Detecting Malicious Temporal Alterations of ECG signals in Body Sensor Networks

Hang Cai, Krishna K. Venkatasubramanian

Dept. of Computer Science,
Worcester Polytechnic Institute,
Worcester, MA, 01609
{hcai, kven}@wpi.edu

Abstract. Electrocardiogram (ECG) sensor is one of the most commonly available and medically important sensors in a Body Sensor Network (BSN). Compromise of the ECG sensor can have severe consequences for the user as it monitors the user's cardiac process. In this paper, we propose an approach called *SIgnal Feature-correlation-based Testing (SIFT)* which is used to detect temporal alteration of ECG sensors in a BSN. The novelty of SIFT lies in the fact that it does not require redundant ECG sensors nor the subject's historical ECG data to detect the temporal alteration. SIFT works by leveraging multiple physiological signals based on the same underlying physiological process (e.g., cardiac process) – arterial blood pressure and respiration. Analysis of our case study demonstrates promising results with ∼98% accuracy in detecting even subtle alterations in the temporal properties of an ECG signal.

1 Introduction

Emerging Body Sensor Networks (BSNs) have demonstrated great potential in a broad range of applications in healthcare and wellbeing. The fact that BSNs collect sensitive data and provide valuable information to caregivers and users makes them attractive targets for tech-criminals to exploit. One such threat is *sensor compromise*, which we define as the unauthorized modification of sensor output (i.e., measurement) to relay incorrect patient health data to the base station. The modification of sensor output can be done in several ways including installation of malware on sensors that modify the readings [2], and inducing arbitrary signals into sensor circuitry leading to erroneous readings [8].

Electrocardiogram (ECG) is one of the most widely deployed sensors on individuals. Any compromise of ECG sensor or surreptitious alteration of the sensor output can pose extreme consequences to a person's health from missed diagnosis and delayed treatment. In general, compromising an ECG sensor in a BSN allows the adversary to alter its signal in two possible ways: (i) *temporal* alteration, which modifies the timing information of ECG complex (e.g., inter-beat-interval); and (ii) *morphological* alteration, which modifies the shape of the ECG. In [1], we proposed a method to detect the morphological alterations of ECG signal in BSNs. In this paper, we present a complementary work on detecting temporal alterations of ECG sensor output due to adversarial compromise.

Temporal alterations can be used to modify a regular ECG signal to imply atrial fibrillation (irregular heart rhythm) or atrial tachycardia (abnormally high heart rhythm) or vice-versa.

Recent years have seen some 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, 5, 7, 12]. Such BSNs naturally require considerable sensor-redundancies, where multiple sensors of the same type (e.g., accelerometers) measure the same limb movement. However, they might not be applicable when we consider ECG sensors, since for usability reasons typically there is only one ECG sensor in a BSN. Alternatively, history-based anomaly detection approaches have also been proposed in [13]. However, the human body is too dynamic for the past to effectively determine the current patient state at all times.

In this regard, we present a novel methodology for detecting temporal alteration of ECG sensor output in a BSN called **SIgnal Feature-correlation-based Testing (SIFT)**. It works by generating a subject-specific model by correlating their ECG sensor output with synchronously measured arterial blood pressure (ABP) and respiration (RESP) signals. As ABP measures the same physiological phenomena as ECG — the cardiac process, consequently, the inter-beat intervals in ECG and inter-systolic peak intervals in ABP are highly correlated. Further, both ABP and RESP signals affect the ECG inter-beat intervals through the autonomic nervous system [4, 11], which is reflected in the observation of several frequency-domain features in inter-beat-interval sequence, ABP and RESP signals. Therefore, any alteration of the temporal properties of ECG signal by an adversary, if not reflected as a commensurate change in the ABP and RESP signals, is considered as evidence of compromise. The analysis of SIFT demonstrates promising results with ∼98% accuracy in detecting even subtle ECG signal alterations for both healthy subjects as well as subjects with cardiac conditions.

