Sensor Data Analytics for Intrusion Detection

Tech Tesfay,

Prof. Anna Scaglione

Outline

- **Motivation** for cyber-physical intrusion detection
- Reconnaissance activity identification using:
  - **Fog computing:** at the network age using Thévenin source impedance
  - **Cloud computing:** at the control centre using data from multiple input sources
- **Grand vision:** automated threat detection by leveraging data from other sources

Motivation

"Whatever can go wrong, will go wrong" Murphy's law.

- There will be security breaches no matter how much protection is put in place.
- Even worse, most utilities have not put security in place.
- **Example**: Ukraine power grid attack, Stuxnet malware, US power grid breach report ....
- Attacks - system diverges from the *safe operating limits* [1].

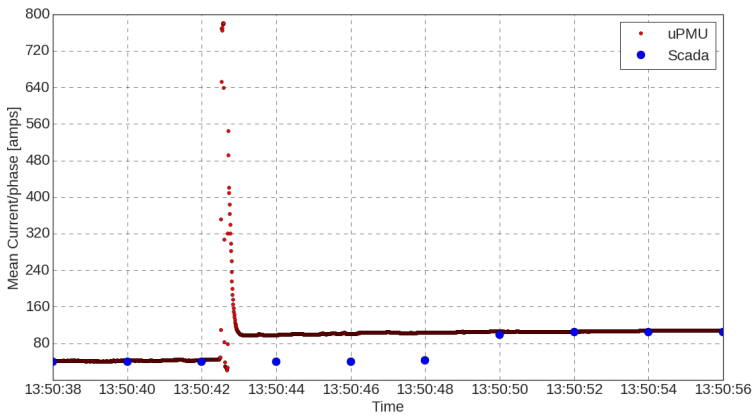
Solution: Put (additional) security measures to counter such attacks?

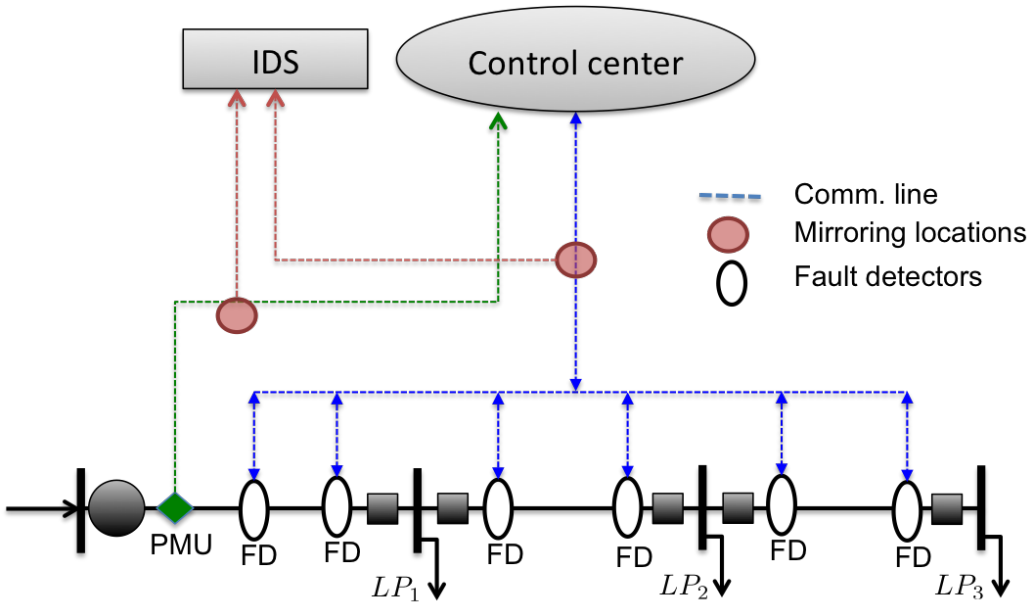
[1] **cardenas2011attacks**

Our approach

Use **intrusion detection** to detect malicious activities:

- Leverage knowledge of the physical laws governing the safe operating limits.
- Use high resolution (μPMU) measurements.
- Use mirrored SCADA packet.





Not so easy to accomplish!

Intrusion detection is a challenging task given the following challenges:

Challenges

- Insufficient number of μPMUs (lack of full system observability),
- Need for real-time analysis (latency of centralized analytics),
- Inaccuracy of the grid parameters in the database (time-varying/human errors)
- Designing appropriate rules to correlate data from different sources and output the correct security status of the grid

Hierarchical intrusion detection architecture

Fog computing

- data analysis at the network edge (local rules),
- near real-time analysis (1 sec in our case),
- prioritizing communication of eventful segments,

Cloud computing

- co-analysis of data from multiple sensors (central rules)
- event localization and categorization (natural vs malicious anomaly)

