Pseudonymization Approach in a Health IoT System to Strengthen Security and Privacy Results from OCARIoT Project

Sérgio Luís Ribeiro(✉) and Emilio Tissato Nakamura(✉)

CPQD, Campinas, Brazil
{sribeiro,nakamura}@cpqd.com.br

Abstract. Regarding security and privacy in Internet of Things (IoT), especially in a digital health system, is necessary to guarantee that user rights are respected. This requires an approach that considers security-in-depth strategy established on risk-based results, actors, their privacy and the entire ecosystem, including the applications and platform. This paper presents an approach to strengthen the security and privacy aspects, using different security layers based on cryptographic, pseudonymization and anonymization technics to protect the processed, stored and transmitted data. The approach present at this paper was developed and applied in a digital health platform in the Project OCARIoT.

Keywords: Security · IoT security · Pseudonymization · Privacy · Digital health

1 Introduction

Security and privacy in Internet of Things (IoT), especially in a digital health system, it is important to understand the potential threats to that system, and add appropriate defenses accordingly, as the system is designed and architected. In the OCARIoT (Smart Childhood Obesity Caring Solution using IoT Potential) [1], it is not different, acceptances still depend on the certainty that the security and privacy rights are respected, in the whole system – device, applications and platform providers that requires a rigorous security-in-depth strategy.

It is why is important to design from the start with security and privacy in mind (Security by Design-SbD [2] and Privacy by Design-PbD [3] principles) because understanding how an attacker might be able to compromise a system helps to make sure appropriate mitigations are in place from the beginning.

This responsibility grows even more in Internet of Things (IoT) systems, since the fusion between the human, the digital and the physical exists. There are daily elements flowing digitally between heterogeneous components that are processing, transmitting and storing data.

Although security and privacy have a long history, it had been evolving since the very beginning when it surged in the information security science, where the main objective is to prevent sensitive information from reaching the wrong people, while making sure that the right people can in fact get it. However nowadays, with the advent

R. Doss et al. (Eds.): FNSS 2019, CCIS 1113, pp. 134–146, 2019.
https://doi.org/10.1007/978-3-030-34353-8_10

of new rules and laws related to privacy, such as GDPR [4], LGPD [5], Privacy Act [6] and others, the current evolving world, demands new and effective way to protect the privacy.

This paper presents an approach to strengthen the security and privacy, using security layers such as cryptographic, pseudonymization and anonymization elements to protect the processed (Data-In-Use, DIU), stored (Data-At-Rest, DAR) and transmitted (Data-In-Motion, DIM) data.

2 Anonymization and Pseudonymization Concepts

Anonymization and pseudonymization are two techniques that are recommended by the GDPR because they reduce risk 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. Anonymizing data allows organizations to use the data for marketing and research, while protecting individuals from data exposure. However, since true anonymization is difficult to achieve, most businesses choose to use pseudonymization techniques [7].

Privacy protection is direct 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" [8].

When done properly, anonymization can place data outside the scope of the GDPR. Some anonymization techniques, highlighted by the GDPR´s Article 29 Working Party (WP) issued Opinion 05/2014, includes [7]:

i. **Noise Addition:** adding a level of imprecision to the original data. For example, a patient's weight shows 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. Typically, differential privacy works by adding some noise to the data and the amount of noise added is a trade-off – adding more noise makes the data more anonymous, but it also makes the data less useful [9].
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 São Paulo State instead of specifying a specific town, like Campinas.

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. Provided that 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." [8].

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, that is considered a way to correlate various information to identify a person, often referred to as quasi-identifiers or indirect identifiers.

Applying such a method, and nothing else, might be called simple pseudonymization. Frequently, security and privacy controls designed to prevent the unauthorized re-identification of data are applied on top of simple pseudonymization to create strong pseudonymization [10].

3 Security and Privacy in OCARIoT Platform

This section presents some important issues related to security and privacy in order to protect the processed (Data-In-Use, DIU), stored (Data-At-Rest, DAR) and transmitted (Data-In-Motion, DIM) data by the OCARIoT Platform and to comply with privacy regulations like European GDPR (General Data Protection Regulation) [4] and Brazilian LGPD (*Lei Geral de Proteção de Dados*) [5].

The focus is on some particular topics: pseudonymization, data encryption and protection against side-channel attacks that are considered one of the most common attack in the privacy scenario. Other topics that complete the holistic view for the security and privacy in OCARIoT Platform, that are under development, especially the results from the risk assessment [11].