2 Problem Statement and System Model

Formally speaking, let x be the signal the adversary is trying to alter, then the goal is to find a means of detecting if this signal x has been temporally altered to x', solely-based on a set of reference signals $Y = \{y_1, y_2, ...y_n\}$ such that, each y_i, where $1 \leq i \leq n$, shares certain common features with x, either in the time or frequency-domain or both. In our case x is the ECG signal, while Y is a set with $n = 2$ elements: ABP signal and RESP signal.

In terms of the **system model** we assume the BSN is comprised of a number of wearable medical sensors capturing physiological signals from patients, especially the ECG, ABP and RESP sensors. These sensors continuously 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. Our ECG compromise detection system is deployed at the base station.

In terms of the **threat model** we assume the primary goal of adversaries is compromising the ECG sensor and temporally altering its output using side-channel attacks such as [8]. Once the ECG sensor is compromised, it may generate erroneous output at any time. We assume that ABP and RESP sensors are secure and will not be attacked.

3 SIFT: An Approach for ECG Temporal Alteration Detection

In this section, we introduce our approach to the detection of temporal alteration of ECG sensor output called **Signal Feature-correlation based Testing (SIFT)**. Figure 1 shows the basic operation of SIFT. It consists of three steps: (1) feature generation, (2) training, and (3) detection.

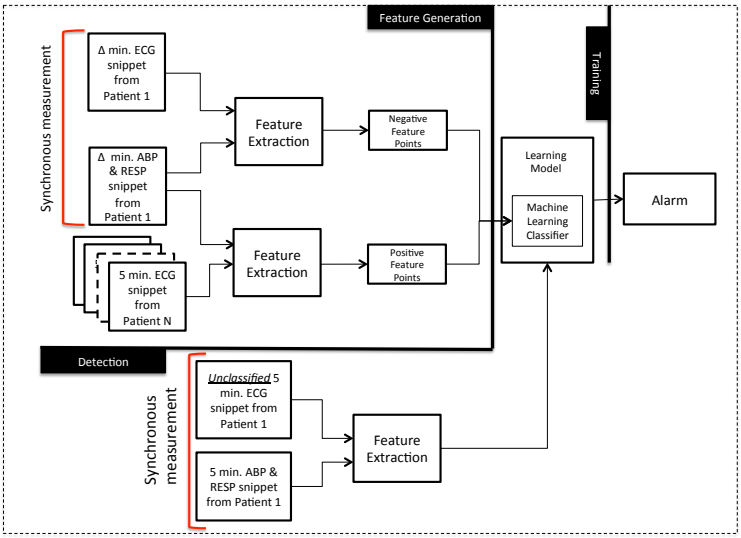

Fig. 1. Signal Feature-correlation based Testing for the Detection of Temporal Alteration of ECG Sensor Output

Feature Generation: We view compromise of ECG sensors, with the intention of providing incorrect data about the subject, to manifest itself as *temporal changes* in the output ECG signal. Temporal changes are associated with the interval between two consecutive R-peaks being misreported. Therefore, we first transform the ECG signal into a series of inter-beat-intervals by detecting the R-peaks and calculating the time difference between two consecutive R-peaks. The RR-tachogram thus produced forms our *candidate signal*. We then extract feature points from this candidate signal with two other *reference signals* derived from ABP and RESP signals. These feature points will then be used to train a subject-specific model and used to detect ECG alterations. In all, we extracted a set of 13 features from candidate and reference signals, which can

be classified into two categories: (1) *Time-Domain Features*, which include (i) correlation coefficient of the RR- and SS-intervals obtained from ECG and ABP snippets; (ii) average RR-interval duration; and (iii) average SS-interval duration. (2) *Frequency-Domain Features*, which include (i) difference in frequency at which Mayer wave is observed in the power spectrums of RR-tachogram and SBP signal; (ii) difference in frequency at which the RSA wave is observed in the power spectrums of RR-tachogram and RESP signal; (iii) highest, lowest and average power in LF band of magnitude squared coherence (MSC[1]) between RR-intervals and SBP signal; (iv) highest, lowest and average power in HF band of MSC between RR-intervals and RESP signal; and (v) total number of peaks in the LF and HF bands of MSC between RR-intervals and SBP signal and RR-intervals and RESP signal, respectively.

