ELPA: A NEW KEY AGREEMENT SCHEME BASED ON LINEAR PREDICTION OF ECG FEATURES FOR WBAN

Emna Kalai Zaghouani[*†], *Abderrazak Jemai*[‡], *Adel Benzina*[*], *Rabah Attia*[*]

[*] Laboratory of Electronic Systems and Communication Network, TUNISIA Polytechnic School
[†] TELNET Innovation Labs, Telnet Holding Tunis, Tunisia
[‡] Laboratoire LIP2, Faculté des Sciences de Tunis, 1060 Belvédère Tunis, Tunisie
Emna.kalai@telnet-consulting.com, Abderrazak.Jemai@insat.rnu.tn, Adel.Benzina@isd.rnu.tn,Rabah.attia@enit.rnu.tn

ABSTRACT

In this paper, we propose a novel key agreement scheme called ECG Linear Prediction key Agreement (ELPA) with the properties of plug-n-play and transparency to secure inter-sensor communication in Wireless Body Area Networks (WBANs). ELPA is a new physiological based key agreement scheme allowing two nodes belonging to the same WBAN to agree on a symmetric key from ECG signal features. The paper introduces the use of Linear Prediction Coding (LPC), which has always been used for a compression purpose, in hiding the cryptographic key. In fact we prove that concealing the symmetric key using this tool ensures high security level while keeping low computational complexity and communication overhead compared with the state of the art.

Index Terms— WBAN, communication security, ECG features, LPC, communication overhead

1. INTRODUCTION

WBAN, a promising new research area, is defined as a network consisting of intelligent, low-power, micro technology sensors and actuators which can be placed on the body, providing timely data [1]. The WBAN nodes are interconnected and connected to a coordinator node via a wireless communication technology like Bluetooth and Zigbee [1]. Considered as a mistrusted channel, and while WBAN deals with sensitive medical data, securing the inter-sensor communications remains a major issue. Communication is encrypted with a shared symmetric key thats why a secure management and agreement of this shared symmetric key must be performed. Usual methods to secure inter-sensor communications have been based on symmetric key distribution or predeployment mechanisms. However the former requires secure key distribution scheme and the latter requires a set up and rekeying for every node management, whereas nodes in WBAN must be removed, added and tuned in a secure and transparent manner : without the user participation. The recent proposed solution, to deal with key management problem, is physiological based security. The main idea of physiological based key agreement is to allow nodes belonging to the same WBAN to agree on a symmetric cryptographic key from the common physiological features. The collected features at different parts of the body are enough similar for an authentication purpose but not exactly identical to generate the same key sequence. This is due to noise caused by analog to digital converter and muscles contractions. The already used technique to overcome this problem is the fuzzy vault scheme [2]. The major weakness of this method is the size of the vault, a random set constructed to conceal the features, which is closely correlated to communication overhead, primary cause of energy consumption. In this paper we investigate the use of linear prediction coding (LPC) to achieve feature set locking minimizing greatly the communication overhead. ELPA does not require any hiding of the features in a vault set which reduces greatly the message size. The proposed algorithm uses the already developed feature extraction stage AC/DCT [3]. ELPA is first assessed in the MIT-BIH Normal Sinus Rhythm Database [4] containing recordings of healthy subjects and second in the MIT-BIH Arrhythmia Database [4]containing records affected by heart diseases to evaluate the robustness of the proposed method. The remainder of this paper is organized as follows. Section 2 presents an overview of the usual methods of securing inter-sensor communication and the existing works in the new promising area: biological key. In section 3 we present the whole proposed scheme and detail each system bloc. Finally we estimate the performance of the proposed system compared with the most known one the PSKA [5] in term of overhead and security.

