WPI-CS-TR-14-04 July 2014

SIFT: Multi Physiological Signal Feature Correlation-based Sensor Compromise Detection in Body Sensor Networks

by

Hang Cai and Krishna K. Venkatasubramanian

Computer Science
Technical Report
Series

WORCESTER POLYTECHNIC INSTITUTE

Computer Science Department
100 Institute Road, Worcester, Massachusetts 01609-2280

SIFT: Multi Physiological Signal Feature Correlation-based Sensor Compromise Detection in Body Sensor Networks

Hang Cai, Krishna K. Venkatasubramanian
Department of Computer Science
Worcester Polytechnic Institute
Worcester, MA, 10609
{hcai, kven}@wpi.edu

ABSTRACT

A Body Sensor Network (BSN) consists of a set of sensing devices deployed on a user (patient) typically for health monitoring purposes.The fact that BSNs provide sensitive information to caregivers about the user's health makes them attractive targets for tech-criminals to exploit. In this work, we focus on BSN sensor compromise detection especially in the cases physiological signal monitoring sensors. A compromised sensor may generate erroneous data which may cause the incorrect interpretation of user health leading to wrong diagnosis and treatment. Detecting sensor compromise is a non-trivial task in BSNs. Traditional sensor compromise detection relies on one of two general techniques: (1) using redundant sensors and some form of voting to determine sensor accuracy, or (2) using past sensor values for predicting the expected current sensor value range. However, in a BSN due to usability limitations we cannot expect the presence of redundant physiological sensors of the same kind. Further, the dynamic nature of the human body precludes the reliance on historical sensor values for predicting the current range of values.

In this paper, we present *SIFT*, a novel methodology to address the problem of sensor compromise without relying on either redundant sensors or historical sensor values. SIFT leverages the fact that in BSNs physiological signals based on the same underlying physiological process (e.g., cardiac process) are inherently correlated i.e., they share similar features with each other. Given this group of correlated signals, any unnatural alteration in one of the correlated signal will not be reflected in the other signals in the group. As SIFT uses signals from multiple sensors it does not require node redundant sensors. Further, as the correlated signals are measured in a synchronous fashion, the current state of the patient's physiology is automatically taken into account, thus overcoming a typical problem with relying on techniques that use past sensor values. We illustrate the operation of SIFT through a case study where we detect the compromise of a electrocardiogram (ECG) sensor using two related signals arterial blood pressure (ABP) and respiration (RESP) sensors as reference. Analysis of our case study demonstrates promising results with over 98% accuracy in detecting even subtle ECG signal alterations for both healthy and unhealthy patients.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection

1. INTRODUCTION

Emerging Body Sensor Networks (BSNs) have demonstrated great potential in a broad range of applications in healthcare and wellbeing. A BSN consists of a set of monitoring sensors deployed on a user, that continuously monitor them and provides this information to a sink entity called the base station for processing and visualization. BSNs intend to improve health outcomes, decrease isolation, reduce health disparities, and substantially reducing costs.

The fact that BSNs collect sensitive data and provide valuable information to caregivers and patients about their health makes them attractive targets for tech-criminals to exploit. As BSNs become increasingly available and liable to be used outside lab settings and care facilities, the threats posed by malicious entities also increases [17]. One such threat is *sensor compromise*, which we define as the unauthorized modification of the sensor firmware. One can easily imagine a situation seen in today's Internet to manifest themselves as a result of sensor compromise. Examples include theft of a person's sensitive health information, ransomware on BSN sensors that only allows the system to function if a fee is paid, or surreptitious alteration of the collected health data over time leading to incorrect interpretation of patient state, wrong diagnosis and treatment, or patient-alarm suppression. BSN sensors being (1) largely wearable, (2) likely shared between users, and (3) removed from the body on a predictable schedule (during showers or at night), are not too difficult to physically compromise, especially for an advanced-persistent adversary. Compromise of the sensor need not even require physical access to the sensors. A poorly designed BSN that allows remote updating of the sensor firmware without appropriate authentication mechanisms can produce the same result. In this work, we focus on sensor compromise detection especially in the cases where the data is being tampered with. We do not consider privacy-loss or ransomware in this work.

Sensor compromise detection can be thought of a detecting anomalous sensor behavior. Recent years have seen considerable work in the domain of anomaly detection in BSNs. These approaches have tried to adapt sensor-redundancy-based methods for detecting faulty sensors in BSNs [3, 4, 8, 15]. Not surprisingly, these solutions have been proposed for BSNs are designed for motion and gait detection. A motion detection BSN naturally requires considerable redundancies in terms of the sensor deployment. Useful as these solutions are for detecting problems with motion sensors, they might not work when we consider physiological sensors in a BSN. This is because, for usability reasons typically there is only one physiological sensor of a particular type in a BSN. Alternatively, behavior-based anomaly detection approaches have also been proposed where if the current physiological signal (i.e., measurement) generated by the sensor is very different from what it has been providing historically, it is considered a compromise [18]. However, the human body is too dynamic for the past to effectively determine the current state at all times. We therefore our *design goals* is to design a method for detecting sensor compromise in a BSN without expecting (1) the presence of redundant sensors of the same kind, and (2) the historical sensor values to be correlated with the current values at all times.

