Ensuring Data Privacy in Automated Supermarkets

Information

Author: Mattia Bertacchini
Supervisor: Lisa Trigiante

January 24, 2024

Abstract

This essay delves into the intricacies of privacy concerns within the realm of Internet of Things (IoT), specifically focusing on the implementation of an innovative automated supermarket akin to Amazon Go. The proposed system integrates artificial intelligence, RFID technology, and ESP cameras to facilitate the shopping experience. Customers effortlessly add items to their virtual carts as they pick them from the shelves, eliminating the need for traditional checkouts and consequently enhancing overall shopping efficiency.

While the project aims to revolutionize the customer experience by reducing shopping time and minimizing the stress associated with item unavailability, it is imperative to address the potential privacy implications associated with such advanced IoT applications. One notable consideration involves the utilization of customer-specific wearables RFID tags, such as bracelets, to facilitate seamless tracking and interaction with the automated supermarket system.

This essay explores the necessity of managing data privacy in the depicted context and addresses the implications of the General Data Protection Regulation (GDPR) on data consent requirements. Furthermore, it investigates strategies to anonymize and aggregate data, allowing for permissible use in statistical analysis without compromising individual privacy. The overarching objective is to strike a balance between technological innovation and safeguarding customer privacy, ensuring the ethical and responsible deployment of IoT solutions in the realm of automated supermarkets.

Contents

1. Introduction

In the real of Intelligent Internet of Things, security and privacy are distinct yet interconnected concepts.

Indeed, even though the IoT allows great interconnection of devices enhancing the user experience in a variety of application fields, it embeds security blind spots that can leave the devices susceptible to hacking thereby leading to security and privacy concerns for the users.

As introduced by journal Elsevier (1) security threats to harm consumers are: (a) unauthorized access and improper use of personal information (e.g. information about frequent visited locations), (b) promotion of attacks on other systems (c) increase of security risks. Therefore it involves protecting devices, networks and data from unauthorized access using encryption, authentication or access control preventing from significant consequences like data theft or disruption of data service.

Privacy concerns in IoT applications pertain to the collection, storage, processing and sharing of personal data generated by IoT devices and systems. Indeed, involved devices collect vast amount of sensitive information about costumers like location data, behavior patterns and preferences. It is essential to ensure that users' personal data is treated securely and in accordance with privacy regulations, protecting them from unauthorized access and ensuring that they are used only for lawful purposes and previously agreed with the users themselves.

This essay seeks to comprehensively address the main limitations of existing work which can be found in the IoT project described in the Abstract and aims to find suitable solutions, that will allow it be respect GDPR requirements (2) (3) (4).

2. Security Issues

As defined above, security issues in the IoT field regard problems related to device hacking and intrusion. Therefore solutions will focus on finding a way to protect devices from external agents and intruders which might be more difficult than expected due to hardware constrained capabilities.

Nowadays, addressing these challenges and ensuring the security of IoT products and services should be a top priority. Not by chance, the case study under consideration was born from the intent to improve and fasten the shopping experience of a market. Therefore, the user should feel that he is controlling any information related to them if we want the new technology to be tempting.

2.1 Security vulnerabilities

The first thing that should concerns us is the question: why are IoT devices so vulnerable?

As Elsevier Journal explains (1), the answer involves several reason. First of all components are often unattended making physical attack easy. Second, wireless communication, mostly adopted in this field, makes eavesdropping extremely simple. Finally, devices low capabilities in terms of both energy and computing resources make IoT implementations often unable to implement complex schemes supporting security.

2.2 Security concerns in case study

Given what defined so far, here’s the list of security concerns in the case of interest.

(a) *Identification*: Providing identity to an adversary can be a serious threat and that’s why a smart device should understand if it should or should not do so. In terms of our application, it is necessary to identify registered RFID tags for system activation.

