# Towards Improving Privacy Control for Smart Homes: A Privacy Decision Framework (Hacia la Mejora del Control de Privacidad para Hogares Inteligentes: Un Marco de Decisión de Privacidad)

**Abstracto:**

En los hogares inteligentes, los usuarios no tienen suficientes opciones para expresar sus preferencias de privacidad y decidir quién puede ver sus datos, cuándo se pueden usar, y qué parte de los datos quieren compartir y qué parte no quieren. Proponemos un marco que utiliza algoritmos de aprendizaje automático para determinar con qué propósito, con quién y con qué nivel de detalles se compartirá la información.

**Publicado en:**[16a Conferencia Anual de 2018 sobre Privacidad, Seguridad y Confianza (PST)](https://ieeexplore.ieee.org/xpl/conhome/8498146/proceeding)

**Fecha de la conferencia:** 28-30 de agosto de 2018

**Fecha de adición a IEEE *Xplore* :** 01 de noviembre de 2018

**Información del ISBN:**

**Número de acceso de INSPEC:** 18201346

**DOI:**[10.1109 / PST.2018.8514198](https://doi.org/10.1109/PST.2018.8514198)

**Editorial:**IEEE

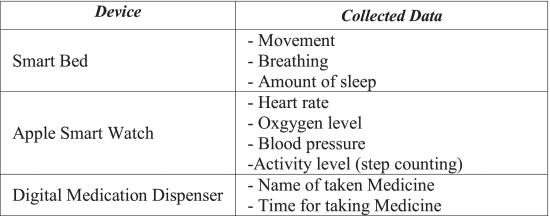
**Lugar de la conferencia:** Belfast, Reino Unido

**SECCIÓN I.**

## **Introducción**

Los hogares inteligentes ofrecen utilidad, economía y comodidad [1] a través de dispositivos inteligentes. La tecnología de Internet de las cosas (IoT) proporciona conectividad entre dispositivos inteligentes en hogares inteligentes [2]. En este documento estudiamos posibles problemas de privacidad en el entorno del hogar inteligente y la falta de control de los usuarios sobre su privacidad. El alcance de nuestro trabajo es desarrollar un marco de decisión de privacidad que ayudará a los usuarios a expresar sus preferencias de manera fácil y adecuada y elegir la privacidad deseada en consecuencia. En este marco, se aplicarán técnicas de aprendizaje automático para clasificar la información en una categoría sensible y no sensible. Con la clasificación de datos, la configuración de privacidad del usuario se puede establecer correctamente. En este documento, proporcionamos un escenario de caso de uso de vida independiente. Una encuesta reciente [3]muestra que el 95% de las personas mayores de 75 años quieren vivir solo en sus propios hogares. Vivir independientemente en la propia casa mejora la salud mental, la memoria y ayuda al bienestar emocional y físico. Deben tomarse las medidas apropiadas si estos ancianos muestran signos anormales en sus comportamientos y actividades, porque estos signos vitales pueden implicar que se han llevado a cabo algunas actividades inusuales. En este escenario, elegimos tres dispositivos inteligentes que son: Smart Bed, Smart Watch y Digital Medication Dispenser. Estos dispositivos recogen información médica de los usuarios, ya que se muestra en la Tabla I. La vida independiente es un escenario de ejemplo para nuestro marco de decisión de privacidad.

**TABLA I** Los dispositivos inteligentes recopilan información médica

[[](https://ieeexplore.ieee.org/mediastore_new/IEEE/content/media/8498146/8514154/8514198/kesha.t1-1019-large.gif)](https://ieeexplore.ieee.org/mediastore_new/IEEE/content/media/8498146/8514154/8514198/kesha.t1-1019-large.gif)

**SECCIÓN I.**

## **Problemas de privacidad en casas inteligentes**

Los dispositivos inteligentes capturan datos sobre los usuarios mediante el uso de diferentes sensores, como la luz, el sonido y la temperatura. Existen diversas tecnologías de comunicación, como la comunicación por línea eléctrica (PLC), ZigBee y el modelo celular [4] para transmitir los datos capturados a través de la red IP al SmartThings Hub para su almacenamiento. Con el permiso del usuario, los datos se pueden entregar para diferentes usos. Debido a los problemas de privacidad en los hogares inteligentes, las personas pueden tener inquietudes sobre la divulgación de su información privada [5], y pueden dudar en usar los dispositivos inteligentes y vivir en hogares inteligentes.

