VIRTA LABS™

Detecting Anomalous Behavior in Infusion Pumps from the AC Power Line

August 2015

Andrew DeOrio, Ph.D.
Benjamin Ransford, Ph.D.
Denis Foo Kune, Ph.D.
Kevin Fu, Ph.D.

Executive Summary

This report summarizes our study of anomaly detection techniques using readings from the AC power cord on an infusion pump. Our goal was to measure the effectiveness of our power analysis techniques in identifying anomalous pump behavior that could include intentional tampering via cyber-physical attacks. Our system was able to detect simulated cyberattack scenarios with close to 100% accuracy. The tunable model is configurable to generate near-zero false positives.

We conducted the study with Virta Labs PowerGuard™ v1.0 hardware and CloudGuard™ v0.5 software at the Virta Labs facility in Ann Arbor, Michigan. The devices under test were a CareFusion Alaris 8100 infusion pump with a PC unit model 8015, a Hospira Plum A+, and a Hospira LifeCare PCA. The experiments described in this whitepaper leverage Virta Labs' expertise in behavioral modeling, signal processing, and threat analysis.

Background and Threat Model

Infusion pumps deliver intravenous drugs to patients at a prescribed dose and rate. A cyber-physical attack may alter both parameters. We have demonstrated an effective anomaly-detection method that can automatically identify unusual infusion rates. The threat model is based on NIST's draft document on Medical Device Security for wireless infusion pumps from November 2014 (https://nccoe.nist.gov/projects/use_cases/medical_devices). The threat actors span a broad range from "honest-but-curious" users who may accidentally trigger a dose change, to sophisticated attackers intending to cause harm through a cyber-physical attack; both could affect patient outcomes. The attack vectors on the infusion pump include the network interface and the push-button human interface. We modeled a cyber-physical attack on

an infusion pump as a large increase in infusion rate. On the large volume pumps the increase could be as small as 90ml/hr and as large as 989ml/hr, and on the syringe pump, the increase was 19.9 ml/hr.

PowerGuard™ is Virta Labs' proprietary power-monitoring product that constantly observes a plugged-in device's power consumption at a fine granularity, performing basic signal processing on board and sending semi-processed signals to Virta Labs' CloudGuard™ analysis service. CloudGuard™ conducts further signal analysis, including activity recognition via machine learning, and flags signals that may indicate anomalous behavior or compromise.

Experiment Setup

We studied three infusion pumps (Figure 1), a CareFusion Alaris 8100 large volume pump attached to an Alaris 8015 PC unit, a Hospira Plum A+ large-volume infusion pump, and a Hospira LifeCare PCA syringe infusion pump. We recorded signals collected directly from PowerGuard devices as they monitored the medical equipment.

Figure 1: Photos of infusion pumps tested. Left: Alaris 8100 Large Volume pump with an Alaris 8015 PC unit. Center: Hospira Plum A+ large-volume pump, infusion bag type. Right: Hospira Lifecare PCA, syringe type. Also shown: PowerGuard™ pass-through power outlet.

For each pump, we first established a set of training data by recording only the pump's normal behavior with the PowerGuard™ over a period of 20 minutes to 24 hours depending on the experiment set up. From this dataset, we created a mathematical model of normal behavior using a proprietary machine-learning training workflow. Once the model was created from the training data, we commenced "live" measurement of the pump, a mode in which new measurements are constantly compared to the trained model using a variety of metrics. If a series of new measurements were sufficiently outside the normal range for a set of the relevant metrics, our detectors automatically deemed the pump's behavior as anomalous.

Metrics and Features

PowerGuard™ uses a variety of time- and frequency-domain components measurements on the AC power line. The main challenges in model development are choosing the right set of features to characterize a device's behavior, and collecting enough data from instances of that

device to avoid overfitting the model (making it overly specific and not generalizable) to any one particular instance of the device. We developed a single anomaly-detection model per device and a single machine-learning classifier model per device. The features and statistical methods were applicable to all three pumps, although the exact parameters of the model varied between devices. We used these computed models to differentiate normal from abnormal activity.

Normal Scenarios and Abnormal/Attack Scenario

We initially profiled normal behavior on the large-volume pumps (CareFusion Alaris 8100 and Hospira Plum A+) as a low infusion rate of 50ml/hr, and on the syringe type pumps (Hospira LifeCare PCA) as a low infusion rate of 0.1ml/hr. A “normal” infusion rate depends on the drug being delivered, the patient, and other clinical decision processes, but in keeping with the NIST threat model, we focused on scenarios in which an adversary would be able to change a low infusion rate to an inappropriately high infusion rate. We further modeled normal behavior to include infusion rates of 10 ml/hr through 500 ml/hr at regular intervals.