In this regard, we present *SIFT*; a novel methodology for sensor compromise detection that leverages the fact that in BSNs physiological signals based on the related underlying physiological process are inherently correlated, i.e., they share similar features among them. To identify if a sensor has been compromised, SIFT observes its generated physiological signal (called *candidate signal*) and compares it with correlated physiological signals (called *reference signals*) generated by a set of distinct sensors. In normal situations, both candidate and reference signals produce features that are very similar to each other. However, any alteration of the candidate signal (due to compromise) results in features that are not observed in one or more of the reference signals, thus indicating compromise. The fact that the candidate and reference signals are generated from distinct sensors eliminates the need for redundant sensors in SIFT. Further, as both candidate and reference signals are measured in a synchronous fashion, the current state of the user's physiology is automatically considered every time we execute SIFT making it adaptive in nature.

We illustrate the capabilities of SIFT by instantiating it. We have developed a SIFT-based approach to detect compromise of a electrocardiogram (ECG) sensor, where sensors measuring arterial blood pressure (ABP) and respiration (RESP) signals are used as reference. We chose ABP because it is a measure of the same physiological cardiac phenomena as ECG. Consequently, the inter-beat intervals in ECG and inter-systolic peak intervals in ABP are high correlated. Further, both ABP and RESP signals affect the ECG inter-beat intervals through the autonomic nervous system. Therefore, any alteration of the temporal properties of ECG time-series (inter-beat intervals) by an adversary can be detected by *not* observing a commensurate change in the ABP and RESP signals. Analysis of our case study demonstrates promising results with over 98% accuracy in detecting even subtle ECG signal alterations for both healthy and unhealthy patients. The contributions of this paper are two fold: (1) design of the SIFT methodology for adaptive, sensor-redundancy-free

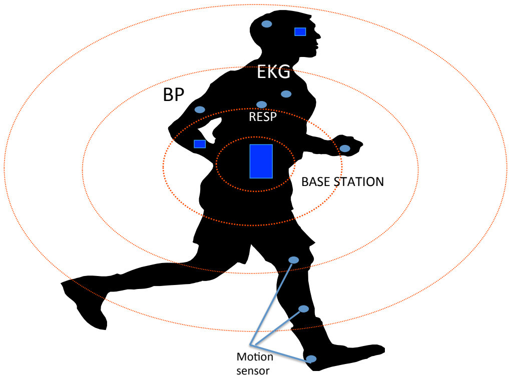


Figure 1: Body Sensor Network

The rest of the paper is organized as follows. Section 2 provides some background information. Section 3 presents our main idea behind SIFT. Section 4 presents a case study that illustrates how SIFT will work in practice. Section 5 presents the performance and security analysis of the case study. Section 6 discusses some of the implications of this work. Finally, Section 7 and 8 present the related work and conclusions, respectively.

2. SYSTEM MODEL

For this work we assume the BSN is comprised of a number of wearable sensing devices (referred to as *sensors* in the rest of the paper) capturing and collaboratively processing physiological signals on users. Particularly, we assume to have electrocardiogram (ECG), respiration (RESP), and blood pressure (ABP) sensors in the network. These devices collect health and contextual data at regular intervals and forward it over a *single-hop* network to a highly capable base station for further processing. Moreover, we assume there is at most one sensors of a kind. We refer to the user on whom the BSN is deployed as the *patient*. In the rest of the paper, we use the term sensors and devices interchangeably.

In terms of the *threat and trust* model, we assume that communication between the sensors and the base station (usually wireless) is fault-free, trustworthy, and secure and uses any of myriad schemes available for this purpose such as [16]. The individual sensors however can be compromised by an external adversary. Further, the channel between each sensor and the base station is assumed to be secure using key distribution techniques such as [1]. Once a sensor is compromised, it may generate erroneous data at any time. We assume that if adversaries compromises a specific sensor they cannot compromise the sensors that measure signals that are correlated with it. For simplicity of discussion in the rest of the paper we use the term sensor to mean a sensing device and assume each sensing device in the BSN measures only one type of physiological signal. Consequently, in the rest of the paper we present the notions of sensor compromise and alteration of the physiological signal it generates (i.e., measures) in an interchangeable manner.

2.1 Problem Statement

The goal of this paper is to propose a method for sensor compromise detection in BSNs using correlated physiological signals. Formally speaking, let p be the signal the adversary is trying to alter (by compromising the sensor measuring it). Then the goal of this work is to find a means of detecting if this signal is being altered to p', solely-based on a set of unaltered signals $Q = \{q_1, q_2, ...q_n\}$ such that, each q_i, where $1 \leq i \leq n$ is correlated with p, that is, it shares certain common features with the p, either in the time or frequency-domain or both.