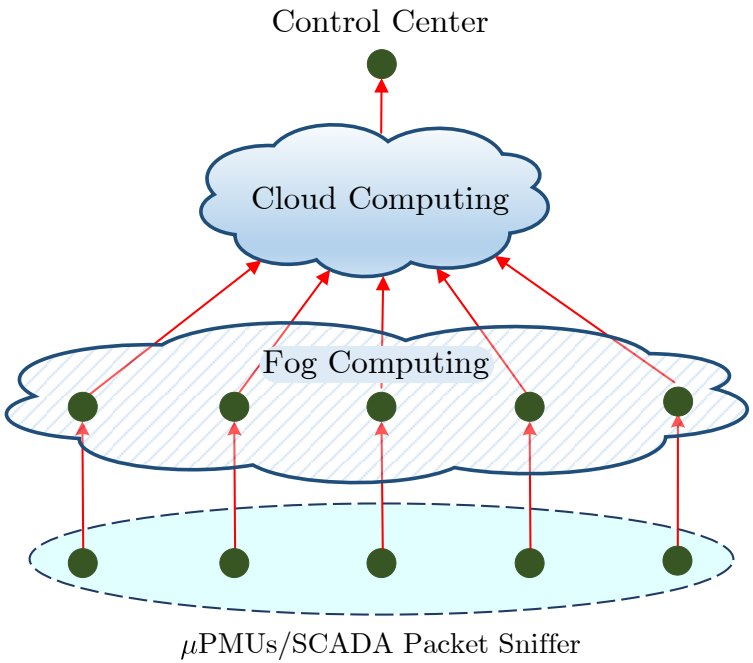


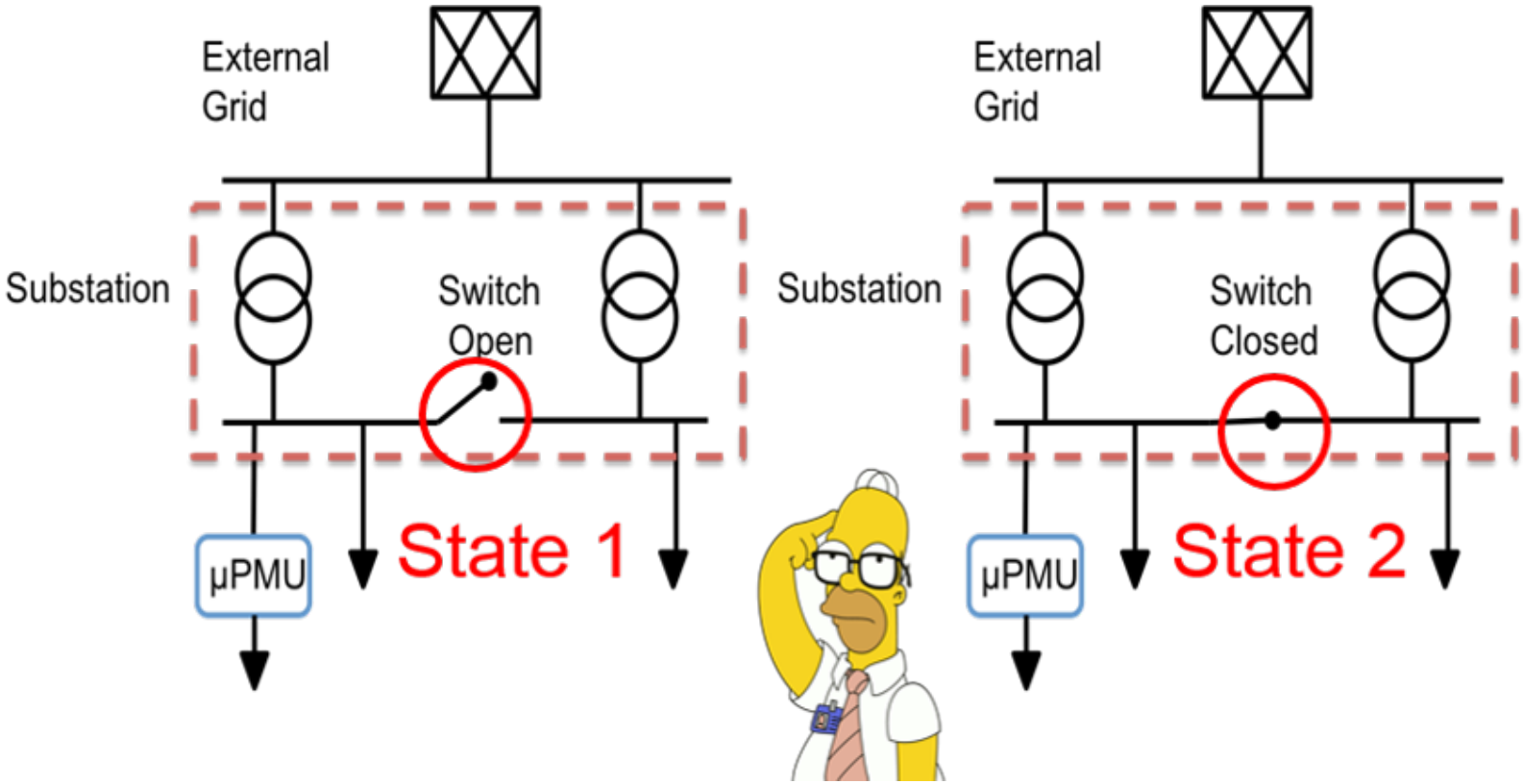
Figure: Intrusion Detection Architecture.

