T.C

ALTINBAS UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

A REVIEW OF THE ISSUES AND CHALLENGES IN IoT SECURITY USING MACHINE LEARNING TECHNIQUES

Liberty Adams OMONKHOA

Master of Science

Supervisor

Asst. Prof. Abdullahi Abdu IBRAHIM

Istanbul, 2021

A REVIEW OF THE ISSUES AND CHALLENGES IN IoT SECURITY USING MACHINE LEARNING TECHNIQUES

by

LIBERTY ADAMS OMONKHOA

Electrical and Computer Engineering

Submitted to the Graduate School of Science and Engineering

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2021

The thesis titled “A Review of The Issues & Challenges in IoT Security Using Machine Learning Techniques” prepared and presented by Liberty Adams Omonkhoa was accepted as a Select Degree Thesis in Department.

_____

Asst. Prof Abdullahi Abdu IBRAHIM

Supervisor

Thesis Defense Jury Members:

Asst. Prof. Abdullahi Abdu IBRAHIM	Faculty of Engineering and Architecture, Altinbas University	_____
Asst. Prof. Muhammad ILYAS	Faculty of Engineering and Architecture Altinbas University	_____
Asst. Prof. Sewale MASADAQ TAHA TAHA	Facultyof Communication, Beykent University	_____

I certify that this thesis satisfies all the requirements as a thesis for the degree of .....

Approval Date of Institute of Graduate Studies:
____/____/____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Liberty Adams Omonkhoa

DEDICATION

This research thesis is dedicated to all members of my family comprising of my parents, siblings, nephews, nieces and in-laws.

ACKNOWLEDGEMENTS

I cannot express enough thanks to the members and staff of the Electrical and Computer Engineering Department of Altinbas University for their support and encouragement. Worthy of mention is my supervisor, Dr. Abdullahi Abdu Ibrahim and the entire project defense committee.

My completion of this project could not have been accomplished without the support of my classmates, Emma, Ali, Amar and Idris. Special thanks also go to my longtime friend and brother, Martins, thanks for sticking by me when I needed you most.

My biggest and final thanks go to God Almighty, the creator of our universe for his endless love, mercies, and grace.

ABSTRACT

A REVIEW OF THE ISSUES AND CHALLENGES IN IoT SECURITY USING MACHINE LEARNING TECHNIQUES

OMONKHOA, LIBERTY ADAMS

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Dr. Abdullahi Abdu Ibrahim

Co-Supervisor:

Date: July/2021

Pages: (68)

The Internet of Things or IoT is no longer a technological promise. The machine-to-machine connection on a large scale is already a reality in more advanced economies and its expansion occurs at an astonishing speed. There will be many companies that supply this software and hardware in the IoT and the final consumers will be the biggest holders of this type of technology. IoT is an emerging computing paradigm that integrates the “things” of the real world into the technological world. There are many advantages and opportunities that the IoT can offer, however, there is a point that still raises many doubts and concerns both suppliers and consumers: how will the security and privacy of the data be guaranteed?

The concern with security and privacy is an argument that was, is, and will always be valid in any society, area or technology. IoT is exactly the same. This is seen as the future that will transform our interactions with objects and with our own life. Where a large network of connected objects will act intelligently. However, with this large network of connected devices and the huge data flow it will generate, the security and privacy of data on the IoT will be a fundamental challenge. According to a study by Gartner [5], the concern with security and privacy are the biggest obstacles to the adoption of IoT. This dissertation aims to understand how IoT organizations and consumers are addressing the challenge of data privacy and security. The objective is to give an empirical perspective on this theme, both from the approach of organizations that supply IoT

products, from consultants in the area of information technologies, as well as from the end users of these same products.

.

Keywords: IoT, Security, Machine learning.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

Table 4.6: IoTPC Items by Extracted Factors 58

Table 4.7: Inference Details of Each Generated Module 60

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIPC	:	Adopted Informtion Privacy Concern
AoT	:	Analytics of Things
CACP	:	Carlifonia Consumer Privacy
CFA	:	Confirmatory Factor Analysis
EFA	:	Exploratory Factor Analysis
FTC	:	Federal Trade Commission
GDPR	:	General Data Protection Regulation
GIS	:	Geographical Information Systems
GPS	:	Global Positioning System
ICT	:	Information and Communication Technology
IoT	:	Internet of Things
IoTPC	:	Internet of Things Privacy Concerns
IERC	:	IoT European Research Council
IUIPC	:	Internet User Information Privacy Concern
MUIPC	:	Mobile User Information Privacy Concern

1. INTRODUCTION

The perception of innovation as an economic-technological process has been the main driver of actions to stimulate industry in more developed countries. Within the various industries existing in a developed economy, the Information and Communication Technologies (ICT) sector has shown itself to be of significant importance. It is more and more frequent to identify economic gains from new products, processes and differentiated business models from sectors related to Information and Communication Technologies. Several factors such as the reduction in the cost of hardware, the growth of data collection and transmission devices, the increase in data processing and storage capacity, combined with the applications of artificial and cognitive intelligence, have contributed to enable the emergence of new technologies and, among them, the Internet of Things (IoT) stands out.

The Internet of Things or IoT is no longer a technological promise. The machine-to-machine connection on a large scale is already a reality in more advanced economies and its expansion occurs at an astonishing speed. Data from Horwitz [1] and Cisco Company [2] indicates that IoT will focus on health and safety equipment monitoring in 2021, and there will be more than 28.5 billion devices connected by 2022. There will be many companies that supply this software and hardware in the IoT and the final consumers will be the biggest holders of this type of technology. IoT is an emerging computing paradigm that integrates the "things" of the real world into the technological world. It is a concept in which our everyday devices and objects are equipped with sensors capable of communicating with each other intelligently without the need for explicit human intervention. A "thing", in the context of IoT, is a connected object that can be, for example, a person with a heart monitor, an industrial tank with level sensors, a car with sensors that warn tire pressure, a light bulb, public lighting of a city, a home outlet, or any other natural or man-made object.

There are many advantages and opportunities that the IoT can offer, however, there is a point that still raises many doubts and concerns both suppliers and consumers: how will the security and privacy of the data be guaranteed?

In a report published by HP [3], 70 % of IoT devices have serious security flaws and are quite susceptible to unauthorized attacks. About 80% of the devices raised privacy concerns. This large percentage is due to the fact that the devices collect some type of personal information, such as name, address, date of birth, health information and even credit card numbers, being even more worrying because these devices are based on applications mobile or cloud services. The study (Figure 1) analyzed the ten most common types of IoT devices, including televisions, webcams, home alarms, home thermostats, among others, all connected to some type of cloud computing, as well as mobile applications that allow their remote control. An average of 25 vulnerabilities were found on each device, totaling 250 vulnerabilities. In total, there are privacy flaws, lack of information transport encryption, insecure Web interface, insufficient authorization, insecure firmware authorization mechanisms and inadequate software protection.

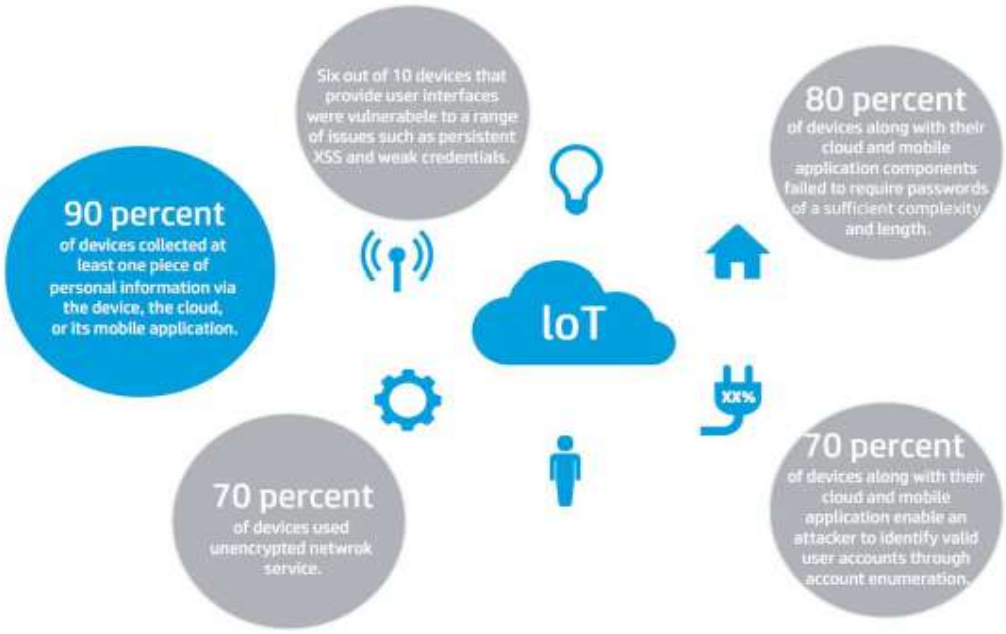


Figure 1: Results of the HP study on IoT [3]

That said, IoT vendor organizations need to substantially improve their security policies and barriers, so that consumers feel that their personal data is secure, and that they only benefit from using IoT.

For Alecrim (2016), the IoT industry needs to define and follow criteria that guarantee the availability of services (including rapid recovery in cases of failures or attacks), protection of

communications (which, in corporate applications, must include strict protocols and audit processes), setting standards for data privacy and confidentiality (no one can access data without proper authorization), integrity (ensuring that data will not be unduly modified), among others. Considering all these aspects is far from a trivial task. In addition to the technological challenges themselves, the industry needs to address each point taking into account global conventions and the legislation of each country. Several market segments already deal with such issues, but this is a work in constant development.

Faced with this reality, HP decided to create the project [4] (Open Web Application Security Project), a nonprofit and internationally recognized entity that contributes to improving the security of IoT device software, gathering important information that allows assess security risks and combat forms of attacks over the internet.

1.1. Framework

The concern with security and privacy is an argument that was, is, and will always be valid in any society, area or technology. IoT is exactly the same. This is seen as the future that will transform our interactions with objects and with our own life. Where a large network of connected objects will act intelligently. However, with this large network of connected devices and the huge data flow it will generate, the security and privacy of data on the IoT will be a fundamental challenge.

According to a study by Gartner [5], the concern with security and privacy are the biggest obstacles to the adoption of IoT. The same study points out that in 2020, 60 % of companies' cybersecurity budgets will be allocated to fast and efficient detection and response approaches. It also provides that, in the same year, at least one organization that supplies IoT products will be held responsible by its national government for vulnerabilities in the level of cybersecurity of its products. This exposes the vulnerability of the system and the importance of security in the IoT.

Although IoT is one of the great buzzwords in the current technological market, 87 % of consumers have no idea what this term means, according to a study by Altimeter [6] on the use of IoT. However, consumers know that organizations are creating more and more connected devices, and that these devices offer a large and unsettling window on personal data and information.

According to the study by Altimeter (2015), consumers are familiarized with the data implications of personal fitness control devices, connected cars, or connected appliances. On average (48 %), a consumer is not comfortable with the idea of sharing his personal data with organizations that

supply IoT devices, and even less does he like the idea of those organizations selling that same personal data to other companies (58 %).

Securing user data is becoming more challenging with improvements in IoT technology. More and more devices are connected, and this increase in connectivity and information sharing results in less control over both data and the devices themselves.

In this scenario, the satisfaction of security and privacy requirements plays a fundamental role. Such requirements include data confidentiality and authentication, access control within the IoT network, privacy and trust between users and "things", as well as the application of security and privacy policies. In addition, the high number of interconnected devices increases scalability problems. That said, a flexible infrastructure is needed, capable of dealing with security threats in such a dynamic environment. This dissertation aims to understand how IoT organizations and consumers are addressing the challenge of data privacy and security. The objective is to give an empirical perspective on this theme, both from the approach of organizations that supply IoT products, from consultants in the area of information technologies, as well as from the end users of these same products.

1.2. Problem Formulation

IoT is an increasingly present future. As a result of the connection between devices, IoT can make people's lives more comfortable and rational. However, in spite of all the benefits that IoT can bring, it still shows a great concern on the part of its users: How will the security and privacy of the data be guaranteed? This is probably the most difficult challenge for IoT.

Suddenly, everything from refrigerators to doors is connected, and while these devices make life easier and people's time management much more effective, they also create new attack vectors for hackers. Day-to-day devices and the connection between them will become even more common than Smartphones are today, and will have access to very sensitive personal data, such as credit cards or personal identification numbers. As the number of IoT devices increases, so do concerns about data privacy and security.

According to a survey by Thomson [7], 20 % of mobile applications used to control IoT devices do not have data encryption. In addition, none of the applications that have been analyzed have mutual authentication between the client and the server, which poses great risks to the user.

The [7] report tested 15 interfaces. Of these, 10 showed vulnerabilities that could, among other situations, allow an attacker to remotely unlock the user's home. For Symantec, hackers who gain access to the home network, for example, breaking into a weakly encrypted Wi-Fi connection, have more attack vectors at their disposal.

For Hernández [8], many other serious situations can occur due to the lack of privacy and security that the IoT presents. An example of the need for security and data privacy in the IoT happens in Smart homes, or smart houses: A network of sensors for domestic lights, if read by hackers can inform some personal aspects of the lives of the people who live in that house. Information such as how often a room is occupied, times when people are at home and when the house is empty. Temperature sensors can create an identical mapping of personal life by informing the time when a person usually takes a shower. The very nature of these devices makes them vulnerable.

Hernandéz [8] also states that IoT devices do not have the security features of traditional information technology (IT) equipment (servers, routers, etc.). The deoxyribonucleic acid (DNA) of current IoT devices is linked to a critical factor: it is essential to guarantee a competitive and low cost for products that are, after all, for the mass market. This is a challenge, as the information generated by the IoT is essential to bring better services and better management of the devices.

1.3. Objectives of Study

In an attempt to address the issue related to data privacy in the IoT, this thesis has the general objective of understanding how some of the suppliers, consultants and consumers of the IoT are addressing the challenge of data privacy and security. Since this theme is one of the biggest challenges of the IoT, its clarification could be an asset for the market, where consumers and suppliers are included, giving a perspective on how this challenge is approached by both.

The specific objectives of this dissertation are:

A. Analyze how some of the IoT vendor organizations are addressing the issues of privacy and security barriers on their devices;
B. Analyze how data privacy in IoT products is seen by analysts in the area;
C. Clarify with your IT consultants, your understanding and vision for data privacy in the IoT.

D. Compare the approach of suppliers in relation to data privacy in their IoT products with the view of analysts and consultants for the same topic.

1.4. Methodology

The methodology is a "set of methods and techniques that guide the elaboration of the scientific investigation process", so what is intended is to describe the methods and techniques used in the development of this dissertation [9].

In the field of organizational studies, several methodological approaches can be used, both quantitative and qualitative. Note, however, that the choice of either type must be associated with the objective of the research and that both have specific characteristics, advantages and disadvantages. On the other hand, it is often possible to use different methods in a combined way, using more than one source for data collection, combining the qualitative and the quantitative. The methodology / language applied must be appropriate to the public for whom it is intended, being more accessible if it is a child audience or more specialized if it is a scholarly audience or knowledgeable of the subject [10].

The definition of methodology for the elaboration of the dissertation went through investigation methods, which can be classified in several ways, however one of the most common distinctions is between qualitative and quantitative research methodologies [11].