2. RELATED WORK

All the security requirements for communications can be fulfilled if a key is successfully and securely distributed. Usually used, predeployed keying mechanism was always the primary solution to secure communication in a Wireless sensor Network. It consists of distributing initial keys to all the participating sensors in the set-up phase [6]. The trend has then

This work is supported by Telnet Innovation labs and the PASRI program funded by the European Union.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

been moved to symmetric keying agreements; most known are SPINS [7]. Even if the aforementioned techniques are attractive due to their energy efficiency, limitations have been exhibited in the used key sharing protocol. Later many studies explored the alternative of using a public-key infrastructure for key distribution. But public key distribution cannot be recommended for WBAN because of their high computational complexity and low power efficiency. The recent trend for secure key management is the use of biometric sensed signal, not available to all other kinds of wireless networks, to generate the symmetric cryptographic key. As nodes belonging to the same WBAN have the advantage of measuring the same signal, key can be extracted and shared in a transparent manner. The use of biometric key was first introduced by [8]. The author describes the biometric measurement, randomness of biometric signals and the use of error correcting codes as the measurement of biometrics is never perfect. [9] propose to use the heart inter-pulse interval IPI which can be extracted from multiple sensors such as electrocardiogram ECG, photoplethysmogram PPG, heart sounds, blood pressure waves. IPI cannot satisfy the distinctiveness criteria and has to be collected in a synchronous way [9]. [10] describes the idea of, rather than using the ECG or PPG signal as a cryptographic key, using it to hide a random generated key. The idea was fully described and implemented in [5]. The agreement process is explained hereafter. On the transmitter side: (1)generating a polynomial by a random choice of its coefficients,these coefficients construct the random key to hide;(2) generating a feature vector obtained from the physiological signal;(3) locking the features by projecting them on the chosen polynomial;(4) adding random points to construct the vault;(5) sending the vault. An intruder MITM, intercepting the vault, cannot distinguish the random values from the legitimate ones. Adding chaff points is known by fuzzy vault scheme and was first presented by [2].On the other side the receiver generates its own version of the feature vector, since it can access the same physiological signal, it starts to unlock the vault by extracting the common features. It can subsequently reconstruct the polynomial coefficients so the key. However it is proven that security strength of PSKA heavily depends upon the vault size [5], a major issue closely correlated with the communication overhead primary cause of energy consumption. Hu et al. [10] prefers later not to use the polynomial locking. Based only on sending features on an ordered way, the sender can check up the compliance of the relevant common set and send an acknowledgement to receiver. Based also on concealing the features in a large random set, OPFKA shows also its weakness in communication overhead.

3. ECG LINEAR PREDICTION KEY AGREEMENT

The purpose of our conceived algorithm ELPA is to perform a secure and transparent node pairing by generating a symmetric key based on the ECG captured signal. We assume that every node, including the coordinator node, has the ability to measure the ECG signal. The main contribution of ELPA system is the use of LPC [11] as a technique to hide the shared symmetric key also called session key. This is meant to replace the adding of chaff points to reduce the communication overhead. If a transmitter node N_T requests to establish a connection with a receiver node N_R, which can be a coordinator one, N_R has to verify that N_T can generate the same session key without sending it. This proves that the two nodes belong to the same BAN. The key sharing process works as detailed in Figure 1: the transmitter begins by measuring an ECG sequence for a fixed time T_s. The vector X containing N samples is then calculated. The features F_1, a representative set of the captured ECG, are so extracted. F_1 is the principle data to construct the session key. F_1 will be linearly predicted to produce two main parameters: A, the LPC coefficients vector and E the residual error of predicting F_1. A will be sent to the receiver and E will be transformed by the key generation process to generate a unique 128 bits session key.

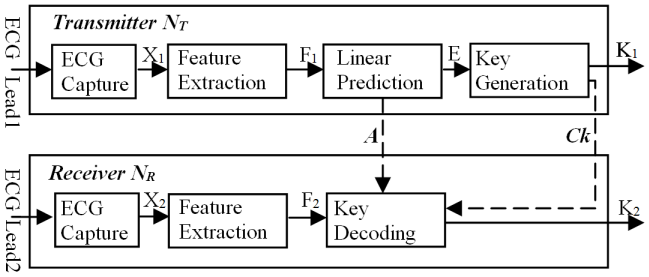

Fig. 1. ECG linear prediction-based key agreement.