3.1 Identities and Accesses Premises

OCARIoT Platform – Is a Platform that deals with children information related to health and habits information that can be collect manually via Application, Web Dashboard or via IoT sensors that can be personal sensor, such as smart band or environmental sensor.

OCARIoT Platform 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 Platform is accessed by the following entities:
  - Application: children.
  - Web Based Dashboard: parents, healthcare professionals, educators, and platform admin.
- Real children or natural children have a correspondent identification in the system (Child ID).

- OCARIoT Platform does not identify natural children. Real parents or natural parents have a correspondent identification in the system (Parents ID).
- OCARIoT Platform does not identify real parents.
- Healthcare professionals do not need to know the natural children.
- Data Privacy Officer (DPO) is the school representative that has the access to the natural children and parent's information.
- Platform Administrator (PA) is the entity that has access to the OCARIoT Platform and input data into the system. This can be done by terminal or script.
- PA takes data from DPO to include them into 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.

3.2 Considerations

There are some considerations about the usage of Child ID and Parents ID:

- The child to access the app uses Child ID.
- Child ID is used by the healthcare professional to insert child data into the OCARIoT Platform.
- The parents to access the specific child data use parents ID.
- Healthcare professional uses Child ID to insert child related health data into the system.
- Healthcare professional does not need to know the Parents ID.
- There is a need to recover or reissue Child ID or Parents ID.
- Only DPO has knowledge about the natural child.
- DPO interacts with PA. PA interacts with the OCARIoT Platform 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 to the child to input his Child ID, and other are 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 to the OCARIoT Platform to store the Child ID (only its correspondent hash (Sect. 2), what increases the security but interfering in the usability.

In the second approach, the healthcare professional can hear the Child ID and select it from the system menu, what make this user-friendlier, but on the other hand decreases security since a table of Child ID (instead of the hashed Child ID) must be stored in the OCARIoT Platform.

The same reason about technology, usability and security can be used to the use of groups of children. One option is to create groups by selecting the available Child IDs directly in the OCARIoT Platform interface. This approach has advantages in usability, but has to store the Child IDs into the system, what decreases security.

Other option is to creating groups without the OCARIoT Platform 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, what decreases usability.

3.3 Initial Setup

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

i. Data Privacy Officer (DPO)/Platform Administrator (PA) choose an ID for the child, Child ID.
ii. DPO/PA creates 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 the access to the natural children and parents, there is a need to DPO to pass the information to the PA to be inserted into the OCARIoT Platform.

This is a process outside from the OCARIoT Platform, and can be as simple as a spreadsheet or another external process.

3.4 Identity and Accesses

The basic components for OCARIoT Platform identity and accesses are in the Fig. 1.

The entities accessing the OCARIoT Platform 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 Platform. Besides that, the School and Home, are the entities where the Sensors are located, sending information (automatically) to the Platform

The main elements in the Fig. 1 are described as:

1. DPO has the children information that is not included in the Platform.
2. DPO/PA choose a generated Child ID for each child, e.g. BR1, BR2, …
3. OCARIoT stores only the hashed Child ID.
4. DPO/PA creates 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 the specific child data.
9. Healthcare professional does not need to know the child or Parents ID.
10. There is a need to recover or reissue Parents ID.
11. Analytics is performed over personal data database.
12. Personal data database uses encryption.
13. Children database is not in the OCARIoT Platform.
14. Children access app using the Child ID.
15. Healthcare professional uses Child ID to insert child data into the Platform.

