Department: Head
Editor: Oliver Amft, amft@computer.org
Editor: Kristof Van Laerhoven, kvl@eti.uni-siegen.de

Privacy Risk Awareness in Wearables and the Internet of Things

Ismini Psychoula
De Montfort University

Liming Chen
Ulster University

Oliver Amft
FAU Erlangen-Nürnberg

***Abstract*—Day to day interactions with wearable and pervasive systems lead to collected data that capture various aspects of human behavior and enable machine learning algorithms to extract extensive information about users. We discuss privacy risk awareness, and ways to preserve privacy and integrate it in current frameworks.**

PERVASIVE COMPUTING has penetrated every aspect of everyday life from smart wearable devices to smart homes and cities. Smart wearables are already becoming a daily necessity as people rely on them to meet their needs and goals. Users can continuously track their physiological parameters like heart rate, calories, quality of sleep, and daily activities such as water intake. It is now common for users to make payments using their smartwatches and smartphones, or give commands to voice assistants for tasks they need to perform. The data collected from these services are used to allow service providers to better understand their users, improve the quality of services and offer personalization. For example, Transport for London used Wi-Fi access points to track mobile phones of commuters and, based on the data, analyze journeys in the transportation network. The tracking initiative helped to determine busy routes, congestion forecasts, and alternative travel recommendations, which ultimately increase efficiency of the transportation service [1]. Singapore is another nation that is leveraging wearable and IoT technologies to offer smart healthcare to their citizens. To combat the spread of COVID-19, Singapore has introduced 'TraceTogether,' a mobile application for community-driven contact-tracing [2]. Contacts are monitored by exchanging identifiers via Bluetooth handshake between two smartphones running the app.

 Published by the IEEE Computer Society

The identifiers are shared with a health authority if a user gets infected with COVID-19 to inform potentially infected individuals [3].

The above examples highlight that sensors, devices, and data are transforming societies and offer benefits from science to services, including transportation, safety, healthcare. However, the same data that offer so many benefits are increasing security and privacy concerns because data analytics can disclose personal data, including behavior, preferences, and health state. For example, researchers recently found that 19 out of the 24 medical apps they sampled leak data with 55 entities, including third and fourth parties [4]. Lately, the list of emerging data leaks and privacy risks have been growing as evidenced by the constant privacy breaches reported in the news [5].

PREDICTIVE AND QUANTITATIVE PRIVACY HARMS

Governing bodies are forced by the increasing public awareness to advance regulations and controls on new technologies. For instance, operating systems and web browsers are adding user controls, blocking cookies and setting new rules on tracking [6]. Despite these efforts, as technology and application advance so do the predictive and quantitative privacy harms and risks that emerge. Predictive privacy harms refer to inappropriate inclusion of, or predictive analysis, from an individual's personal data without their knowledge or expressed consent. The issue with predictive harms is that the continuous collection and analysis of detailed personal data can be used to re-identify and profile users to predict their behavior [7]. However, ubiquitous data such as those collected for wearable and IoT devices are not recognized as personal data in all of the existing directives. Quantitative privacy harms refer to threats posed by modern technologies relating to how often and to what extend private data are collected. The harms are brought on in cases where there is excessive and ubiquitous data collection [8]. Users expect the majority of their data will remain private. For example, users of an arbitrary smartphone app might not care about sharing their location once, but they will expect that they are not tracked throughout a whole day. Wearable and IoT devices are at particularly high risk of creating harm because they enable services to gather personal data (e.g. behaviour, health) that does retain economical and social value even tens of years after it was acquired.

The main privacy challenges in wearable and IoT ecosystems include obtaining consent for data collection, allowing users to to meaningfully and intelligibly control and choose the data they share, meanwhile ensuring the use of collected data is limited to the stated purpose [9]. The potential for misuse arises from the pervasive tracking of habits, behaviors, and locations through sensors and user interaction over a long period. There are new risks to personal safety as a result of the prevalence of wearable and IoT systems [10]. For example, mapping a user's running route from fitness tracking data can expose the user's home address as it is usually the starting and endpoint of the map. By combining physical activity and heart rate measurements, fitness level and trend, as well as diseases are identifiable [11, 12].

Here, we discuss privacy awareness and requirements that should be considered to integrate privacy into wearable and IoT services. We describe a privacy risk-aware framework that works like a data firewall around the personal wearable and IoT device ecosystem. Our system can guide technology developers and service providers to incorporate privacy risk awareness, design principles and mitigation methods in the life-cycle of service design and development, thus enabling privacy-preserving data management.

