Nguyen, Ngu; Jähne-Raden, Nico; Kulau, Ulf; Sigg, Stephan

Representation Learning for Sensor-based Device Pairing

Published in:
2018 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2018

DOI:
10.1109/PERCOMW.2018.8480412

Published: 02/10/2018

Document Version
Peer reviewed version

Please cite the original version:
Nguyen, N., Jähne-Raden, N., Kulau, U., & Sigg, S. (2018). Representation Learning for Sensor-based Device Pairing. In 2018 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2018 (pp. 508-511). [8480412] Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/PERCOMW.2018.8480412

Representation Learning for Sensor-based Device Pairing

Ngu Nguyen*, Nico Jähne-Raden†‡, Ulf Kulau†, Stephan Sigg*
*Aalto University, Finland
†Technische Universität Braunschweig, Germany
‡Medizinische Hochschule Hannover, Germany
Email: le.ngu.nguyen@aalto.fi, nico.jaehne-raden@plri.de, kulau@ibr.cs.tu-bs.de, stephan.sigg@aalto.fi

***Abstract*—The emergence of on-body gadgets has introduced a novel research direction: unobtrusive and continuous device pairing. Existing approaches leveraged contextual information collected by sensors to generate secure communication keys. The secret information is represented throught hand-engineered features. In this paper, we propose a learning method based on Siamese neural networks to extract features that signify on-body context while separating off-body devices.**

I. Introduction

The advance of mobile technology and sensor manufacturing has facilitated the popularity of wearable devices. Nowadays, smartwatches are widely-used in daily activities due to their convenience and communication capability. In sport, sensor-equipped shoes are utilized to monitor and evaluate performance of both amateur and professional players. In healthcare, implant devices have been standardized and employed for a long time. Furthermore, research community in smart-textile envisionages the popularity of intelligent platforms embedded in clothes. The increasing number of device types with various use cases has posed a challenge: securely pairing them to form an adhoc network. The PIN approach was a common solution but it is obtrusive because of required user's input. Moreover, it is difficult to use with appliances that are lack of interactive interfaces such as implant devices.

Recently, contextual information has emerged as a source to produce secret keys for on-body device pairing. On-body devices have built-in sensors to collect contextual data. For instance, an off-the-shelf smartwatch can measure physiological information of its user, such as movement, heart-rate, and skin temperature. Body movement, e.g walking, is utilized as a source of key generation [1, 2, 3]. Nevertheless, the existing approaches can be employed only when users are moving and they rely on hand-drafted features. To resolve the first issue, heartbeats can be exploited when users stay still, as suggested in activity recognition [4]. In case of the latter, representation learning [5], which have been succeeding in a wide range of applications, is a promising approach to extract characteratic information from sensing data.

In this paper, we present a method to learn an optimal feature vector for contextual data in device pairing. If two sensors are on the same body, the learned representation promotes similarity of collected data. Otherwise, it aims to separate the features extracted from off-body devices. In order to fulfil the aforementioned objective, we propose to use a regularized Siamese network [6] whose branches are denoising autoencoders [7]. Our approach is usable not only for acceleration data but also for other modalities since the network architecture is applicable to various data types.

II. On-body Device Pairing

The popuparity of sensor-equipped wearable appliances has introduced a novel research problem: on-body device pairing. The existing protocols [8, 9, 1, 2, 3] capture sensing data from multiple devices at the same time to form characteristic sequences, called fingerprints. In the context of user authentication, these fingerprints are compared with a template database to recognize legitimate users or detect impostors. On the other hand, in device pairing, they are used to establish a secret key for device-to-device secure communication.