We modeled the attack scenario on the large volume pumps (CareFusion Alaris 8100 and Hospira Plum A+) as a high infusion rate at the maximum configurable rate of 999 ml/hr, and on the syringe type pump (Hospira LifeCare PCA) as a high infusion rate at the maximum 20 ml/hr for a simulated morphine sulphate 5 ml/hr syringe. In one case, we also modeled an anomalous infusion rate of 100 ml/hr compared to a normal infusion rate of 10 ml/hr.

Feature Selection to Differentiate Normal from Abnormal Behavior

A feature is *informative* if it takes one set of values for one type of signal—e.g., signals from normal operation—and another set of values for a different type of signal—e.g., signals from abnormal operation. Figure 2 shows raw PowerGuard™ measurements of a single feature from each infusion pump, namely the amplitude of the 60 Hz component of the AC power line over time. (60 Hz is the dominant frequency in utility AC power in the United States.) The plots show data for two pump activities; a normal infusion rate for each pump is shown in blue. A cyber-physical attack is modeled as an attempt to increase the infusion rate to the maximum configurable rate for each device; data from these rates is shown in red.

The lines in Figure 2 isolate a single component of the much noisier raw time-domain signals we collected from the infusion pumps. As the plots show, under different infusion rates, the infusion rate affects the minimum (min), maximum (max), and mean amplitude values of this 60 Hz frequency component. The plots also show periodic behavior we observed at this frequency; peaks occur roughly every 40 seconds. For this study, we therefore selected the two most informative features from the set we measured: min and max 60 Hz amplitude, computed in rolling windows of 100 seconds to capture periodic behavior.

Figure 2: Amplitude of the 60 Hz band, which is directly proportional to power consumption caused by electromechanical actuation (pumping). Left: Alaris 8100 pump with 8015 PC. Middle: Hospira Plum A+. Right: Hospira Lifecare PCA. In all plots, blue represents a "normal" infusion rate (50 ml/hr for Alaris 8100 and Hospira Plum, 0.1 ml/hr for Hospira PCA). Red represents an "abnormal" infusion rate (999 ml/hr for Alaris 8100 and Hospira Plum, 20 ml/hr for Hospira PCA).

Results

We applied the features developed in the previous section to train and test automated models. The following plots show sample data from our anomaly-detection workflow. In these examples, the infusion pump is set to deliver drugs at different rates, including the modeled cyber-physical attack infusing at much higher rates than the trained normal rate.

Case 1: CareFusion Alaris 8100, low infusion rate (50 ml/hr) and maximum infusion rate (999 ml/hr)

Figure 3 is a visualization of an infusion pump anomaly-detection model that the PowerGuard™ toolchain produces. The X- and Y-axes plot the two features used by this model, the minimum and maximum of the 60 Hz component of the power line in rolling 100-second windows. Each dot represents one example measured from the device over 100 seconds. Normal data collected from the pump at a 50 ml/hr delivery rate (white dots) were used to train a one-class support vector machine (SVM) model (red line). Next, we evaluated the model with novel normal 50 ml/hr infusion rates (green dots), which fall inside the red boundary, indicating no anomaly. Finally, we evaluated the model on novel measurements that simulated abnormal (attack) activity at a 999 ml/hr rate (red dots). These red dots fall outside the red boundary, indicating an anomaly. Our error rate on the training data was 1.0%, with 0.0% error for normal test data (false positives) and 0.0% error on abnormal test data (false negatives). We note that the trade-off between false negatives and false positives is configurable in the model, and later we show that a 0.0% false positive rate is possible.

