

Analyzing and Safeguarding Human-Kinesics Information in Wirelessly Instrumented Space

Kirti Chawla

Department of Computer Science
University of Virginia
Charlottesville, 22903, USA
kirti@cs.virginia.edu

Abstract—This paper presents a systematic study of privacy leakage in the wirelessly instrumented space. Rapid advances in RFID and sensing technologies enable the construction of wirelessly instrumented space for variety of applications including assisted living, smart manufacturing facilities, intelligent appliances, and more. The wireless and pervasive nature of these technologies provides an opportunity to glean context-specific and sensitive information leading to privacy issues. To address these concerns, we construct a kinesics recognition system to analyze the information present in such a space. Our experiments based on multiple subjects show that user kinesics are a reliable indicator of user’s identity. Consequently, we propose a few mitigating measures to reduce the privacy leakage in the proposed system.

Keywords – Human-activity, Kinesics, RFID, WISP, Privacy Leakage, Classifiers

I. INTRODUCTION

Recent advances in RFID and sensing technologies, have enabled the construction of wirelessly instrumented space for numerous applications that encompasses our home, work place, and more [2, 8]. Small form factor, minimal power requirement, and wireless operability of RFID tags, readers, and sensors, facilitate their placement in a variety of orientations and locations, thus making them pervasive.

The user interacts with the wirelessly instrumented space either actively (through devices) or passively (through activity recognition), to meet her requirements. Numerous case studies have shown the beneficial applications of the system [3]. While the gains are immediately visible, we argue that the privacy of user becomes an “exploitable” commodity. To address the problem of benefitting the user, while keeping her privacy intact, we propose a systematic study to learn information present, as available to an adversary, in such an instrumented space. Consequently, we suggest two possible mitigating measures to alleviate the ensuing privacy concerns.

This paper is organized as follows. Section II briefly discusses the related work. In section III, we present the design of a kinesics recognition system – our representative wirelessly instrumented space. Section IV focuses on the problem of privacy leakage via kinesics. We present experimental methodology, and detailed evidence in section V, suggest two mitigating techniques to address the privacy concerns in section VI, provide future work direction in section VII, and conclude in section VIII.

II. RELATED WORK

In order to provide complete context for the discussion, pertinent research studies are highlighted that relate to our work. Wilson et al describe an unsupervised technique based on data gathered from ubiquitous sensors to label anonymous activity episode [5]. By using this technique, certain unmapped activities of the monitored user can be mapped, thereby reducing the occurrence of anonymous activity in order to better serve the user. Logan et al design an instrumented house based on a sensing platform consisting of 900 multi-modal sensors to compare the usefulness of different sensors in different scenarios [6]. A study done by Wu et al, proposes the use of video-feed, RFID, and a dynamic Bayesian network to detect the use of different objects and recognize the activity performed by the monitored user. They claim high (~80%) activity-recognition rates sampled on real-world data [1]. In a case study involving a hospital based instrumented space, hidden markov model based activity recognition is used to determine the requirement of a patient and alert the medical staff appropriately [4]. Additionally, Ravi et al propose activity recognition based on accelerometer data by using base-level and multi-level classifiers [7].

These efforts relate the positive outcome of the current research with our work. However, Srinivasan et al describe a sophisticated FATS (Fingerprinting And Timing based Snooping) attack, which enables an adversary to listen to wireless communication between the sensors and correlate it to various human activities [9]. Furthermore, Saponas et al present a study, wherein a sensor-enabled Nike sports shoe coupled with a receiver-enabled iPod Nano can be used to provide location information, thereby compromising privacy of the user [10]. Moreover, Juels et al enumerate privacy threats due to RFID, based on action, association, location, preference, and more, of the user [11]. This shows that the privacy of user can be compromised in a variety of ways. Therefore, the use of such “double-edged”¹ technology to construct a wirelessly instrumented space is prone to serious privacy concerns.

It should be noted here that the system designed to collect user data for legitimate purposes can also be compromised to reveal interesting patterns of the user activities. Thus, the aforementioned research studies and the systems, methods, and applications resulting out of them are subject to this compromise.

¹ RFID and sensing technology can be put to good as well as bad use.

III. DESIGN OF A KINESICS RECOGNITION SYSTEM

To motivate a systematic study of privacy issues in a wirelessly instrumented space, we adopt an approach of constructing a representative space, and learning the nature of information that is present. We present a detailed design description of a “*kinesics*”² recognition system, wherein user manipulates the physical and virtual environment based on simple kinesics.

A. WISP Glove

WISP (Wireless Identification and Sensing Platform) is battery-less programmable RFID sensor device based on TI MSP430F2132 microcontroller architecture that operates in UHF (Ultra High Frequency: 902-928 MHz) band, and is compliant with EPC (Electronic Product Code) Gen2 protocol [12, 18]. It permits operation of numerous onboard sensors by harvesting energy from radio-field emitted by a RFID reader.

We construct a WISP glove and use the onboard accelerometer sensor to gather acceleration data in 3-axes. This serves as an embodiment of kinesics measuring system. Fig. 1 depicts the design of the WISP glove.



Figure 1. WISP glove prototype

We use this prototype to interface with a kinesics recognition algorithm, which gathers the acceleration values emitted during the kinesics and analyzes them.

B. Kinesics Recognition Algorithm

The user wearing a WISP glove transmits 3-axes acceleration values when performing a kinesics. These values are received by the proximate RFID reader and fed to a kinesics recognition algorithm that distinguishes different types of kinesics, and maps the kinesics to programmable events. In order to distinguish kinesics from each other, the system monitors acceleration values, and uses a threshold-based algorithm to assign weights to different kinesics, with the kinesics having highest weight marked as the resultant kinesics. Fig. 2 illustrates the threshold-based algorithm that is used to determine the resultant kinesics.

```

Input: 3-axes Acceleration Values ( $X, Y, Z$ ), Acceleration Thresholds ( $T_1, T_2, \dots, T_i$ ), Sampling Threshold ( $S_T$ )
Output: Resultant Kinesics  $K_R$  from the set ( $K_1, K_2, \dots, K_j$ )

while ( $ID < S_T$ )
{
    CurrX[ID] =  $X$ , CurrY[ID] =  $Y$ , CurrZ[ID] =  $Z$ 

    if ((CurrX[ID] <  $T_1$ ) and (CurrY[ID] <  $T_2$ )) then
    {
        INCREASE-WEIGHT( $K_1$ )
    }
    else if ((CurrX[ID]  $\geq T_1$ ) and (...) and (CurrY[ID] <  $T_2$ )) then
    {
        INCREASE-WEIGHT( $K_2$ )
    }
    ...
    else
    {
        INCREASE-WEIGHT( $K_j$ )
    }
}

Update ID
}

 $K_R = \text{Max}(K_1, K_2, \dots, K_j)$ 

Return  $K_R$ 

```

Figure 2. Pseudo-code of threshold-based kinesics recognition algorithm

Using the kinesics recognition algorithm, it is relatively straightforward to isolate kinesics based on their acceleration values. Each such isolated kinesics then corresponds to a rule in the algorithm, which serves to increase the weight of the kinesics, when encountering a real time feed of acceleration values.