3. SIFT FOR SENSOR COMPROMISE DETECTION

In this section we present the SIFT methodology, which aims to identify if the physiological signal from a sensor in the BSN has been altered from its actual values. SIFT works by extracting characteristic features from the candidate signal and checking how coherent they are with the features obtained from one or more synchronously measured reference signals. The reference signals are chosen such that they (1) emanate from the same or a related underlying physiological process as the candidate signal and/or (2) the process generating the reference signal has a direct influence on the candidate signal. Assuming the sensors measuring the reference signals are not compromised, we should be able to identify the compromise and alteration of the candidate signal.

3.1 SIFT: Overview

Let $c_{x,t}$ be the candidate signal measured for a duration Δ starting at time t by sensor x. Let $B_{t'} = \{b_{1,t'}, b_{2,t'}, ...b_{i,t'}, ..., b_{n,t'}\}$ be a set of reference signals measured for a duration Δ[1] starting at time t by sensors with index $1 \leq i \leq n$. SIFT works as follows:

- *Feature Generation:* When the base station has Δ time-units of candidate and reference signals (called signal *snippets* from now on), it first performs extract specific features from them. Note that it is possible that some form of preliminary transformation of the candidate and reference signals is needed before feature extraction can take place. For example, if we are interested in the temporal properties of ECG signals then we might want to extract the inter-beat intervals (also known as RR-intervals) from it. The features are chosen such that they characterize the inter-dependence of the candidate signal and at least one of the reference signals. We are able to choose such features because the candidate and reference signals are related based on an underlying physiological process. We define $V = \{v_1, v_2, ..., v_m\}$ as set of features (m-dimensional feature vector) of interest for a signal snippet, where each $v_i = f_u(c'_{x,t}, b'_{k,t})$ and $1 \leq i \leq m$, t is the starting point of the signal measurement, $1 \leq k \leq n$ is the index of the reference signal, f_u is a function in a set of functions $F = \{f_1, f_2, ...f_z\}$ that generates the features from the candidate and reference signals.

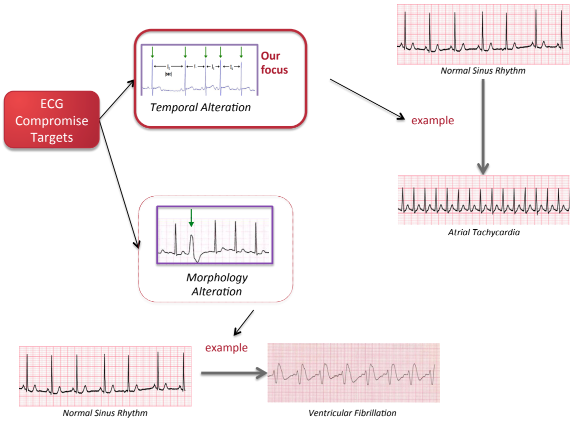


Figure 2: Types of ECG Sensor Compromise in BSN

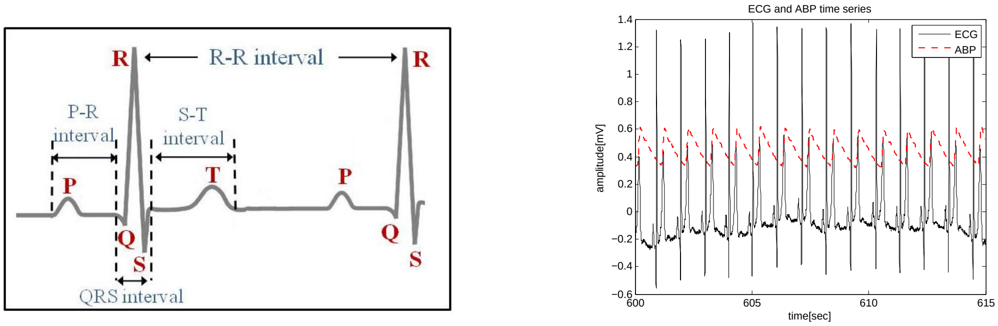
- *Training:* In order to identify altered signals, we train a user-specific supervised machine learning model. Therefore, for each user we generate a set P, which contains a collection of m-dimensional feature points obtained from synchronously measured candidate and reference signals from the same user. The features in P form the *legitimate* points within the system. To these features we add another set of tuples Q, which are generated as described in the three previous steps, except feature points in Q are generated from candidate and reference signals that are *not* from the same user. The points in the set Q form the *illegitimate* points for our learning algorithm. The learning algorithm essentially determines one or more feature thresholds which will then be used to determine the alteration of any new candidate signal snippet.

- *Evaluation:* Once the classification is done, whenever we want to identify if a ECG sensor is compromised or not, we collect Δ duration of candidate and reference signals, transform them, extract features from them, and input them into the model, which then outputs a binary decision on whether the signal has been altered or not, which then indicates sensor compromise in our system.

In a real BSN, the entire process will be executed at the base station, which continually receives data (in a secure manner without tampering) from the sensors measuring the candidate and reference signals in the network. If the data from the candidate sensor is deemed altered, an *alarm* will be raised to inform the patient and their caregivers. We now move on to demonstrate the operation of SIFT in detecting compromised ECG sensors by identifying temporal alteration in actual ECG signals using signals from arterial blood pressure and respiration sensors.

