Privacy Everywhere: a Mechanism for Decision Making and Privacy Assurance in IoT Environments

Completed Research

Leandro Prado de Andrade
Federal University of São Carlos
leandro.andrade.nti@gmail.com

Sergio Donizetti Zorzo
Federal University of São Carlos
zorzo@ufscar.br

Abstract

Increase in number of Internet of Things (IoT) devices and the diversity of forms of interaction can make it tiring to ask the user to make a privacy decision whenever it contacts a new IoT device or some device tries to access certain information. Due to sensitivity of many of information captured by IoT devices, it is necessary that data captured by IoT devices be processed before being sent to the cloud services. This work aims to present the Privacy Everywhere mechanism for decision making and guarantees user privacy in IoT environments. In validation of mechanism, accuracy check of the Allow/Deny and Privacy Action neural networks that make up the mechanism presented an accuracy of 88.02% and 86.67%, respectively. According to the results, the Privacy Everywhere mechanism is able to help users preserve the privacy of their data in IoT environments.

Keywords

Internet of Things, IoT privacy, decision making, privacy preferences.

Introduction

Through technological advances, society has moved towards the paradigm "always connected". A growing variety of devices is becoming capable of connecting to a network, capturing information around them, and sending and receiving information. In this context is inserted the Internet of Things which is the latest term to describe the evolutionary trend of devices to become more intelligent, more aware of their context, more able to react to their context and more communicative (Rosner, 2016). Each device is identifiable through its embedded computing system and is capable of interacting within an existing infrastructure.

Security strategies are based on existing vectors of attack, however, with emerging technologies these vectors are changing, leading to increased vulnerability to data exfiltration (Goel et al., 2017). Increase in the number of IoT devices and the possibility of integration of services and information cross-breeding create numerous scenarios and possibilities for collecting user data. Asking user to make privacy decisions every time they contact a new device or some service try to access certain information may be unfeasible or very tiring. Automated privacy decisions are interesting in these types of scenarios.

IoT solutions, in most cases, require an IoT cloud service, where access control, information processing, configuration management of devices, and other functions are done. Cloud computing is transforming and redefining the design and procurement of information technology infrastructure and software (Gupta et al., 2017). Since IoT cloud services are not focused on preserving user privacy, data captured by IoT devices needs to be handled before being sent to IoT cloud services.

This work aims to present the Privacy Everywhere mechanism for privacy assurance and decision making in IoT environments. The target of this mechanism is to make privacy decisions by users in these environments and perform privacy actions on data collected by IoT devices before data is sent to IoT clouds.

The paper is structured as follows: Section 2 presents the related work. Section 3 shows the data gathering with users in order to obtain information to train the neural networks that make up the Privacy Everywhere mechanism. Section 4 describes the Privacy Everywhere mechanism. Section 5 presents the mechanism validation and results. Session 6 shows the discussion and implications of this work. Finally, this paper concludes with conclusion and future works.

Related Work

IoT, big data and online platforms have made security and privacy even more important, such that they are now board-of-director-level and corporate-wide concerns (Lowry, Dinev and Willison, 2017). This importance is due to the sensitivity of the large amount of data generated and processed. From the perspective of privacy calculus, the individual's privacy is interpreted as an exchange where individuals disclose their personal information as long as there are benefits from this exchange (Xu et al., 2009). In the literature, some authors have presented solutions to deal with privacy issues in the data collection of users.

Dinev et al. (2013) proposed and tested an information privacy framework and indicated as relevant to information privacy: anonymity, secrecy, confidentiality and control. The Privacy Everywhere mechanism also considers it important to anonymize the data in the scenarios in which this action is required. The proposed mechanism also provides secrecy and confidentiality of sensitive data, disclosing only the information that users want to disclose.

Henze et al. (2016) introduced an integrated solution for privacy compliance focused on end users and cloud service developers. This solution features privacy compliance points that encrypt data before sending it to the cloud. The issue that arises from this encryption is that not all data types need an encryption action. On this issue, the Privacy Everywhere mechanism acts through a query to the neural network Privacy Action that, according to the scenario in question, will inform the mechanism if it should: not perform any privacy action, notify the user, anonymize the data or encrypt the data.

Jayaraman et al. (2017) proposed an approach where storage of data collected by IoT devices are stored encrypted on servers that have as security requirements: secure connection between the gateway and the IoT server, secure persistence in data storage and control of data access on servers. The proposed Privacy Everywhere mechanism also proposes secure connection and control of data access on servers, but in a simplified way.

Chow (2017) proposed a framework called Privacy Stack to address user privacy issues in IoT environments. The Privacy Stack consists of Awareness, Inference, Preferences and Notification layers. Awareness layer concerns the user's awareness a data collection is being performed. In the Inference layer, a balanced approach and not rely only on the users to understand the possible inferences of the data collected is proposed. Preferences layer refers to individuals' privacy preferences. Finally, in the Notification layer ways of notifying the user are treated and all previous layers must be considered.