C. Representative Kinesics

We have constructed a set of three representative hand kinesics to harness the kinesics recognition system. To get consistent kinesics data, we require the user wearing WISP glove to perform their kinesics by keeping their hand parallel to reference plane, and perpendicular to the antenna. This requirement is illustrated in Fig. 3.

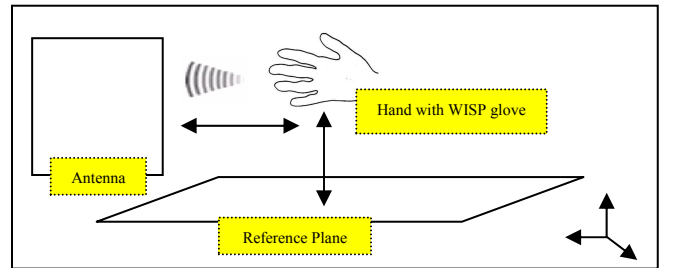


Figure 3. Position requirement to gather user kinesics data

The nomenclature and description of the three hand kinesics is as follows:

- **PPLT kinesics:** In this kinesics, user tilts the hand towards left, while keeping it parallel to the reference plane. It is named as “*perpendicular parallel left tilt*”.

² Non-verbal behavior related to movement, either of any part or the body as a whole.

- **PPRT kinesics:** To perform this kinesics, the user tilts the hand towards right, while keeping it parallel to the reference plane. We name it “*perpendicular parallel right tilt*”.
- **PPDT kinesics:** This kinesics requires, the user to tilt the hand towards down, while keeping it parallel to the reference plane. It is called “*perpendicular parallel down tilt*”.

For each kinesics, we sample acceleration data based on 10 unique points, over a period of three seconds. In section V, we show that even with only a few (~10) points, kinesics data retains distinguishing features.

D. Illustrative Applications

We associate the kinesics recognition algorithm with a set of programmable events, which can be triggered on detection of pertinent kinesics. For this purpose, we rely on the kinesics recognition algorithm to determine the exact kinesics required. We then use the judgment of the algorithm to invoke the programmable events. Fig. 4 depicts this concept in more detail.

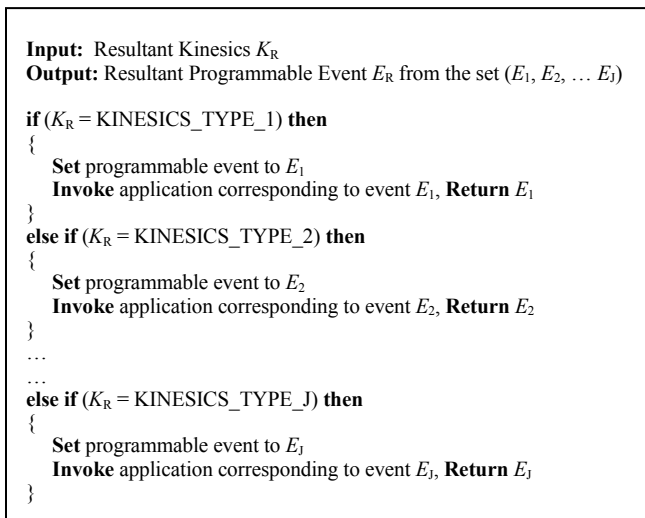


Figure 4. Kinesics mapped to programmable events

In order to discuss meaningfully about the privacy issues, we have constructed three simple yet diverse applications. This allows us to show the possible implications of misuse of privacy of the user, by suggesting that the adversary may gain access and exercise control in a manner that is similar to the user. The description of these applications is as follows:

- **Change IM client status message:** In this application, the user performs PPLT kinesics to modify the existing IM client status message to a custom message. We have targeted Yahoo Messenger IM client for the construction of this application. When the kinesics recognition algorithm determines that the resultant kinesics is PPLT kinesics, it then uses the system registry to fetch the key `Software\Yahoo\Pager\Profiles\<User>\Custom Msgs`, to get

the handle for custom messages for the particular user. Using this handle, we then update it to a custom message and post the update to the running IM client via posting a windows message (based on the WIN32 api model) [20].

- **Lock workstation:** While using this application, the user performs PPRT kinesics to lock the workstation. We have determined that the dynamic link library `user32.dll` provides a method `LockWorkstation()`. When the kinesics recognition algorithm determines that the kinesics is PPRT kinesics, it calls this method to initiate locking of workstation.
- **Control lights/appliances:** This application requires, the user to perform the PPDT kinesics to control various lights and/or appliances. We have integrated the X10 home automation modules with the kinesics recognition algorithm [21]. Particularly, we have relied on the X10 CM17A module (serial port based X10 signal transmitter) and the TM751 module (power line based X10 signal receiver). X10 modules can be installed on the power line and communicate with each other using a custom protocol. This protocol enables the user to map large number (~255) of devices by using a collection of X10 modules. In a typical setup, when the kinesics recognition algorithm determines that the kinesics is PPDT kinesics, it sends an appropriate command from the X10 CM17A module to the X10 TM751 module. In our setup the lights and/or appliances are attached to the X10 TM751 module, while the kinesics recognition algorithm controls the X10 CM17A module. By repeating the PPDT kinesics, the user can turn on or off various lights and/or appliances.

Fig. 5 illustrates the component of our experimental kinesics recognition system. It is relatively straightforward to construct newer set of applications that can be invoked based on the programmable events pertaining to a given kinesics. The intention in this effort has been to provide an ideal background of representative applications to discuss the implications of the lack of privacy, and at the same time, we deliberately resist not to offer complete and exhaustive set of applications.

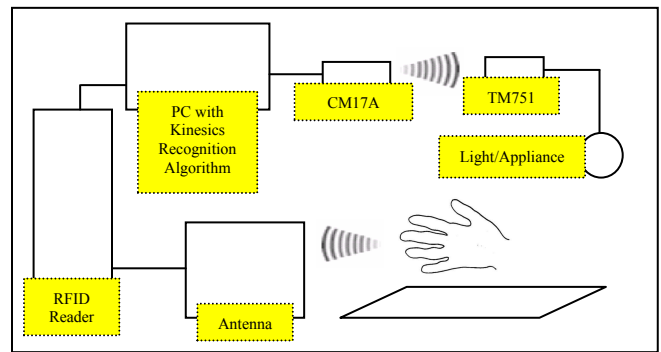


Figure 5. Components of the kinesics recognition system

IV. PRIVACY LEAKAGE VIA KINESICS

To study the privacy leakage through kinesics, we use the kinesics recognition system as the test bed. As previously mentioned, we have constructed three representative kinesics that are used to enable three different applications. Here, we note here that a WISP device can be queried up to a distance of 4.5 meters, with the possibility that the RFID reader and antenna are under adversarial control. Therefore, it is conceivable that the legitimate system and the adversary have similar view of the user kinesics. Using these assumptions, we put forth a hypothesis about the privacy leakage through user kinesics.