Training: In order to account for the individual variation in the physiological processes, we build a *subject-specific* model for each subject on whom we tested our system. To train the models, we first extract the aforementioned 13-dimensional features from Δ minutes (the time for which data needs to be collected to train the model) of synchronously measured ECG, ABP and RESP signals from the same subject and label these as negative class points (which indicates the three signals are from the same patient). The feature extraction is done using sliding window of size $w < \Delta$, which is moved over the three synchronously measured signals. Each w-sized window of data thus produces one feature point for the system. We then extract the aforementioned features using snippets from ECG signals with ABP and RESP signals from different patients and label these as positive class points (which indicates the ECG signal is from the different patient but ABP and RESP signals are from the same patient). Once the negative and positive points are collected, we feed them into a machine learning classifier to generate a subject-specific model.

Detection: After model training stage is completed, we can use the trained model for a subject to decide if any newly received snippet of ECG signal has been temporally altered or not. Again, we use feature generation method for w-sized long synchronously measured ECG and ABP and RESP snippets to generate a feature point, and then feed this feature point to our subject-specific model. Then the model will output a label for this feature point as negative or positive. If the point is deemed positive, we raise an alarm. Note that we have to set w to a value greater than or equal to 5 minutes because it is the recommended duration needed to produce clear Mayer and RSA waves [9]. This means SIFT needs at least 5-minutes of subject data to be able to determine signal alteration. Developing alternative mechanisms for reducing the time needed to generate alerts (i.e., Δ) is part of our future work for this project.

[1] Magnitude squared coherence (MSC) is the measure of spectral coherence and 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.

4 Validation

Our goal with the validation was to demonstrate two things: (1) the ability of our approach to detect changes 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 historical ECG data from a subject.
Dataset: We collected 28 subjects' data from the MIT PhysioBank Fantasia and MGH databases [6]. The Fantasia database is made up of healthy subjects, while the MGH database mainly contains data from subjects with specific cardiac conditions (i.e., ailment). We categorize these subjects into three types: (1) *Normal* subject type indicates subjects who did not suffer from any cardiac conditions and had normal sinus rhythm ECG, which consists of 6 males and 7 females with an average age of 44.46 (std 25.52). (2) *Abnormal* subject type indicates subjects with consistent tachycardia or bradycardia, which consists of 4 males and 2 females with an average age of 61.4 (std 19.25). (3) *Mixed* subject type indicates subjects whose ECG signal showed both normal as well as tachycardia or bradycardia rhythms, which consists 5 males and 4 females with an average age of 44.78 (std. 20.39).
Detection Results: In our experiments, we select Naive Bayes as our classifier to train the model. We set $\Delta = 60$ minutes and $w = 5$ minutes to produce the feature points. Furthermore, we compared the results of SIFT with an approach that analyzed historical RR-intervals to detect ECG alteration at any given time. This case is represented by the label **RR-only** in the results.

Figures 2, 3 and 4 show the box-plots for balanced accuracy (BAC), false positive (FP) and false negative (FN) rates of our detection system. In terms of detection accuracy, we can see that RR-only is reasonably accurate (average BAC of $\sim$87.41%). However, it has a considerably higher spread (compared to our approach). The RR-only approach performs best for subjects in the Abnormal set mainly because subjects in this set displayed unhealthy ECG (the variations of the RR-interval in this group is considerable high) and therefore it was easy to detect changes to these. In the case of Normal subject type the variations of box plot were much larger because the variations of the RR-interval is comparably smaller. Finally, in the case of the subjects in Mixed set the performance was worst both in terms of median BAC as well as the spread because subjects in this set exhibited ECG that was both normal as well as abnormal rhythms.