The Privacy Stack framework proposed by Chow (2017) encouraged to be added to the Privacy Everywhere mechanism ways of notifying the user. Nevertheless the mechanism proposed in this paper does not notify the user in all situations but only in situations where user notification is considered interesting. A comparison between the works presented in this section and the mechanism proposed in this paper can be visualized in Table 1.

Characteristics	Dinev et al. (2013)	Henze et al. (2016)	Jayaraman et al. (2017)	Chow (2017)	This work
Allows control of the dissemination of collected data about users	✓	✓	✓	✗	✓
Control of data disclosure can be automated	✗	✓	✓	✗	✓
User is notified when personal information is collected	✗	✗	✗	✓	✓

Secure connection and control of data access on servers	✗	✓	✓	✗	✓
Is able to perform a privacy action on collected data	✓	✓	✓	✓	✓
Analysis is performed in data collection scenario before performing a privacy action	✗	✗	✗	✗	✓
Computational cost of the privacy action in collected data is considered	✗	✗	✗	✗	✓

Table 1. Comparison between the works presented in this section

As can be seen in Table 1, the development of the Privacy Everywhere mechanism taken into account characteristics that are not addressed by the related works.

User Privacy Preferences in IoT Environments

To develop the mechanism, study was carried out with 136 undergraduate students of the Federal University of Alfenas in order to understand users' privacy preferences. For definition of sample size, Cochran's formula (1977) was used, highlighting three variables that directly affect the size of a sample: size of population (when known), margin of error and level of confidence.

Sample calculation used in this work consisted of 90% trust rating, 7% error margin and a total population of 3737 individuals. The value obtained by the calculation was 134 individuals, as there are 4 groups of scenarios, 136 individuals were interviewed in order to work with a multiple value of 4.

Students were randomly selected in order to obtain greater heterogeneity in the research. Each participant received a brief introduction on IoT concept and also received a questionnaire with 24 scenarios containing an IoT data collection. In each scenario participants answered whether they allow or deny sending of information and how comfortable they felt with data collection in the scenario in question. The average time spent by students to answer the questionnaire was 14 minutes. Table 2 shows the demographic distribution of these students.

Age	**Participants**	**Percentage (%)**
<21	45	33.1
21 to 25	73	53.7
26 to 30	11	8.1
31 to 35	4	2.9
>35	3	2.2
Sex		
Female	69	50.7
Male	67	49.3

Graduation area	**Participants**	**Percentage (%)**
Biological sciences	59	43.4
Exact sciences	38	27.9
Human sciences	39	28.7
Level of privacy concern		
Very worried	47	34.5
intermediary	84	61.8
unconcerned	5	3.7

Table 2. Participants demographic distribution

Scenarios were built from combination of factors that influence user privacy decision making considered relevant by authors such as Lee and Kobsa (2016) and Naeini et al. (2017). Table 3 presents factors with their respective possible values, used to create the scenarios in this work.

Factors	**Possible values**
Location	Private, not private
Data type	Presence, video, location, audio, personal preferences, personal information
Benefit	User, other
Retention	Forever, Purpose satisfied
Shared	Yes, no

Table 3. Influential factors in privacy decision making and their possible values used to create the scenarios

In this research, after presenting each scenario and asking the respondent to respond whether he allows or denies sending information, was also asked how the interviewee felt about data collection presented. Comfort of participants was measured on a 5-point Likert scale from "Very confortable" to "Very uncomfortable". The statistical representation that shows the relationship between the factors present in scenarios and the level of comfort of participants with collected data can be visualized in Figure 1. For example, 41% of participants felt uncomfortable when the data collection happened in a private place.

As can be seen in Figure 1, users were more uncomfortable when data collection occurred in private places. Regarding the type of data, users were more comfortable when the collection was performed by presence sensors and showed greater discomfort when their personal information was requested. Participants were a bit more comfortable when the data collection was for their benefit. Regarding retention time, there was no significant difference between retention forever and retention until the purpose was satisfied. Participants were also slightly more comfortable in scenarios where the data collected would not be shared.

