# Fitness Trackers and Wearable Devices: How to Prevent Inference Risks?

Ilaria Torre
ilaria.torre@unige.it

Frosina Koceva
frosina.koceva@gmail.com

Odnan Ref Sanchez
odnan.sanchez@tnt-lab.unige.it

Giovanni Adorni
giovanni.adorni@unige.it

Department of Computer Science, Bioengineering, Robotics and Systems Engineering (DIBRIS)
Genoa, Italy

## ABSTRACT

Wearable and personal devices are becoming more and more part of people's everyday lives. These devices produce enormous amount of personal data which are handled by third parties as authorized by the user. However, such third parties may be able to infer sensitive information using the collected personal information. In this paper we present a case study based on fitness trackers and we sketch our model for privacy management and inference prevention. For this study, we built a Bayesian Network and used it to compute the risk of inferring unknown data. Using the simulated case we show the feasibility of inferring some private data from a set of personal data available to a third party as authorized by the user (i.e., sensor data and profiling data provided by the user while registering for the service). This paper provides a step towards the open issues of privacy and security management in the field of ubiquitous devices.

## CCS Concepts

•**Security and privacy** → Human and societal aspects of security and privacy - Privacy protection;

## Keywords

User support service; Privacy management; User attribute inference; Wearable devices; Fitness tracker, Internet of Things, Personal data manager

## 1. INTRODUCTION

As the advent of Internet of Things (IoT) has come, an estimate of 16.3 billion devices has already been connected to the network as of 2015 and the market is expected to steadily increase and reach 26.3 billion devices by the year 2020 [8]. Reaching an estimated compound annual growth rate of 49%, IoT in the field of health and wellness (e.g., health trackers, medicine dispensers, wellness devices, etc.) is anticipated to achieve the fastest growth in the IoT market [8]. This tremendous increase of personal devices is able to exponentially increase an individual's personal data that will be processed and shared among third parties.

Today, wearable devices such as fitness trackers (provided for example by Fitbit, Jawbone, Apple, Garmin, etc.) have caught the attention of many users due to their promise of improving their active and healthy lifestyle. Fitbit trackers [12], for example, are able to monitor the individual's steps, sailed distances, calories burned, active minutes, hourly activity and stationary time. This tracker also reminds the user to move and tracks her/his sleeping habits. All these are possible due to its numerous sensors (i.e., GPS, 3-axis accelerometers, 3-axis gyroscope, digital compass, optical heart rate monitor, altimeter, ambient light sensor, vibration motor, etc.) built in the tracker. Fitbit users can also share their activities on social networks and challenge their friends in a competition which further encourages them to be active. In this study, we focus on evaluating fintess trackers, as a working example.

One of the major state-of-the-art issues of IoT concerns the security and privacy of users. In the past years, several approaches and solutions have been proposed to improve the user security and privacy by preventing unauthorized access of data generated by IoT devices. However, recent articles [15] [27] report about privacy breaches coming from the supposedly trusted third parties which are granted by the user's permission. A relevant risk is that third parties may be able to derive personal and sensitive information of the users by processing (e.g., through data mining techniques, machine learning algorithms, predictive filtering, etc.) the collected user information.

The use of personal data managers (PDM) is emerging as one of the ways to manage and control heterogeneous personal devices. These tools are conceived as a gateway to all third parties trying to access sensor data of a specific user. They can be located in a user's device, computer or in a personal cloud (e.g., InPUT Project [16]). They typically include security features, however, current PDMs are mostly focused on data minimization, authorization and authentication control, which do not prevent inference derivation.

In this respect, we proposed a model to extend PDMs [28] [29] in order to secure the user from undesired inference derivation of sensitive information. Basically it defines an



*BODYNETS 2016 December 15–16, Turin, Italy*



DOI: 10.1145/1235

BODYNETS 2016, December 15-16, Turin, Italy

DOI 10.4108/eai.15-12-2016.2267791

Adaptive Inference Discovery Service (AID-S), designed to be integrated in a PDM in order to manage information prevention. It is aimed to compute the inference risk measured as the probability to infer new data from a set of shared data. This information can be used to take further actions such as alerting the user or transforming personal data items before they are shared. State-of-the-art solutions and techniques are also integrated in the approach.

In this paper, we briefly present our model and then we describe an instance of it focused on preventing inferences when using fitness wearable devices such as Fitbit and the other trackers mentioned above. In this implementation, we use Bayesian Networks for computing the inference risk.

