Novel Efficient and Privacy-Preserving Protocols For Sensor-Based Human Activity Recognition

Zakaria Gheid, Yacine Challal

Novel Efficient and Privacy-Preserving Protocols For Sensor-Based Human Activity Recognition

Zakaria Gheid*§, Yacine Challal*‡
*Ecole nationale supérieure d'informatique
Laboratoire des Méthodes de Conception des Systèmes, Algiers, Algeria
‡Centre de Recherche sur l'Information Scientifique et Technique, Algiers, Algeria
Email: §z_gheid@esi.dz , ‡y_challal@esi.dz

Abstract—**Human activity recognitin (HAR) has become an important emerging field of application for sensor networks (SN) technologies. Nevertheless, the pervasiveness of SN in everyday life has given rise to new privacy concerns especially when mining personal sensed data in external environments. From that perspective, many research works have proposed cryptography-based techniques so as to tackle SN privacy issues, yet have costed significant degradations in computational-time efficiency. In this work, we propose a novel privacy-preserving Knn classification protocol to be used in HAR process and that is based on a novel privacy-preserving protocol that aims to assess similarity between personal recorded activities and ~~extern~~ patterns using the cosine similarity metric. We build our proposals without any cryptographic schemes in order to provide a high efficient recognition service.**

Index Terms—**Human activity recognition (HAR), sensor networks (SN), Knn classification, cosine similarity, privacy, efficiency.**

1. Introduction

The last recent years have seen a huge advancement in sensing and communication technologies, leading to a ubiquitous computing era where sensor networks (SN) are becoming smarter, more connected and allowing to track anything, anytime and everywhere.

These developments in SN span a wide range of applications that support innovative services in all areas of life. Particularly, human activity recognition (HAR) was an emerging research field that aims to mine pervasive data streams collected by wearable and implantable sensors so as to provide more understand of human activities and behaviours. This may improve the quality of individual life in several aspects, ranging from daily assisted living to leisure applications. For instance, most elderly people prefer to stay in their own homes as they age [1], but living individually can be scary as a simple fall may induce injuries, which is fatal for their lives if not assisted early. To shed some light on this, Centers for Disease Control and Prevention reported that 2.5 million of older people fall each year, but less than half could tell their doctors [2]. Therefore, an emergency HAR system that recognizes falls and abnormal activities using SN may save many elderlies' lives. From another side, people wanting to stay well and be healthy need to follow a lifestyle management program, which should include a self-monitoring of caloric intake related to their daily physical activities. For this purpose, a sensor-based HAR system may be useful so as to track daily activities, set reminders and give recommendations [3]. Likewise in public security, transportation and urban management, tracking people activities and mobility could be exploited to great social benefits [4].

However, the pervasive nature of sensor-based HAR systems raised in privacy concerns surrounding tracking people's actvities and locations. These concerns encompass especially storing, communicating and mining sensed data in external environments.

From the perspective of research, many proposed works [5], [6] have implemented cryptographic schemes, such as the Paillier [7] and ElGamal [8] cryptosystems in order to tackle the privacy issues in SN. Nevertheless, these cryptogaphy-based security measurments costed a significant degradation in response time, trading so security and performance. Such a trade-off may be intolerable in emergency situations where instant decision is vital.

In this work we propose a diferent approach that enhance both security and performance measures. Using two novel proposed protocols that are free from cryptography, we aim to add a significant improvement to the recognition process of HAR systems. The contribution of this work can be summarized as follows

- We propose (Π-CSP+): a novel privacy-preserving and efficient cosine similarity protocol that aims to asses similarity between sensed activities and external patterns.
- We integrate the above proposed (Π-CSP+) in a novel privacy-preserving and efficient Knn classification protocol named (Π-Knn) so as to classify the sensed activities according to extern patterns held by a service provider.
- We make evaluations of our proposals proving their high security as well as their efficiency level comparing to other proposals.

The rest of the paper is organized as follows. Section 2 presents preliminaries used to introduce our proposals. In Section 3, we highlight the privacy concern raised by the HAR classification process and present our privacy-preserving proposed protocols. In Section 4, we give a formal security proof of our protocols using the real/ideal simulation paradigme and Section 5 is devoted to the performance evaluation accross different experimental tests. In Section 6, we provide a literature survey of related works and we discuss their lacks. We conclude by summarizing the contributions of this work.

2. Preliminaries

In this section we present preliminaries and building blocks used later to implement our proposals.

2.1. Sensor-based HAR

Sensor-based HAR systems aim to retrieve information about performed activities from raw sensor data gathred by wearable or implantable sensor networks (SN). The general structure of a sensor-based HAR process encompasses four main steps.

- In **preprocessing** phase, sensed data signals pass by different filters so as to remove frequencey noises while preserving useful information.
- During **segmentation**, continuous sensed data streams are splitted according to fixed time-series.
- During **Feature exraction**, different methods are applied to each time window (a set of time-series) transforming the large data raws into vectors of quantitative features (mean, variance,...etc.).
- **Classification** consists of applying different methods on the set of feature vectors in order to recognize the performed activities.

