Privacy Protection with Pseudonymization and Anonymization In a Health IoT System

Results from OCARIoT

Sérgio Luís Ribeiro
CPQD
Campinas, Brazil
sribeiro@cpqd.com.br

Emilio Tissato Nakamura
CPQD
Campinas, Brazil
nakamura@cpqd.com.br

Abstract— **This paper presents the implementation of a users' privacy protection approach in a health Internet of Things (IoT) system. It is composed of a set of security layers based on cryptography, pseudonymization and anonymization techniques applied to processed (Data-In-Use, DIU), stored (Data-At-Rest, DAR) and transmitted (Data-In-Motion, DIM) data. Regarding security and privacy in IoT systems, especially in digital health systems, it is necessary to guarantee that the user rights are respected. This requires a security-in-depth strategy established based on risk-based results, every interconnecting actors, their security and privacy requirements and the specific aspects of the entire ecosystem, including the applications and platform. The presented privacy protection approach was developed and applied in a digital health platform, OCARIoT.**

Index Terms—**pseudonymization, anonymization, risk, privacy, security, IoT security, digital health.**

I. Introduction

An interconnected and integrated system is exposed to a wide range of risks, in a way that the developer has the responsibility to design and build it covering from the typical usage to the possible abuses and attackask attempts. Considering brute force or injection attacks to fraud, every system needs to be prepared to provide security, privacy, safety, resilience, and reliability. This responsibility grows even more in the Internet of Things (IoT) systems, that bring new security and privacy perspective resulting from the fusion between the human, the digital and the physical. Especially in health systems, there are daily elements flowing digitally between heterogeneous components that are processing, transmitting and storing data. Each of these components represents an attack point in a way that a digital incident directly reflects in the physical aspects of life that can ultimately affect the human being.

Security and privacy are essential, and in the OCARIoT (Smart Childhood Obesity Caring Solution using IoT Potential) [1], it is stated that "acceptance of the pilot still depend on the certainty that the security and privacy rights are respected, in the whole system – device, applications and platform providers", requiring a rigorous security-in-depth strategy. This includes the understanding of the potential threats to the system and the usage of the most appropriate defenses accordingly, including design and architecture. Besides risk assessment, Security by Design (SbD) [2] and Privacy by Design (PbD) [3] principles enable the system building with security and privacy in mind since the beginning, creating the necessary trust to be effectively used by the users in a real environment.

Although security and privacy has a long history, it had been evolving since the very beginning when it surged in the information security science. Global contexts like digital transformation, the advent of new technologies, digital-physical-human fusion, new cyberattack techniques, and emerging threats turn the protection more complex and challenging, demanding the evolution. In addition, new rules and laws related to privacy, such as European General Data Protection Regulation (GDPR) [4], Brazilian *Lei Geral de Proteção de Dados Pessoais* (LGPD) [5], Privacy Act [6], among others, demands a new and effective way to protect the privacy.

This paper presents the implementation of an approach to strengthen security and privacy in a health IoT system, using different security layers such as cryptography, pseudonymization and anonymization elements to protect the processed (Data-In-Use, DIU), stored (Data-At-Rest, DAR) and transmitted (Data-In-Motion, DIM) data.

II. Anonymization and Pseudonymization

Privacy protection is directly related to personal data, that GDPR defines as "any information relating to an identified or identifiable natural person ('data subject'). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" [7].

Anonymization and pseudonymization are two techniques that are recommended by the GDPR because they reduce the risk level and help in compliance with the data protection obligations. The main feature is to reduce the linking between the individual and the data, mainly after a data breach.

Anonymization is the permanent removal of any information that may serve as an identifier. Once a data set has been anonymized, it is impossible to identify individuals from it. This technique is usually used by organizations for marketing and research purposes, without the need for reaching the individuals. When done properly, anonymization can place data outside the scope of the GDPR. Anonymization is not primarily suitable to

2471-7819/19/$31.00 ©2019 IEEE
DOI 10.1109/BIBE.2019.00169

OCARIoT because of its nature to dealing with specific children data that requires historical, comprehensive and analytical data. Beyond that, data in OCARIoT is dynamic, interacting with different entities such as healthcare professionals, parents, educators, technology providers, and the children. Despite that, anonymization must be used to process general data used to create grouped statistics and overall views based on analytics.