The remainder of the paper is structured as follows. Section 2 presents the background and related works. Section 3 sketches the model based on our framework and Section 4 evaluates the case of fitness trackers using a Bayesian Networks for computing the inference probability. Finally, Section 5 concludes the paper with possible future works.

## 2. RELATED WORKS

In this section, first we present an overview of approaches to manage personal/sensed data shared among third parties. Then, we mention the state-of-the-art related works concerning inference risks when using wearable IoT devices.

***Privacy management***

Privacy concerns have given rise to new proposals for empowering users with control over their data. Several projects are being carried out to develop platforms which allow users to control their personal data and the policies to make them available to third parties. PDMs belong to such solutions [30, 34, 7]. These platforms typically require users to specify what they want to hide from or share with a third-party application. Access is allowed only if the corresponding permission is granted by the user. However, regulating the access of one's data cannot deny the possible disclosure through inference. Inference attacks are based on the integration and correlation of known data about an individual which leads to the discovery of private data. The inference is done by linking sensitive information to the knowledge that may be available to a third party as a background knowledge, common sense or domain-specific knowledge [23]. In this respect, a framework that is similar to our model is ipShield [6]. As in our approach, it monitors the inference risks, but, it is limited only to the sensors on the phone.

Several techniques can be found in the literature to address inference attacks. Most of them concern anonymization and transformation [2]. However, such techniques cannot be always deployed. For example, [26] analyze physiological signals from wearable devices. They show that several inferences can be derived (e.g. detecting variation of heartbeat, respiratory disorders) but they explain that, in these cases, defensive techniques that are based on data transformation and obfuscation may be critical since they can compromise diagnostic services. For instance, the interbeat interval in ECG is typically between 300ms to 2,000ms [9] and it should not loose accuracy. In this case, only approaches that preserve the accuracy and the characteristic of the signal are appropriate.

***Personal data leakage from a singular data source***

The popularity of wearable devices, fitness devices and health care monitoring systems has increased the detail level, volume, collection, exchange and sharing of personal user data among applications. While on the one hand there are various health and wellness benefits, the disadvantage is that the shared data is raising privacy concerns. For instance, the results of [19] indicate that keystroke inference attacks using typing-induced motion data captured by smartwatch gesture classification using accelerometer and gyroscope data is highly effective. Furthermore, physiological data shared for healthcare studies can also be used to infer addictions like smoking or drinking [11]. For instance, by using custom hardware and analyzing the data from wrist-worn sensors, it has been shown how to successfully classify smoking gestures [22], eating gestures [10] or smoking cessation [25]. Motion sensor data from smart phones held in hand has been shown to provide enough information to impersonate a stylus [1] and a mouse [33]. From data sensed by activity trackers, [32] examine whether a predefined event, e.g., walking from office to a nearby coffee shop, can be identified from among many other routine inference of sensitive human behavior events. Through real-world experiments, they found that pedometer readings that are captured even at a coarse granularity, such as one minute, enable the accurate inference of events such as grocery shopping, walking to coffee shop, visiting gym, etc.

***Personal data leakage from aggregated data of different sources***

Even though singular data could appear harmless and non-sensitive, assembling each one with other data from different sources can result in a so-called global inference attack and global inference chains [13]. This risk, based on the correlation and aggregation of data from heterogeneous sources and even from self-generated content, is often not grasped by users. In our previous work [4, 5] on inference attacks in social systems, we analyzed the inadequacy of permission-based protection for users that accorded their permission to a subset of data. The experimental study showed the possibility for an adversary to aggregate user data discovered from various sources into a more complete profile, and infer further data that the user had not shared. Moreover, we found that the risk of information leakage in online social networks is directly correlated with the number of shared personal data and with the number of user profiles owned by the user over time and then forgotten. This risk can be applied also to user data that are collected from IoT devices and shared, suggesting that the higher the number of devices owned by the user and running background applications (e.g., smart home gadgets, fitness appliances, etc.), the higher the risk that a huge amount of historical personal data are collected with associated risks of information leakage.

For instance, Fitbit provides users with the possibility of sharing their fitness statistics on Tweeter, i.e. #fitstats tweets are created together with a link to the user's Fitbit dashboard. Based on the correlation of the tweeter profile public data with Fitbit dashboard public data a possible de-anonymization attack and behavior inference could be performed. Furthermore, the user's average step counts are shared by default unless user opts-out, thus a user's behavior can be easily inferred as described by [32].

