# Detecting Signal Injection Attack-based Morphological Alterations of ECG Measurements

Hang Cai
*Department of Computer Science*
*Worcester Polytechnic Institute*
*Worcester, MA 01609*
*Email: hcai@wpi.edu*

Krishna K. Venkatasubramanian
*Department of Computer Science*
*Worcester Polytechnic Institute*
*Worcester, MA 01609*
*Email: kven@wpi.edu*

***Abstract*—In this paper, we present an approach to detecting signal injection-based morphological alterations of ECG measurements in Body Sensor Networks (BSN). Signal injection attacks target, the usually unprotected, analog sensing interface of the sensors in a BSN and induce arbitrary signals in them. Signal injection is very dangerous because can be stealthily mounted on unsuspecting BSN users from close proximity (for example in a public place). Inducing morphological alterations in ECG measurements can have profound consequences for the user, as an adversary can easily make a person who is experiencing cardiac arrhythmia appear to be normal and thus cause immediate or long-term harm to their health.**

**To detect signal injection-based morphological alterations, we leverage the idea that multiple physiological signals based on the same underlying physiological process (e.g., cardiac process) are inherently related to each other, i.e., have common features. Any adversarial alteration of one of the signals will not be reflected in the other signal(s) in the group. Therefore, to detect the morphological alterations in ECG measurements, we use arterial blood pressure (ABP) measurements. Both ECG and ABP measurements are alternative representation of the cardiac process. Our approach demonstrates promising results with over 90% accuracy in detecting even subtle ECG morphological alterations for both healthy subjects and those with cardiac conditions.**

## I. Introduction

Body Sensor Networks (BSNs) have demonstrated great potential in a broad range of applications in healthcare and well-being. A BSN contains a number of diverse, low-cost, wireless embedded sensing devices (henceforth referred to as *sensors*) that form a **distributed wireless network** around the user [1]. The sensors monitor various physiological signals from the user and wirelessly forward them to a sink entity, called the *base station*. The base station processes the sensor measurements, displays them to the user, and may forward the measurements to a medical cloud for long-term storage and for access by caregivers.

Unlike traditional hospital-based medical systems, BSNs allow their users to be ambulatory and go about their daily lives. This makes them particularly susceptible to *attacks that can be mounted in proximity (within a few meters) to the user*. Numerous examples of proximity-based attacks have already been reported including triggering malware in a sensor by sending a specific combination of inputs [2], fitness monitors like fitbit being loaded with malware through open Bluetooth ports [3], and signal injection to manipulate sensor measurements [4]. In this work we focus on signal injection attacks. *Signal injection attacks* are sensory-side-channel attacks that target the analog interface of the sensors with the aim of inducing arbitrary signals into the sensing circuitry through electromagnetic induction (EMI). A direct consequence of a signal injection attack is that leads to the sensor generating incorrect measurements of the user's health. In [4], the authors demonstrated the ability of using EMI to induce arbitrary signals into the leads of an electrocardiogram (ECG) sensor. This allowed the adversary to misrepresent a potentially dangerous arrhythmia in the user's heart and report it as being normal. Similar signal injection attacks have been seen in other domains as well such as smartphones. For example, researcher were able to obtain unauthorized access to Apple SIRI and Google Now systems by inducing signals on the headphone chord of the smartphone [5].

Signal injection attacks are especially dangerous for several reasons. (1) They can *introduce backdoors* even in an otherwise secure system by targeting the sensing interface, which is typically not protected [4]. (2) They can be mounted in a *stealthy* manner, for instances, in crowded areas such as malls or concerts. In the context of BSNs, an adversary mounting a signal injection attack affects the capability of the sensors from collecting and forwarding the accurate user health state. A successful signal injection attack can lead to an adversary manipulating sensor measurements, which then might lead to *harm* either from unnecessary medical interventions by users or incorrect diagnosis and treatment by the physician. It is therefore imperative that signal injection attacks be detected as quickly as possible.

In this paper, we want to detect signal injection attacks on ECG sensors in a BSN. We particularly focus on ECG sensors because: (1) ECG is one of the most commonly available sensors in a variety of medical monitoring BSNs, and (2) it has already been demonstrably compromised through signal injection as shown in [4]. In general, an ECG measurement has two key characteristics that can

be manipulated by signal injection attacks *temporal* and *morphological* [6]. The manipulation of temporal characteristics involves modifying the timing information of the ECG complex (e.g., inter-beat-interval), while the manipulation of the morphological characteristics involves modifying the shape of the ECG complex.

In our previous work [7], we proposed an approach that can detect the malicious temporal alteration of the ECG signal. It required building a model by correlating different but correlated signals (arterial blood pressure and respiration) to detect temporal ECG alterations. However, the model developed was limited to features that captured only the timing properties of the ECG signal and did not have the ability to capture morphological alterations. Since an adversary can change the morphological characteristics of ECG signal without changing any temporal characteristics, in this work our **goal** is to design an approach to detect the morphological alterations of ECG measurements.

In this regard, we leverage the fact that different physiological signals generated by the same underlying physiological process are inherently correlated, i.e., they share similar features among them. For example, electrocardiogram and blood pressure are different manifestations of the cardiac process and the two signal time-series track each other. To identify if the ECG sensor's measurement has been morphologically altered, we train a model, at the base station, that captures the commonalities of the ECG measurement with that of another correlated physiological signal, the arterial blood pressure (ABP) signal. Under normal situations, both ECG and ABP, which are measured synchronously, produce features that are not observable when the ECG signal is altered without a corresponding change in the user's physiology, thus indicating alteration. The *advantage of our detection approach* is that it does not require redundant ECG sensors nor does it rely on keeping user medical history. Analysis of our approach demonstrates promising results with over 90% accuracy in detecting even subtle morphological alterations in ECG signals. The **contributions** of this paper are three-fold: (1) the design of an approach for morphological alteration detection in ECG measurements, and (2) implementation of the detector using real ECG and ABP data from the MIT PhysioBank database [8], and (3) demonstration of the robustness of the approach in the presence of a variety of (simulated) signal injection attacks.