Some anonymization techniques, highlighted by the GDPR´s Article 29 Working Party (WP) Opinion 05/2014, include [8]:

i. Noise Addition: adding a level of imprecision to the original data. For example, a patient's weight might show a range of +/- 7 kg., rather than a precise number.
ii. Substitution/Permutation: replacing information with other values. For example, a patient's height of 100 cm might be stored as "blue."
iii. Differential Privacy: converting individual user data into something unidentifiable by bundling and blurring it in one way or another.
iv. Aggregation/K-Anonymity: a "hiding in the crowd" concept where if each individual is part of a larger group, then any of the records in the group could correspond to a single person. For example, a data set might contain information about people in the Ceará State instead of specifying a specific town, like Canoa Quebrada.

Pseudonymization is defined in the GDPR as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information. Such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person." [7]. In other words, pseudonymization commonly refers to a de-identification method that removes or replaces direct identifiers (names, phone numbers, government-issued ID numbers, etc.) from a data set, but may leave in place data that could indirectly identify a person (often referred to as quasi-identifiers or indirect identifiers).

Applying only that method might be called simple pseudonymization. As there are different techniques to revert the pseudonymization, it is recommended that security and privacy controls designed to prevent the unauthorized re-identification of data would be applied on top of simple pseudonymization to create strong pseudonymization [9].

III. OCARIoT Scenario

The scenario to protect the data in OCARIoT is a result of risk assessment and includes a set of security and privacy controls, from encryption to security operations [10]. In this paper, we highlight the aspects related to the users' identification in OCARIoT, especially related to the main actors that have access to different sets of information: child, parents, Data Privacy Officer (DPO), platform administrator, healthcare professional, and educator.

A. Premises for OCARIoT identities and accesses

OCARIoT manages children's information related to health and habits that are collected manually via application, web dashboard or via IoT personal sensors like a smart band or environmental sensor. OCARIoT is accessed by a set of different entities and needs an access control policy accordingly to the data, privacy requirements and the entity. Some premises are:

- OCARIoT application is accessed by the children, while the web dashboard is accessed by parents, healthcare professionals, educators, and platform admin.
- Real children or natural children have correspondent identification in the system (Child ID).
- OCARIoT does not identify natural children.
- Real parents or natural parents have correspondent identification in the system (Parents ID).
- OCARIoT does not identify real parents.
- Healthcare professionals do not need to know the natural children.
- DPO is the school representative that has access to natural children's and parents' information.
- Platform Administrator (PA) is the entity that has access to the OCARIoT and input data into the system. This can be done by terminal or script.
- PA takes data from DPO to include them in the system.
- DPO/PA need to link the natural child to his Child ID / pseudonym.
- DPO/PA need to link the parents to the correspondent child, in the natural and ID form.

B. Macro steps for initial setup

This initial setup is related to the parents' and children's ID creation. There are four macro steps:

i. DPO/PA choose an ID for the child, Child ID.
ii. DPO/PA creates an ID for the parents, Parents ID.
iii. DPO/PA links Child ID to Parents ID.
iv. DPO/PA links sensors to the Child ID.

As only DPO has access to the natural children and parents, there is a need for DPO to provide the information to the PA to be inserted into the OCARIoT. This is a process outside of the OCARIoT and can be as simple as a spreadsheet or another external process. There are some considerations about the usage of Child ID and Parents ID:

i. The child to access the app uses Child ID.
ii. Child ID is used by the healthcare professional to insert child data into the OCARIoT Platform.
iii. The parents use Parents ID to access specific child data.
iv. Healthcare professional uses Child ID to insert child-related health data into the system.
v. Healthcare professional doesn´t need to know Parents ID.
vi. There is a need to recover or reissue Child ID/Parents ID.
vii. Only DPO has knowledge about the natural child.
viii. DPO interacts with PA, who interacts with the OCARIoT using the information provided by DPO.

Regarding the use of Child ID by the healthcare professional, a process must define how he will take the knowledge about the Child ID. One possibility is to ask the child to input his Child ID. The second one is to ask for the Child ID to the child and select it directly from the system. There are technological, usability and security issues related to each approach. In the first option, it is not necessary for the OCARIoT to store the Child ID, but only the correspondent hash (Section IV), which increases the security but interfering in the usability. In the second approach, the healthcare professional can hear the Child ID, selecting it from the system menu, in a user-friendlier way. The drawback is that it can decrease the security since a table of Child ID must be stored, instead of the hashed Child ID.

Similar issues would be considered for the use of groups of children. One option is to create groups by selecting the available Child IDs directly in the OCARIoT interface. This alternative has advantages in usability but has to store the Child IDs into the system, which decreases security. Another option is to create groups without the OCARIoT knowing the Child IDs, only their correspondent hashed versions. In this case, PA must input each Child ID or use a script to create groups, which decreases usability.