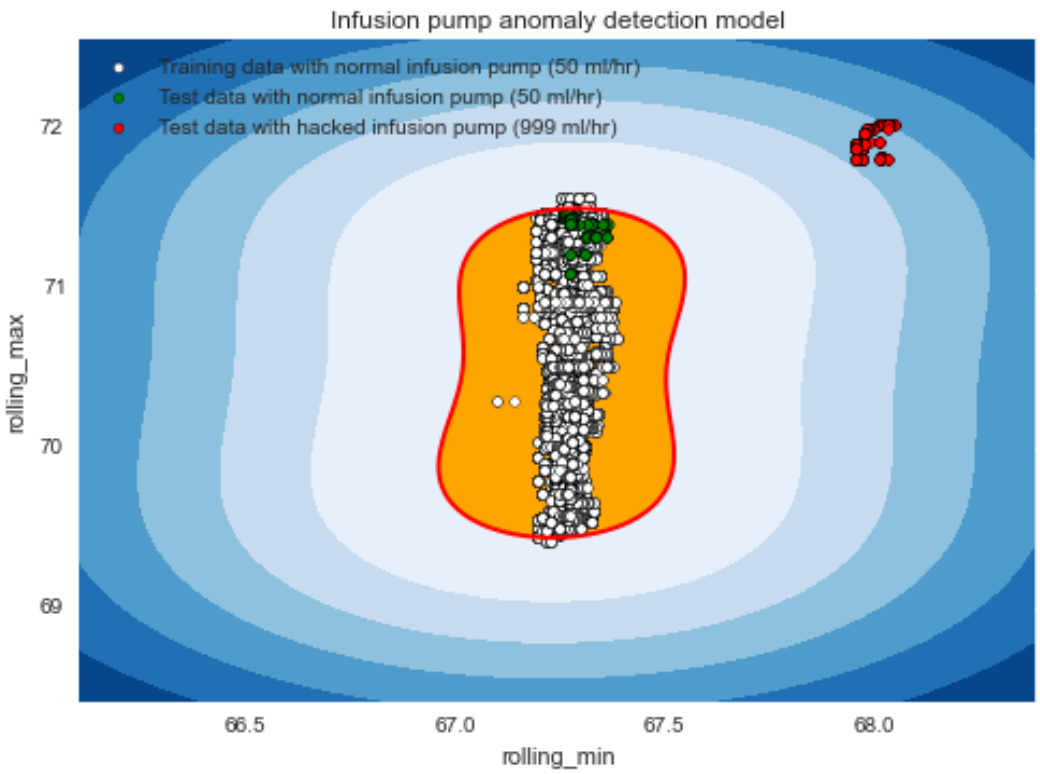

Figure 3: CareFusion Alaris 8100, normal 10 ml/hr infusion rate (white dots) vs. abnormal 999 ml/hr (red dots).

Case 2: CareFusion Alaris 8100, multiple infusion rates (10 to 500 ml/hr) and maximum infusion rate (999ml/hr)

In our next experiment, we updated the anomaly-detection model to work with multiple normal infusion rates, ranging from 10 ml/hr to 500 ml/hr (blue and white dots). The following plot again shows rolling minimum and rolling maximum feature measurements on the X- and Y-axes. The learned model is shown with a red line. Compared to Figure 3, the “normal” region has expanded to encompass a wider array of normal infusion rates. We again observed that examples from a 999 ml/hr simulated abnormal infusion rate fell outside the boundary and were marked as anomalies. The error on our training data was 0.1% (false positives), and error on the test data was 0.0% (false negatives).

