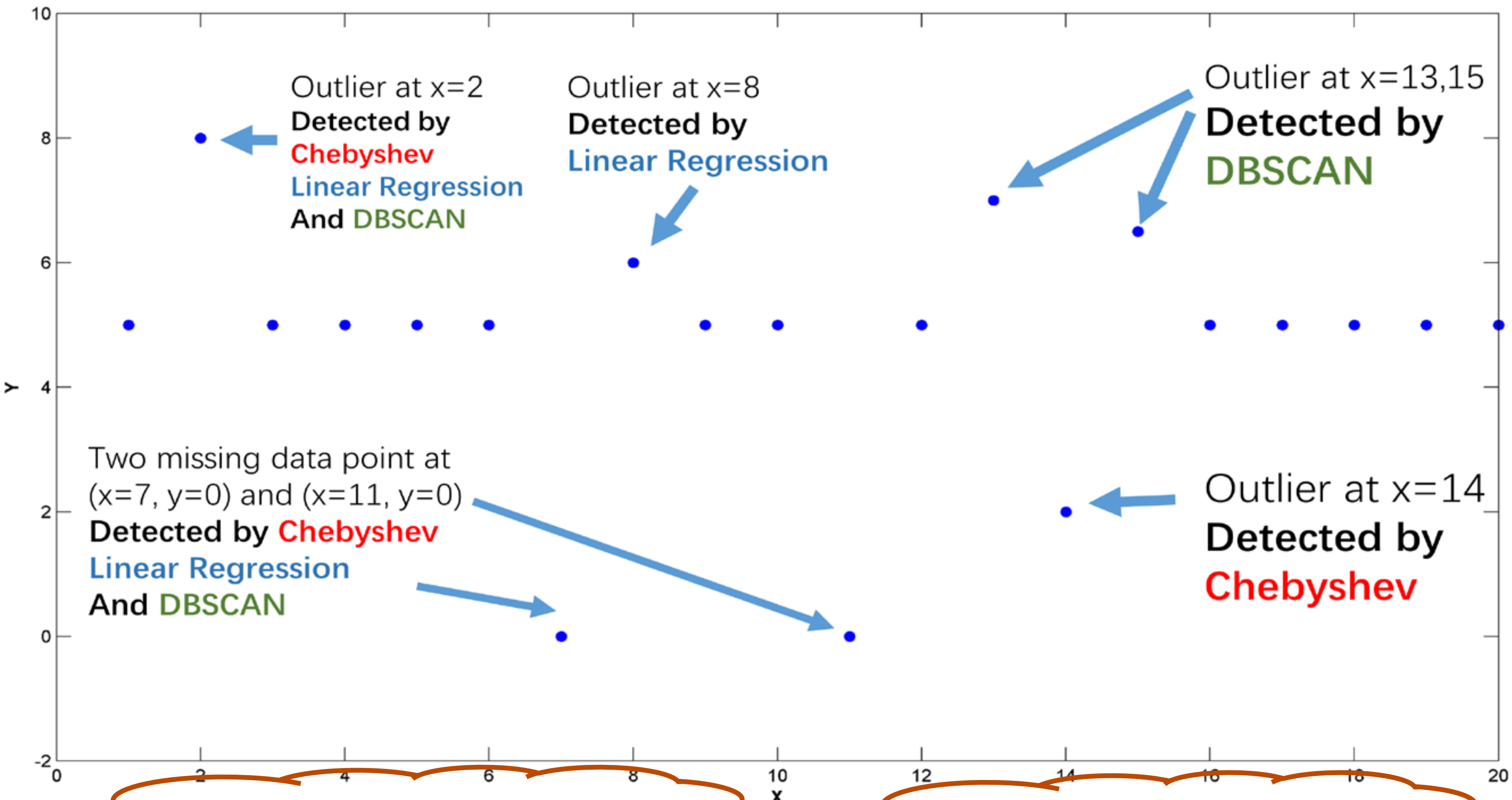


Fig 3: LSTM Auto encoder Model

$$
\begin{aligned}
\mathbf{i}_t &= \sigma\left(W_{xi}\mathbf{x}_t + W_{hi}\mathbf{h}_{t-1} + W_{ci}\mathbf{c}_{t-1} + \mathbf{b}_i\right), \\
\mathbf{f}_t &= \sigma\left(W_{xf}\mathbf{x}_t + W_{hf}\mathbf{h}_{t-1} + W_{cf}\mathbf{c}_{t-1} + \mathbf{b}_f\right), \\
\mathbf{c}_t &= \mathbf{f}_t\mathbf{c}_{t-1} + \mathbf{i}_t \tanh\left(W_{xc}\mathbf{x}_t + W_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c\right), \\
\mathbf{o}_t &= \sigma\left(W_{xo}\mathbf{x}_t + W_{ho}\mathbf{h}_{t-1} + W_{co}\mathbf{c}_t + \mathbf{b}_o\right), \\
\mathbf{h}_t &= \mathbf{o}_t \tanh(\mathbf{c}_t).
\end{aligned}
$$

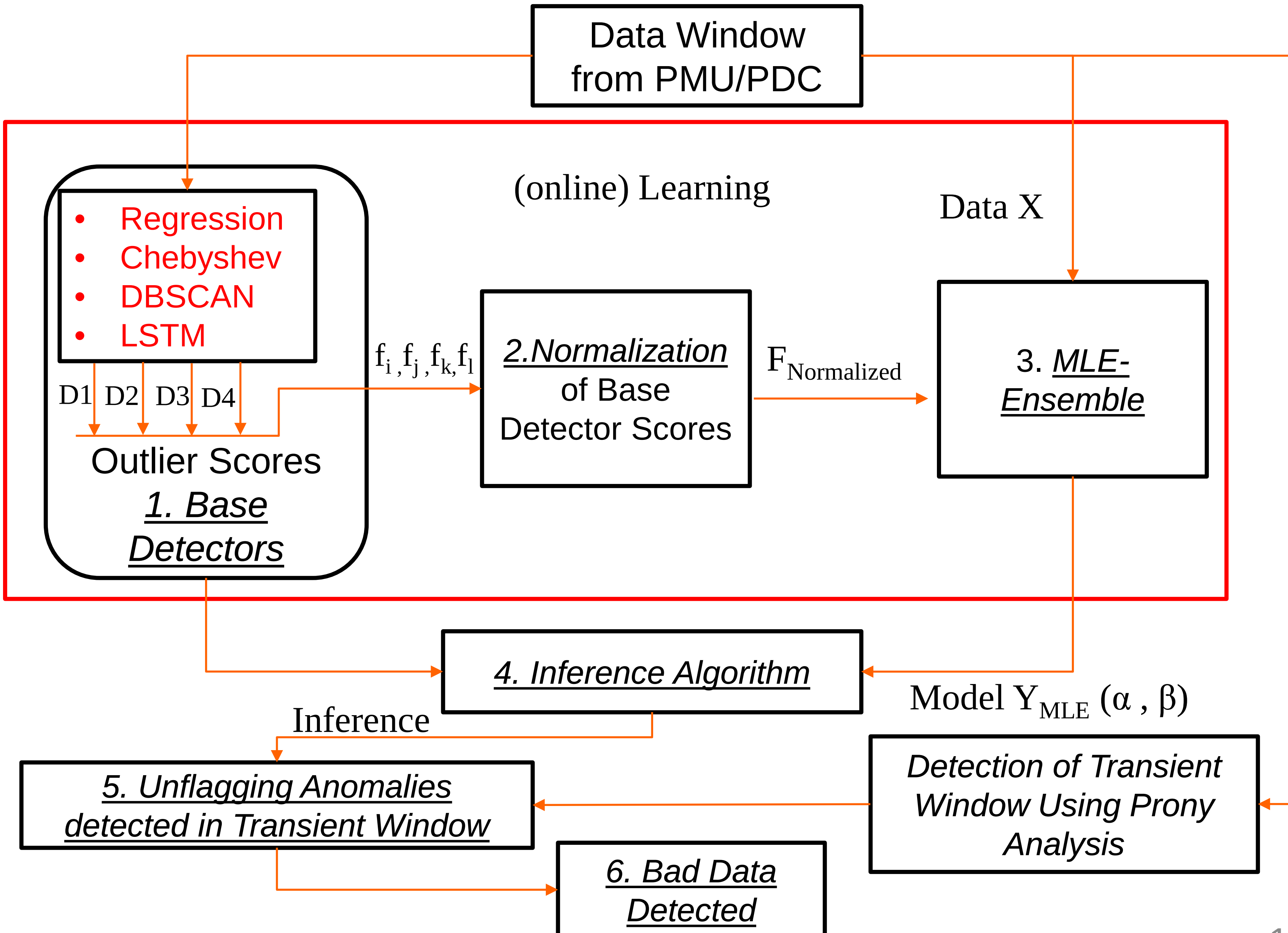


No Single Winner!