(b) *Authentication*: Passive RFID tags do not have the capabilities to exchange a huge number of messages with an authentications server. This makes authentication process much more complex.

(c) *Data Integrity*: Even if in current state of art data integrity is not the primal issue, considering the future implementation of payment methods, it will have to get to the top of the priority list of concerns. Data Integrity involves of common surveillance methods to protect useful information from cybercriminals and to prevent external interference during transmission and reception. Methods that ensure data accuracy and reliability include checksums and cyclic redundancy checks (CRC) used through error detection mechanism.

(d) *Data Confidentiality*: Ensuring data confidentiality is paramount for user trust and security, employing diverse mechanisms to deter unauthorized disclosure. In the context of IoT-based devices, this safeguards sensor networks from revealing sensor node data to adjacent nodes and prevents unauthorized readers from accessing label data transmissions.

(e) *Access Control*: Access control deals with permissions in the usage of resources of a wide IoT network. It should focus on IoT capabilities because factors affecting access-control decision are heavily context-dependent.

(f) *Data privacy*:With the escalating volume of data within IoT settings, the pressing challenge of potential data usage for purposes beyond their original intent becomes increasingly critical to address. Within the IoT ecosystem, comprising devices, sensors, readers, and applications, diverse data types pertaining to individuals are continuously gathered as they navigate through such environments. The aggregation of this data raises the risk of individual identification. Data collected via object identifiers, sensor readings, and IoT system connectivity can inadvertently expose details about individuals, including their behaviors, whereabouts, interests, and other personal preferences stored for streamlined system functionality.
This is the main point of our study case and the bridge to Section 3. Indeed, should an individual intercept all data sent to the MQTT broker, they could potentially conduct analytics on user preferences. Thus, even if data is accessed, intruders should be unable to comprehend its contents.

It is worth saying that each architectural layer, as shown in Table 1, presents specific issue that must be take into account. At the network layer threats are classified with risk levels ranging from Low to Medium due to the known disadvantages of wireless data transfer standards. Assigning an IP address to a smart device connected to the Internet provides the function of transmitting network data. Deficiencies in these devices provide a platform through which hackers and government agencies can access the network, monitor users, and access other connected devices for a variety of purposes. The threats at the application layer are classified with Medium security risk level due to the large number of users and the data to be stored and processed in the layer, in addition to the known vulnerabilities of virtualization.

Then, malicious code injection and denial-of-service (DOS) attacks, having high risk levels, are found in both application and network layers.

3. Privacy Issues

In IoT, a multitude of devices collect, analyze, and transmit a huge variety of sensitive data across vast networks. This influx of data, teeming with valuable insights and personal information, demands protection from potential adversaries. Simultaneously, it is imperative that user is aware of the way his data is being precessed within the ecosystem.

3.1 Profiling

For the discussed project, the main privacy threats is *Profiling*.

Profiling stands as a pivotal tool employed to personalize e-commerce experiences, encompassing diverse facets such as personalized newsletters and targeted advertisements. Through the intricate interplay with an array of profiles and data repositories, organizations diligently amass pertinent information to refine their marketing strategies. With the continuous evolution of the Internet of Things, the landscape witnesses an exponential surge in data sources on a daily basis. This surge not only amplifies the volume of data collected but also catalyzes qualitative shifts within the data. As organizations delve deeper into data acquisition, previously inaccessible dimensions of individuals' personal lives are brought to light, fundamentally altering the qualitative fabric of collected information.

Profiling could lead to privacy violations if the data is used for unsolicited ads, price discrimination, and social engineering. The challenge is in balancing the interests of the user with the requirements of the user’s privacy when creating and analyzing data profiles.

Here’s the evident connection with the study case: if someone manages to access shopping history of a user, it won't take that long to come up with the described issue.

But there’s more, considering the application we’re in, another huge problem comes to mind. Indeed, being able to access user data, will let intruder access fundamental information like frequently visited stores, allowing him to identify sensitive data like user locations and shopping habits.