In this work we focus on securing the classification phase as all other phases are performed locally. Let HARP denote the problem of classifying activities in a HAR process. We define HARP as follows [9]

Definition 1 (HARP)**.** *Let $A = \{a_1, ..., a_k\}$ denote set of activities' labels and $W = \{w_0, ..., w_n\}$ a set of n time windows equal in size. We assume each t_i includes a set of time series $S_i = \{s_0^i, ..., s_m^i\}$ from the m measured attributes. Then, the HAR problem returns to find a mapping function $f : S_i \mapsto A$ such that $f(S_i)$ is as similar as possible to the activity performed in t_i*

2.2. K-nearest neighbors (Knn) classifiers

Knn algorithm is one of the instance-based [10] classifiers that could be used to classify activities according to trainning examples called instance space. Using a distance/similarity metric, Knn classifies a new instance (represented by a feature vector) by locating the k nearest instances (neighbours) having a same class in the instance space, then, labelling the unknown instance with the same class label of the located neighbours. In this work, we leverage the use of activity patterns as instace space instead of personal training examples so as to avoid the training phase required by such classifiers. Let D= $\{(x_1, y_1), ..., (x_n, y_n)\}$ denote a set of instance space involving n activity patterns where x_i and y_i correspond to the pattern data and the pattern class respectively. Assume $z = (Xz, Yz)$ a new activity instance where Xz denotes the extracted feature vector and Yz the activity class we are searching for. We define the set of points x for which a function f reaches its largest value as

$$\underset{x}{\text{argmax}} f(x) = \{x | \forall y : f(x) \geq f(y)\}$$

and we define I, the identity function as

$$x \in \{true, false\} \mapsto I(x) = \begin{cases} 1, & \text{if } x \text{ = true} \\ 0, & \text{if } x \text{ = false} \end{cases}$$

A detailed implementation of Knn is given in Algorithm1.

```
Algorithm 1: knn classification
Input  : $D$, $z$ and $k$, where:
         $D = \{(x_1, y_1), ..., (x_n, y_n)\}$, $z = (Xz, Yz)$
         and $0 < k \leq n$.
Output: $yz$, the class label of $z$.
  Compute d($Xz, x_i$), the distacnce/similarity between
    $z$ and every object in $D$.
  Select $Dz \subseteq D$, the set of $k$ closest objects to $z$.
  $Yz = \underset{c}{\text{argmax}} \sum_{(x_i, y_i) \in Dz} I(c = y_i)$.
```

2.3. Cosine similarity metric

Cosine similarity is a statistical metric used to assess similarity in vector space model. It operates by measuring the cosine (cos) of the angle between two vectors, thus, the more it is closer to 1 the more vectors are similar. Assume $\vec{a} = (a_1, .., a_n)$ and $\vec{b} = (b_1, ..., b_n)$ two numerical vectors. Let $(\vec{a} \cdot \vec{b})$ denote the scalar product and $\|\vec{a}\|$ (resp.$\|\vec{b}\|$) denote the Euclidean norm. Cosine similarity between $\vec{a}$ and $\vec{b}$ is measured by

$$\cos(\vec{a}, \vec{b}) = \frac{(\vec{a} \cdot \vec{b})}{\|\vec{a}\| \|\vec{b}\|} \tag{1}$$

while the scalar product is get by

$$(\vec{a} \cdot \vec{b}) = \sum_{i=1}^{n} (a_i \times b_i) \tag{2}$$

Notice that when we deal with normalized vectors, the cosine metric is shortened to the scalar product itself. Assume $\hat{a} = (\vec{a}/\|\vec{a}\|)$ and $\hat{b} = (\vec{b}/\|\vec{b}\|)$ the normalized representation of $\vec{a}$ and $\vec{b}$ repectively. Then

$$\cos(\vec{a}, \vec{b}) = (\hat{a} \cdot \hat{b}) \tag{3}$$

In this work, we use the cosine metric as a similarity function within the Knn process (see instruction 1, Algorithm 1). We leverage the use of this metric because of its high accuracy level when evaluated in such a context [11].

3. Novel protocols for efficient and privacy-preserving HAR

In this section, we highlight the privacy concern raised by the HAR classification process, then, we introduce two novel protocols that aim to preserve personal data privacy with a low time-computation cost.

3.1. Privacy problem statement

In a context when a HAR system does the classification step according to extern patterns, it should collaborate with a patterns service-provider in order to assess similarity between a new recorded activity and each class patterns according to a distance/similarity metric (see task 1 of Algorithm 1). To clarify this, let $P1$ and $P2$ denote respectively a service-provider of activity patterns and a HAR system. Assume $\vec{p_j} = (p_{j,1}, ..., p_{j,n})$ the pattern of the activity class j held by $P1$ and $\vec{z} = (z_1, ..., z_n)$ a new activity recorded by $P2$. As we chose to use the cosine metric (see Section 2.3) because of its high accuracy level[], we formalize the collaboration between $P1$ and $P2$ as

$$\cos(\vec{p_j}, \vec{z}) = (\hat{p_j} \cdot \hat{z_n}) \tag{4}$$

where $\hat{p_j}$ and $\hat{z_n}$ denote the normalized representation of $\vec{p_j}$ and $\vec{z}$ respectively. Such a collaboration specifying that one party ($P1$ or $P2$) should disclose its vector to its collaborator (see equation 2) is a privacy issue for both parties since $P1$ may provide a commercial service and $P2$ is recording personal private data such as location. Thus, in order to allow this computation while preserving data privacy for both collaborator parties, we propose Π-CSP+, a novel privacy-preserving and efficient cosine similarity protocol that will be used later to implement Π-Knn, a novel efficient and privacy-preserving Knn protocol for a HAR classification context.