Lack of training data

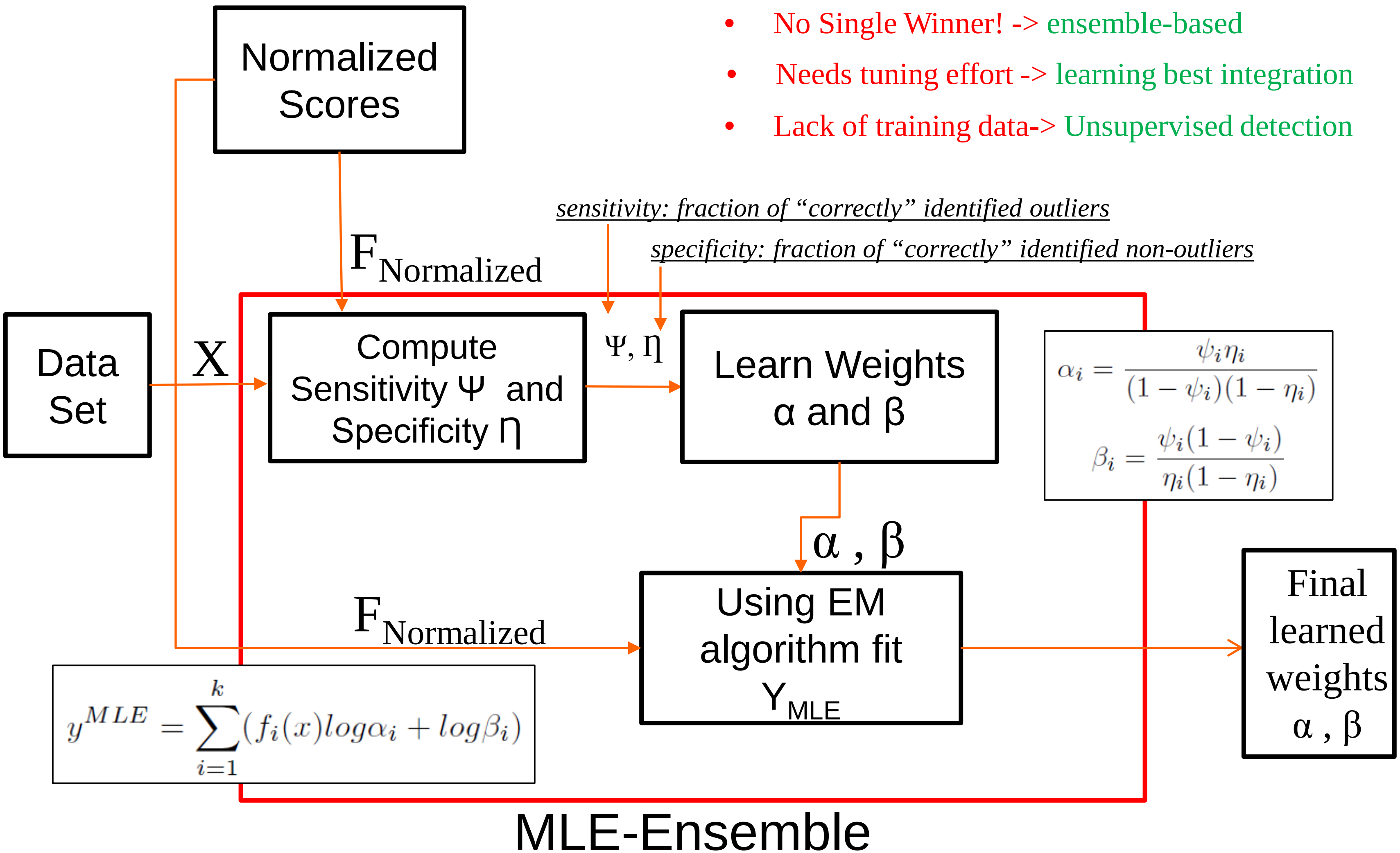
Needs tuning effort

Anomaly Detection with Ensemble



Maximum Likelihood Estimator (MLE)

- No Single Winner! -> ensemble-based
- Needs tuning effort -> learning best integration
- Lack of training data-> Unsupervised detection



Performance Metrics for Ensemble Based Technique

Given a PMU detector D and PMU data X, denote the actual anomaly data set as B_T , and the anomaly reported by D as B_D, the performance of D is evaluated using three metrics as follows.

Precision: Precision measures the fraction of true anomaly data in the reported ones from D, defined as

$$Precision = \frac{|B_D \cap B_T|}{|B_D|}$$

Recall: Recall measures the ability of D in finding all outliers, defined as

$$Recall = \frac{|B_D \cap B_T|}{|B_T|}$$

False Positive: False positive (FP) evaluates the possibility of false anomaly data detection; the smaller, the better

$$FP = 1 - \frac{|B_D \cap B_T|}{|B_D|}$$

Simulation results for SyncAD

RTDS simulated PMU data (1.5 hours)

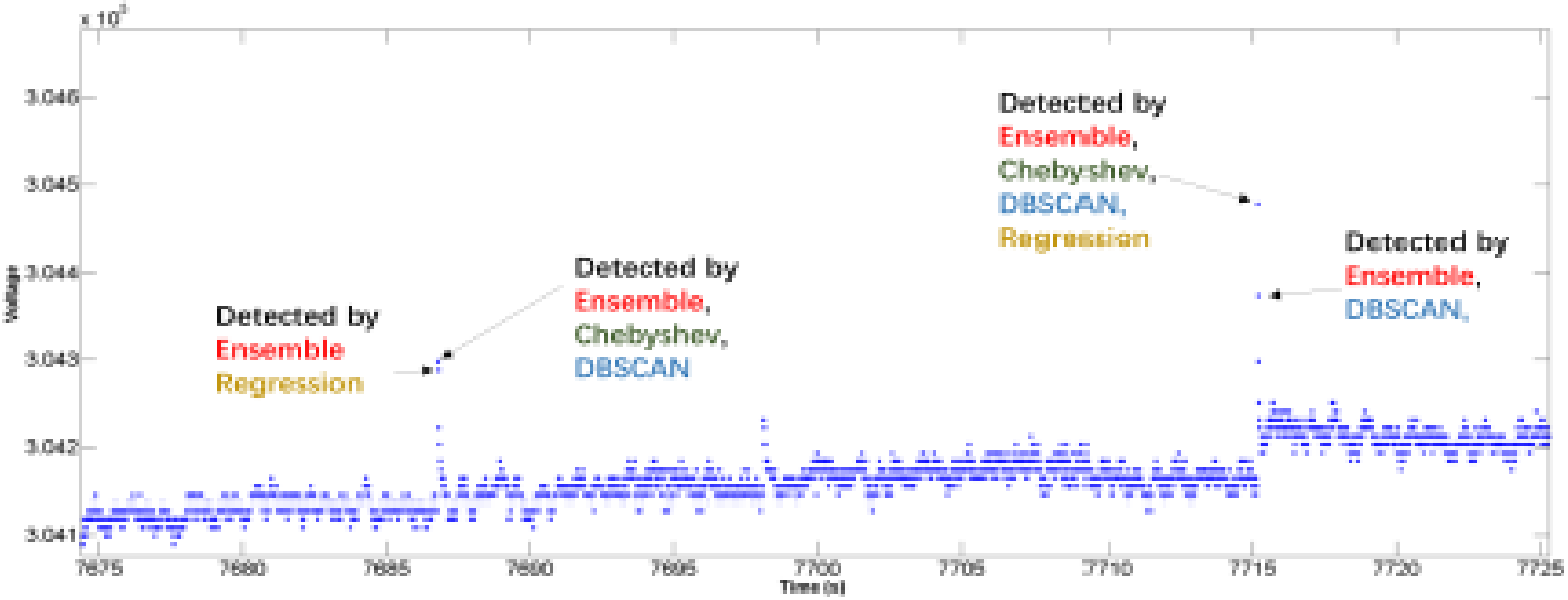
	Recall	Precision	False positive
Linear Regression	0.9021	0.8565	0.1435
DBSCAN	0.8821	0.8821	0.1179
Chebyshev	0.9154	0.8754	0.1246
LSTM	0.9298	0.8554	0.1446
MLE ensemble	**0.9351**	**0.8913**	**0.1087**