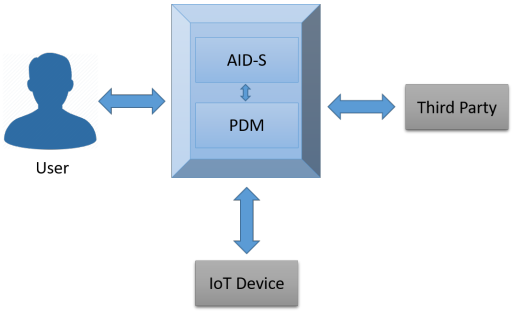


**Figure 1: General overview of our proposed model.**

# 3. INTERACTION MODEL

This section describes the model we defined to prevent the risk of personal data disclosure through inference attacks. Figure 1 shows the general overview of our model which is in accordance with the framework described in in [29].

As shown in Figure 1, the <PDM + AID-S> function block acts as a gateway among a user, an IoT personal device and a third party. The interaction among these entities is controlled by the <PDM + AID-S> function block which ensures that i) the user is made aware about privacy risks related to her/his privacy settings ii) a common privacy agreement is defined among them. PDMs can be placed in the user phone, laptop, or in personal clouds [16] and AID-S works as a service requested by the PDM to accomplish reasoning tasks which may be integrated in the PDMs or being asked as a separate cloud service.

## 3.1 Interaction Workflow

This part describes briefly the workflow of the system which is derived from the previous work in [29]. The interaction workflow consists of 4 phases, namely, Statement Declaration, Inference Check, Recommendation and Negotiation. The flow starts as soon as a third party requires for the first time to access a set of personal data to provide its services to the user.

### *3.1.1 Statement Declaration*

To start, a Statement is sent to the PDM by the third party. Formally, a Statement is a declaration wherein the third party specifies the list of data items it requires and processes, their quality, granularity, and which of them will be shared with other applications. A detailed example of a Statement is provided by My Data Store PDM [30].

When the PDM receives this Statement, it examines it and compares it with the user privacy settings that are already stored in the user profile. The related steps depend on how a PDM operates as in [30, 34, 7]. After the PDM approves that a Statement is compliant with the user privacy settings, it sends a request to AID-S for an inference check.

### *3.1.2 Inference Check*

This phase is aimed to measure the risk that a third party is able to infer sensitive information by processing the collected data gathered from the list of data stated in the Statement. An ideal representation of inference probabilities based on correlations among user features is defined in the inference matrix below.

DEFINITION 1. *We define an inference matrix, $I_{j,k}$, as the set of all the probabilities that a user attribute can be inferred given the combination of data stated in the Statement.*

$$I_{j,k} = \begin{pmatrix} P(a_1|c_1) & \cdots & P(a_1|c_k) \\ \vdots & \ddots & \vdots \\ P(a_j|c_1) & \cdots & P(a_j|c_k) \end{pmatrix} \quad (1)$$

*where:*

- $a_j$*: is the* $j^{th}$ *user data that could be inferred*
- $c_k$*: is the* $k^{th}$ *combination of user data*
- *D: user data set,* $a_j \in D$ $(1 \leq j \leq |D|)$
- $P\{D\}$*: power set of D,* $c_k \in P\{D\}$ $(1 \leq k \leq |P\{D\}|)$
- $P(a_j|c_k) \in [0, 1]$

AID-S refers to this matrix when a request is issued from the PDM to check if there is a risk of possible inference. The key is to find combinations of data listed in the Statement that are highly correlated with other non-shared personal data, since this results in high probability to infer such data. However, the real risk for user privacy depends also on the user perception about the possible disclosure of such data. Some individuals could rate $a_j$ as a sensitive data while others could be less worried about its disclosure. The user perception about possible disclosure is expressed in the user profile (with her/his privacy settings). Thus, in case $P(a_j|c_k)$ reaches more than the set threshold for $a_j$ disclosure, this information is sent to the PDM for managing recommendation to the user (see the next phase).

Notice that the probabilities in the matrix can be computed according to several algorithms (e.g., data mining techniques, machine learning algorithms, predictive filtering, etc.) on proper data sets, and/or by exploiting correlations and conditional probabilities that are already available in the literature. In this study, we exploit the correlations found in other works, referenced in Table 1.

From the description above, it is clear that the Inference Check is carried out by using Statement data only (i.e., metadata). This means that AID-S does not store and process personal user data, but only their description. This enables AID-S to be easily integrated into many PDM platforms with no need to be trusted by the users it itself as a new third-party application.