4. SIFT-BASED ECG SENSOR COMPROMISE DETECTION

[1] In actuality, the candidate signal and each of the reference signal measurements are sent to the base station at regular intervals δ_k, where k is the index of the interval. Starting at time t, the base station waits till it has enough data, i.e., $\sum_k \delta_k = \Delta$ before it executes the compromise detection.



(a) (b)

Figure 3: (a) Annotated P, Q, R, S, T waves in a typical ECG signal, and (b) ECG and ABP signals

In this section, we apply the SIFT methodology to detect a compromised electrocardiogram (ECG) sensor in a BSN. Compromise of ECG sensors, with the intension of providing incorrect data about the patient, can manifest itself in two ways: through temporal changes or through morphological changes to the ECG signal. Temporal changes are associated with the interval between two consecutive R peaks being consistently misreported. For example, due to temporal changes, a regular ECG signal might be modified to imply atrial fibrillation (irregular heart rhythm) or atrial tachycardia (abnormally high heart rhythm) or vice-versa. Morphological changes are associated with changes in the shape of the ECG signal. Morphological changes also indicate arrhythmia. Examples include ventricular fibrillation (erratic, disorganized firing of impulses in the ventricles) and ventricular tachycardia (rapid rhythm preventing the heart from filling with blood properly). Figure 2 shows the two classes of ECG compromise. In this work, *we primarily focus on compromise that results in the inducing of temporal changes in the ECG signal.* This is done in two steps: (1) identifying reference signals, and (2) identifying appropriate features for the learning model.

4.1 Identifying Reference Signals

As we are focused on temporal properties of ECG signals, we first transform the ECG signal in to a series of inter-beat-intervals by detecting the R-peaks in our snippet and calculating the time difference between two consecutive R-peaks. In Figure 3 (a), shows a snippet of a normal ECG signal. Normally, a single ECG complex consists of P,Q,R,S and T waves. Simply put, the P-wave is observed during atrial depolarization (which causes the blood to be pushed to the ventricles), the QRS complex (see Figure 3 (a)) is observed during the rapid depolarization of the right and left ventricles (which causes the blood to be pushed out of the ventricles 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 (see Figure 3 (a)). The RR-interval refers to the beat-to-beat variations in heart rate and is a measure of heart-rate. Our approach to identifying the reference signals is to determine related physiological signals that affect the RR-intervals in some manner. In this regard, we specifically look at two reference signals: arterial blood pressure (ABP)[2] and respiration (RESP) rhythms. The reason we choose these two signals is because RR-intervals in a ECG signal are inextricably linked to ABP and RESP signals in time and frequency domains.

As ECG and ABP signals are both measures of the cardiac process (see Figure 3 (b)). A systolic peak in the ABP signal is typically precede by a R-peak in the ECG the time-domain. The RR-interval and peak systolic peak interval (SS-interval) in the ABP signals are highly correlated as they are measures of the same phenomenon (ventricular compression) [13]. Further, each of these three physiological signals in the body are controlled by our autonomic nervous system. The autonomic nervous system can be divided into two parts: sympathetic (or vagal) and parasympathetic. The sympathetic part is responsible for the body's fight or flight response. Conversely, the parasympathetic system is responsible for the body as rest [12]. In resting humans, the RR-interval fluctuations in the heart-rate are due to both sympathetic and parasympathetic activity.

Spectral analysis of RR-intervals is typically used to estimate the effect of the sympathetic and parasympathetic modulation of the RR-intervals. There are two main frequency bands of interest in the RR power spectrum, which are believed to mostly reflect the subject's sympathetic and parasympathetic activity. One is in the 0.04 to 0.15 Hz band (also known as the Low-Frequency (LF) band), another one is in the 0.15 to 0.4 Hz band (also known as the High-Frequency (HF) band) [10]. In the LF band, one can observe what are known as Mayer waves, which are synchronous with a blood pressure oscillations, and exhibit significant coherence with sympathetic nerve activity. In the HF band, one can observe what is known as Respiratory Sinus Arrhythmia (RSA) waves, which are synchronous with the respiration oscillations, a mainly parasympathetically mediated process. Consequently, both Mayer waves and RSA waves can be observed in the spectra of the ABP and RESP signals, respectively, in a manner synchronous with their observation in the RR power spectrum [12].

4.2 Feature Generation

[2] ABP is a continuous measurement of blood pressure with systolic pressure peaks and diastolic pressure troughs.

Now that we have identified our the reference signals, the next step is to identify appropriate features between the candidate and reference signals. The goal is to capture the various physiological signals that have any correlation or affect the RR-intervals observed in the ECG signal. In general there are two classes of features we consider here. We separated the features into two groups instead of dealing with one one large set because both groups are fundamentally different in how they are related to the RR-intervals. Feature Group 1 is made up of time-domain features that measure the correlation between RR-intervals and other manifestations of inter-beat intervals (in this case SS-interval). While in the Feature Group 2 is made up frequency-domain features the capture the effect of related physiological signals on the RR-intervals themselves (see Table 1).

