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

Authors : Chuadhry Mujeeb Ahmed, Jianying Zhou, Aditya P. Mathur

Conference: ACSAC '18: 2018 Annual Computer Security Applications Conference

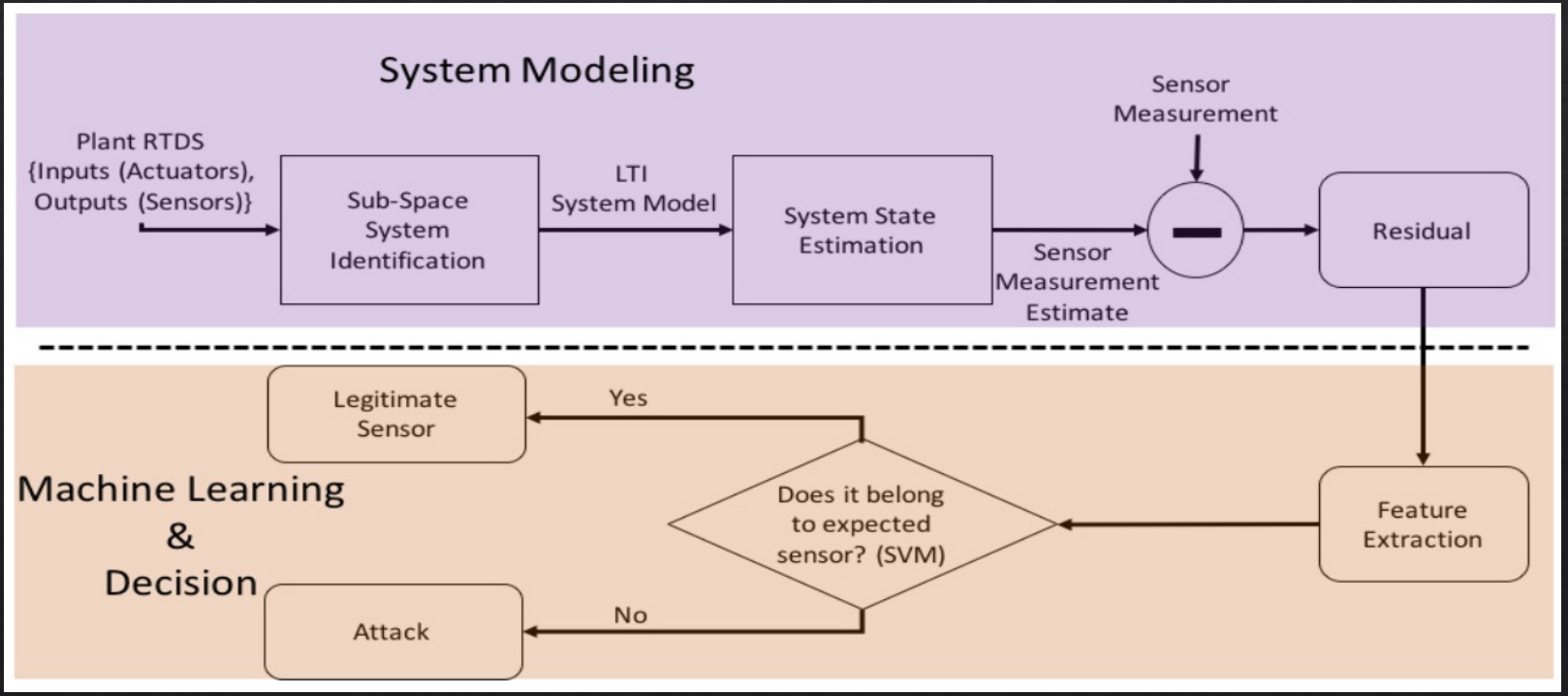
Shahab Nikkhoo

What is the problem?