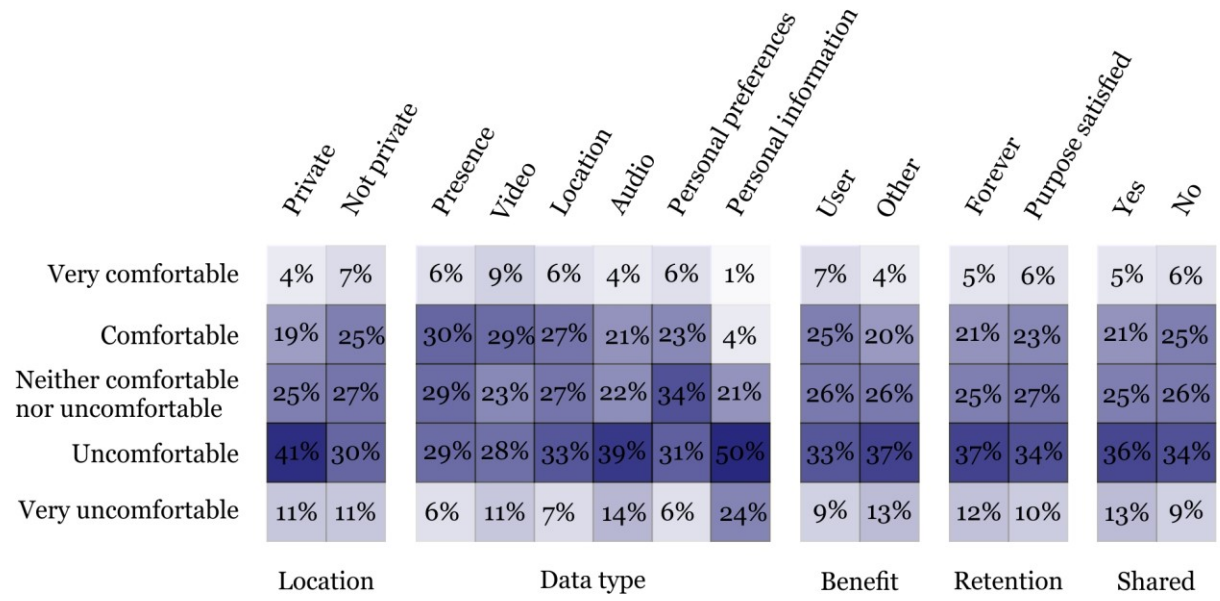

Figure 1. Statistical representation that shows the relationship between the factors present in scenarios and the comfort level of participants. Cells with higher values are darker

After the research with students, the second part of this work was started: the 96 data collection scenarios in IoT environments that were used to carry out the research with the 136 students were presented to five professionals in the area of computer networks and information security. The professional team that participated in this second stage of this work was composed of 2 specialists, 2 masters and 1 PhD.

In each of the scenarios presented to the professionals they answered, in order to preserve the privacy of user with a lower possible computational cost, which is the appropriate privacy action for each scenario in

question. The privacy actions to be carried out on collected data that could be indicated by the professionals were: not to perform any privacy action, notify the user, anonymize the data or encrypt the data.

The answers obtained from the research with the students were used to construct the Allow/Deny neural network training set, while the answers provided by the team of professionals of computer networks and information security were used to construct the training set of Privacy Action neural network. The two neural networks are components of the Privacy Everywhere mechanism presented in this paper.

Privacy Everywhere Mechanism

This section presents the architecture, functionality and operation of the Privacy Everywhere mechanism for user decision-making and privacy actions on data collected by IoT devices.

Privacy Everywhere Neural Networks

The inclusion of two neural networks in the mechanism was necessary so that the mechanism could automate the privacy decision making by users and could also automate the privacy actions to be performed on the data collected by IoT devices. These neural networks are multilayer perceptron (MLP), which are networks composed of several interconnected neurons with weights assigned to the connections. These networks present in the Privacy Everywhere mechanism have the same set of inputs, but differ in the outputs (responses provided). The Allow/Deny neural network presents as possible outputs allow or deny, while the Privacy Action neural network presents as possible outputs: do not perform any privacy action, notify the user, anonymize the data and encrypt the data. Allow/Deny neural network was trained with the training dataset generated from the research with the students carried out in this work. Privacy Action neural network was trained with the training dataset generated from the research with the professionals. The two neural networks that make up the mechanism can be visualized in Figure 2.

