

SU-PhysioDB: A Physiological Signals Database for Body Area Network Security

Duygu Karaođlan Altop

Department of Computer
Science and Engineering

Sabancı University, Istanbul Turkey
Email: duygu@sabanciuniv.edu

Albert Levi

Department of Computer
Science and Engineering

Sabancı University, Istanbul Turkey
Email: levi@sabanciuniv.edu

Volkan Tuzcu

Department of Pediatric Cardiology
Istanbul Medipol University,

Istanbul Turkey
Email: vtuzcu@gmail.com

Abstract—This paper presents a new physiological signals database, SU-PhysioDB, that contains simultaneous measurements of electrocardiogram (ECG), blood pressure (BP) and body temperature (BT) signals. SU-PhysioDB can be used to evaluate the performance of the security mechanisms designed for the communication among the biosensors within Body Area Networks (BANs). We present a detailed description of our SU-PhysioDB database along with providing a performance comparison of two specific physiological parameter generation techniques using a public database and our SU-PhysioDB database. Results show that our SU-PhysioDB database is a promising option to be used while evaluating the performance of a bio-cryptographic security infrastructure designed for BANs.

Keywords—Physiological Signals; Electrocardiogram; Blood Pressure; Body Temperature; Body Area Networks; Bio-cryptography

I. INTRODUCTION

Biosensors of the Body Area Networks (BANs) aim at providing pervasive healthcare by collecting and communicating highly sensitive medical information. They are often used for monitoring mission critical processes, because of which they are susceptible to targeted attacks. For instance, pacemakers were targeted to reveal a patient's ECG and to actuate an untimely electrical shock [1]. Besides, if the sensitive personal medical information that the biosensors collect from their hosts is available to malicious entities, it can be misused, which may result in privacy loss, yielding discrimination and/or abuse. For instance, Positive Nation magazine has reported in 2005 that a HIV-diagnosed care worker is not only suspended and dismissed from work but also his health status is made public knowledge in his hometown [2].

Due to the extreme limitations of these biosensors, lightweight and secure key management infrastructures are required for BANs. For this reason, biosensors use physiological parameters that are generated from different vital signals (i.e., ECG, BP) to protect the exchanged private health information. Key establishment requires the communicating entities to have something only they know and it is the physiological signals of the individuals in the context of using bio-cryptography in BAN security. For instance, ECG is commonly recognized as one of the most reliable physiological signals used in the intra-BAN network communication security [3]–[5].

In this paper, we present SU-PhysioDB, a new public and free physiological signals database that is constructed for the purpose of securing the intra-BAN communication. Its main objective is the evaluation of robust physiological parameter generation techniques. We compare the performance of two specific physiological parameter generation techniques, (i) time-domain inter-pulse interval (IPI) based physiological

parameter generation and (ii) frequency-domain cross power spectral density (CPSD) based physiological parameter generation technique, using the ECG and BP signals retrieved from the publicly available and widely used PhysioBank MIMIC II Waveform database [6] (PhysioBank-MIMIC-DB) and our SU-PhysioDB database. The reason behind this comparison is to demonstrate the accuracy of our SU-PhysioDB database.

The rest of this paper is organized as follows. In Section II, we present the details of our SU-PhysioDB database. Section III overviews two existing physiological parameter generation techniques that we use to evaluate the performance of our SU-PhysioDB database. In Section IV, a performance comparison is provided on the two existing physiological parameter generation techniques using the PhysioBank-MIMIC-DB and our SU-PhysioDB databases. Finally, Section V concludes the paper.

II. SU-PHYSIODB

SU-PhysioDB database includes approximately 20 minutes of simultaneous recordings of ECG, BP and 2 BT signals (from different body locations) that are collected from 166 volunteers of age in between 18 and 46 (49 of which are female) in Sabancı University by TÜBİTAK 114E557 project team. Figure 1 shows the testbed constructed for this purpose, and Figure 3 shows an example set of captured signals.

Biopac MP150 Data Acquisition System [7] was used to digitally record physiological data at a sampling rate of 4000 Hz using Biopac's Acknowledge Software (ACQ4.3): (i) raw ECG data were collected using the Biopac ECG100C ECG amplifier module, with disposable ECG electrodes aligned in standard configuration (one apiece to below right/left collarbone and below abdomen), (ii) raw BP data were collected using Biopac NIBP100D non-invasive continuous BP acquisition device, with finger cuff sensors, and (iii) raw BT data were collected using Biopac SKT100C Temperature Amplifier Module, one from the near-finger area and one from the near-collarbone area, as shown in Figure 2.