- CPS devices has several vulnerabilities
- Sensors in physical layer
- Adding more sensors

How we can solve it

- Some how find a characteristic of a sensor
- Use it as a fingerprint
- The characteristic must be unique



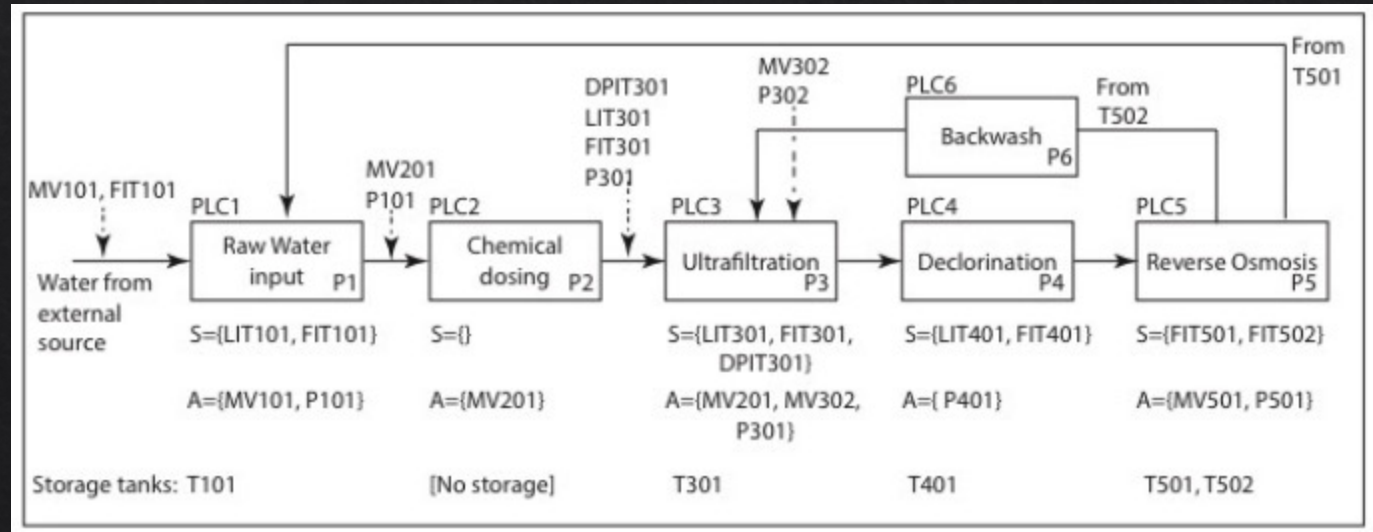
Sensors

- They are sensing the world
- We are always dealing with noise
  - Manufacture imperfection
  - Environment
- How can noise be extracted?
  - We need to estimate the true value

- Why not using cryptographic solutions?

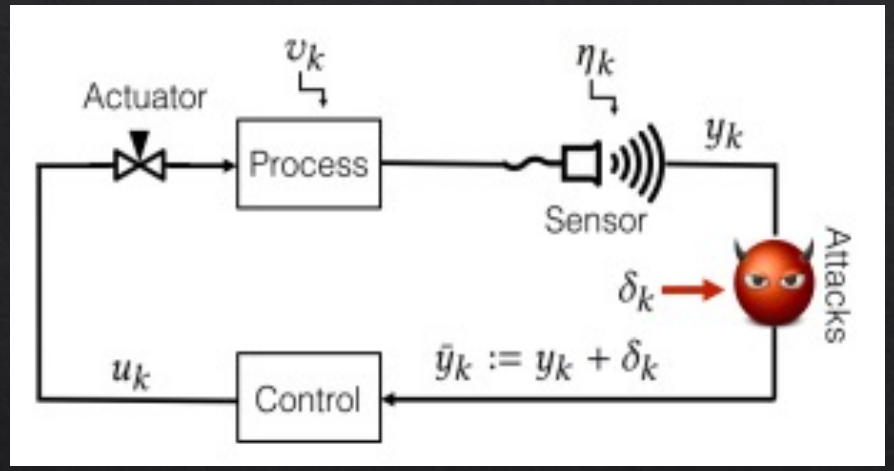
Assumptions and Testbed

- Secure Water Treatment Testbed (SWaT)
  - Data collected for seven days of normal operation
  - Four days with several attacks
  - 18 sensor
- Attacks started after certain amount of time
- Attacker has access to model and sensor data