1.4.1. Methodology used

Since the dissertation study theme is a complex one, involving mainly interpretation, but also showing a tendency towards quantification, a greater emphasis was therefore placed on mixed methodology. This methodology uses both sense and understanding and also measurement.

In a more detailed way, it is a case study [12, 13], the case study is considered a strategy in which qualitative or quantitative data can be used (fieldwork, files, reports, interviews , observations, etc.), and it is especially suitable when we try to understand, explore or describe complex events and contexts when several scenarios are involved.

Yin [13], states that this approach adapts to research in education, when the researcher is faced with complex situations, in such a way that it makes it difficult to identify the variables considered important, when the researcher seeks to find interactions between relevant factors specific to that entity , when the objective is to describe or analyze the phenomenon, which is directly accessed, in a deep and global way, when the researcher intends to apprehend the dynamics of the phenomenon, the program or the process and when the researcher seeks answers to the “how? " and the “why?”, such as “why is the IoT having a high global impact on society”, or “how are consultants looking at the data privacy of IoT devices?”.

The case study analysis will be a rigorous way of grounding concepts. The use of this method will allow to generalize the theory, which implies the provision of valuable information for the proposed approach.

Thus, this methodology aims to meet the objectives that this dissertation proposes, being:

- Analyze how some of the organizations providing IoT are addressing the issues of privacy and security barriers on their devices;
- Analyze how data privacy in IoT products is seen by analysts in the area;
- Clarify with your IT consultants, your understanding and vision for data privacy in the IoT.
- Compare the approach of suppliers in relation to data privacy in their IoT products with the view of analysts and consultants for the same topic.

1.5. Dissertation Structure

The dissertation is structured by five chapters, in order to allow a better handling / understanding of it.

The **Theoretical Framework** appears in Chapter 2, which aims to frame the theoretical framework on the theme of IoT, which includes its history, everyday applications, the Analytics of Things, examples of communication technologies between IoT devices, the description of alliances created to influence the norms and standards in the IoT, and finally, the legal and social limitations of the IoT, which include data security and protection, data privacy and legislation.

Chapter 3 describes the methodological approach used, in particular the **Case Study** method, that is, all the steps, procedures and strategies carried out for the preparation of the dissertation, and the subsequent activities involved. This includes the research strategy, the design of the questionnaire, and the steps taken to perform the analysis of data privacy documents on the IoT.

The **Activities developed** described in Chapter 4, describe the activities developed during this dissertation. An analysis is made of the results obtained by applying the questionnaires made to organizations involved in the IoT market, as well as a description of the results obtained in the analysis of data privacy documents in the IoT.

The **Conclusions and Perspectives of Future Work** appear in Chapter 5, where the main conclusions of the analysis of the results obtained are presented, as well as the perspectives of future work that may continue the study carried out.

2. LITERATURE REVIEW

In this chapter, a brief introduction to IoT is first made, where the main topics involved in this theme will be described. Emphasis will be placed on IoT, where they are included and their history and applications in everyday life, giving the specific example of Smart homes.

Below is an explanation of the Analytics of Things, and the way it is largely responsible for taking advantage of the potential of IoT.

The following point refers to the types of wireless communication that exist between IoT devices, including, Zigbee, Thread, Wi-Fi, Bluetooth, Zigbee RFID and NFC.

Next, the existing alliances in the IoT market are described to influence norms and standards. Alliances include and AllSeen, Open Interconnect Consortium (OIC), Thread Group, Industrial Internet Consortium (IIC), IEEE P2413 and Apple Homekit.

This chapter ends with an approach to the legal and social limitations surrounding the IoT, including data security and protection, data privacy, and legislation and regulations in Europe and the USA.

2.1. Internet of Things

The Internet has definitely changed our daily lives, allowing quick access to information that we didn't have before. But if the "people's" Internet can be considered a real revolution, the IoT can provide us with much more.

Defining the term Internet of Things (IoT) can be a bit challenging, as the concept undergoes several changes according to the approach adopted. In general terms, the IoT is a new approach in relation to the interconnection of technologies and objects through computer networks, providing the definition of the concept of global device network [1]. However, the IoT depends on the technological process for its continuous evolution, so the IoT cannot be treated as a new disruptive technology, but as a paradigm of computing that is constantly evolving [2].

Figure 2.1 below, exemplifies the concept of IoT, where anything, device or person, can be connected to any network - from anywhere - to use any service in a given context or time. For example, a teacher (anyone) with his Smart Watch (any device), can send an email (any service)

from his classroom (anywhere), via mobile internet (any network), about research grants (any context).

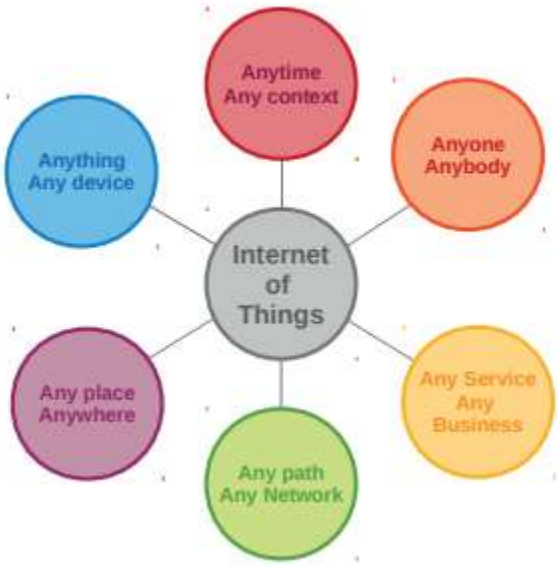


Figure 2.1: Representation of the Internet of Things [4]

Conceptually, the Internet of Things is based on three pillars related to the ability of intelligent objects: (1) the ability to identify itself (any object can self-identify), (2) the ability to communicate (any object can communicate) and (3) ability to interact (any object can interact). Through these three pillars, it is possible to build networks with: interconnected objects, end users and among others [5].

2.2.1. Smart homes

Smart home is a home with smart devices connected together and designed to offer or distribute a range of services inside and outside the home through a variety of network devices. The Smart homes was valued at USD 79.13 billion in 2020 with estimated value at USD 313.95 billion by 2026 and grow at a CAGR of 25.3 % over the forecast period (2021 - 2026) [6]. Although the devices do not require a high-speed Internet connection, to take advantage of all the features of a Smart home, a permanent wide-band Internet connection is required.

A Smart home can truly make life easier for its inhabitants, with facilities ranging from remote control of the air conditioning, lights and washing machine via Smartphone (Figure 2.2). Smart refrigerators can create automatic inventories of all the products they contain and check for the

existence of any expired expiry date, thus notifying the user through the Smartphone. A smart coffee machine can warn you when you run out of water, or when you need cleaning. Intelligent thermostats can regulate the temperature of the house to a pleasant temperature even before its inhabitants arrive home from a day's work. Intelligent light control allows you to schedule the hours when certain lights turn on or off, the level of light emitted, as well as regulating the way the lights react to motion detection.

Figure 2.2: Smart homes [7]

2.2.2. Smart Cities

There are many concepts given to Smart Cities and several are the characteristics that scholars understand as necessary that make a city “Smart”. Smart city is an intelligent community based on efficient, sober and planned growth that makes a conscious effort to use information and communication technologies (ICT), IoT and geographical information systems (GIS) to transform life and work within its territory in a meaningful and fundamental way, instead of following an incremental way” [8]. Traffic and public transport can be much better managed with real-time information on pavement conditions, congestion, weather conditions and availability of parking to be collected from multiple sensors. Likewise, lighting can be much more responsive to the needs of city dwellers, and levels of environmental and noise pollution can be better monitored and communicated. In Smart Cities (Figure 2.3), there is an interaction of data from different sources in different layers. That is, the data generated by personal devices connect to the information

generated by the sensors of a smart city, receiving data on, for example, public transport schedules, or roads with lesser car traffic.

Figure 2.3: Smart City [9]

2.2.3. Wearables and Portables

Some of the current IoT icons are wearables and portables, which include bracelets, glasses or smart watches (Figure 2.4). They can track and record physical activity, such as exercise, eating, sleeping, or other activities, such as reading, etc. Wearables have been highly used as an innovative technology in healthcare, for their ability to continuously record vital statistics and observations in real time such as blood pressure remotely. They can monitor a patient's condition and notify family members, health care providers or emergency services connected to the system of potential risk incidents, such as falls, diet changes, or temperature changes.

Figure 2.4: Wearables (Smart Shirt, Smart Watch, Smart Glasses, and Fitness Tracker) [10]

Wearables and portables are some of the icons of the current IoT. These products naturally collect sensitive user information for their operation, as in the case of fitness trackers, collect information on heart rate, body temperature, calories expended, number of steps taken, etc. This information is subsequently stored on the servers of the device providers (according to the standards of their product privacy policies), is sent to the products' own apps, to health service providers or health insurers, and can also be sent to networks social. This information, although sensitive and confidential, is shared, as shown in Figure 2.5 with a series of channels, calling into question several risk factors.

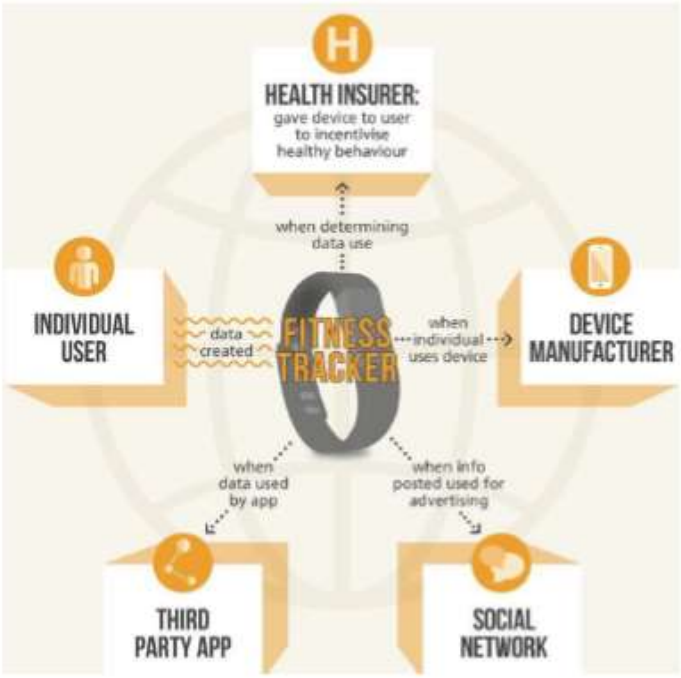


Figure 2.5: Channels of data sharing by a Fitness Tracker [11]

As the IoT evolves, more and more people will have their activities and behaviors recorded and analyzed, so there is an increasing need for suppliers to inform consumers about who stores data on their personal devices, how it is stored, how it is used and to whom it is disclosed. Privacy principles require people to keep control of their data, as well as to be able to choose to leave the "smart" environment without incurring negative consequences.

The full impact of IoT on our privacy can become more evident when its capabilities are combined with other innovations that shape our current and future world, and that control not only our activities, behaviors and preferences, but also our own emotions and thoughts.

2.3. Analytics of Things

With the advent of IoT, all kinds of "things" - from thermostats, cars, even clothes - are able to communicate with each other and generate data, information and actions that promise to transform a wide variety of products and services.

In order to capitalize on this era, it is crucial to understand the Analytics of Things (AoT), and how to combine Machine Intelligence and Human-decision-making. The term AoT thus points out that IoT devices generate a large amount of data, and that the same data needs to be analyzed to be useful (Figure 2.6).

IoT and Big Data are basically two sides of the same coin. Managing and extracting value from data coming from IoT is the biggest challenge that companies face. Organizations must establish an adequate analysis / infrastructure platform to analyze the data coming from the sensors. It must also be borne in mind that not all data are relevant.

An IoT device generates continuous streams of data in a scalable way, so organizations must know how to deal with these streams and except actions on them. These actions can be, for example, correlation of events, calculations of metrics, statistics and descriptive predictions.

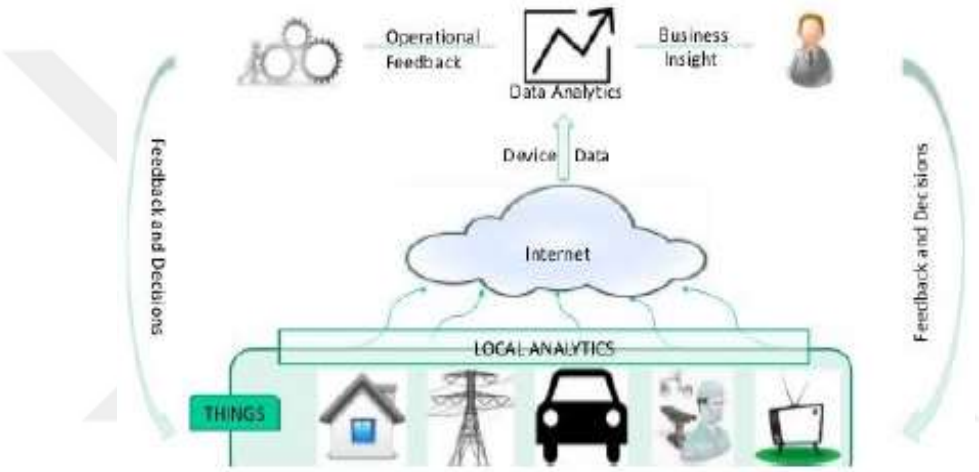


Figure 2.6: AoT architecture [12]

2.3.1. Value and purpose of the Analytics of Things

The analytics and analysis of the IoT appears for a wide range of purposes. The primary virtue of analytics on connected devices is the ability to aggregate data from multiple devices and make temporal comparisons, leading to better decision making by users.

Comparing the use of a resource as important as energy is, for example, a key analytical approach to connected data. For Davenport [13], AoT can be used for several purposes, including:

- **Descriptive analytics** - collect insights about past performance and look at the reasons for success or failure, in order to develop statistical models that explain variations.
- **Detection of anomalies** - identify unusual situations, such as too high a temperature, or an image of a person in an area that must be uninhabited.
- **Predictive analytics** - use predictive analytics to detect possible problems on devices, even before they occur.

- **Optimization** - use data and sensor analysis to optimize a process.
- **Prescriptive analytics** - use sensors and other types of data to tell workers what to do, such as the use of temperature and soil sensors for prescriptive planting by farmers.
- **Situational awareness** - Bring together seemingly disconnected events to form a logical explanation, such as when a series of car oil temperature readings, combined with a drop in fuel efficiency, indicate that an oil change may be necessary.

2.3.2 Evolution of IoT

The Internet has evolved in many ways in the past few years, so it has become capable of connecting billions of devices around the world. These devices are of different sizes and contain computational processing capacity different from each other, in addition to supporting different types of applications [14].

Over the past few decades, the Internet has progressed from a simple statistical repository of hypertext documents linked to a dynamic universe of interactions of humans, machines and network applications [15]. Figure 2.7 exemplifies this evolution.

According to Lemos [16], the term "Internet of Things" was used for the first time at a conference in Procter Gamble (P&G), in 1999, by Kevin Ashton, in a lecture about the use of Radio-Frequency Identification devices (RFIDs) in order to explain that "Things" could be any type of object capable of obtaining information through its own means.