[1] **jamei2016micro**

Reconnaissance activity identification using fog computing

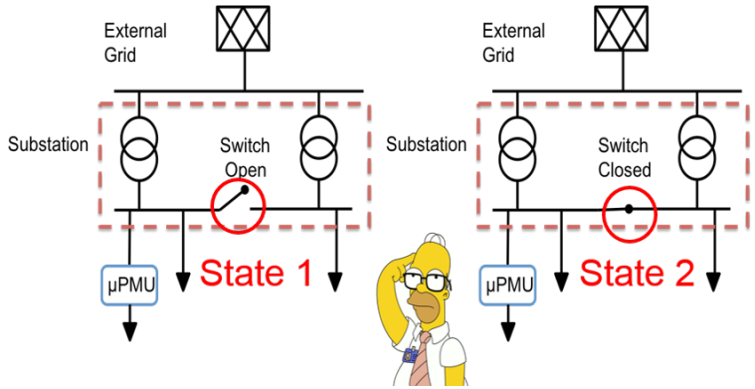
Case study: Reconnaissance through Thévenin estimation

Normally-open switch at a substation is a good point for attacker to gauge its controllability over SCADA network by toggling the switch status.



Can we identify this specific reconnaissance activity?

Case study: Reconnaissance through Thévenin estimation



Insight

- The Thévenin impedance upstream seen from a distribution substation is dominated by the transformer impedance
- **Implication:** The upstream Thévenin impedance for "closed-switch" is almost half of the value when the switch is "open"

Goal

- Online Tracking of Thévenin source impedance using after transformer substation μPMU data

Related work on Thévenin estimation:

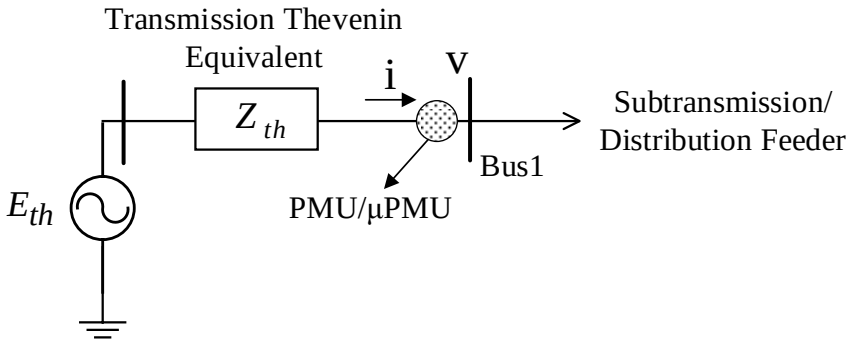
- Least-square methods [1,2,3,4,5].
- Addressing the quasi steady-state adverse effect on Thévenin estimation[6,7].
- Thévenin circuit estimation in a three-phase unbalanced distribution grid [8] using RMS volatge values.

[1] **vu1999use**
[2] **smon2006local**
[3] **tsai2008line**
[4] **parniani2006voltage**
[5] **arefifar2009online**
[6] **abdelkader2012online**
[7] **alinejadline**
[8] **hart1986characterising**

Our Contributions

- First to utilize Thévenin parameter for reconnaissance activity identification
- Online estimation of Thévenin parameters in a balanced/unbalanced grid,
- Proposing a robust algorithm for non-stationary and correlated data.
- Removing the inaccurate common assumption of constant Thévenin voltage angle over a short window

Thévenin Equivalent Circuit



$$\mathbf{v}[k] = \mathbf{E}_{th}[k] - \mathbf{Z}_{th}[k]\mathbf{i}[k] \quad (1)$$

Figure: Transmission Grid Thévenin Equivalent Seen from Substation

In the sequence domain, assuming transposed lines in the transmission level:

$$\begin{bmatrix} v_0[k] \\ v_1[k] \\ v_2[k] \end{bmatrix} = \begin{bmatrix} 0 \\ E_1[k] \\ 0 \end{bmatrix} - \begin{bmatrix} Z_0[k] & 0 & 0 \\ 0 & Z_1[k] & 0 \\ 0 & 0 & Z_2[k] \end{bmatrix} \begin{bmatrix} i_0[k] \\ i_1[k] \\ i_2[k] \end{bmatrix}. \quad (2)$$

Estimation: Unbalanced Grid

- Taking advantage of unbalanced data
- Assuming $Z_1[k] \approx Z_2[k]$:

$$Z_0[k] = -\frac{v_0[k]}{i_0[k]}, \quad Z_1[k] = -\frac{v_2[k]}{i_2[k]}, \quad E_1[k] = \frac{v_1[k]i_2[k] - v_2[k]i_1[k]}{i_2[k]}. \quad (3)$$

- Estimation at each instant of time only depends on the measurements of that time-instant.

