THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Noise Matters: Using Sensor and Process Noise Fingerprint to Detect Stealthy Cyber Attacks and Authenticate sensors in CPS

Chuadhry Mujeeb Ahmed, et al.

https://doi.org/10.1145/3274694.3274748

Presented by Ellis Thompson

Overview – *What I aim to cover?*

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

- What is the concept?
  - Use hardware noise characteristics to create unique patterns for a sensor and then classify the resultant data
- How is this done?
  - Leveraging noise of a sensor output vs expected output
- What is the proposed solution?
  - A support vector machine (SVM) technique to classify signals as to either belong to a sensor or not
- How was it tested?
  - On a plethora of sensors in a SWaT testbed

Concept – *Signals are noisy*

The resultant signals sent by a sensor are inherently noisy.

This noise can be a result of:

- Electrical noise in transmission
- Electrical noise from DC offset
- Frequency noise
- Variations in Manufacturing
- Temporal noise
- Readout noise
- Spatial noise
- Offset noise

Patterns *are* unique to sensors/setups, not always possible to identify source but overarching result is pattern dependent

Concept – *Architecture: Model*

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

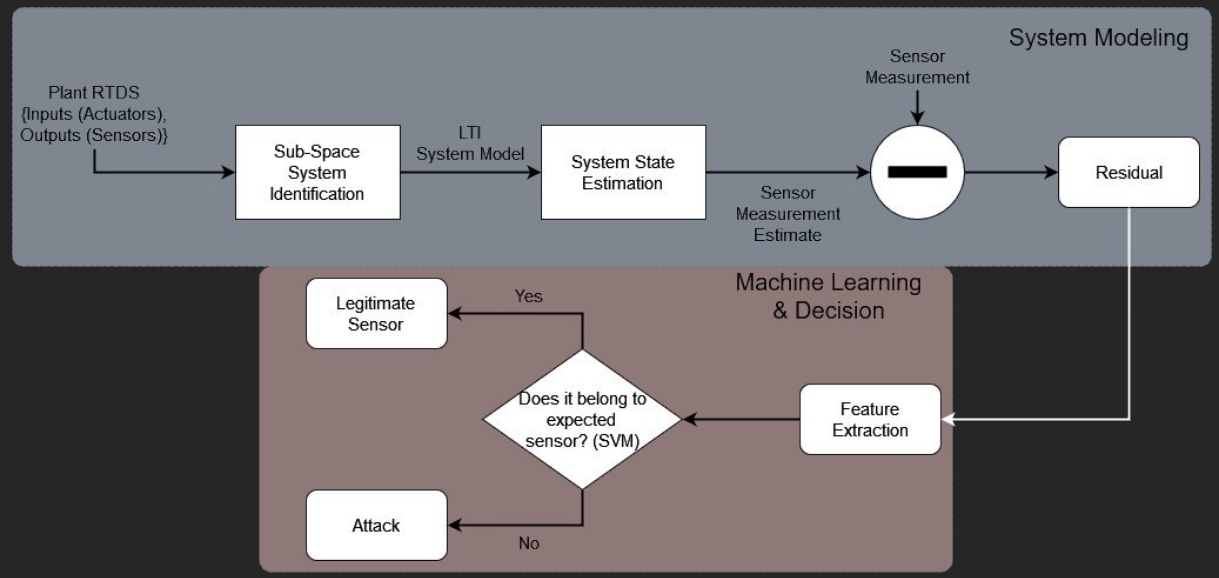
- Comparing the residual expected outcome from a model to the real-world output through the use of an SVM.

Attacks are a Man-in-the-middle, sensor spoofing attacks in the form

$$\bar{y}_k := y_k + \delta_k = (Cx_k + \eta_k) + \delta_k$$

If x_k is the system state and η_k is genuine noise.