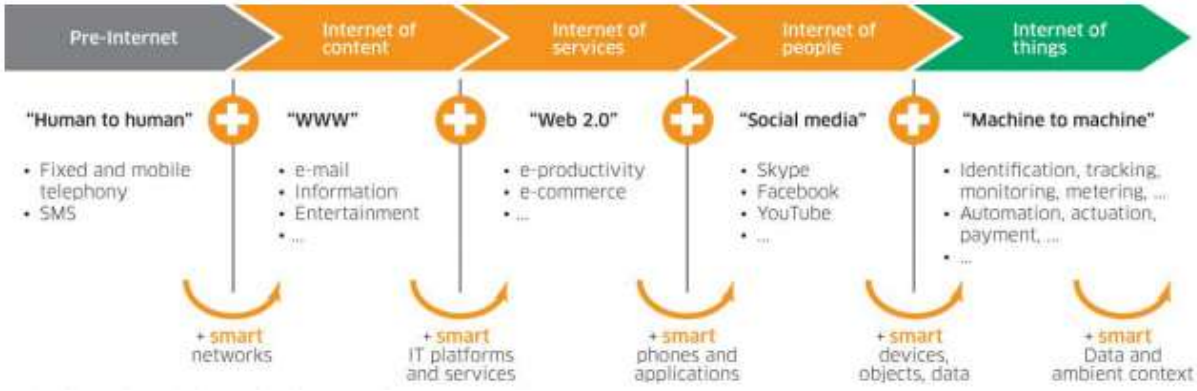


Figure 2.7: The evolution of internet [15]

2.3.3 Architecture

The Internet of Things consists of a three-layer architecture, these being perception, network and application, but some researchers do not consider that these layers fully explain the IoT

architecture and end up adding two more layers to this architecture, the middleware and business [17-20].

- **Perception Layer:** it is the lowest level layer in the IoT architecture, in this, there are devices that perform the collection of information, such as barcode readers, radio frequency identification (RFID) devices, cameras, global positioning system (Global Positioning System - GPS), sensors, among others [17].
- **Network Layer:** this is responsible for transmitting and processing all information, the network layer transfers and processes the information coming from the perception layer. Communication is made over wired or wireless networks, using technologies such as Bluetooth, infrared, ZigBee, among others [17, 19].
- **Middleware layer:** this is responsible for receiving data from the network layer aiming at the management of services and information storage [18, 19].
- **Application Layer:** this is responsible for executing the final presentation of the data. The application layer receives information processed by the middleware or network layer and provides global management of the presented application [18].
- **Business Layer:** this is the highest-level layer in the IoT architecture, being responsible for managing applications and services. It generates graphs, business models, flowcharts, etc., from the data received by the application layer. Based on the analysis of the data, this layer will help to determine future actions and business strategies [19].

2.3.4 IoT Reference Model

The variability of devices provided by IoT brings about several scenarios for smart environments. Moreover, these scenarios are often classified by the entities and by the pattern of information transfer in these environments. The IoT reference model considered in this work was proposed by [2], which were based on the IoT views of the International Telecommunications Union (ITU) and the IoT European Research Council (IERC). This reference model considers that people or objects are interconnected anywhere, anytime, by a network participating in any type of service.

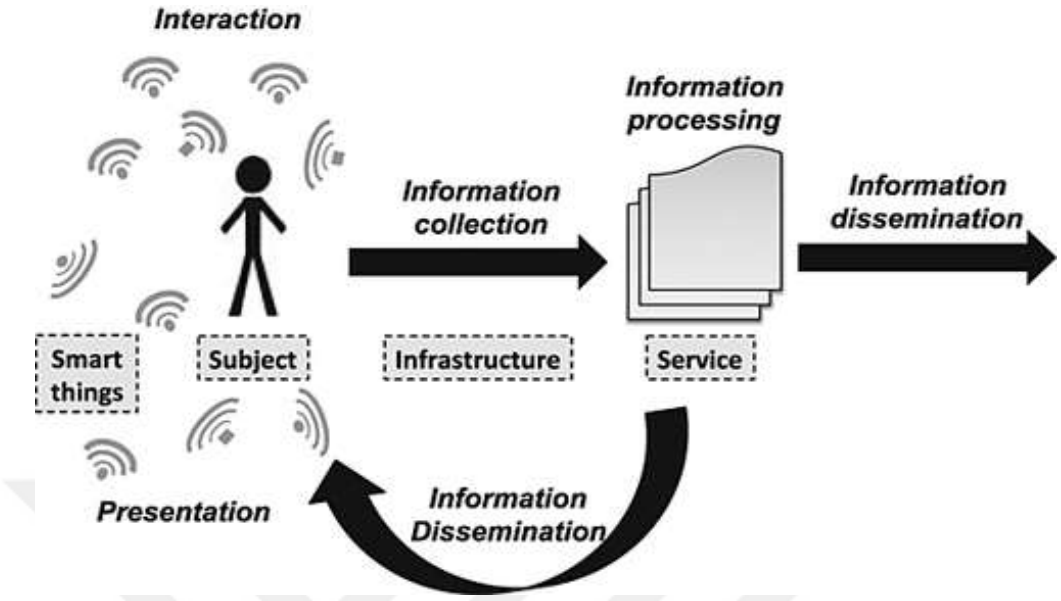


Figure 2.8: IoT Reference Model [2]

Figure 2.8 presents this model, indicating that smart things can be any everyday gadgets or appliances capable of collecting, processing and transmitting information in a given environment. These devices take data about the users who use the IoT infrastructure, obtaining as a final product the dissemination of information to users, as well as to other systems or IoT devices [2]. The use of this model served as a basis for investigating privacy concerns in different IoT scenarios.

2.4 Privacy

Sharing personal information in an uncontrolled manner may eventually lead to breaches of privacy. The greater the number of information made available, the more difficult it will be to control privacy over that information [21].

Warren and Brandeis [22] defined the concept of privacy as the right to be left alone; Westin [23] defines privacy as the individual, groups or institutions' right to determine for themselves when, how and what information about them will be made available to third parties.

According to Hong and Landay [24], privacy can be defined as the users' privilege to determine for themselves when, how and to what extent information about them is made available to third parties. Similarly, Rodota [25] describes the concept of privacy as the possibility of the individual having mechanisms at their disposal that are capable of controlling the use of their information.

Clarke [26] defines privacy as the individual's interest in sustaining a personal space, without interference from another person or organization. In addition to stating, that privacy has several other dimensions, such as:

- Personal Privacy: Also referred to as body privacy, it refers to the integrity of the individual's physical state involving procedures that include compulsory immunization, blood transfusion, compulsory provision of samples of body fluids and body tissue, compulsory sterilization, among others, which should not be carried out without the consent of the interested party.
- Privacy of Personal Behavior: it relates directly to all aspects of behavior, in particular, to more sensitive issues such as preferences, sexual habits, political, religious activities, etc.
- Privacy of Personal Communications: The subject establishes an exchange of information using various communication tools, so that he is not monitored.
- Personal Data Privacy: individuals claim that their personal information should not be automatically available to third parties, even if another individual or organization has access to it. The interested party must have at least substantial control over his or her personal information.

The definition of privacy adopted for this work is the one put forward by Westin [23], presenting the right of the individual, groups or institutions to determine for themselves when, how and what information about them will be made available to third parties.

2.4.1 History of Privacy

Discussions about privacy have been going on for many years. Several historical events have brought about changes in the perspective of privacy concerns [27]. Michael [28] reports that privacy issues can be found from the year 1361, when a judge in England established preventive measures regarding acts of intrusion and curiosity in communications.

In the 19th century, Warren and Brandeis [22] defined the concept of privacy as “the right to be left alone”, motivated in large part by the reporters of the time, who took photographs of people without their consent.

Privacy was discussed again in the 1960s, when it was discovered that the government was using automated data processing to more effectively catalog its citizens. During the Second World War,

already with the practice of Nazism, it was possible to detail public records about the entire population of the invaded cities, thus allowing the Nazis to easily find the entire Jewish people. Faced with this scenario, many European nations started to worry about having data protection laws, in order to avoid any incorrect or abusive use of stored information [27].

Over the years, several efforts have been made to create legislation that would be able to guarantee citizens' privacy preferences. One of those efforts can be highlighted as the American law created in the 1970s, responsible for presenting fair practices of use of information (Fair Information Practice Principles - FIPPs) [29]. At first, the creation of FIPPs were based on the work of Westin [23], who originally:

a) Openness and Transparency: there should be no secret registration practices for individuals, that is, the user must be aware of all the information being collected about him, as well as the content of that information.
b) Individual Participation: this principle guides the existence of mechanisms by which individuals can contest the validity and request data corrections.
c) Collection Limitation: this principle is related to data collection from individuals, reporting that this collection must be proportional to its use, that is, only the necessary information should be collected, thus not carrying out excessive information collection.
d) Data Quality: It is necessary to ensure that the data collected is considered relevant for the specific purposes for which it was collected, in addition to checking, whenever possible, for updates of this information, thus ensuring that the information will always be complete and accurate.
e) Usage Limitation: after the data is properly collected and validated by the previous steps, it is necessary to ensure that this information is used correctly, therefore, this principle aims to verify that the data is being used according to the specific purpose for which it was collected, and if the people who are manipulating this information have authorization for this task.
f) Reasonable Security: security guarantees considered adequate must be implemented according to the sensitivity of the data collected, in order to guarantee the security of the information made available.

g) Responsibility: the last principle concerns the fact that information holders must be responsible for complying with the other principles, that is, since an individual's personal information is in the possession of third parties, they must guarantee the application of all principles presented.

These principles were a starting point in history for the creation of legislation on personal privacy, resulting in the creation of new laws over the years, according to the context, culture and other factors that can influence privacy.

2.4.2 Data Protection Legislation

When analyzing the legal issues at a more current time, it is possible to mention some countries that started to worry about these regulations. Privacy in Europe obtained a legal standard with the general data protection regulation (General Data Protection Regulation - GDPR). This regulation is made up of a set of rules on privacy and data protection aimed at protecting European citizens from breaches of privacy.

Organizations that keep personal data of European users under their control should implement technical and organizational measures to adjust to the privacy regulations, such as maintaining the highest level of privacy in the treatment of data by default, not allowing the treatment of any type of data outside the specified legal context, clearly disclose any type of data collection, its purpose, storage time and whether they will be shared with third parties, among other measures.

Inspired by the European data protection law, the state of California approved, in 2018, California Consumer Privacy (CACP). The law aims to increase citizens' rights in relation to their digital privacy, giving users some rights, such as:

- Right to know all the data collected by the organizations.
- Right to refuse to sell your personal information.
- Right to delete personal data already published.
- Right to know the commercial purpose of data collection.

2.4.3 The USA Data Protection

In November 2013, the Federal Trade Commission, FTC (2013), an independent agency of the US government (United States of America) dedicated to promoting consumer protection, organized a public workshop to explore privacy issues and consumer safety, both at home (including home automation, smart appliances, etc.) and on the street (including fitness devices, autonomous cars, etc.), posed by the growing connectivity of devices. This workshop was attended by students, industry representatives and consumer protection groups. This served to inform the Commission about developments in this area.

After this workshop, in January 2015, the FTC released a 71-page report, which describes the best practices for IoT, in order to meet consumer protection, while still allowing all the benefits of IoT fully realized.

According to Whaley [30], the FTC recommends that device manufacturers take concrete measures to protect the security of information collected through these same devices.

These measures include:

- Think about the safety of the device from the beginning of its construction, and not just an afterthought;
- Train the organization's employees on the critical nature of the information, and ensure that it is managed at a high level within the organization itself;
- Ensure that, when service providers or consultants outside the organization are involved, that they are able to maintain appropriate security measures, as well as correct supervision;
- Monitoring connected devices throughout their life cycle and providing security patches to cover known risks.

Whaley [30]'s report also encourages suppliers to consider "data minimization", limiting consumers' data collection in the first place, and retaining them only for a certain period of time, instead of keeping them indefinitely. This reduces the risk of a company with large volumes of consumer data, becoming a very attractive target for hackers, as well as the risk that consumer data will be used in a way that they do not consent. According to data minimization recommendations,

the FTC intends to be flexible, and organizations can choose between three options: not to store any data; only store data limited to the areas necessary to provide the service offered by the device; or filter the collected data.

The FTC also recommends that suppliers inform consumers about the collection and use they make of their data, as well as providing consumers with choices about how their information will be used, particularly when the volume of data collected is greater than the user can reasonably expect.

The report concludes that any specific legislation on IoT is still very premature, given the rapid evolution of technology.

2.4.4 Turkish Data Protection

The Turkish data protection law (Law No. 6698, of March 24, 2016, of the Civil Code), was also inspired by European Union legislation and aims to create rules for the collection, storage, processing and use of personal information. The law provides for the processing of personal data, including the use of digital media, both for legal entities and individuals, aiming to protect the fundamental rights of freedom and privacy.

Among the various items of this new legislation, some points are highlighted:

i. Data Processing: according to the regulations, the processing of personal information can only be carried out with the consent of its owner or for specific purposes, such as compliance with legal obligations, public administration, carrying out studies by research organizations, among others.
ii. Security and Good Practices: the law states that agents responsible for the collection and processing of private data must pay attention to technical and administrative security policies in order to guarantee the protection of these private data against use by unauthorized persons and use in accidental situations.
iii. National Data Protection Authority: this authority is a competent body integrated with the indirect federal public administration, with the task of taking care of the protection of personal data under the terms of the legislation, developing guidelines for the National Policy for the Protection of Personal Data and Privacy, to supervise and apply punishments in cases of non-compliance with the law, among others.

iv. Personal Data Protection Authority and Organization: the council aims to propose strategic guidelines, in addition to providing resources for the elaboration of the National Policy for the Protection of Personal Data and Privacy. It is also your responsibility to suggest actions to the inspection authorities, as well as to carry out studies on the subject. The council is composed of the board and the presidency. The board is composed of 9 representatives drawn from diverse areas. Five members of the Board of Grand National Assembly of Turkey is selected by the four members of the President. Presidency consists of the Vice President and service units.

The law requires companies and organizations to adapt to its principles within a maximum period of up to 18 months from the sanctioned law, thus ensuring that all personal information collected, processed and made available complies with the imposed guidelines.

2.5. Synthesis

By reading the theoretical framework, it is possible to understand the concepts associated with the main theme of the dissertation, data privacy in the IoT, and subsequently the activities developed, which aim to give an understanding of how suppliers, consultants and consumers of the IoT are addressing the challenge of data privacy and security.

The first part of the framework refers to the theme of IoT, which includes examples of its use and its purpose. Then, the explanation of the types of communication existing between the IoT devices is made, with the objective of giving an understanding of the way these objects communicate and "talk" to each other, with their advantages and adjacent security and privacy dangers. The review of currently existing norms and standards highlights the way in which organizations that supply IoT products are positioning themselves in the market in view of their business and competitive strategies. Then, the legal and social limitations reveal the two main barriers to a greater adoption of IoT by consumers, which are the security and privacy of their data, which are also the biggest challenges for suppliers. Finally, the legislation in force in Europe and the USA allows us to give an idea of how the theme of IoT is being approached by European and North American governments.

2.6. Final considerations

In the next chapter, concepts considered fundamental for this work were presented. In this sense, we sought to absorb an in-depth knowledge about privacy, in addition to acquiring general knowledge about the scenario studied, Internet of Things. It is worth noting that the concepts presented in this chapter added the possibility of clearly identifying and exploring privacy problems in IoT environments.