C. Basic components for identity and accesses

The basic components for OCARIoT's identity and accesses are in Figure 1. The entities accessing the OCARIoT represented in the figure are the child, Platform Administrator (PA), parents and healthcare professional. The Data Protection Officer (DPO) and school are other entities that do not access OCARIoT.

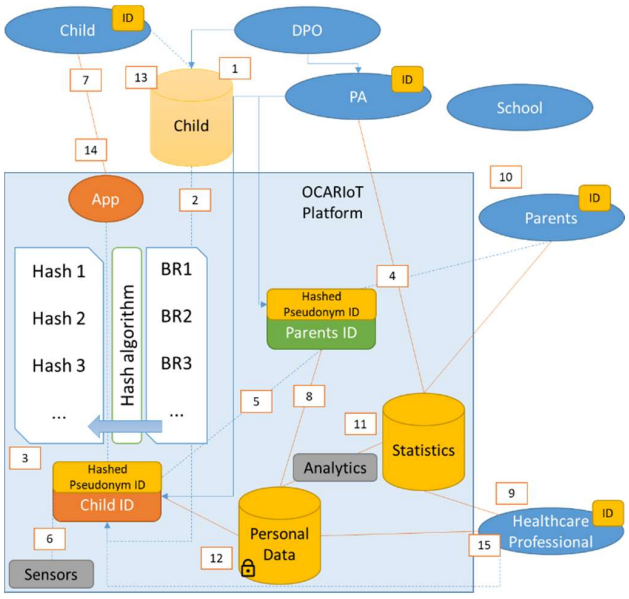

Figure 1: Basic components for identity and access control.

The main elements in Figure 1 are described as:

1. DPO has children's information that is not in OCARIoT.
2. DPO/PA chooses a system-generated Child ID for each child, e.g. BR1, BR2, ...
3. OCARIoT stores only the hashed Child ID.
4. DPO/PA creates an ID for the parents (process to be defined).
5. DPO/PA links Children ID to Parents ID.
6. DPO/PA links sensors to the Child ID.
7. The child to access the app uses Child ID.
8. Parents ID is used to access specific child data.
9. Healthcare professional does not need to know the Child ID or Parents ID.
10. There is a need to recover or reissue Parents ID.
11. Analytics is performed over the personal data database.
12. Personal data database uses encryption.
13. Children database is not in the OCARIoT.
14. Children access the app using the Child ID.
15. Healthcare professional uses Child ID to insert child data into the OCARIoT.

IV. Pseudonymization in OCARIoT Platform

OCARIoT needs pseudonymization instead of anonymization methods because there a requirenment to relink the Child ID and Parents ID to the real or natural user. The pseudonym in the OCARIoT is a set of characters that represents a natural user, like BR1, BR2 or 12. The child and parents use these pseudonyms as an identification or login to access the platform. The basic authentication method is the password.

Considering that a security incident can leak the database, including the identifications, an additional layer of security is to don´t store the IDs directly in the OCARIoT. In this case, what is stored in the platform is the hashed pseudonyms. This increases security because a potential leakage doesn't provide direct access to the Child ID and Parents ID, only to a hash with 256 bytes as a result of SHA-256 function [11]. An issue about this additional security is the affected usability since there are no lists of IDs provided to the user. Instead of that, the user needs to input his ID, just like traditional login methods.

There are three main types of cryptography algorithms: secret key, public key, and hash functions. Unlike secret key and public key algorithms that are based on secret keys, hash functions, also called message digests or one-way encryption, have no key. Hash is a fixed-length value resulting from computing on the plaintext that makes it impossible for either the contents or length of the plaintext to be recovered [11]. Hash algorithms are effective because of the low probability that two different plaintext messages will yield the same hash value.

In OCARIoT, the hash algorithm must be used every time the user inputs his pseudonym (Child ID or Parents ID). The OCARIoT than compares the hashed value calculated at that time with the stored hashed pseudonyms. Every data related to each child is linked to his correspondent hashed pseudonym.The link between the Child ID and the natural child is outbound from the OCARIoT, under the responsibility of the DPO.

A. Pseudonymization method for Child ID

DPO needs to know the Child ID in order to link it to the child, sensors, and parents. This process needs to be defined since it includes an external method for the DPO to perform the

links and the OCARIoT that doesn't know who the child is. The PA configures the linking. Healthcare professional needs to know the Child ID to insert child data into the OCARIoT. Once the Child ID is inputted, the OCARIoT performs the hash function to generate the hashed Child ID that is compared with the stored hashed Child ID.