Los dispositivos inteligentes interactúan entre sí y muestran y analizan los datos que reciben para intercambiar información con los usuarios o proporcionar servicios cómodos [6]. Un problema es que los usuarios no tienen diversas opciones de configuración u almacenamiento para los dispositivos inteligentes en el hogar inteligente. Sin embargo, los procesos de configuración Zero-Conf se emplean en hogares inteligentes, lo que obliga a los usuarios a aceptar la configuración predeterminada del proveedor para seguridad y privacidad. En este caso, los usuarios no tienen muchas oportunidades para configurar sus propios dispositivos. Además, diferentes usuarios (p. Ej., Privacidad no preocupada, pragmatista de la privacidad y fundamentalista de la privacidad) tienen diferentes expectativas de privacidad [7]. Si no se considera la expectativa de privacidad de un usuario durante la recopilación de datos, se puede divulgar información de identificación personal (PII). PII es cualquier dato que pueda divulgar cierta información sobre un usuario que ayude a identificar a un individuo específico [8].

Necesitamos configuraciones de privacidad que ayuden a los usuarios a comprender qué información han capturado los dispositivos y a los usuarios a controlar el intercambio de información. Existen algunas soluciones para los problemas de privacidad en los hogares inteligentes, como configurar su configuración de privacidad por sí mismos, pero con esta solución existían algunos desafíos. En esta solución, los usuarios deben etiquetar toda la recopilación de datos e identificar qué información es sensible para ellos y cuál no lo es. Sin embargo, la precisión de la configuración de privacidad depende de la entrada del usuario y todas las cargas recaen en los usuarios. La entrada del usuario es impredecible y puede dejar de etiquetar los datos en cualquier momento. Otra solución [9]se basa en el aprendizaje activo que los usuarios etiquetan algunos datos seleccionados, y un clasificador de aprendizaje automático etiqueta otros datos en función de los datos etiquetados por el usuario. La limitación del enfoque en [9] es que los usuarios no eligen con quién compartir una información en particular. Por ejemplo, los usuarios pueden preferir no compartir su ubicación con una compañía eléctrica, por lo que deben etiquetar los datos de ubicación como datos confidenciales; sin embargo, es posible que quieran compartirlo con sus amigos, pero en este caso, no tienen opciones para elegir qué dispositivos recopilan qué datos, quién puede acceder a esta información y con qué fines.

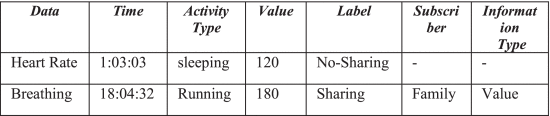
**SECCION II.**

## **Solución**

Para proporcionar a los usuarios un control de privacidad adecuado, presentamos un nuevo marco para el control de privacidad en hogares inteligentes. En este marco, los usuarios pueden etiquetar solo algunos datos seleccionados y, según las preferencias del usuario, el aprendizaje automático puede clasificar el resto de los datos recopilados por los dispositivos inteligentes. Con este marco de decisión de privacidad, esperamos resolver el problema de privacidad del usuario al tener: (1) más opciones para que los usuarios expresen sus preferencias, (2) abordar la falta de conocimiento técnico de los usuarios de dispositivos domésticos inteligentes porque no necesitan tener conocimiento sobre el aspecto técnico de los dispositivos inteligentes.