3. METHODOLOGY

3.0 METHODOLOGICAL STRATEGY

This chapter aims to refer to the existing methodological strategies, and to what type of strategy used for the elaboration of the thesis.

The methodological strategy should be seen as a set of methods or rules that aim to delineate a certain path that leads to an end, in a process of knowledge transmission, directed to the target audience and understood by the recipients, applying specific research processes of to lead to adequate and credible results.

3.1. Qualitative methodology

Qualitative research, used to interpret phenomena, occurs through the constant interaction between observation and conceptual formulation, between empirical research and theoretical development, between perception and explanation [31, 32]. The qualitative method is useful and necessary to identify and explore the meanings of the studied phenomena and the interactions they establish, thus enabling to stimulate the development of new understandings about the variety and depth of social phenomena [33].

A qualitative research refers to a generic term to group several research strategies that share certain characteristics. In this type of investigation, the data collected are called qualitative, which means rich in descriptive phenomena in relation to people, places and conversations, and complex statistical treatment. The questions to be investigated are not established through the operationalization of variables but are rather formulated with the aim of studying phenomena in all their complexity in a natural context.

Qualitative methods suggest that the researcher is in the field work, observes, makes value judgments and analyzes. In qualitative research, it is essential that the investigator's interpretive capacity never loses contact with the development of the event. Another characteristic aspect of qualitative research is that it directs aspects of the investigation towards cases or phenomena in which the contextual conditions are not known or are not controlled.

Some examples of the methods to be used in this dissertation are research actions, case studies, among others, and their source of evidence includes observation, interviews, documents and texts.

3.2. Quantitative methodology

Quantitative research is characterized by its performance at the levels of reality and aims to identify and present data, indicators and observable trends. This type of investigation is generally appropriate when there is the possibility of collecting quantifiable measures of variables and inferences from samples of a population. This investigation uses numerical measures to test hypotheses, through rigorous data collection, or looks for numerical patterns related to everyday concepts. In a later stage, the data are subject to statistical analysis, through mathematical models (or own software), in order to test the raised hypostasis. As such, its use is generally linked to experimental or quasi-experimental research.

In organizational studies, quantitative research allows the measurement of opinions, reactions, habits and attitudes in a universe, through a sample that represents it statistically. It can be said that a cause-effect relationship is then established and phenomena are predicted.

Stake [34] points out that in the usual quantitative models, the researcher exerts an effort to limit his / her personal interpretation function, from the beginning of the research design until the data is statistically analyzed. It is a period that must be guided by the absence of values. In quantitative research, the questions look for the relationship between a small number of variables. The effort goes towards the operationalization of these variables and to reduce the effect of interpretation to a minimum, until the data are analyzed. Here it is important that the interpretation does not change the direction of the investigation.

The main methods of quantitative research are survey, correlational, causal-comparative and experimental. Action research, historical research, case study, focus group, ethnography and grounded theory are recognized as the main forms of qualitative approach.

3.3. Research Strategy

The most central feature of qualitative research, in contrast to quantitative research, is that it puts the emphasis on the individual perspective of what is being studied. Describing the complexity of

the problem, one can contribute to the process of change, enabling the understanding of various peculiarities.

It appears that the Case Study is an investigation method that can be applied to both the quantitative and qualitative approaches. This is because, this is the methodology that should be used to understand, explore or describe complex events and contexts.

Many researchers seek answers to "how?" and “why”, when the researcher intends to relate factors of a certain entity, when the researcher intends to analyze, describe, or know the dynamics of the phenomenon, program or process under study. Thus, the selection of this method can be justified on the basis of Yin [35]: “case studies are preferable when the proposed questions are ‘how’ or ‘why’, when the researcher has little control over the events and when the question is centered on a contemporary phenomenon with a real context”.

The three principles that contribute to good data collection are: (i) Use multiple sources of evidence; (ii) build, throughout the study, a database; (iii) forming an evidence chain.

Based on these premises, the research question posed “At IoT, how do suppliers and consumers of products approach data privacy?” it is an explanatory question that intends to explore the topic of data privacy in the IoT. This objective presupposes the collection of varied information; however, it is important to emphasize that the researcher has no influence on the data, proceeding only to its collection and analysis.

It is through the literature review that it is intended to understand and explore the existing articles as well as to define some concepts that will help a better understanding of the dissertation.

The definition of the IoT concept, the contextualization of its applications in the market, the issues surrounding legal and social limitations, as well as the exploration of communication between devices and the existing norms and standards, allow a more comprehensive context in relation to the objectives to which this dissertation is proposed.

This thesis aims to understand how IoT supplier organizations, consultants, and end users are addressing the challenge of data privacy and security. The sources of data collection that can be used in a case study are usually of three types: (1) Surveys (Questionnaires and interviews), (2) various documents and (3) through observation.

Bearing this premise in mind, the study methodology was based, in a first phase, on the investigation of the theme through questionnaires addressed to suppliers of IoT products and technologies.

In a second phase, using the Contextual Analysis tool, patterns, relationships and trends were identified in information obtained from documents in the area.

3.4. Questionnaire Design

The use of qualitative models suggests that the researcher is in the fieldwork, observes, makes value judgments and analyzes. In qualitative research, it is essential that the investigator's interpretive capacity never lose contact with the development of the event.

3.4.1. Scope of investigation

The first operational concern related to completing the questionnaire was directed to the conception and reliability of the questions, in confirming a correct approach to the theme and the national framework.

Thus, the careful choice of questions was based on the following criteria:

- Framework with the area of activity of organizations, in the field of IoT
    a) Why the IoT is having a high global impact on society;
    b) In what way is the behavior of organizations in the IoT.
- Framework with the national panorama
    a) How is the development of IoT in Turkey going?
- Framework with the main theme of the dissertation, data privacy
    a) What is the vision of organizations for data privacy;
    b) How is the process of adapting the privacy of users' data going on when organizations are developing IoT products?

Bearing in mind the fact that the questions in the questionnaire are open-ended, it will be the responsibility of the researcher to analyze them and draw conclusions about them.

3.5 Privacy Measurement Instrument in IoT Environments

The development process of the instrument application scenario is presented. Here, the description of the proposed instrument is exemplified and the perception of privacy of IoT users is defined in accordance with the proposed Internet of Things Privacy Concerns (IoTPC) instrument.

This work aimed to outline the construction of an instrument capable of measuring privacy through a new scale of privacy concerns for IoT environments. This instrument will promote cooperative research efforts, allowing other researchers to use it to carry out tests and adjustments to their research, as well as more clarity in the formulation and interpretation of research questions. This work also describes the elaboration of a privacy inference module for privacy negotiation mechanisms in IoT environments that exemplifies one of the possible applications of the instrument.

3.5.1. Instrument Application Scenario

Since several instances are likely to be seen in IoT makes it nearly impossible to develop an instrument capable of covering all likely situations. To build a scenario for the conception and application of the proposed instrument, some studies related to the current literature were observed. Among them, Lee and Kosha [36] who examined the factors that affect security of IoT users. In this study, instances were modeled to illustrate and collate those factors. From the proposed scenarios, five parameters were used in a different way:" Where"," What"," Who"," Reason" and" Persistence". "Where" refers to the place where the information is collected; "What" is the type of information collected;" Who" is the agent responsible for collecting the information; "Reason" is the purpose of collecting such information; and "Persistence" is the frequency with which this information is collected.

Considering these parameters proposed by Lee and Kobsa [36], the instrument's scope of action in this work was delimited so that there was better accuracy. To this end, 64 micro-scenarios of IoT were analyzed considering the research of [36], and 25 scenarios were selected, considering the aspects of context coverage and heterogeneity of devices to compose a general futuristic scenario about IoT, which served as the basis for the construction of the instrument's items. The futuristic scenario presented below in Table 3.1 as an example mirrors daily IoT usage scenario for a student.

Table 3.1: Micro scenarios used to create the futuristic general scenario.

Code	Scenario
1	Your bedroom curtains want access to your alarm times to open when you wake up
2	Your bed wants to collect information about your sleep
3	The cafeteria requests data from your cell phone to prepare coffee according to your taste preferences
4	The refrigerator wants access to your bank details so that it can buy food as soon as it is out of stock
5	A sports store wants access to your favorite sports in order to advertise products for you
6	Your cell phone wants information about the last time you visited your doctor
7	Your doctor wants access to your health data to monitor your health
8	The gym wants to collect information about your measurements and weight
9	The academy wants your music preferences to play only music you like
10	Your phone wants access to your calendar to adjust the ring level according to your tasks
11	Your city wants access to your cell phone to notify you of changes in weather
12	Your backpack asks for access to the items you carry in it daily, so it can let you know when you forget something
13	Your home air conditioner wants to know your schedule for the day to adjust the temperature accordingly and save energy
14	Your car wants access to your destination to calculate the most economical route
15	Your car wants information about its itinerary to determine if it has the range for the route
16	A gas station wants information about your car's consumption
17	The gas station air compressor wants information about your car model to determine the proper pressure from the caliper
18	Your entertainment system wants access to your most recent news searches to select relevant programming
19	The university restaurant wants information about your food preferences

20	Your department wants information on dates of work and exams
21	The water cooler wants to collect information about how often you use it
22	Your chair wants to know your height and weight, for a better fit and according to your physical conditions
23	Your chair in the office wants to know your identity to keep track of your weight
24	Your office air conditioner wants to know your temperature preferences to adjust the temperature
25	Your classroom wants to collect facial expressions from you during lessons

This scenario was used in the design of the instrument described in the next section. Therefore, the IoTPC was built to act in the scenario presented, ensuring better accuracy of the instrument, but not limiting its performance strictly to the proposed scenario.

3.5.2. Description of the Instrument

The Internet of Things Privacy Concerns (IoTPC) instrument is capable of empirically measuring users' privacy in the narrated IoT futuristic scenario. The IoTPC is composed of 17 items distributed within five dimensions and was developed based on already existing privacy instruments such as the IUIPC, MUIPC and AIPC, all based on the CFIP, as shown in Figure 3.1.

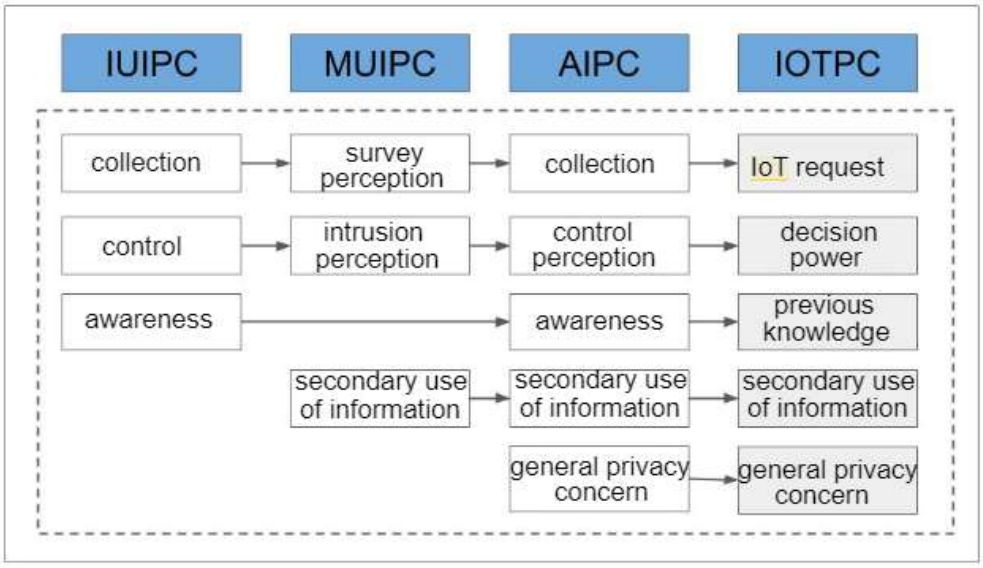


Figure 3.1: IoTPC Development

For the construction of the instrument, it was necessary first, to examine and clearly define its dimensions so that, later, its items were well documented. During their development process, the dimensions of IUIPC, MUIPC and AIPC were chosen as the starting points, Figure 3.2 illustrates these aggregations.

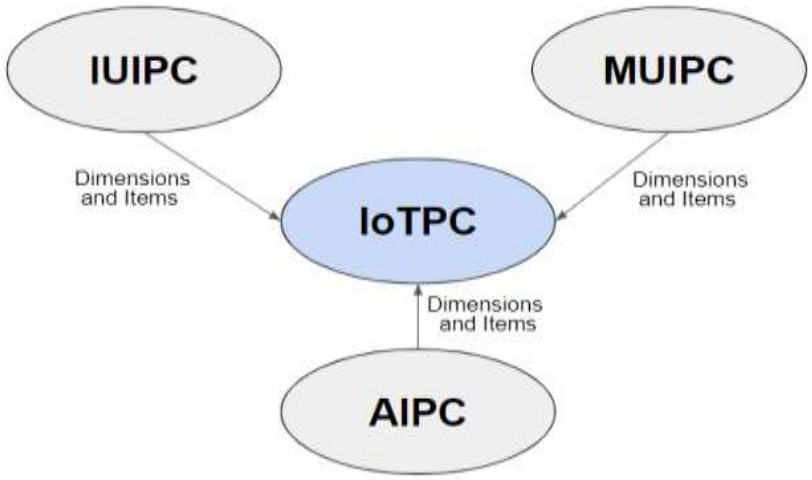


Figure 3.2: Aggregations of Selected Instruments

3.5.3. Dimensions of the IoTPC

The dimensions of the IUIPC, MUIPC and AIPC were analyzed in order to verify if they reflected some characteristic of privacy in the IoT environments, such as the dimension of secondary use of information, aggregated from the MUIPC. This dimension reveals that, normally, users believe their security is compromised when third parties use their data without their permission.

It, too, may reflect privacy concerns in IoT environments, as devices collect various data and may possibly share those data with third parties.

The dimensions adopted by the IoTPC are described below and were designed based on the dimensions of the IUIPC, MUIPC and AIPC.

1. **IoT requests:** this dimension of the instrument is defined by the requests made by IoT devices to their users, that is, the collection of personal data carried out by the devices. Malhotra et. al. [37] define that the concept of collection is linked to the degree that a given individual is concerned with the amount of personal data held by third parties. However, the heterogeneity of possible devices within an IoT scenario and the fact that the vast majority of these devices ubiquitously collect information from their users highlight the concept of surveillance that is also incorporated into this dimension.

   Peissl et al. [38] define surveillance as the act of observing an individual's listening or recording. In a futuristic IoT environment, devices may eventually use the collected information to track and watch over their users.
2. **Decision Power:** This dimension is defined as the interest of users in controlling their personal data in IoT environments. According to Malhotra et al. [37], the concept of control can be defined as the fact that the individual has an interest in controlling or, at least, significantly influencing the use of their personal data. Oftentimes, in an IoT environment, users are uncomfortable with not having one hundred percent control over their private data, as well as a sense of infringing on their privacy preferences.