The existing approaches relied on correlation of movement data when devices were hold or worn by the same user. To correct a limited number of errors in fingerprints, error correcting codes and fuzzy cryptography are applied to create the identical secret key. Mayrhofer [8] proposes the candidate key protocol in which a user shaked devices together for a few seconds. One device hashes the acceleration sequence and then sent the hashed vector along with random salt to the other party. The latter looks up in its own processed data. If a match is found, the vector is appended in a key pool. As soon as a sufficient number of matched vectors is achieved, the key pool itself is hashed and becomes the shared secret key. Groza and Mayrhofer [9] later improved the protocol with heuristic tree and hashed heuristic tree to overcome the analysis of hash values performed by attackers. Walkie-Talkie [1] is another scheme that exploits correleted acceleration sequences collected by sensors when a user is walking. The authors used independent component analysis and low-pass filtering to remove undesired movements. Acceleration values are quantized as 0 or 1 conditioned on whether they are lower or higher than a threshold region. Bit sequences are then xorred to obtain the secret key. In BANDANA [2], keys reflect the difference between mean and instantaneous gait cycles. The approach exploits only acceleration along the z-axis. To further promote similarity of fingerprint generated on the same body, bits produced from low difference between mean and instantaneous gait are discarded. The Inter-Pulse-

Interval protocol [3] exploits the random residual by which individual steps (left and right) differ from the mean gait cycle in time domain. The key is formed from the first bits of a graycode of fingerprints.

All of the aforementioned approaches require the user to move, either by shaking the devices or walking. Thus, they expose information that can be captured and analysed by attackers, for example, with a highspeed camera. Furthermore, there exists a gap when the user stays till (e.g. sitting, standing, or lying). To handle these issues, heartbeats recorded by on-body accelerometers [10, 4, 11] can be the source of secret key generation. To handle the new data type, we propose to utilize a regularized Siamese network [6] for feature learning.

III. Regularized Siamese Networks

We would like to combine Siamese networks [12] and denoising autoencoders [7] to learn a optimal representation for sensor-based device pairing. The Siamese architecture encourages the discrimination of devices worn by different users while an autoencorder acts as an efficient fingerprint extraction mechinism. The network architecture is pictured in Figure 1, which is called a regularized Siamese network and was applied to extract speaker-specific information from audio data [6].

The Siamese network was presented by Bromley *et al.* [12] to verify hand-written signatures. The main components includes two neural network that share identical weight parameters. Given two inputs $\mathbf{x}_1$ and $\mathbf{x}_2$, these networks produce two corresponding outputs $\mathbf{o}_1 = f(\mathbf{x}_1)$ and $\mathbf{o}_2 = f(\mathbf{x}_2)$, respectively. The parameters are trained in such a way that the distance $d(\mathbf{o}_1, \mathbf{o}_2)$ reflects a similarity relation. In each branch of a Siamese network, the ouput is not a vector of posterior probabilities as in classification problems, but it is considered as a feature vector. Let $C = \{C_1, ..., C_K\}$ be the set of K classes in the data, o_1 the ouput of a reference sample x_1 of C_i, o_2 the ouput of another sample x_2 of the same class, and o_3 the ouput of a sample x_3 from any C_j where $i \neq j$. The goal of training a Siamese network is to maximize the disimilarity of inter-class samples while minimizing that of intra-class ones. For example, if $\mathbf{x}_1$ and $\mathbf{x}_2$ are signature images of the same person while $\mathbf{x}_3$ is that of an attacker, $d(\mathbf{o}_1, \mathbf{o}_2) < d(\mathbf{o}_1, \mathbf{o}_3)$ and $d(\mathbf{o}_1, \mathbf{o}_2) < d(\mathbf{o}_2, \mathbf{o}_3)$. In the training process, instead of using individual samples, *positive* and *negative* pairs of samples are required. Positive pairs contain two samples of the same class while negative ones include samples of different classes.