The rest of the paper is organized as follows. Section II presents the related work. Section III discusses the system and threat model along with the problem statement. Section IV presents the background for ECG and ABP. Section V presents the main idea of our approach. Section VI presents the parameter selection process for our approach. Section VII presents the security analysis. Finally, Section VIII presents the conclusions. In the rest of the paper, we use the terms *subject* and *user* interchangeably.

## II. Related work

Not much work has been done in the domain of detecting signal injection attacks. In [4] the authors present several preventive solutions for signal injection attacks. However, these solutions require the sensor hardware to be upgraded through improved shielding and adaptive filtering techniques, which is hard to do without increasing the complexity and cost of the limited capability sensors. Therefore, we need a *detection solution* that executes at the base station (which typically has considerably more computational power) and can identify signal injection attacks through analysis of the measurements

Work on detecting anomalous sensor measurements has largely focused on the benign case of fault detection. Fault detection in sensors in a BSN has involved the adaptation of sensor-redundancy-based methods from wireless sensor network domain to the BSNs [9], [10], [11], [12]. However, almost all the approaches are designed for motion and gait monitoring BSNs, and these kinds of BSNs naturally require considerable sensor-redundancies (multiple sensors of the same type). In [13] the authors identify faults in a sensor by correlating its data with different sensors measuring related stimuli. Specifically, the paper focuses on detecting permanent faults in ECG signals based on ventricular pressure signals. Their approach builds a rule-table for various combinations of blood pressure and heart rates and determines if the observed data fall within these expected bounds and, if not, then the sensors are deemed faulty. This approach uses a simple cardiac output model to determine the relationships between heart rate and blood pressure, therefore it does not work if an adversary deliberately replaces legitimate ECG measurements with another signal having similar heart rate characteristics.

Finally, in [14], we presented a very preliminary version of morphological alteration detection approach. In this work, we present a more complete picture with detailed security analysis and performance results.

## III. System Model, Threat Model, and Problem Statement

**System Model:** We assume the BSN consists of a number of wearable medical devices (i.e., *sensors*) including ECG and ABP sensors (See in Figure 1). These sensors are low-capability devices that collect physiological data from the user at regular intervals and forward that data to a highly capable sink entity, which we refer to as the *base station*, for processing and storage. The base station provides a root of trust for our system. We assume it is not susceptible to attacks. Sensors in a BSN typically connect with the base station over a wireless network [15]. For the purposes of this work, we assume the communication between the sensors and the base station is trustworthy, and secure. This can be achieved by deploying any of one of the numerous solutions proposed to address the secure communication problem in a BSN setting such as [15], [16], [17]. Consequently, any

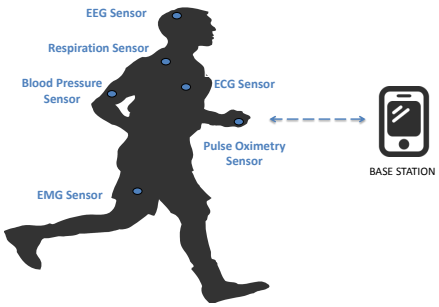



Figure 1: **Body Sensor Network**

alteration of the measurements has to occur at or near the source (i.e., the ECG sensor).

**Threat model:** We assume the adversaries mounting signal injection attacks posses the several characteristics. (1) Adversaries are located in *relative proximity* to the user (victim) and the attacks are mounted locally as opposed to remotely over the Internet. (2) The adversaries are assumed to be able to mount only *scalable attacks*, that is, indiscriminate signal injection through EMI on ECG sensors in the BSN. (3) Due to the non-targeted nature of the attacks, the signal injection attacks are assumed to affect a subset of the sensors in the BSN which *does not* include the ABP sensor. (4) The non-targeted nature of the attacks, also means that we assume the adversary has no prior information on the user including their past medical history or records. Morphological alterations can be implemented using signal injection in three general ways: (1) by introducing arbitrary noise to the original ECG measurements; (2) by replaying historical ECG measurements stolen from the user in the past as current measurements; and (3) by replacing the actual ECG measurements with measurements belonging to another user. In the case of introducing arbitrary noise, the user and their caregivers will immediately be able to see the noise and can therefore ignore the measurements. Replaying historical ECG measurement would require adversaries to access to a user's past medical records, which they do not have. *Consequently, in this paper we focus on the third case, where actual ECG measurements are replaced with measurements belonging to another user.*

**Problem Statement:** Our goal is to develop an approach for detecting alteration of the morphological characteristics of an electrocardiogram (ECG) measurement in a BSN using a synchronously obtained measurement of an *inherently trustworthy* reference sensor, the arterial blood pressure (ABP) sensor.

