INL/RPT-23-74252-Revision-0

Signal Decomposition for Intrusion Detection in Reliability Assessment in Cyber Resilience Summary Report

August 2023

Paul W Talbot, Dylan James McDowell, Bri Rolston, Tyler Lewis

Changing the World's Energy Future

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/RPT-23-74252-Revision-0

Signal Decomposition for Intrusion Detection in Reliability Assessment in Cyber Resilience Summary Report

Paul W Talbot, Dylan James McDowell, Bri Rolston, Tyler Lewis

August 2023

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

SUMMARY FOR ANNUAL REPORT

Project Title: Signal Decomposition for Intrusion Detection in Reliability Assessment in Cyber Resilience

Project Number: 21A1050-024FP

Initiative: Secure and Resilient Cyber-physical Systems

Core Capabilities:

- Cyber and Information Sciences
- Applied Mathematics

Total Approved Amount: $1,287,000 over 3 years

Principal Investigator: Paul Talbot, INL

Co-investigators:

- Bri Rolston, INL
- Dylan McDowell, INL
- Hany Abdel-Khalik, Purdue University

Collaborators:

- None

Impact Statement:

Detecting subtle data manipulation through digital signal characterization in automated workflows, coupling physics- and data-driven algorithms.

Project Description:

The complexity involved in ensuring cyber resilience for physical process interactions in connected systems such as energy grids increases dramatically as the coupling between processes becomes more direct and responsive. An example of this growing complexity is seen in integrated energy systems (IES), in which various processes such as nuclear heat generation and commodity production are directly coupled for increased responsiveness to highly variable signals such as market pricing and electricity demand. This causes the potential attack surface of the coupled processes to be larger than those of the two processes employed independently.

Securing these complex systems requires a two-fold monitoring approach that involves both cybersecure monitoring for potential malicious incursions and physics monitoring for system tampering. Physics monitoring includes analyzing the signals within the system for anomalous behavior. Such analyses have proven insufficient when based solely on data-driven machine learning and artificial intelligence (MLAI) techniques or on low-level model comparisons. Previous efforts at Purdue University indicated that combining high-fidelity models with MLAI algorithms could serve as the basis for a software tool that detects anomalies in physical processes.

The present work built upon that concept, developing an advanced library for signal decomposition and analysis by using both MLAI and high-fidelity physics algorithms to enable greatly enhanced anomaly detection capabilities—especially in regard to detecting false data injections. This software can be used as part of a secure imbedded intelligence system designed under Consequence-driven Cyber-informed Engineering for complex coupled systems. The aforementioned advanced library provides the foundation for online and posteriori analysis of digital signals for the purpose of detecting potential malicious tampering in digital signals that represent physical processes. Demonstrations conducted throughout the development highlight the effective use of characterization algorithms to detect signal perturbations—particularly triangle-attack-style perturbations—in three wide-ranging applications: seismic monitoring, nuclear thermal-hydraulics system simulation, and custom manufacturing.

Talent Pipeline:

- Yeni Li, student at Purdue University
- Tyler Lewis, student at Purdue University

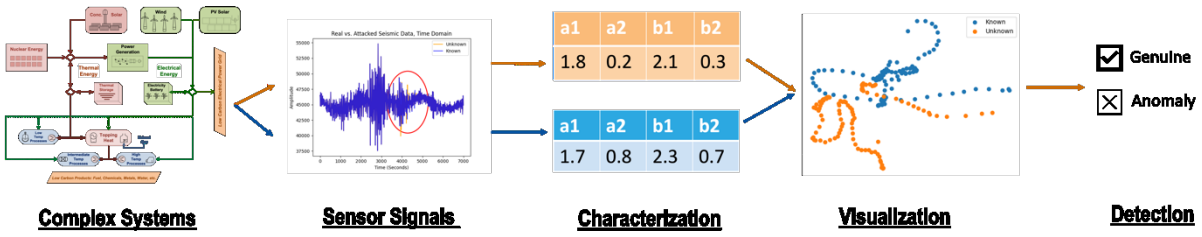
External Peer-Reviewed Publications:

- Yeni Li, Arvind Sundaram, Hany S. Abdel-Khalik & Paul W. Talbot (2022) Real-Time Monitoring for Detection of Adversarial Subtle Process Variations, Nuclear Science and Engineering, 196:5, 544-567, https://doi.org/10.1080/00295639.2021.1997041
- Yeni Li, Paul W. Talbot & Hany S. Abdel-Khalik. (2022). A Novelty Detection Workflow for Nuclear System Monitoring, 2022 ANS Winter Meeting, Pheonix, AZ.