A. Privacy Leakage Hypothesis

It is natural that the user is required to perform a variety of kinesics to avail different services from a kinesics recognition system. Furthermore, it is possible that the same system is used by multiple users. The system gathers the acceleration data from the user and applies certain thresholds to distinguish between different kinesics. As long as the data is amenable to thresholds, the system, by default, does not distinguish between different users. However, it still gathers the data from all the users. Thus, the system may potentially leak privacy through user kinesics in two different scenarios:

- **SUMK scenario:** As we know that a user needs to perform multiple kinesics to get different services from the kinesics recognition system. The difference in 3-axes acceleration values for different kinesics, then serve to reveal about difference in kinesics. This scenario is called as the “*Single User Multiple Kinesics*” scenario. Formally, we suggest the following:

$$\forall I, U \exists J, K [X_1, Y_1, Z_1]_{\{J, U\}} \neq [X_1, Y_1, Z_1]_{\{K, U\}} \quad (1)$$

Where,
 $[X_1, Y_1, Z_1]$ = Kinesics data vector
 I = Sample point
 U = User
 J, K = Different kinesics

- **MUSK scenario:** The system can be used by multiple users. We note that the 3-axes acceleration data from different users are different due to their different physical attributes such as the default acceleration of different hands, preference towards certain tilt, and more. We claim that, if two users perform same kinesics differently, then the difference in kinesics data may potentially reveal about the difference in identity of the user. We have named this scenario as the “*Multiple User Single Kinesics*” scenario. Formally, we say:

$$\forall I, J \exists U, V [X_1, Y_1, Z_1]_{\{J, U\}} \neq [X_1, Y_1, Z_1]_{\{J, V\}} \quad (2)$$

Where,
 $[X_1, Y_1, Z_1]$ = Kinesics data vector
 I = Sample point
 U, V = Different users
 J = Kinesics

We note the possibility of single user having different kinesics data for the single kinesics and multiple users having different kinesics data for multiple kinesics. To get a meaningful response from the kinesics recognition system, the user is constrained to provide the kinesics data. Thus, any difference in kinesics data for the single user performing single kinesics will conform to the Gaussian distribution, wherein multiple samples of single kinesics from the single user enable the determination of standard deviation.

In the multiple user multiple kinesics scenario, it is evident that the kinesics data will be different, which follows from the principle of MUSK scenario. Thus, by reasoning about the two privacy leakage scenarios, allows us to draw conclusions for the remaining scenarios as well.

B. Classification of Kinesics

The focus on the privacy leakage hypothesis rests on the difference in the user kinesics. To understand this, we formulate the difference in user kinesics as a classification problem based on threshold values and type of kinesics. We use five base-level classifiers to distinguish the three kinesics and the users.

- **Naïve Bayes Classifier [13]:** This classifier is based on the Baye’s theorem, wherein the probability that the given kinesics data represents a particular kinesics, from a set of possible kinesics, is dependent on the probability that the data belongs to a particular kinesics even before viewing the kinesics data, the probability that the given kinesics data is seen, assuming it belongs to a particular kinesics, and the probability that the given kinesics data will appear. This is represented formally as under:

$$\forall I, P(K | [X, Y, Z]) = \frac{P(K) \times \prod_{i=1}^N P([x_i, y_i, z_i] | K)}{P([X, Y, Z])} \quad (3)$$

Where,
 $[X, Y, Z]$ = Kinesics data vector with values $([x_1, y_1, z_1], [x_2, y_2, z_2], \dots, [x_N, y_N, z_N])$
 K = class of kinesics
 N = number of sample points
 $P(K|[X, Y, Z])$ = Probability that the given kinesics data belongs to a particular kinesics
 $P(K)$ = Probability that the kinesics belongs to a particular kinesics even before viewing the kinesics data
 $P([x_i, y_i, z_i]|K)$ = Probability that the given kinesics is seen, assuming it belongs to a particular kinesics
 $P([X, Y, Z])$ = Probability that the given kinesics data will appear

Note that the aforementioned probabilities are estimates and not exact probabilities, which must be derived after viewing the kinesics data. Also, classification based on this classifier is favorable to the kinesics data only when the correct kinesics class K has the highest value of the numerator in equation (3) and is not based on goodness of estimates.

- **CN2 Classifier [14]:** Rules are generated to classify data through an induction system in this classifier. The

task of the classifier is to output “*if-then-predict*” rules that aid in reducing the search space and classify the data. Formally, it is represented as follows:

$$\forall I, \text{if } [X_1, Y_1, Z_1] \text{ then predict } K \quad (4)$$

Where,

$[X_1, Y_1, Z_1]$ = Kinesics data vector
 K = Kinesics class

The classifier is an iterative algorithm, which progresses until no further favorable kinesics data vectors can be found matching to a particular kinesics class. Also, it prunes the useless or less favorable rules from the rule-set and maintains the size of rule-set based on a user-defined metric.

- **kNN Classifier [15]:** The kinesics data vector is classified to the same kinesics class of its k nearest neighbors. The parameter k tradeoffs the probability of the non-Bayes decision for the classless kinesics data vector with the posterior probabilities of the actual class of the kinesics data vector. Let $([X_1, Y_1, Z_1], K_1), ([X_2, Y_2, Z_2], K_2), \dots, ([X_J, Y_J, Z_J], K_J)$ be J pairs of kinesics data vector and kinesics class. A new pair $([X, Y, Z], K)$ arrives with the constraint that only the kinesics data vector can be measured, with the requirement that it must be classified to one of the already known kinesics classes. Formally we say:

$$\forall I, [X', Y', Z'] \in S_{NN}; |S_{NN}| = k \quad (5)$$

$$\text{if } \text{MIN}\{\delta([X_1, Y_1, Z_1], [X, Y, Z])\} = \delta([X', Y', Z'], [X, Y, Z])$$

Where,

$[X', Y', Z']$ = Candidate nearest neighbor kinesics data vector
 $[X_1, Y_1, Z_1]$ = Kinesics data vector
 S_{NN} = Set of nearest neighbor
 k = Classifier parameter for determining number of nearest neighbor
 δ = Distance function
 MIN = Minimum function

It should be noted here that the distance function varies depending upon nature of the data. In our case, we use Euclidean distance as the distance function.