An autoencoder [7] is an unsupervised learning technique in which a feedforward non-recurrent neural network is trained to reproduce an input. It maps an input vector $\mathbf{x}$ to a hidden representation $\mathbf{o} = f(\mathbf{x}) = s(\mathbf{W}\mathbf{x} + \mathbf{b})$, where $\mathbf{W}$ is the weight matrix, $\mathbf{b}$ is the bias vector, and s is the activation function. Then, the resulting representation $\mathbf{o}$ is reconstructed to $\mathbf{x}^* = g(\mathbf{o}) = s(\mathbf{W}'\mathbf{x} + \mathbf{b}')$. It is possible that $\mathbf{W}' = \mathbf{W}^T$. The parameters $\mathbf{W}$, $\mathbf{W}'$, $\mathbf{b}$, and $\mathbf{b}'$ are optimized through minimizing the reconstructed error: $\mathbf{W}, \mathbf{W}', \mathbf{b}, \mathbf{b}' = \text{argmin} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(\mathbf{x}^{(i)}, \mathbf{x}^{*(i)})$, where n is the number of training samples and $\mathcal{L}$ is the loss function. The principle motivates that autoencoders can be applied to extract features from sensing data. We train a standard autoencoder as described above to reproduce heartbeat acceleration data. An example of original and reconstructed signal is showed in Figure 2. It illustrates that our model can reduce noise while amplifying peaks, which is useful for heartbeat detection and fingerprint generation.

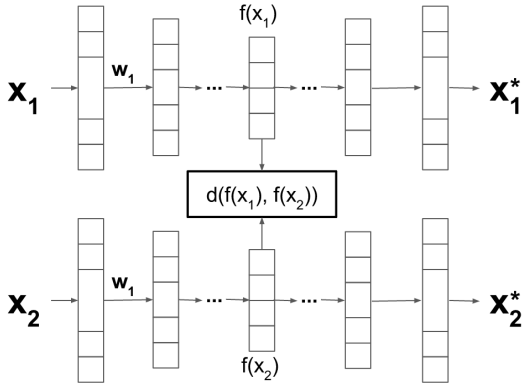

Fig. 1. Siamese auto-encoder

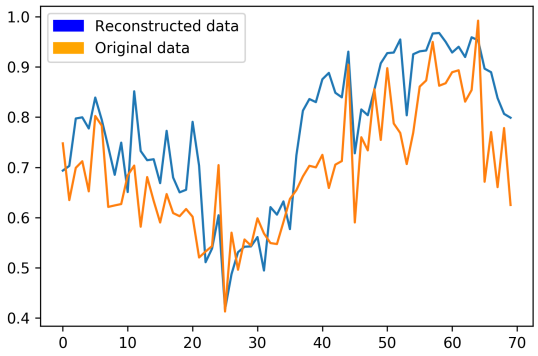

Fig. 2. Original and reconstructed acceleration signal (with auto-encoder)

IV. Ballistocardiography

The heart is a muscular organ which circulates blood throughout the whole body. In a single heartbeat, the ejection of blood into the great vessels produces subtle and repetitive motions. This is a physiological information that can be captured by non-invasive devices from the surface of the body. Ballistocardiography (BCG) [10] is the field that measures and analyses the signal. With the development of sensing technology, BCG can be performed through accelerometers embedded in objects that are in contact with human [13] or smartphones [4].

The BCG signal has been proved its usefulness in user authentication [14] and activity recognition [4]. In the former application, Vural *et al.* [14] investigated the identification of individuals with an accelerometer placed on the sternum. In their study, the subjects were instructed to sit stably. A three-axis accelerometer was placed over the chest to record

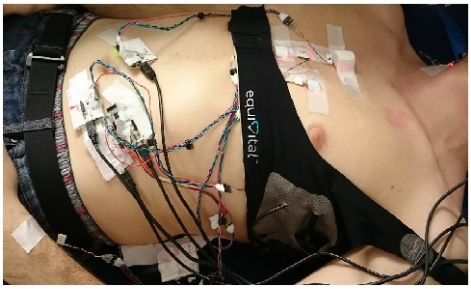

Fig. 3. Sensors are distributed on the subject's chest, neck, wrist, and back

tiny movements of the body caused by heartbeats. For feature extraction, the authors split each heartbeat into two regions, then computed the spectrogram matrix of each region before concatenating them. Fifty bins with highest relative entropy were selected as features for individual identification. A Gaussian mixture model was trained for each subject and a background model was generated to detect impostors. In the field of activity recognition, Hernandez *et al.* [4] demonstrated that wearable motion sensors could identify wearers and recognize their still body posture (sitting, standing, and lying). They attached commercial off-the-shelf devices (smartphone and smartglasses) to two body locations (head and wrist) for movement data collection. From each 10-second windowed data, the features were extracted: raw amplitude, 200-bin histogram of amplitude values, and shape descriptors (angles and distances between five descriptive points of each heartbeat). A linear Support Vector Machine was trained and tested in a cross-subject manner.