Tests on the RTDS simulated PMU data (1.5 hours, 5% bad data points, 5%-10% range)

	Recall	Precision	False positive
Linear Regression	0.7854	0.7655	0.2345
DBSCAN	0.7216	0.7015	0.2985
Chebyshev	0.8125	0.7542	0.2458
LSTM	0.8298	0.7754	0.2246
MLE ensemble	**0.8912**	**0.9021**	**0.0979**

Tests on the RTDS simulated PMU data (1.5 hours, 10% bad data points, 10%-20% range)

Results with SyncAD using Real PMU Data

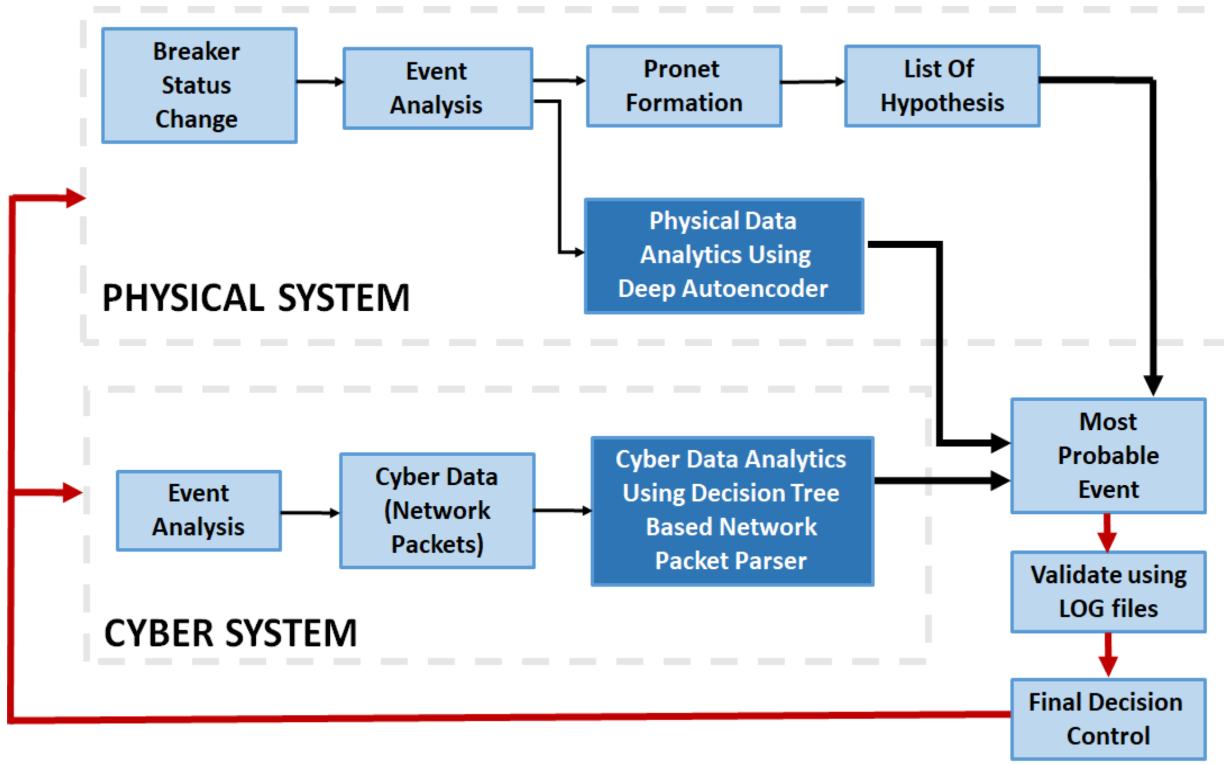


- What is resiliency? How do we measure resiliency?
- How PMU data analytics enable resiliency?
- Use Case I: PMU based Anomaly/ Event Detection
- **Use Case II: PMU based Failure Diagnosis**
- Use Case III: Data-driven Resiliency Analysis
- Summary and Moving Forward

Use case II: Cyber-physical Data Analytics in Protection Failure

- Protection Mal-operation is #1 concern according to NERC

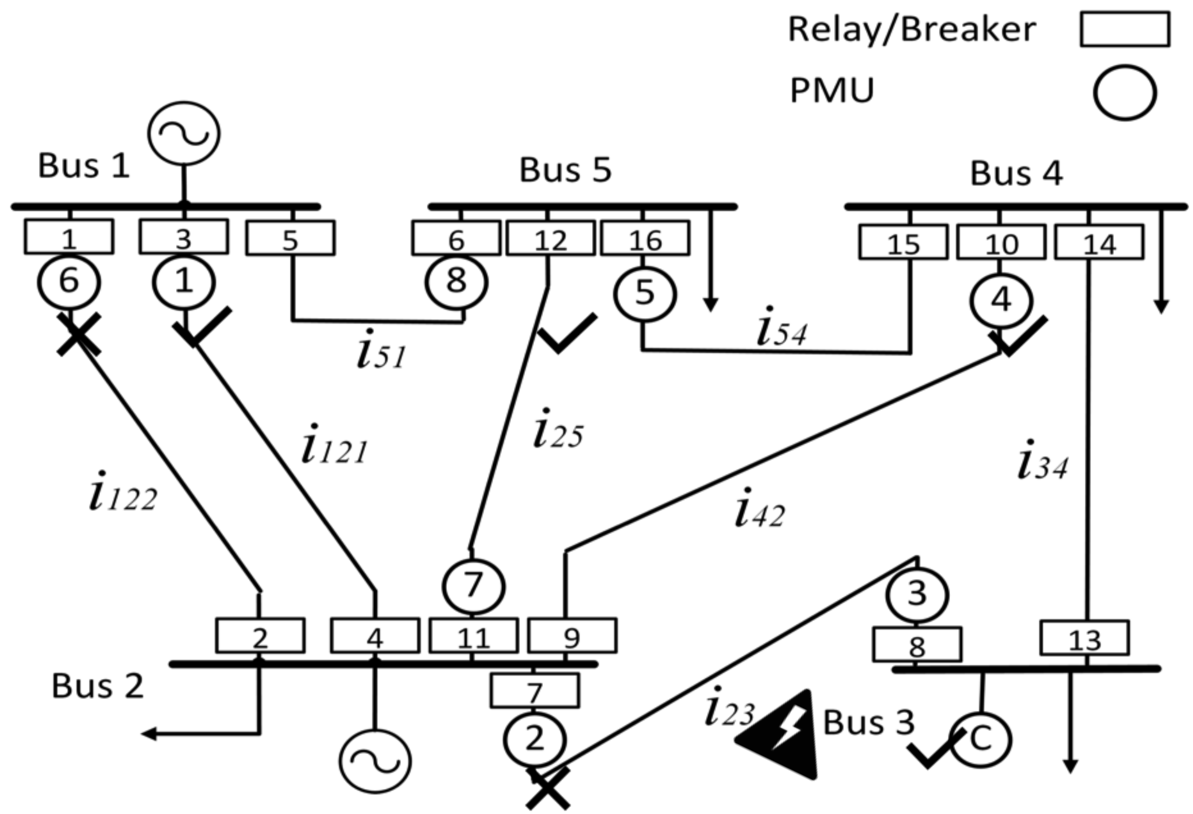
- Protection and associated control is becoming more digital



Abnormal Operation

A fault occurs on line 2-3 Relays 7 and 8 are expected to open their corresponding breakers but relay 7 doesn’t respond

To compensate relay’s 7 malfunction, relays 1, 3, 10 and 12 should open their corresponding breakers but relay 1 malfunctions.