- **SVM Classifier [16]:** Using this classifier, the kinesics data vector and kinesics class is treated as a pair of values. Each kinesics data vector is 3-dimensional real vector. The task of this classifier is to find a maximum margin hyperplane, which divides the pair of values into different kinesics classes. Formally, we suggest the following:

$$\forall I, \{([X_1, Y_1, Z_1], K_1) | [X_1, Y_1, Z_1] \in \mathfrak{R}^3, K_1 \in K\} \quad (6)$$

Where,

$[X_1, Y_1, Z_1]$ = Kinesics data vector
 K = Set of kinesics classes
 K_1 = Instance of kinesics class
 \mathfrak{R} = Set of real numbers

Thus, a hyperplane can be defined as a set of kinesics data vectors $[X_1, Y_1, Z_1]$ that satisfies the following relation:

$$\forall I, w \cdot \{[X_1, Y_1, Z_1]\} - h = 0 \quad (7)$$

Where,

w = normal vector perpendicular to the hyperplane
 $[X_1, Y_1, Z_1]$ = Kinesics data vector
 h = number of hyperplanes
 \bullet = dot product

The optimization problem then is to minimize w , and h , wherein h hyperplanes classify the kinesics data vectors into different kinesics classes.

- **C4.5 Classifier [17]:** In this classifier, construction of the decision tree from a set of kinesics data vector. This set is repeatedly split based on the given classification test. In our case, this classification test is the kinesics class. Formally, this can be described as follows:

$$\forall I, ([X_1, Y_1, Z_1], K_1); [X_1, Y_1, Z_1] \in D, K_1 \in C \quad (8)$$

Where,

D = Set of kinesics data vector
 C = Set of kinesics classes
 $[X_1, Y_1, Z_1]$ = Kinesics data vector

We note that the threshold values correspond to kinesics data vector, while types of kinesics can be represented using the kinesics classes. Thus, by formulating the difference in kinesics of the user as a classification problem, enables us to distinguish different kinesics either from the single or multiple users, as described in SUMK and MUSK scenarios. Hence, kinesics serves as a reliable indicator of the identity of the user. Furthermore, the kinesics can be revealed to an adversary, who has control of a RFID reader and an antenna to query the wirelessly instrumented space. Therefore, user kinesics is the source of privacy leakage in such an environment.

V. EXPERIMENTAL EVALUATION

We perform a set of qualitative experiments to find the evidence for privacy leakage hypothesis. These experiments provide detailed methods for the classification of user kinesics for SUMK and MUSK scenarios. We construct the five classifiers as mentioned in section IV, and map them to the requirements of SUMK and MUSK scenarios. This allows us to quantify the information that can be leaked to an adversary. In section VI, we show that the construction of such adversarial argument facilitates in the strengthening of privacy of kinesics and identity of the user. We now layout the system configuration and methodology for the experimental evaluation.

A. System Configuration

For our experiments, we have used the following hardware and software configuration in the experimental setup:

TABLE I. EXPERIMENTAL TEST BED

System Type	System Configuration			
	Hardware		Software	
Workstation	CPU	Intel Core 2 Duo at 2 GHz	OS	WinXP
	RAM	2 Gigabytes	Prog. Support	C#
	Hard Disk	100 Gigabytes	LOC	1190
RFID Backend	Reader	Impinj UHF 2 Speedway Reader	Protocol	LLRW 1.01
	Antenna	Linear		
WISP Glove	Sensor	WISP accelerometer sensor	Firmware version	4.1DL
	Glove	Standard glove		

LOC – Lines of Code

It should be noted here that the kinesics recognition system runs on the workstation, and interprets the accelerometer data generated from WISP glove through the RFID backend. It is also enables the programmable events based on given kinesics, as discussed in section III.

B. Methodology

We adopt a consistent set of rules and methods across experiments to minimize any methodology-related error in the results.

As discussed in section III, we have constructed three representative hand kinesics that involves tilting hand in three different directions. Furthermore, we asked 10 volunteer subjects (five male and five female) to perform these kinesics. For each subject, we have gathered 10 iterations of 10 sample points over three kinesics for a total of 300 samples. We call the set of 300 samples per subject as a “*template*”. For each of the gathered template, we determine the AVG, and STDEV metrics, to understand and place bound on the nature of data. On the basis of these metrics, we correlate the templates of two male subjects to determine the difference in their kinesics. Also, we correlate the templates of two female subjects for the same purpose.

Additionally, we correlate the template of a male subject with the template of a female subject to highlight the difference in kinesics across gender. Also, we separately perform correlation between different kinesics for a given male and female subject. This enables us to determine differences between one kinesics to another, within that subject.

In section IV, we introduced the classification of user kinesics based on the five base-level classifiers. We employ these classifiers to determine differences of kinesics between two male subjects, two female subjects, and a male and a female subject. To simulate the arrival of new samples from a test subject, whose template is already developed, we synthesize the new samples from the template. This is done by using a simple transformation that involves subtracting an amount of data per axes, which preserves the unique values in new samples, while maintaining the new samples within the standard deviation of the template. Formally, the new

samples synthesized from the proposed technique are described as follows:

$$\forall I, [X_1 - \frac{\sigma_x}{S_x}, Y_1 - \frac{\sigma_y}{S_y}, Z_1 - \frac{\sigma_z}{S_z}] \quad (9)$$

Where,

$[X_i, Y_i, Z_i]$ = Kinesics data vector

$\sigma_x, \sigma_y, \sigma_z$ = Standard deviation of kinesics data per axis

S_x, S_y, S_z = Sample size of kinesics data per axis

It is conceivable that more sophisticated techniques exist to synthesize new samples from the template. Our intention is to present a motivating example. Furthermore, we use this technique with the classifiers to determine the similarity of the new samples and the template of a male and a female subject, respectively. The resultant effect of the technique is illustrated in Fig. 6:

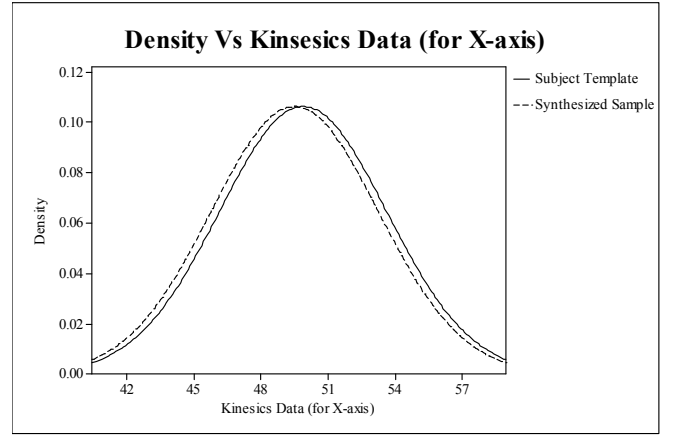


Figure 6. Distribution of subject template and synthesized sample

We note that by using (9), distribution of the synthesized sample is “*left-shifted*” slightly with respect to the subject template. When data from synthetic sample arrives at the classifier that is trained with subject template, the result indicates the similarity of kinesics data, while treating the synthetic sample as a real time update that belongs to the subject template. Furthermore, these experiments enable us to satisfy the requirements of SUMK and MUSK scenarios.