- *Feature Group 1:* As the ABP and ECG signals represent the cardiac process, the RR-intervals from the ECG signal and systolic peak intervals (SS-intervals) from the ABP signal should be strongly correlated to one another. Therefore, our first feature group consists of three features: (i) correlation coefficient of the RR- and SS-intervals series obtained from ECG and ABP snippets, (ii) average RR-interval duration, and (iii) average SS-interval duration.
- *Feature Group 2:* As the RR-intervals in the ECG signal are affected by both ABP and RESP signals in the LF and HF band, our second feature group consists of the following: (i) difference in frequency at which Mayer waves are observed in RR-intervals and ABP; (ii) difference in frequency at which RSA wave is observed in RR-interval and RESP; (iii) highest, lowest and average power in LF and HF bands of magnitude squared coherence (MSC) between RR and ABP and RR and RESP, respectively; and (iv) total number of peaks in the LF and HF bands of magnitude squared coherence (MSC)[3] between RR and ABP and RR and RESP, respectively.

Note that, we tried combinations of features from the Feature Group 1 and Feature Group 2, but did not get significantly better results than Feature Group, hence the rest of our discussion will not consider such cases.

4.3 Training

Once the features from either of the groups have been extracted, we use them to train a supervised machine learning model. As for the learning algorithm use the Naive Bayes classifier. The models are generated specifically for the user on whom the BSN is to be deployed. To train the model we first extract the appropriates features (3 in Feature Group 1 and 10 in Feature Group 2) from the snippets of synchronously measured ECG, ABP and RESP signals and label these as legitimate points. We then perform the aforementioned feature generation using snippets of altered ECG signals with synchronously measured ABP and RESP and label these as illegitimate points (see Section 5 for details).

Table 1: Feature Summary

Source	Feature Type
Feature Group 1	Correlation of RR- and SS-interval
	Average RR-interval duration
	Average SS-interval duration
Feature Group 2	Mayer wave frequency difference in the RR and ABP spectrum
	RSA wave frequency difference in the RR and RESP spectrum
	Highest power in LF of RR and ABP
	Lowest power in LF of RR and ABP
	Average power in LF of RR and ABP
	Highest power in HF of RR and RESP
	Lowest power in HF of RR and RESP
	Average power in HF of RR and RESP
	Peak number in LF of RR and ABP
	Peak number in HF of RR and RESP

Once the feature generation and model training stages have been complete, we can use the trained model to decide if any newly received snippet of ECG signal has been altered temporally or not. We do this by first extracting n features from the ECG snippet and synchronously measured ABP and/or RESP snippets (depending upon the feature group) and feeding them to our patient-specific model. The model then assigns a label to this n-dimensional (where n is 3 or 10 depending upon the feature group) feature point, as legitimate or illegitimate. If the point is deemed illegitimate, we raise an alarm.

Table 2: User Summary

Type	#	Male	Female	Avg. Age (yrs.)
Normal	13	6	7	44.46
Abnormal 1	10	5	5	62.1
Abnormal 2	7	2	5	75

5. VALIDATION

In this section, we validate our SIFT-based approach to detect compromised ECG sensors. Our goal with the validation was to demonstrate two things: (1) the ability to detect changes (even subtle ones) in the temporal properties of ECG signals induced by an adversary, and (2) the inability of an attacker to deceive SIFT using synthetic ECG signals derived from past ECG signal snippets from the user.

By *subtle changes* we mean where an adversary replaces an ECG snippet with another very similar one. For example, an actual ECG snippet with normal sinus rhythm being replaced with another normal ECG snippet with another person. From a safety standpoint this might not be an issue as the overall diagnosis of the patient's health is not affect, but this is a security issue as we are not receiving the actual patient's data. Moreover, a clever adversary might not introduce drastic changes into a compromised sensor, but try to induce changes slowly to avoid detection. Hence, it is important to be able to detect even small changes in this manner.

[3] MSC is the measure of spectral coherence — a measures the causality between the two signals. The MSC of two signals signal $x(t)$ and signal $y(t)$ is defined as follows: $C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)*P_{yy}(f)}$, where, $P_{xx}(f)$ and $P_{yy}(f)$ denotes the power spectral densities of signal $x(t)$ and signal $y(t)$ respectively, and $P_xy(f)$ denotes the cross power spectral density of these two signals.

Experimental Setup: The first step in validating ECG compromise detection using SIFT is to train a user-specific learning model. For this work we collected data belonging to 30 patients from the MIT PhysioBank Fantasia and MGH databases [5]. We chose these databases as they provided all three signals of interest to us, all sampled at 250Hz. Furthermore, the Fantasia database is made up of healthy users, while MGH database mainly contains data from patients with specific ailments. We then searched the MGH database to specifically chose users whose ailment manifested itself in temporal variation in the measured ECG signal. We categorize these list of 30 users into three user types based on their ECG: Normal, Abnormal 1 and Abnormal 2. *Normal* user type indicates users who did not suffer from any ailments and had normal sinus rhythm ECG. *Abnormal 1* user type indicates users whose ECG signal showed both normal as well as temporally abnormal rhythms, while *Abnormal 2* user type indicates users with tachycardia (fast resting heart-rate) and bradycardia (slow resting heart-rate). Table 2 shows the various statistics on the user population we used for our experiments.