3.2. Π-CSP+: privacy-preserving and efficient cosine similarity protocol

In order to introduce our proposed Π-CSP+, let us consider two parties $P1$ and $P2$ having respectively $A = \{\vec{a_1}, ..., \vec{a_s}\}$ and $B = \{\vec{b_1}, ..., \vec{b_p}\}$ sets of object vectors and want to securely assess similarity between their objects. Assume for $1 \leq i \leq s$ and $1 \leq j \leq p$: $\vec{a_i}$ and $\vec{b_j} \in \mathbb{R}^n$ and they have the same structure. In order to shorten our focus to the privacy concern in the scalar product (see section 3.1), we consider both parties collaborate with normalized vectors (see section 2.3). Let $\hat{A}$ and $\hat{B}$ denote the normalized sets of A and B respectively. We define $M_R[s \times s]$, $M_A[s \times n]$ and $M_B[n \times p]$ as matrix tools used during the privacy-preserving scalar product process, where M_R is a random noise, M_A involves the s normalized object vectos get from $\hat{A}$ and put as rows and M_B includes the p normalized object vectors get from $\hat{B}$ and put as columns. Assume M_R is an invertible matrix, $(s, n, p) \in N^{3*}$ such as: $\{1 < s < n, 0 < p < s\}$. The detail of Π-CSP+ implementation is provided in algorithm 2.

Algorithm 2: Π-CSP+, a Privacy-preserving and Efficient Cosine Similarity Protocol

```
Input  : $A = \{\vec{a_1}, ..., \vec{a_s}\}$ $P1$ object vectors
         $B = \{\vec{b_1}, ..., \vec{b_p}\}$ $P2$ object vectors
Output: (For $P1$ only)$M_{AB}[s \times p]$ containing the
         cosine similarity results, where
         $M_{AB}[i, j] = cos(\vec{a_i}, \vec{b_j})$
Preprocessing: $P1$ and $P2$ compute respectively
    $\hat{A} = \{\hat{a_1}, ..., \hat{a_s}\}$ and $\hat{B} = \{\hat{b_1}, ..., \hat{b_s}\}$ the sets of
    normalized vectors from $A$ and $B$.
Step 1 by $P1$
    Generates a random invertible matrix $M_R[s \times s]$
    Puts $A$'s elements as rows in a matrix $M_A[s \times n]$
    Performs $(M_R \times M_A)$ and sends the result matrix
    $(M_{RA})$ to $P2$
Step 2 by $P2$
    Puts $B$'s elements as columns in a matrix $M_B[n \times p]$
    Performs $(M_{RA} \times M_B)$ and sends back the result
    matrix $(M_{RAB})$ to $P1$
Step 3 by $P1$
    Performs $(M_R^{-1} \times M_{RAB}) = (M_A \times M_B) = M_{AB}$
    which is the searched cosine similarity matrix.
```

3.3. Π-Knn: privacy-preserving and efficient Knn classification protocol

In what follows, we introduce Π-Knn in which we make calls to Π-CSP+ presented above in order to securely perform the collaboration task (see task 1 of Algorithm 1) needed by the Knn process when dealing with an extern service provider. Assume D= $\{(x_1, y_1), ..., (x_s, y_s)\}$ a set of instance space held by a service provider denoted $P1$ and involving s activity patterns where x_i and y_i correspond to the pattern data and the pattern class name respectively. In order to adapt the similarity evaluation task within the knn process to Π-CSP+ presented above, we devide each time window w_i, which is considered as a time unit for one classification (see section 2.1), into p sub-windows. Thereby, in each classification we will consider p recorded activities each of which has a separate extracted feature vector.

Assume $Z =< (Xz_1, Yz_1), ..., (Xz_p, Yz_p) >$ a set of p recorded activities in a HAR system denoted $P2$, such as for $1 \leq j \leq p$: $Xz_j \in \mathbb{R}^n$ is the feature vector of the observation j and Yz_j denotes the correspondant activity class we are searching for. For the correctness purpose, we assume x_i and Xz_j have the same structure whenever $1 \leq i \leq s$ and $1 \leq j \leq p$ and we define I the identity function as defined above (see section 2.2). The deatiled implementation of Π-Knn is provided in algorithm 3.