   According to Peissl et al. [38], perceived intrusion can be defined as an invasive act that disturbs tranquility or loneliness. The concept of perceived intrusion also applies to the IoT scenario, since its devices may request information from its users, eventually disturbing them or even removing them from your tranquility.
3. **Previous Knowledge:** This dimension is related to all the knowledge about privacy practices that the IoT user already carries with them. It is possible, here, to use the concept of awareness to define this dimension, in which awareness is the degree to which an individual is concerned with their knowledge of privacy practices related to organizational information [37].
4. **Secondary Use of Information:** sometimes the data provided by users for a particular purpose are used differently without their consent, for example, by profiling individuals and sending marketing messages to them, without their authorization [39]. This practice reflects the privacy concerns in IoT environments and was also taken into account when constructing the dimensions of the IoTPC. Secondary use of information is a privacy

dimension that is beyond the scope of the user's decision-making power, so these two dimensions cannot be confused.

5. **General Privacy Concerns:** Responsible for representing items such as “I am concerned about the threats to my personal privacy today” and “Compared to others, I am more sensitive about how IoT devices handle my personal information”, this dimension aims to capture the individual's overall privacy concerns in an IoT scenario [40].

3.5.4. IoTPC Items

Using the same strategy to create the dimensions of the IoTPC, the items related to each dimension were also generated based on the aforementioned instruments, undergoing changes for the context of the IoT. The definition of such items is considered the instrument design methodology, which were evaluated in order to verify if they reflected some privacy feature in IoT environments.

The IoTPC items were described by a code, thus, aiming to give a unique identification for each one. The code was composed of three pieces of information, namely: (I) Instrument of Origin, (II) Dimension of Origin and (III) Item Numbering. To exemplify which instrument contributed to the construction of that item, the first three characters of the code represent the (I) Instrument of Origin, for example, if the instrument that gave rise to that item was extracted from the Measuring Mobile Users’ Concerns for Information Privacy (MUIPC), the first three characters of the code of this item were formed as "MUI". The second part of the code identification is related to the (II) Dimension of Origin, being represented by: Control (Co), Awareness (Aw), (Collection (Coll), (Perceived Surveillance (Ps), Perceived Intrusion (Pi), Secondary Use of Information (Sui), General Privacy Concerns Information (Gen). Finally, in the last part of the code, a number from 1 to 17 is assigned, for better ordering of items within the scale. Table 3.2 presents the items designed for this instrument.

Table 3.2: IoTPC items

Item	Scenario
IUICo1	Privacy in IoT environments is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used and shared.

IUIAw2	IoT devices that seek information must disclose how data is collected, processed and used.
IUIAw3	It is very important to me that I am aware of and knowledgeable about how my personal information will be used.
IUIColl4	It usually bothers me when IoT devices ask me for personal information.
IUIColl5	When IoT devices ask me for personal information, I sometimes think twice before giving it.
IUIColl6	It bothers me to provide personal information to so many IoT devices.
MUIPs7	I believe my mobile device's location is monitored at least sometime.
MUIPs8	I'm worried that IoT devices are collecting too much information about me.
MUIPs9	I'm concerned that IoT devices might monitor my activities through my mobile device.
MUIPi10	I feel that as a result of using IoT devices, others know about me more than I would like to, causing me discomfort.
MUPi11	I believe that as a result of using IoT devices, information about myself that I consider private is now accessible to others more than I would like.
MUISui12	I am concerned that IoT devices may use my personal information for other purposes without notifying me or obtaining my authorization.
MUISui13	When providing personal information to use IoT environments, I am concerned that IoT devices may use my information for other purposes.
MUISui14	I am concerned that IoT devices may share my personal information with other entities without my authorization.
AIPGen15	Compared to other people, I'm more sensitive about how IoT devices handle my personal information.
AIPGen16	For me, the most important thing is to keep my privacy intact when using IoT devices.
AIPGen17	I am concerned about the threats to my personal privacy today.

The organization of the items in relation to the dimensions of the IoTPC was as follows: the Decision Power dimension encompasses the items IUICo1, MUIPi10 and MUIPi11. In the Prior Knowledge dimension, the items it contains are IUIAw2 and IUIAw3. The IoT Requests

dimension contains the items IUIColl4, IUIColl5, IUIColl6, MUIPs7, MUIPs8 and MUIPs9. The Secondary Use contains the items MUISui12, MUISui13 and MUISui14. Finally, the General Concerns dimension contains the items AIPGen15, AIPGen16 and AIPGen17. Table 3.3 illustrates this organization.

Table 3.3: Organization of IoTPC Items by Dimensions

Dimensions	Items
IoT requests	IUIColl(4,5 and 6), MUIPs7, MUIPs8 and MUIPs9
Decision power	IUICo1, MUIPi10 and MUIPi11
Previous knowledge	IUIAw2 and IUIAw3
Secondary Use of Information	MUISui12, MUISui13 and MUISui14
General Privacy Concerns	AIPGen15, AIPGen16 and AIPGen17

3.6. Design of the IoT Privacy Concern (IoTPC) Inference Learning Module

3.6.1. Initial considerations

Here, the construction of the IoTPC Learning inference module and the results obtained with its evaluation are presented. In section 3.6.2, the inference module IoTPC Learning is described, as well as its conversion rules, while section 3.6.3 presents the techniques applied.

3.6.2. IoTPC Learning Inference Module

In order to use the IoTPC to assist in a privacy preservation mechanism in IoT environments, an IoT scenarios inference module called IoTPC Learning was built. Next, the construction process of this module is described, reporting its development, as well as the techniques used.

In the context of inference modules, conversion rules were created aiming to transpose the answers given by the research participants into binary answers, with this, it was possible to perform the inference of IoT scenarios through the IoTPC. The information and services requested by the scenarios used in the construction of the IoTPC were classified as follows: (**I**) Critical Information

- CI, (**II**) Simple Information - SI, (**III**) Critical Service - CS and (**IV**) Simple Service - SS. An overview of the conversion rules can be seen in Table 3.3.

Table 3.3: Overview of the conversion rules

	Critical Information (CI)	**Simple Information (SI)**	**Critical Service (CS)**	**Simple Service (SS)**
Completely agree	No	No	No	No
Partially agree	No	No	No	No
I don't know how	No	Yes	Yes	No
Partially Disagree	Yes	Yes	Yes	Yes
Completely disagree	Yes	Yes	Yes	Yes

Critical Information (CI): refers to scenarios that want to collect sensitive data; –documents, bank records among others –, as in:" The refrigerator wishes to have access to your bank data so that it can buy food as soon as it is out of stock".

Simple Information (SI): refers to the collection of less sensitive information, such as music preferences, food preferences, among others, such as the scenario "Your office/school's air conditioning want to know your temperature preferences to adjust the temperature".

Critical Service (CS): Similar to the classification of information, some of the IoT scenarios, in addition to informing the type of information collected, also indicate the purpose of this collection.

SC addresses scenarios whose purpose generates savings in resources, such as fuel, energy, among others, for example: "The thermostat in your house wants to know your schedule for the day to adjust the temperature accordingly and save energy."

Simple Service (SS): This on the other hand, refers to less sensitive services, relating to scenarios that use user information for less important tasks, such as "The gym wants your preferences musicals to play only songs of your choice".

The answers given by the users were associated with an output ”Yes” or ”No”, providing or not the requested information, determining the inference results for the scenario linked to that item of the instrument.

In some cases, a single IoTPC item may reflect the inferential response for more than one IoT scenario, and it is possible to obtain more than one service rating or information for those scenarios. Some of the instrument's items reflect general privacy characteristics, which is why they were associated with all 25 IoT scenarios used in the construction of the IoTPC. An overview of the classification of scenarios according to the type of service or information can be seen in Table 3.4.

Table 3.4: Classification of Scenarios According to the Type of Service or Information

Items	Construction Dimensions	Extracted Factors	Scenario	Type of Information or Service
1*	Decision power	Caution	1 to 25	SI, CI, SS, CS
2	Previous knowledge	Caution	1	SS
3*	Previous knowledge	Caution	1 to 25	SI, CI, SS, CS
4	IoT requests	IoT requests	2	SI
5	IoT requests	IoT requests	4	CI
6	IoT requests	IoT requests	3, 11, 12, 13, 24	SS, CS
7	IoT requests	Decision power	1 to 25	SI, CI, SS, CS
8	IoT requests	IoT requests	8, 9	SS
9	IoT requests	IoT requests	5, 10	SS
10	Decision power	Decision power	6, 7, 18	SS, CS
11	Decision power	Decision power	14, 15, 16, 17	SS, CS
12	Secondary Use	Caution	25	SS
13	Secondary Use	Caution	21, 22, 23	SI
14	Secondary Use	Caution	19, 20	SI
15*	General Concerns	Decision power	1 to 25	SI, CI, SS, CS
16*	General Concerns	Decision power	1 to 25	SI, CI, SS, CS

17*	General Concerns	Decision power	1 to 25	SI, CI, SS, CS

3.6.3. Techniques Applied

The first step in the construction of IoTPC Learning was to generate a database with the inference results obtained from the conversion rules, which led to the choice of a machine learning algorithm that was able to infer the responses to these scenarios according to the privacy preferences given by the IoTPC.

The decision tree was selected in this work as a learning model due to the small size of the database and the fact that most attributes are descriptive, containing only two possible outputs, yes or no.

A decision tree is defined as a classification procedure that aims to repetitively partition a set of data into smaller subdivisions, based on a set of tests defined in each branch of the tree [41].

A graphical example of the decision tree can be seen in Figure 3.3, where it is exemplified when a tennis player must go out to play according to the climate. Note that, on the left branch of the tree, when it is sunny and there is high humidity, he should not go out to play, however, when it is sunny and the humidity is normal, he can practice his sport normally. And similarly, the rules extend to cloudy and rainy weather, as shown by the central and far right branches of the figure.

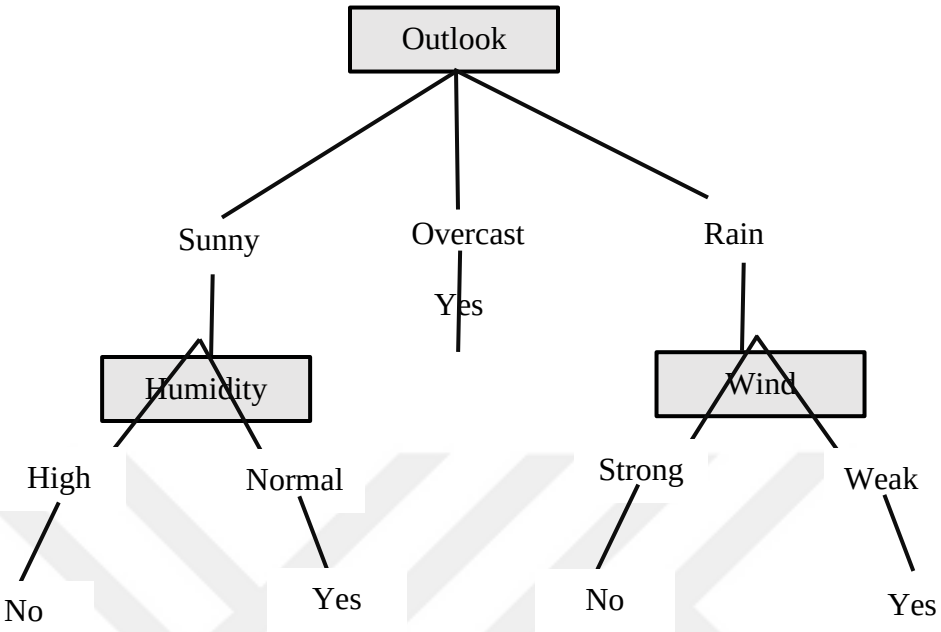


Figure 3.3: Decision Tree for Playing Tennis [42]

The Weka machine learning software [43] was used to generate the learning models, however, some precautions and adaptations were performed before generating the decision trees. The pruning technique was applied to avoid overfitting [44], in addition to a resampling filter, in order to balance the classes. Finally, cross-validation with ten folds was used for training and testing the database [44].

For each of the 25 scenarios, a corresponding decision tree was generated, whose inference was based only on the users' personal attributes, without allowing the other 24 scenarios to influence the decision making.

4. RESULTS AND DISCUSSION

This chapter describes all the activities developed for the elaboration of the dissertation, which include the analysis of the results obtained in the questionnaires submitted to various people present in the IoT usage, as well as the description of the results obtained by the analysis of the data privacy document in the IoT industry.

4.1. Questionnaires Results Obtained

The first stage of the activities developed in the dissertation was based on the elaboration and presentation of questionnaires with questions about the IoT and the respective data privacy, to several supplier organizations in the area.

The three fundamental moments during the data analysis phase: description, analysis and interpretation. The description corresponds to the writing of texts resulting from the original data registered by the researcher. Analysis is a process of organizing data, where essential aspects must be highlighted, and key factors identified. Finally, interpretation concerns the process of obtaining meanings and inferences from the obtained data.

In the same order of ideas, the analysis model in qualitative research that consists of three stages: data reduction, data presentation and conclusions and verification. Data reduction concerns the process of selecting, simplifying and organizing all the data obtained during the investigation. The presentation of data refers to the moment when the information is organized and compacted so that the researcher can quickly and effectively see what is happening in the study. The third and last moment corresponds to drawing conclusions from all the information collected, organized and compacted, which depends on the amount of notes taken, the methods used and, mainly, the researcher's experience in this field. The present research study, it can be mentioned that:

- The first phase of data reduction, concerns the simplification and organization of each answer to the questionnaire, in the case of companies that submitted the completed

questionnaire, of simplification and organization of the information referring to the documentation provided by the other companies;

- The second phase of data presentation, refers to the compacting and summarizing of the questionnaire responses, as well as summarizing the documentation received and filling in the questionnaires referring to companies that could not fill it out;
- The last phase, the extraction of conclusions, concerns the analysis of the responses to the questionnaires, by the investigator, in order to draw conclusions and patterns about them.

4.2 Evaluation of IoTPC Results

In this section, the results obtained with the evaluation of the proposed instrument are described in detail. In section 4.2.1, the results of the IoTPC assessment related to the study are presented and, finally, in section 4.2.2, the results of the IoTPC assessment pertinent to the study are broken down.

The instruments found in the literature served as the basis for this work. It used the generic items in their constructions. Due to this characteristic in common between the studied instruments, two versions of the IoTPC were developed and evaluated, thus, aiming to verify which of the two versions would generate better results – one that produced generic items in its construction and the version of the instrument that addressed the context studied directly in its items.

4.2.1 The Study

In this study [45], the IoTPC items remained generic, as stated in section 3.5.4, aiming to observe whether the evaluation results obtained with the construction of a generic instrument would obtain better results when compared to an instrument with more specific items.

4.2.1.1. Population Data

The location chosen for the execution of this study was the Informatics Laboratories belonging to the Altınbaş Üniversitesi, Istanbul, Turkey. A sample of 165 students was taken by the researcher, these being undergraduate and graduate students. The age of the participants who carried out the research ranged from 15 years (1.1 %) to 51 years (0.6 %), with the majority of the sample being aged 20 years (18 %) and the others 32 and 33 years old, representing 1.6% of the total age. Figure 4.1 exemplifies this information.

Figure 4.1: Age of Participants

The male population was 88 participants (53.3 %) against 72 participants (43.6%) of the female population while 5 (3 %) prefer not to say. Figure 4.2 exemplifies this information.