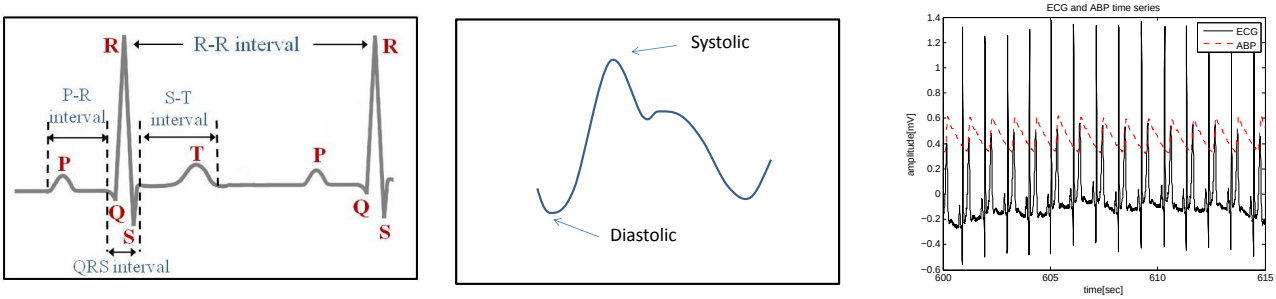
## IV. Background

In this section, we provide some background information on the principal signals that we consider for this work, i.e., electrocardiogram (ECG), arterial blood pressure (ABP) signals. ECG is the measurement of the electrical representation of the cardiac process of a person. As shown in Figure 2a, and ECG signal is made up of peaks and trough combinations which is made up of five elements named P, Q, R, S and T waves. The P wave is observed during atrial depolarization (which causes the blood to be pushed to the ventricles), the QRS complex is observed during the rapid depolarization of the right and left ventricles (which causes the blood to be pushed out of the ventricles and into the lungs and the rest of the body), and the T wave is the depolarization of the ventricles. The time difference between two R peaks is known as an RR-interval. The RR-interval refers to the beat-to-beat variations in heart rate and is a measure of heart rate. Atrial blood pressure (ABP), on the other hand, is the continuous measurement of blood pressure and can be measured non-invasively [18] much like ECG. As shown in Figure 2b, a typical atrial blood pressure contains the trough which is diastolic blood pressure and the peak which is systolic blood pressure. Diastolic troughs occur near the beginning of the cardiac cycle and systolic peaks occurs when the ventricles contract. As ECG and ABP signals are both measures of the cardiac process and both controlled by our autonomic nervous system and they track each other. For example, an R peak in the ECG signal will typically be followed by a systolic peak in the ABP signal as both represent the compression of the ventricles that results in the blood being circulated through the entire body via the Aorta (see Figure 2c). Similarly, a pathologies in the cardiac process that results in abnormal ECG wave form is also reflected in the ABP signal [19].

## V. Approach

In this section, we introduce our approach for the detecting of morphological alterations of ECG sensor measurements because of signal injection attacks. Our approach works by training a user-specific supervised-learning model that includes features that capture the inter-relationship between synchronously measured ECG and ABP signals from a particular user. The model also includes features collected from ABP measured from the user and ECG measured from several different users (thus modeling the attack where a user's ECG is replaced with someone else's as part of the signal injection). Once the model has been trained, we then use features from snippets of synchronously measured ECG and ABP signals from the user and feed it into the model. The model generates an alert if it determines that the signals came from two different users. Figure 3 shows our system setup. Its contains three main steps: (1) extracting features that capture the inter-relationship between ECG and ABP, (2) training a user-specific model, and (3) detecting altered ECG measurements based on the newly received ECG and ABP snippets.

As we are particularly interested in identifying morphological alterations shape of the ECG signal based on the ABP signal, it is essential to be able to capture the shape of the ECG and ABP signal in tandem. Inspired by the idea of phase space reconstruction, which was originally used to de-




(a) P, Q, R, S, T waves in a typical ECG (b) Systolic Peak & Diastolic Trough in ABP (c) ECG and ABP superimposed

Figure 2: ECG and ABP signal

lineate the nonlinear behavior of a dynamic system [20], we generate a *portrait* of the ECG and ABP signals. A portrait allows us to specify the instantaneous state of the the two signals over time. We define a *portrait* as an $n$-dimensional representation of the relationship of several time-series in one multi-dimensional space. To generate a portrait, first, we measure $w$ time-units synchronously measured ECG and ABP signals and normalize them. Normalization is needed as the magnitude and units of ECG and ABP signals are different. Formally, let $a(t)$ and $e(t)$ be the normalized ABP and ECG signals at time $t$, where $1 \leq t \leq w$. Then we create a 2-dimensional portrait, $P$, through the function $f(t) = (a(t), e(t))$, where again $1 \leq t \leq w$. Figure 4 shows an example of a portrait of ECG and ABP signals with annotations of where their characteristic peaks lie in it.

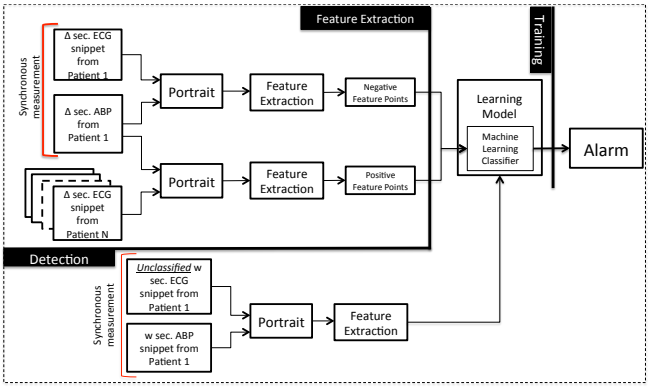



Figure 3: **Detecting Morphological Alterations of ECG Sensor Measurements**

### *A. Feature Extraction*

Once a portrait is created, the next step is to extract appropriate features from the portrait that captures the inter-relationship between the ECG and ABP signals. Based on the work in [21], [22], we extract a total of *eight features*. We categorize these eight features into two different types of features: (1) matrix features (three features) and (2) geometric features (eight features).