System modeling

- We have dynamic system



$$\begin{cases} x_{k+1} = Ax_k + Bu_k + v_k, \\ y_k = Cx_k + \eta_k. \end{cases}$$

System modeling

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

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + L_k\left(\bar{y}_k - C\hat{x}_k\right)$$

$$e_k := x_k - \hat{x}_k$$

$$P_k := E[e_k e_k^T]$$

$$\lim_{k\to\infty} P_k = P$$

$$L_k = L := \left(APC^T\right)\left(R_2 + CPC^T\right)^{-1}$$

$$\delta_k = \beta[Ce_k + \eta_k] + \alpha$$

System modeling

- What is residual?

$$r_k := \bar{y}_k - \hat{y}_k$$

- Validation

Sensor	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
RMSE	0.0363	0.2867	0.0346	0.0113	0.0520	0.0313	0.2561	0.0200	0.0612	0.2267	0.0014	0.0096	0.0670	0.0082	0.0267	0.0037	0.0595	0.0035
(1-RMSE)*100%	96.3670	71.3273	96.5409	98.8675	94.8009	96.8656	74.3869	98.0032	93.8757	77.3296	99.8593	99.0377	93.3031	99.1821	97.3313	99.6251	94.0537	99.6501

Machine learning

- Features
  - Time domain
  - Frequency domain

- SVM is used as 2-class classifier and 1-class classifier
- Training took seven real-time days

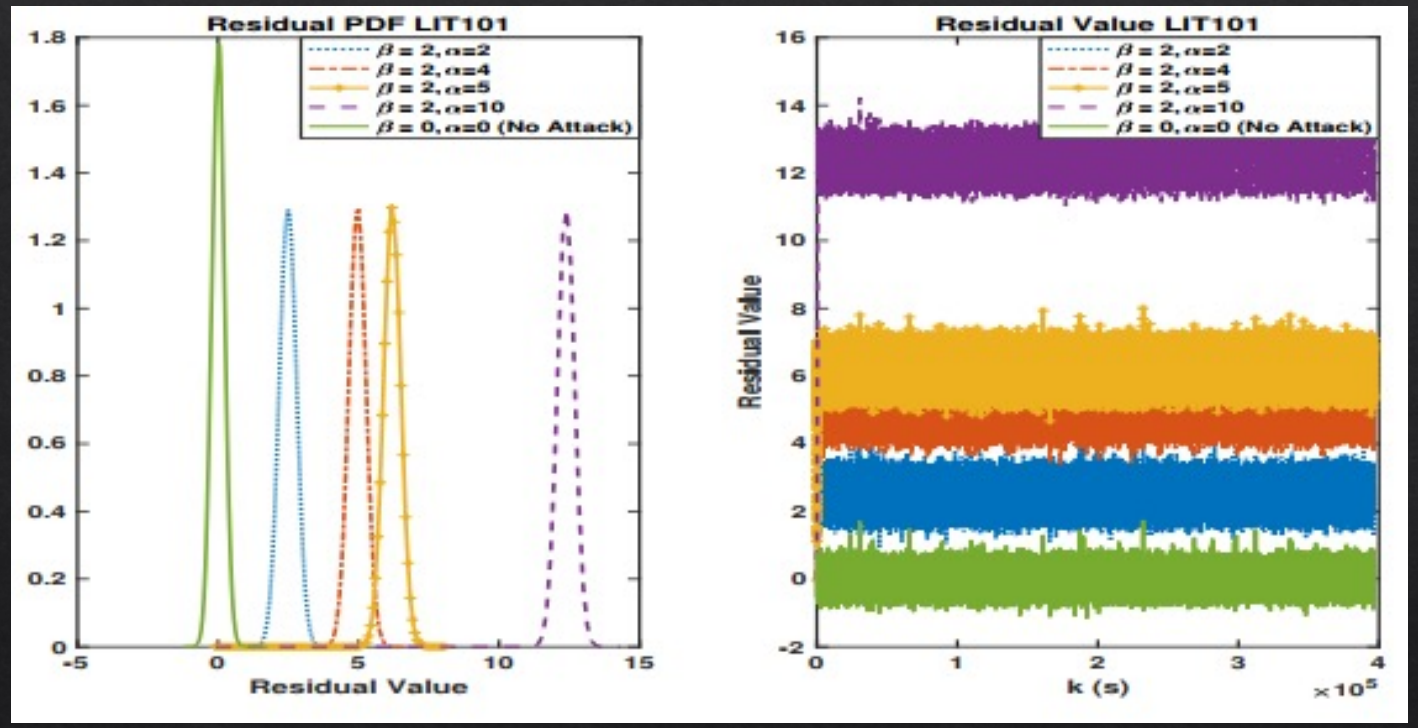
Feature	Description
Mean	$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
Std-Dev	$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x}_i)^2}$
Mean Avg. Dev	$D_{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N}\lvert x_i - \bar{x}\rvert$
Skewness	$\gamma = \frac{1}{N}\sum_{i=1}^{N}(\frac{x_i - \bar{x}}{\sigma})^3$
Kurtosis	$\beta = \frac{1}{N}\sum_{i=1}^{N}(\frac{x_i - \bar{x}}{\sigma})^4 - 3$
Spec. Std-Dev	$\sigma_s = \sqrt{\frac{\sum_{i=1}^{N}(y_f(i)^2) * y_m(i)}{\sum_{i=1}^{N} y_m(i)}}$
Spec. Centroid	$C_s = \frac{\sum_{i=1}^{N}(y_f(i)) * y_m(i)}{\sum_{i=1}^{N} y_m(i)}$
DC Component	$y_m(0)$