C. Results

Using the proposed experimental methodology, we have performed detailed experiments for the SUMK and MUSK scenarios. Our results provide strong evidence in support of the privacy leakage hypothesis. For the sake of brevity, we only report the AVG, and STDEV metrics for the PPLT kinesics of the ten subjects (S1, S2, ... S10). However, it should be noted that other kinesics (PPRT, and PPDT) provide similar insights. Also, we report correlation data, classification accuracy, and confusion matrices for the type of subjects discussed in previous section.

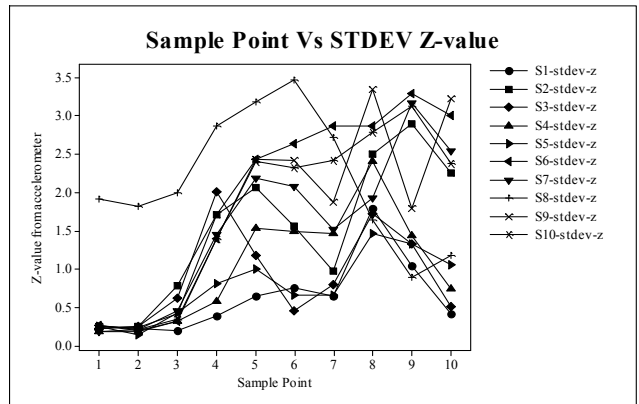
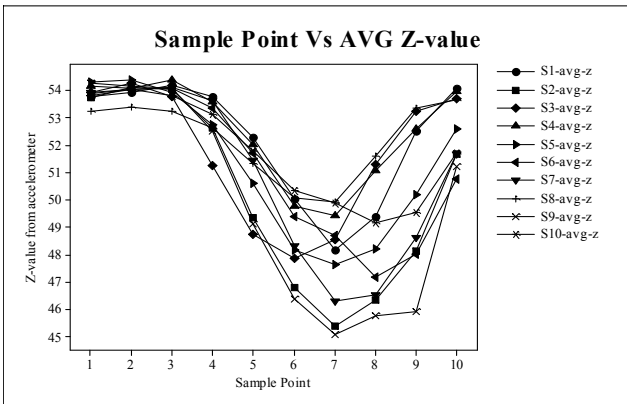
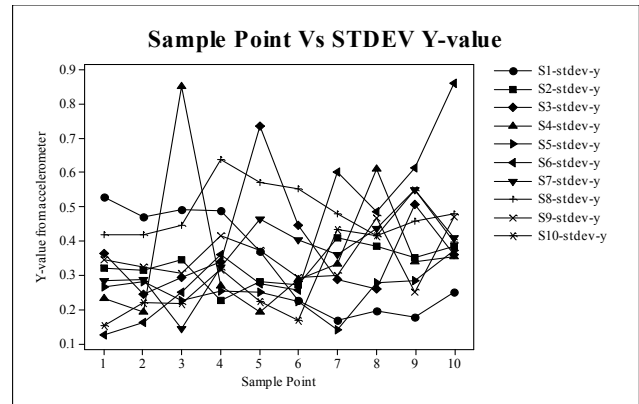
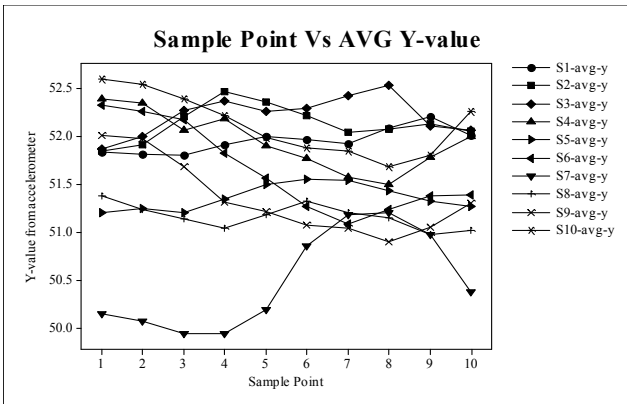
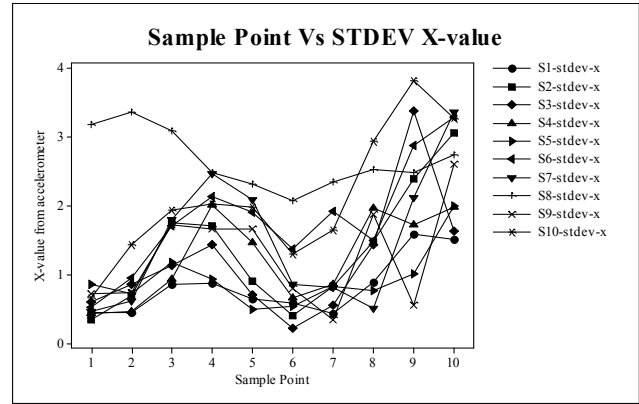
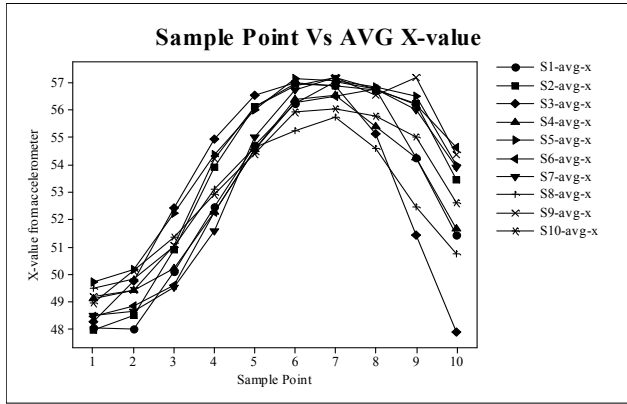


Figure 7. AVG, and STDEV metrics for PPLT kinesics of 10 subjects in $[X,Y,Z]$ -axis

In Fig. 7, we note that for PPLT kinesics, 10 subjects can be distinguished based on just 10 sample points along the Y -axis. Furthermore, acceleration data for the 10 subjects, along the X -axis and the Z -axis is tightly bound enough to not yield any meaningful distinguishing features about PPLT kinesics. Also, each sample point represents average of 10 iterations of data collected from 10 subjects. Moreover, by analyzing the STDEV metric for PPLT kinesics, it can be concluded that every subject contributes a distinguishing variation to the kinesics. This aspect of the analysis is particularly important due to the

fact that subjects were asked to perform the kinesics in a constrained manner. By using the principle of symmetry, it is conceivable that similar distinguishing features exist for PPRT kinesics. For PPDT kinesics, distinguishing features exist along the X -axis and Z -axis. Indeed, the data provides such evidence. We have excluded the data of PPRT and PPDT kinesics for the sake of brevity.

Based on the experimental evidence, it should be clear that user kinesics serve as a means to privacy leakage. It also strengthens MUSK scenario, showing that different subjects have distinguishing features in their kinesics.