External Intellectual Property

- Purdue, INL (2021), SONAR [source code], https://github.com/yenili/sonar (private repository)

Figure and Caption:



SONAR anomaly detection algorithm workflow. Mapping complex digital signals to a characterization space enables detection of otherwise undetectable perturbations.

1. SCIENTIFIC AND TECHNICAL ACCOMPLISHMENTS

Mapping signals into a characterization space greatly enhances the capability to detect subtle anomalies. The primary technical accomplishments of this activity include extending anomaly detection algorithms to the development of workflows via a novel software known as the Signal-Oriented Network Anomaly Recognition (SONAR) tool. SONAR was built using the Risk Analysis Virtual Environment (RAVEN) framework developed and maintained at Idaho National Laboratory. RAVEN provides user-interactive workflow management for automated formatting, construction, and analyses—as supported by in-depth documentation and data visualization tools. Using SONAR, a user can train a flexible model on a known

dataset, then test for similarity with other unlabeled, potentially anomalous signals. The mechanism used for determining the level of similarity—which classifies new datasets as either *genuine* or *anomalous*—relies on distance metrics as well as data-based time-series signal decomposition techniques that map signals from the time domain to domains of signal *features* (or *characteristics*). The advantage of the feature domain is that subtle behavior is often exaggerated in feature space, thus more clearly differentiating genuine signals from anomalous ones. SONAR uses distance measures (e.g., cosine distance) to gauge the similarity between signals in characterization space. Furthermore, the software is designed to accommodate many feature identification techniques such as dynamic mode decomposition, Fourier analysis, and wavelet decomposition.

SONAR offers a parameter sweeping capability that assists in parameter tuning when fitting characterization models to large datasets. The demonstrations highlighted in SONAR's documentation showcase its ability to tune the characterization model so as to recognize genuine signals, as well as the ability to compare against an array of new signals simultaneously. SONAR's use of the decomposition space to differentiate anomalies from genuine signals enables advanced, high-precision detection of subtle and long-range anomalies—a capability fine-tuned by SONAR's modular automated workflows.

Example Case 1: Seismic Signal

To illustrate the use of SONAR, consider the signal represented in Figure 1. For this demonstration, we used samples of seismic activity data, which are relevant to IES. Perturbations in seismic signals are potentially easy to mask, thanks to their noisy, high-variance signals in typical behavior (see Figure 1). Unidentified perturbations in signals might, for example, force a generating plant to shut down, thereby disrupting the energy supply and resulting in high costs. The blue line in the figure represents the genuine data on which we trained our model. The orange line represents a subtle adversarial attack conducted over a long timeframe. This attack is generated by altering the mean of the signal while overlaying the noise from the original signal (a form of replay attack). For verification, we constructed many such subtle intrusions to evaluate the robustness of our selected algorithm at differentiating intrusions from a set of genuine seismic signals pertaining to similar events.

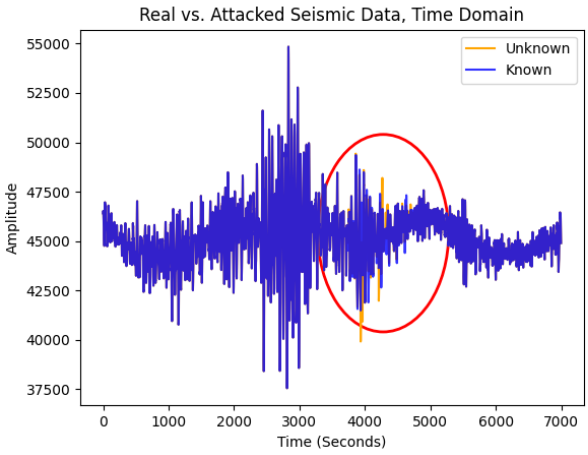


Figure 1: Seismic Data with a Subtle Perturbation

The first step in using SONAR to characterize and classify signals is to decompose the data. Via SONAR, users have access to many characterizing algorithms. For the present demonstration, we selected the randomized window decomposition characterizing algorithm, wherein many random snapshots from the

