Questions and Answers

Q1

From the diagrams, I got the impression that electrical noise is inserted directly onto the DC electrical output of the PV panels BEFORE it gets to the inverter, and then the AC output of the inverter is being monitored to look for that exact same noise somewhere before the grid connection point… and then if a variance is detected, the presumption is an attack is taking place? From some of your pictures, it also seemed like the monitoring could take place ANYWHERE before the feeder head, and that you could actually detect WHICH inverter on a distribution network was being attacked (if you know the unique noise signal and/or the network model)? I'm assuming you are only looking at the electrical signal, and NOT the telemetry being reported by the inverters. (Or, cross comparing the electrical values with the reported telemetry.) Is that correct?

A1

We are superimposing the Dynamic Watermarking (DW) signal onto the control inputs (if it is available such as lab inverter experiments) or reference (if we don't have access to the control inputs such as commercial inverters).
Then, we deploy a general-purpose cyber-attack defense methodology called "Dynamic Watermarking (DW)" to detect cyber-attacks on sensor measurements. And we keep real-time monitoring of any statistical changes in the sensor measurements. These DW asymptotic tests can be converted into statistical tests by choosing two appropriate thresholds. It is established that if the reported sequence of measurements passes both DW Test 1 and DW Test 2, then any malicious sensor present could not have distorted the actual measurement value beyond adding a zero power signal to the ambient noise.

Q2

You mention that the system would also detect a malfunctioning sensor is very interesting. Could the Secondary Observer then become a surrogate remedy for a bad sensor for an extended time ? (until repair of the device or comm channel).

A2

We developed secondary corrective control in conjunction with the general-purpose cyber-attack defense methodology called "Dynamic Watermarking (DW)". Using it as a surrogate remedy for a bad sensor for an extended time may be possible given certain extensions. This is because if there is a bad sensor in the system, it reports a false value with high-level noise to the controller.

Q3

What is a practical attack vector creating the (current) sensor attacks (Harmonics Injection and Amplitude Reduction)? 2. Can the method distinguish between sensor fault and the hall sensor attack?

A3

There are so many potential practical attack vectors in real-world applications including physical tampering, sensor measurement interception, sensor spoofing, network interference, etc.
Also, we developed an auxiliary test to distinguish between sensor fault and attack; however, for both cases, it would be better to safely shut down the system and check the entire system for system safety reasons.

Q4

What is the difference between failure detection and cyber attack within system identification?

A4

The system identification and the detection of cyber-attacks or failures are different.
We are using system identification for overcoming the requirement to know the system model. This is because an accurate model of the system is unknown due to the nonlinearity of the system as well as the complexity of the connected grid. For the detection of cyber-attacks or failures, we keep monitoring the system behavior in real-time using our general-purpose cyber-attack defense methodology called "Dynamic Watermarking (DW)".

Q5

Why using thresholds if signals can swing back and forth during a transient behavior for example?

A5

In our DW theory, the security tests are asymptotic tests. These asymptotic tests can be converted into statistical tests by choosing two appropriate thresholds.
We initiate the detection tests after steady-state.

Q6

What if watermarking signal generator is used as an attack?

A6

We are superimposing the private random excitation signals "watermark" onto the control inputs for malicious sensor measurements case (Actuators are honest).
The statistics of the watermark can even be disclosed to other nodes in the system; however, its actual realization is not revealed to any other node in the system.
The question raised means the actuators are being attacked (Actuators are malicious). Even in this case, our DW method works if and only if the sensor measurements are honest.
However, it would be challenging to protect the system if both actuators and sensors are malicious.