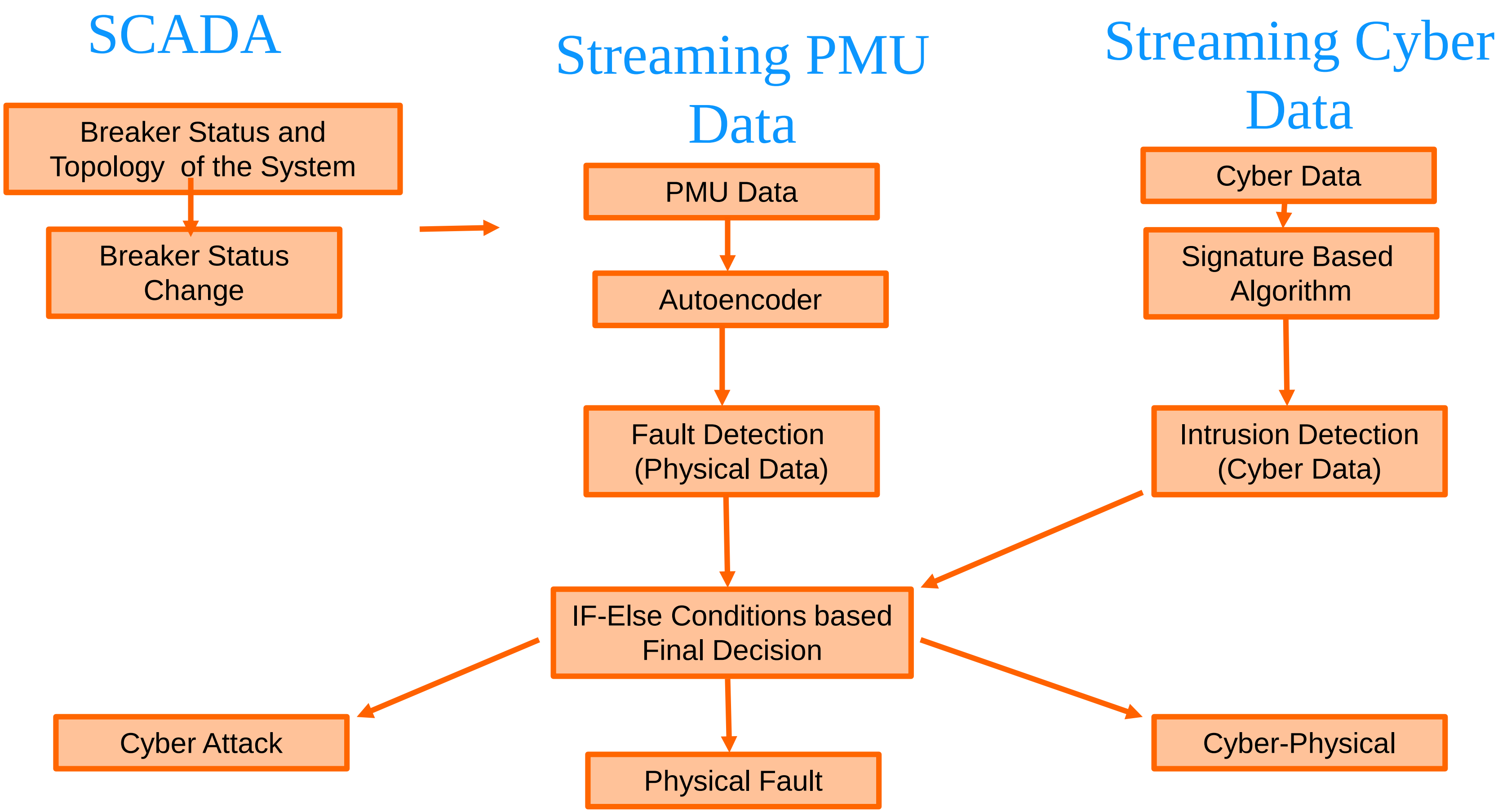


Cyber Physical Security Analytics for Anomalies in Transmission Protection Systems

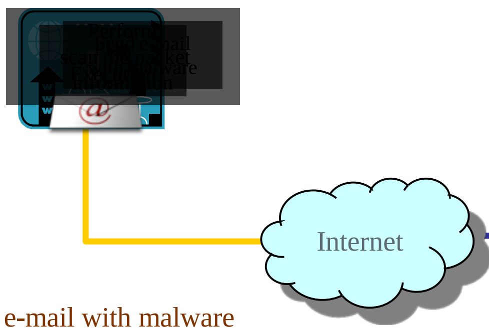
Hypothesis Generation

Hypothesis #	Location of fault	Initial Incident	Consequential Incident
Actual Scenario	Line 2-3	Breaker 8 tripped Relay 7 malfunctioned	Breakers 3,10,12 tripped Relay 1 malfunctioned
Hypothesis 1	Line 2-4	Breaker 10 tripped Relay 9 malfunctioned	Breakers 3,8,12 tripped Relay 1 malfunctioned Relay 6 Tripped
Hypothesis 2	Line 2-1-2	Breaker 3 tripped Relay 4 malfunctioned	Breakers 8,10,12 tripped Relay 1 malfunctioned Relay 6 Tripped
Hypothesis 3	Line 1-5	Breaker 6 tripped Relay 5 malfunctioned	Relay 2, 3, 4 malfunctioned Breakers 8,10,12 tripped
Hypothesis 4	Line 2-5	Breaker 12 tripped Relay 11 malfunctioned	Breakers 3, 8, 10 tripped Relay 1 malfunctioned Relay 6 Tripped

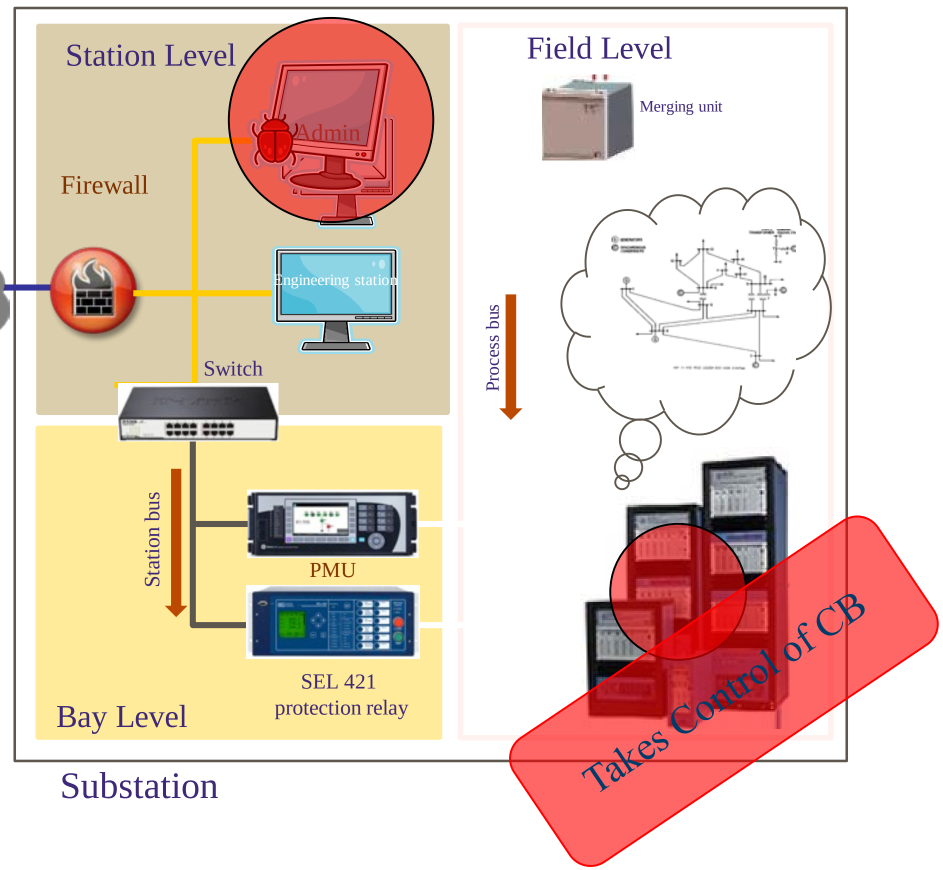
Data Analytics For Event Classification



Simulating Cyber Attack on a Relay



1. Attacker sends an e-mail with malware
2. E-mail recipient opens the e-mail and the malware gets installed quietly
3. Using the information that malware gets, hacker is able to take control of the e-mail recipient's PC and get access of two-level password
4. Analysis IEC 61850 protocol(GOOSE, SMV packet) information and relay setting file
5. Manipulate MMS packet and relay configuration session information
6. Takes control of circuit breaker or change the setting of relay



Detect Intrusion Using Cyber Data From Relay.