data are characterized using support vector decomposition. This is one of many data characterization methods, each of which has its own set of strengths and weaknesses. For instance, Fourier-based techniques are optimal for periodically repeating data, and wavelets excel at characterizing localized phenomena. In the present example, randomized window decomposition was selected for the flexibility it affords in analyzing the principal components (or features) in windows of the data.

Once the training data are decomposed into feature space, the algorithm uses the fitted features as a standard against which to compare unknown signals. Each unknown signal is decomposed in the same manner as the training signal, then a distance is calculated between these based on the desired characterization space selected by the user. This demonstration uses the cosine distance, measured as the angle between the training and unknown signals in characterization space. SONAR labels low-distance signals as *genuine* and high-distance ones as *anomalous*. Figure 2 illustrates the training signal and the example perturbed signal in characterization space, underscoring the value of the SONAR approach. In Figure 1, we see a barely perceptible anomaly that traditional methods are unlikely to flag as an anomaly. However, the difference is pronounced in Figure 2. Though the two signals in characterization space are similar, they nonetheless clearly differ, and SONAR correctly labels the perturbed data as *anomalous*.

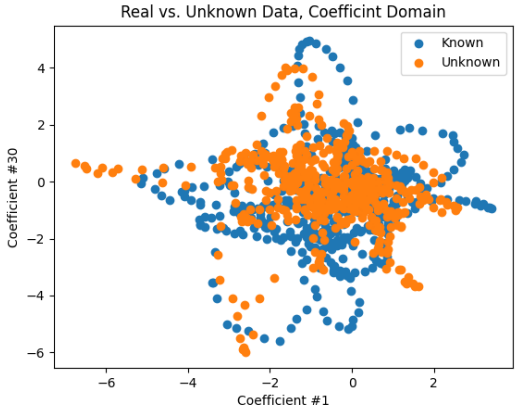


Figure 2: Coefficient Space of the Genuine and Spoofed Signals

To evaluate the effectiveness of our trained classification model, we generated several perturbed signals and then tested our algorithm's ability to classify them correctly. Characterization and distance were determined for each signal, such that we could gauge SONAR's performance via confusion matrix (see Figure 3). This matrix demonstrates SONAR's built-in parameter sweeping protocol in which the distance metric, decomposition parameters, and data window size can be automatically fine-tuned for the user.

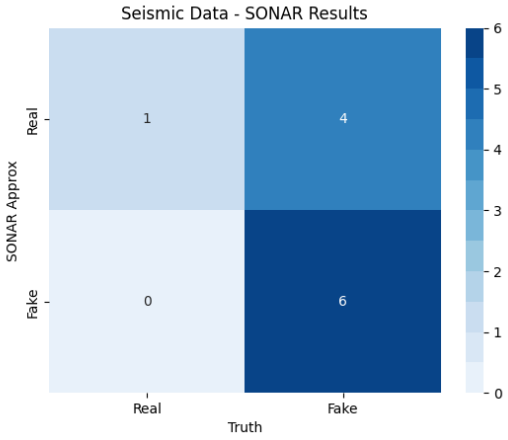


Figure 3: Confusion Matrix for SONAR Algorithm as Applied to the Seismic Data Study

Early investigations into distance metrics led to publications (e.g., [Li, 2022]) that explored preliminary usage of the SONAR classification algorithm for detecting subtle anomalies when modeling a pressurized-water reactor simulated in RELAP5, a thermal-hydraulics system analysis code. In using SONAR on this model data, the outlet temperature, a model output, undergoes denoising followed by decomposition, thus enabling high-confidence anomaly detection. The RELAP5 data used in that research represent another of SONAR's documented demonstration cases. Figure 4 shows the same three figures shown separately for the seismic case: the time-domain signal with a subtle injected anomaly, the SONAR-decomposed signals in feature space, and the confusion matrix. In this small-data, low-noise example, SONAR identifies the spoofed signals with 100% success. In fact, SONAR's performance can be guaranteed for this dataset—any deviation from the genuine signal will reflect a wide margin between it and the spoofed signals, making characterization straightforward.