By sending only the A vector the secret session key is hidden and protected from eavesdroppers and intruders attacks. To adjust and correct the few fault bits we choose to protect the generated key K_1 by the mean of BCH coding. The BCH suffix is then sent to the receiver and added to the receiver key version K_2.On the receiver side and in synchronous manner, ECG is captured and feature extraction is performed. The key decoding stage reconstructs the 128 bits session key. In fact having its own feature version and helped by the LPC coefficients generated on the transmitter side, the receiver can reconstruct his own version of the symmetric key. We address by our proposed method the communication overhead issue which is caused essentially by adding the chaff points, random feature vectors. We show that in this case overhead is limited.

3.1. Feature Extraction

Feature extraction is extracting the relevant data from the captured signal capable of identifying nodes belonging to the same WBAN system. Authors in [3] prove that AC/DCT technique is a promoting solution. It is proven that autocorrelation coefficients embed similarity features among

records of the same subject. Algorithm1 explains the method

Algorithm 1 Feature Extraction

1: The sender and receiver collect ECG signal for Ts
2: Signal filtering
3: Compute AC = [AC_1,,AC_L] , L auto-correlation coefficients
4: Compute DCT = [DCT_1,,DCT_L] DCT transformation of the AC vector.
5: Truncate DCT to F_q the first K elements of DCT vector with q = 1 if transmitter and q = 2 if receiver

involving 4 main stages: signal acquisition, signal preprocessing, AC calculation and finally DCT calculation. The signal is first captured on the transmitter and receiver side (two different ECG leads) in a synchronous manner for the same period T_s. T_s is chosen so that multiple pulses are included in the windowed signal. The signal is then filtered for noise reduction using a median filter. The auto-correlation coefficients are calculated (1):

$$\hat{R}_x[m] = \sum_{i=0}^{N-|m|-1} x[i]x[i+m] \quad (1)$$

Where $x[i]$ is the windowed ECG, $x[i+m]$ is the time-shifted version of the windowed ECG with a time lag of $m = 0, 1, ..., L-1$, $L << N$. A frequency domain transformation is then performed via discrete cosine transform. The DCT vector is estimated as (2) :

$$Y[u] = G(u) \sum_{i=0}^{N-1} y[i] \frac{\pi cos(2i+1)u}{2N} \quad (2)$$

For the AC/DCT method $y[i]$ is the auto correlated ECG obtained from (1). $G[u]$ is calculated from the equation cited in [3]. This stage reduces greatly the auto-correlation dimensionality. In fact only $K << L$ non zero DCT coefficients will contain significant information. This is illustrated by Figure 2.

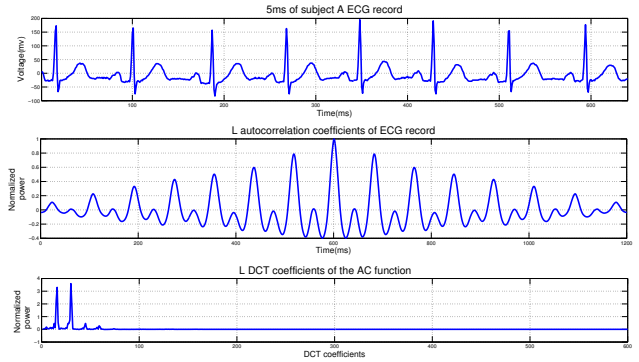

Fig. 2. Dimensionality reduction and feature extraction.