Let y_k be the true sensor part and $\bar{y}_k$ be a constructive false part, δ_k becomes the attack vector



Concept – *Architecture: Model (Residual Part)*

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

- The residual is described as $r_k \coloneqq \bar{y}_k - \hat{y}_k$ (output - predicted)

Yielding the vector:

$$r_k = C\left\{\sum_{i=0}^{k-2}(A-LC)^i(v_{k-i-1} - L\eta_{k-i-1})\right\} + \eta_k$$

Where A, C are state spaces of the model, L is the gain matrix, v is a control input

If you haven't guessed yet: η, the noise, is obtained as the *fingerprint* this is known from observation of the system.

Concept – *Attacker Model*

There are some notes on the attacker:

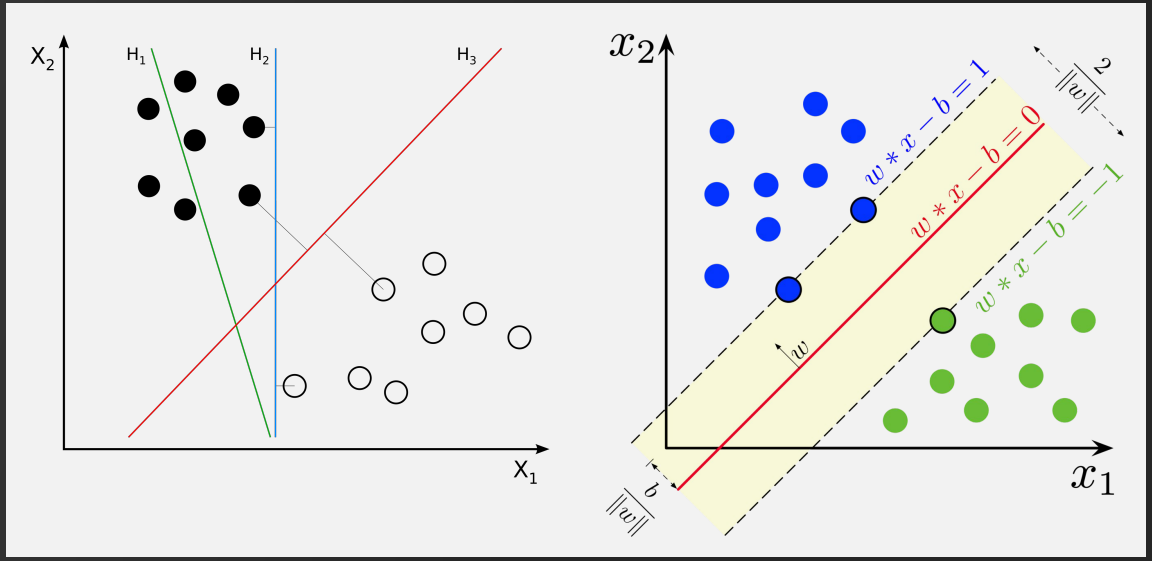
- Assumed attackers already have access to $y_{k,i} = C_i x_k + \eta_{k,i}$
- Assumed the attacker knows the system dynamics
- Replay attack is not considered (as sensor noise is preserved)

2 Types of attacks are then considered:

- Generic spoofing attacks – Attacker arbitrarily applies some vectors
- Stealthy attacks – Attacker samples from noise

Solution – SVM

One method for classification (and regression) problems



Non-probabilistic bilinear classifier
Maximum-margin hyperplane (its p-dimensional)

Solution – SVM (features)