TABLE II. CORRELATION OF KINESICS DATA FOR SUBJECTS

Subjects						
Subject 1 and 3 (Male/Male)			Subject 10 (Male)			
	PPLT	PPRT	PPDT	LR	RD	DL
X	0.808	0.651	0.860	-0.872	-0.922	0.955
Y	0.244	0.446	0.653	0.934	-0.918	-0.972
Z	0.753	0.830	0.677	0.896	0.940	0.980
Subject 5 and 9 (Female/Female)			Subject 6 (Female)			
X	0.990	0.943	0.498	-0.712	-0.279	0.721
Y	-0.764	0.814	0.898	0.818	-0.565	-0.918
Z	0.971	0.948	0.949	0.527	0.156	0.867
Subject 4 and 7 (Male/Female)			Subject 2 (Female) – T/S			
X	0.963	0.784	0.832	1.000 (LL)	1.000 (RR)	1.000 (DD)
Y	-0.880	-0.780	0.912	1.000 (LL)	1.000 (RR)	1.000 (DD)
Z	0.889	0.811	0.901	1.000 (LL)	1.000 (RR)	1.000 (DD)

(LR – PPLT, PPRT), (RD – PPRT, PPDT), (DL – PPDT, PPLT), (LL – PPLT, PPLT), (RR – PPRT, PPRT), (DD – PPDT, PPDT), (T/S – Template/Synthetic)

TABLE III. CLASSIFICATION ACCURACY AND CONFUSION MATRIX OF KINESICS FOR SUBJECTS

Subjects					
Subject 8 (Male) – Self					
	Naïve Bayes	C4.5	SVM	CN2	kNN
LRD	0.880	0.907	0.830	0.943	0.997
L_C, L_W	(92, 8)	(97, 3)	(83, 17)	(99, 1)	(100, 0)
R_C, R_W	(94, 6)	(98, 2)	(86, 14)	(90, 10)	(99, 1)
D_C, D_W	(78, 22)	(77, 23)	(79, 21)	(94, 6)	(100, 0)
Subject 1 and 3 (Male/Male)					
LRD	0.773	0.580	0.687	0.597	0.767
L_C, L_W	(88, 12)	(85, 15)	(89, 11)	(90, 10)	(89, 11)
R_C, R_W	(71, 19)	(25, 75)	(44, 56)	(25, 75)	(64, 36)
D_C, D_W	(73, 27)	(64, 36)	(72, 28)	(64, 36)	(77, 23)
Subject 5 and 9 (Female/Female)					
LRD	0.723	0.730	0.737	0.720	0.737
L_C, L_W	(84, 16)	(87, 13)	(81, 19)	(91, 9)	(90, 10)
R_C, R_W	(56, 44)	(59, 41)	(60, 40)	(58, 42)	(58, 42)
D_C, D_W	(77, 23)	(73, 27)	(79, 21)	(67, 33)	(73, 27)
Subject 4 and 7 (Male/Female)					
LRD	0.790	0.780	0.753	0.787	0.767
L_C, L_W	(77, 23)	(70, 30)	(64, 36)	(68, 32)	(57, 43)
R_C, R_W	(98, 2)	(99, 1)	(100, 0)	(99, 1)	(100, 0)
D_C, D_W	(62, 38)	(65, 35)	(62, 38)	(69, 31)	(73, 27)
Subject 2 (Female) – T/S					
LRD	0.790	0.912	0.810	0.940	1.000
L_C, L_W	(90, 10)	(99, 1)	(79, 21)	(100, 0)	(100, 0)
R_C, R_W	(72, 28)	(94, 6)	(69, 31)	(94, 6)	(100, 0)
D_C, D_W	(75, 25)	(82, 18)	(95, 5)	(88, 12)	(100, 0)

(L_C – PPLT Correct, L_W – PPLT Wrong, R_C – PPRT Correct, R_W – PPRT Wrong, D_C – PPDT Correct, D_W – PPDT Wrong)(LRD – PPLT, PPRT, PPDT), (T/S – Template/Synthetic)

In Table II. first column, we correlate the acceleration data of three kinesics in 3-axes for different combination of subjects. We find that for the two male subjects (subject 1 and 3), PPLT kinesics does not correlate well for the Y-axis. Thus, Y-axis correlation value distinguishes the PPLT kinesics of the two subjects. This result matches with the analysis, where 10 subjects are shown to have distinguishing PPLT kinesics based on the Y-axis acceleration data. Furthermore, we see that the PPRT kinesics has low correlation value along the Y-axis, thus providing the evidence for the principle of symmetry. We note that for the PPDT kinesics, low correlation values exist along the Y-axis and the Z-axis. Initially, this may appear as a contradiction to the claim, where we suggest that distinguishing features exist along the X-axis and the Z-axis. However, we compare only two particular subjects, instead of all the subjects. It is conceivable that there are subjects that have distinguishing features along the X-axis and the Z-axis. In general, we suggest that the axes, which have low correlation value among all the axes, are an indicator that distinguishing features exist along those axes. The correlation analysis provides further evidence for the MUSK scenario.

In second column of the Table II, we separately, correlate the acceleration data for three kinesics of a male and a female subject (subject 10 and 6). Low correlation values along any axes enable us to distinguish between the two kinesics for that subject. The correlation analysis corroborates the SUMK scenario. Furthermore, in case of the female subject (subject 2), high correlation values for the subject template and the synthetic sample, shows the similarity of same kinesics (LL, RR, DD) taken from different datasets.

Table III, shows the classification accuracy derived from five base-level classifiers for different combination of the subjects. When a male subject (subject 8) kinesics is self compared, the classification accuracy derived from all the classifiers, is high. Consequently, classification of the subject template with itself over all the kinesics can be treated as an “accuracy” threshold, against which other classification accuracy results can be compared. We note here that LRD corresponds to PPLT, PPRT, and PPDT kinesics.

Additionally, we measure the ability of the classifiers to correctly classify the kinesics of the subjects through a confusion matrix. In a confusion matrix, for each kinesics, we report the correct classification (L_C, R_C, D_C), and the wrong classification (L_W, R_W, D_W). When a male subject (subject 8) kinesics is self compared, we see that all the classifiers have high values for the correct classification ($L_C = 92, R_C = 94, D_C = 78$ for Naïve Bayes Classifier), and low values for the wrong classification ($L_W = 8, R_W = 6, D_C = 22$ for Naïve Bayes Classifier). Thus, the confusion matrix provides a rigorous indicator, wherein high value for the correct classification corresponds to subjects being more similar than different. Alternatively,

it can be seen that confusion matrix for a male subject and a male subject comparison (subject 1 and 3), a female subject and a female subject comparison (subject 5 and 9), and a male subject and a female subject comparison contains relatively low values for the correct classification. Hence, suggesting that subjects are different. We note that this difference is based on the difference in the kinesics of the subjects. Furthermore, in the case of comparison between a female subject template, and the synthetic sample (subject 2), the confusion matrix indicates high values for the correct classification. This further strengthens the previously mentioned correlation results for the same subject.