**Matrix Features:** Matrix features describe the distribution of points in the portrait that capture the shape of the

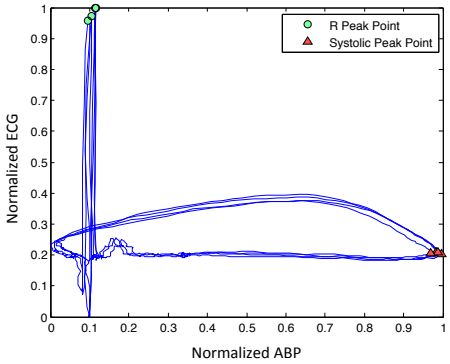



Figure 4: **A typical ECG and ABP portrait**

ECG signal with respect to the shape of the ABP signal. To obtain these features, we view the portrait under an $n \times n$ grid and count the number of points from the portrait that fall into each cell in the grid. We store this information in an $n \times n$ matrix, $C$, in which each element $c(i,j)$ is the count of the number of points in the corresponding grid element $(i,j)$, where $i,j \leq n$. We chose $n = 50$ for generating the matrix $C$. From the matrix $C$, we *extract three features*. (i) *Spatial Filling Index (SPI)*: to obtain the spatial filling index we first generate a matrix S, from matrix C, given by (1).

$$S = \begin{bmatrix} (\frac{c(1,1)}{k})^2 & (\frac{c(1,2)}{k})^2 & \dots & (\frac{c(1,n)}{k})^2 \\ (\frac{c(2,1)}{k})^2 & (\frac{c(2,2)}{k})^2 & \dots & (\frac{c(2,n)}{k})^2 \\ \vdots & \vdots & \ddots & \vdots \\ (\frac{c(n,1)}{k})^2 & (\frac{c(n,2)}{k})^2 & \dots & (\frac{c(n,n)}{k})^2 \end{bmatrix}, k = \sum_{i,j=1}^{n} c(i,j) \quad (1)$$

Each element $S(i,j)$ represents the square of the probability that a portrait point falls into the grid $(i,j)$. Then the spatial filling index can be calculated as $SPI = \sum_{i,j=1}^{N} S_{i,j}$. (ii) *Standard deviation of Column Averages of matrix $C$*: we calculate the column average of matrix $C$ as $H = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} C_{i1} & \frac{1}{n}\sum_{i=1}^{n} C_{i2} & \dots & \frac{1}{n}\sum_{i=1}^{n} C_{in} \end{bmatrix}$. Given $H$ we take the standard deviation of the values of $H$. (iii) *AOC formed by the Column Averages*: this feature is obtained by computing the integral of the curve that $H$ forms.

**Geometric Features:** Geometric features describe the absolute and relative location of certain *characteristic points* of the signals in the portrait. The characteristic points are the points that represent the important characteristics of the signal. For example, the characteristic points can be the important peaks and troughs in the signal, such as P, Q, R, S and T points in the ECG signal, systolic and diastolic points in the ABP signal; or simply the consecutive points near the R peak in the ECG signal and the consecutive points near the systolic peak in the ABP signal. These geometric features captures how well the two signals track each other, which they should given the emanate from the same physiological process. In this work we only use geometric features-based on two characteristic points, the R peaks in the ECG signal and the systolic peaks in the ABP signal as they were sufficient distinguishing power by effectively capturing the state of the two signals.

To identify where the characteristic points lie in the portrait, we first perform peak detection for each $w$ seconds synchronously measured ECG and ABP signals, to get both R peaks and systolic peaks and label them. Note that, depending upon duration of $w$, the portrait can have multiple characteristic points from ECG and ABP in it. Overall, we extract *five geometric features* based on the labeled characteristic points in the portrait. (i) *Average of the Angles (w.r.t. x-axis) between R peaks*: let $(x_r, y_r)$ denote an R peaks in a portrait and $m$ denote the total number of R peaks in a portrait, then the average of the angles for the R peaks is given by $\frac{1}{m}\sum_{i=r_1}^{r_m} arctan((x_i, y_i))$. (ii) *Average of the Angles (w.r.t. x-axis) between Systolic peaks*: let $(x_s, y_s)$ denote the systolic peaks in a portrait and $q$ denote the total number of systolic peaks in a portrait, then the average of the angles for the systolic peaks is given by $\frac{1}{q}\sum_{i=s_1}^{s_q} arctan((x_i, y_i))$. (iii) *Average Distance between R peaks and the Origin*: let $(x_r, y_r)$ denote an R peaks in a portrait and $m$ denote the total number of R peaks in a portrait, then average distance is given by $\frac{1}{m}\sum_{i=r_1}^{r_m}\sqrt{(x_i^2 + y_i^2)}$. (iv) *Average Distance between Systolic peaks and the Origin*: let $(x_s, y_s)$ denote the systolic peaks in a portrait and $q$ denote the total number of systolic peaks in a portrait, then the average distance is given by $\frac{1}{q}\sum_{i=s_1}^{s_q}\sqrt{(x_i^2 + y_i^2)}$. (v) *Average Distance between the R peaks and their corresponding Systolic peaks*: typically, there is one systolic peak is preceded by one R peaks, therefore, for a R peak $(x_{r_i}, y_{r_i})$, let $(x_{s_i}, y_{s_i})$ denote the corresponding systolic peak, where $1 \leq i \leq m$. Then, the average distance between the R peaks and their corresponding systolic peaks is given by $\frac{1}{m}\sum_{i=1}^{m}\sqrt{(x_{r_i} - x_{s_i})^2 + (y_{r_i} - y_{s_i})^2}$