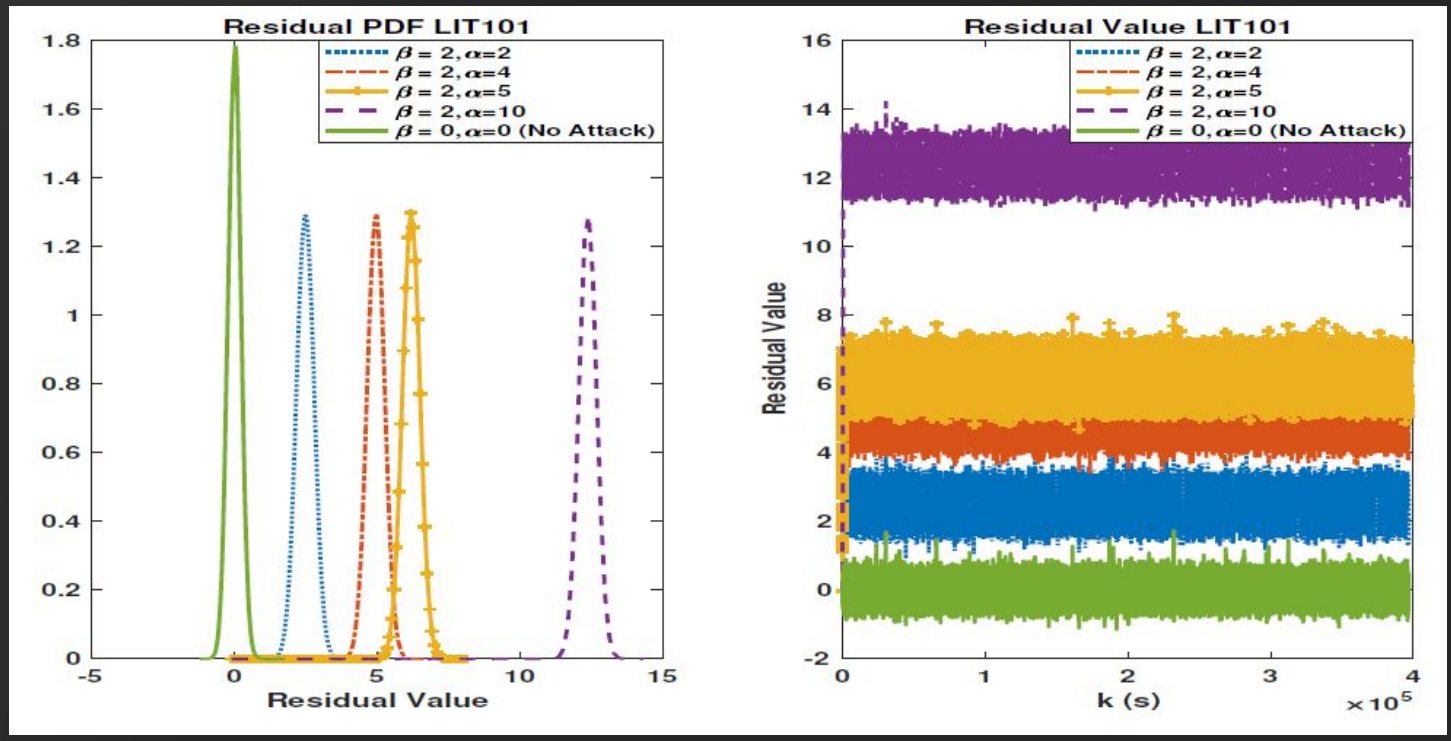
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

- Mean
- Standard Deviation
- Mean Average Deviation
- Skewness (measure of symmetry)
- Kurtosis (measure of *tailedness/how peaked or flat a distribution is*)
- Spectral Standard Deviation (based on frequency characteristics)
- Spectral Centroid (based on frequency characteristics)
- DC component (DC noise)

Preliminary Results: Eval of Residual

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

During an *arbitrary* attack the residual value deviates significantly with small changes.