Estimation: Balanced Grid

The only non-trivial equation is:

$$v_1[k] = E_1[k] - Z_1[k] i_1[k] \tag{4}$$

Assumption: The resistive part of the Thévenin impedance is negligible compared to the inductive part.

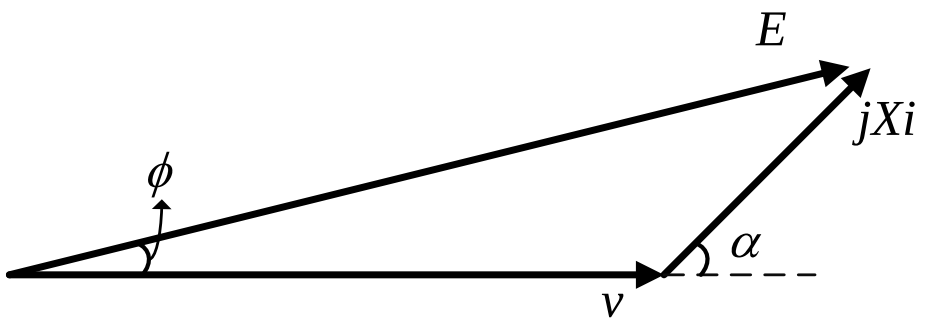


Figure: Phasor Diagram of the Equivalent Thévenin Circuit for Balanced Grid.

Let $A[k] = |E[k]|$ and i_{im} be the imaginary component of the current, then we have:

$$\underbrace{A^2[k] - v^2[k] - X^2[k]|i[k]|^2 + 2 i_{im}[k] X[k] v[k]}_{r(A,X;k)} = 0 \tag{5}$$

Estimation: Balanced Grid

we form the following M over-determined homogeneous set of quad. equations:

$$\underbrace{\begin{pmatrix} r(A,X;k-M+1) \\ r(A,X;k-M+2) \\ \vdots \\ r(A,X;k) \end{pmatrix}}_{\vec{\mathbf{r}}(A,X;k)} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{4}$$

Minimize the squared-norm of the vector $\vec{\mathbf{r}}(A,X;k)$:

$$\min_{\boldsymbol{\theta}[k]} f(A,X;k) = \frac{1}{2}||\vec{\mathbf{r}}(A,X;k)||^2 \tag{5}$$

where $\boldsymbol{\theta}[k] = [A[k], X[k]]^T$.

Estimation: Balanced Grid

Advantages of our formulation:

- Having the assumption of constant Thévenin voltage phase angle over a window of M samples is not needed
- Reporting phasor angles relative to the voltage phasor angle removes the effect of off-nominal frequency

Estimation: Balanced Grid

The *Levenberg-Marquardt Algorithm* (LMA) is used to solve the non-linear least square problem.

Advantages:

1. handling close to rank-deficient matrices,
2. better performance compared to Gauss-Newton for a bad initial guess.

LMA

Algorithm 1: LMA at time instant k

Input: $\vec{\mathbf{r}}(A, X; k)$, and an initial guess $\boldsymbol{\theta}_0[k]$
Output: Thévenin parameters at time k
begin
  flag $\leftarrow 1$;
  initialize $\rho < 1$, λ, and ϵ ;
  $\boldsymbol{\theta}[k] \leftarrow \boldsymbol{\theta}_0[k]$;
  while *flag=1* **do**
    $\mathbf{J} = \nabla \vec{\mathbf{r}}(\boldsymbol{\theta}; k)$;
    $\mathbf{P}_{LM} \leftarrow -(\mathbf{J}^T\mathbf{J} + \lambda \text{diag}(\mathbf{J}^T\mathbf{J}))^{-1}\mathbf{J}^T\vec{\mathbf{r}}(\boldsymbol{\theta}; k)$;
    $\boldsymbol{\theta}_{new}[k] \leftarrow \boldsymbol{\theta}[k] + \mathbf{P}_{LM}$;
    if $f(\boldsymbol{\theta}_{new}; k) < f(\boldsymbol{\theta}; k)$ **then**
      $\lambda \leftarrow \rho\lambda$;
      $\boldsymbol{\theta}[k] \leftarrow \boldsymbol{\theta}_{new}[k]$;
    else
      $\lambda \leftarrow \frac{\lambda}{\rho}$;
    if $f(\boldsymbol{\theta}; k) < \epsilon$ **then**
      flag $\leftarrow 0$;
  $\phi[k] \leftarrow \sin^{-1}(X[k]i_r[k]/A[k])$
  return $E[k]$, $X[k]$;