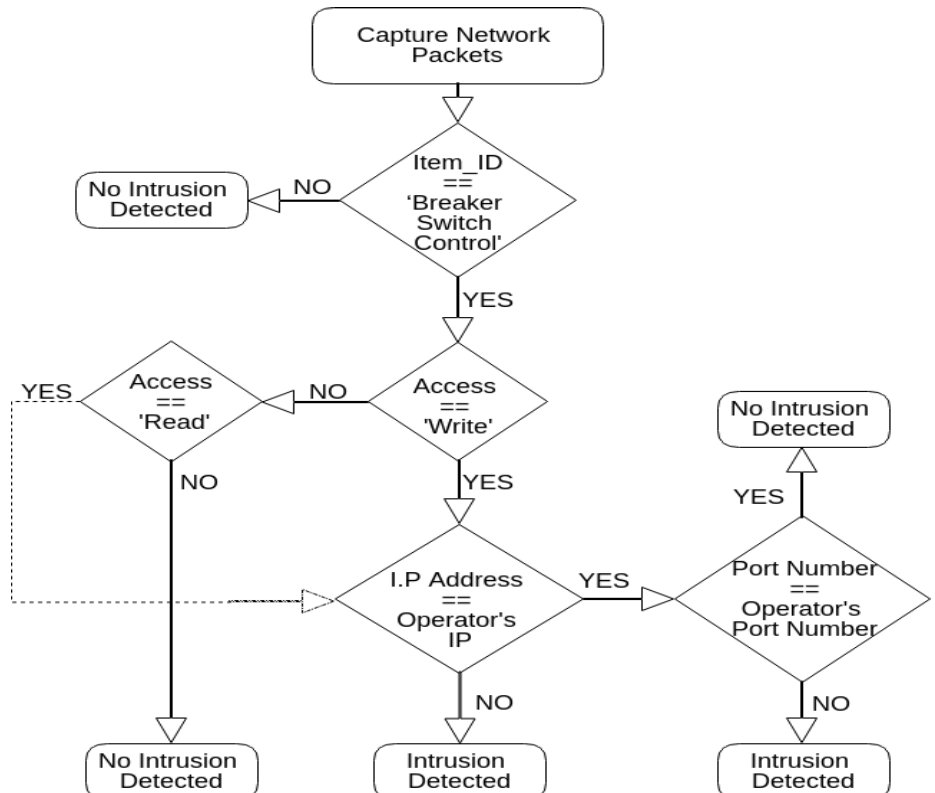
Relay IP address: 192.168.0.16 || Operator IP address: 192.168.0.23 || Unauthorized IP address:192.168.0.14

No.	Time	Source	Destination	Protocol	Length	Info
2296	126.405616	192.168.0.14	192.168.0.16	MMS	229	confirmed-RequestPDU
2297	126.409243	192.168.0.16	192.168.0.14	MMS	84	confirmed-ResponsePDU
2298	132.293425	192.168.0.14	192.168.0.16	MMS	229	confirmed-RequestPDU
2299	132.296947	192.168.0.16	192.168.0.14	MMS	84	confirmed-ResponsePDU
2300	137.581544	192.168.0.14	192.168.0.16	MMS	229	confirmed-RequestPDU
2301	137.645231	192.168.0.16	192.168.0.14	MMS	84	confirmed-ResponsePDU
2302	141.453519	192.168.0.14	192.168.0.16	MMS	229	confirmed-RequestPDU
2303	141.456890	192.168.0.16	192.168.0.14	MMS	84	confirmed-ResponsePDU
2304	145.213451	192.168.0.14	192.168.0.16	MMS	229	confirmed-RequestPDU
2305	145.216523	192.168.0.16	192.168.0.14	MMS	84	confirmed-ResponsePDU
2306	151.245001	192.168.0.14	192.168.0.16	MMS	229	confirmed-RequestPDU

<u>Attack Scenario For Relay</u>

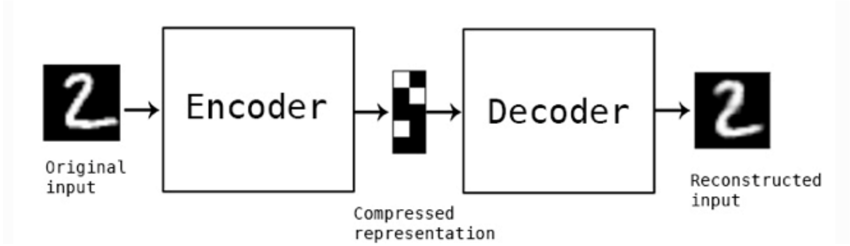
Communication between Relay and Un-authorized IP Address-(Attacker)

Detecting an Intrusion :



Detect Intrusion Using Physical Data From PMU

Algorithm Description :



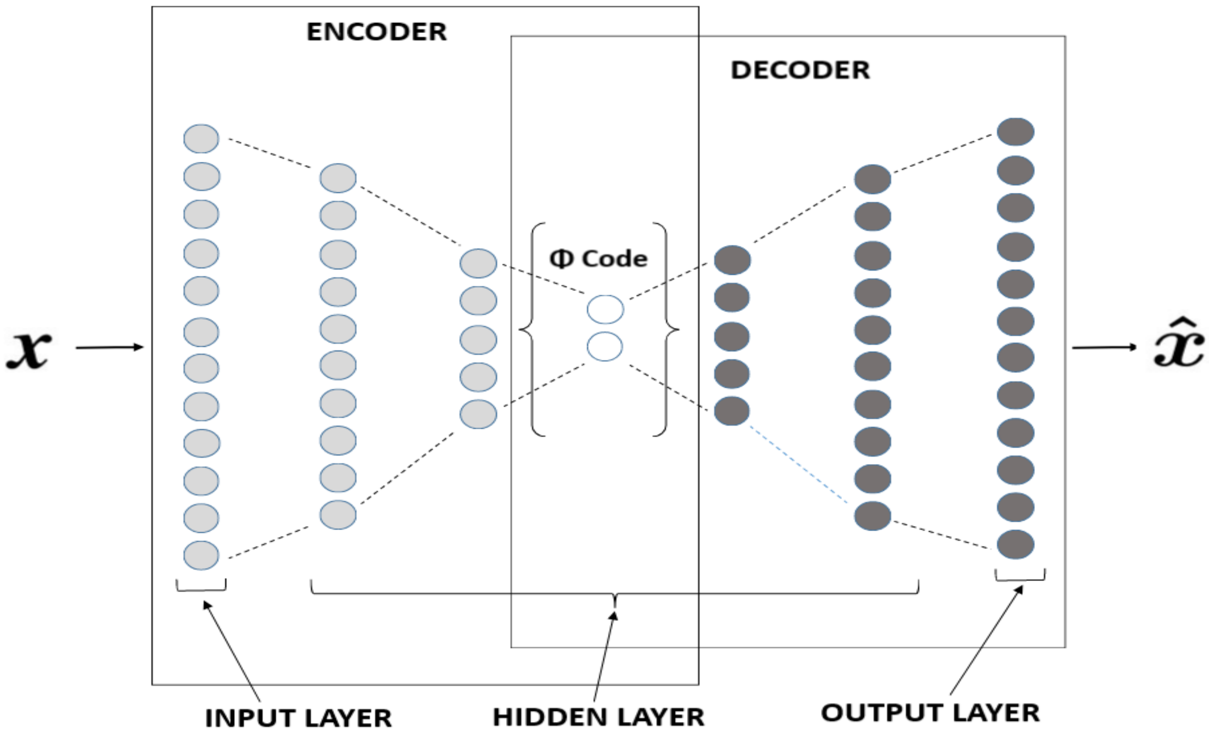
- Basic Idea : Reconstruction of input feature vector with minimum loss (Mean Square Error)

- Train the algorithm on input data consisting of no anomalies.
  Output Result : Reconstructed input feature vector with low MSE.