3.2 Vulnerabilities that lead to Privacy Issues

Let’s now take into account what might lead to Profiling or any other kind of privacy threats.

(a) *Weak programming practices*: Many researches have reported that a high number of firmware are released with known vulnerabilities like backdoors, root users as prima access and the lack of Secure Socket Layer (SSL) usage. This might make the adoption of strong programming practices nearly useless. Indeed, those security weaknesses can be exploited to cause buffer overflows, informations modifications or main unauthorized access to a certain device.

(b) *Data leakage*: Celik et al. (5) conducted an analysis, finding out that 138 out of 230 SmartThings applications expose at least

Layer	Concerns	Threat level
Perception	**Eavesdropping**. Within the RFID technology, an attacker could easily sniff out the confidential information like passwords or any other data flowing from tag-to-reader or reader-to-tag which makes it vulnerable.	High
	RF Jamming. RFID tags can also be compromised by kind of a DoS attack in which communication through RF signals is disrupted with an excess of noise signals.	Medium
Network	**Denial of Service (DoS) attack**. The kind of attack in which the network is flooded with a useless lot of traffic by an attacker, resulting in a resource exhaustion of the targeted system due to which the network becomes unavailable to the users.	High
	Malicious code injection. This is a serious kind of attack in which an attacker compromises a node to inject malicious code into the system which could even result in a complete shutdown of the network or in the worst case, the attacker can get a full control of the network.	High
Application	**Spear-Phishing attack**. It is an email spoofing attack in which victim, a high ranking person, is lured into opening the email through which the adversary gains access to the credentials of that victim and then by a pretense retrieves more sensitive information.	Low

Table 1
Security concerns for each architectural layer.

one piece of sensitive data such as device info, device state, user input via the Internet or messaging services. This problem can also be found in our project, which publishes via MQTT device information like IP and user information like RFID tag ID. Therefore, knowing the ID of a user, an attacker can easily locate a person or monitor his buying preferences.

(c) *Improper encryption*: Since complex cryptographic function can result in large overhead for resource-constrained IoT smart things, interest in ultra-lightweight algorithms is rapidly erasing. These consist in light but secure encryption algorithms optimized for low-powered hardware. However, the power dissipation of the derive can be measured while performing encryption and later statistically analyzed to recover the secret key, thus compromising the hardware. Countermeasures incur significant power and performance overheads and therefore are now suitable for lightweight cryptographic primitives.
Although power dissipation problem, the usage of ultra-lightweight algorithms might be suitable for our application. Since it does not add large overhead and does not require much power, it allows real-time functionalities, which are a primal necessity for the proper usage of the technology developed, while sufficiently increasing the level of security.

4. Countermeasures

Before moving to GDPR analysis, having defined the threats within the realm of IoT concerning both Privacy and Security, it becomes imperative to delineate potential countermeasures applicable to our study aimed at mitigating and resolving these concerns.

Among all possible countermeasures like access and authentication controls, security protocols, intrusion detection, single sign-on, establishing trust, security awareness, privacy by design, and security tools, we deduced the crucial ones for our application.

4.1 Access and authentication controls

As previously mentioned, using authentication and access controls techniques is a crucial necessity for limiting access and avoiding unwanted intruder. Different types of access control exists:

(a) *Role-Based Access Control*: Different roles, like admin, cild or guest, are assigned to each user. Only administrators are allowed to make changes, access control policies, add new users and organize the devices connected.

(b) *Location-Based Access Control*: If not physically near the devices, some user functionalities might be restricted or made unavailable.

(c) *Supervisory Access Control*: Enables a user who might otherwise be restricted from device usage to access the device only in the presence of another authorized user. This might be useful to avoid undesired inventory movements, allowing warehouse worker to make inventory movements or access data only when a supervisor is present.