Model Training: In order to train a model for a user we need legitimate and illegitimate points. For generating the legitimate points for Feature Group 1, we collect ECG, ABP signals for 55 minutes. We then use a sliding window of 5 minute[4] interval on the ECG, ABP signals to produce 66, *5-minute signal snippets*. Each of these snippets then produces one legitimate 3-dimensional feature point. This results in a total of 66 legitimate points for each user over the entire dataset. In order to add illegitimate points to the model, we combined each user's ABP snippet from every 5 minute sliding window with an ECG snippet from an another user's randomly selected 5 minute window to generate a 3-dimensional feature point. As we want to be able to distinguish between a user's ECG from a variety of other ECG signals, we generate 290 illegitimate points for each user. Once the legitimate and illegitimate feature points have been generated, we train our Naive-Bayes classifier using these features and use 10-cross validation to test the model built. For Feature Group 2 we train our Naive-Bayes model similarly, except the features are in the frequency domain involving ECG, ABP and RESP signals and the features generated were 10-dimensional.

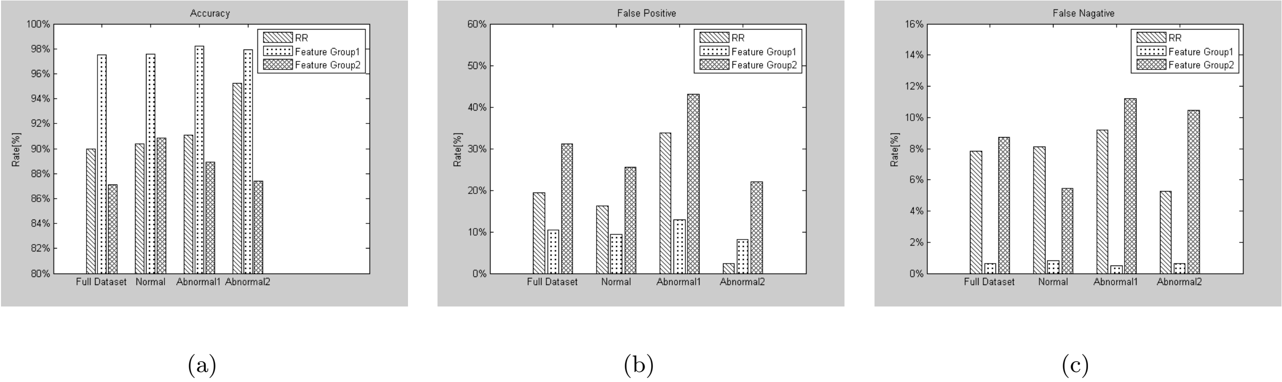
As we are dealing with temporal alteration of ECG signals, any such alterations should be visible through analysis of RR-intervals themselves. Therefore, in addition to the performance measurement of Feature Group 1 and Feature Group 2, we compared the results of our features with a non-SIFT approach where we simply used historical RR-interval average to detect ECG alteration. This case is represented by the label **avgRR** in the results.

Performance Analysis: Figure 5 (a) (b) and (c) shows the detection accuracy, false positive and false negative rates of our detection system. We define false positive as the case where an actual ECG snippet was classified as altered and false negative as the case where an altered ECG snippet was classified as unaltered. In terms of detection accuracy and error results we can see that Feature Group 1 consistently outperforms others. Further, It can be seen that avgRR has a reasonable level of accuracy by itself. This is not surprising as we are focused on detecting temporal changes in the ECG signal. The biggest difference between Feature Group 1 and avgRR is the degree to which the former can distinguish subtle variations in the temporal properties of an ECG signal as seen in the false negative results for the Normal case in Figure 5 (c). The only situation where avgRR is better than Feature Group 1 is in the rate of false positives when the underlying ECG signal being modified is characterized by tachycardia or bradycardia as seen in Abnormal 2 in Figure 5 (b). In all other cases Feature Group 1 outperformed avgRR. Both FG 1 and avgRR outperform Feature Group 2 in all cases. This shows us that the affect of ABP and RESP signals of the temporal properties of ECG is not as strong as some of the other indicators. However, given the accuracy rate for Feature Group 2 is close to 90% the performance is not particularly poor. The performance penalty is in error rates and also much higher computation overhead for Feature Group 2 over Feature Group 1 and avgRR.