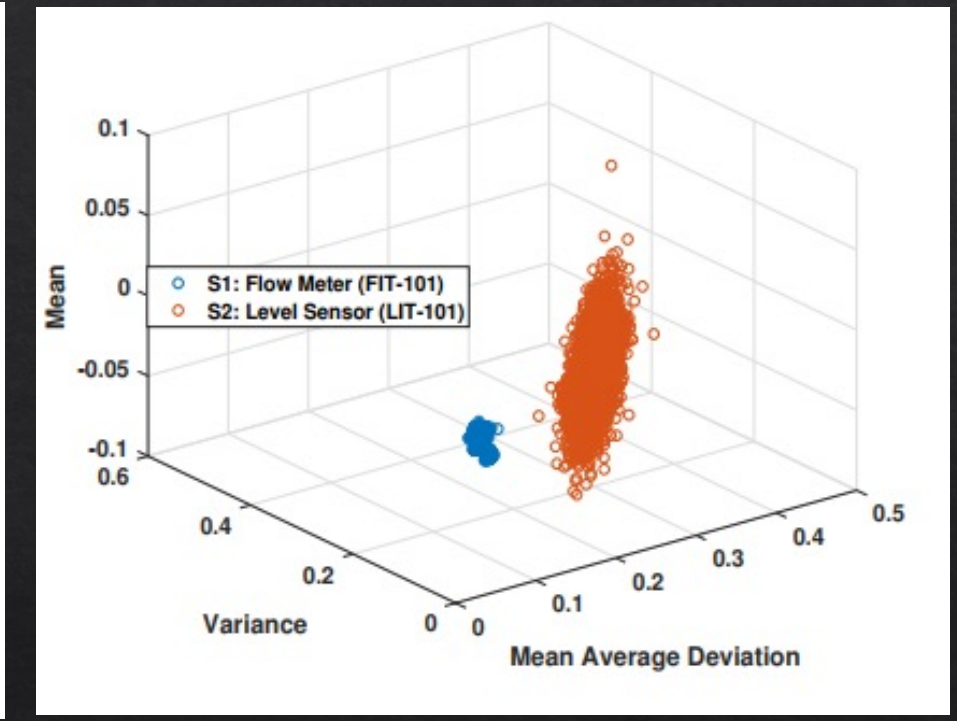
Attacks

- Types of possible attacks
  - External
  - Internal
- They run 41 attacks
  - Sensor
  - actuators