In this paper, we perform experiments on a 14-subject dataset collected by Jähne-Raden *et al.* [11]. The subjects were reported to not have any cardiovascular disease. They were students from Technische Universität Braunschweig. Their age range was from 21 to 34. Five of the subjects were female. Each subject was instructed to lie stably on a bed. Sensor arrays were attached to nine positions over a subject's chest, neck, wrist, and back as showed in Figure 3. Each array contained four accelerometers of different manufacturers, which were highly sensitive and had the sampling rate of up to 480Hz. The ground-truth data was collected by a BCG-based bed sensor and a medical ECG device.

We implemented an algorithm to extract heartbeat regions in acceleration data (see Algorithm 1). Our algorithm analyses a window of acceleration values to find a potential heartbeat. There are two tuneable parameters: window length and heartbeat distance. First, the input sequence is segmented into fixed-length windows and overlapping is possible. Then, the minimum value is located, which can be considered at the center peak in a heartbeat. After that, a region is expanded to cover the whole beat. This algorithm returns a list of starting and ending locations.

Algorithm 1: Algorithm of finding heartbeats in acceleration data

Input: a sequence of z-axis acceleration data $\mathbf{x}$, window length l, and beat distance d
Output: a list $\mathbf{l}$ of starting and ending positions (s_i, e_i) of found heartbeat regions

```
while not at end of x do
    extract a window w at the current position c;
    find position p of the minimun value in w;
    if beat found then
        extend around p to obtain (s_i, e_i);
        append (s_i, e_i) to l;
        jump to the next position c = c + d;
    else
        slide the window;
    end
end
```

V. Experiments

A. Model Specification

We experimented with acceleration data from the neck and the wrist since they are farthest pairwise locations in the dataset [11]. Using Keras [1] and Theano [2], we implemented a Siamese network and an autoencoder on heartbeat acceleration samples. Each input of the models contained three consecutive heartbeats. In case of the Siamese neural network, its branches are multilayer perceptrons (32 hidden units and rectified linear unit activation function). We split the pairwise data into training and testing set (75% and 25%, respectively) Other hyperparameters include: Euclidean distance as distance function, RMSProp as optimization algorithm, contrastive loss function, batch size 16, and 40 epochs. For the second model, we trained the autoencoder on wrist data and then evaluated on neck data. Its hyperparameters are: 32 hidden units, sigmoid activation function, Adam optimization algorithm, mean squared error loss function, batch size 128, and 1000 epochs.

B. Results and Discussion

In the task of discriminating on- and off- body sensor, our Siamese network achieves the following results: precision of 0.74 (training data) and 0.61 (testing data) and recall of 0.81 (training data) and 0.67 (testing data). We use the trained autoencoder to extract fingerprints from heartbeat acceleration data on two body locations (neck and wrist). Each fingerprint vector $\mathbf{f}$ are then transformed into a binary sequence $\mathbf{f}_b$. The transformation is: $\mathbf{f}_b(i) = 1$ if $\mathbf{f}(i) > \text{mean}(\mathbf{f})$ and $\mathbf{f}_b(i) = 0$ otherwise. This process can be performed independently in each wearable device. For evaluation, we compute Hamming distance between $\mathbf{f}$ on the same subject (two locations: neck and wrist) and on different subjects. Figure 4 displays the average Hamming distance in both cases, along with standard