### *3.1.3 Recommendation*

When it is over a given threshold, the inference risk is relayed by the PDM to the user. As stated above, this threshold specifies users preferences of sharing a data item. Privacy settings are stored in a user profile. They can be configured by the users on the PDM (as for example in [6]) or estimated by the PDM from a set of user-configured data items [20]. Recommendation techniques notify about risks and provide suggestions about personal data that should not be shared with the third party in order to lower the probability of inferring $a_j$.

### *3.1.4 Negotiation*

In this phase, the third-party Statement can either be refused or accepted entirely/partially by the user, depending on the recommendation provided by the PDM about the privacy settings. The negotiation among the user, the third party and the <PDM + AID-S> goes on until a satisfying

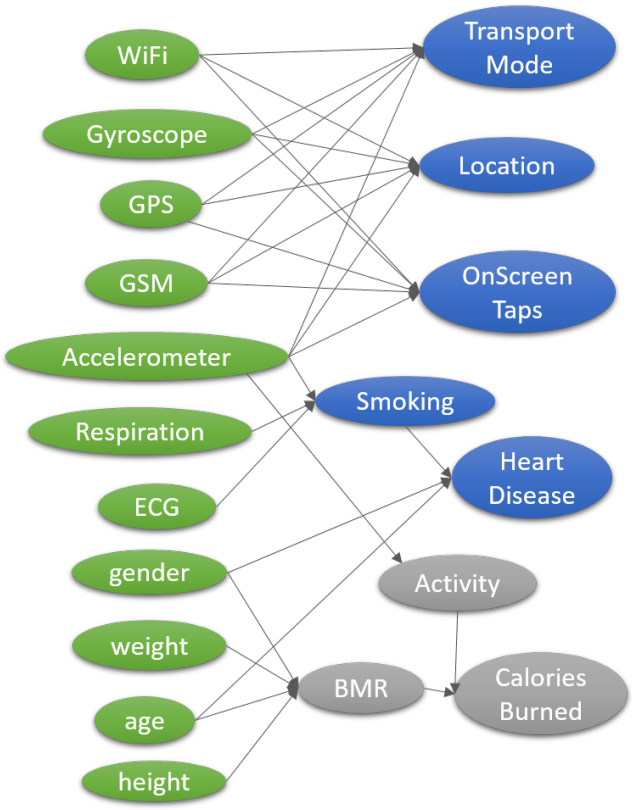


**Figure 2: An example of a Bayesian Network graphical model in the case of fitness wearable devices.**

trade-off between both preserving *privacy* and the *utility* of the third-party service is gained.

# 4. A CASE EXAMPLE WITH FITNESS DEVICES

In this section we present a case example from fitness devices. By simulating sensed and shared data from common fitness trackers (e.g., Fitbit, Jawbone and Bioharness), we computed the inference probability of a set of unknown personal data.

To compute the inference probability (inference risk), we built a Bayesian network (i.e., a graphical probabilistic model which represents dependencies among nodes) as shown in Figure 2. Then, we simulated the availability of some data (i.e. evidences in parent nodes) and through propagation, we obtained the posterior probability of child nodes.

Parent nodes on the left side of Figure 2 (green nodes) represent personal user data (i.e., sensor data and user attributes) required in the hypothetical third-party Statement (e.g., Fitbit, Jawbone and Bioharness Statements). Arcs represent dependencies among these nodes and the child nodes on the right (blue and grey nodes). *BMR*, *Activity*, and *Calories Burned* (grey nodes) are user attributes which are computed by the third party as part of its service (to be exact, the grey nodes in the figure are from Fitbit) and are not considered for inference computation since they are computed by Fitbit and this is known by the users in this particular example.

## 4.1 Bayesian network initialization

Bayesian Networks represent a set of variables and their conditional dependencies via a directed acyclic graph. They are built on the Bayes theorem. With reference to Matrix $I(x, y)$, Bayesian Network is used as follows:

$$P\left(a_j|c_i\right) = \frac{P\left(c_i|a_j\right) \cdot P\left(a_J\right)}{P\left(c_i\right)} \tag{2}$$

where:

- $a_j$: is one of the inferable child nodes (blue nodes)
- $c_i = \{b_1, b_2, ..., b_l\}$: is a combination of parent nodes $b$ (combination of green nodes), with $l$: the number of incoming arcs from $c_i$ to $a_j$
- $a, b \in D$

### 4.1.1 Prior Probabilities

As stated, the objective of the network in this study is to compute the probability that a third party may infer $a_j$ (blue node) if it knows $c_i = \{b_1, b_2, ..., b_l\}$ (green nodes). To initialize the network, the prior probability that the third party knows each $b_i$ is necessary. It represents the chance for the third party to know $b_i$. In our scenario, which assumes that the third party has no knowledge about the user when the flow starts, the prior probability $P(b_i)$ depends on the willingness of the user to release this data. $P(b_i)$ values ranges from 0 to 1 in increasing order proportional to the user's willingness to share this information. That is, $P(b_i) = 1$ means that the user is 100% willing to share this information, $P(b_i) = 0$ means the user never wants to share this information, and between these values are the user's perception towards the uncertainty on her/his willingness.