### B. Model Training

We use a supervised-learning-based approach to construct (train) the user-specific model, which requires as input two classes of 8-dimensional feature points referred to as negative and positive class points. The *negative class points* capture the situations where the ECG and ABP originate from the same user, while the *positive class points* capture the situations where the ECG of the user is replaced with someone else's. In this regard, we collect $\Delta$ time-units of synchronously measured ECG and ABP signals from the user whose model we are training. The feature extraction for the negative class points is done by sliding window of size $w < \Delta$, over the time-series of the ECG and ABP signals. Each $w$ sized window of data produces one portrait, and one 8-dimensional feature point is then generated from this portrait. To generate positive class points, we build portraits using $\Delta$ minutes snippets the user's ABP and ECG belonging to several different users, and then extract 8-dimensional features by sliding a window of size $w$ over the time-series. Once the negative and positive class points are collected, we feed them into a machine learning classifier to generate a user-specific model. In this paper, we used a polynomial kernel-based Support Vector Machine (SVM) as our machine learning classifier. We chose SVM because it is well understood, relatively easy to understand, and has excellent tool support, properties that are essential for implementation of limited capability base stations where the model will be executing.

### C. Generating Alerts

Once the model is trained, we can then use the trained model to decide if any newly received snippet of ECG measurements have been maliciously altered in the morphological sense. In this regard, we collect $w$ time-units of newly measured ECG and ABP signals from the user, generate a portrait and extract the 8-dimensional feature from it. Then we feed this feature point into the user-specific model. The model will then output a positive or negative label for this feature point. If the feature point is deemed to be positive, we consider this $w$ second ECG signal snippet to be altered and alert the user.

### D. Model Retraining

Our approach relies on the inter-relationship between ECG and ABP signals to operate. If a patients physiology changes over time, the models have to adapt as well. In our current design, the model is trained in an offline fashion with only the alert generation happening online. This means that the model has to be re-trained every so often in order to capture the current state of the patients health. One approach is to automate the re-learning based on a schedule. However, choosing the inter-relearning interval has to be done carefully. Too short an interval would lead to unnecessary re-learning and too long an interval would result in increased errors. Determining the optimal model re-training frequency for our work is probably a user dependent parameter. For relatively healthy users the retraining need not happen often, while for individual cardiac conditions, the training has to be done more frequently depending upon the actual condition, how acute it is, and any medications they might be taking. The calculation of optimal model retraining

is a non-trivial problem in its own right and out of scope for this paper.

## VI. Parameter Selection

In this section, we illustrate how we select the two most important parameters of our system $\Delta$, the amount of data needed to train the model (i.e., *training time*), and $w$, the amount of data needed to test for malicious alteration of ECG signals (i.e., *testing time*). We begin with a discussion our dataset, followed by performance metrics for identifying how well we are performing for various parameter choices. Finally, we discuss the parameter selection itself.

### A. Dataset

In this work, we collected data belonging to 26 subjects (i.e., users) from the MIT PhysioBank Fantasia and MGH/MF database [8]. We chose these particular subjects from these databases because the availability of both ECG and ABP signals for them. Furthermore, the Fantasia database is made up of healthy subjects, while the MGH/MF database mainly contains data from subjects with specific cardiac conditions. We searched the MGH/MF database to specifically choose subjects whose cardiac condition manifested itself in morphological variation in the measured ECG signal (e.g., atrial fibrillation, ectopic beats etc.). Table I shows the statistics on the patient population we used to train and test our ECG morphological alteration detectors. We categorized the patients in the dataset into two types based on their ECG signals: (1) *Normal* subject, which only includes subjects who did not suffer from any ailments and had a normal sinus rhythm ECG; (2) *Abnormal* subject, which only includes subjects whose ECG showed morphological abnormalities. For each of the 26 subjects we had on average about 41 minutes of usable ECG and ABP data. From these 26 subjects, we picked 18 subjects (9 normal subjects and 9 abnormal subjects) to form a group $G$, referred to as the *training group*. The data for the subjects in the training group is used for training and validation of the model. We used the remaining 8 patients (3 normal subjects and 5 abnormal subjects) to form a group $H$, referred to as *external group*, which is used for testing our model's detection capabilities.

Table I: Subject Data Summary

| Type | Total # | Male | Female | Avg. Age (years) | Std. Age (years) |
|---|---|---|---|---|---|
| Normal | 12 | 5 | 7 | 46.5 | 25.5 |
| Abnormal | 14 | 8 | 6 | 73 | 7.9 |

### B. Metrics

We use the following metrics to train our model and validate it: false positive rate, false negative rate, and balanced accuracy rate. We define *false positive rate* (FP) as the fraction of the cases in which an unaltered ECG sensor output is misclassified as altered. Similarly, we define *false negative rate* (FN) as the fraction of the cases where an altered ECG sensor output is misclassified as unaltered. We define *balanced accuracy rate* (BAC) as the sum of half of the true negative rate (the fraction of the unaltered ECG sensor output properly classified as unaltered) and half of the true positive rate (the fraction of the altered ECG sensor output properly classified as altered). The reason we use BAC is that it avoids inflated performance estimates on imbalanced datasets [23]. This is important given that we have an imbalanced sample with many more positive examples than negative examples during the training phase. Even though we compute these metrics for each subject in our dataset, we validate our approach using summary statistics of these metrics over all subjects.

### C. Selecting Training and Testing Times

Given the safety-critical nature of signal injection attacks on BSNs, we want to detect potential alterations to the ECG signals very quickly in order to avoid harm to the user. Therefore we need to minimize $w$ which establishes the time elapsed between the alteration of the measurement and its detection. Further, even though the training is offline, we need to make sure that it can be done quickly so that the user need to endure long interruptions in BSN operation. Consequently, the training time ($\Delta$) has to be as short as possible as well.