[1] keras.io
[2] deeplearning.net/software/theano/

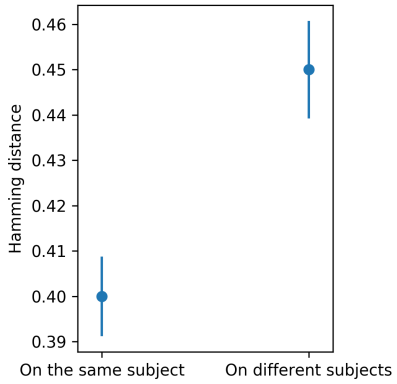

Fig. 4. Hamming distance on the same subject (two locations: neck and wrist) and on different subjects (lower is better)

errors. From the figure, the heartbeat fingerprints of the same user is more similar than on different users, 60% and 55% respectively. That suggests that an error correcting code technique can be applied to derive the secret key for secure on-body device pairing, for example as employed in [2].

VI. Conclusion

In this paper, we present a method based on Siamese networks and autoencoder to learn a fingerprinting scheme in device pairing with heartbeat acceleration data. If two devices are on the same subject, the learned fingerprints promotes similarity of collected data. Otherwise, it aims to separate the data extracted from different users. In future, we plan to implement the regularized Siamese network to improve the fingerprint quality.

References

[1] W. Xu, G. Revadigar, C. Luo, N. Bergmann, and W. Hu, "Walkie-talkie: Motion-assisted automatic key generation for secure on-body device communication," in *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2016.

[2] D. Schürmann, A. Brüsch, S. Sigg, and L. Wolf, "BANDANA – Body Area Network Device-to-device Authentication using Natural gAit," in *IEEE PerCom*, Mar. 2017, pp. 190–196.

[3] Y. Sun, C. Wong, G.-Z. Yang, and B. Lo, "Secure key generation using gait features for body sensor networks," in *IEEE BSN, 2017*, 2017, pp. 206–210.

[4] J. Hernandez, D. J. McDuff, and R. W. Picard, "Bioinsights: Extracting personal data from still wearable motion sensors," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, June 2015, pp. 1–6.

[5] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2013.50

[6] K. Chen and A. Salman, "Extracting speaker-specific information with a regularized siamese deep network," in *Proceedings of the 24th International Conference on Neural Information Processing Systems*, ser. NIPS'11. USA: Curran Associates Inc., 2011, pp. 298–306. [Online]. Available: http://dl.acm.org/citation.cfm?id=2986459.2986493

[7] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: ACM, 2008, pp. 1096–1103. [Online]. Available: http://doi.acm.org/10.1145/1390156.1390294

[8] R. Mayrhofer, "The candidate key protocol for generating secret shared keys from similar sensor data streams," in *European Workshop on Security in Ad-hoc and Sensor Networks*. Springer, 2007, pp. 1–15.

[9] B. Groza and R. Mayrhofer, "SAPHE: simple accelerometer based wireless pairing with heuristic trees," in *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia*. ACM, 2012, pp. 161–168.

[10] I. Starr, A. J. Rawson, H. A. Schroeder, , and N. R. Joseph, "Studies on the estimation of cardiac output in man, and of abnormalities in cardiac function, from the hearts recoil and the bloods impacts; the ballistocardiogram," *The American Journal of Physiology*, vol. 127, no. 1, Nov. 1939.

[11] N. Jähne-Raden, U. Kulau, L. Wolf, and M. Marschollek, "Poster abstract: Heartbeat the odds – a novel digital ballistocardiographic sensor system," in *Proceedings of the 15th annual international conference on Embedded Networked Sensor Systems*. ACM, nov 2017, accepted for publication.

[12] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a "siamese" time delay neural network," in *Proceedings of the 6th International Conference on Neural Information Processing Systems*, ser. NIPS'93. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993, pp. 737–744. [Online]. Available: http://dl.acm.org/citation.cfm?id=2987189.2987282

[13] W. K. Lee, H. Yoon, D. W. Jung, S. H. Hwang, and K. S. Park, "Ballistocardiogram of baby during sleep," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 7167–7170.

[14] E. Vural, S. Simske, and S. Schuckers, "Verification of individuals from accelerometer measures of cardiac chest movements," in *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG)*, Sept 2013, pp. 1–8.