3.2. Linear Prediction Coding

The core of the proposed scheme is linear prediction. Mostly used in audio signal and speech processing, linear prediction tries to find parameters of a linear model that could reproduce the most faithfully the original signal [11]. The basic block is a FIR filter of appropriate order p. In the present case, the DCT coefficients of the AC ECG signal are linearly predicted. As explained in Figure 1 the F_1 vector is predicted to generate two main components E and A where E is the error of prediction also called the excitation and A the coefficients of the prediction filter. The prediction error E is calculated as described in (3):

$$e_n = e(n) = y(n) - \hat{y}(n) = \sum_{i=0}^{p} a(i)y(n-i) \quad (3)$$

Where $\hat{y}(n)$ is the predicted signal and $a(i)$ are the coefficients of the prediction where $a(0)$= 1 and p the prediction order. To recover the signal the two components: LPC coefficients and excitation signal are mandatory. One of the important decisions that usually have to be made in the linear prediction is the determination of the optimal order of prediction [11].

In this paper and for most ECG signals we found that an order higher than 4 show no significant improvement in modeling quality. The vector A is sent to the receiver and residual prediction error E is transformed to construct the 128 bits symmetric key.

3.3. Key generation

The purpose of this stage is to transform the resulting prediction error composed of K elements to a 128 bits vector and to protect the generated key with a BCH coding. It is evident that most residual error signals swing around zero. So we choose to adopt a pulse-code train transformation. A pulse-code train is defined as positive values are coded as +1 and negative ones as zero.

Algorithm 2 Key Generation

1: **for** each element of E = [e_1,,e_K] **do**
2: **if** $e_i \geq 0$ **then**
3: $I_i \leftarrow +1$
4: **else**
5: $I_i \leftarrow 0$
6: **end if**
7: **end for**
8: Start BCH encoding
9: **for** each k elements of K_1 = [I_1,, I_K] **do**
10: BCH encode
11: Save the n- k last elements BCH code in Ck_p
12: **end for**
13: Concatenate Ck_1,..,Ck_P to construct Ck

As the proposed algorithm is based only on the comparison of the 128 bits keys K_1, K_2 and as a single bit error

is enough to reject legitimate nodes, its recommended to include an error correcting code. Several codes are used in the literature and their choice is a compromise between complexity and robustness. In the actual use case a BCH coding is sufficient to perform error correction with variable corrective power. For each k codeword an n-k suffix is calculated in the encoding phase. The n-k suffixes will be than concatenated to construct Ck vector. Ck is then sent to the receiver to adjust the eventual errors in the generated key. Figure 3 shows the data stream to be sent after key generation and BCH coding. ID_s and ID_R are the sender and receiver identifiers and No

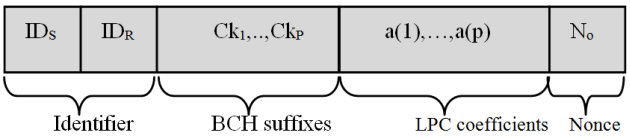

Fig. 3. Data stream after key generation and BCH coding.

is a nonce, random generated number to guarantee freshness of the sent message.

3.4. Key decoding

The receiver node N_R starts by extracting his own version of the features F_2 from the sensed ECG. It can then reconstruct the prediction error. The excitation signal is a subtraction of the original signal and the predicted one as explained in (1). The original signal is the vector F_2 slightly different from the original one F_1 and the predicted one is as explained in (3), where $a(i)$ are the received coefficients A and y is receiver version of the features. Algorithm 3 explains key decoding procedure in details. Key generation is followed by BCH (n,k) decoding to adjust the eventual errors in the calculated key.