Algorithm 3: Π-Knn, a Privacy-preserving and Efficient Knn classification protocol

```
Input  : $D = \{(x_1, y_1), ..., (x_s, y_s)\}$: $x_{i,(1 \le i \le s)} \in \mathbb{R}^n$
         $Z =< (Xz_1, Yz_1), ..., (Xz_p, Yz_p >$:
         $Xz_{j,(1 \le j \le p)} \in \mathbb{R}^n$
         $1 < s < n$
         $0 < k \le s$
         $0 < p < s$
Output: $< Yz_1, ..., Yz_p >$, the correspondant class
         label of each observation within $Z$
Step 1 by ($P1 \cup P2$)
1: Compute Π-CSP+($D$, $Z$), the cosine similarity matrix
   using Π-CSP+.
Step 2 by $P1$
2: for ($j = 1$; $j <= p$; $j + +$) do
3:     Select $Dz_j \subseteq D$, the set of $k$ patterns having the
       highest similarity rate in the column $j$ of the cosine
       similarity matrix got from task 1.
4:     $Yz_j = \underset{c}{\text{argmax}} \sum_{(x_i, y_i) \in Dz_j} I(c = y_i)$.
5: end for
6: return $< Yz_1, ..., Yz_p >$ to $P2$
```

4. Security analysis

In this section, we give a security analysis of our proposals according to the real/ideal simulation model [12], which provides strong security guarantees [13].

4.1. Security preliminaries

4.1.1. Multiparty computation (MPC). Given a set of participants that want to jointly compute the value of a public function f relying on their private data. Let P_1,...,P_n denote the participants and v_1,...,v_n their private data respectively. We call $f(v_1, ..., v_n)$ an MPC model [13].

4.1.2. Adversary model. The security analysis we give later depends on an allowed behaviours of corrupted parties. In an MPC model, we can distinguish, according to the allowed behaviours, two types of adversaries namely passive and active [13].

- **Passive adversary** (semi-honest). When a collaborating party is corrupted by such an adversary, it still follows the protocol specifications provided that it is allowed to analyse all information it gathered during the execution.
- **Active adversary** (malicious). A Party corrupted by such an adversary is alowed to randomly deviate from the protocol specifications, yet there are two common behaviours: a) aborting the protocol untimely or b) injecting fake inputs.

4.1.3. Assumptions & Notations.

- Let Π denote a multiparty protocol executed by P_1 and P_2 in order to evaluate the function f such as

$$\begin{array}{ll} f : [s \times n] \times [p \times n] & \to [s \times p] \\ \quad (M_A, M_B) & \mapsto M_A \times M_B \end{array}$$

- We call security parameters the set $\{s, n, p\}$ denoted $param$ and defined as $\begin{cases} 1 < s < n \\ 0 < p < s \end{cases}$
- Let $view_X^\Pi(param,\ M_A,\ M_B)_i$ denote the set of messages get by the party $P_{i,i \in \{1,2\}}$ during the execution X of Π on inputs M_A, M_B and security parameters $param$.
- Let $out_X^\Pi(param,\ M_A,\ M_B)_i$ denote the output of the party P_i from the execution X of the protocol Π on inputs M_A, M_B and security parameters $param$ and let $out_X^\Pi(param,\ M_A,\ M_B)$ denote the global output of all collaborating parties from the same execution of Π, where

$$\begin{aligned} out_R^\Pi(param[], M_A, M_B) = \\ out_R^\Pi(z, param[], M_A, M_B)_1 + \\ out_R^\Pi(z, param[], M_A, M_B)_2 \end{aligned}$$

4.2. Security definition

In this subsection, we give a definition of secure MPC according to real/ideal simulation paradigm.

4.2.1. Security model. In what follows, we introduce the real/ideal execution models.

- During a **real execution model** denoted R of the protocol Π on inputs M_A, M_B and security parameters $param$, we consider the presence of a real adversary denoted A, which behaves according to sme adversarial model (passive, active) while corrupting the party P_i. At the end of the execution R, the uncorrupted party denoted P_j outputs whatever specified in Π and the corrupted P_i outputs any random function of $view_R^\Pi(param,\ M_A,\ M_B)_i$.
- During an **ideal execution model** denoted L of the protocol Π on inputs M_A, M_B and security parameters $param$, we consider the presence of a trusted party denoted T that receives inputs of $P_{i,i \in \{1,2\}}$ in order to evaluate f in the presence of an ideal adversary denoted S. Assume S is corrupting the party P_i, handle its inputs and behave according to some adversial model (passive, active) before sending them to T. By the end, the uncorrupted party denoted P_j outputs what was received from T and the corrupted P_i outputs a random function of $view_L^\Pi(param,\ M_A,\ M_B)_i$.

4.2.2. Secure MPC protocol. Under the real/ideal pradigm, we consider that Π is secure if for any real adversary A that attacks Π and behaves according to some adversary model, there exists an ideal adversary S that can emulate it such that any effect on Π acheived by A could also be acheived by S while behaving according to the same adversarial model. Let $\overset{d}{\equiv}$ denote the distribution equality. We formilze this security definition as

$$\{out_R^\Pi(param, M_A, M_B)\} \overset{d}{\equiv} \{out_L^\Pi(param, M_A, M_B)\} \quad (5)$$

4.3. Security proof by simulation

Relying on definitions given above, in this subsection we provide a security proof of Π-CSP+ (see algorithm 2) and Π-Knn protocol (see algorithm 3).

4.3.1. Π-CSP+ security proof.

Theorem 1 (Π-CSP+ security)**.** *The* Π-*CSP+ detailed in Algorithm 3 is a secure MPC protocol in the presence of a malicious adversary.*

Proof. In order to prove the theorem 1, we give a separate simulation of the case where a malicious adversary corrupts $P1$ and the case where it corrupts $P2$. We assume that if both parties are corrupted we are not required to provide security measurments. Let A, S and T denote respectively a real active adversary, an ideal active adversary and a trusted third party. Let Π denote the Π-CSP+.