Numerical Results: Unbalanced Grid

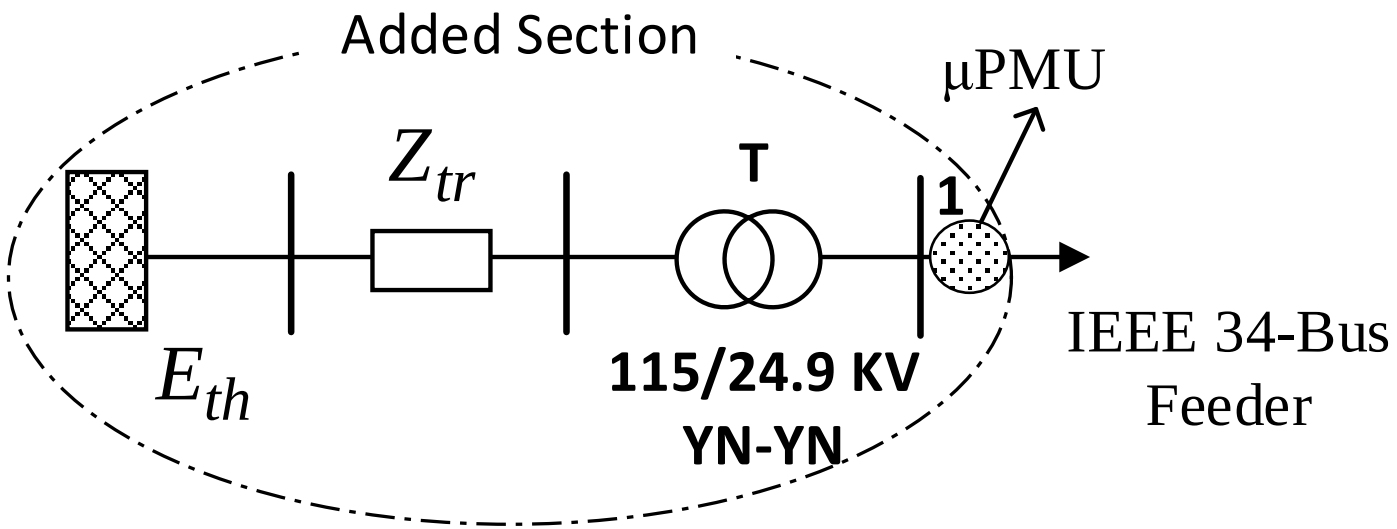


Figure: Modified IEEE-34 Bus Test Case

Numerical Results: Unbalanced Grid

	Estimated	Actual
Z_0	$2.5533 + j9.4392$	$2.5716 + j9.4320$
Z_1	$2.9922 + j10.92$	$2.99 + j10.8901$

Numerical Results: Balanced Grid

New-England test-case, load ramp event of +2%/sec at load bus 16.