Algorithm 3 Key decoding

1: The receiver receives the message:[ID_s,ID_r,Ck,A]
2: Compute $\hat{E}$ = [$\hat{e}_1$,,$\hat{e}_K$] the prediction error result of LPC synthesis of the extracted F_2 and the received A.
3: **for** each element of $\hat{E}$ = [$\hat{e}_1$,,$\hat{e}_K$] **do**
4: **if** $\hat{e}_i \geq 0$ **then**
5: $\hat{I}_i \leftarrow +1$
6: **else**
7: $\hat{I}_i \leftarrow 0$
8: **end if**
9: **end for**
10: **for** each k elements of $\hat{I}$ = [$\hat{I}_1$,, $\hat{I}_K$] **do**
11: Concatenate $\hat{I}_p$ and Ck_p to form Ny_p the noisy message
12: BCH decode of Ny_p to reform the original I_p
13: **end for**
14: Reconstruct K_2 = I_p,p=1..K/k

4. PERFORMANCE EVALUATION

This section is dedicated to evaluate the proposed algorithm ELPA by addressing three main characteristics: 1) Security; 2) Computational complexity; 3) Communication overhead. The test bed ECG signals are obtained from the Physionet database. We will assess the generated key by performing a series of experiments on two sets of databases: the MIT-BIH Normal sinus Rhythm Database [4] containing 36 ECG recordings sampled at 128 Hz, two ECG lead records for each subject and the MIT-BIH Arrhythmia Database [4] containing 48 from different subjects sampled at 360Hz, each record consists of two ECG leads.

4.1. Security

The linear prediction locking make it very difficult to adversaries to know the key agreed upon. Only nodes which have the capabilities of measuring the same ECG signal can unlock the established symmetric key. Thanks to temporal variation of the collected data, the system is protected from replay attacks. If an intruder tries to capture and replay the exchanged message containing essentially the LPC parameters, it will be discarded by any receiver due to the difference in the captured ECG signal, so in the predicted parameters. Our proposed system is also protected from MITM attacks. As the system is based on linear prediction, having only the LPC parameters cannot allow to recover the signal. The eavesdroppers must have the original signal to synthesize the prediction error.

4.2. Computational complexity

The main contribution of the ELPA algorithm is the outstanding reduction of the computational complexity and communication overhead compared with the previous proposed solutions as PSKA [2]. PSKA is based on the locking of conventional key using a mathematical tool: polynomial root finding. The performance of the system is optimized when using higher orders of the polynomial which means a higher computational cost. In the proposed PSKA algorithm polynomial order can reach 12. On the other side, locking the session key in the ELPA algorithm consists only on a linear prediction. LPC parameters are simply determined by minimizing the mean square value of the residual error. It leads to solve linear equations by using LevinsonDurbins algorithm. As the used prediction order is 4, we can prove easily that solving four linear equations is much less complex than Lagrangian interpolation algorithm for decoding polynomials from their projections.

4.3. Communication overhead

The security strength of PSKA completely depends on the vault size. Firstly this can cause collision between features which leads to false positives. Secondly a chaff point takes in

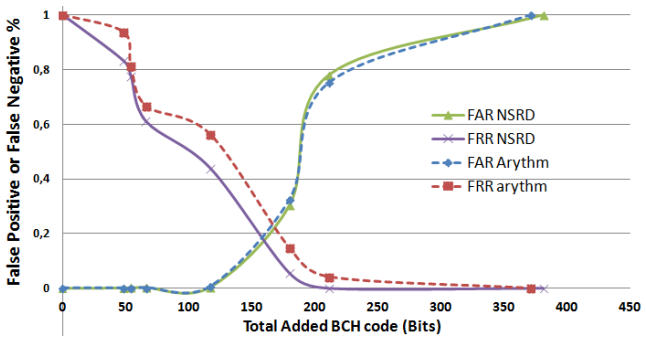

Fig. 4. ELPA False Positive and False Negative rates.