Accordingly, the prior probabilities can be computed starting from the privacy settings in the user profile managed and stored by PDM. We call this measure "*disclosure Propensity Index*" (PI). It is computed similarly to [6] with a mixed user and system-driven approach:

i) *User-driven*: since it exploits the privacy settings provided by the user for predefined categories of personal data (specific for each PDM),

ii) *System-driven*: since computations are performed to estimate PI from the privacy settings in the user profile. In fact, personal data that have to be rated by the user have generally a granularity lower than the nodes in the Bayesian Network (in order to be user friendly).

Thus, for PI computation, a mapping schema is exploited first to associate user profile privacy settings to the network nodes. Then, privacy settings in the user profile are converted into PI values normalized to [0 1]. PI values of profile items that correspond to parent green nodes can be directly used as prior probability P($b_i$). For $b_i$ nodes that do not have a computed PI value, prior probability can be estimated using PI values of their children in the network by performing a weighted average as described in the following equation:

$$P\left(b_i\right) = \frac{1}{n}\sum_{j=1}^{n} w_j PI(a_j) \tag{3}$$

where:

- $PI(a_j)$: is PI computed for node $a_j$

- $w_j$: is the weight that represents the strength of the correlation between $b_i$ and $a_j$; $w_j \in [0, 1]$
- $n$: is the number of the child nodes of $b_i$.

For instance, if a user has a computed PI for his *Location* information in Figure 2, this value should be inherited by *Accelerometer*, *WiFi*, *Gyroscope*, *GPS* and *GSM*. However, if there is also a PI value for *Transport mode*, which is as well inherited from *Accelerometer*, *WiFi*, *Gyroscope*, *GPS*, *GSM*, the prior probability of the GPS, for example, is given by the weighted average of *Location* PI and *Transport mode* PI.

#### 4.1.2 Conditional probabilities

In addition to prior probability, Bayesian Networks require input conditional probabilities on child nodes in order to be initialized. In this study we considered the following child nodes whose correlations with parent nodes have been reported in the literature (references in Table 1): *Transport Mode*, *Location*, *On-Screen Taps*, *Smoking*, and *Heart Disease* (blue nodes in Figure 2). We have normalized these correlations and have used them to specify the conditional probabilities of the child nodes.

Bayesian networks showed to be very robust to missing data and thus are useful when the knowledge about some nodes is not precise, as in our case [17].

Below, we provide brief descriptions of the child nodes and of the correlations used to specify the conditional probabilities. In the table we report only the highest correlation value among the set of personal data combinations in Figure 2.

*Transport mode*: identifies whether the user is being stationary, walking, running, biking or using motorized transport [24]. Using a Hidden Markov Model classifier, the researchers were able to identify the transport mode with at least 93.6% of accuracy using GPS and Accelerometer. They also provided the inference accuracy based on the combinations of further sensors: *GSM*, *WiFi* and *Gyroscope*. The classifier was designed to work on mobile phone and thus, there are different sensing modalities on which combination and classification accuracy could increase with respect of the basic modality represented by *GPS* and *Accelerometer*.

*Location*: identifies where the user is (e.g., hospital, home, office, etc.). It may be inferred by different combinations of sensors with different accuracy [3, 18, 21]. The highest accuracy is obtained by using the accelerometer information, with 200m radius of accuracy by using a trajectory inference model [14].

*Smoking*: detects if the user is a smoker or not. It is detected with a gesture recognition algorithm by processing data from wristband detecting arm trajectory with an accuracy of 95.7% [22]. This user attribute has also been inferred by exploiting *ECG*, *respiration* and *accelerometer* data with different accuracy values [26].

*Heart Disease*: (or, specifically, prediction of stroke) has been studied in relation to several factors - age, systolic blood pressure, use of antihypertensive therapy, diabetes mellitus, cigarette smoking, prior cardiovascular disease, atrial fibrillation, left ventricular hypertrophy by electrocardiogram - and identified with different correlation coefficients [31]. For this study, we use only *Smoking*, *gender*, *age* as the parent nodes since these are available in the fitness trackers.