- **Case 1:** $P2$ **is corrupted by** A**.** Then, the allowed behaviour of $P2$ is only injecting fake inputs (M_B) (Because aborting the protocol untimely will stop the execution of Π-CSP+ and so, has no meaning). Assume $P2$ sends a fake M_B. In this case, S can emulate A by just handling the fake M_B and sends it to T, which performs compuation and sends back M_{AB} to $P1$. Thereby, completing the simulation. At the end, the views of $P2$ through ideal and real executions are decribed as

$$view_L^\Pi(param, M_A, M_B)_2 = \{M_B\} \quad (6)$$

$$view_R^\Pi(param, M_A, M_B)_2 = \{M_B, M_{RA}\} \quad (7)$$

  But, since M_{RA} will contain $((s \times s) + (s \times n))$ unknowns opposite to $(s \times n)$ equations, thus according to security parameters defined in $param$, M_{RA} will not involve any information for $P2$ and can be considered as a random noise. Hence, the view of $P2$ in the real execution could be reduced as

$$view_R^\Pi(param, M_A, M_B)_2 = \{M_B\} \quad (8)$$

  Thus, relying on (6) and (8) we get

$$\begin{aligned}\{out_R^\Pi(param, M_A, M_B)_2\} &\overset{d}{\equiv} \\ \{out_L^\Pi(param, M_A, M_B)_2\}&\end{aligned} \quad (9)$$

  On the other hand, $P1$ will output M_{AB} in real execution, which is the same output received from T in ideal process. Recall that $P1$ is uncorrupted, thus it outputs what was specified in the protocol. This means that

$$\begin{aligned}\{out_R^\Pi(param, M_A, M_B)_1\} &\overset{d}{\equiv} \\ \{out_L^\Pi(param, M_A, M_B)_1\}&\end{aligned} \quad (10)$$

  Through (9) and (10), we proved by simulation that all effects achieved by a real malicious adversary corrupting $P2$ can also be acheived in an ideal process. In this case, Π-CSP+ is a secure MPC protocol.

- **Case 2:** $P1$ **is corrupted by** A**.** Then, it can inject fake inputs (M_A) or abort the protocol in step 2. But, since $P2$ does not require any output, the abort of $P1$ will have no effect. Assume $P1$ sends a fake M_A. In this case, S will emulate A by handling the fake M_A and just sends it to T in order to complete the simulation. By the end, the views of $P1$ through ideal and real executions are decribed as

$$view_L^\Pi(param, M_A, M_B)_1 = \{M_A, M_{AB}\} \quad (11)$$

$$view_R^\Pi(param, M_A, M_B)_1 = \{M_A, M_{AB}, M_{RA}, M_{RAB}\} \quad (12)$$

  Like in the precedent case, we can reduce (12) since M_{RAB} will involve $(s \times p)$ equations and $(n \times p)$ unknowns, so, according to security parameters defined in $param$, it can not reveal any information for $P1$. Likewise, since M_R is a random noise, M_{RA} could also be get from (11), thus, we reduce it from (12). Hence, the view of $P1$ in the real execution could be shortened as

$$view_R^\Pi(param, M_A, M_B)_1 = \{M_A, M_{AB}\} \quad (13)$$

  Thus, from (11) and (13) we get

$$\begin{aligned}\{out_R^\Pi(param, M_A, M_B)_1\} &\overset{d}{\equiv} \\ \{out_L^\Pi(param, M_A, M_B)_1\}&\end{aligned} \quad (14)$$

  Regarding the uncorrupted $P2$, as it does not require any output, it will not receive any information in ideal execution, which is the case for real execution since the only message get during Π is M_{RA} that does not involve any information according to security parameters $param$. Consequently, we can deduce from (14) that any effect acheived by a real malicious adversary corrupting $P1$ can also be acheived in an ideal process. This means that in this case, Π-CSP+ is a secure MPC protocol.

□

Note 1 (Secure re-execution)**.** *We consider both* $P1$ *and* $P2$ *having a probability* $p > 0$ *to change their inputs* (M_A *and* M_B) *in each execution* (X) *of* Π-*CSP+. Under such assumption, we ensure the secure re-execution of* Π-*CSP+ for* $t_{(t>0)}$ *times by the same parties.*

4.3.2. Π-Knn security proof.

Corollary 1 (Π-Knn security)**.** *The Π-Knn protocol detailed in Algorithm 3 is a secure MPC protocol in the presence of a malicious adversary.*

Proof. As the call to Π-CSP+ is the only multiparty task within Π-Knn (see algorith 3), we can deduce the security of Π-Knn relying on theorem 1 proved above. □

5. Performance analysis

In this section we evaluate the computation performance of Π-CSP+, which in turns, reflects the performance of Π-Knn protocol. To do this, we assess the ability of Π-CSP+ to handle feature vectors of high size (s) extracted from activities recorded in time windows having a short length(l). This evaluation aims to prove the adequacy of Π-CSP+ for activity recognition systems that require a high accuracy level besides a quick decision make.