PRIVACY AWARENESS AND PREFERENCES

In order to design privacy-aware systems and services, understanding a user's preferences and concerns is of great importance. Factors such as social relationships, transparency of the mechanisms, context and who is collecting the data are important determinants. Different studies have found that individuals thought monitoring in personal spaces and by an unknown entity or the government was unacceptable. Another finding was that photo and video monitoring cause privacy concerns regardless of the context [13].

Although information privacy, data sharing and control are universal issues, the precise concerns and responses to data sharing requests depend on the trustworthiness of the wearable and

IoT services and user characteristics, including the user's culture and age [14]. The extent of these concerns depend on factors such as the type of data collected, retention time, purpose of data collection, trust in the service provider and perceived value of the data collected [15]. In regards to individuals' privacy preferences for wearable sensors, studies show that users want to have control of the data they produce. Especially for people who are aware of the risks and have privacy concerns, data control has a significant influence on their acceptance of technology [16].

When sharing personal data, there are various competing factors like those discussed previously that are involved in each decision, such as motivations, purposes, and personal preferences. There are situations and contexts under which the users are concerned and reluctant to share data, while in other situations not as much. So to address privacy issues, regulations and privacy-aware mechanisms, user preferences and perceived risks are important factors for technology acceptance.

PERSONALIZING PRIVACY

We need frameworks that will allow users to personalize their privacy based on their use and privacy expectations. Also, to raise awareness, it would be useful to develop tools that will help the users select what type of data to provide in each service or apply data transformations and privacy-enhancing techniques before sharing them.

We gathered the following key requirements for the design of privacy risk-aware frameworks around wearable and IoT ecosystems:

Privacy Risk Awareness: There is a need for methods that will provide continuous privacy risk assessments. In user-centered services, this could be achieved by calculating the privacy risk of a user based on the accumulated knowledge that exists about him from the data collected, based on the services and devices he currently uses. The privacy risk calculations should also include the potential insights that could be gained from combinations of separately shared data items. The challenge is to evaluate privacy in terms of risk-benefit trade-offs that will enable the users to make informed decisions about data sharing. Also, there is a need for user-friendly tools and interfaces that explain risks and benefits, and advise the user.

Data Control: Users should be able to share the data collected by their wearable and IoT devices with the third parties they prefer and negotiate the conditions for the sharing. Instead of asking for decisions at each sharing occasion, sharing should take place according to user preferences set in advance. The user preferences will specify the conditions under which data will be shared, such as type of data (e.g., photos, sensor readings), retention time, granularity and volume of the data.

Data Transformation: Privacy-preserving mechanisms that will transform the data according to their privacy risk, user preferences, and the data-sharing agreements are another important requirement. For instance, current anonymity [17] and differential privacy [18] techniques could be modified to include individuals' privacy risk.

Privacy and Security: Through the life-cycle of the data from collection to storage and analysis data need to be kept secure and applications should preserve the privacy of individuals by keeping data safe from observation, copying and malicious adversaries.

A PRIVACY RISK AWARE FRAMEWORK

Our privacy risk-aware framework to address the above requirements is based on a Trusted Privacy Mediator. The Trusted Privacy Mediator can be thought of as a private cloudlet with data storage located in the user's domain (i.e., a computer or smartphone). The framework attempts to enable users to make privacy-aware decisions about sharing personal data and to automatically transform data according to their sensitivity and user preferences. We consider an example of providing care within a smart home (Figure 1). Here, data owners are smart home users that sign up to services. Their smart home setup might include wearable and IoT health monitoring devices, and actuators that remotely control appliances. Service providers collect and store user data and create machine learning detection/prediction models either for their own use or for data consumers, who are care providers that use the services to gain insights, monitor patients and offer remote diagnosis.