PSKA 36 bits and a vault size for stronger security can reach 5000. OPFKA, proposed later, enhances the PSKA algorithm however the results in terms of memory storage and communication overhead are still unsatisfying as the algorithm in [10] is based also on concealing features in a vault set. The main contribution of the proposed scheme is that it can replace the hiding of the features with chaff points by using linear prediction coding as locking method. The sender has only to send the LPC generated parameters. Using 2 bytes to code the $a(i)$ coefficients seems to be adequate in the present case. If we choose a prediction order equal to four, vector A takes only 8 bytes. Adding the BCH suffixes consumes only (n-k) * K/k bits. For example for a BCH (64 , 36) for a key length of 144 only 14 bytes are added.

This is to be compared with the communication overload of the PSKA or the OPFKA algorithms, that reach rates in Kbytes of communication overhead [10]. Results in terms of False Acceptance Rate FAR and False Rejection Rate FRR still promoting compared with the PSKA ones. PSKA reach an optimum FAR and FRR of 0.2 for a polynomial order equal to 14. Figure 4 shows the results of experiments done on two sets of database. The Normal Sinus Rhythm Database containing records from 18 subjects. The developed algorithm can reach an FAR and FRR equal to 0.2 for an added BCH bits not exceeding 20 Bytes. We have also tested the performance of our system on the arrhythmia database containing 48 records of subjects having heart disease. This second step of assessment is mandatory as the aim of our algorithm is to be implemented in health monitoring system worn mostly by unhealthy subjects. Figure 4 illustrates the robustness of the proposed system ELPA to arrhythmia problems.

Compared with the previous solutions, the ELPA algorithm ensures a higher level of security at a lower computational complexity and communication overhead while conserving the same FAR and FRR rates.

5. CONCLUSION

In this paper we present a secure and lightweight key agreement scheme called ELPA. ELPA allows two nodes, including the coordinator node, to agree upon a symmetric key based on the similarity in the measured ECG signal at different leads. The major contribution of the conceived algorithm is the use of LP Coding to hide the common features. We have proved by assessing the algorithm in two different databases that ELPA is minimizing greatly the energy consumption by reducing the communication overhead. We have also proved that the proposed method keeps a high security level by resisting MITM and replay attacks.

REFERENCES

[1] S.Movassaghi, M.Abolhasan, J.Lipman, and D.Smith, "Wireless body area networks: A survey," *Communications Surveys & Tutorials, IEEE*, 2014.

[2] A. Juels and M. Sudan, "A fuzzy vault scheme," *Designs, Codes and Cryptography*, vol. 38, no. 2, 2006.

[3] Y. Wang, F. Agrafioti, D. Hatzinakos, and K.N. Plataniotis, "Analysis of human electrocardiogram for biometric recognition," *EURASIP journal on Advances in Signal Processing*, vol. 2008, pp. 19, 2008.

[4] "Normal sinus rhythm and arrhythmia databases of mitbih," www.physionet.org/physiobank/database.

[5] 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 Trans.*, vol. 14, no. 1, 2010.

[6] P Boyle and T Newe, "Security protocols for use with wireless sensor networks: A survey of security architectures," in *Wireless and Mobile Communications, 2007. ICWMC'07.* IEEE, 2007.

[7] A. Perrig, R. Szewczyk, JD Tygar, V. Wen, and D.E. Culler, "Spins: Security protocols for sensor networks," *Wireless networks*, vol. 8, no. 5, pp. 521–534, 2002.

[8] S. Cherukuri, K.K Venkatasubramanian, and S.KS Gupta, "Biosec: A biometric based approach for securing communication in wireless networks of biosensors implanted in the human body," in *Parallel Processing Workshops, 2003.* IEEE, 2003.

[9] S. Bao, C.CY. Poon, Y. Zhang, and L. Shen, "Using the timing information of heartbeats as an entity identifier to secure body sensor network," *Information Technology in Biomedicine, IEEE Trans.*, vol. 12, no. 6, 2008.

[10] C. Hu, X. Cheng, F. Zhang, D. Wu, X. Liao, and D. Chen, "Opfka: Secure and efficient ordered-physiological-feature-based key agreement for wireless body area networks," in *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 2274–2282.

[11] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.