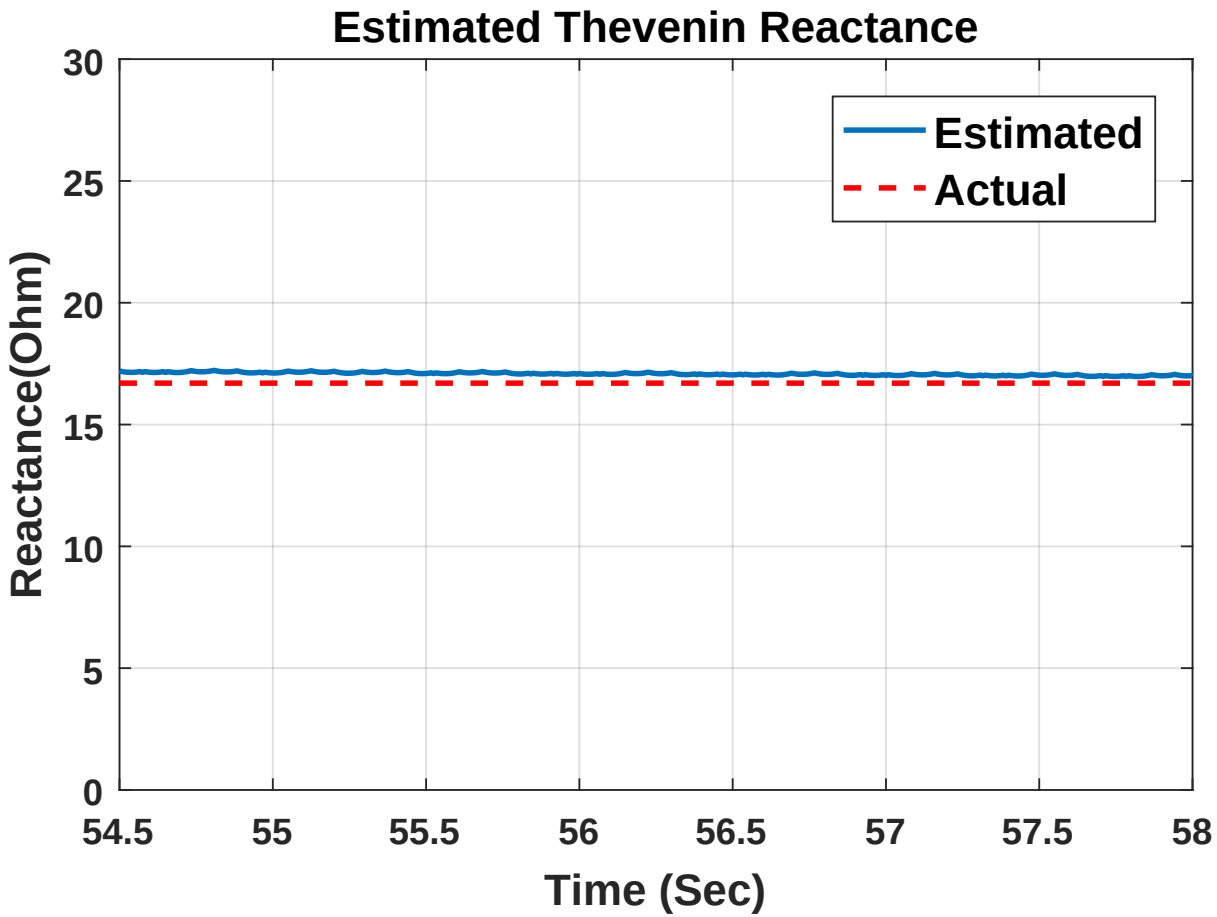


Figure: Estimated Thévenin Reactance Seen from Bus 16 of New England Test Case Using LMA Method.

Application : Reconnaissance Activity Identification

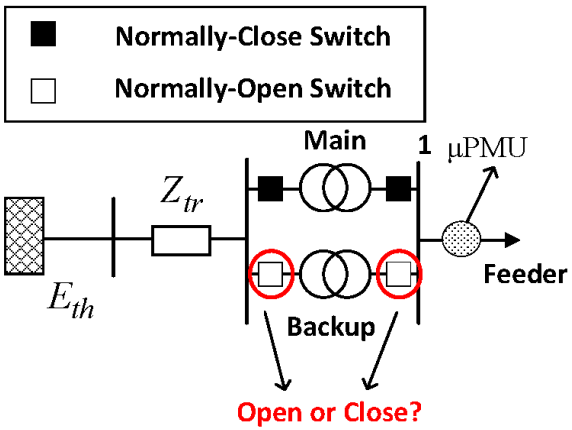


Figure: Substation Main-Spare Transformer Setup

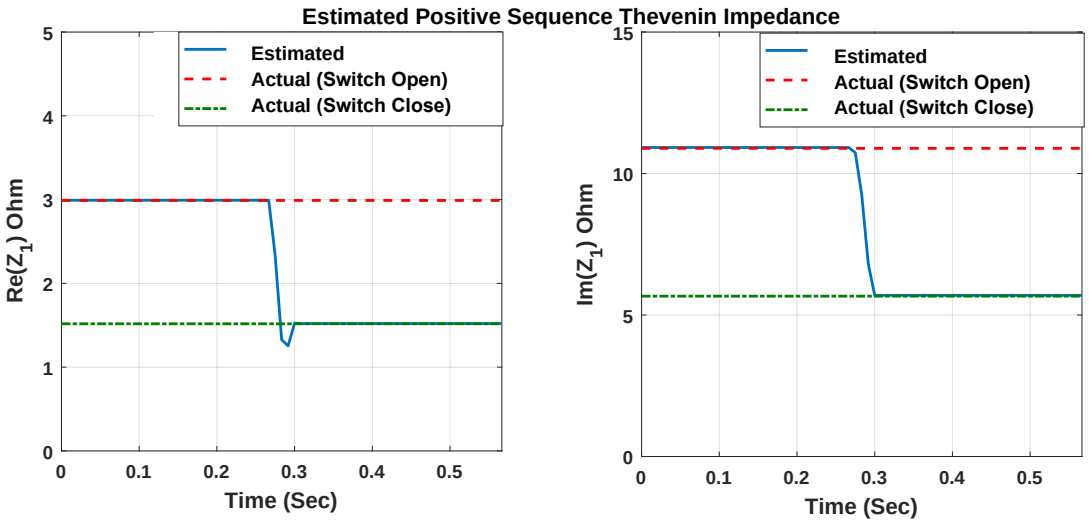


Figure: Estimated Thévenin Source Impedance

Reconnaissance activity identification using cloud computing

Steps:

- Analyse event using data from multiple μPMU
- Integrate SCADA data in the analysis for event categorization (natural vs malicious anomaly).