1. Initial Set Up

The Trusted Privacy Mediator acts as a proxy

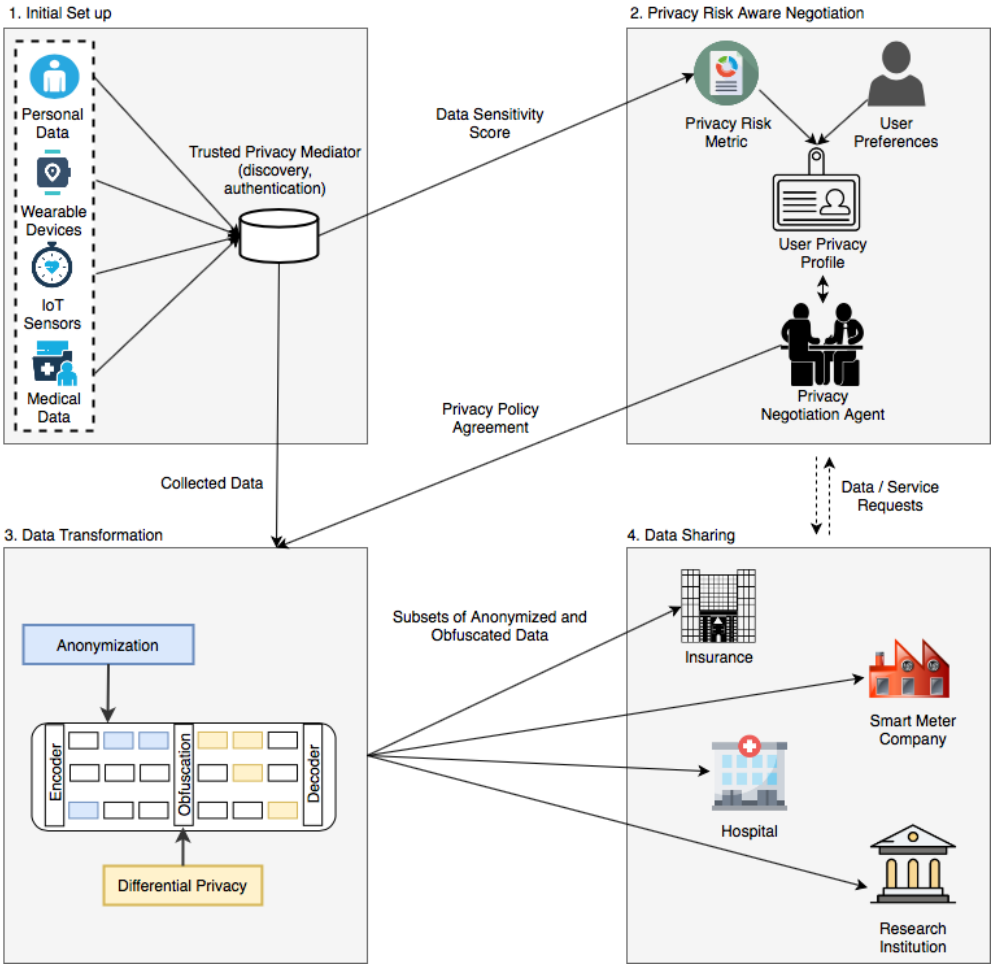


Figure 1. Components and phases of the Privacy Risk Aware Framework. (1) The Trusted Privacy Mediator is installed on the user's device, where it discovers and authenticates devices, and assigns sensitivity weights to the collected data. (2) The privacy risk metric component receives the sensitivity weights and derives the privacy risk of the user. The privacy risk is then combined with user preferences to create a user privacy profile. In turn, the privacy profile is used as a basis for the negotiation and creation of custom privacy policy agreements. (3) The custom privacy policies are sent to the data transformation component that anonymizes and obfuscates the data according to the policy. The data transformation component is based on an adaptation of a deep learning autoencoder model that uses one encoder with multiple decoders to create the transformed subsets. (4) Subsets of transformed data are shared to the corresponding service provider.

between the user's personal data and the service providers. The Trusted Privacy Mediator discovers and authenticates the wearable and IoT devices owned by the user and collects their data. For instance, the Trusted Privacy Mediator will attempt to connect to the smart blood pressure cuff, smartwatch, smart thermostat and other devices in the local networks.

2. Privacy Risk Aware Negotiation

The Privacy Negotiation Agent offers fine-grained control over the data sharing, aligned with the user's preference. At the same time, services booked by the user need to get supplied with sufficient data to function. Custom privacy policy agreements are the basis for data sharing. The initial privacy preferences are set up by users via a mobile app or web portal. As setting up fine-grained privacy preferences can be tedious, public templates and recommendations can be deployed (more advanced versions of the framework could

also implement agents that are able to learn a user's privacy preferences over time).

Subsequently, once the inventory has been established, a Privacy Risk metric is used to calculate the privacy risk of the user based on the sensitivity of the data items and the insights that could be gained from combinations with previously shared data. The User Privacy Profile is utilized by the Privacy Negotiation Agent to negotiate customized privacy policy agreements with service providers on behalf of the user and offer recommendations on data sharing. Each service that requests data will share its privacy policy that specifies what type of data it expects and at what granularity, as well as how long the data will be stored. The policy might also include other conditions related to the use of data, along with potential rewards and benefits for the user. Thus, for each new service or data request, the Privacy Negotiation Agent will aim to find an agreement with the service provider in a way that provides a beneficial privacy risk-trade off for the user and is in line with the user's privacy preferences.