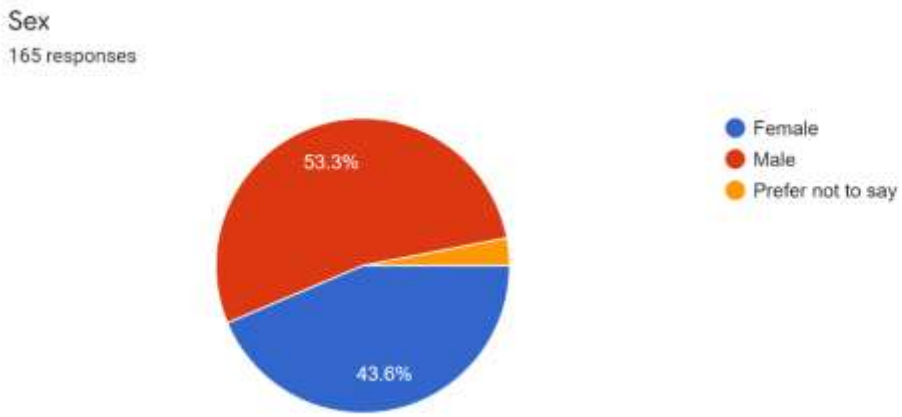
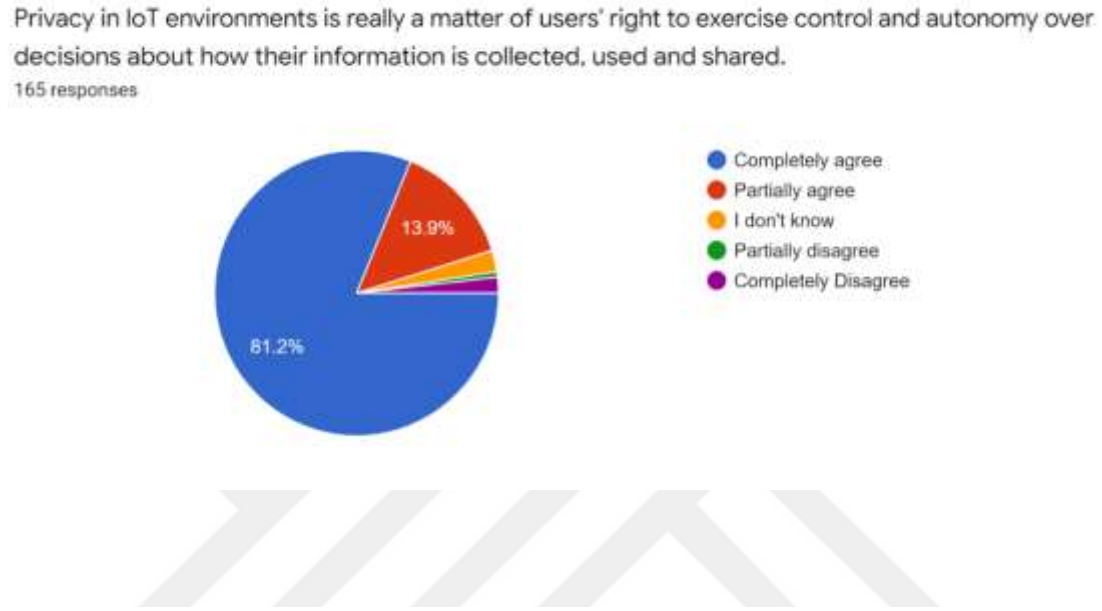
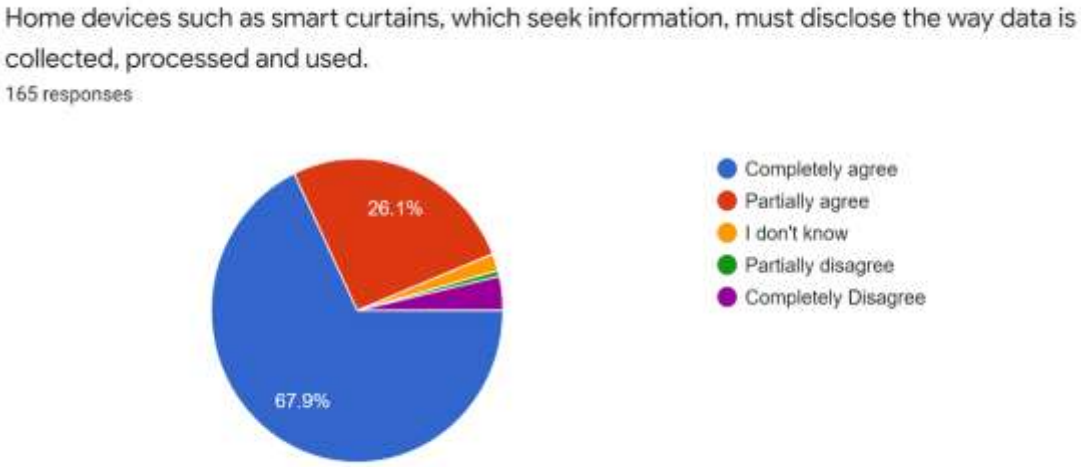


Figure 4.2: Sex of Participants

4.2.1.2 Other Evaluated Parameters

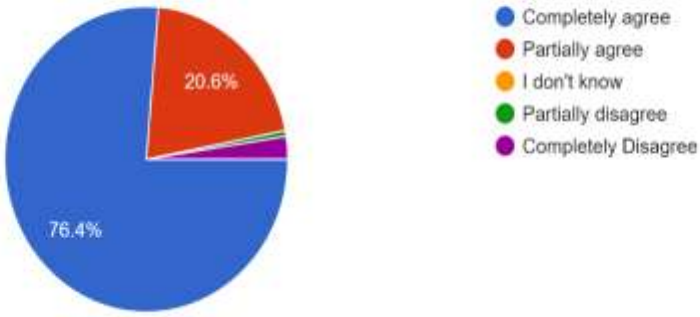
These following parameters are evaluated to complete the IoTPC Leaning module.





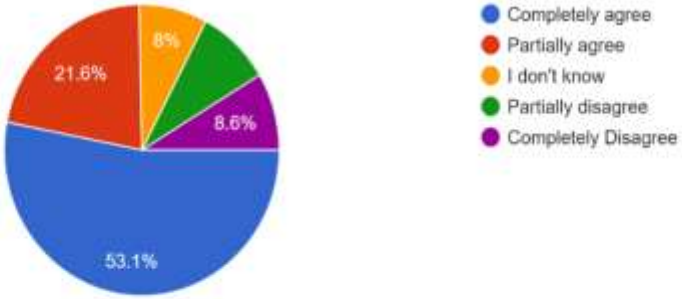
It is very important to me that I am aware and knowledgeable about how my personal information will be used

165 responses



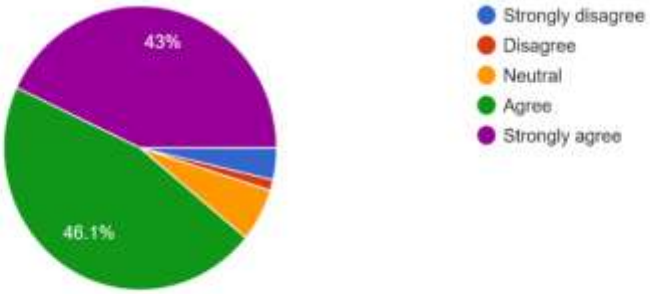
It usually bothers me when my bed asks for information about my sleep

162 responses



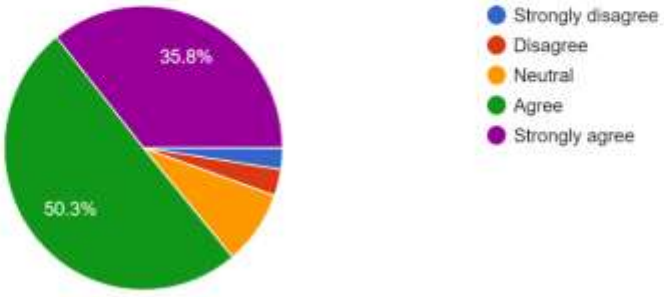
When devices like my refrigerator ask for my bank information, I usually think twice before providing them

165 responses



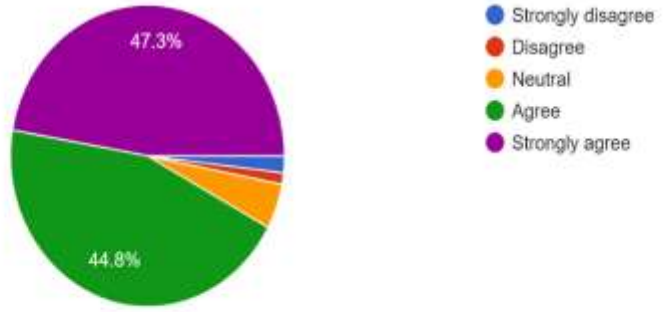
It bothers me to provide personal information to devices such as coffee machines, air conditioners, thermostats, among others.

165 responses



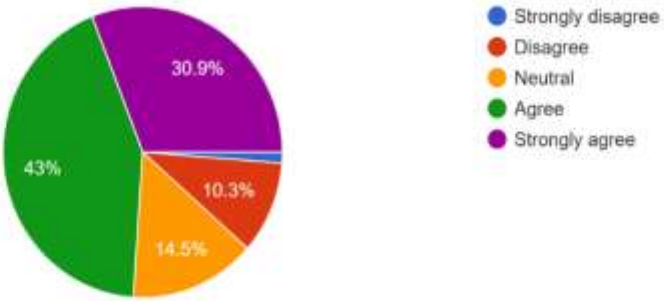
I believe that the location of my mobile device is monitored at some point in time.

165 responses



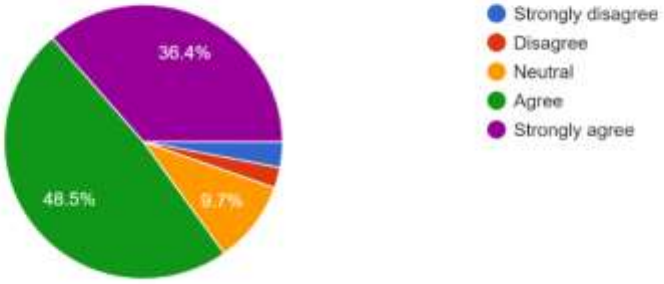
I'm concerned that my school/institute/university is collecting a lot of information about me

165 responses



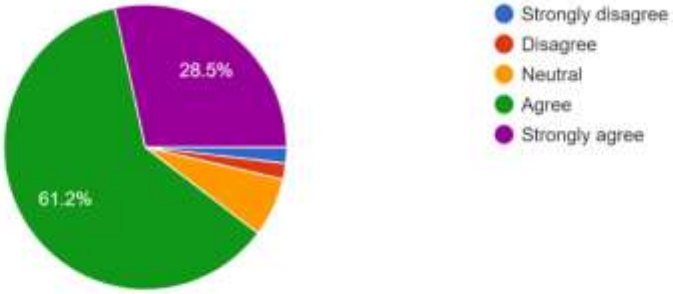
I am concerned that some stores may be monitoring my activities via my mobile device

165 responses



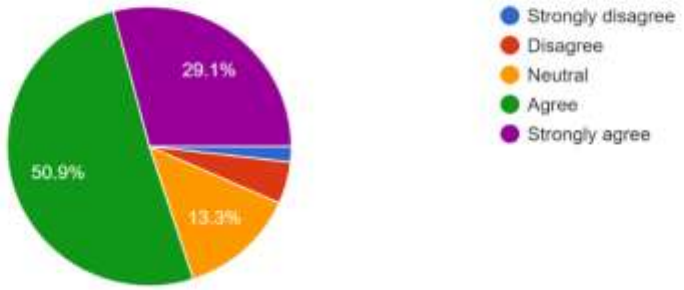
I feel that as a result of using entertainment, hospital and commercial devices, others may know more about me than I am comfortable with

165 responses



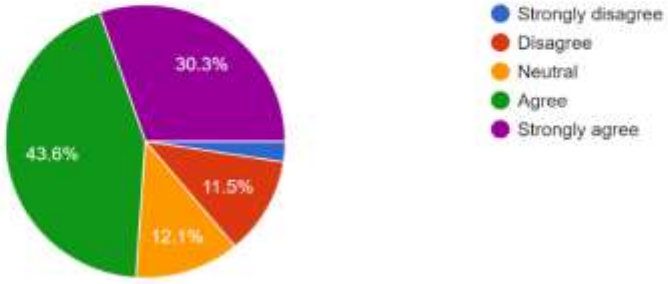
I believe that, as a result of the use of vehicular consumption devices, vehicle range and tire pressure, the information about me that I consider... more accessible to others more than I would like

165 responses



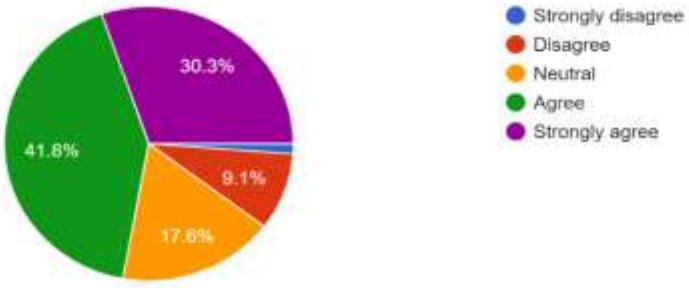
I am concerned that my school/institute/university may use my personal information for other purposes without notifying me or obtaining my authorization

165 responses



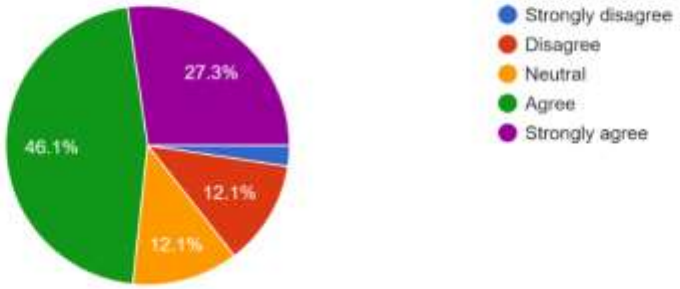
When I provide personal information to use IoT environments, I am concerned that devices such as my chair or drinking fountain may use my information for other purposes

165 responses



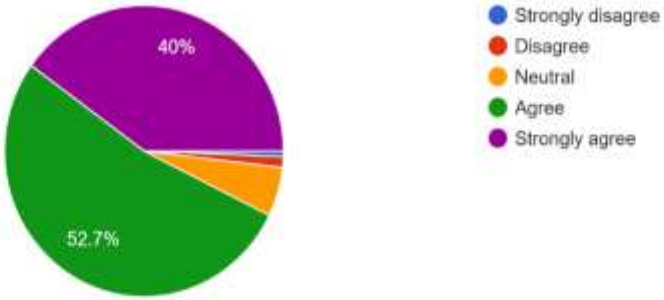
I am concerned that my department or the university restaurant, may share my personal information with other entities without obtaining my permission