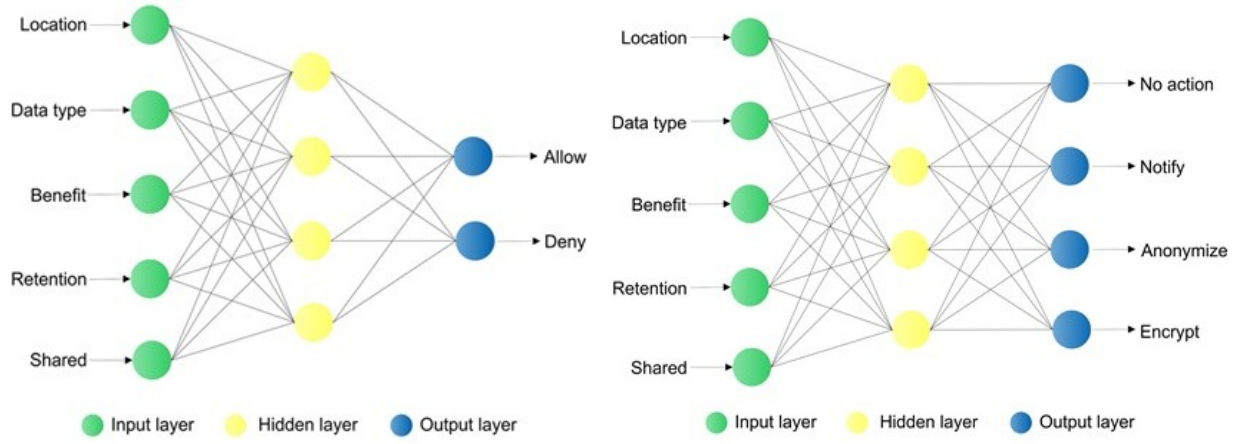

Figure 2. Allow/Deny neural network (left) and Privacy Action neural network (right)

The variables used in the input layer for the training of neural networks are the variables that are used to create the scenarios and can be seen at Table 3.

Privacy Everywhere Arquitecture and Operation

The Privacy Everywhere mechanism is made up of the Allow/Deny and Privacy Action neural networks; by the privacy-IoT-analyzer, privacy-IoT-notify, privacy-IoT-anonymize and privacy-IoT-encrypt modules; and also by a mobile application that is responsible for receiving the notifications sent to the user. The Privacy Everywhere engine has two modes of operation: automatic mode and manual mode. In the automatic mode of operation, privacy decisions and privacy actions are performed by the mechanism.

In the manual mode of operation, the user must respond whether or not to allow the data to be sent and also whether the data (if sent) is to be anonymized or encrypted.

IoT devices collect information that is sent through the privacy guarantee mechanism before it is sent. This mechanism contain a module called pivacy-IoT-analyzer. This module sends in the JavaScript Object Notation (JSON) format the query to the neural network Allow/Deny with the scenario in which the information request is inserted. If the answer of the neural netwok Allow/Deny is to deny sending, the stream ends. On the other hand if the response of the neural network Allow/Deny is to allow the sending, the neural network Privacy Action is then queried. If the response is no privacy action being taken, the information is sent to the IoT cloud to be processed, stored, or passed on to some user request or some IoT device. Performed this procedure, the mechanism action in this request is terminated. If the response is to notify, the flow is directed to the privacy-IoT-notify module. If the answer is anonymize, the flow is directed to the privacy-IoT-anonymize module. And if the response is encrypt, the stream is directed to the privacy-IoT-encrypt module. Different outputs are motivated by the search to make a privacy decision that is sufficient for the scenario in question and that is achieved with the lowest computational cost possible. If computational cost were not considered, the best alternative in most cases would be to anonymize or encrypt the data before sending.

The Figure 3 shows the flow analysis architecture of a data request in the Privacy Everywhere mechanism in automatic operation mode before data is sent.

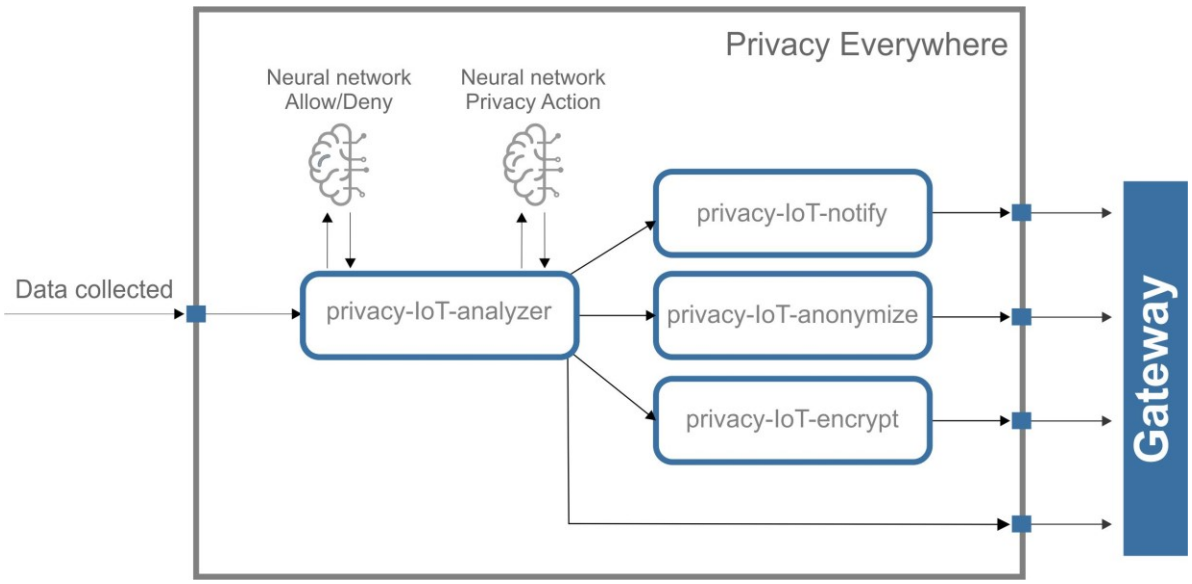

Figure 3. Flow analysis architecture of a data request in the Privacy Everywhere mechanism in automatic operation mode before data is sent