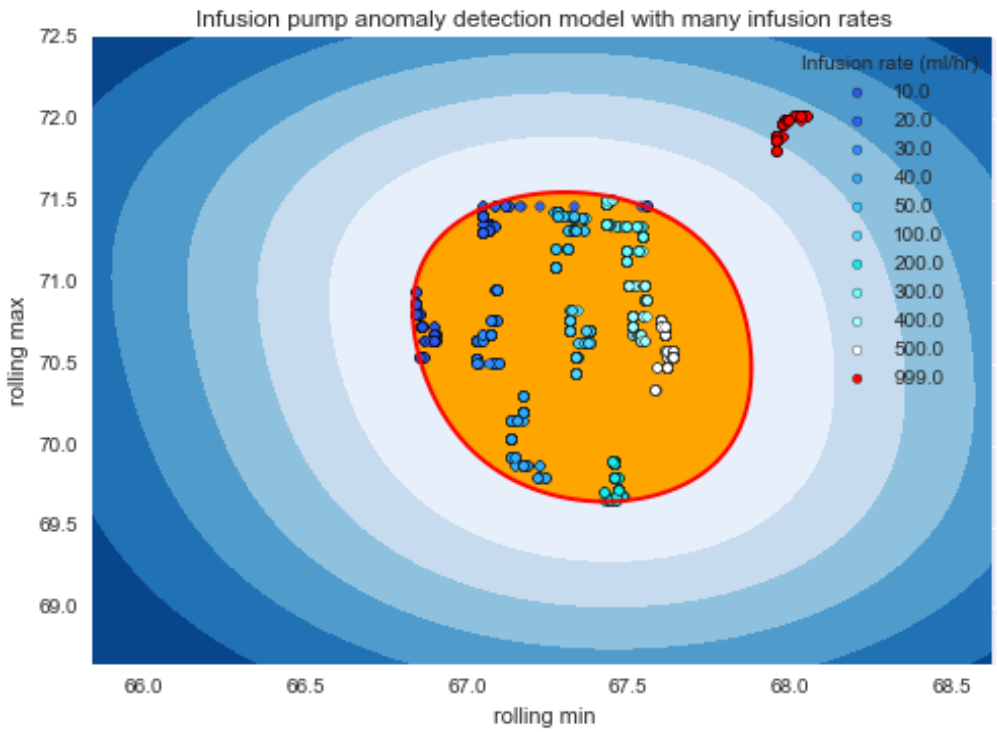

Figure 4: CareFusion Alaris 8100, multiple normal infusion rates (10 ml/hr to 500 ml/hr, white dots) vs. abnormal 999 ml/hr (red dots)

In addition to anomaly detection, we also built a regression model to predict the infusion rate using only features measured from the AC power line. The input to this model is the rolling min feature, measured over 100 seconds. The output is the predicted infusion rate, which is a continuous value in contrast to the discrete "normal vs. abnormal" output of the anomaly detector. This plot visualizes the regression model, shown by the blue line. The data used to build the model is shown with black dots. The model had a residual sum-of-squares value of 14,658 and a variance of 0.82. Thus, we can see that there is a relationship between rolling minimum of the power consumption's 60 Hz band and the infusion rate.

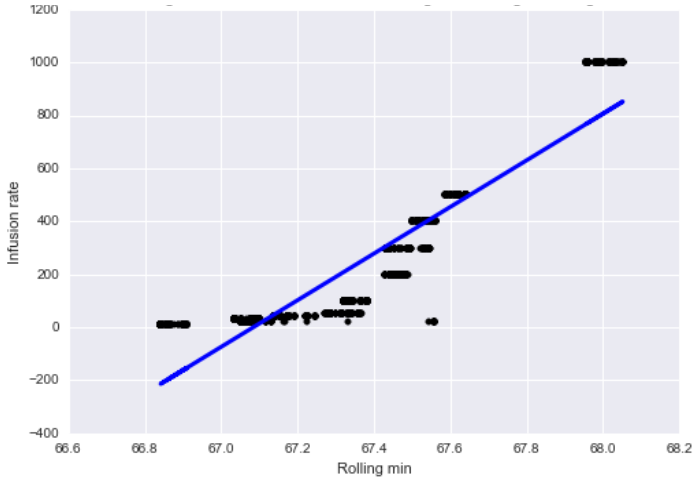

Figure 5: Relationship between infusion rates and the value of one of our metrics (rolling min).

Case 3: CareFusion Alaris 8100, discrimination between normal (10 ml/hr) and low anomalous (100 ml/hr) rate

In this anomaly-detection experiment, we simulated the case of an 80-year old patient who received an accidental fatal infusion rate of potassium chloride (BBC: *http://www.bbc.com/news/uk-england-birmingham-12552718*). We trained an anomaly detection model with the patient's prescribed infusion rate, 10 ml/hr (white dots). Then, we tested on data at her fatal infusion rate, 100 ml/hr (red dots). Our model accurately detected these latter data points as anomalies, with 0.5% error rate on the training data (false positives) and 0.0% error on the test data (false negatives).