- Test the algorithm on input data consisting of anomalies.
  Output Result : Reconstructed input feature vector with high MSE.

- We want our algorithm to have high MSE on input data consisting of anomalies and low MSE on input data consisting of no anomalies.

Detect Intrusion Using Physical Data From PMU



Architecture Of
Stacked Autoencoder

Loss Function : Mean Squared Error
Optimizer : ADAM

x : Input Feature Vector

$\hat{x}$: Reconstructed Output Feature Vector

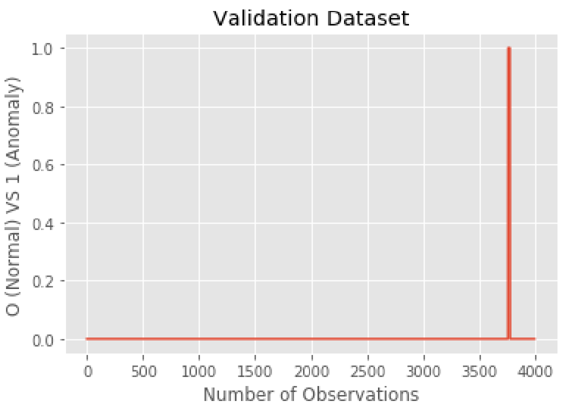
Detect Intrusion Using Physical Data From PMU

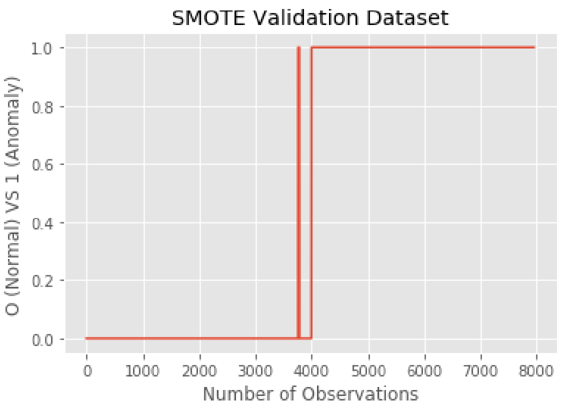
Dataset Description :

Dataset	# PMU Readings (Total : 37500)
Training Dataset (No Fault)	22250
Testing Dataset (No Fault)	11250
Validation Dataset (Fault)	4000

Types Of Validation Dataset:

Validation Dataset	PMU Readings (# Normal Instances)	PMU Readings (# Anomalous Instances)
Type 1	3979	21
Type 2 (Synthetic Minority Oversampling -SMOTE)	3979	3979





Detect Intrusion Using Physical Data From PMU

Evaluation Metrics

The intersection between actual values and predicted values yield four possible situations:

- True Positive (TP): Positive instances correctly classified.
- False Positive (FP): Negative instances classified as positive.
- True Negative (TN): Negative instances correctly classified as negative.
- False Negative (FN): Positive instances classified as negative.

Classification Measures:

Accuracy is calculated as the number of correctly classified instances over total number of instances evaluated.

$$\text{Accuracy} = \frac{TP + TN}{Total\ instances}$$

Precision is the percentage of correctly predicted instances over the total instances predicted for positive class.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall is the percentage of correctly classified instances over the total actual instances for the positive class.

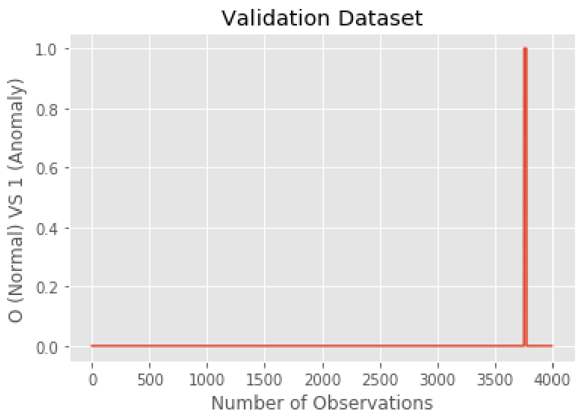
$$\text{Recall} = \frac{TP}{TP + FN}$$

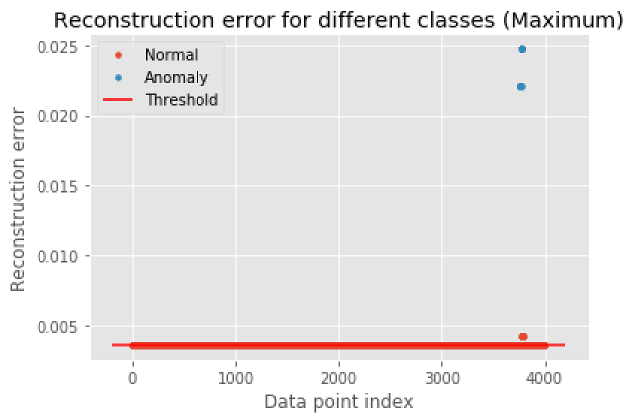
F-Measure is a measure of test accuracy.

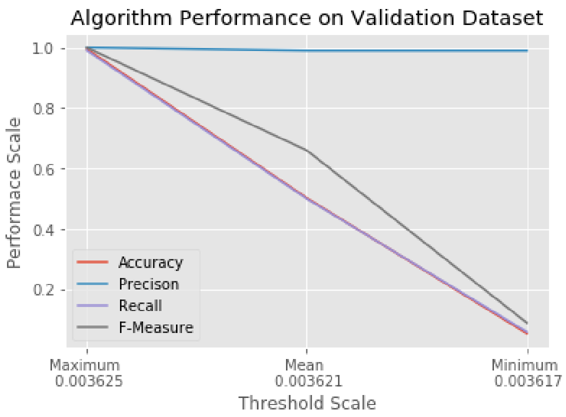
$$\text{F-measure} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Detect Intrusion Using Physical Data From PMU

Autoencoder Evaluation On Type 1 (Validation Dataset)

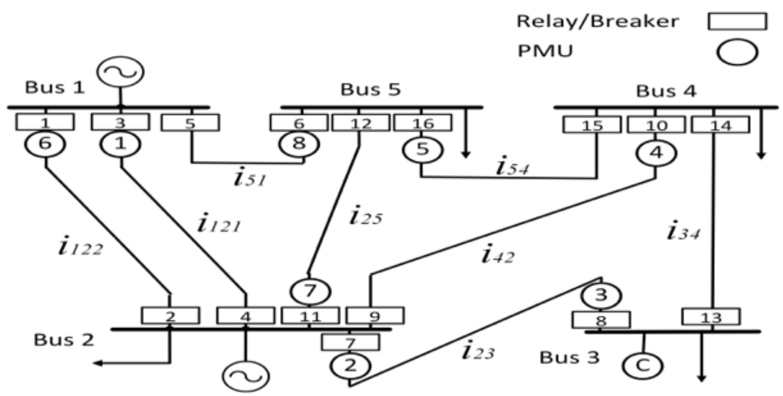






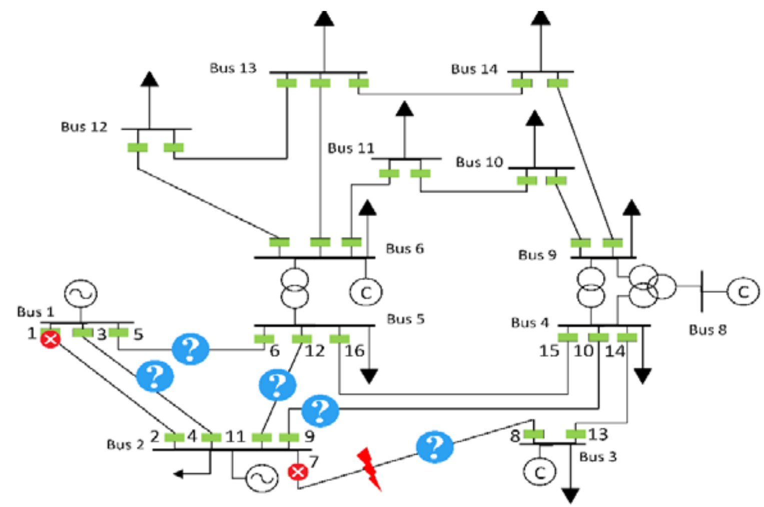
Threshold (Test Data)	Accuracy	Precision	Recall	F-Measure
0.003617 (Minimum)	5.50%	0.99	0.06	0.09
0.003621 (Mean)	50.25%	0.99	0.50	0.66
0.003625 (Maximum)	99.48%	1.0	0.99	1.00