However, we can see that using SIFT the detection performance and spread is considerably better than using RR-only approach in terms of median BAC, FP, FN. For *Normal* subject type, our approach provides 98.46% BAC on average with average FP at 2.44% and FN at 0.65%. Not surprisingly the performance degrades a bit when we consider subjects with cardiac conditions. For the *Mixed* subject type, the average BAC of SIFT is 96.39%. However, the average FP increases to 6.06% with the average FN at 1.15%. We suspect the reason for this increase is twofold: (1) on detailed examination of the data, some of the subject's ECG, ABP and RESP signals had considerable measurement errors, and (2) to a lesser degree, physiological signals of subjects in the Mixed set display both normal and abnormal rhythms and this decreases the classifier performance. In

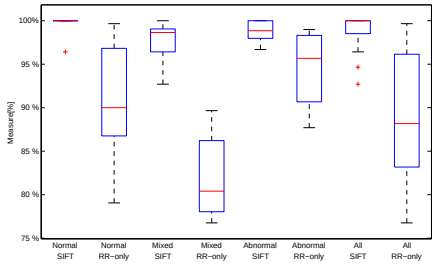

Fig. 2. Balanced Accuracy Rate for Our Approach and RR-only features

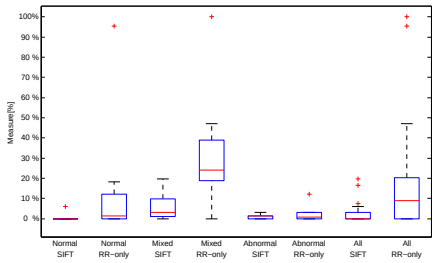

Fig. 3. False Positive Rate for Our Approach and RR-only

the future, we plan to work on improving the proposed system to reduce the second category of errors. For the *Abnormal* subject type, where the subjects display consistently tachycardia or bradycardia, we found the average BAC to be 98.81% with FP and FN at a much better 1.26% and 1.11%, respectively. These results demonstrate that our approach can accurately detect temporal alterations in ECG signal without sensor redundancy and considers the current state of the subject in its operation. Additionally, our approach can distinguish even subtle ECG signal temporal alterations for both healthy and unhealthy subjects. By subtle changes we mean when an adversary replaces an ECG snippet with another very similar one. For example, an actual ECG snippet with normal sinus rhythm being replaced with an ECG snippet from another person with normal sinus rhythm.

Attacks using Synthetic ECG: We add another layer of analysis to the capability of our approach in detecting ECG compromise by evaluating if it can be fooled by using synthetic ECG signals obtained from generative models parameterized with a subject's own ECG data. In this regard, we used ECGSYN [10] a well-known synthetic ECG generator, which has been shown to generate clinically relevant synthetic ECG signals given a set of input parameters. We trained the ECGSYN model with actual subject's ECG data collected over a number of intervals from 5 minutes to 20 minutes from the dataset used to train the subject-specific model. This simulates the case where the attacker knows a portion of the ECG signals used to train the subject-specific model and uses it to alter the current ECG signal. Based on our experiments we find that if we use 10 minutes ECG data to train the ECGSYN model, we were able to detect

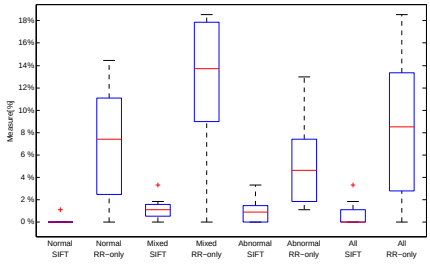

Fig. 4. False Negative Rate for Our Approach and RR-only

the alteration of ECG (replacement of actual ECG with synthetic ECG from ECGSYN) in 91.07% of the cases, giving us FN at 8.93%. Not surprisingly, the FN goes up to 9.82% and the accuracy drops down to 90.18% as the amount of data available for training the ECGSYN model is doubled to 20 minutes. Despite the extreme assumption of the attacker having access to a portion of the same data as our approach has trained its model, the approach works with over 90% accuracy. This result shows that our approach is robust even to an adversary who have access to a subject's ECG data.