We note that the experimental evidence supports the proposed privacy leakage hypothesis. Therefore, it can be concluded that the user kinesics are a reliable indicator of the identity of the user.

VI. MITIGATION TECHNIQUES

The purpose of privacy leakage hypothesis, and the related experimental evaluation was to show that user kinesics may disclose identity of the user to an adversary. We now focus on the mitigating techniques to address the problem of safeguarding kinesics data, and the identity of the user.

A. Mitigation Technique I: Conceal

Our first mitigation technique involves encryption of the acceleration data emitted by the WISP accelerometer sensor. Considering that the WISP device is a resource constrained platform, stronger cryptographic techniques may not be immediately applicable. Therefore, we rely on a OTP based symmetric encryption scheme that provides the required security, without overcharging the WISP device of its performance. Fig. 8 illustrates the proposed scheme in more details:

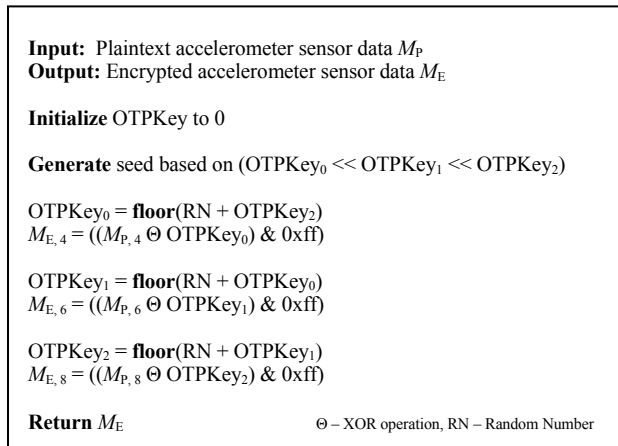


Figure 8. OTP based symmetric encryption scheme on WISP

An EPC Gen2 compliant RFID reader expects 96-bit EPC ID from an EPC Gen2 compliant RFID tag [19]. We note that a WISP device emulates an EPC Gen2 compliant

RFID tag, wherein accelerometer sensor data values are encoded as EPC ID and returned to the RFID reader. In particular, WISP device responds with a 16-byte message to the querying RFID reader. The structure of the message is given as follows:

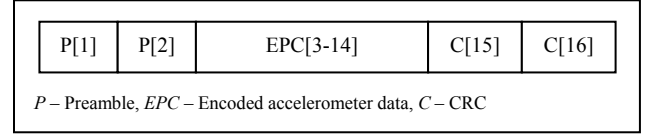


Figure 9. 16-byte response message of the WISP device

We note that the WISP device encodes accelerometer sensor data in the 4th, 6th and 8th byte of the EPC field as depicted in Fig. 9. In Fig. 8, we have described an OTP based symmetric encryption scheme, which operates on these sensor data byte locations, before 16-byte response message is returned by WISP device to the RFID reader. We evaluate the effectiveness of the OTP symmetric encryption scheme by comparing the accelerometer data along *Y*-axis for the PPRT kinesics as follows:

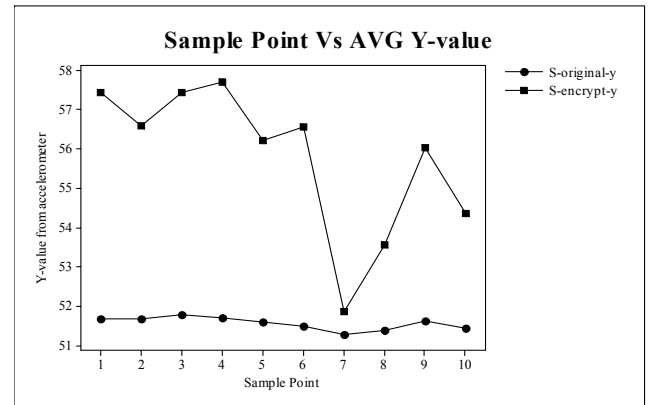


Figure 10. PPRT kinesics data along *Y*-axis with/without proposed encryption scheme

In Fig. 10, the difference between the actual and the encrypted PPRT kinesics data along *Y*-axis can be seen. When the proposed encryption scheme is active on the WISP device, the RFID reader, and an adversary view the encrypted kinesics data. Thus, concealing the kinesics data that is transmitted by the WISP device, makes it harder for the adversary to gain private information about the user. We note that our relatively simple encryption scheme can be replaced with other sophisticated schemes based on rigorous security proof e.g. RSA, and AES. Our intention to propose the scheme is to introduce the idea, without rigorously proving it to be secure.

B. Mitigation Technique II: Confuse

The 16-byte response message returned by the WISP device to a RFID reader, contains fixed byte pattern consisting of the WISP device ID. This forms the basis of our second mitigation technique.

In this scheme, we propose saturating the environment with the WISP devices with same WISP device ID. With careful placement of the WISP devices around the kinesics recognition system, we construct a WISP “cage” that enables us to collect correct kinesics data, while at the same time providing the adversary with perturbed samples of kinesics data. Fig. 11 illustrates the concept:

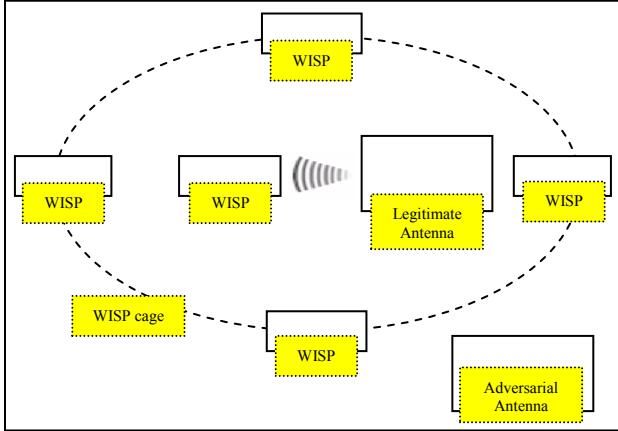


Figure 11. Structure of WISP cage

To evaluate this technique, we gather PPRT kinesics data along Y-axis with and without WISP cage. The results of the evaluation are shown as under:

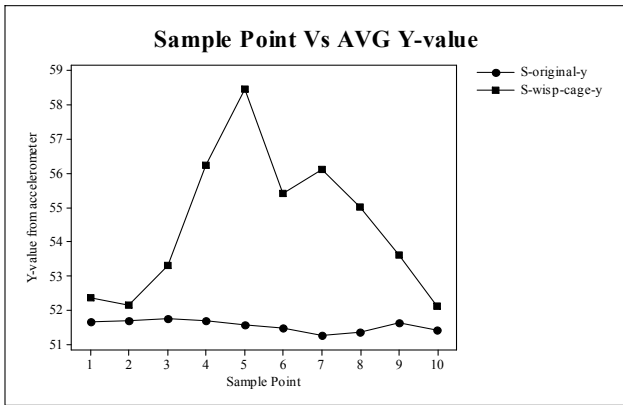


Figure 12. PPRT kinesics data along Y-axis with/without proposed technique

We have used one WISP device with same device ID as WISP glove. As shown in Fig. 12, the adversary gathers perturbed samples of the user kinesics data. The adversary then has to extract meaningful data from the collected sample. Thus, by deploying the WISP devices around the WISP glove with the same device ID, the task of retrieving correct kinesics data becomes harder for the adversary.