To see which window size, $w$, works best, we set $\Delta$ to a fixed value of 20 minutes and test our ECG morphological alteration detector with the data from users in the training group. We tried several values for $w$ and eventually settled on 3, 6, and 9 seconds. This is because values smaller than 3 seconds did not capture each meaningful features from the portraits produced, and values of 10 seconds and greater were considered too slow for detecting safety-critical attacks. For each window size, $w$, we generated a set of negative class feature points by using 20 minutes of synchronously measured ECG and ABP signal from the same subject. We generated the set of positive class feature points by combining each subject's ABP snippet with a randomly selected ECG snippet from every other subject in the training group. The resulting negative and positive class feature points were then fed into SVM classifier for training purposes. We used 10-fold cross-validation to validate the model trained. Table II shows the average BAC, FP, and FN for different window sizes, $w$. We can see that the average BAC rate of the subject-specific models for three different window size $w$ are all considerably high. Overall, we can see that the balanced accuracy rate of the ECG morphological alteration detector when we set $w$ from 3 seconds to 9 seconds, is only slightly worse off. We therefore chose $w = 3$ seconds, as it provided us with the best responsiveness for our detector.

To select the training time, $\Delta$, we evaluated our detector by training it for 3 different durations: 10, 15, and 20 minutes (with a fixed $w = 3$ seconds). Again, for each value of $\Delta$, we generated a set of negative class and positive class

Table II: Balanced Accuracy Rate for Different $w$ with fixed $\Delta = 20mins$

| Window Size $w$ | Avg. FP | Avg. FN | Avg. BAC |
|---|---|---|---|
| 3 secs | 5.33% | 5.11% | 94.78% |
| 6 secs | 5.64% | 1.73% | 96.32% |
| 9 secs | 6.37% | 1.02% | 96.31% |

feature points using data from group $G$. Then we used SVM classifier to train the subject-specific model for each patient in group $G$, respectively. We used 10-fold cross-validation to validate the model built. Table III shows the average BAC, FP and FN for different values of $\Delta$. Overall the difference in accuracy for the three different $\Delta$ value is 1.5% when we move from 10 minutes to 20 minutes, however FP and FN is comparably low when we set $\Delta$ as 20 minutes. Further, the more data we have the better models we can create overall. Hence, we set $\Delta = 20$ while training our model.

Table III: Balanced Accuracy Rate, FP and FN for Different $\Delta$ with fixed $w = 3secs$

| Training Data $\Delta$ | Avg. FP | Avg. FN | Avg. BAC |
|---|---|---|---|
| 10 mins | 9.44% | 2.63% | 93.96% |
| 15 mins | 5.37% | 3.61% | 95.51% |
| 20 mins | 5.33% | 5.11% | 94.78% |

Figure 5 and 6, show the box-plots for balanced accuracy (BAC), false positive (FP) and false negative (FN) rates when we performed 10-cross-validation of the detector of the morphological alteration of ECG data using user data from the set $G$ given $\Delta = 20$ minutes and $w = 3$ seconds. We see that the Abnormal subject group has a slightly higher spread compared to the Normal subject group with respect to the reported BAC, FP and FN. This is reasonable as the Normal subject group consists only of the subjects with a Normal Sinus Rhythm, however, the patients in Abnormal subject group have various types of ECG signals. For *Normal* subjects, our detector provides a 96.41% BAC on average with an average false positive rate and an average false negative rate at 6.39% and 0.79%, respectively. Not surprisingly the performance degrades slightly when we consider subjects with cardiac conditions. For *Abnormal* subjects, the average BAC rate is 93.56% with an average false positive rate and an average false negative rate at 7.22% and 5.65%, respectively. Overall, the validation results show that the model trained by the morphological alteration detector is very accurate with a 94.78% BAC on average with an average false positive rate of 5.33% and an average false negative rate of 5.11%. Given the model, we next present the security analysis of our approach by simulating various signal injection attacks on the ECG sensor that results in morphological alteration of the its measurements.

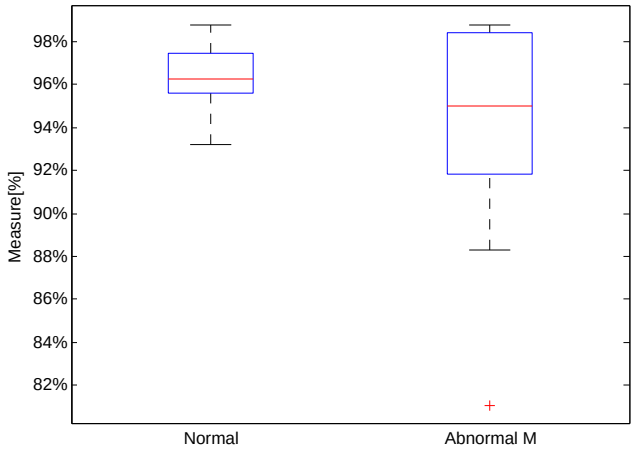



Figure 5: **Validation of ECG Morphological Alteration Detection w.r.t. BAC**

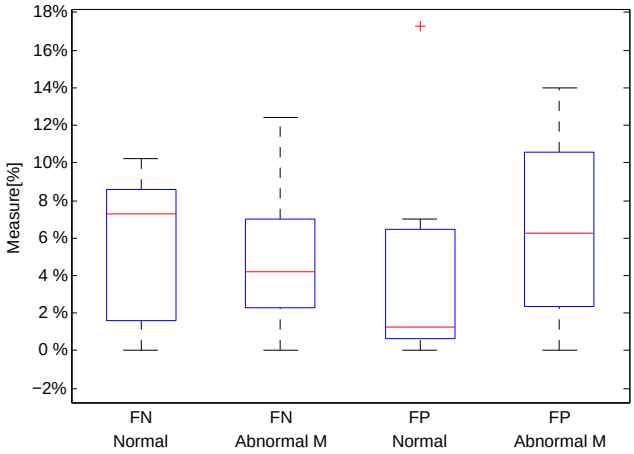



Figure 6: **Validation of ECG Morphological Alteration Detection w.r.t. FP and FN rates**