Decision Based On Data Analytics And Validation Using Additional Non-Streaming Data



Scenario	Location of Fault	Initial incident	Consequential incident
	Line 2-3	Breaker 8 tripped Relay 7 malfunctioned	Breakers 3, 10, 12 tripped Relay 1 malfunctioned Relay 6 tripped
Scn 1	Line 2-4	Breaker 10 tripped Relay 9 malfunctioned	Breakers 3, 8, 12 tripped Relay 1 malfunctioned Relay 6 tripped
Scn 2	Line 2-1-2	Breaker 3 tripped Relay 4 malfunctioned	Breakers 8, 10, 12 tripped Relay 1 malfunctioned Relay 6 tripped
Scn 3	Line 1-5	Breaker 6 tripped Relay 5 malfunctioned	Relays 2,3,4 malfunctioned Breakers 8, 10, 12 tripped
Scn 4	Line 2-5	Breaker 12 tripped Relay 11 malfunctioned	Breakers 3, 8, 10 tripped Relay 3 malfunctioned Relay 6 tripped

Cyber Physical Security Analytics for Anomalies in Transmission Protection Systems



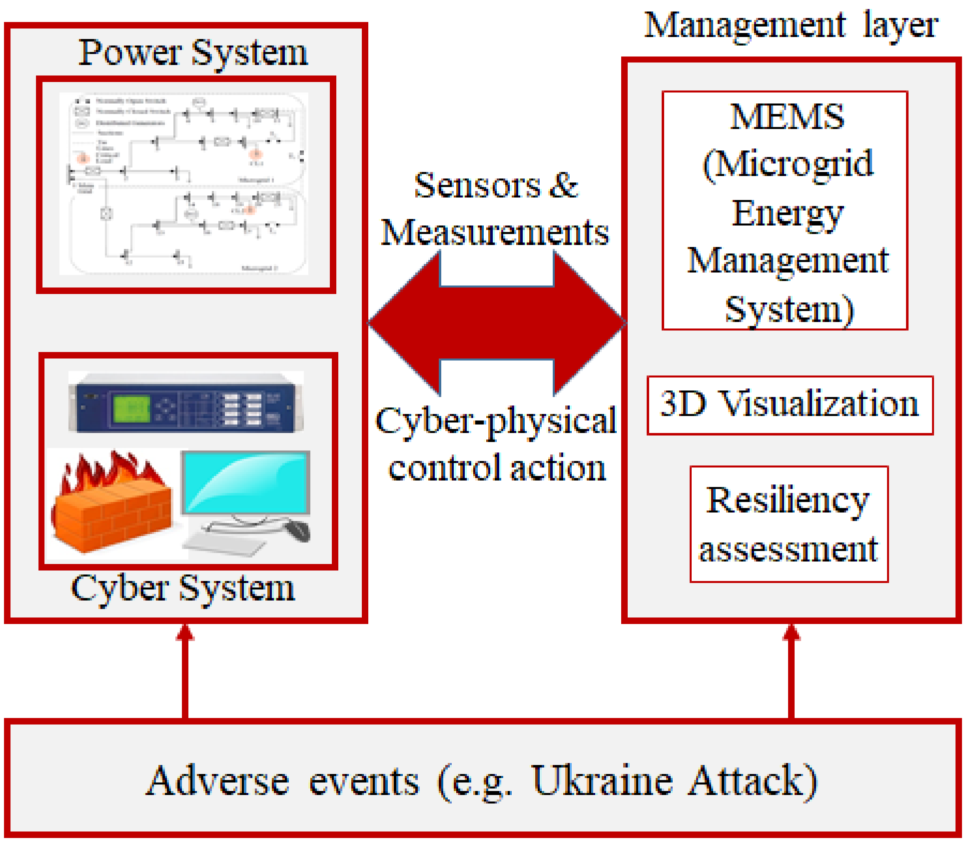
- PMU 2 and 3 show highest MSE among all PMUs
- it can be determined that most probably the fault could have occurred in the line from bus 2 and 3

- What is resiliency? How do we measure resiliency?
- How PMU data analytics enable resiliency?
- Use Case I: PMU based Anomaly/ Event Detection
- Use Case II: PMU based Failure Diagnosis
- **Use Case III: Data-driven Resiliency Analysis**
- Summary and Moving Forward

Cyber-Physical Modeling and Visualization for Microgrid Resiliency (S-82)

- Create *accurate models* of physical and cyber microgrid and interface them to obtain holistic cyber-physical system (CPS) model
- Demonstrate cyber-physical resiliency metrics and performance of microgrid with adverse events
- Develop *a 3D visualization framework* for enhanced situational awareness

CPS MODEL



- **Model of microgrid based on Miramer microgrid in *OpenDSS*, power simulator**
- **Cyber/ communication model of microgrid in *Mininet*, a**

Tools

Power System

- Real-Time simulation tools including RTDS, OPAL-RT
- Offline simulation tools including steady state and dynamic tools

Communication System

- Simulation tools such as NS-3, Mininet
- Emulation tools such as CORE, DeterLab

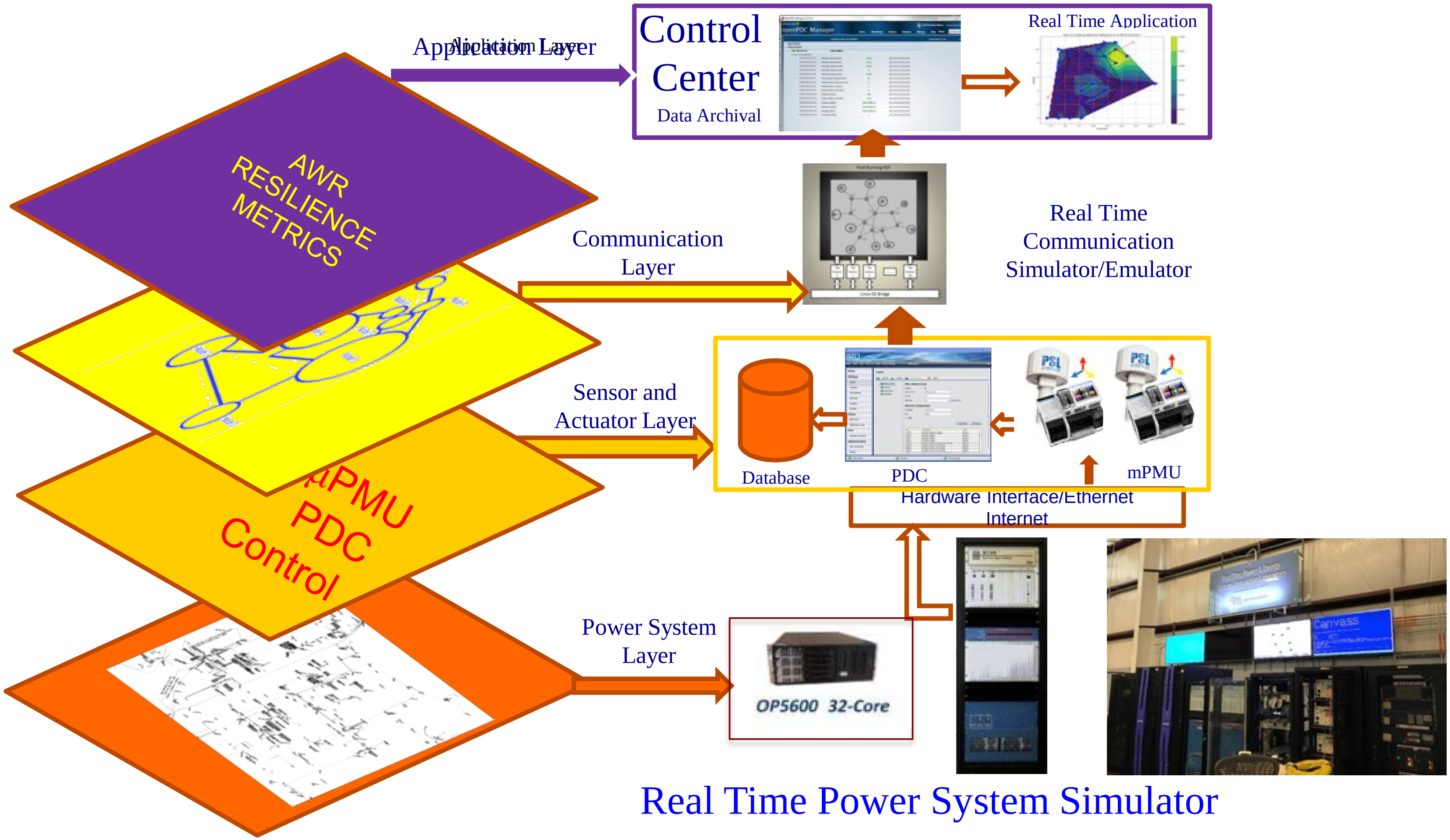
Security Tools

- Cyber-Attack tools and implementations
- Defense and visualization tools such as IDS systems