Attack Sequence Number	Start Time	End Time	Attack Point	Start State	Attack	Expected Impact or Attacker Intent	OC-SVM Detection Accuracy
3	28/12/2015 11:22:00	11:28:22	LIT-101	Water level between L and H	Increase by 1 mm every second	Tank Underflow; Damage P-101	100%
7	28/12/2015 12:08:25	12:15:33	LIT-301	Water level between L and H	Water level increased above HH	Stop of inflow; Tank underflow; Damage P-301	100%
8	28/12/2015 13:10:10	13:26:13	DPIT-301	Value of DPIT is <40kpa	Set value of DPIT as >40kpa	Backwash process is started again and again; Normal operation stops; Decrease in water level of tank 401. Increase in water level of tank 301	100%
10	28/12/2015 14:16:20	14:19:00	FIT-401	Value of FIT-401 above 1	Set value of FIT-401 as <0.7	UV shutdown; P-501 turns off; UV did not shutdown; P-501 did not turn off	100%
11	28/12/2015 14:19:00	14:28:20	FIT-401	Value of FIT-401 above 1	Set value of FIT-401 as 0	UV shutdown; P-501 turns off	100%
16	29/12/2015 11:57:25	12:02:00	LIT-301	Water level between L and H	Decrease water level by 1mm each second	Tank Overflow	100%

Results


Residual PDF LIT101
β = 2, α=2
β = 2, α=4
β = 2, α=5
β = 2, α=10
β = 0, α=0 (No Attack)
Residual Value
Residual Value LIT101
β = 2, α=2
β = 2, α=4
β = 2, α=5
β = 2, α=10
β = 0, α=0 (No Attack)
Residual Value
k (s)
×10^5


S1: Flow Meter (FIT-101)
S2: Level Sensor (LIT-101)
Mean
Variance
Mean Average Deviation

Results

- Attack detection delay

↓ Sample Size / Sensor →	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18
60	95.1097%	94.6%	96.5839%	94.6427%	93.423%	95.9593%	94.5721%	94.6205%	94.5721%	100%	94.5721%	99.1053%	96.3729%	94.9045%	94.5721%	94.5721%	94.6189%	94.5721%
120	95.2125%	95.0253%	96.4684%	94.6821%	94.9547%	96.6162%	94.5721%	94.5721%	94.5721%	100%	94.5721%	98.8507%	97.2138%	95.5884%	94.5721%	94.5721%	95.8626%	94.5721%
250	95.2391%	95.4374%	96.6003%	94.8492%	96.4977%	96.8431%	94.5721%	94.5721%	94.5721%	100%	94.5721%	98.5122%	97.5922%	96.7%	94.5721%	94.5721%	97.0962%	94.7944%
500	95.0856%	96.0438%	96.8309%	95.7221%	97.2827%	96.7693%	94.5722%	94.5722%	94.5722%	100%	94.5722%	97.666%	97.8029%	97.5565%	94.5722%	94.8665%	97.9261%	95.64%
2000	96.3816%	96.2719%	95.9978%	95.0384%	96.1897%	95.7237%	94.545%	97.5877%	95.6689%	100%	94.5724%	96.1897%	97.8618%	96.3542%	94.4901%	96.6009%	98.1908%	96.9846%

- How well the technique works?

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%

Discussion

- How about the fingerprint of aged sensor or replaced sensor ?
- What about computation power and time ? Does it worth it ?
- Were the sensors in the experiment sensing same thing ?
- Does those attacks really match the assumption ?