Privacy Everywhere engine makes use of Message Queuing Telemetry Transport (MQTT), which is a protocol for sending and receiving messages, to notify the user. The privacy-IoT-notify module sends a notification through an MQTT topic to the user's Privacy Everywhere mobile application before sending the information to the requestor and terminating the request handling. The privacy-IoT-anonymize module performs the anonymization of the user data contained in request before sending the information to the requestor and terminating the flow. And the privacy-IoT-encrypt module performs data encryption before sending the information to the requester and terminating the stream.

As seen in Figure 4, the Privacy Everywhere mechanism is responsible for making the privacy decisions by user by authorizing or denying the submission, controls the data acquisition, and performs the appropriate privacy action for each scenario in question before the data is sent to the IoT cloud. It remains the responsibility of the IoT cloud to perform data processing and access control.

The request for the information can start from a user or from some IoT device authorized by IoT cloud. Being authorized by the cloud is a primary requirement for requesting data. After proper authorization by

IoT cloud, the request arrives at IoT device. The Privacy Everywhere mechanism acts on the return flow of the data request, allowing or denying the submission and performing appropriate privacy action for the scenario in question before the requested information is returned to requester.

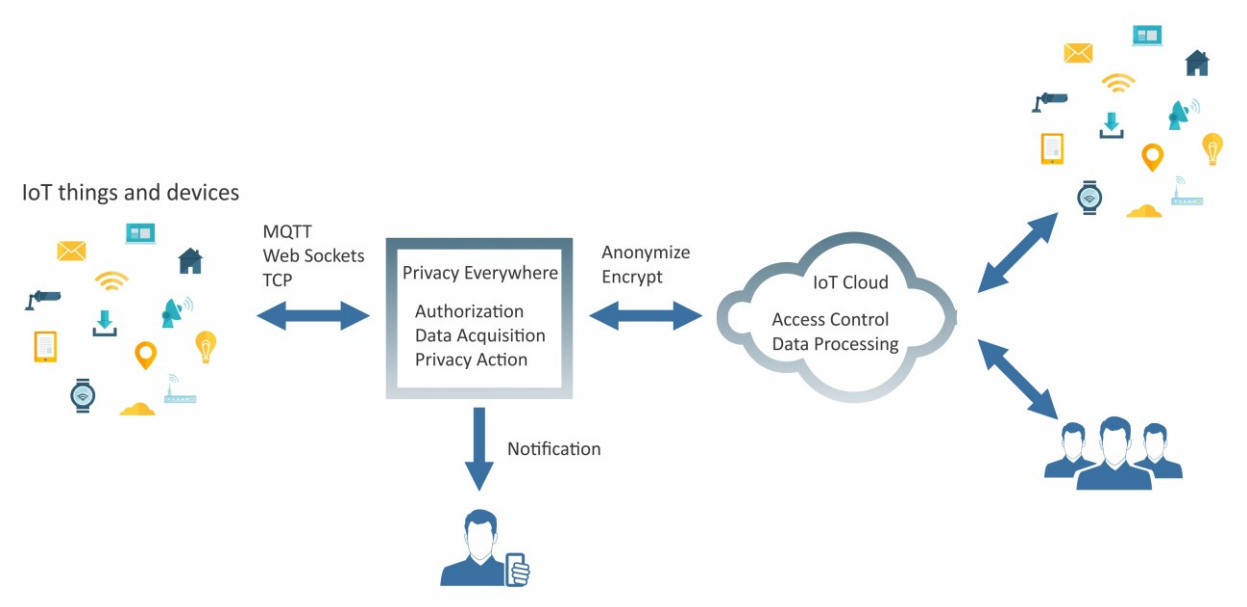

Figure 4. Overview of the Privacy Everywhere mechanism for decision making and privacy assurance in IoT environments

The Privacy Everywhere engine can handle data requests made over the MQTT protocol, Web Sockets, and over the TCP protocol. Notifications are sent to the users' mobile application via the MQTT protocol.