All of the collected physiological data are securely stored on a personal computer, together with the age and gender of the volunteer and the acquisition time. Expectedly, we have not recorded either the names of the volunteers or any other personal data about them. This study has been approved by Sabancı University Academic Ethics Committee.

III. OVERVIEW ON THE EXISTING PHYSIOLOGICAL PARAMETER GENERATION TECHNIQUES

In this section, we provide an overview of the time-domain IPI-based and the frequency-domain CPSD-based physiologi-

cal parameter generation methods, which are described in our previous works [8], [9]. The general infrastructure of these techniques are visualized in Figure 4.



Fig. 1: Physiological signal acquisition testbed



Fig. 2: Physiological signal acquisition process



Fig. 3: An example for the simultaneously captured physiological signals. Uppermost two are the ECG, upper-middle two are the first BT, lower-middle two are the BP, and the lowermost two are the second BT of two simultaneous measurements from two different volunteers.

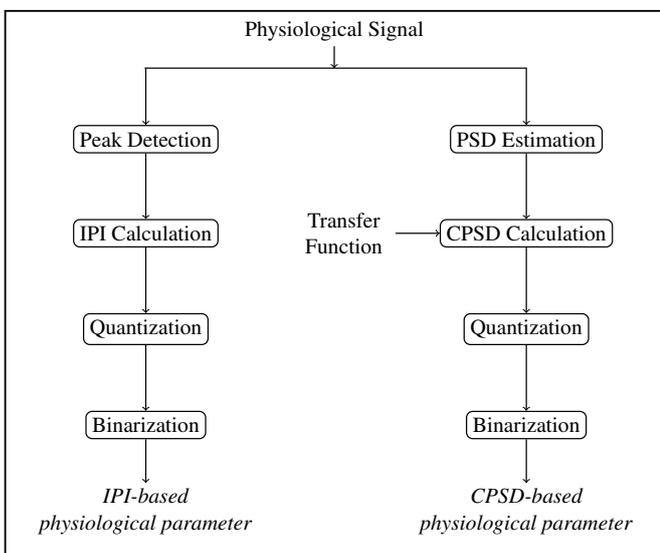


Fig. 4: Overview of our physiological parameter generation techniques

Physiological parameter formation process in the time-domain IPI-based physiological parameter generation technique is as follows: First of all, the peak locations of the sensed physiological signals are detected. Then, the initial IPI sequences of length l are calculated by taking the time difference among the adjacent peak locations. After that, these initial IPI sequences are aggregated by dividing them into groups of g and summing up the elements of each group. The reason behind this operation is to decrease the effect of the measurement errors. Thereafter, the aggregated IPI sequences are quantized using a circular uniform quantization method (with a step size of s), in order to further decrease the effect of the measurement errors. Finally, 128 bit binary sequences are computed by binarizing the quantized IPI sequences. The reason behind this operation is to decrease the error margin of the physiological parameters that are generated in different BANs.

Physiological parameter formation process in the frequency-domain CPSD-based physiological parameter generation technique is composed of two phases: (i) initialization phase, and (ii) operational phase. In the initialization phase, before the deployment of the biosensors, transfer functions are estimated for each pair of the sensed physiological signals using Welch's averaged periodogram method [10], and each of these transfer functions are embedded to the relevant sensing devices. The operational phase starts with the estimation of the power spectral densities (PSD) for the sensed physiological signals. Then, the initial CPSD sequences of length l are calculated by multiplying the estimated PSDs with the relevant embedded transfer functions. After that, the range of the calculated CPSD sequences is adjusted in order to reduce the wide range into a manageable size. Thereafter, the initial CPSD sequences are aggregated by dividing them into groups of g and summing up the elements of each group. Finally, the aggregated CPSD sequences are quantized, again, using a circular uniform quantization method (with a step size of s) and 128 bit binary sequences are computed by binarizing the quantized CPSD sequences. Similar to the time-domain IPI-based physiological parameter generation technique, in the frequency-domain CPSD-based physiological parameter generation technique, the aggregation and quantization operations are performed in order to decrease the effect of the measurement errors, while the binarization operation is performed in order to decrease the error margin of the physiological parameters that are generated in different BANs.