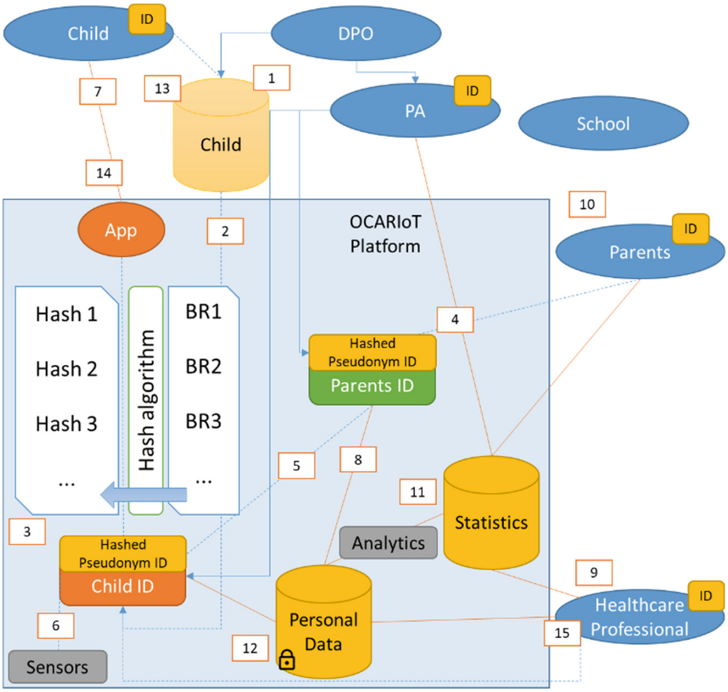

Fig. 1. Basic components for identity and access control.

4 Pseudonymization in OCARIoT Platform

OCARIoT Platform cannot use anonymization because of its nature to dealing with specific children data that requires historical, comprehensive and analytical data. Beyond that, data in OCARIoT Platform is dynamic, interacting with different entities such as healthcare professionals, parents, educators, technology provider and the children and also OCARIoT Platform needs methods to relinking the Child ID and Parents ID to the real or natural user.

The pseudonyms in the OCARIoT Platform are a set of characters that represents a natural user, like BR1, BR2 or as 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 Platform. In this case, what is stored in the platform is the hashed pseudonym. An issue about this additional security is the affected usability, since there are no lists of IDs provided to the user, for instance. Instead of that, the user needs to input his ID, just like traditional login methods.

In the case of healthcare professional inserting a child data into the system, or in the case of creating a group of children, the IDs must be previously known and inserted, and not chosen from a provided list of choice in the OCARIoT Platform interface.

What is stored in the OCARIoT Platform is the hashed pseudonym instead of the pseudonym. This creates an additional layer of security since 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 [12].

There are three types of cryptography algorithms: secret key, public key, and hash functions. Unlike secret key and public key algorithms, hash functions, also called message digests or one-way encryption, have no key. Hash is a fixed-length value resulting from a computing on the plaintext that makes it impossible for either the contents or length of the plaintext to be recovered [12]. Hash algorithms are effective because of the extremely low probability that two different plaintext messages will yield the same hash value.

In the OCARIoT Platform case, the hash algorithm applies over the Child ID and Parents ID, creating a 256 bytes value that represents the hashed pseudonyms. As it is not possible to revert a hash to the plaintext (hash is a one-way encryption), the hash algorithm must be used every time the user inputs his pseudonym (Child ID or Parents ID). The OCARIoT Platform 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 Platform. DPO has this responsibility

4.1 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 Platform that does not 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 Platform. Once the Child ID is inputted, the OCARIoT Platform 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, BR3, … 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 will be used by the entities to interact with the OCARIoT Platform.

To add an additional layer of security, the pseudonym will be stored in the OCARIoT Platform as a hash (hashed Child ID). A SHA-256 hash algorithm [12] applied in the original Child ID to generate the hashed Child ID.

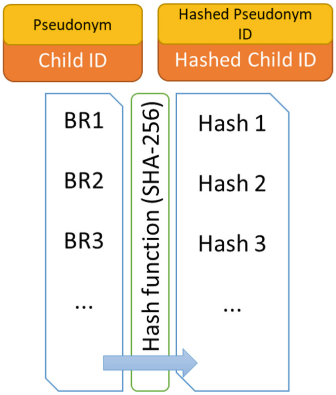

Fig. 2. Pseudonymization method for Child ID.

A hashed Child ID is the only information stored inside OCARIoT Platform (related to the Child ID). 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 password. The OCARIoT Platform using a similar approach for the identity protects the password, but using salt to increasing security, Fig. 3 shows the method to secure the passwords. Using salt, although the same password had been chosen by two different children (BR1 and BR2), they generate different hashes to be stored.

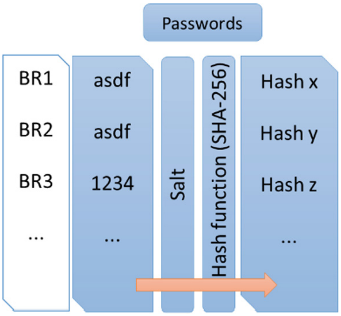

Fig. 3. Password protection for Child ID.

4.2 Pseudonymization Method for Parents ID

The pseudonymization method for Parents ID uses the same techniques from 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.