Privacy Everywhere Application Interface

The Privacy Everywhere mobile application is designed for user to receive notifications from IoT systems and devices when the result of data request analysis is to notify the user. This mobile application also allows the user to choose between automatic mode of operation of the mechanism and manual mode of operation. In manual operation mode user must inform upon request whether he allows or denies the sending of the information and what is the desired privacy action (in manual mode the available privacy actions to be performed are anonymize or encrypt). Users can also subscribe to or unsubscribe from the MQTT topics available and simulate the Privacy Everywhere engine behavior in this mobile application.

Figure 5 shows the Privacy Everywhere application interface. In "Decide for me" mode, privacy decisions are automated by the Privacy Everywhere engine, saving the user time and effort. In the "I will decide" mode, the automation of the decision-making is interrupted and the user is then consulted when a data request is made and must decide whether or not to allow the data to be sent and and whether the data needs to be anonymised or encrypted.

In the Privacy Everywhere application the user can also view which places are MQTT topics available (each location has its own MQTT topic), and the user can subscribe or unsubscribe from the topics in order to receive or not receive notifications.

Simulation of the Privacy Everywhere mechanism is done inserting values in the variables used for the training of neural networks: where data collection occurs; what type of data is collected; who benefits from the collection; for how long this data is retained and whether the data is shared or not. After the inputs to simulation were provided, user can observe if the mechanism allows or denies the sending of the data and what is the privacy action done in informed scenario: do not perform any privacy action, notify the user, anonymize the data or encrypt the data.

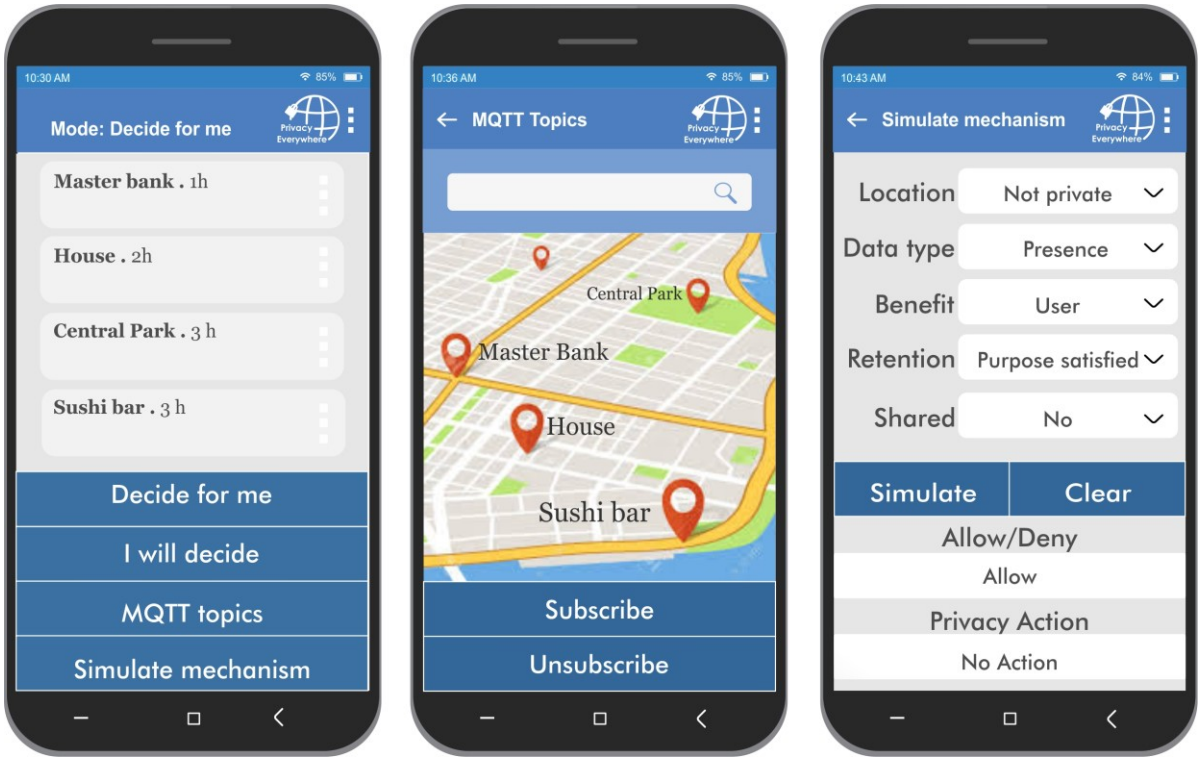

Figure 5. Privacy Everywhere application interface

Mechanism Validation and Results

In data gathering for construction of the mechanism, IoT scenarios were presented to undergraduate students and professionals in computer networks and information security. An example of IoT scenario used in the questionnaires can be seen below:

"The security company SEG+ wants to access the presence sensor of your residence in order to register in their system the most frequent times that their clients are home. The data will not be shared and will be stored until the satisfaction of its purpose."