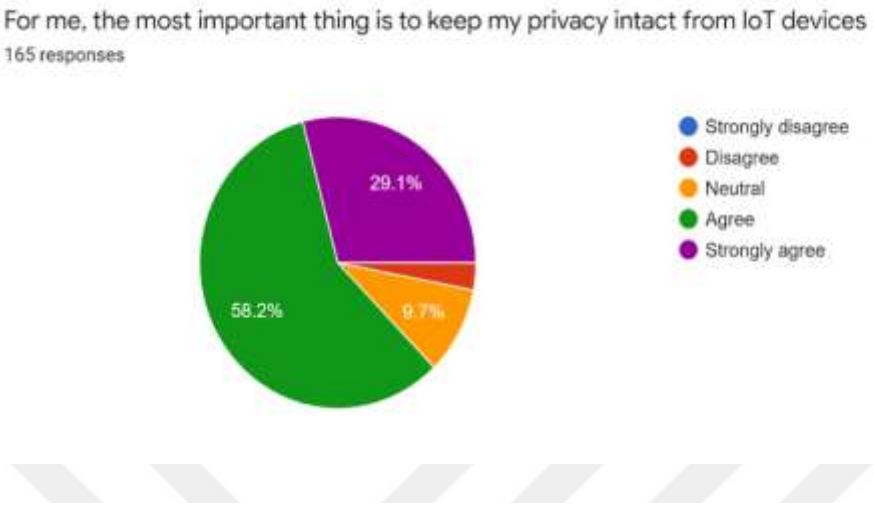
165 responses

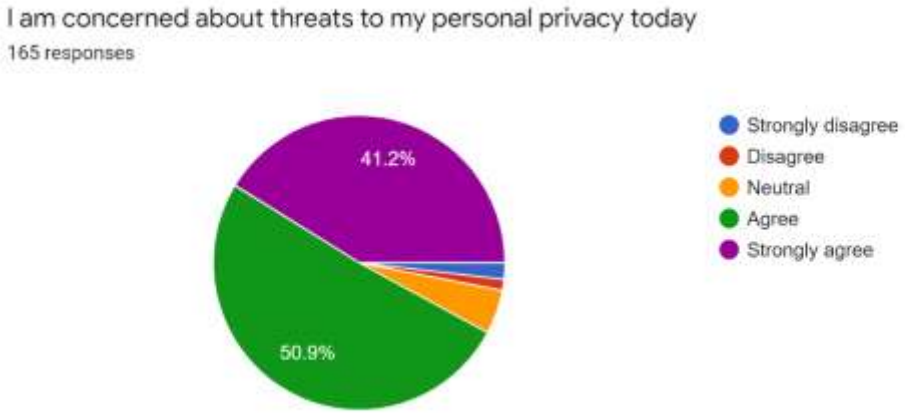


Compared to others, I am more sensitive about how IoT devices handle my personal information

165 responses







4.2.1.3 Instrument Evaluation

A) Reliability Analysis: When testing the reliability of the IoTPC, the Cronbach's Alpha measurement returned to a value of (α = 0.911), guaranteeing the reliability of the instrument according to the metrics [46].

The analysis of Cronbach's alpha, which is so significant, was evaluated in terms of how much each item of the IoTPC contributes to the result, presenting its mean and its standard deviation. The statistical details of each item contributing to Cronbach's Alpha can be seen in Table 4.1, such as mean, standard deviation, corrected total item correlation and the change in Cronbach's Alpha if an item from the instrument was excluded.

Table 4.1: Cronbach's Alpha Analysis

Item	Average	Standard deviation	Corrected Total Item Correlation	Cronbach's Alpha if Item is Deleted
IUICo1	4.44	0.958	0.498	0.909
IUIAw2	4.41	1.131	0.330	0.913
IUIAw3	4.74	0.705	0.502	0.909
IUIColl4	3.82	1.232	0.620	0.905
IUIColl5	4.11	1.212	0.562	0.907
IUIColl6	3.87	1.245	0.634	0.905
MUIPs7	4.28	1.280	0.332	0.914
MUIPs8	3.85	1.352	0.790	0.900
MUIPs9	3.87	1.372	0.644	0.905
MUIPi10	3.80	1.364	0.624	0.905
MUPi11	3.97	1.211	0.536	0.908
MUISui12	4.30	1.022	0.702	0.904
MUISui13	4.21	1.112	0.697	0.903
MUISui14	4.34	1.031	0.559	0.907
AIPGen15	3.34	1.569	0.757	0.901
AIPGen16	3.59	1.383	0.444	0.911
AIPGen17	3.89	1.380	0.773	0.900

It can be seen that the average scores of individual items ranged from 3.34 for the AIPGen15 item –" Compared to other people, I am more sensitive about how IoT devices handle my personal information" – to 4 .74 for IUIAw3 – "It is very important to me that I am aware and knowledgeable about how my personal information will be used".

In the Corrected Total Item Correlation, responsible for exemplifying how much each item influences the total value of Cronbach's Alpha, the results respected the minimum value of 0.300, varying their scores from 0.330 for the item IUIAw2 - "IoT devices that seek information should disclose how data is collected, processed and used" – at 0.790 for item MUIPs8 – "I'm worried that IoT devices are collecting too much information about me".

B) Sample Adequacy: the KMO adequacy measure presented a result of (KMO = 0.774), considered a valid value for carrying out the Exploratory Factor Analysis (EFA). Bartlett's spherical test showed a result of (T<0.001), concluding that the variables are significantly correlated for the realization of the EFA.

Another important point to note during the EFA is the commonalities. During the application of EFA, variables can be defined as a linear combination of common factors that will explain a portion of the variance of each variable, but a small portion that summarizes this total variance cannot be explained [47]. According to Hair [48], the portion explained by common factors is called commonalities. The commonalities are responsible for showing how much each item inherits from the total variance, that is, how much of the total variance is being loaded into each item. Table 4.2 presents the details of the loading of each item in relation to the total variance.

Table 4.2: Communalities

Item	Common Variance	Total Variance Inherited by Each Item
IUICo1	0.422	0.418
IUIAw2	0.403	0.371
IUIAw3	0.558	0.355
IUIColl4	0.681	0.703
IUIColl5	0.752	0.424
IUIColl6	0.687	0.619
MUIPs7	0.457	0.183
MUIPs8	0.742	0.716
MUIPs9	0.675	0.514
MUIPi10	0.548	0.422
MUPi11	0.632	0.645
MUISui12	0.848	0.612
MUISui13	0.756	0.611
MUISui14	0.632	0.477
AIPGen15	0.803	0.780

AIPGen16	0.553	0.275
AIPGen17	0.732	0.709

C) Initial Factor Extraction: when performing the initial factor extraction, the first technique used was the accumulated variance, which can be seen in Table 4.3. It is noted that, with the extraction of 3 factors for the 17 items of the instrument, the accumulated variance reaches 60.138 %, which is already considered sufficient. Kaiser's rule resulted in the extraction of five factors among the 17 items that obtained an eigenvalue above 1. The graphic analysis of the Scree Test was also performed indicating that, from the extraction of three factors for the 17 items of the instrument, the accumulated variance was not as significant. Figure 4.3 shows this abrupt drop in eigenvalues.

Table 4.3: Total Variance Explained

Number of Factors	Total	% Variance	% Cumulative
1	7.334	43.140	43.140
2	1.532	9.014	52.154
3	1.357	7.984	60.138
4	1.234	7.261	67.398
5	1.014	5.967	73.366
6	0.759	4.466	77.832
7	0.633	3.725	81.557
8	0.603	3.549	85.106
9	0.561	3.298	88.404
10	0.439	2.584	90.987
11	0.365	2.150	93.137
12	0.299	1.762	94.899
13	0.275	1.619	96.518
14	0.221	1.301	97.819
15	0.193	1.134	98.953

16	0.101	0.593	99.545
17	0.077	0.455	100.000

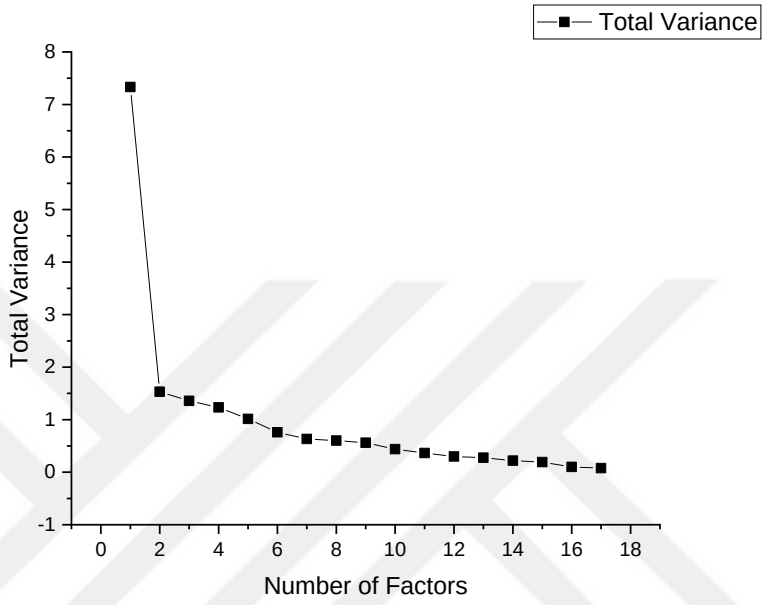


Figure 4.3: Scree Test

According to these analyses, the results obtained by the accumulated variance and Scree Test techniques indicated that the ideal number of factors to be extracted from the IoTPC was three.

With this information, the first factor matrix was generated by extracting three factors. It can be noted, from Table 4.4, that in this first matrix generated most of the items of the IoTPC were loaded in Factor 1, obtaining a low load in the other factors.

Table 4.4: Factor Matrix

Item	Factor 1	Factor 2	Factor 3
MUIPs8	0.830	-0.093	-0.135
AIPGen15	0.804	-0.323	0.169
AIPGen17	0.796	-0.202	0.185
MUISui12	0.743	0.245	0.028
MUISui13	0.739	0.229	-0.111

IUIColl6	0.693	-0.063	-0.367
MUIPs9	0.692	-0.182	-0.044
IUIColl4	0.678	-0.314	-0.381
MUIPi10	0.647	-0.007	0.060
IUIColl5	0.585	-0.071	-0.276
MUISui14	0.584	0.337	0.150
MUIPi11	0.568	-0.288	0.489
IUICo1	0.519	0.355	0.149
IUIAw3	0.518	0.268	-0.125
AIPGen16	0.460	0.053	0.245
MUIPs7	0.339	0.052	0.255
IUIAw2	0.367	0.482	-0.058

D) Rotation Factor: With the application of the Varimax rotation factor to this same matrix, it can be observed, in Table 4.5, that there is a more uniform distribution of the IoTPC items within the obtained factors. It is observed that factor 1 is related to the dimension of IoT Requests, since the items loaded in this factor belong to this dimension. Factor 2 obtained loaded variables originating from the dimensions of IoT Requests, Decision Power and General Privacy Concerns. Finally, factor 3 got loaded items belonging to the Previous Knowledge, Secondary Information Use and Decision Power dimensions.

Table 4.5: Rotating Factor Matrix

Item	Factor 1	Factor 2	Factor 3
MUIPs8	0.677	0.379	0.336
AIPGen15	0.556	0.674	0.127
AIPGen17	0.493	0.643	0.227
MUISui12	0.386	0.470	0.585
MUISui13	0.478	0.244	0.568
IUIColl6	0.722	0.116	0.291
MUIPs9	0.562	0.403	0.190

IUIColl4	0.817	0.177	0.066
MUIPi10	0.398	0.402	0.320
IUIColl5	0.595	0.130	0.230
MUISui14	0.167	0.325	0.586
MUIPi11	0.181	0.781	0.041
IUICo1	0.117	0.283	0.569
IUIAw3	0.324	0.099	0.491
AIPGen16	0.133	0.423	0.280
MUIPs7	0.045	0.366	0.218
IUIAw2	0.098	-0.001	0.601

E) Researcher's Interpretation: Factor 1 consists of the items IUIColl4, IUIColl6, MUIps8, IUIColl5 and MUIps9. This new dimension of the IoTPC clearly reflects the concern of IoT users regarding the data collected by IoT devices and the way in which these devices watch over them. Due to the fact that all items loaded in this factor belong to the IoT Requests dimension, factor 1 was also called IoT Requests.

Factor 2 appears in the items MUIPi11, AIPGen15, AIPGen17, AIPGen16, MUIPi10 and MUIPs7. This new dimension is strongly related to two main aspects, the first is the fact that IoT users are concerned with controlling their personal information in relation to IoT devices; and the second fact is related to general privacy issues, as all items that addressed this dimension were loaded in this factor. Given this, the second factor extracted from the IoTPC was also called Decision Power.

Factor 3 appears in items IUIAw2, MUISui14, MUISui12, IUICo1, MUISui13 and IUIAw3. This new dimension reflects the concern of IoT users with regard to IoT devices providing information collected by them to third parties, without the authorization of their users, in addition to being concerned with obtaining prior knowledge regarding the good privacy practices. Thus, the third factor extracted from the IoTPC was called Caution. Table 4.6 illustrates the organization of all IoTPC items in the extracted factors.

Table 4.6: Organization of IoTPC Items by Extracted Factors

Factors	Item
IoT requests	IUIColl4, IUIColl6, MUIPs8, UIUColl5 and MUIPs9
Decision power	MUIPi11, APIGen15, AIPGen16, AIPGen17, MUIPi10 and MUIPs7
Caution	IUIAw2, MUISui14, MUISui12, IUICo1, MUISui13 and IUIAw3

4.3. Results Analysis IoTPC Learning Assessment Results

In this chapter, the construction of the inference module IoTPC Learning and the results obtained with its evaluation are presented. As described in section 3.6.2, the inference module IoTPC Learning, its conversion rules and applied techniques were presented. Then in section 4.3, the results achieved from the learning model of IoTPC Learning are exemplified.

In order to use the IoTPC to assist in a privacy preservation mechanism in IoT environments, an inference module of IoT scenarios called IoTPC Learning was built. The construction process of this module is described below, reporting its development, as well as the techniques used.

The results [49] of the 25 decision trees generated to infer the IoTPC scenarios are presented in a grouped way below. As can be seen in Figure 4.8, of the 70 training data used for the learning model, approximately 81.65 % were correctly predicted, representing 57.4 responses. As for the wrong predictions, a percentage of about 18.35 % was generated, which is equivalent to 12.6 responses. In general terms, the average accuracy level of the 25 models generated was 81.65 %.

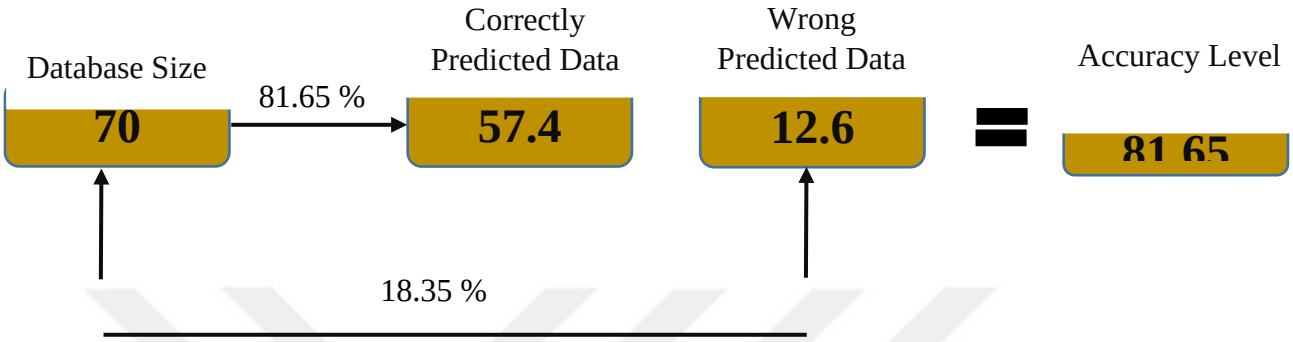


Figure 4.8: Learning process accuracy level

Some interesting points are identified through the analysis of each of these models. In Figure 4.9, it can be clearly observed that the level of correct predictions is higher than the level of incorrect predictions and that, in specific cases, some scenarios obtained the same value of correct and wrong inferences, thus as the same accuracy value, which can be exemplified by scenarios 14 and 15. This similarity of values was due to the fact that, among the 25 scenarios used, some deal with similar situations. In this example, scenario 14.