## VII. Security Analysis

In this section, we first address the viability of our approach in detecting signal injection attacks on ECG sensors that result in the morphological alteration of the ECG measurement. As mentioned earlier in the threat model, with respect to signal injection attacks we are focusing on detecting the scenarios where adversaries replace the original user's measurements with that of another user. In this regard, we attacked the 18 user-specific models we trained. Our approach works through analysis of ECG measurements at the base station. We therefore simulated the signal injection attack on a sensor in two ways. (1) We simulated a signal injection attack that resulted in the replacement legitimate ECG measurements of a user with that of another user in the training group (i.e., users in the set $G$). (2) Here we simulated a signal injection attack that resulted in the replacement legitimate ECG measurements of a user with that of users in the external group (i.e., users in the set $H$).

Before we delve into the attack details, we introduce two key notation used in this section. We define $T_{learn}$ as a time interval for which we collected ECG and ABP data from a user to build their detection model. The duration of $T_{learn}$ is the same as the training time $\Delta$, i.e., 20 minutes. $T_{curr}$ is the current time interval when we are utilizing our approach to detect ECG measurement alterations. For our analysis we set $T_{curr}$ to 15 minutes and we assume that an adversary alters the entire $T_{curr}$ interval of ECG measurement. Note that, since our testing time is only 3 seconds, as long as the adversary replaces 3 seconds or more of actual ECG measurement, we should be able to detect it.

### A. Evaluating the accuracy of the model

We first tested our approach to see if it can correctly classify a user's ECG signals after the user-specific model is trained. This is very important to ensure that our trained user-specific model is accurately able to identify yet unseen data from the same user. Therefore, for each user, we obtained $T_{curr} = 15$ minutes of synchronously ECG and ABP signals. The two resulting signal time series were then divided into 300, 3-second intervals, each of which produced a portrait and one 8-dimensional feature point. These 300 feature points were then input into the model, which then labeled them as positive or negative. Ideally, we should get all negative labels for the points, as they are from the same user. Overall, when average over the 18 user-specific models, our approach achieved an average detection accuracy rate of 97.37% in detecting unmodified ECG data, which demonstrates that our approach indeed has a low false alarm rate. This result also shows that in most cases the ECG and ABP signals did not change over time demonstrating that the model learned in an accurate representation of the user's cardiac process.

### B. Replacement using Measurement from Training Group

We next consider the case where the attacker modifies $T_{curr}$ duration of a user's ECG time-series with ECG measurements from another user in the training group $G$. This experiment simulates the case where the adversary has access to the ECG measurements of users in the training group. Thus, for a given user, we replaced each of the consecutive 3 second ECG snippets in their $T_{curr}$ with a randomly selected 3-second ECG measurement snippet obtained from a randomly selected user in the training group. The modified ECG measurement was then fed into the user-specific model along with the legitimate (i.e., unmodified) ABP signal measured in $T_{curr}$ for the user. The model then produced a label for each 3-second ECG snippet of the modified ECG measurements. In aggregate over all our 18 user-specific models, our approach achieved an average detection accuracy rate at 93.79%. This demonstrates that even if an adversary has access to the ECG data of the subjects from the training group $G$, our approach can still detect it with considerable accuracy.

### C. Replacement using Measurement from External Group

We then consider the case where the attacker is modifies $T_{curr}$ duration of a user's ECG time-series with ECG measurements from another user in the external group $H$. This experiment simulates the case where the adversary replaces legitimate ECG measurements with those of yet unseen users. Thus, for a given user, we replaced each of the consecutive 3 second ECG snippets in their $T_{curr}$ with a randomly selected 3-second ECG measurement snippet obtained from a randomly selected user in the external group. The modified ECG measurement was then fed into the user-specific model along with the legitimate (i.e., unmodified) ABP signal measured in $T_{curr}$ for the user. The model then produced a label for each 3-second ECG snippet of the modified ECG measurements. In aggregate over all our 18 user-specific models, our approach achieved an average detection accuracy rate at 90.09%. This demonstrates that even an adversary has access to the ECG data of the subjects from the external group $H$, our approach can still detect it with considerable accuracy. It is not surprising that the performance of this case is a little worse than the previous case, because the adversary is trying to feed heretofore-unseen ECG measurements into the model. However, the detection accuracy loss is little, demonstrating the robustness of our model and our approach.

Table IV summarizes the performance of our ECG signal injection detector.

Table IV: Summary of Security Analysis Performance

| Attack Scenario | Detection Rate |
|---|---|
| No Replacement | 97.73% |
| Replacement w/ Training Group Measurements | 90.09% |
| Replacement w/ External Group Measurements | 93.79% |

## VIII. Conclusions

In this paper, we presented a novel approach to detect signal injection-based morphological alteration of ECG measurements in a Body Sensor Networks (BSN). Our approach leveraged the similarity of ECG to another signal that measure the cardiac process, arterial blood pressure signal, and building a model for it. Analysis of our approach demonstrated promising results with 90% accuracy in detecting ECG morphological alterations. In the future, we plan to extend this work in several directions. (1) To build a combined detector for both temporal and morphological alterations of ECG measurements. (2) To implement the detectors on an actual base station platform such as the amulet system [24] and evaluate the performance and computational cost. (3) To find optimal ways to alert users into action as a result of detecting signal injection attacks.