Figure 2 shows the pseudonymization method for Child ID. Child ID is a code generated by the system (e.g. BR1, BR2, or something randomized) that is chosen by the DPO/PA linking it to the natural child. This is a pseudonym know by the child and by DPO and it will be used by the entities to interact with OCARIoT. In order to add an additional layer of security, the pseudonym will be stored in OCARIoT as a hash (hashed Child ID), using the SHA-256 hash algorithm [11].

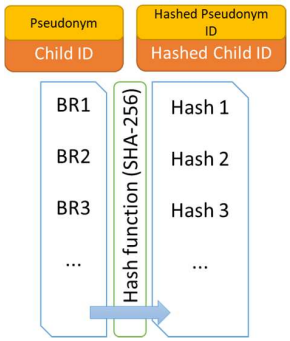

Figure 2: Pseudonymization method for Child ID.

A hashed Child ID is the only information stored inside OCARIoT. Every operation that uses the Child ID (e.g. BR1, etc.) performs a hash operation to compare the identity. The authentication method for the child to access the application is a password. OCARIoT uses a similar method used for the identity to protects the password but using salt to increase security by improving the randomness (Figure 3). Using salt, although the same password had been chosen by two different children (BR1 and BR2), they generate different hashes to be stored.

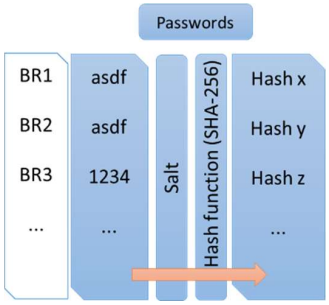

Figure 3: Password protection for Child ID.

B. Pseudonymization method for Parents ID

The pseudonymization method for Parents ID is the same from the Child ID. The Parents ID is used by the parents to access the OCARIoT dashboard to access the information about their children. Besides the parents themselves, only the DPO needs to know the linking between the Child ID and the Parents ID. The parents' pseudonym, Parents ID, can be chosen by the parents (easier to guess) or created by the OCARIoT (more difficult to remember). A SHA-256 hash algorithm [11] is applied to the pseudonym to generate a hashed Parents ID. This hashed version of Parents ID is the only information stored inside OCARIoT Platform in order to provide an additional layer of security. Similarly to the Child ID, every operation that uses the Parents ID performs a hash operation to compare the identity (Figure 4).

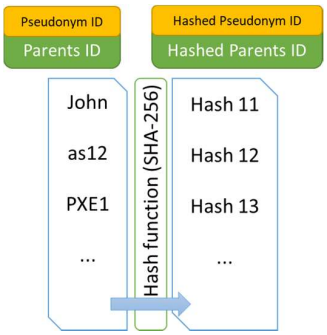

Figure 4: Pseudonymization method for Parent ID.

C. Data encryption in OCARIoT Platform

OCARIoT architecture is based on microservices, which implies data distribution. Data encryption is a necessary additional layer of security that applies to all personal data, reinforced by GDPR and LGPD. With encryption, a security incident like a leakage does not represent a direct compromise of children's personal data confidentiality.

Data encryption is not necessary for all databases and applies to DIM and DAR. OCARIoT process, transmit or store different types of information. Some of them are considered personal data, such as questionnaires, notification, recommendation, reports, prescriptions, manually inputted personal data (like weight, height, restrictions), sensors data and logs. Others are not considered personal data, such as the environmental data, credentials or commands.

An application-level protocol like Transport Layer Security (TLS) [12] is suitable to protect DIM between internal components and between OCARIoT and users. For DAR, there are two possibilities: (i) in the application-level or (ii) in the database-level. In the application-level, cryptography functions are used directly by the application on the server-side. In the database-level, there is the possibility to use a native encryption function in the database system or use third-party cryptography components integrated into the database system.

Symmetric cryptography algorithms are best suitable for DAR encryption since they are faster compared to asymmetric cryptography and a large range of computer processors supporting hardware-based cryptography. Advanced Encryption Standard (AES) is the standard algorithm for encryption [13]. Using symmetric cryptography has a challenge related to key management. Important aspects need to be addressed like [14]: (i) the generation, changing and destruction of encryption keys; (ii) how and where the encryption keys are stored; (iii) how the keys are protected.

V. Security Strategy for Health Systems

In the OCARIoT context, an attack flow can start in the children's smart sensors, app or dashboard, alongside with the server-side services. An exploited vulnerability in any of these elements can lead to unauthorized access to information and leakage, modification or destruction. Proper OCARIoT protection must prevent direct attacks against all its components, including users, hardware and software.

The strategy for OCARIoT is to reduce the attack surface by minimizing the use of personal information. Pseudonymization difficult the linking to the natural child and the stored information encryption protects against the data leakage. To protect OCARIoT from attacks against, for instance, the mobile device operating system, there are security techniques to run the app in containers and use secure hardware modules to store sensitive information like cryptographic keys.