(d) *Reactive Access Control*: Similar to point (c), if a user attempts to use a device they do not have permission to use, the application will notify a more privileged user, like the administrator, for permission in real-time.

In addition to what just defined, Fernandes et al. (6) proposed a method of restricting access to IoT sensitive data. Creating a system called FlowFence, they allows programs to control use of data so that only authorized devices can access them.

In our specific context, FlowFence could potentially facilitate the differentiation between user information, which may be encrypted due to its sensitive nature, and item movements, which might remain unencrypted. This distinction is crucial as it ensures that sensitive user data is adequately protected while allowing for the tracking of item movements within the system. It's worth noting that the ability to access movement data without the capability to correlate it with specific users renders such information futile for potential attackers, thus enhancing overall system security.

4.2 Intrusion detection

IoT smart devices should be able to defend themselves from outsider attacks. Therefore, they should be able to use intrusion detection tools, identifying abnormal activities and make user aware of the critical event occurred.

For example, when a malware attack is identifies in one given device, users should be notified immediately to facilitate immediate action, such as password changes or the implementation of other essential precautions, aimed at preventing potential theft of personal information.

Still referring to the context we are analyzing, it is possible to see how the possibility to rapidly change a user ID can be useful. Another suitable solution is, as mentioned, a user notification that invites him to contact the appropriate person to change the RFID badge. As depicted in the following paragraph, the use of a token might come useful considering future implementation of online payment methods.

4.3 Establishing trust

Establishing trust is essential for the successful implementation of IoT. It encompasses users' perceptions and experiences while engaging with IoT systems. Feelings of vulnerability and the perception of being subjected to external control can significantly undermine the adoption and efficacy of IoT-based applications and services.

Xie & Wang (7) presented the concept of mutual trust for inter-system security in IoT by creating an item-level access-control framework. It establishes trust from creation to operation and the IoT transmission phase. This trust is established by two mechanisms; the creation key and the token. Any new device which is created is assigned a creation key by an entitlement system. The device manufacturer must request for this key. The token is generated by the manufacturer or current owner, and this token is combined with the RFID identification of the device. This mechanism ensures that the permissions are changed by the same device if a new owner is appointed, or it will be used in a different department of the same company, thus reducing the overheads of the new owner. Owners can change these tokens, provided the previous token is availed, to replace permissions and access control to the previous token.

5. GDPR analysis

GDPR stands for General Data Protection Regulation, which is a legal framework that establishes guidelines for the collection and processing of personal data of individuals within the European Union (EU) and the European Economic Area (EEA). It also applies to organizations outside the EU and EEA that offer goods or services to, or monitor the behavior of, EU data subjects (2) (3) (4). GDPR entails several principles and requirements that impact IoT projects, including:

(a) *Lawfulness, fairness and transparency*: It is required to have a valid legal basis for processing personal data, such as consent, contract, legal obligation, vital interest, public interest, or legitimate interest. It is also necessary to inform the data subjects about the purpose, scope, duration, and recipients of the data processing, and respect their rights to access, rectify, erase, restrict, object, and port their data.

(b) *Purpose limitation*: Personal data must only be collected and processed for specific, explicit, and legitimate purposes, and