The fact that Feature Group 1 has an overall accuracy of close to 98% shows the capability of using the SIFT approach for compromise detection without relying on redundant sensors or historical data, thus meeting our *design goals*. Further, the false negatives are 0.60% shows that Feature Group 1 is very good at discerning even subtle changes in the ECG signal. It even detects a replacement of a normal ECG signal with another normal looking ECG signal from another person. However, the cost is in the false positives, which are bit high. However, we believe that this might be a consequence of the quality of data we used. Manual analysis of the ABP signals found that in many situations the ABP signals were not recorded properly and had considerable noise in them. This resulted in missing SS-peaks and poor correlation between RR-intervals and SS-intervals for the same user. We believe the false positive rate will be much lower with a less noisy data source, which will be the focus of our future work.

Security Analysis: We have seen the SIFT-based ECG compromise detection is very effective in detecting sensor compromise in a BSN, even when the ECG has been subtly altered. In this section, we add another layer of analysis to the capability of SIFT in detecting ECG compromise by evaluating if it can be fooled by using generative models parameterized with a user's own ECG data from the past. In this regard, we used ECGSYN [11] a well-known synthetic ECG generator, which has been shown to generate clinically relevant synthetic ECG signals. ECGSYN can be parameterized for anyone by collecting a their ECG data and extracting temporal (e.g., average and standard deviation of the heart-rate, ratio of Mayer/RSA waves) and morphological (e.g., intensity of the P,Q,R,S,T waves and their angles from a reference point) properties from it [13]. We used ECG snippets for each of our 30 user's from the PhysioBank database to parameterize ECGSYN and introduced synthetic ECG signal instead of the user's actual ECG data. Figure **??** shows examples of a user's actual and synthetic ECG data generated using ECGSYN. The low false negatives of our approach helped considerably in this regard as it allowed us distinguish all but 4 of the 30 synthetically generated snippets. This shows that our approach is robust to even adversaries who have access to the user's ECG data.

[4] We chose 5 minutes because it gave us the best results without requiring excessive data alteration detection latency



(a) (b) (c)

Figure 4: (a) Accuracy Rate (b) False Positive Rate and, (c) False Negative Rate for for Feature Group 1, Feature Group 2 and avgRR features

In this work, we have shown that our approach can detect any alteration of ECG signals, including (but not limited to) switching of a healthy ECG stream with someone else's healthy data stream. This is significant because from a fault detection stand-point (which assumes a benign environment) the only thing that matters is that data being collected correctly represents the user's state. However, from a security stand-point that fact that a normal ECG signal was replaced with another indicates adversarial presence within the system, which can have long term safety consequences for the patient.

6. DISCUSSION

The SIFT methodology has been designed for detecting compromised sensors in a BSN in a malicious environment. However the solution has other implications as well. The most obvious is the ability to detect faults in benign environments. For example, in our case study if the ECG sensor experiences a fault that leads to incorrect measurement of the underlying ECG of the patient, then it can be detected as well. However to be used for fault detection, the thresholds in the learning model may have to be tuned appropriately to reduce false positives depending upon how rare the faults are.

Further, the ability of SIFT-based approaches to differentiate signals generated from two different users (as is the case with ECG signals using blood pressure and respiration signals) can be leveraged to determine if the sensors are on the same person. This is particularly useful in making sure that malicious entities are not mounting person-in-the-middle attacks between the sensors and the base station within the BSN network.

Interestingly, the SIFT methodology is very similar to smart-alarm algorithms designed in the medical informatics domain [14], where data from multiple physiological signals are combined to determine the state of the patient (rather than relying on simple threshold-based alarms for individual physiological signals). However, SIFT may not be usable in the same way because if the patient's underlying physiological process changes it will be visible in all its physiological signal manifestations. Hence, a genuine change in the underlying physiological process which may have clinical significance would not be detected as it has no security significance.

Finally, as noted earlier, SIFT relies on the a patient-specific model of the correlation between various related physiological signals to function. If a patient's physiology changes, the models have to adapt as well. The current design of SIFT cannot adapt to this change automatically. The model has to be re-calibrate every so often in order to capture the current state of the patient's health. One could automate the re-learning based on a schedule. However, choosing the inter-re-learning 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.

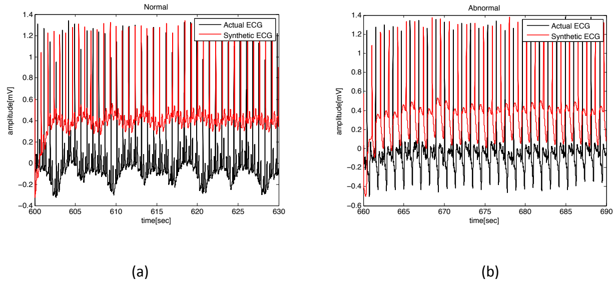


Figure 5: Actual and Synthetic ECG signal snippets for (a) Healthy User, and (b) User with Tachy-Brady Syndrome

7. RELATED WORK

Most of the work in this domain has been on detecting faulty sensors in wireless sensor networks. Over the years researchers have developed numerous solutions in this regard [2,6,7,18,19]. However, most of the fault detection schemes are based on two main assumptions: (1) the network has a large number of sensors with identical functionality deployed, and (2) for a given stimuli, the sensors in the same neighborhood should have the same or similar sensed values. Given these assumptions, the approaches cluster the nodes into different "subnets" according to their location and compare the similarity of the sensor readings with others nearby based on a pre-defined threshold.