## REFERENCES


[1] M. Kermani, M. Zhang, A. Raghunathan, and N. Jha, "Emerging frontiers in embedded security," in *VLSI Design and 2013 12th International Conference on Embedded Systems (VLSID), 2013 26th International Conference on*, Jan 2013, pp. 203–208.

[2] A. Uluagac, V. Subramanian, and R. Beyah, "Sensory channel threats to cyber physical systems: A wake-up call," in *Communications and Network Security (CNS), 2014 IEEE Conference on*, Oct 2014, pp. 301–309.

[3] "'10-second' theoretical hack could jog Fitbits into malware-spreading mode," http://www.theregister.co.uk/2015/10/21/fitbit_hack/.

[4] D. Foo Kune, J. Backes, S. S. Clark, D. B. Kramer, M. R. Reynolds, K. Fu, Y. Kim, and W. Xu, "Ghost Talk: Mitigating EMI signal injection attacks against analog sensors," in *Proceedings of the 34th Annual IEEE Symposium on Security and Privacy*, May 2013. [Online]. Available: https://spqr.eecs.umich.edu/papers/fookune-emi-oakland13.pdf

[5] "Hackers show Google Now and SIRI can be hacked via Radio waves," https://www.hackread.com/how-to-hack-google-now-and-siri/.

[6] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *Biomedical Engineering, IEEE Transactions on*, vol. 50, no. 3, pp. 289–294, 2003.

[7] H. Cai and K. K. Venkatasubramanian, "Detecting malicious temporal alterations of ECG signals in body sensor networks," in *Network and System Security*. Springer, 2015, pp. 531–539.

[8] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[9] K. Duk-Jin and B. Prabhakaran, "Motion fault detection and isolation in body sensor networks," *Pervasive and Mobile Computing*, vol. 7, no. 6, pp. 727–745, 2011.

[10] D.-J. Kim, M. H. Suk, and B. Prabhakaran, "Fault detection and isolation in motion monitoring system," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 5234–5237.

[11] H. Sagha, J. del R Millan, and R. Chavarriaga, "Detecting and rectifying anomalies in body sensor networks," in *2011 International Conference on Body Sensor Networks*, 2011, pp. 162–167.

[12] S. Galzarano, G. Fortino, and A. Liotta, "Embedded self-healing layer for detecting and recovering sensor faults in body sensor networks," in *Systems, Man, and Cybernetics, 2012 IEEE International Conference on*, Oct 2012, pp. 2377–2382.

[13] A. Mahapatro and P. M. Khilar, "Fault diagnosis in body sensor networks," *International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM)*, vol. 5, pp. 252–259, 2013.

[14] H. Cai and K. K. Venkatasubramanian, "Poster: Detecting malicious morphological alterations of ECG signals in body sensor networks," in *ACM/IEEE International Conference on Information Processing in Sensor Networking*, 2015.

[15] K. K. Venkatasubramanian, A. Banerjee, and S. K. S. Gupta, "PSKA: Usable and secure key agreement scheme for body area networks," *Trans. Info. Tech. Biomed.*, vol. 14, no. 1, pp. 60–68, Jan. 2010.

[16] A. Banerjee, S. K. S. Gupta, and K. K. Venkatasubramanian, "PEES: physiology-based end-to-end security for mhealth," in *Proceedings of the 4th Conference on Wireless Health*, ser. WH '13, 2013, pp. 2:1–2:8.

[17] M. Rostami, A. Juels, and F. Koushanfar, "Heart-to-heart (h2h): authentication for implanted medical devices," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, ser. CCS '13, 2013, pp. 1099–1112.

[18] "The clearsight system," http://www.edwards.com/eu/products/mininvasive/pages/clearsightsystem.aspx, accessed: 2016-02-09.

[19] "Abnormal EKGs and Corresponding Arterial Waveforms," http://www.dynapulse.com/educator/WebCurriculum/Chapter\%203/Abnormal\%20EKG\%20and\%20Waveform.htm.

[20] M. S. M. Krishnan, D. N. Dutt, Y. Chan, and V. Anantharaman, "Phase space analysis for cardiovascular signals," in *Advances in Cardiac Signal Processing*. Springer, 2007, pp. 339–354.

[21] T. Rocha, S. Paredes, P. de Carvalho, J. Henriques, and M. Antunes, "Phase space reconstruction approach for ventricular arrhythmias characterization," in *Engineering in Medicine and Biology Society, 2008. 30th Annual International Conference of the IEEE*. IEEE, 2008, pp. 5470–5473.

[22] O. Malgina, J. Milenkovic, E. Plesnik, M. Zajc, and J. F. Tasic, "ECG signal feature extraction and classification based on R peaks detection in the phase space," in *GCC Conference and Exhibition (GCC), 2011 IEEE*. IEEE, 2011, pp. 381–384.

[23] D. R. Velez, B. C. White, A. A. Motsinger, W. S. Bush, M. D. Ritchie, S. M. Williams, and J. H. Moore, "A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction," *Genetic epidemiology*, vol. 31, no. 4, pp. 306–315, 2007.

[24] J. Sorber, M. Shin, R. Peterson, C. Cornelius, S. Mare, A. Prasad, Z. Marois, E. Smithayer, and D. Kotz, "An Amulet for trustworthy wearable mHealth," in *Workshop on Mobile Computing Systems and Applications (HotMobile)*, February 2012, pp. 7:1–7:6.