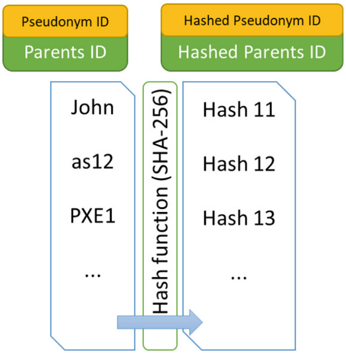

Fig. 4. Pseudonymization method for Parent ID.

The parent's pseudonym, Parents ID, can be chosen by the parents (easier to guess) or created by the OCARIoT Platform (more difficult to remember). A SHA-256 hash algorithm [12] 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 (related to the Parents ID) in order to provide an additional layer of security. Every operation that uses the Parents ID performs a hash operation to compare the identity (Fig. 4).

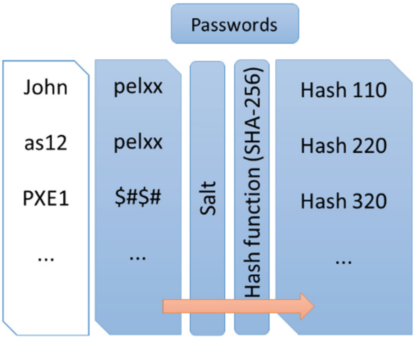

Fig. 5. Password protection for Parents ID.

Authentication method used by the parents to access the OCARIoT dashboard is based on the Parents ID and password Fig. 5. Each parent´ correspondent passwords is stored in a salted hash format.

4.3 Data Encryption in OCARIoT Platform

The OCARIoT Platform architecture is based on microservices, which implies data distribution. The data encryption is not a requirement to all databases. There are the

GDPR and GDPL compliance requirements and the results of a risk assessment that direct the adequate security controls implementation by OCARIoT Platform. The risk assessment is under development and is an OCARIoT Project result.

Data encryption is a necessary additional layer of security that applies to all personal data. With encryption, a security incident like a leakage does not represent a direct compromise of children personal data confidentiality. According to the GDPR, personal data is "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" [8].

OCARIoT Platform 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.

Data encryption applies to Data-In-Motion (DIM) and for Data-At-Rest (DAR). Identifying every points of attack in the OCARIoT Platform 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 agent exploring a set of vulnerabilities in one or more OCARIoT Platform components, turning a threat into a security incident, what causes impacts.

An application-level protocol like Transport Layer Security (TLS) [13] is suitable to protect DIM between internal components and between OCARIoT and users.

There are two possibilities to provide DAR encryption: in the application-level or in the database-level. In the application-level, cryptographic functions are used directly by the application in the server-side. In the database-level, there is the possibility to use a native encryption function in the database system or use third-part cryptography components integrated to 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 [14].

Using symmetric cryptography has a challenge related to the key management. Important aspects need to be addressed like [15]: (i) the generation, changing and destruction of encryption keys; (ii) how and where the encryption keys are stored; (iii) how the keys are protected.

5 Protection Against Side-Channel Attacks

Side-channel attacks are a class of physical attacks in which an adversary tries to exploit physical information leakages such as timing information, power consumption, or electromagnetic radiation. Since they are non-invasive, passive and can generally be performed using relatively cheap equipment, they pose a serious threat to the security of most cryptographic hardware devices. Such devices range from personal computers to small embedded devices such as smart cards and RFIDs (radio frequency identification devices [16].

In the OCARIoT Platform context, an attack flow can start in the children smart sensors, OCARIoT app or OCARIoT dashboard, alongside with the OCARIoT 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 Platform protection must prevent direct attacks against its components.

The protection against side-channel attacks in OCARIoT is more related to the sensors, mobile devices and hardware. In the software perspective, an appropriate memory management, especially for cleaning, is important. Processing OCARIoT information could lead to some timing, power or electromagnetic attacks.

The strategy for OCARIoT Platform 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 the OCARIoT Platform from attacks against, for instance, mobile device operational system, there are security techniques to run OCARIoT app in containers and use secure hardware modules to store sensitive information like cryptographic keys.

Security layers in OCARIoT Platform includes: (i) identification, authentication and authorization, (ii) data encryption, (iii) communications security and (iv) physical and asset security (by the cloud provider), this paper only present the data encryption security layer. Other security layers will be highlighted by the risk assessment [11] that is under development, including software development security, security engineering, security assessment and testing and security operations.