In recent years, researcher have tried to adapt these redundancy-based methods to the domain of BSNs [3,4,8,9,15]. Almost all the work has been done for BSNs that naturally require considerable sensor redundancies, i.e., motion monitoring

BSNs. Useful as these solutions are for detecting faults with motion sensors (which have considerable redundancy), they might not work when we consider physiological sensors in a BSN, as typically there is only one sensor of a particular type. In [9], the work closest to us in spirit, 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 signal. The approach builds a rule-table for various combinations of blood pressure and heart-rates and determines if the observed data in within these expected bounds, if not, then the sensors are deemed faulty. This approach uses a simple cardiac output model to determine the relationships between the heart-rate and blood pressure, therefore does not work if an adversary deliberately replaces the legitimate ECG signal with another signal having similar heart-rate characteristics.

8. CONCLUSIONS

In this paper we presented SIFT, a novel methodology to address the problem of sensor compromise in BSNs. The main idea was to leverage the fact that physiological signals in the human body are inherently correlated. We illustrated the capabilities of SIFT using a case study where we detect the compromise of a ECG sensor using blood pressure and respiration sensors as reference. Analysis of our case study demonstrated promising results with over 98% accuracy in detecting even subtle ECG modifications when the temporal characteristics of the inter-beat intervals from the ECG and blood pressure signals were used as features of choice. In the future, we plan to extend this work in the following directions: (1) using the SIFT methodology to detect morphological changes in ECG signal, and (2) understand SIFT itself better by evaluating whether a symmetric relationship exists between the candidate and reference signals such that alteration of a reference signal can be detected using the candidate signal.

9. REFERENCES

[1] 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*, WH '13, pages 2:1–2:8, New York, NY, USA, 2013. ACM.

[2] J. Chen, S. Kher, and A. Somani. Distributed fault detection of wireless sensor networks. In *Proceedings of the 2006 workshop on Dependability issues in wireless ad hoc networks and sensor networks*, pages 65–72. ACM, 2006.

[3] K. Duk-Jin and B. Prabhakaran. Motion fault detection and isolation in body sensor networks. *Pervasive and Mobile Computing*, 7(6):727–745, 2011.

[4] 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 (SMC), 2012 IEEE International Conference on*, pages 2377–2382, Oct 2012.

[5] A. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. M. RG, J. E. Mietus, G. B. Moody, C.-K. P. C-K, and H. E. Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):215–220, 2000.

[6] M. Hajibegloo and A. Javadi. Fast fault detection in wireless sensor networks. In *Digital Information and Communication Technology and it's Applications (DICTAP), 2012 Second International Conference on*, pages 62–66. IEEE, 2012.

[7] P. Jiang. A new method for node fault detection in wireless sensor networks. *Sensors*, 9(2):1282–1294, 2009.

[8] 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*, pages 5234–5237. IEEE, 2012.

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

[10] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz. Heart rate variability standards of measurement, physiological interpretation, and clinical use. *European heart journal*, 17(3):354–381, 1996.

[11] P. McSharry, G. Clifford, L. Tarassenko, and L. Smith. A dynamical model for generating synthetic electrocardiogram signals. *Biomedical Engineering, IEEE Transactions on*, 50(3):289–294, March 2003.

[12] P. E. Mcsharry and G. D. Clifford. Models for ECG and RR Interval Processes. In G. D. Clifford, F. Azuaje, and P. Mcsharry, editors, *Advanced Methods And Tools for ECG Data Analysis*, chapter 4. Artech House Publishers, 2006.

[13] S. Nabar, A. Banerjee, S. K. S. Gupta, and R. Poovendran. Gem-rem: Generative model-driven resource efficient ecg monitoring in body sensor networks. In *Body Sensor Networks (BSN), 2011 International Conference on*, pages 1–6, May 2011.

[14] A. Roederer, J. Holmes, M. Smith, I. Lee, and S. Park. Prediction of significant vasospasm in aneurysmal subarachnoid hemorrhage using automated data. *Neurocritical Care*, pages 1–7, 2014.

[15] 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*, pages 162–167, 2011.

[16] K. K. Venkatasubramanian, A. Banerjee, and S. K. S. Gupta. PSKA: Usable and secure key agreement scheme for body area networks. *Information Technology in Biomedicine, IEEE Transactions on*, 14(1):60 –68, Jan. 2010.

[17] K. K. Venkatasubramanian and S. K. S. Gupta. Physiological value-based efficient usable security solutions for body sensor networks. *ACM Trans. Sen. Netw.*, 6(4):31:1–31:36, July 2010.

[18] T. Zahra and S. Mohsen. A trust-based distributed data fault detection algorithm for wireless sensor networks. In *Proceedings of International Workshop on Internet and Distributed Computing System*, 2008.

[19] C. Zhang, J. Ren, C. Gao, Z. Yan, and L. Li. Sensor fault detection in wireless sensor networks. In *Proceeding of the IET International Conference CCWMC'09*, pages 66–69, 2009.