IV. EXPERIMENTS

In this section, we compare the performance of the time-domain IPI-based and the frequency-domain CPSD-based physiological parameter generation techniques in terms of the randomness, distinctiveness, error rates and temporal variance of the generated physiological parameters, using two different databases: (i) PhysioBank-MIMIC-DB, and (ii) SU-PhysioDB. The former database includes ECG and BP signals obtained from the publicly available and widely used PhysioBank MIMIC II Waveform database [6]. This database contains thousands of recordings of multiple physiological signals, including one or more ECG signals, continuous arterial BP signals, fingertip photoplethysmogram signals, and respiration signals, with additional waveforms (up to 8 simultaneously) as available. On the other hand, the details of the latter database are given in Section II.

It is important to note here that we have regulated the physiological parameter generation techniques in order to reduce the effect of having different sampling rates for these different databases. The time-domain IPI-based and the frequency-domain CPSD-based physiological parameter generation techniques described in Section III are suitable for the PhysioBank-MIMIC-DB database. As for the SU-PhysioDB database, on one hand, in the time-domain physiological parameter generation technique, we have divided the IPI sequences by 100 and performed the flooring operation, just after computing the initial IPI sequences. On the other hand, in the operational phase of the frequency-domain physiological parameter generation technique, we have kept the first 128 bits and performed the upcoming operations on those values, just after multiplying the PSD estimations with the hardcoded transfer functions.

As provided in our previous works [8], [9], 10 different starting points are used to compute the IPI and the CPSD sequences with step sizes of $s \in \{2, 3, 4, 5\}$, $s \in \{6, 7, 8, 9\}$, $s \in \{2, 3, 4, 5, 6, 7\}$ and $s \in \{8, 9, 10, 11, 12, 13\}$, respectively for the IPI-based and the CPSD-based physiological parameters that are generated using the physiological signals from the PhysioBank-MIMIC-DB and our SU-PhysioDB databases:

- IPI sequences of length 32, 64 and 128 are used as they are ($g = 1$ and $l = 32$ or $l = 64$ or $l = 128$),
- IPI sequences of length 64 and 128 are divided into groups of 2 ($g = 2$ and $l = 64$ or $l = 128$),
- IPI sequences of length 128 are divided into groups of 4 ($g = 4$ and $l = 128$),
- CPSD sequences of length $l = 128$ are used as they are ($l = 128$ and $g = 1$), and
- CPSD sequences of length $l = 128$ are divided into groups of 2 or groups of 4 ($l = 128$ and $g = 2$ or $g = 4$).

We analyzed the performances of the physiological parameter generation methods using Matlab R2014b. As discussed one by one below, results of the evaluation metrics, whose definitions can be found in [8], [9], show that the physiological parameters generated using the ECG and BP signals retrieved from the PhysioBank-MIMIC-DB and SU-PhysioDB databases do have equivalent performances. It is an obvious indicator that the SU-PhysioDB database can be used to evaluate the performance of the security infrastructures design for BANs.

First of all, randomness of the generated physiological parameters are evaluated using the Shannon's entropy, as in our previous work [8], [9]. Table II shows the average randomness of the generated time-domain IPI-based and frequency-domain CPSD-based physiological parameters. Results reveal that the physiological parameters have comparable randomness values when generated using the PhysioBank-MIMIC-DB and our SU-PhysioDB databases.

Secondly, distinctiveness of the generated physiological parameters are evaluated using the average Hamming distance metric, as in our previous work [8], [9]. Figure 5 shows the average differences among the generated time-domain IPI-based and frequency-domain CPSD-based physiological parameters. Results show that the physiological parameters have comparable distinctiveness values when generated using the PhysioBank-MIMIC-DB and our SU-PhysioDB databases.

Thirdly, error rates of the generated physiological parameters are evaluated using the Equal Error Rate (EER) met-

ric, as in our previous work [8], [9]. Figure 6 shows the Receiver Operating Curves (ROCs) for the generated time-domain IPI-based and frequency-domain CPSD-based physiological parameters. Besides, the lowest EER achieved by the time-domain IPI-based and the frequency-domain CPSD-based physiological parameter generation methods are given in Table I. Results reveal that the physiological parameters have comparable EER values when generated using the PhysioBank-MIMIC-DB and our SU-PhysioDB databases.