*Activity*: can be automatically detected by the tracker, which is set or manually logged by the user after the activity. It can be calculated from, at least, the three-axis accelerometer that provides motion patterns and the energy expenditure of activity expressed by Metabolic Equivalents (METs). We assume that the parent node should only be accelerometer since it is not crucial in this example and the major focus is on the inference of the blue nodes.

*Basal Metabolical Rate (BMR)*: the value is calculated by the third-party using its equations. For example, Fitbit uses the Mifflin-St Jeor equation, which involves age, height, weight and gender [12].

*Calories Burned*: is estimated by the BMR, activity tracker and activity record manually.

For *BMR*, *Activity*, and *Calories Burned* (shown in gray nodes), we assume a 100% inference probability since they can always be computed by Fitbit given the required parent nodes.

It is worth noting that the significance of correlations among nodes and the related probability levels depend on the sample size used in each specific study that we used as references. For initializing our network, when the literature did not provide enough details, we estimated this value with the available data.

**Table 1: Simulation: PI estimation for a given user, inference probability, References in the literature.**

| **Inferred Attributes** | PI | Inference Probability | **Cond. Prob. References** |
|---|---|---|---|
| Transport mode | 0.9 | 95% | [24] |
| Location | 0.3 | 97% | [3, 18, 21] |
| On-Screen Taps | 0.4 | 80% | [3, 18, 21] |
| Smoking | 0.3 | 90% | [26, 22] |
| Heart Disease | 0.1 | 68% | [31] |

### 4.2 Inference computation

The computation of the posterior probabilities are defined in this subsection. The main steps from the user settings to the inference computation are briefly explained below.

#### 4.2.1 User Privacy Settings

User Privacy settings are managed by the PDMs. However, the task of privacy preferences and threshold estimations are managed by AID-S.

In this case study, we assume that the PDM has acquired the user preference settings and then shares them with AID-S as needed for inference computation. For example, PDM provides AID-S with the following user privacy settings. User Setting =[$Transport\ Mode = 4.5$, $Location = 1.5$, $On-Screen\ Taps = 2$, $Smoking = 1.5$, $Heart\ Disease = 0.5$] in a 0 to 5 range.

#### 4.2.2 PI Computation

AID-S computes the PI values as described in Section 4.1.1. In this example, the network nodes in AID-S are the same as the PDM user setting and thus AID-S simply normalizes the data $\in$ [0 1]. In other cases some mappings are required as mentioned above. The obtained values will be the PI for the blue nodes as indicated in Table 1. To estimate the prior probabilities of the green nodes based on blue nodes, Eq. 3 is used with an assumption of equal weights.

#### 4.2.3 Belief Propagation

Once the Bayesian Network is initialized, it can be used to estimate the inference risks for a given set of evidences. In

our example, evidences correspond to the personal data required by the third party in the Statement. In other terms, data that are available with certainty=1 are named evidences. Thus, we have an evidence when $P(b_i) = 1$. Given some evidences, the Bayesian Network operates by propagating beliefs throughout the network.

For example, simulating the case of a Statement which requires the user to share all data of parent green nodes except two nodes, e.g., $GSM$ and $WiFi$, we have ($P(b_i) = 1$) for all the required data. The result of the simulation shows that all the child nodes have updated probabilities which represent the risk that the third party may infer them, exploiting the correlations among the requested data. The inference probabilities that are returned in this example are shown in Table 1.

One of the advantages of using Bayesian networks is that they optimally works when an inferred node is used to condition another inferred node. For example, given that the third party knows data sensed from ECG, respiratory sensor and accelerometer, it has 0.9 probability to infer whether the user is a smoker or not (i.e., P(smoking=true|ECG=true, respiratory=true, accelerometer=true) = 0.9). Afterwards, given the further evidence about age and gender combined with the smoking probability, it allows a third party to compute the inference probability of heart diseases.

Several algorithms are available to compute inference measures for limited subsets of the Inference Matrix. A working example is from [32], which computes the probability of inferring the user behavior (e.g., walking, running, etc.) and the user's typical paths (e.g., going to coffee shop, grocery, outdoors, etc.) only by exploiting the steps per minute computed from a fitness IoT pedometer. It has been reported that as their threshold value, $\epsilon$ - denoting the Euclidian distance between the two time-series (sense sequence, query sequence), varies, it could infer the user behavior with at least 50% accuracy, thus, $P(a|c_k) = 0.50$.