Security layers in OCARIoT include (i) identification, authentication, and authorization, (ii) data encryption, (iii) communications security and (iv) physical and asset security (by the cloud provider). Other security layers will be highlighted by the risk assessment [10] that is under development, including software development security, security engineering, security assessment, and testing and security operations.

VI. Conclusions

This paper presents the implementation of an approach to strengthen security and privacy using different security layers such as cryptography, pseudonymization, and anonymization.

The application had been in OCARIoT, which is in the design and development process. Besides the integration between different assets from the partners, OCARIoT also integrates third-party components, sensors, and infrastructure.

Securing an IoT system requires a rigorous security-in-depth strategy. OCARIoT deals with critical personal data related to health and children. As each system is different from the other, identifying every point of attack in a particular context and understanding the interactions between different components is important to define the right security strategy. This is a result of risk assessment, which calculates the probability of threat agents exploring a set of vulnerabilities in one or more OCARIoT components, turning a threat into a security incident, which causes impacts.

In adherence to privacy-by-design, security-by-design, and privacy regulations like GDPR and LGPD, OCARIoT considers that "acceptance of the pilot still depend on the certainty that the security and privacy rights are respected, in the whole system – device, applications and platform providers." [1].

Consequently, it is crucial to implement an approach to strengthen security and privacy using different security layers, based on risk and implementing controls that encompass encryption, pseudonymization, and anonymization techniques to protect processed, stored, and transmitted data is required.

The presented privacy protection with anonymization and pseudonymization in a health IoT system is one of the necessary layers in a security architecture in order to provide better security and privacy that will improve the acceptance by users regarding data security and privacy.

Acknowledgment

The authors acknowledge the financial support given to this work, under OCARIoT project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731533 and the RNP under No 3007. Thanks to the partners TECNALIA, UPM, CERTH, UNP, CVE, SERMAS, EA, UNIFOR, ATLA, CPQD, NUTES, and UCE. This paper reflects only the author's views and the Agencies are not responsible for any use that may be made of the information contained therein.

References

[1] OCARIoT Project. *"Smart Childhood Obesity Caring Solution using IoT potential"*, Available: https://ocariot.eu/. Accessed June 2019.

[2] OWASP. "*Security by design principles*". Available: https://www.owasp.org/index.php/Security_by_Design_Principles. Accessed July 2019.

[3] IAPP. "*Privacy by design 7 fundamentals principles*". Available: https://iapp.org/resources/article/privacy-by-design-the-7-foundational-principles/. Accessed July 2019.

[4] European Commission. "*2018 reform of EU data protection rules"*. Available: https://ec.europa.eu/commission/priorities/justice-and fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en. Accessed June 2019.

[5] Lei Nº13.709. *"Lei sobre a proteção de dados pessoais"*. Available: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm. Accessed June 2019.

[6] Privacy Act. *"Privacy Act of 1974"*. Available: https://www.justice.gov/opcl/privacy-act-1974. Accessed July 2019.

[7] GDPR. "*General Data Protection Regulation*. Art4. GDPR Definitions". Available: https://gdpr-info.eu/art-4-gdpr/. Accessed June 2019.

[8] PERRY, B. "*Pseudonymization, Anonymization & GDPR"*. Available: https://medium.com/@brperry/pseudonymization-anonymization-gdpr-3dc8405dd465. Accessed June 2019.

[9] Privacy Analytics. "*Comparing Pseudonymization and Anonymization Under the GDPR"*. Available: http://www.privacy-analytics.com. Accessed July 2019.

[10] RIBEIRO, S.L., NAKAMURA, E.T. *"A Privacy, Security, Safety, Resilience and Reliability Focused Risk Assessment In a Health IoT System - Results from OCARIoT Project"*, June 2019. IEEE Global Internet of Things Summit (GIoTS), ISBN: 978-1-7281-2171-0, Arhus, Denmark.

[11] NIST. *"Hash Functions"*. Information Technology Laboratory. Available: https://csrc.nist.gov/projects/hash-functions. Accessed May 2019.

[12] OpenSSL. *"Cryptography and SSL/TLS Toolkit"*. Available: https://www.openssl.org/. Accessed June 2019.

[13] NIST. *"Block Cipher Techniques"*. Information Technology Laboratory. Available: https://csrc.nist.gov/projects/block-cipher-techniques/bcm. Accessed June 2019.

[14] Business.Com. *"How to Select the Right Encryption Key Management Solution"*. Available: https://www.business.com/articles/encryption-key-management-considerations/. Accessed June 2019.