TABLE I: Lowest EER achieved by the IPI-based and the CPSD-based physiological parameter generation techniques

Dataset	Method	EER(%)
PhysioBank-MIMIC-DB	Time-domain	4.2
	Frequency-domain	15.3
SU-PhysioDB	Time-domain	4.1
	Frequency-domain	13.4

Finally, temporal variance of the generated physiological parameters are evaluated using the *average temporal ratio* metric, as in our previous work [8], [9]. Table III shows the average temporal ratios of the generated time-domain IPI-based and frequency-domain CPSD-based physiological parameters, where l is the length of the initial physiological parameter sequence, g is the group size and s is the step size. Results show that the physiological parameters have comparable temporal variance values when generated using the PhysioBank-MIMIC-DB and our SU-PhysioDB databases.

V. CONCLUSIONS

In this paper, we present a new public and free physiological signals database, SU-PhysioDB, that is constructed for the purpose of securing the intra-BAN communication. We evaluate the performance of two existing physiological parameter generation techniques, time-domain IPI-based and frequency-domain CPSD-based physiological parameter generation techniques, using the publicly available PhysioBank-MIMIC-DB and our SU-PhysioDB databases. Results show that the physiological parameters generated using the physiological signals retrieved from those different databases have equivalent performances. Therefore, we can state that our SU-PhysioDB database is a prospering option to be used while evaluating the performance of a bio-cryptographic security infrastructure that is designed to secure the BANs. We encourage the use of this database by anyone who works on the area. It is available on <http://people.sabanciuniv.edu/levi/projects/114E557/>.

ACKNOWLEDGMENT

This work was partially supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under grant 114E557.

Duygu Karaoğlan Altop was supported by TÜBİTAK BİDEB 2211-C and Turkcell Academy Technology Leaders Graduate Scholarship Program.

REFERENCES

- [1] D. Halperin, T. Heydt-Benjamin, D. Ransford, S. Clark, B. Defend, W. Morgan, K. Fu, T. Kohno, and W. Maisel, "Pacemakers and implantable cardiac defibrillators: software radio attacks and zero-power defenses," in *Proceedings of the IEEE Symposium on Security and Privacy*. IEEE Press, Oakland, California, USA, 18–21 May 2008, pp. 129–142.

TABLE II: Average Shannon-based Randomness of the IPI-based and CPSD-based physiological parameters. Shannon’s entropy measures the information contained in a message by computing the symbol frequencies. Randomness level of the input sequence increases as the result approaches to 1. Physiological parameters having a Shannon’s entropy greater than or equal to 0.9 are assumed to be random enough to be used as cryptographic keys.

$\begin{matrix} s \\ l, g \end{matrix}$	2	3	4	5	6	7	8	9	10	11	12	13
PhysioBank-MIMIC-DB												
IPI					CPSD							
32, 1	0.93	0.84	0.88	0.86								
64, 1	0.90	0.80	0.82	0.76								
128, 1	0.91	0.80	0.73	0.66	0.98	0.97	0.95	0.96				
64, 2	0.94	0.93	0.90	0.87								
128, 2	0.97	0.93	0.92	0.85	0.99	0.99	0.98	0.99				
128, 4	0.96	0.93	0.92	0.92	0.99	0.99	0.99	0.99				
SU-PhysioDB												
IPI					CPSD							
32, 1	0.97	0.98	0.97	0.94	0.93	0.90						
64, 1	0.96	0.92	0.87	0.83	0.83	0.84						
128, 1	0.98	0.95	0.91	0.87	0.79	0.75	0.99	0.99	0.99	0.99	0.98	0.97
64, 2	0.97	0.96	0.98	0.98	0.98	0.97						
128, 2	0.99	0.98	0.97	0.94	0.93	0.90	0.99	0.99	0.99	0.99	0.98	0.98
128, 4	0.99	0.98	0.96	0.96	0.97	0.98	0.99	0.99	0.99	0.98	0.98	0.99

TABLE III: Temporal variance of the IPI-based and CPSD-based physiological parameters. Temporal variance evaluates the similarity between two physiological parameters that are generated by the biosensors of the same BAN at different time intervals. Physiological parameters having a temporal ratio greater than 1 are assumed to be temporally variant. Underlined values indicate that corresponding physiological parameters have temporal variance.