The evaluation of the mSieve system [26] and its model-based substitution scheme, by using 660 hours of ECG, respiration, location and accelerometer data collected over multiple user studies with over 43 participants, demonstrates that sensitive behavioral inferences, such as onset of stress, smoking, and cocaine use, can be protected while still retaining meaningful utility ($\geq 85\%$ accuracy when privacy sensitivity is high and $\geq 90\%$ on average) of the shared physiological signals in terms of their use for tracking heart rate, breathing irregularities, and detecting conversation episodes.

Other examples are provided in [20]. This patent estimates the information entropy of a non-shared data item and utilizes the information entropy to compute inference probability.

# 5. CONCLUSION AND FUTURE WORKS

Today, having a set of heterogeneous IoT devices for each individual justifies the need for a PDM. Though current designs of PDMs are able to protect data from unauthorized third parties, they are still not sufficient enough to prevent the risk of inference of private information from the permitted data collection of authorized third parties.

The high risk of inferring sensitive information using wearable sensors has been discussed in various works. This paper uses an innovative model based on our framework for inference prevention based on handling privacy statements between the user and a third party. Specifically, this paper describes an Adaptive Inference Discovery Service (AID-S) for the particular example in the domain of fitness devices, as an extension to PDMs in order to assure that private information will not be inferred given the combination of information released to fitness third parties.

We focused in particular on fitness devices, which are gaining popularity in the field of IoT. To implement the reasoner, we used a Bayesian Network, initialized with prior probabilities and conditional probabilities, for parent and child nodes, respectively. Thus, we were able to simulate the posterior probabilities which represent the measure of the risk that a third party may infer a private data item given the set of shared data. This paper aims to show that the leakage of private information from trusted third parties becomes pervasive with the rise of personal IoT devices.

As for future work, we intend to extend the model on different domains, also with other reasoning methods. Concerning the current implementation, next steps are the optimization of weights, $w_i$ and PIs computation from the user profile. We also intend that this paper will be an eye opener for researchers who operate in this field.

# 6. REFERENCES


[1] S. Agrawal, I. Constandache, S. Gaonkar, R. Roy Choudhury, K. Caves, and F. DeRuyter. Using mobile phones to write in air. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, pages 15–28. ACM, 2011.

[2] S. H. Ahmadinejad, P. W. Fong, and R. Safavi-Naini. Privacy and utility of inference control mechanisms for social computing applications. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, pages 829–840. ACM, 2016.

[3] N. Brouwers and M. Woehrle. Detecting dwelling in urban environments using gps, wifi, and geolocation measurements. In *Workshop on Sensing Applications on Mobile Phones (PhoneSense)*, pages 1–5. Citeseer, 2011.

[4] F. Carmagnola, F. Osborne, and I. Torre. Escaping the big brother: An empirical study on factors influencing identification and information leakage on the web. *Journal of Information Science*, 40(2):180–197, 2014.

[5] F. Carmagnola, F. Osborne, and I. Torre. User data discovery and aggregation: the cs-udd algorithm. *Information Sciences*, 270:41–72, 2014.

[6] S. Chakraborty, C. Shen, K. R. Raghavan, Y. Shoukry, M. Millar, and M. Srivastava. ipshield: a framework for enforcing context-aware privacy. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, pages 143–156, 2014.

[7] A. Chaudhry, J. Crowcroft, H. Howard, A. Madhavapeddy, R. Mortier, H. Haddadi, and D. McAuley. Personal data: thinking inside the box. In *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives*, pages 29–32. Aarhus University Press, 2015.

[8] Cisco Systems, Inc. The Zettabyte Era – Trends and Analysis. Technical report, Cisco, 2016.

[9] G. D. Clifford, F. Azuaje, and P. McSharry. *Advanced methods and tools for ECG data analysis*. Artech House, Inc., 2006.

[10] Y. Dong, A. Hoover, J. Scisco, and E. Muth. A new

method for measuring meal intake in humans via automated wrist motion tracking. *Applied psychophysiology and biofeedback*, 37(3):205–215, 2012.

[11] E. Ertin, N. Stohs, S. Kumar, A. Raij, M. al'Absi, and S. Shah. Autosense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, pages 274–287. ACM, 2011.

[12] Fitbit. Fitbit Store, 2016. Last accessed: 25-10-2016.

[13] G. Friedland, G. Maier, R. Sommer, and N. Weaver. Sherlock holmes' evil twin: on the impact of global inference for online privacy. In *Proceedings of the 2011 workshop on New security paradigms workshop*, pages 105–114. ACM, 2011.