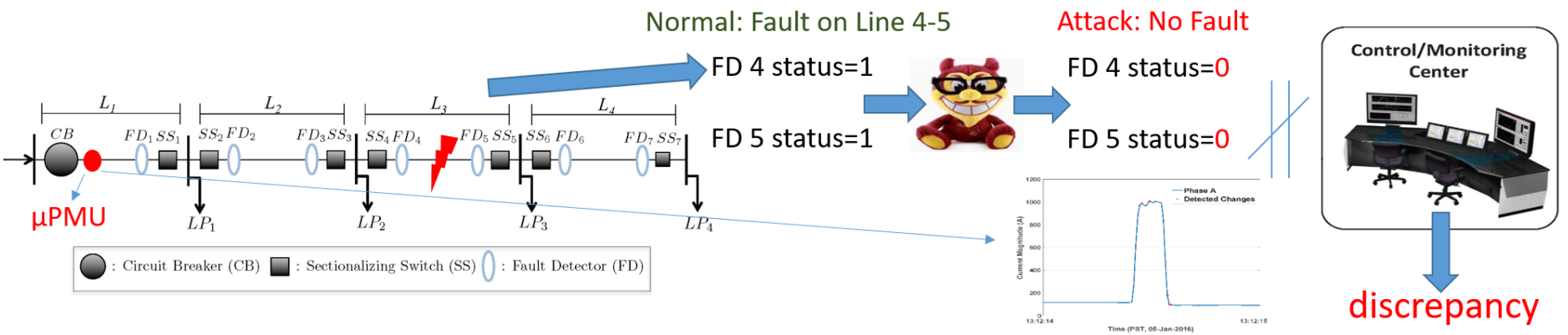
Case study: Reconnaissance identification on FLISR

Goal:

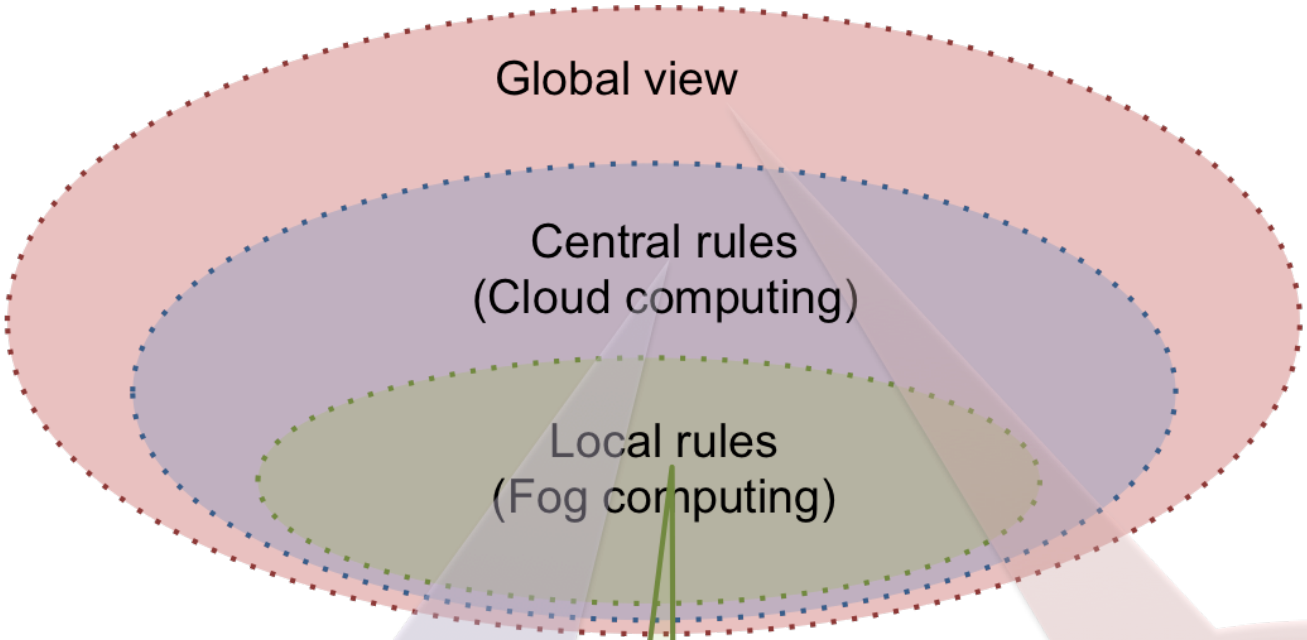
- Run Fault Location, Isolation, and Service Restoration (FLISR) algorithm using μPMU measurements.
- Detect malicious activities through identification of inconsistencies between μPMU data analytics and SCADA data.

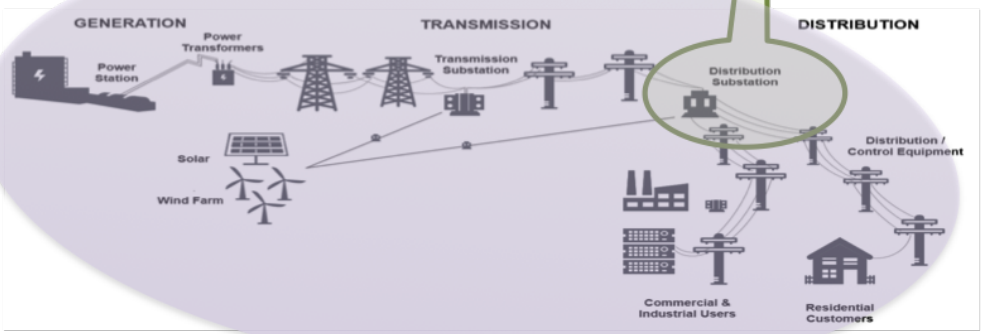
Example: FLISR

- μPMU data analytics identifies fault on line $4-5$.
- Data injection attacks on SCADA data
  - Blocking fault detector packets,
  - Altering detected fault location,
  - Jamming or altering the isolation commands.
- SCADA information inconsistent with μPMU data analysis can inform of ongoing attack.