5 Related work

Most of the work in this domain has been on detecting faulty sensors in wireless sensor networks. 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 stimulus, the sensors in the same neighborhood should have the 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, 5, 7, 12]. Almost all the work done for BSNs requires considerable sensor redundancies, i.e., motion monitoring BSNs. Useful as these solutions are for detecting faults with motion sensors, they might not be applicable when we consider physiological sensors in a BSN, as typically there is only one sensor of a particular type. Finally, in [1], a method to detect only the morphological alterations of ECG signals was proposed. As stated before this work is complementary to our work and needs to be used in conjunction with our work here to provide a full ECG compromise detection system.

6 Conclusions

In this paper we presented SIFT, a novel methodology to detect temporal alteration of an ECG sensor output using its correlation with arterial blood pressure and respiration signals. Analysis of our approach demonstrated promising results

with ∼98% accuracy in detecting even subtle ECG modifications. In the future, we plan to extend this work in the following directions: (1) reducing the minimum time for which data needs to be collected for effective training of the SIFT, (2) implement SIFT on an actual BSN system to evaluate its performance, (3) investigate ways to overcome our assumption that reference signals are not compromised, by using reference signals that are collected from more trustworthy sources such as the base station.

References

1. Cai, H., Venkatasubramanian, K.K.: Detecting malicious morphological alterations of ECG signals in body sensor networks. In: Proceedings of the 14th International Conference on Information Processing in Sensor Networks. pp. 342–343. ACM (2015)
2. Clark, S.S., Ransford, B., Rahmati, A., Guineau, S., Sorber, J., Fu, K., Xu, W.: Wattsupdoc: Power side channels to nonintrusively discover untargeted malware on embedded medical devices. In: Proceedings of USENIX Workshop on Health Information Technologies. vol. 2013 (2013)
3. Duk-Jin, K., Prabhakaran, B.: Motion fault detection and isolation in body sensor networks. Pervasive and Mobile Computing 7(6), 727–745 (2011)
4. Fink, G.: Encyclopedia of Stress, Three-Volume Set, vol. 1. Academic Press (2000)
5. Galzarano, S., Fortino, G., Liotta, A.: Embedded self-healing layer for detecting and recovering sensor faults in body sensor networks. In: Systems, Man, and Cybernetics, 2012 IEEE International Conference on. pp. 2377–2382 (Oct 2012)
6. Goldberger, A., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., RG, R.G.M., Mietus, J.E., Moody, G.B., C-K, C.K.P., Stanley, H.E.: Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. Circulation 101(23), 215–220 (2000)
7. Kim, D.J., Suk, M.H., Prabhakaran, B.: Fault detection and isolation in motion monitoring system. In: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE. pp. 5234–5237. IEEE (2012)
8. Kune, D., Backes, J., Clark, S., Kramer, D., Reynolds, M., Fu, K., Kim, Y., Xu, W.: Ghost talk: Mitigating EMI signal injection attacks against analog sensors. In: Security and Privacy (SP), 2013 IEEE Symposium on. pp. 145–159 (May 2013)
9. Malik, M., Bigger, J.T., Camm, A.J., Kleiger, R.E., Malliani, A., Moss, A.J., Schwartz, P.J.: Heart rate variability standards of measurement, physiological interpretation, and clinical use. European heart journal 17(3), 354–381 (1996)
10. McSharry, P., Clifford, G., Tarassenko, L., Smith, L.: A dynamical model for generating synthetic electrocardiogram signals. Biomedical Engineering, IEEE Transactions on 50(3), 289–294 (March 2003)
11. Pitzalis, M.V., Mastropasqua, F., Massari, F., Passantino, A., Colombo, R., Mannarini, A., Forleo, C., Rizzon, P.: Effect of respiratory rate on the relationships between rr interval and systolic blood pressure fluctuations: a frequency-dependent phenomenon. Cardiovascular Research 38(2), 332–339 (1998)
12. Sagha, H., del R Millan, J., Chavarriaga, R.: Detecting and rectifying anomalies in body sensor networks. In: 2011 International Conference on Body Sensor Networks. pp. 162–167 (2011)
13. Zahra, T., Mohsen, S.: A trust-based distributed data fault detection algorithm for wireless sensor networks. In: Proceedings of International Workshop on Internet and Distributed Computing System (2008)