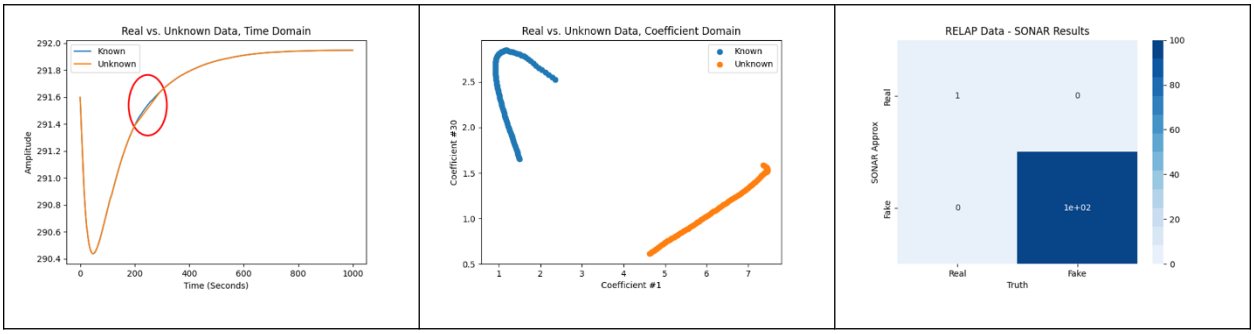


Figure 4: RELAP demonstration: (Left) Time-Domain Signal; (Center) Feature Space; (Right) Confusion Matrix

The final SONAR demonstration dataset in this activity characterizes in-line temperature readings of melt pools during a 3D printing build. Data manipulation in 3D printing can degrade material quality, potentially causing decreased safety as well as losses for a manufacturer. This case also shows that SONAR can be applied to a wide range of data types and quantities, while not strictly requiring a high volume of data (as is otherwise the case for many machine learning applications). Figure 5 gives an overview of the SONAR results for this dataset. Once again, we see that all injected anomalies are identified by SONAR, despite the noise-dominated nature of the data and the relatively low volume thereof.

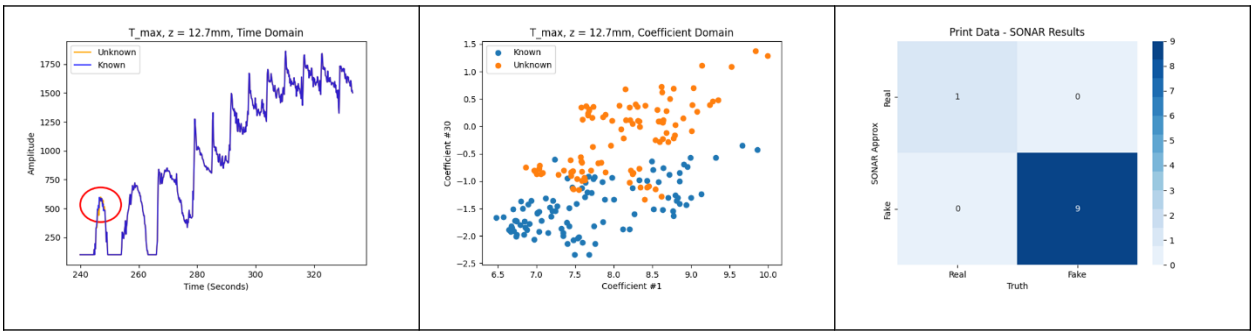


Figure 5: Overview of SONAR for 3D Printing Data: (Left) Time-Domain Signal; (Center) Feature Space; (Right) Results of the SONAR Algorithm.