[14] J. Han, E. Owusu, L. T. Nguyen, A. Perrig, and J. Zhang. Accomplice: Location inference using accelerometers on smartphones. In *2012 Fourth International Conference on Communication Systems and Networks (COMSNETS 2012)*, pages 1–9. IEEE, 2012.

[15] A. Hilts, C. Parsons, and J. Knockel. Every Step You Fake: A Comparative Analysis of Fitness Tracker Privacy and Security. Technical report, Citizen Lab, University of Toronto, 2016.

[16] Input Consortium. In-Network Programmability for next-generation personal cloUd service supporT, 2015. Last accessed: 12-10-2016.

[17] K. Iqbal, X.-C. Yin, H.-W. Hao, Q. M. Ilyas, and H. Ali. An overview of bayesian network applications in uncertain domains. *International Journal of Computer Theory and Engineering*, 7(6):416, 2015.

[18] D. H. Kim, Y. Kim, D. Estrin, and M. B. Srivastava. Sensloc: sensing everyday places and paths using less energy. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pages 43–56. ACM, 2010.

[19] A. Maiti, M. Jadliwala, J. He, and I. Bilogrevic. (smart) watch your taps: side-channel keystroke inference attacks using smartwatches. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 27–30. ACM, 2015.

[20] S. G. Motahari and Q. Jones. System and method for protecting user privacy using social inference protection techniques, Aug. 6 2013. US Patent 8,504,481.

[21] S. Nirjon, R. F. Dickerson, P. Asare, Q. Li, D. Hong, J. A. Stankovic, P. Hu, G. Shen, and X. Jiang. Auditeur: a mobile-cloud service platform for acoustic event detection on smartphones. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, pages 403–416. ACM, 2013.

[22] A. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, and E. Kalogerakis. Risq: Recognizing smoking gestures with inertial sensors on a wristband. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, pages 149–161. ACM, 2014.

[23] M. Prakash and G. Singaravel. An analysis of privacy risks and design principles for developing countermeasures in privacy preserving sensitive data publishing. *Journal of Theoretical & Applied Information Technology*, 62(1), 2014.

[24] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):13, 2010.

[25] N. Saleheen, A. A. Ali, S. M. Hossain, H. Sarker, S. Chatterjee, B. Marlin, E. Ertin, M. al'Absi, and S. Kumar. puffmarker: A multi-sensor approach for pinpointing the timing of first lapse in smoking cessation. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 999–1010, New York, NY, USA, 2015. ACM.

[26] N. Saleheen, S. Chakraborty, N. Ali, M. M. Rahman, S. M. Hossain, R. Bari, E. Buder, M. Srivastava, and S. Kumar. msieve: differential behavioral privacy in time series of mobile sensor data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 706–717. ACM, 2016.

[27] The Guardian. Samsung's voice-recording smart TVs breach privacy law, campaigners claim, 2015. Last accessed: 12-10-2016.

[28] I. Torre, G. Adorni, F. Koceva, and O. R. Sanchez. Preventing disclosure of personal data in iot networks. In *Proceedings of the IEEE 12th International Conference on Signal-Image Technology &Internet-Based Systems*, pages 389–396, 2016.

[29] I. Torre, F. Koceva, O. R. Sanchez, and G. Adorni. A framework for personal data protection in the iot. In *Proceedings of the 11th International Conference for Internet Technology and Secured Transactions*, 2016.

[30] M. Vescovi, C. Moiso, M. Pasolli, L. Cordin, and F. Antonelli. Building an eco-system of trusted services via user control and transparency on personal data. In *IFIP International Conference on Trust Management*, pages 240–250. Springer, 2015.

[31] P. A. Wolf, R. B. D'Agostino, A. J. Belanger, and W. B. Kannel. Probability of stroke: a risk profile from the framingham study. *Stroke*, 22(3):312–318, 1991.

[32] T. Yan, Y. Lu, and N. Zhang. Privacy disclosure from wearable devices. In *Proceedings of the 2015 Workshop on Privacy-Aware Mobile Computing*, pages 13–18. ACM, 2015.

[33] S. Yun, Y.-C. Chen, and L. Qiu. Turning a mobile device into a mouse in the air. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pages 15–29. ACM, 2015.

[34] G. Zyskind, O. Nathan, et al. Decentralizing privacy: Using blockchain to protect personal data. In *Security and Privacy Workshops (SPW), 2015 IEEE*, pages 180–184. IEEE, 2015.