3. Data Transformation

Depending on the privacy policy agreement and the type of service provider (e.g., trusted or not), the Trusted Privacy Mediator will apply different privacy-enhancing mechanisms to the data. Key concerns are to determine the appropriate obfuscation level and dealing with situations where privacy risks are high but the user is willing to share the data. One option for trusted care providers is that the Trusted Privacy Mediator shares an anonymized data subset, where data irrelevant to the service are excluded. For non-trusted service providers, stricter data transformation includes adding random noise, i.e., to provide differential privacy guarantees.

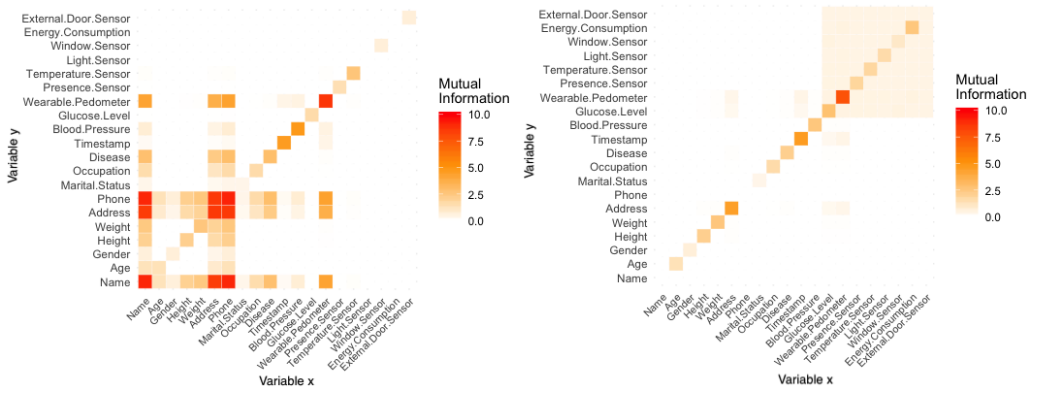
4. Data Sharing A user requests a service, then the service providers send their data requests and data agreement offers to the user. Each service provider might propose numerous offers and services. After agreeing with the service provider on the privacy policy that satisfies both the service provider and the user, the Trusted Privacy Mediator component ensures that the data are appropriately handled by applying data transformation techniques to any sensitive information that the service provider should not receive. At the end of the process, the appropriate de-sensitized data are shared to the service providers.

Following the example of providing care in a smart home, Figure 2 shows the different subsets of data that are shared by the Trusted Privacy Mediator based on the service request (the data transformation and anonymization method with encoders-decoders is described in detail in [19]). In this example, there are two service requests-one to a hospital physician for medical diagnosis, and another to a smart meter company for saving energy costs. To evaluate the data transformations, we used pointwise mutual information as a metric to quantify the information shared between two variables. A $\text{PMI}(x, y) = 0$ indicates that variables x and y are statistically independent.

$$\text{PMI}(x; y) = \log \frac{P(x, y)}{P(x)P(y)}$$

Mutual information reflects the change in the uncertainty of a private variable (i.e., address) due to the observation of a public variable (i.e., energy consumption). It is used to measure the amount of information leaked from a privacy mechanism by comparing the distribution of the original data and that of obfuscated data an adversary has gained access to. Figure 2(a) shows that attributes not relevant for a data consuming hospital physician, such as the variables Energy Consumption and Thermostat state, have been removed by the Trusted Privacy Mediator. On the other hand, data such as Name, Address, Phone Number, Blood Pressure and Wearable Pedometer remain with little to no transformations because they are needed for the service. When sharing data with the smart meter company, Figure 2(b) illustrates that the Trusted Privacy Mediator has anonymized and removed personally identifiable data while sharing obfuscated sensor readings with the service provider.

Advances still have to be made on several aspects of the framework discussed here in order to be able to implement it fully. In particular, methods that are able to understand and learn the user preferences and perform the negotiation to meet their expectations should be investigated further. Another challenge is the development of computational privacy risk metrics that can accurately derive the privacy risk of a user based on the data collected and shared about the user.



(a) Hospital physician (b) Smart meter company

Figure 2. Mutual Information between the original data and the data transformed by the Trusted Privacy Mediator when sharing with (a) hospital physician and (b) smart meter company.

Nevertheless, privacy risk and awareness are increasingly making such endeavors worthwhile.