Regarding the evaluation environment, we make experiments on the same set of vectors using a simulator built in Python and an Intel i5-2557M CPU running at 1.70 GHz and having a 4 GB of RAM.

5.1. Experimental scenarios

We consider a HAR system where a client's activity are recorded during a period of time (w_i) that has a length of (l) time unit. During this period, that we call observation time window, the HAR system exracts from the sensed data raws, every time unit, a new feature vector having the size (s). At the end of (w_i), the HAR system sends the l extracted feature vectors in order to be classified by a service provider considered having ($2 \times l$) patterns. We consider testing separately the effect of the (s) size of feature vectors and the length (l) of the observation time window on the running time of Π-CSP+ throughout 2 experiments respectively E_1 and E_2.

We perform E_1 four times such that in each one we fixe the time window length l to one distinct value from the set $[2, 20]$ and we vary s in the set $[50, 100]$ for each fixed l.

In E_2, we do the opposite by fixing s four times to one value from the set $[50, 100]$ and we vary l for each fixed s in the set $[2, 20]$. Notice that we choose these sets of values so as to respect the security parameters ($param$) used in implementations Π-Knn and Π-CSP+ (see section 3.2).

For the comparison purpose, we make the same experiments on the most recent cosine similarity computation protocol named PCSC [14], which is free from cryptography and asserted to be the most efficient.

We take three samples from each experiment and we plot results of E_1 and E_2 respectively in Figure 1 and Figure 2, besides the running time of the direct cosine similarity computation denoted DCS (direct application of the cosine similarity metric without any secure measurment. see equation (1), section 2.3) that we consider as a running time reference.

5.2. Results & discussion

Through E_1 wa have evaluated the effect of the feature vecors size (s) on the running time of the three similarity computation methods (Π-CSP+, PCSC and DCS). Results illustrated in Figure 1 reveal the high efficiency level of Π-CSP+ running time which remains stable in the neighborhood of $0.0x$ ms for $l \in \{2, 10\}$ and reaches $0.1x$ for $l = 15$ with a slow increasing rate of 4.66% between $l = 2$ and $l = 15$. On the other hand, PCSC running time revealed an increasing overhead with rate of 6% between $l = 2$ and $l = 15$, besides a high distance from the running time of DCS computation reference (> 170 ms), which is highly greater than Π-CSP+ time distance from DCS ($< 0.x$ ms).

In E_2 we made focus on the effect of time window length (l) on running time of the previous three computation methods. Results shown in Figure 2 reveal more clearely the overhead induced in running time of PCSC throughout the three sample sizes $s = \{70, 90, 100\}$. PCSC distance time from DCS reference was increasing continuously (on average of 1200 ms) with an increasing rate that has reached $\approx 10\%$ versus a rate of 7% reached by Π-CSP+ while keeping a short stable distance on average of $0.1x$ ms from DCS running time.

Results of E_1 and E_2 have shown the efficiency of Π-CSP+ computation time regarding the increase of time window length (l) or when dealing with feature vectors having a high size (s). These reusults affirm the adequacy of Π-CSP+ to efficiently secure a classification process of any HAR system that needs a high accuracy level (a high number of features within a vector (s) and a high number of vectors within a short time (l)).

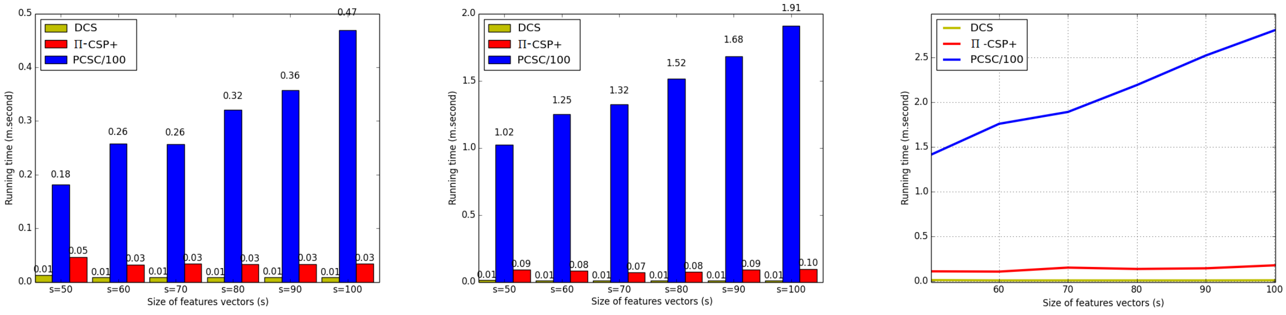
6. Related works

In this state-of-the art section, we review recent works in HAR field and we highlight their privacy lack. Next, we provide a short review of exiting privacy-preserving techniques that might be used to securly assess similarity in a HAR classification process.

6.1. HAR systems

Almost all existing HAR systems make focus on accuracy and reliability of activities' detection without considering data privacy concern.