Students were then asked to respond whether to allow or deny the sent of data for each presented IoT scenario. Professionals were asked what were the appropriate privacy action for each IoT scenario presented. There were 24 scenarios presented to each student and 96 scenarios were presented to each professional. It was generated 3264 student responses and 480 professional responses.

The responses were divided into two datasets: the neural network training dataset of the mechanism and the test dataset. Students responses were divided into 3072 responses (94.12% of responses) for the training dataset of the neural network Allow/Deny and 192 responses (5.88% of responses) to the test dataset. The answers of professionals was divided in 450 answers (93.75% of answers) for the training dataset of the neural network Privacy Action and 30 answers (6.25% of answers) for the test dataset.

For validation of the mechanism, we used the test datasets (that were not used in the training of the neural networks) and it was observed the accuracy of the neural networks of the Privacy Everywhere mechanism. Thus, the mechanism performed 192 predictions with the Allow/Deny neural network and 30 predictions with the neural network Privacy Action. Of the 192 predictions made with the Allow/Deny neural network, 169 were correct. This corresponds to an accuracy of 88.02%. The neural network Privacy Action has succeeded in 26 out of 30 predictions, resulting in an accuracy of 86.67%. The results of the predictions can be visualized in Figure 6.

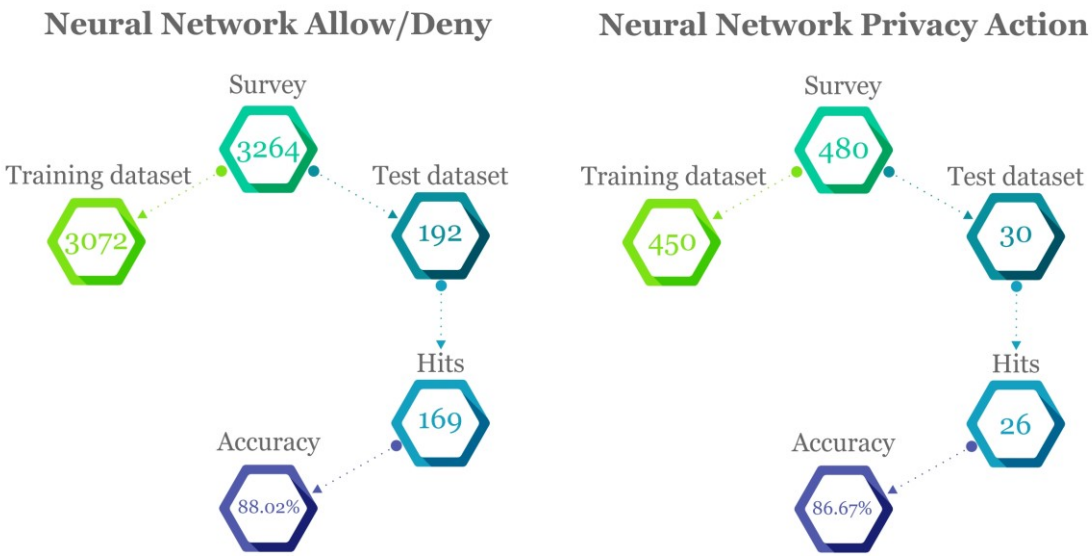

Figure 6. Results of predictions of the neural networks Allow/Deny and Privacy Action

To complete the validation, 30 data requests with random IoT scenarios were sent to the mechanism. With these requests it was possible to observe that: the mechanism was able to analyze the scenario of data collection, make a privacy decision and perform a privacy action; the Privacy Everywhere mobile application received the notifications properly; and the anonymization and encryption functionalities worked properly.

Discussion and Implications

The absence of mechanisms that effectively act to guarantee user privacy in IoT environments demand solutions to be developed. Disclosing private information indiscriminately without proper treatment results in loss of privacy (Adelhamid, Sharman & Bezawada, 2015).

This study represents an attempt to explore the effectiveness in making privacy decisions by user in IoT environments and to perform privacy actions on the data collected by IoT devices. The 88.02% accuracy of the Allow/Deny neural network indicates that this value is satisfactory and sufficient for the mechanism to make a privacy decision by user in environments with a large number of IoT devices and in situations where user does not wish to be disturbed. This study also provides theoretical insights into what factors influence decision making by users.

The 86.67% accuracy of the proposed mechanism's Privacy Action neural network indicates that this valor is sufficient to perform adequate privacy actions on data collected before this data reaches the requester. These privacy actions in IoT environments provide insights into how the privacy risks inherent in the emergence of new forms of interaction and communication can be mitigated.

From a practical perspective, this study has implications in several areas of application - smart homes, video surveillance, healthcare, smart cities, smart mobility, environmental monitoring and other areas that make use of IoT solutions. These types of applications generate a large amount of data flow requiring authorization and due privacy treatment before being sent. The results suggest that both privacy decisions by users and the privacy actions to be executed in each specific scenario can be automated.