6 Future Work

In recent years, it is possible to notice a witnessed and exuberant wave of application possibilities for blockchain technology, since ensuring food safety and global self-sovereign digital identities, until decentralized virtual government management.

A blockchain technology is also appreciated, however aspects related to privacy, in some user case, should be considered and the approach presented at this paper can be used to strengthen security and privacy aspects.

In this way, security assessment and tests should be done in the project to have more experimental details and results to be present in other papers.

7 Conclusion

This paper presents an approach to strengthen the security and privacy, using different security layers such as cryptographic, pseudonymization and anonymization elements to protect the processed (Data-In-Use, DIU), stored (Data-At-Rest, DAR) and transmitted (Data-In-Motion, DIM) data, based on the results of the risk assessment.

This approach, considered that the OCARIoT Platform is in the design and development process by the following OCARIoT Project partners: Tecnalia Research & Innovation, Universidad Politecnica de Madrid, Centre for Research and Technology Hellas/Information Technologies Institute, Unparallel Innovation, Colegio Virgen de Europa, Servicio Madrileño de Salud, Ellinogermaniki Agogi, Universidade de Fortaleza, Instituto Atlântico, CPQD, Center for Strategic Health Technologies and Universidade Estadual do Ceará.

Besides the integration between different assets from the partners, the OCARIoT Platform also integrates third party components, sensors and infrastructure assets that were considered in the risk assessment.

Securing an Internet of Things (IoT) system, such as OCARIoT - due to the personal nature of the data collected - requires a rigorous security-in-depth strategy. In this way, to work with security and privacy in Internet of Things (IoT) is necessary to first look at the context where it will be used. Besides that, the best practices always suggest that is necessary to consider the whole system, starting from the context, then following all the layers, project, devices, applications, information data, network infrastructure and the platform. In the OCARIoT project, it is not different, acceptances still depend on the certainty that the security and privacy rights are respected, in the whole system – device, applications and platform providers.

Although GDPR does not refer to particular information such as: security standards, pseudonymization or anonymization technics. The use of these, provide a better security and privacy that will improve the acceptance by users regarding data security and privacy.

Consequently, 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.

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 No. 731533 and the RNP under No. 3007. 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. https://ocariot.eu/. Accessed June 2019
2. OWASP: Security by design principles. https://www.owasp.org/index.php/Security_by_Design_Principles. Accessed July 2019
3. IAPP: Privacy by design 7 fundamentals principles. https://iapp.org/resources/article. Accessed July 2019
4. European Commission: 2018 reform of EU data protection rules. https://ec.europa.eu/commission/priorities/justice-andfundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en. Accessed June 2019

5. Lei N°13.709: *Lei sobre a proteção de dados pessoais*. http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/L13709. Accessed June 2019
6. Privacy Act: Privacy Act of 1974. https://www.justice.gov/opcl/privacy-act-1974. Accessed July 2019
7. Perry, B.: Pseudonymization, Anonymization & GDPR. https://medium.com/@brperry/pseudonymization-anonymization-gdpr. Accessed June 2019
8. GDPR: General Data Protection Regulation. Art4. GDPR Definitions. https://gdpr-info.eu/art-4-gdpr/. Accessed June 2019
9. Valdez, A.C., Ziefle, M.: The users' perspective on the privacy-utility trade-offs in health recommender systems. Int. J. Hum.-Comput. Stud. **121**, 108–121 (2019)
10. Privacy Analytics: Comparing Pseudonymization and Anonymization Under the GDPR. http://www.privacy-analytics.com. Accessed July 2019
11. 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. In: IEEE Global Internet of Things Summit (GIoTS), Arhus, Denmark, June 2019. ISBN 978-1-7281-2171-0
12. NIST: Hash Functions. Information Technology Laboratory. https://csrc.nist.gov/projects/hash-functions. Accessed May 2019
13. OpenSSL: Cryptography and SSL/TLS Toolkit. https://www.openssl.org/. Accessed June 2019
14. NIST: Block Cipher Techniques. Information Technology Laboratory. https://csrc.nist.gov/projects/block-cipher-techniques/bcm. Accessed June 2019
15. Business.Com: How to Select the Right Encryption Key Management Solution. https://www.business.com/articles/encryption-key-management-considerations/. Accessed June 2019
16. Verbauwhede, I.M.R.: Secure Integrated Circuits and Systems. Integrated Circuits and Systems. Springer, Boston (2010). https://doi.org/10.1007/978-0-387-71829-3