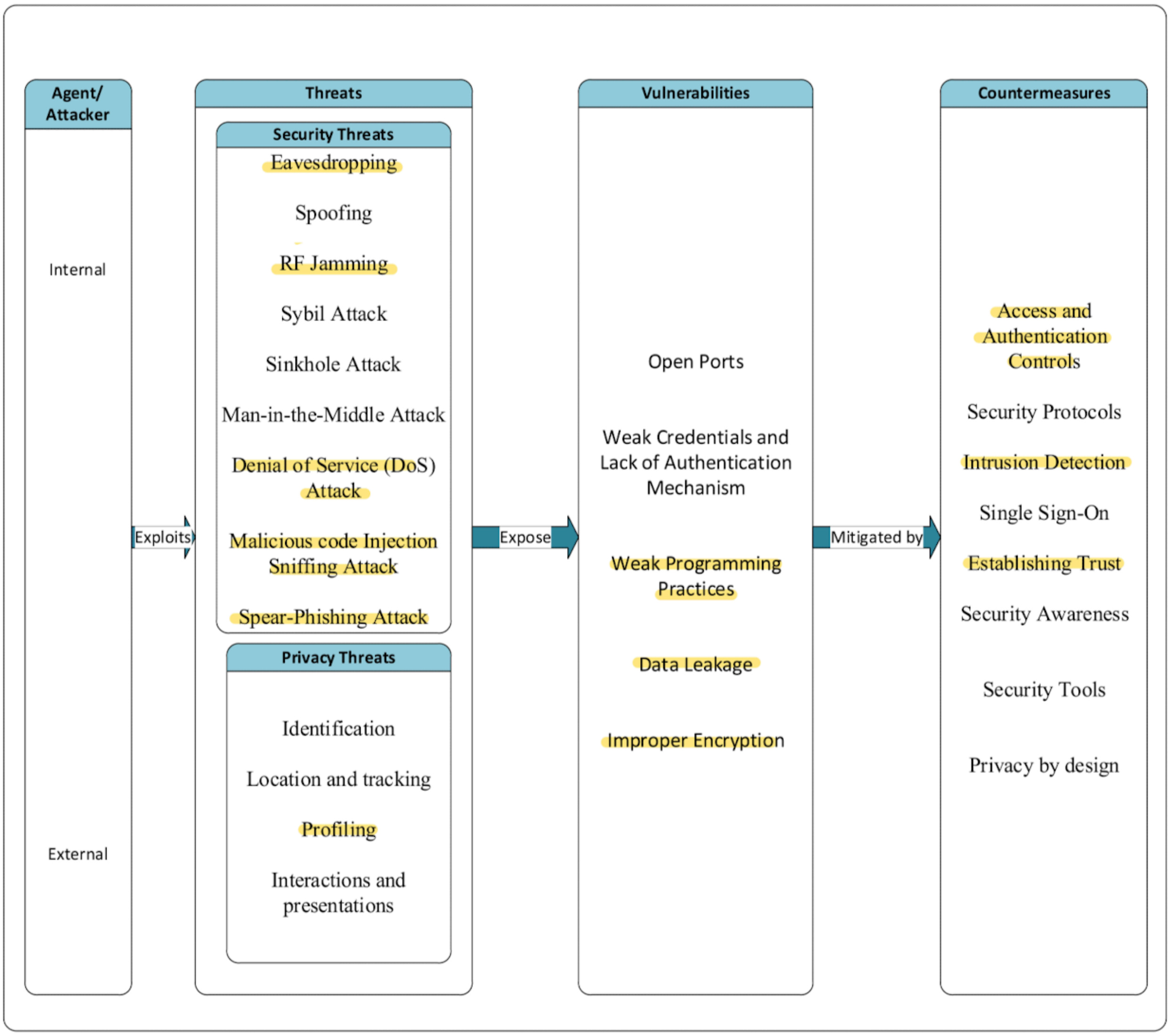

Fig. 1
Image is taken from (1) and shows Security and Privacy threads, vulnerabilities and solutions. Can be seen as summary of what has been delineated so far. Highlighted elements are those which interest the study case.

not used in a way that is incompatible with those purposes.

(c) *Data minimization*: Personal data must only be collected and processed if it is adequate, relevant, and necessary for the purposes for which it is processed.

(d) *Accuracy*: Measures must be taken to ensure that the personal data collected and processed is accurate and up to date, and reasonable steps must be taken to correct or delete inaccurate or outdated data.

(e) *Storage limitation*: Personal data must only be retained for as long as necessary for the purposes for which it is processed, and deleted or anonymized when it is no longer needed.