B. Najafi et al. [15] proposed a physical activity monitoring system based on Kinematic sensors. The system is able to recognize sitting, standing and lying body postures as well as periods of walking with the aim of monitoring elderly people in their daily lives. Authors have focused on accuracy detection of activities but they gave no security and privacy preserving measurements. JC Hou et al. [16] proposed PAS: an open architecture that exploits off-the-shelf technologies to assist elderly people through monitoring their physiological functions, mobility profiles, besides fall detection service and some other assisted-living tasks. Regarding security concerns, PAS incorporated mechanisms to

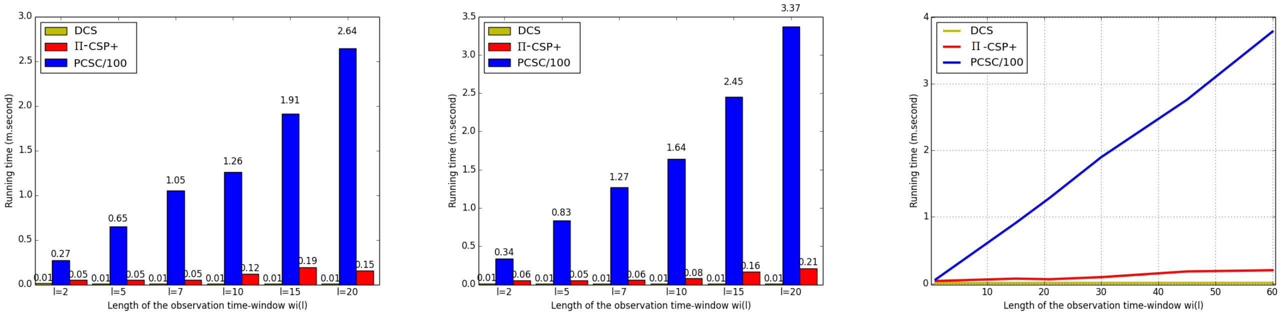


(a) l=2

(b) l=10

(c) l=15

Figure 1. E_1. Effect of the features vectors size (s) running time while fixing the time window length (l).



(a) s=70

(b) $s = 90$

(c) $s = 100$

Figure 2. E_2. Effect of the time window length (l) on running time while fixing the features vectors size (s).

secure both data storage and communication; however, there is no privacy protection of sensed data during analysis and recognition process. S Jiang et al. [17] proposed CareNet: a system prototype for remote physical activity monitoring in healthcare application. CareNet is able to detect falls and launch associated alarms, in addition to provide on-demand video information in order to verify the physical activity results. Privacy protection within CareNet is ensured only through secure communication while there is no privacy protection measurements regarding data analysis. AS Evani et al. [18] proposed a patient activity monitoring system using wearable flex sensors in order to follow patient's routine day-to-day activities. Their system recognize sitting, standing and walking activities as well as inactivity that is considered as abnormal behaviour. With regard to data privacy, no protection measures were embedded.

Recently, Debraj De et al. [19] introduced a fine-grained activity recognition system using multimodal wearable sensors. Authors highlighted the need for detecting complex activities in critical healthcare application. The proposed system was able to recognize 19-in home activities without using sensing modes that induce direct privacy concerns, such as video recording. Although this use of only wearable devices, some sensed data such as GPS localisation could disclose sensitive information which requires strong mechanisms for privacy protection during analysis and recognition process.

6.2. Privacy-preserving similarity evaluation

As the privacy concern of the cosine similarity metric used in HAR classification lives in computing the scalar product (see section 3.1), in this subsection we make interest on existing privacy-preserving scalar product techniques. Throughout a literature review, almost all such techniques are trading security and computational efficiency. By summarizing, there has been two main approaches: a) cryptographic-based techniques [20], [21], [22], [23], [24], [25] that uses cryptographic schemes such as the Paillier [7] and ElGamal [8] cryptosystems in order to provide hard security guarantees while raising in significant degradation in computational-time efficiency and b) noise-based techniques [14], [22], [23], [26] that aim to guard a high efficiency level by using simple arithmetic transofrmations. Nevertheless, existing approches that fall in this category do not provide a security protection for all data types (ex. binary attributes vs numerical attributes), making so, an other trade-off.

Contrary to precedent work, the main contribution of this paper is to provide a high security guarantee as cryptographic techniques level (see section 4) besides a high efficient computational-time service comparing to arithmetic techniques [14] (see section 5).

7. Conclusion

In this paper we have tackled the privacy and efficiency concern in classification step of a human activity recognition (HAR) process by designing two novel protocols. we proposed Π-Knn, a novel knn classification protocol that securely performs the similarity evaluation task between recorded activities and extern patterns based on a novel efficient and privacy-preserving cosine similarity protocol named Π-CSP+. Through a security analysis using the simulation paradigm, we have shown the security guarantees provided by our proposals in the presence of an active adversary. Regarding the performance evaluation, diferent experimental tests have revealed the time-effciency of computations performed by our protocol when compared to other recent proposed method, which reveals its adequacy for situations where a quick decision is critical.

References

[1] Nicholas Farber, Douglas Shinkle, Jana Lynott, Wendy Fox-Grage, and Rodney Harrell. *Aging in place: A state survey of livability policies and practices*. 2011.

[2] Centers for Disease Control, Prevention (CDC), et al. Injury prevention and control, web based injury statistics query and reporting system (wisqars). 2012.

[3] N Cavill and L Ells. Treating adult obesity through lifestyle change interventions. *A Briefing Paper for Commissioners. National Obesity Observatory: Oxford, UK*, 2010.