Vision: Expanding the input scope





Automated security information and event management (SIEM) system with more input data.

The need for automation

Manual operations to secure a power grid network is challenging

- Too many legacy systems with no security features,
- Longer life span and difficult to keep track of security patches,
- Too many type of attacks to sufficiently identify and prepare for each attack,
- Human factor - weakest link!

Automated threat detection and identification is key!

Continuous security

"If you know the **enemy** and know **yourself**, you need not fear the result of a hundred battles." Sun Tzu, The art of war

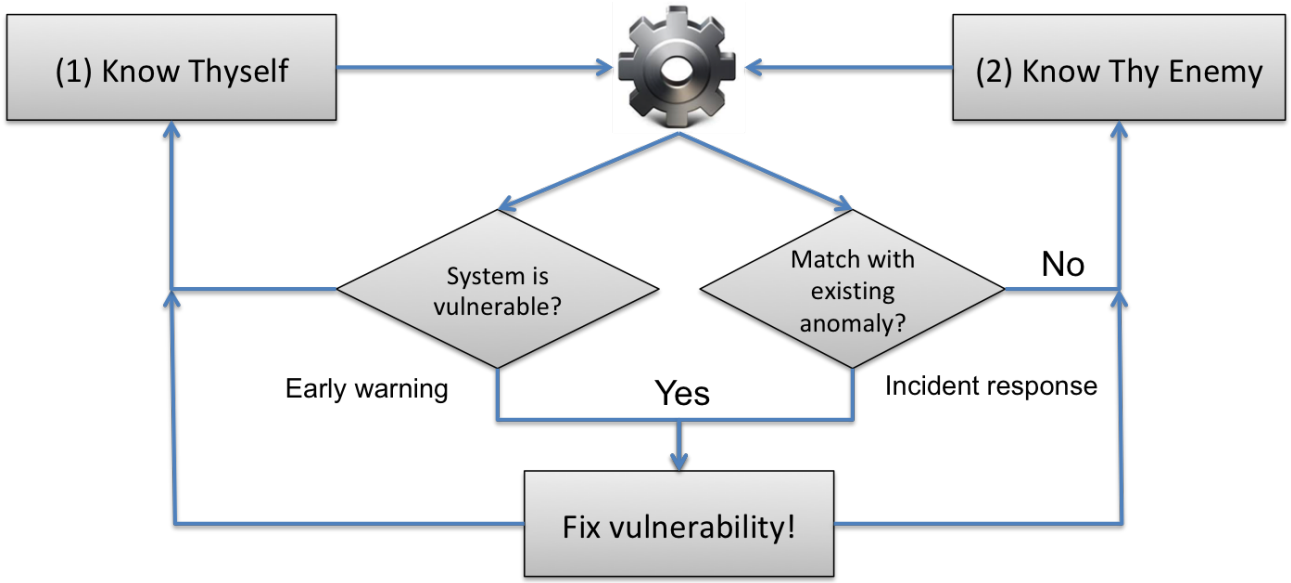


Figure: Automated security information and event management (SIEM)

Comments and/or possible collaborations are welcome!

Related Publications

- Jamei, Mahdi, Emma Stewart, Sean Peisert, Anna Scaglione, Chuck McParland, Ciaran Roberts, and Alex McEachern. "Micro Synchrophasor-Based Intrusion Detection in Automated Distribution Systems: Toward Critical Infrastructure Security." IEEE Internet Computing 20, no. 5 (2016): 18-27.
- Jamei, Mahdi, Anna Scaglione, Ciaran Roberts, Emma Stewart, Sean Peisert, Chuck McParland, and Alex McEachern. "Anomaly Detection Using Optimally-Placed Micro-PMU Sensors in Distribution Grids." IEEE Transactions on Power System (2017).
- Jamei, Mahdi, Anna Scaglione, Ciaran Roberts, Alex McEachern, Sean Peisert, Emma Stewart, and Chuck McParland. "Online Thevenin Parameter Tracking Using Synchrophasor Data." Accepted in IEEE PES GM 2017.
- Jamei, Mahdi, Anna Scaglione, Ciaran Roberts, Emma Stewart, Sean Peisert, Chuck McParland, and Alex McEachern. "Automated Anomaly Detection in Distribution Grids Using uPMU Measurements." In Proceedings of the 50th Hawaii International Conference on System Sciences. 2017.

Thank You!
Questions?