VII. FUTURE WORK

We have provided the privacy leakage theory within the context of a kinesics recognition system. We intend to further develop the theory and the supporting experimental

evidence for a wirelessly instrumented space that includes multi-modal sensors and RFID tags. This will enable us to reason about privacy leakage in more general setup.

Furthermore, we have suggested two of many possible mitigation techniques to address the privacy concerns. We are investigating other potential candidates that can be applied to alleviate the user privacy concerns. Also, we will consider a provably secure encryption scheme as a replacement to the currently proposed encryption scheme.

Additionally, we have not analyzed the relationships between gathered kinesics data, and spatio-temporal aspects of occurrence of events. We intended to include such relationships to learn additional information that may be of interest to an adversary.

VIII. CONCLUSIONS

In this paper, we have presented a kinesics recognition system as the representative wirelessly instrumented space. We have developed three simple yet meaningful user kinesics, and corresponding applications to indicate the potential of the proposed system. Furthermore, we have proposed a privacy leakage theory, wherein user kinesics is shown to be reliable indicator of identity of the user. Also, we have performed detailed experiments based on different subjects involving correlation, and classification of their kinesics. Our results provide strong evidence in support to the proposed theory. Additionally, we have suggested two mitigation techniques to address the privacy concerns. Thus, our work provides a systematic study of user privacy when operating in a wirelessly instrumented space.

REFERENCES

- [1] J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J. M. Rehg, “A Scalable Approach to Activity Recognition based on Object Use”, Proc. IEEE International Conference on Computer Vision (ICCV 07), Atlanta, USA, IEEE Press, Oct. 2007, pp. 1-8.
- [2] E. M. Tapia, S. S. Intille, and K. Larson, “Activity Recognition in the Home Using Simple and Ubiquitous Sensors”, Proc. International Conference on Pervasive Computing (PERVASIVE 04), Vienna, Austria, Springer Press, LNCS, Mar. 2004, pp. 158-175.
- [3] T. Stiefmeier, D. Roggen, G. Tröster, G. Ogris, and P. Lukowicz, “Wearable Activity Tracking in Car Manufacturing”, IEEE Pervasive Computing Magazine (Pervasive Computing 08), IEEE Press, Vol. 7, Issue. 2, Apr.-Jun. 2008, pp. 42-50.
- [4] D. Sánchez, M. Tentori, and J. Favela, “Activity Recognition for the Smart Hospital”, IEEE Intelligent Systems Magazine (Intelligent Systems 08), IEEE Press, Vol. 23, Issue. 2, Mar.-Apr., 2008, pp. 50-57.
- [5] D. H. Wilson, A. C. Long, C. Atkeson, “A ContextAware Recognition Survey for Data Collection Using Ubiquitous Sensors in the Home”, Proc. ACM Conference on Human Factors in Computing Systems (CHI 05), Portland, USA, ACM Press, Apr. 2005, pp. 1865-1868.
- [6] B. Logan, J. Healey, M. Philipose, E. M. Tapia, and S. Intille, “A Long-Term Evaluation of Sensing Modalities for Activity Recognition”, Proc. International Conference on Ubiquitous Computing (UBICOMP 07), Innsbruck, Austria, Springer Press, LNCS, Sep. 2007, pp. 483-500.

- [7] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity Recognition from Accelerometer Data", Proc. Conference on Innovative Applications of Artificial Intelligence (IAAI 05), Pittsburgh, USA, ACM Press, Jul., 2005, pp. 1541-1546.
- [8] D. Lymberopoulos, T. Teixeira, and A. Savvides, "Detecting Patterns for Assisted Living Using Sensor Networks: A Case Study", Proc. IEEE International Conference on Sensor Technologies and Applications (SensorComm 07), Connecticut, USA, IEEE Press, Oct. 2007, pp. 590-596.
- [9] V. Srinivasan, J. Stankovic, and K. Whitehouse, "Protecting your Daily In-Home Activity Information from a Wireless Snooping Attack", Proc. International Conference on Ubiquitous Computing (UBICOMP 08), Seoul, South Korea, ACM Press, Sep. 2008, pp. 202-211.
- [10] T. S. Saponas, J. Lester, C. Hartung, T. Kohno, "Devices That Tell On You: The Nike+iPod Sport Kit", Proc. USENIX Security Symposium (USENIX Security 07), Boston, USA, 2007, pp.55-70.
- [11] S. L. Garfinkel, A. Juels, R. Pappu, "RFID Privacy: An Overview of Problems and Proposed Solutions", IEEE Privacy and Security Magazine (Security and Privacy 05), IEEE Press, Vol. 3, Issue 3, 2005, pp. 34-43.
- [12] A. P. Sample, D. J. Yeager, P. S. Powledge, and J. R. Smith, "Design of a Passively-Powered, Programmable Sensing Platform for UHF RFID Systems", IEEE International Conference on RFID, (RFID 07), Texas, USA, IEEE Press, Mar. 2007, pp. 149-156.
- [13] D. D. Lewis, "Naive (Bayes) at forty: The independence assumption in information retrieval", European Conference on Machine Learning (ECML 98), Chemnitz, Germany, Springer Press, Apr. 1998, pp. 4-15.
- [14] P. Clark, and T. Niblett, "The CN2 Induction Algorithm", Journal of Machine Learning, Springer Press, Vol. 3, No. 4, 1989, pp. 261-283.
- [15] T. M. Cover, and P. E., "Nearest Neighbor Pattern Classification", IEEE Transaction on Information Theory, IEEE Press, Vol. 13, Issue 1, 1967, pp. 21-27.
- [16] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Journal of Data Mining and Knowledge Discovery, Springer Press, Vol. 2, No. 2, 1998, 121-167.
- [17] J. R. Quinlan, "Improved Use of Continuous Attributes in C4.5", Journal of Artificial Intelligence Research, Vol. 4, Issue 1, 1996, 77-90.
- [18] EPCGlobal, "EPC Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860 MHz-960MHz Version 1.2.0", Specification for Air Interface, 2009
- [19] EPCGlobal, "EPCglobal Tag Data Standards Version 1.4", 2009
- [20] Microsoft, "Windows API Reference", Microsoft Corp., www.microsoft.com
- [21] X10, "X10 Home Automation Modules", X10 Inc., www.x10.com