(f) *Integrity and confidentiality*: Personal data must be protected from unauthorized or unlawful access, use, disclosure, alteration, or destruction, using appropriate technical and organizational measures.

(g) *Accountability*: Compliance with the GDPR must be demonstrated by keeping records of data processing activities, conducting data protection impact assessments, appointing data protection officers, and reporting data breaches.

Now, as first step, let's dig into the application studied in order to evaluate data manipulated by devices.

5.1 Application Overview and Data Architecture

For every customer, a comprehensive set of data points is stored, including their name, surname, email, user ID. Moreover, each user's transactional history is documented, encompassing details such as the items purchased, their quantities, and the specific locations where the transactions occurred. Specifically, the application captures an array of information for each item acquired, comprising the item's unique identifier, its description, the quantity procured, as well as the precise longitude and latitude coordinates of the market where the purchase transpired. Furthermore, each user can create a list of favorite articles.

This data architecture enables a granular understanding of user behavior, facilitating insights into their frequented locations and purchasing preferences, including preferred items and other pertinent preferences.

5.2 GDPR and IoT Project Alignment

In the depicted context it's crucial to align data practices with the requirements outlined in the General Data Protection Regulation (GDPR). Here are some key considerations and recommendations regarding GDPR compliance for case study under analysis:

- *Personal data*: Any data that can identify or relate to a natural person, directly or indirectly, is considered personal data under the GDPR. This includes names, surnames, user IDs, email, frequent visited locations, and lists of favorite articles. It also includes online identifiers, such as IP addresses, cookies, RFID tags, and device identifiers. In this context, according to Data Minimization Principle, reviewing data collected for each customer might be a good practice. This means that, exception made for the email that can be used for user identification and login, name and surname of the customer might be removed.
- *Data protection by design and by default*: Data protection principles and measures should be integrated into the design and default settings should be privacy-friendly. This entails, for example, using pseudonymization or encryption techniques to protect the data, limiting data collection and retention to the minimum necessary, and providing data subjects with control over their data. Using, as explained in the previous point, email and User Identifier to associate movements with individual accounts, will prevent useless exposure of personal information like name and surname while still allowing user identification when needed.
- *Data processing agreement*: If third-party services or platforms are utilized, such as cloud providers, payment processors, or analytics tools, a data processing agreement should be established. This agreement specifies the roles, responsibilities, and obligations of each party regarding the data processing, and ensures compliance with the GDPR.
- *User Consent*: According to Lawfulness, fairness and transparency Principle, it is necessary to clearly define the purposes for which customer data are collected, such as facilitating transactions, improving service efficiency, analyzing shopping patterns and understand future availabilities trend. In addition to this, it is also mandatory to obtain explicit consent from customers, giving them the option to opt-in or opt-out of data collection and processing activities.

6. Conclusion

ToDo / capire cosa aggiungere o se terminare con qualcos'altro.

References

(1) A survey on privacy and security of Internet of Things, Mark Mbock Ogonji, George Okeyo, Joseph Muliaro Wafula

(2) IoT regulation: IoT, GDPR, ePrivacy Regulation and more regulations

(3) GDPR and Internet of Things (IoT) - Legal IT group

(4) GDPR Compliance for Internet of Things (IoT) Devices: Privacy in a Connected World

(5) Z. Berkay Celik, et al., Open access to the Proceedings of the 27th USENIX Security Symposium is sponsored by USENIX. Sensitive Information Track- ing in Commodity IoT Sensitive Information Tracking in Commodity IoT, in: USENIX Secur. Symp., 2018

(6) E. Fernandes, J. Paupore, A. Rahmati, D. Simionato, M. Conti, A. Prakash, Flowfence: Practical data protection for emerging iot application frameworks, in: USENIX Security Symposium, 2016

(7) I. Chong, A. Xiong, R.W. Proctor, Human factors in the privacy and security of the internet of things, Ergon. Des. 27 (3) (2019) 5–10