Simulation functionality of the neural networks outputs has been added in the mobile Privacy Everywhere application so that the user can see how the mechanism behaves in certain scenarios. This simulation is necessary so that the user can use with more confidence the automatic mode of the mechanism.

Privacy actions performed by Privacy Everywhere mechanism aim not only to preserve user privacy but also to add an extra layer of security in the exchange of data between users and organizations. Increasing security in this data exchange is also interesting for companies, since the impact of malicious attacks causes serious implications to the bottom and top lines of an organization (Mukhopadhyay et al., 2013).

Conclusion and Future Works

Privacy Everywhere mechanism was presented on this paper. The proposed mechanism was able to help users with privacy issues, saving for they time and effort to make privacy decisions and performing privacy actions on the data collected in environments with high interaction and communication such as IoT environments. Technological evolution and the increasing adoption of integrated and connected solutions are moving towards more and more invasive environments in relation to privacy issues. Privacy Everywhere also presents itself as an additional tool for preserving user privacy.

As an improvement of the mechanism in future works, it is necessary to make a wide data collection with the potential users of the mechanism in order to improve the responses of the neural network in allowing or denying the sending of the data. Regional and cultural differences should also be considered for data collection and for mechanism evaluation. It would also be indicated to improve the mechanism, increase the number of factors for privacy decision making used for the construction of neural networks. This would increase the decision-making possibilities of the mechanism, but would also considerably increase the number of possible and necessary scenarios for the construction of the neural networks.

REFERENCES

Abdelhamid, M., Sharman, R., and Bezawada, R., 2015. "Better Patient Privacy Protection with Better Patient Empowerment about Consent in Health Information Exchanges". WISP 2015 Proceedings. 14.

Cochran, W. G., 1977. “Sampling techniques”. 3. Ed. Westlake Village: John Wiley & Sons, 428p. (Wiley series in probability and mathematical statistics: Applied probability and statistics). ISBN 0-471-16240-X.

Chow, R. 2017. “The last mile for iot privacy”. IEEE Security & Privacy, 15(6):73-76.

Dinev, T., Hu, H., Smith, J., H., and Hart, P., 2013. “Information privacy and correlates: An empirical attempt to bridge and distinguish privacy-related concepts.” In: European Journal of Information Systems, vol. 22(3), pp. 295-316.

Goel, S., Williams, K., J., Zavoyskiy, S., Rizzo, N., S., 2017. “Using Active Probes to Detect Insiders Before They Steal Data.” In: 23rd Americas Conference on Information Systems (AMCIS), Boston, MA, USA.

Gupta, M., Sharman, R., Walp, J., and Mulgund, P., 2018. “Information Technology Risk Management and Compliance in Modern Organizations”. 1st ed. IGI Global, Hershey, PA, USA, pp. 1-360. (doi: 10.4018/978-1-5225-2604-9).

Jayaraman, P., P., Yang, X., Yavari, A., and Georgakopoulos, D., 2017. “Privacy preserving Internet of Things: From privacy techniques to a blueprint architecture and efficient implementation”. Future Generation Computer Systems, vol. 76, pp. 540-549.

Lee, H. and Kobsa, A., 2016. “Undestanding user privacy in Internet of Things environments.” In: IEEE 3rd World Forum on Internet of Things (WF-IoT). Reston, VA, pp. 407-412.

Lowry, P. B, Dinev, T., and Willison, R., 2017. “Why security and privacy research lies at the centre of the information systems (IS) artefact: Proposing a bold research agenda.” In: European Journal of Information Systems, vol. 26(6) pp. 546-563.

Henze, M., Hemerschmidt, L., Kerpen, D., Haubling, R., Rumpe, B., Wehrle, K., 2016. “A comprehensive approach to privacy in the cloud-based Internet of Things”. Future Generation Computer Systems, 56, 701-718. (doi: https://doi.org/10.1016/j.future.2015.09.016).

Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A., and Sadhukhan, S., K. et al., 2013. “Cyberrisk decision models: To insure IT or not?” Decis. Support Syst. 56 (1), 11–26. (doi: https://doi.org/10.1016/j.dss.2013.04.004).

Naeini, P.E., 2017. “Privacy Expectations and Preferences in an IoT World.” In: Symposium on Usable Privacy and Security (SOUPS). Santa Clara, California.

Rosner, G., 2016. “Privacy and Internet of Things”. 1st Ed. O’Reilly Media, pp. 1-62. ISBN: 9781492042822.

Xu, H., Teo, H. H., Tan, B. C., and Agarwal, R. (2009). “The role of push-pull technology in privacy calculus: the case of location-based services”. Journal of Management Information Systems, 26(3), 135-174.