$\begin{matrix} s \\ l, g \end{matrix}$	2	3	4	5	6	7	8	9	10	11	12	13
PhysioBank-MIMIC-DB												
IPI					CPSD							
32, 1	<u>1.18</u>	0.79	0.57	0.42								
64, 1	<u>2.32</u>	<u>1.33</u>	<u>1.57</u>	0.85								
128, 1	<u>3.76</u>	<u>1.80</u>	<u>1.43</u>	0.99	0.14	0.11	0.09	0.08				
64, 2	<u>2.13</u>	<u>1.78</u>	<u>1.11</u>	0.93								
128, 2	<u>7.68</u>	<u>4.80</u>	<u>2.90</u>	<u>1.63</u>	0.13	0.12	0.10	0.09				
128, 4	<u>4.16</u>	<u>2.41</u>	<u>2.10</u>	<u>1.97</u>	0.11	0.10	0.08	0.08				
SU-PhysioDB												
IPI					CPSD							
32, 1	<u>2.07</u>	<u>1.69</u>	<u>1.61</u>	<u>1.73</u>	<u>1.39</u>	<u>1.18</u>						
64, 1	<u>3.22</u>	<u>2.46</u>	<u>1.94</u>	<u>1.35</u>	<u>1.24</u>	<u>1.35</u>						
128, 1	<u>3.14</u>	<u>3.11</u>	<u>2.68</u>	<u>2.40</u>	<u>1.67</u>	<u>1.36</u>	0.24	0.21	0.21	0.18	0.18	0.19
64, 2	<u>2.45</u>	<u>2.33</u>	<u>1.97</u>	<u>1.64</u>	<u>1.67</u>	<u>1.76</u>						
128, 2	<u>3.21</u>	<u>3.63</u>	<u>3.45</u>	<u>2.66</u>	<u>2.52</u>	<u>2.07</u>	0.26	0.24	0.23	0.22	0.19	0.19
128, 4	<u>3.16</u>	<u>2.70</u>	<u>2.31</u>	<u>2.32</u>	<u>2.08</u>	<u>1.91</u>	0.24	0.24	0.23	0.23	0.22	0.23