[4] Daqing Zhang, Bin Guo, and Zhiwen Yu. The emergence of social and community intelligence. *Computer*, (7):21–28, 2011.

[5] Xun Yi, Jan Willemson, and Farid Nait-Abdesselam. Privacy-preserving wireless medical sensor network. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on*, pages 118–125. IEEE, 2013.

[6] Na Li, Nan Zhang, Sajal K Das, and Bhavani Thuraisingham. Privacy preservation in wireless sensor networks: A state-of-the-art survey. *Ad Hoc Networks*, 7(8):1501–1514, 2009.

[7] Pascal Paillier. Public-key cryptosystems based on composite degree residuosity classes. In *Advances in cryptology-EUROCRYPT'99*, pages 223–238. Springer, 1999.

[8] Taher ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms. In *Advances in cryptology*, pages 10–18. Springer, 1984.

[9] Oscar D Lara and Miguel A Labrador. A survey on human activity recognition using wearable sensors. *Communications Surveys & Tutorials, IEEE*, 15(3):1192–1209, 2013.

[10] David W Aha, Dennis Kibler, and Marc K Albert. Instance-based learning algorithms. *Machine learning*, 6(1):37–66, 1991.

[11] Ali Mustafa Qamar, Eric Gaussier, Jean-Pierre Chevallet, and Joo Hwee Lim. Similarity learning for nearest neighbor classification. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, pages 983–988. IEEE, 2008.

[12] Ran Canetti. Security and composition of multiparty cryptographic protocols. *Journal of CRYPTOLOGY*, 13(1):143–202, 2000.

[13] Yehuda Lindell and Benny Pinkas. Secure multiparty computation for privacy-preserving data mining. *Journal of Privacy and Confidentiality*, 1(1):5, 2009.

[14] Rongxing Lu, Hui Zhu, Ximeng Liu, Joseph K Liu, and Jun Shao. Toward efficient and privacy-preserving computing in big data era. *Network, IEEE*, 28(4):46–50, 2014.

[15] Bijan Najafi, Kamiar Aminian, Anisoara Paraschiv-Ionescu, François Loew, Christophe J Büla, and Philippe Robert. Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. *Biomedical Engineering, IEEE Transactions on*, 50(6):711–723, 2003.

[16] Jennifer C Hou, Qixin Wang, Bedoor K AlShebli, Linda Ball, Stanley Birge, Marco Caccamo, Chin-Fei Cheah, Eric Gilbert, Carl A Gunter, Elsa Gunter, et al. Pas: A wireless-enabled, sensor-integrated personal assistance system for independent and assisted living. In *High Confidence Medical Devices, Software, and Systems and Medical Device Plug-and-Play Interoperability, 2007. HCMDSS-MDPnP. Joint Workshop on*, pages 64–75. IEEE, 2007.

[17] Shanshan Jiang, Yanchuan Cao, Sameer Iyengar, Philip Kuryloski, Roozbeh Jafari, Yuan Xue, Ruzena Bajcsy, and Stephen Wicker. Carenet: an integrated wireless sensor networking environment for remote healthcare. In *Proceedings of the ICST 3rd international conference on Body area networks*, page 9. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008.

[18] AS Evani, B Sreenivasan, JS Sudesh, M Prakash, and J Bapat. Activity recognition using wearable sensors for healthcare. In *the 7th International Conference on Sensor Technologies and Applications (SENSORCOMM 2013)*, pages 173–177, 2013.

[19] Debraj De, Pratool Bharti, Sajal K Das, and Sriram Chellappan. Multimodal wearable sensing for fine-grained activity recognition in healthcare. *Internet Computing, IEEE*, 19(5):26–35, 2015.

[20] Bart Goethals, Sven Laur, Helger Lipmaa, and Taneli Mielikäinen. On private scalar product computation for privacy-preserving data mining. In *Information Security and Cryptology–ICISC 2004*, pages 104–120. Springer, 2004.

[21] Wenliang Du and Mikhail J Atallah. Privacy-preserving cooperative statistical analysis. In *Computer Security Applications Conference, 2001. ACSAC 2001. Proceedings 17th Annual*, pages 102–110. IEEE, 2001.

[22] Wei Jiang, Mummoorthy Murugesan, Chris Clifton, and Luo Si. Similar document detection with limited information disclosure. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 735–743. IEEE, 2008.

[23] Mummoorthy Murugesan, Wei Jiang, Chris Clifton, Luo Si, and Jaideep Vaidya. Efficient privacy-preserving similar document detection. *The VLDB Journal-The International Journal on Very Large Data Bases*, 19(4):457–475, 2010.

[24] Hiroaki Kikuchi, Kei Nagai, Wakaha Ogata, and Masakatsu Nishigaki. Privacy-preserving similarity evaluation and application to remote biometrics authentication. *Soft Computing*, 14(5):529–536, 2010.

[25] Dexin Yang, Baolin Xu, Bo Yang, and Jianping Wang. Secure cosine similarity computation with malicious adversaries. In *Computer Networks & Communications (NetCom)*, pages 529–536. Springer, 2013.

[26] Jaideep Vaidya and Chris Clifton. Privacy preserving association rule mining in vertically partitioned data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 639–644. ACM, 2002.