- "Your car wants access to its destination to calculate the most economical route."

And scenario 15

- " Your car wants information about its itinerary to determine if it has autonomy for the route."

These deal in general with fuel economy in vehicles; therefore, it was expected that the models for these two scenarios would be similar.

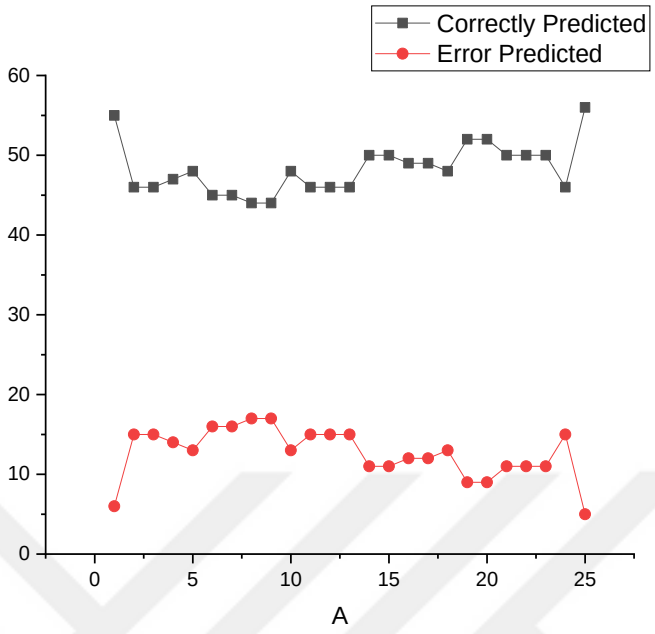


Figure 4.9: Individual relationship of the number of data correctly predicted in each scenario

Among the decision trees generated, in addition to the percentage of instances classified correctly, the accuracy of these classifications was observed. The detailed results of each generated model can be seen in Table 4.7.

The structure of IoTPC Learning is divided into three privacy factors extracted from the exploratory factor analysis: IoT requests, decision power and caution. Thus, the 25 decision trees developed were distributed among these three factors, making it possible to obtain a control over which dimension belongs to the requests made to the inference module.

The results obtained with IoTPC Learning may present a lower level of accuracy in relation to existing modules in some privacy negotiation mechanisms, such as Privacy Application [50]; however, some points should be highlighted.

Table 4.7: Inference details of each generated model

Decision tree	Correctly Predicted	Wrongly Predicted	Precision	Correctly Ranked Instances (%)
1	55	6	0.94	90.16
2	46	15	0.75	75.4
3	46	15	0.74	75.4
4	47	14	0.79	77.04
5	48	13	0.79	78.68
6	45	16	0.73	73.77
7	45	16	0.73	73.77
8	44	17	0.72	72.13
9	44	17	0.72	72.13
10	48	13	0.79	78.68
11	46	15	0.74	75.4
12	46	15	0.74	75.4
13	46	15	0.74	75.4
14	50	11	0.81	81.96
15	50	11	0.81	81.96
16	49	12	0.82	80.32
17	49	12	0.82	80.32
18	48	13	0.78	78.68
19	52	9	0.85	85.24
20	52	9	0.85	85.24
21	50	11	0.82	81.96
22	50	11	0.82	81.96
23	50	11	0.82	81.96
24	46	15	0.74	75.4
25	56	5	0.91	91.8

In Privacy Application and other privacy negotiation mechanisms, the learning process of their inference modules is given by the continued use of the mechanism, having the user answer a series of inference questions until the mechanism has the ability to start inferring correctly. With the integration of IoTPC Learning to these mechanisms, this initial calibration would not be necessary, since together – the developed instrument and the learning model – they are already able to indicate which information is considered private or not in that environment. Another point to note is that the validations used in the construction of the IoTPC Learning, such as the pruning technique and class balancing, ensure a greater degree of reliability for the module. Once IoTPC Learning is acting on a privacy trading mechanism, its learning capability becomes continuous. As IoTPC Learning is used, it adds the ability to learn from new users, thus gradually improving your predictions.

4.4 Final considerations

In this chapter, the construction process of the inference module IoTPC Learning was addressed. Some important points are worth highlighting about this process, such as the fact that the conversion rules to define whether or not the user would make information available in a given scenario were created by an expert in privacy in environments IoT, in this case, the researcher responsible for the present work. An evident limitation occurred in relation to the learning model, despite having obtained satisfactory results of accuracy, it would be interesting to have a larger database to carry out the training of the model, thus generating more accurate results in the same way, more personal attributes about the users, such as, for example, political position, financial condition, region where they live, among others, which would help to draw a better profile of these users, in addition to contributing to a greater precision of the learning model.

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

This work presented the process of building an instrument capable of measuring the concern with the privacy of users within IoT scenarios; the development of an IoT scenario inference module and its results. Thus, it offers, as a contribution, the construction of a new instrument, duly validated, capable of measuring such concerns, in addition to an option for an IoT scenario inference module to assist mechanisms for privacy negotiation.

The analysis of the instrument evidenced the classification of the items evaluated into three factors, in addition to verifying the degree of reliability of the instrument. The three factors allowed us to stratify privacy concerns into three large groups, the first group is represented by users who are concerned with how IoT devices collect their personal information and how they can use this collection to monitor them. The second group is represented by users who are concerned with controlling their personal information and who, because of this concern, consider themselves more sensitive with regard to their personal privacy. Finally, the third group is represented by users who are concerned about the fact that IoT devices are sharing their personal information with third parties without their authorization. This undue action ends up leading users to learn good practices on how to maintain greater privacy for their personal data.

The IoTPC Learning analyzes presented the construction of 25 learning models, with each model representing the inference result corresponding to an IoT scenario. The module also indicates an initial inference solution for privacy trading mechanisms, which can generate feedback on the privacy dimensions most requested by IoT devices. This use case of using the instrument showed that the module can be useful in other privacy trading mechanisms, providing integration options of IoTPC Learning with these mechanisms, thus resulting in significant improvements and even new functionalities.

Despite the efforts employed, this work faced some limitations. The fact that the context of privacy studied in this work is that of IoT ended up taking the development of the instrument to a specific scenario, therefore, not contemplating all the possibilities of violation of privacy possible in an IoT environment. The small size of the database and the relatively low number of features,

counting only on age, gender and education, also influenced the construction and validation of the inference module.

5.2 Future works

Future work could aim at the application of Confirmatory Factor Analysis (CFA) in order to confirm the data obtained through EFA; in addition to verifying whether the developed inference module obtains a higher level of accuracy when compared to modules from other privacy trading mechanisms over the course of its use.

Another possible approach would be to use machine learning to verify if the classification algorithms group IoTPC user profiles according to the factors extracted by the EFA

REFERENCES

[1] L. A. Omonkhoa and Dr. A. A. Ibrahim, "A Review Of The Issues And Challenges In IoT Security Using Machine Learning Techniques," April 2021. [Online]. Available: http://www.warse.org/IJETER/static/pdf/file/ijeter28942021.pdf.

[2] Ziegeldorf, J.H., O.G. Morchon, and K. Wehrle, Privacy in the Internet of Things: threats and challenges. Security and Communication Networks, 2014. **7**(12): p. 2728-2742.

[3] Oriwoh, E. and M. Conrad, 'Things' in the Internet of Things: towards a definition. International Journal of Internet of Things, 2015. **4**(1): p. 1-5.

[4] Perera, C., et al., Context aware computing for the internet of things: A survey. IEEE communications surveys & tutorials, 2013. **16**(1): p. 414-454.

[5] Miorandi, D., et al., Internet of things: Vision, applications and research challenges. Ad hoc networks, 2012. **10**(7): p. 1497-1516.

[6] Intelligence, M. SMART HOMES MARKET - GROWTH, TRENDS, COVID-19 IMPACT, AND FORECASTS (2021 - 2026). 2021 [cited 2021 March 8th]; Available from: https://www.mordorintelligence.com/industry-reports/global-smart-homes-market-industry.

[7] Wheatley, M. Intel Survey: Smart homes to be a Commonplace by 2025? . 2015 [cited 2021 March 8th]; Available from: http://realtybiznews.com/intel-survey-Smart-homes-to-be-commonplace-by-2025/98730580/.

[8] Briefs, R. What Are Smart Cities? 2020 [cited 2021 March 7th]; Available from: https://www.cbinsights.com/research/what-are-smart-cities/#what.

[9] Philippines, I. Smart Cities. 2015 [cited 2021 March 7th]; Available from: http://www.IoTphils.com/solutions/Smart-cities/.

[10] Team, G. 6 Forms of Wearable Technology You Must Know Right Now. 2020 [cited 2021 March 7th]; Available from: https://www.42gears.com/blog/6-wearable-technologies-you-must-know-right-now/.

[11] America, C.F.o. National Telecommunications and Information Administration. 2016 [cited 2021 March 6th]; Available from: https://www.ntia.doc.gov/files/ntia/publications/cfa_doc_iot_comments.pdf.

[12] Awasthi, G. Internet of Things - An Architectural Perspective. . 2015 [cited 2021 March 7th]; Available from: http://www.slideshare.net/gawasthi22/internet-of-things-arch-perspective.

[13] Davenport, T. The Analytics of Things. . 2015 [cited 2021 March 7th]; Available from: https://www.linkedin.com/pulse/analytics-things-tom-davenport.

[14] Huang, Y. and G. Li. A semantic analysis for internet of things. in 2010 International Conference on Intelligent Computation Technology and Automation. 2010. IEEE.

[15] Moeed, S.A. and A.A. Kumar, Internet of Things (IoT) - Internet Evolution. International Journal of Engineering Trends and Technology (IJETT), 2017(Special): p. 50-55.

[16] Lemos, A., A comunicação das coisas: teoria ator-rede e cibercultura. São Paulo: Annablume, 2013. **310**.

[17] Wu, M., et al. Research on the architecture of Internet of Things. in 2010 3rd international conference on advanced computer theory and engineering (ICACTE). 2010. IEEE.

[18] Aazam, M. and E.-N. Huh. Fog computing and smart gateway based communication for cloud of things. in 2014 International Conference on Future Internet of Things and Cloud. 2014. IEEE.

[19] Khan, R., et al. Future internet: the internet of things architecture, possible applications and key challenges. in 2012 10th international conference on frontiers of information technology. 2012. IEEE.

[20]. Zhang, M., F. Sun, and X. Cheng. Architecture of internet of things and its key technology integration based-on RFID. in 2012 Fifth International Symposium on Computational Intelligence and Design. 2012. IEEE.

[21] Ghani, N.A. and Z.M. Sidek, Controlling your personal information disclosure. Proceedings of ISP, 2008. **8**.

[22] Brandeis, L. and S. Warren, The right to privacy. Harvard law review, 1890. **4**(5): p. 193-220.

[23] Westin, A.F., Privacy and freedom. Washington and Lee Law Review, 1968. **25**(1): p. 166.

[24] Hong, J.I. and J.A. Landay. An architecture for privacy-sensitive ubiquitous computing. in Proceedings of the 2nd international conference on Mobile systems, applications, and services. 2004.

[25] RODOTÀ, S., A vida na sociedade da vigilância (coord. Maria Celina Bodin de Moraes). Rio de Janeiro: Renovar, 2008: p. 23-232.

[26] Clarke, R., Internet privacy concerns confirm the case for intervention. Communications of the ACM, 1999. **42**(2): p. 60-67.

[27] Langheinrich, M. Privacy by design—principles of privacy-aware ubiquitous systems. in International conference on Ubiquitous Computing. 2001. Springer.

[28] Michael, J., Privacy and human rights: an international and comparative study, with special reference to developments in information technology. 1994: Aldershot,[England): Dartmouth.

[29] Cate, F.H., The failure of fair information practice principles. Consumer protection in the age of the information economy, 2006.

[30] Whaley, E. The "Internet of Things" - Is Legislation Coming? . 2015 [cited 2021 March 7th]; Available from: http://www.troutmansanders.com/the-internet-of-things--is-legislation-coming-02-03-2015/.

[31] McNeill, P., Research methods. 2006: Routledge.

[32] Bulmer, M., Sociological research methods. 1977: Transaction Publishers.

[33] Bartunek, J.M. and M.-G. Seo, Qualitative research can add new meanings to quantitative research. Journal of Organizational Behavior, 2002: p. 237-242.

[34] Denzin, N.K., Introduction: Entering the field of qualitative research. Handbook of qualitative research, 1994.

[35] Yin, R.K., Design and methods. Case study research, 2003. **3**(9.2).

[36] Lee, H. and A. Kobsa. Understanding user privacy in Internet of Things environments. in 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT). 2016. IEEE.

[37] Malhotra, N.K., S.S. Kim, and J. Agarwal, Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. Information systems research, 2004. **15**(4): p. 336-355.

[38] Čas, J., et al., Introduction: Surveillance, privacy and security, in Surveillance, Privacy and Security. 2017, Routledge. p. 1-12.

[39] Smith, H.J., S.J. Milberg, and S.J. Burke, Information privacy: Measuring individuals' concerns about organizational practices. MIS quarterly, 1996: p. 167-196.

[40] Buck, C. and S. Burster, App information privacy concerns. 2017.

[41] Friedl, M.A. and C.E. Brodley, Decision tree classification of land cover from remotely sensed data. Remote sensing of environment, 1997. **61**(3): p. 399-409.

[42] Nash, J. A Tutorial to Understand Decision Tree ID3 Learning Algorithm. 2017; Available from: https://nulpointerexception.com/2017/12/16/a-tutorial-to-understand-decision-tree-id3-learning-algorithm/.

[43] Frank, E., et al., Weka-a machine learning workbench for data mining, in Data mining and knowledge discovery handbook. 2009, Springer. p. 1269-1277.

[44] Aggarwal, C.C. and C.R.D. Clustering, Algorithms and Applications. 2014, CRC Press Taylor and Francis Group.

[45] Lopes, B. Internet-of-Things-Privacy-Concerns. 2019; Available from: https://github.com/brunolp15/Internet-of-Things-Privacy-Concerns/blob/master/IoTPC/Estudo-1.csv.

[46] Streiner, D.L., Being inconsistent about consistency: When coefficient alpha does and doesn't matter. Journal of personality assessment, 2003. **80**(3): p. 217-222.

[47] Rezende, M.L., L.P.d.S. Fernandes, and A.M. Rodrigues, Utilização da análise fatorial para determinar o potencial de crescimento econonômico em uma região do sudeste do Brasil. Economia e Desenvolvimento, 2007(19).

[48] Hair, J.F., et al., Multivariate data analysis . Uppersaddle River. 2006, NJ: Pearson Prentice Hall.

[49] Lopes, B. Internet of Things Privacy Concerns. 2019; Available from: https://github.com/brunolp15/Internet-of-Things-Privacy-Concerns. [50] Pereira Couto, F.R. and S. Zorzo, Privacy negotiation mechanism in Internet of Things environments. 2018.