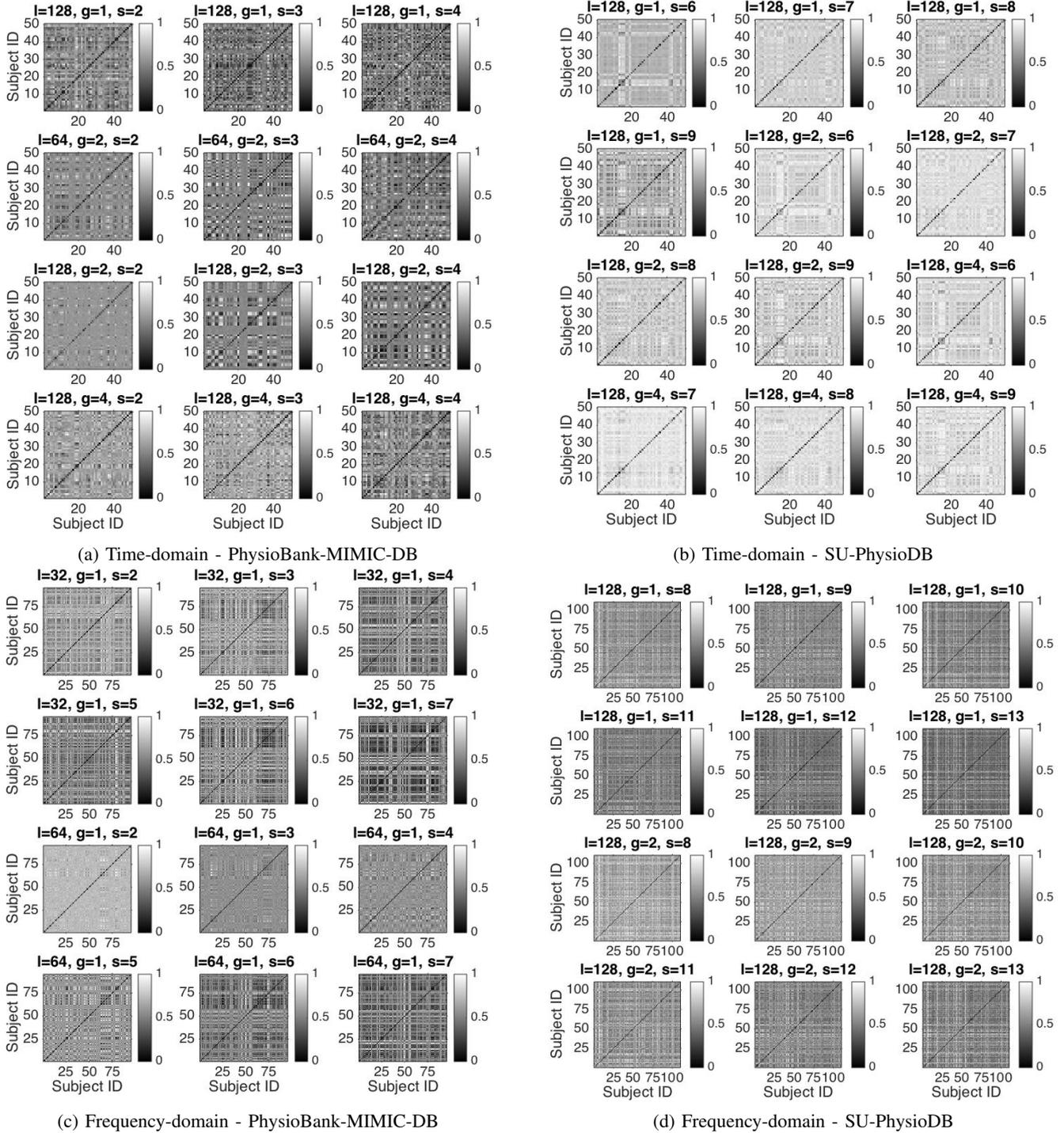
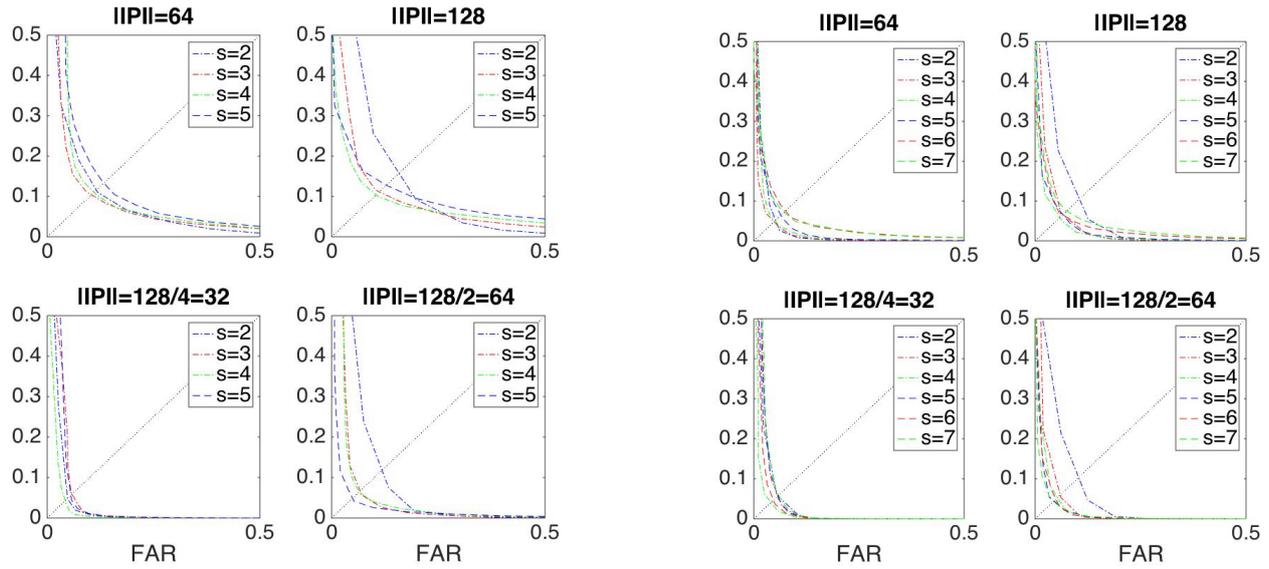
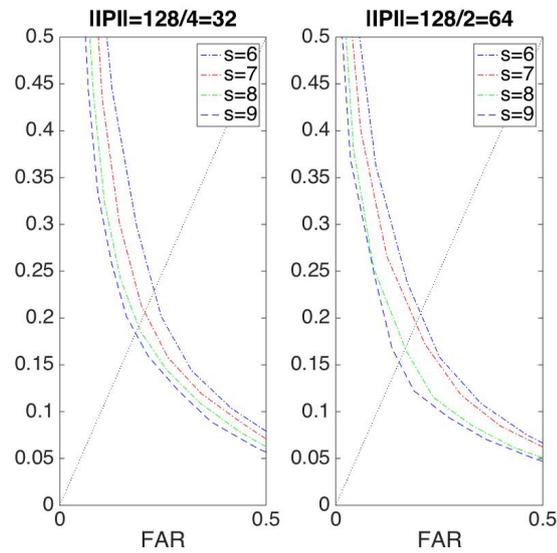