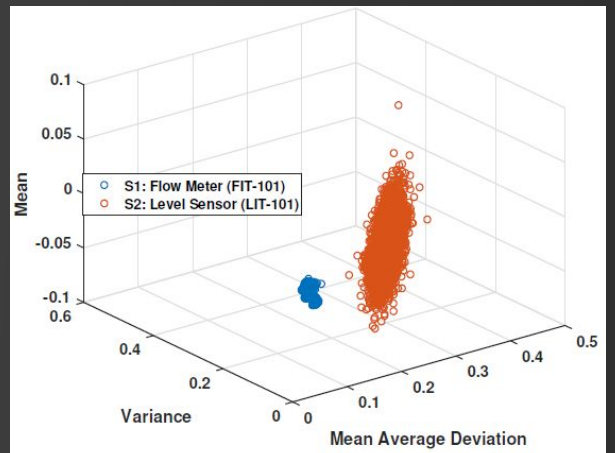


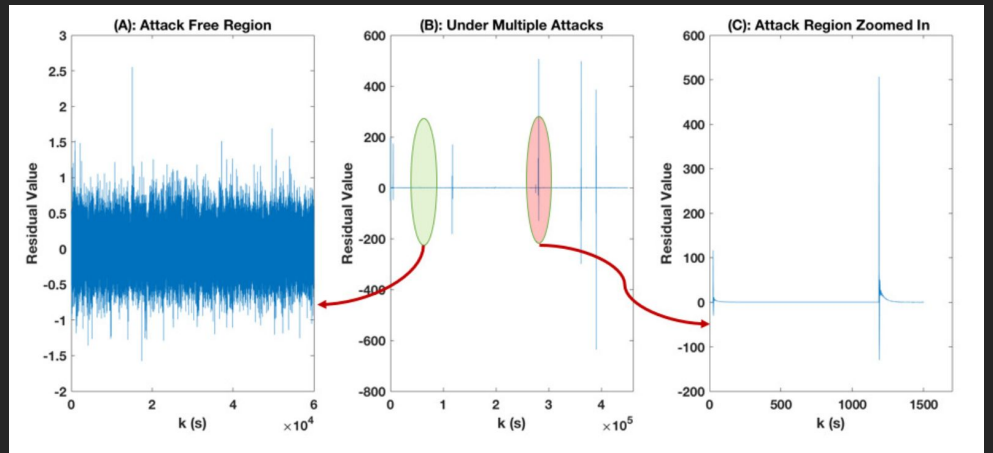
Results: SWaT

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

State-of-the-art Water Treatment testbed

- RQ1: Proof of Fingerprint – Clear that a fingerprint exists
- RQ2: Attack Detection Delay – 120 samples (2 minutes) of data achieves 98% accuracy, 60 samples achieves 95%
- RQ3: How does train/test data size effect identification – Sample sizes of 2-15 had little variance i.e. the approach is robust





Results: SWaT – RQ4

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

RQ4: How well does it actually perform?

Fairly well (TPR/NR = True Positive/Negative Rate)

One class(OC) out performs Multi Class (MC)

Sensor	Atk. seq. [a]	Attacked [b]	Detected [c]	MC-SVM TNR	MC-SVM TPR	OC-SVM TNR	OC-SVM TPR
DPIT-301	8	8	5	99.65%	62.5%	86.3%	88.88%
LIT-101	3,21,30,33,36	27	24	97.88%	88.88%	89.4%	93.54%
FIT-101	None	27	22	99.49%	81.48%	94.2%	80.64%
LIT-301	7,16,26,32,41	37	29	91.41%	78.37%	88.7%	80.95%
FIT-301	None	37	22	91.55%	59.45%	88.85%	78.57%
LIT-401	25,27,31	35	20	92.09%	57.14%	89.5%	77.5%
FIT-401	10,11,39,40	12	8	99.86%	66.66%	91.6%	73.3%

Closing – Questions & Discussion Points

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

First: **Any questions?**

Second: **Some points to discuss**:

- Do we think vehicular sensors could also produce distinct noise? How would camera noise differ?
- SWaT has a pretty slow update interval (~1second), would we be able to detect attacks quicker with a vehicles faster update interval?
- This was tested on water sensors, could this scale to vehicles/other applications?