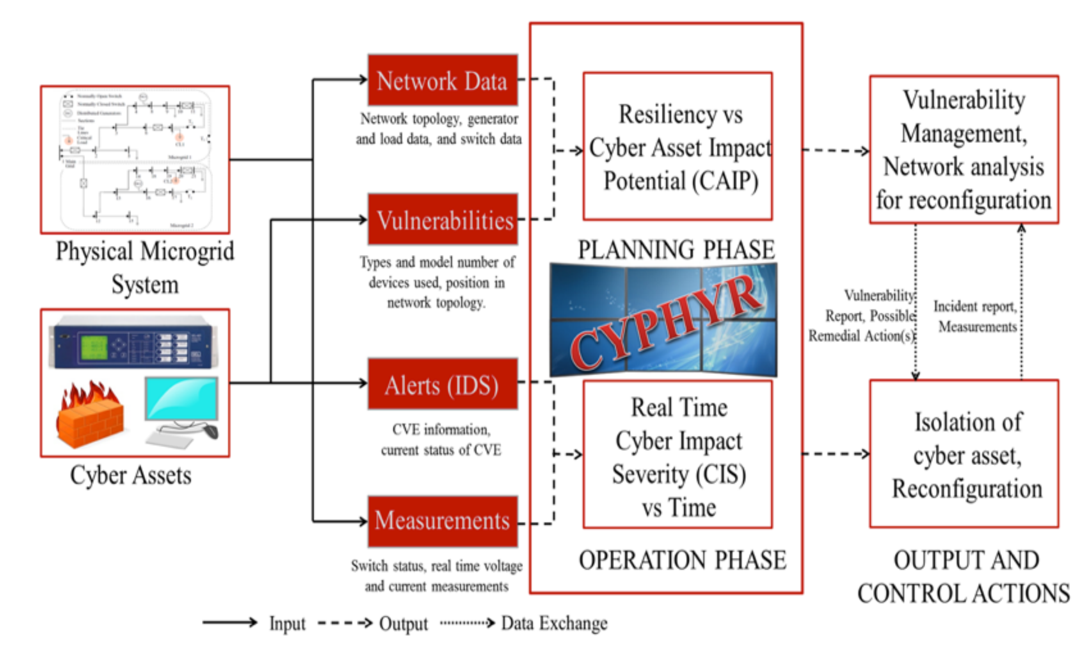
IPC, TCP/IP, Remote Encapsulation

Proxy interface, TCP/IP

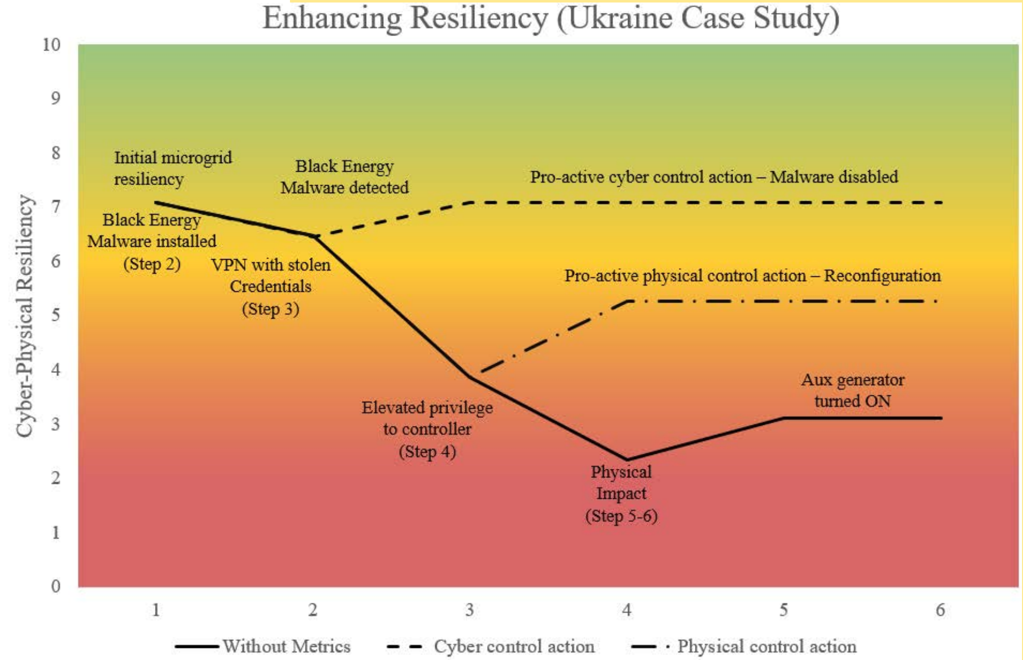
Test Environment



Real Time Power System Simulator



CyPhyR: Cyber-Physical Resiliency Tool



- What is resiliency? How do we measure resiliency?
- How PMU data analytics enable resiliency?
- Use Case I: PMU based Anomaly/ Event Detection
- Use Case II: PMU based Failure Diagnosis
- Use Case III: Data-driven Resiliency Analysis
- **Summary and Moving Forward**

Takeaway #1: Resiliency is a Complex Problem

Power Grid Resiliency = **Flexible Infrastructure** × **Secure Cyber Infrastructure** × **Resilient Power Control Applications**

Flexible Infrastructure

Multiple switch
Macrogrid
Minigrid
Microgrid
Nanogrid
Graceful disintegration and interconnection
Flexible management and control of resources
Economic and market incentive

Secure Cyber Infrastructure

Communication
Authentication
Encryption
Computation
Access control
Attestation
Forensics
Patch management
Software Audits
System Management
Intrusion Detection
Event Monitoring/Analytics
Security Assessment

Resilient Power Control Applications

Generation
Automatic Generation Control
Governor Control
Automatic Voltage Regulation
Protection
Transmission
State Estimation
VAR Compensation
Protection
Distribution
Load Shedding
Protection
Advanced Metering Infrastructures

- Resiliency metric is a MCDM problem
- Resiliency is characteristics of the system

Takeaway #2: Finding Match in Data Analytics Techniques and Power System Problems is VIT

- Data Analytics and machine learning approaches needs to be applied after analyzing the power system problem carefully. Finding match between machine learning strength and power system problem to be solved is important.
- Machine learning is only applicable in data-rich problems if no system model is available (e.g. forecasting)
- If model is available with rich data set, typically it will be two step approach: apply machine learning to narrow down your possible options and refine it with model based approach (e.g. event detection)
- Machine learning will not give a good results based on state of the art for highly complex and dynamic problems (e.g. transient stability, contingency analysis).
- Validation and metric is important for these evolving solution technologies

Takeaway#3: Get Involved in PMU Data Analytics and Applications

NASPI White Paper on Data Quality Requirements for PMU based Control Applications

IEEE Synchrophasor based Power Grid Operation as part of Bulk Power System Operation. White paper on a) Challenges and Solutions in Implementing PMU based Applications in Control Center) and b) Quality-Aware Applications

https://sgdril.eecs.wsu.edu/workshop_conferences/real-time-data-analytics-for-the-resilient-electric-grid/

Thank You

Acknowledgement: PSERC, DOE, NSF, INL

Anurag K. Srivastava
anurag.k.srivastava@wsu.edu