We are entering a new era where the types and quantities of personal information being collected and shared are growing exponentially. The shifts in the way services harvest, and users generate, data create unlimited possibilities but also add new privacy challenges. As mobile, wearable, and IoT-based services expand, they must incorporate privacy from the conception phase in expectation of harms and provide measures for ongoing assessments.

The privacy landscape will keep changing as users become more aware of privacy risks and demand more privacy protection from service providers. Privacy risk-aware frameworks can change the way data are stored, processed and shared. Potentially, we will move to an even more decentralized way of collecting, managing and sharing data in which each service provider will have to subscribe to a user's individual data rather than the other way around. Therefore, privacy awareness and risk are key to address the privacy protection challenge.

ACKNOWLEDGMENT

This work has been funded by the European Union Horizon 2020 MSCA ITN ACROSSING project (GA no. 676157)

References

[1] Transport for London (TfL), *Review of the tfl wifi pilot*, 2017. [Online]. Available: http://content.tfl.gov.uk/review-tfl-wifi-pilot.pdf.

[2] Government Technology Agency Singapore, *Tracetogether*, 2020. [Online]. Available: https://www.tracetogether.gov.sg.

[3] ——, *Privacy-preserving cross-border contact tracing*, 2020. [Online]. Available: https://bluetrace.io.

[4] Q. Grundy, K. Chiu, F. Held, A. Continella, L. Bero, and R. Holz, "Data sharing practices of medicines related apps and the mobile ecosystem: Traffic, content, and network analysis," *bmj*, vol. 364, p. l920, 2019.

[5] BBC News, *Data breaches*, 2020. [Online]. Available: https://www.bbc.co.uk/news/topics/c0ele42740rt/data-breaches.

[6] "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation)," *OJ*, vol. L 119, pp. 1–8, 2016-05-4.

[7] K. Crawford and J. Schultz, "Big data and due process: Toward a framework to redress predictive privacy harms," *BCL Rev.*, vol. 55, p. 93, 2014.

[8] D. Gray and D. Citron, "The right to quantitative privacy," *Minn. L. Rev.*, vol. 98, p. 62, 2013.

[9] W. Zhou, Y. Jia, A. Peng, Y. Zhang, and P. Liu, "The effect of iot new features on security and privacy: New threats, existing solutions, and challenges yet to be solved," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1606–1616, 2018.

[10] F. T. Commission *et al.*, "Internet of things: Privacy & security in a connected world," *Washington, DC: Federal Trade Commission*, 2015.

[11] M. Altini, P. Casale, J. Penders, and O. Amft, "Cardiorespiratory fitness estimation in free-living using wearable sensors," *Artificial intelligence in medicine*, vol. 68, pp. 37–46, 2016.

[12] J. M. Radin, N. E. Wineinger, E. J. Topol, and S. R. Steinhubl, "Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the usa: A population-based study," *The Lancet Digital Health*, 2020.

[13] H. Lee and A. Kobsa, "Understanding user privacy in internet of things environments," in *Internet of Things (WF-IoT), 2016 IEEE 3rd World Forum on*, IEEE, 2016, pp. 407–412.

[14] D. Singh, I. Psychoula, J. Kropf, S. Hanke, and A. Holzinger, "Users' perceptions and attitudes towards smart home technologies," in *International Conference on Smart Homes and Health Telematics*, Springer, 2018, pp. 203–214.

[15] P. E. Naeini, S. Bhagavatula, H. Habib, M. Degeling, L. Bauer, L. F. Cranor, and N. Sadeh, "Privacy expectations and preferences in an iot world," in *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*, 2017, pp. 399–412.

[16] I. Psychoula, D. Singh, L. Chen, F. Chen, A. Holzinger, and H. Ning, "Users' privacy concerns in iot based applications," in *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, IEEE, 2018, pp. 1887–1894.

[17] L. Sweeney, "K-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[18] C. Dwork, A. Roth, *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[19] I. Psychoula, E. Merdivan, D. Singh, L. Chen, F. Chen, S. Hanke, J. Kropf, A. Holzinger, and M. Geist, "A deep learning approach for privacy preservation in assisted living," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, IEEE, 2018, pp. 710–715.

Ismini Psychoula is a PhD candidate in the School of Computer Science and Informatics at De Montfort University, United Kingdom. Contact her at p16028128@my365.dmu.ac.uk.

Liming Chen is a professor of Data Analytics at Ulster University, United Kingdom. Contact him at l.chen@ulster.ac.uk.

Oliver Amft is a full professor of Digital Health at the Friedrich-Alexander University of Erlangen-Nürnberg, Germany. Contact him at amft@computer.org.