Fig. 5: Distinctiveness of physiological parameters. In each sub-subfigure, diagonal and non-diagonal cells hold the average Hamming distances between the physiological parameters generated from the same host and from different hosts, respectively. The value at the cell (x, y) is the average Hamming distance between the physiological parameters of the x^{th} and y^{th} hosts and the darkness of this cell represents the proximity of these physiological parameters. In order for the generated physiological parameters to be distinctive for different users, Hamming distances between the physiological parameters that are generated from the same host (diagonal cells), should be as dark as possible, i.e. close to 0, and Hamming distances between the ones that are generated from different hosts (non-diagonal cells), should be as light as possible, i.e. close to 1. Results show that the average Hamming distance between each of the IPI-based and CPSD-based physiological parameters derived by different BANs' biosensors is quite large, while the average Hamming distance between each of the IPI-based and CPSD-based physiological parameters derived simultaneously by the same BAN's biosensors is very low.



(a) Time-domain - PhysioBank-MIMIC-DB

(b) Time-domain - SU-PhysioDB



(c) Frequency-domain - PhysioBank-MIMIC-DB

(d) Frequency-domain - SU-PhysioDB

Fig. 6: ROCs of physiological parameters

- [2] L. term illness at work, Available at <https://www.theguardian.com/money/2006/sep/06/discriminationnetwork.discrimination>.
- [3] K. K. Venkatasubramanian, A. Banerjee, and S. K. S. Gupta, "PSKA: Usable and secure key agreement scheme for body area networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 1, pp. 60–68, 2010.
- [4] F. Miao, L. Jiang, Y. Li, and Y.-T. Zhang, "Biometrics based novel key distribution solution for body sensor networks," in *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*. IEEE Press, Portland, Minneapolis, MN, 3–6 September 2009, pp. 2458–2461.
- [5] J. Shi, K.-Y. Lam, M. Gu, M. Li, and S.-L. Chung, "Towards energy-efficient secure communications using biometric key distribution in wireless biomedical healthcare networks," in *Proceedings of the International Conference on Biomedical Engineering and Informatics (BMEI)*. IEEE Press, Portland, Tianjin, China, 17–19 October 2009, pp. 1–5.
- [6] D. Kreiseler and R. Bousseliot, "Automatisierte EKG-Auswertung mit Hilfe der EKG-Signaldatenbank CARDIODAT der PTB," *Biomedizinische Technik/Biomedical Engineering*, vol. 40, no. 1, pp. 319–320, 2009.
- [7] B. Systems, Available at <http://biopac.com>.
- [8] D. Karaođlan Altop, A. Levi, and V. Tuzcu, "Towards using physiological signals as cryptographic keys in body area networks," in *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*. ICST, Belgium, 2015, pp. 92–99.
- [9] —, "Deriving cryptographic keys from physiological signals," *To appear in Pervasive and Mobile Computing*, doi: 10.1016/j.pmcj.2016.08.004, 2016.
- [10] P. D. Welch, "The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.