In summary, the novel algorithms and workflow deployed by SONAR have demonstrated the capacity to detect subtle anomalies via streamlined workflows applicable across a range of disciplines. The several demonstration cases detailed here and documented in the SONAR code base prove the utility of this tool for a wide array of data types, use cases, and physics, from industrial-scale, noise-dominated data to small-scale experimental and simulation work.

2. BENEFITS TO THE DEPARTMENT OF ENERGY

With many electricity price projections indicating a highly competitive generation market, the U.S. Department of Energy has dedicated significant funding to exploring alternate uses of nuclear-produced heat, aside from electricity generation. An example of this funding is the cross-cutting technologies IES program, in which advanced and existing nuclear reactors are considered for coupling with heat users, such as hydrogen electrolysis, water desalination, district heating, chemical refineries and processes, pulp and paper manufacturing, and many others. However promising these connections are, they introduce an larger cyberattack surface than do traditional, electrical-power-focused nuclear generating plants.

To guard against cyberattacks on Industrial and Automation Control Systems (IACS) and Operational Technology, resources are an oft-focused layer of firewall-style digital cyber-informed engineering by design. Most approaches to cybersecurity risk management in IACS and Operational Technology environments use known cybersecurity best practices to identify, mitigate, or defend against cyberattacks. There are very few approaches in which the reverse is true (i.e., using engineering and operations best practices to improve IACS cybersecurity risk profiles and ensure process resilience). The fact that conventional engineering process monitoring techniques (e.g., anomaly detection and pattern recognition) were not designed with an adversary in mind renders them ineffective at detecting process manipulations. The past decade saw numerous papers demonstrating how these techniques can be bypassed by physics payloads designed by attackers with domain-specific system knowledge. Replay-style attacks based on an approach such as Larsen's "triangles" are proficient at bypassing conventional engineering process monitoring.

This project resulted in the building of software tools, as well as a demonstration of their use in identifying anomalous behavior in—or significant changes to—the *physical* process (i.e., unexpected physics related to cyberattacks or other activities originating from automation technology), using engineering data and capabilities as opposed to cybersecurity tools. The approach was demonstrated across three wide-ranging physics disciplines and is positioned for use in nuclear IES design, affording increased resilience. The software design is sufficiently modular to open the door to researching additional signal analysis algorithms and deploying them for specific physical systems in order to enhance anomaly detection capabilities and accuracy.

3. PROGRAM DEVELOPMENT ACCOMPLISHMENTS

The algorithms and workflows developed and demonstrated in this research activity feature a wide range of potential programmatic applications. Employing subtle anomaly detection workflows opens up significant opportunities for securing wireless security, with embedded signal checking potentially assisting in limiting data injection in wireless communications. Representing wireless communications as complex coupled digital signals would enable analysts to train and classify signals as they are received.

The anomaly detection workflow is also useful in root cause analyses. In addition to identifying anomalous signals in real time, sensor behavior can be analyzed after the fact in order to determine when anomalous behavior may have occurred and to assist in exposing data manipulation activities.

We also anticipate continued research work in the IES programmatic space, as well as in opportunities for seed or full LDRD proposals that pertain to non-traditional heat applications for nuclear energy as well as to cyber resilience. While research on IES cybersecurity is in its early stages, anomalous signal detection can be included as an aspect of secure embedded intelligence by design.

4. RESEARCH OUTPUTS

Published Submissions

- Yeni Li, Paul W. Talbot & Hany S. Abdel-Khalik. (2022). A Novelty Detection Workflow for Nuclear System Monitoring, 2022 ANS Winter Meeting, Pheonix, AZ. INL/CON-22-68064.
- Yeni Li, Arvind Sundaram, Hany S. Abdel-Khalik & Paul W. Talbot. (2022). Real-Time Monitoring for Detection of Adversarial Subtle Process Variations, Nuclear Science and Engineering, 196:5, 544-567, https://doi.org/10.1080/00295639.2021.1997041, INL/JOU-21-62163.

Planned Submission

- Paul Talbot, Dylan McDowell, Tyler Lewis, Yeni Li, Xingyue Yang, Hany Abdel-Khalik (September 2023 proposed) Signal-Oriented Network Anomaly Recognition, *Mathematics* Special Issue: *Models and Algorithms in Cybersecurity*.