SmartThings Hub, que ofrece una forma simplificada de controlar los dispositivos inteligentes conectados, almacena todos los datos recopilados de los dispositivos inteligentes. En este marco, podemos usar toda la información del usuario de SmartThings Hub. Todos los datos recopilados se basan en la información que capturaron de los dispositivos inteligentes clasificados en diferentes categorías, como información médica, información geográfica e información de características personales. Al clasificar esta información en diferentes grupos, los datos recopilados estarán más organizados y los usuarios pueden etiquetar los datos más fácilmente. Los usuarios pueden expresar sus preferencias identificando los valores de propósito, información y suscriptor (ver Tabla II ).

**TABLA II** Un conjunto de datos de ejemplo de suscriptor e información para fines médicos

[[](https://ieeexplore.ieee.org/mediastore_new/IEEE/content/media/8498146/8514154/8514198/kesha.t2-1019-large.gif)](https://ieeexplore.ieee.org/mediastore_new/IEEE/content/media/8498146/8514154/8514198/kesha.t2-1019-large.gif)

Propósito (por qué se debe compartir la información): los datos recopilados se pueden usar para fines generales, como los datos de salud se pueden usar para investigación médica o para fines más específicos de un usuario, como una compañía eléctrica que usa los datos del usuario para mejorar la disponibilidad de su servicio para el usuario. Por lo tanto, el propósito del intercambio de información puede afectar su decisión de compartir datos.

Suscriptor (a quien se debe compartir la información): muchos suscriptores como la compañía eléctrica, la compañía de agua y otros usuarios como los amigos del usuario solicitan datos recopilados de hogares inteligentes. En nuestro marco, los usuarios tienen esta opción para especificar quién puede acceder a su información y a quién no se le permite.

Información (qué datos confidenciales se deben compartir): los usuarios también tienen la opción de decidir sobre la granularidad del intercambio de datos.

El clasificador de aprendizaje automático ayudará al proceso de determinar compartir o no compartir la información del usuario. Planeamos emplear el aprendizaje activo basado en grupos [10] para la clasificación de datos por dos razones: (1) puede ayudar a minimizar la carga de etiquetar tantas instancias (2) los usuarios tienen acceso a un grupo de datos no etiquetados y pueden etiquetar una cierta cantidad de datos. Al usar este marco, los usuarios pueden especificar sus preferencias para un cierto número de instancias sin etiquetar y, en función de sus respuestas, el modelo aprenderá las preferencias del usuario y etiquetará el resto de los datos sin etiquetar. Un algoritmo que realiza el aprendizaje activo basado en grupos es Naïve Bayes [11]. El clasificador Naïve Bayes puede incorporar conocimiento previo y datos con valores faltantes y proporciona un modelo interpretable. Sin embargo, no realiza métodos discriminatorios. El otro algoritmo que se puede usar para el aprendizaje activo es Support Vector Machines (SVM). SVM tiene un éxito significativo en las tareas de aprendizaje del mundo real debido a diferentes razones, como la reducción de la necesidad de etiquetar instancias y tener un buen rendimiento en datos no vistos. Cuando los usuarios etiquetan algunos de los datos antiguos que recopilamos de SmartThings Hub, podemos emplear una clasificación de aprendizaje activa para etiquetar el resto de los datos. Después de que los usuarios expresen sus preferencias sobre compartir su información para diferentes propósitos (en este caso de uso, con fines médicos),

**SECCION IV.**

## **Discusión**

Con este marco, los usuarios pueden expresar sus preferencias y elegir conscientemente la privacidad deseada. Según las preferencias, los datos de capacitación se prepararán para la clasificación de sensibilidad de datos. Después de que el clasificador de aprendizaje automático aprenda las preferencias del usuario, será más fácil mantener la privacidad del usuario. Una ventaja significativa del uso del aprendizaje automático es que si los usuarios agregan un nuevo dispositivo inteligente que recopila nuevos datos, el aprendizaje automático puede adoptar y clasificar los datos en una categoría correcta.