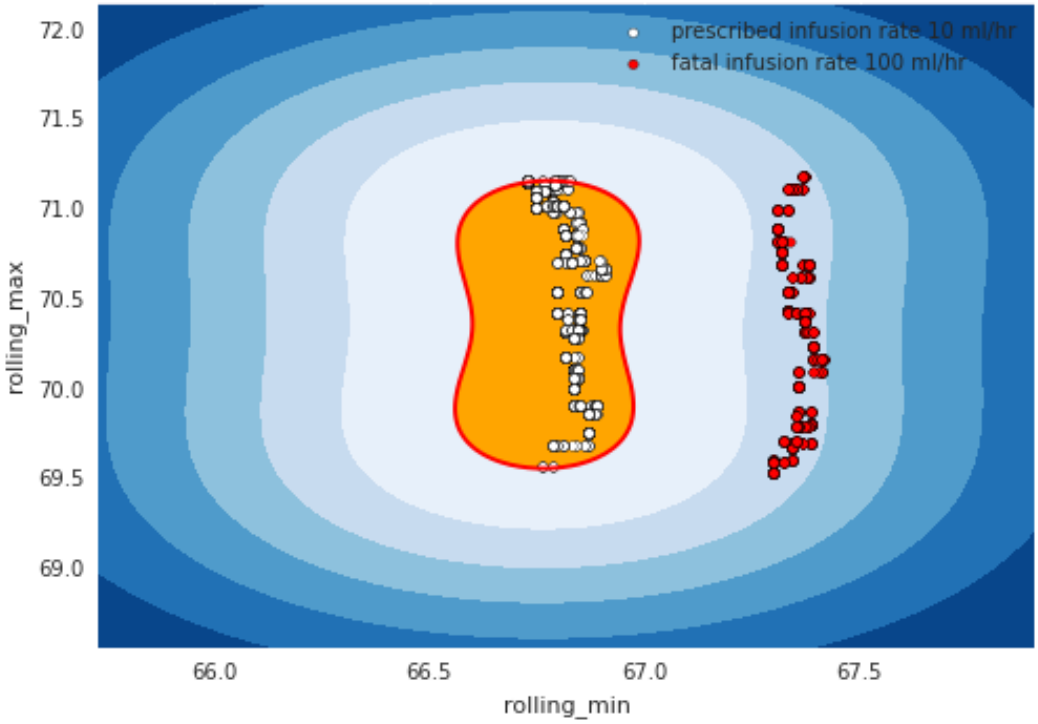

Figure 6: CareFusion Alaris 8100. Normal infusion rate of 10 ml/hr (white dots) compared to abnormal infusion rate of 100 ml/hr (red dots).

Generalization of Results to Hospira Pumps

In order to make our model more resilient to false positives, we explored another machine learning strategy, classification. In contrast to anomaly detection, which trains only on normal data, our classifier model was trained on both normal and abnormal data, including infusion rates of 50 ml/hr, 500 ml/hr and 999 ml/hr. Additionally, we modified one of the features, using a rolling 5th percentile of the 60 Hz amplitude instead of a rolling min. The resulting model is less

sensitive to variations from the cluster of white dots in the normal data group. This model had 0.0% false positives and 0.0% false negatives. Thus, we found that by training with both normal and abnormal data, we were able to delineate the different even more clearly.

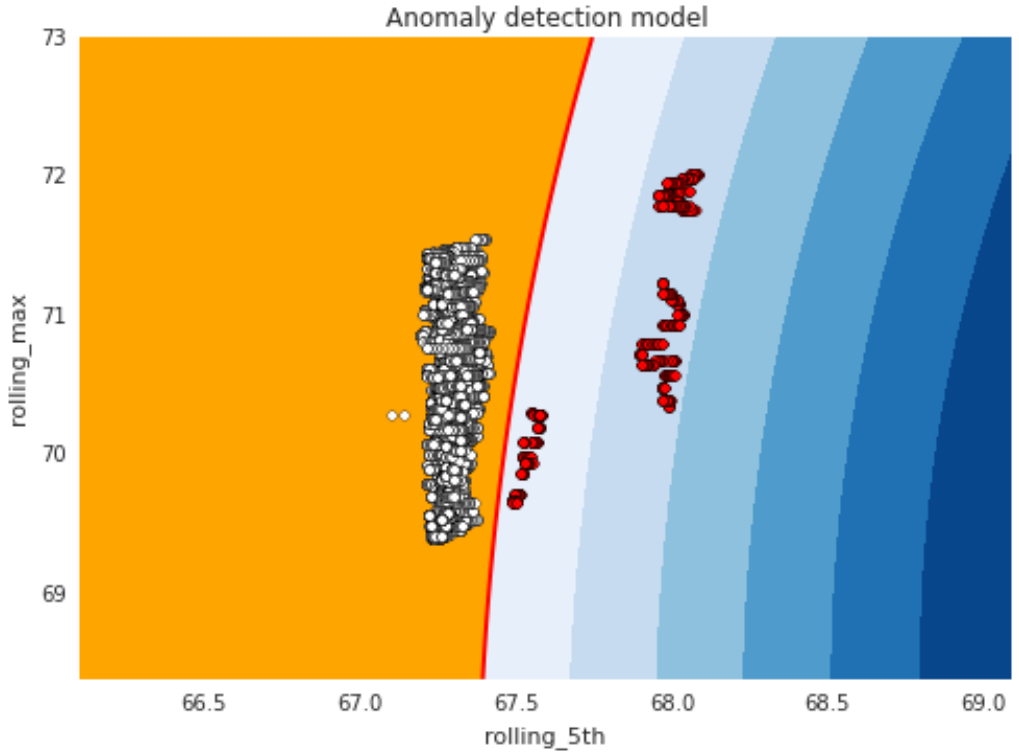

Figure 7: CareFusion Alaris 8100. Training on both normal and abnormal datasets. Normal infusion rate of 10 ml/hr (white dots) compared to the attacker scenario of 500 ml/hr and 999 ml/hr (red dots).

As an example of classifier models we developed, the following figure shows a visualization of the classifier regions with normal (yellow) and abnormal (white to blue) spaces. Each model was trained using a set of low and high infusion rates. After training, a separate set of data, again with both low and high infusion rates, was used for testing. White dots represent normal data at expected low infusion rates, while red dots represent simulated hacker activity at maximum infusion rate. In all three cases, the system correctly identified normal readings and anomalies with zero false negatives and zero false positives. The PowerGuard™ tools automatically built a simple model for the infusion pumps. More complex devices, including general-purpose computing platforms, yield much more complicated models under automated PowerGuard™ model creation.

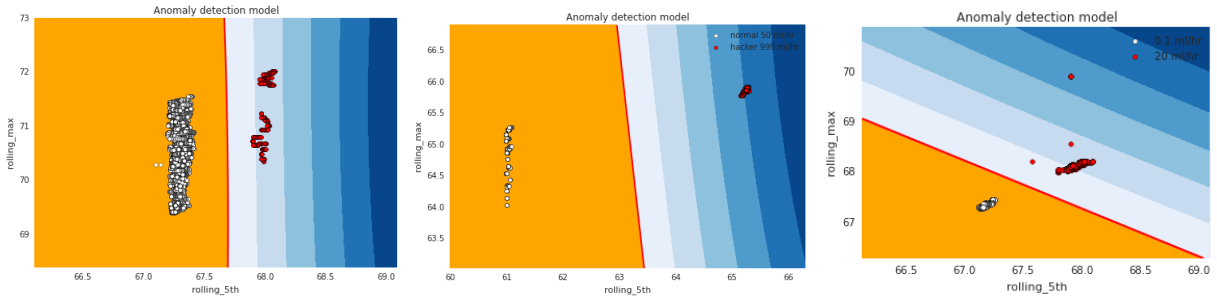

Figure 8: Classifier models for infusion pumps created by training the same PowerGuard™ model with data from several devices. Training data included both low infusion rates and high infusion rates, a technique that lowered false positives to zero in our experiments. Left to right: Alaris 8100, Hospira Plum A+, Hospira PCA.

The quality of our anomaly-detection models is affected by the completeness (coverage) of our device-specific training. An ideal anomaly-detection model has low false-positive and false-negative rates. Training data that covers a diverse range of a device's possible use cases will result in a higher-quality model. Thus, we plan to expand our coverage in two dimensions: more devices, and more behaviors of each device.

Summary of Results

According to our study, infusion pump patterns of AC power consumption at the wall are highly correlated with the infusion rate. Our tools automatically generated a model that discriminated with high confidence among the various infusion rates with low rates of false positives (1.0% and lower) and false negatives (close to 0.0%). PowerGuard™ tools and techniques can be used to detect adversarial interference with normal pumping activity.

Acknowledgements

We would like to thank Michael Holt for his assistance in running the experiments and data collection.

About Virta Labs

Virta Laboratories, Inc. (Virta Labs) detects malware and software execution anomalies on mission-critical devices by only observing signals on the AC power line. Our patent-pending PowerGuard technology builds on years of academic security and machine-learning research on power analysis. PowerGuard constantly measures a protected asset's power consumption patterns for malware and other run-time anomalies and sends alerts to a centralized reporting console. PowerGuard works literally "outside the box" to spot costly problems quickly without interfering with operations.

Contact us for more information:
Virta Laboratories, Inc.
web: https://www.virtalabs.com/
email: info@virtalabs.com
phone: +1 (888) 314-6887
mail: 1327 Jones Drive, Suite 110, Ann Arbor, MI 48105