PRIVACY PRESERVATION MODEL FOR DATA EXPOSURE IN ANDROID SMARTPHONE USAGE

ANIZAH BINTI ABU BAKAR

UNIVERSITI SAINS MALAYSIA

2021

PRIVACY PRESERVATION MODEL FOR DATA EXPOSURE IN ANDROID SMARTPHONE USAGE

by

ANIZAH BINTI ABU BAKAR

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

November 2021

ACKNOWLEDGEMENT

First and foremost, I would like to express my gratitude to my supervisor, Dr. Manmeet Mahinderjit Singh and my co-supervisor, Dr. Azizul Rahman Mohd Shariff for giving me continuous guidance and assistance throughout the research. Their advice has made me able to achieve the research objectives and motivated me to finish this research. Secondly, I would like to thank my parents and friends for giving me continuous support to complete this research successfully. They helped me in many aspects during my studies at the university. My special appreciation to my friend, Othman Jalil, who helped me by recruiting the participants to get involved in the experiment and usability testing, contributed to this research's results and findings. Next, I would like to express my appreciation to Research University Grant (RUI) for providing financial support to complete this research. Last but not least, the next appreciation goes to the participants who volunteered to participate in the experiment and usability testing that helped by sharing their data for this research purpose. I am very thankful to all the people who were involved directly and indirectly in completing this study.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

R_{APP}	Risk of an application
Σ_{PL}	Summation of permission levels in an application
N_{P}	Number of permissions requested by an application
ΣW_{UD}	Summation of the weight of user data accessed or collected by an application
T_{SIZE}	Total size allocated by an application in the smartphone storage
R_{APPCAT}	Risk of an application category
$\Sigma_{PLa,b,\ldots,n}$	Summation of permission levels in an application category
$N_{Pa,b,\ldots,n}$	Number of permissions requested by an application category
$\Sigma W_{UDa,b,\ldots,n}$	Summation of the weight of user data accessed or collected by an application category
$T_{SIZEa,b,\ldots n}$	Total size allocated by an application category in smartphone storage
P_{u}	Privacy exposure level of a user
$\bar{X}_{P}$	Mean of permissions of applications in a smartphone
ΣW_{UDi}	Summation of the nth weight of user data accessed or collected by applications
ΣT_{SIZEi}	nth total size allocated by applications in the smartphone storage

LIST OF ABBREVIATIONS

IoT	Internet of Things
CRMA	Continuous Risk Monitoring and Assessment
CARTA	Continuous Adaptive Risk and Trust Assessment
PRMM	Privacy Risk Mitigation
PRI	Privacy Risk Indicator
SMS	Short Message Service
GPS	Global Positioning System
LTE	Long-Term Evolution
NFC	Near Field Communication
RFID	Radio-Frequency Identification
ABE	Attribute-based Encryption
PRiMo	Privacy Risk Model
AMoDaC	App-sensor Mobile Data Collector
IDE	Integrated Development Environment
APK	Android Package Kit
SDK	Software Development Kit
VBA	Visual Basic for Application
MB	MegaBytes

LIST OF APPENDICES

MODEL PEMELIHARAAN PRIVASI BAGI PENDEDAHAN DATA DALAM PENGGUNAAN TELEFON PINTAR ANDROID

ABSTRAK

Statistik menunjukkan terdapat 6378 juta pengguna telefon pintar. Penggunaan aplikasi dalam telefon pintar mendedahkan pengguna kepada risiko privasi. Penyelidikan sedia ada mempunyai kekurangan dalam memformalkan model matematik yang berkeupayaan untuk mengira risiko kedua-dua aplikasi sistem dan aplikasi pengguna. Alat pengumpul data pelbagai aspek juga tiada untuk memantau pengumpulan data pengguna and risiko yang dikemukakan oleh setiap aplikasi. Selain itu, tiada penanda aras tahap risiko yang dapat memaklumkan pengguna tentang perbezaan tahap risiko yang boleh diterima dan tidak boleh diterima dalam penggunaan telefon pintar. Bagi menangani isu risiko privasi, satu model privasi formal, dikenali sebagai PRiMo menggunakan struktur pokok dan pengetahuan kalkulus untuk mengira risiko setiap aplikasi, risiko yang dikemukakan oleh setiap kategori aplikasi, dan risiko privasi secara keseluruhan yang dialami oleh pengguna telefon pintar telah dicadangkan. PRiMo telah ditanam dalam satu pengumpul data pengesan aplikasi mudah alih dikenali sebagai AMoDaC untuk menganalisa data pengguna yang dicapai oleh aplikasi telefon pintar melalui kebenaran yang diberikan. Berdasarkan analisa AMoDaC, aplikasi alat & utiliti/produktiviti mengemukakan risiko paling tinggi berbanding kategori aplikasi yang lain. Tambahan pula, 29 pengguna menghadapi risiko rendah dan boleh diterima, manakala 2 pengguna menghadapi risiko sederhana. Kadar keberkesanan dan ketepatan sistem adalah 96.8%. Berdasarkan keputusan, satu penanda aras telah dikemukakan dengan membuat perbandingan antara hasil yang diperoleh dari PRiMo dengan sukatan ujian yang sedia ada. Berdasarkan penanda aras risiko yang dikemukakan, pengguna yang menghadapi risiko kurang dari 25% dianggap sebagai selamat, manakala pengguna yang menghadapi risiko sederhana, tinggi, dan sangat tinggi perlu mengambil langkah yang sepatutnya. AMoDaC membantu dalam menyedarkan pengguna tentang risiko yang mereka alami ketika mmenggunakan telefon pintar mereka dan mendedahkan jumlah data yang dikutip oleh aplikasi. Ia juga mengesyorkan langkah pengurangan di mana akan melarang permintaan izin yang berisiko. Secara umum, ia memupuk kesedaran penggunaan yang dapat melindungi pengguna dari digodam dan membantu masyarakat dalam mengurangkan jenayah siber disebabkan oleh pendedahan data peribadi.

PRIVACY PRESERVATION MODEL FOR DATA EXPOSURE IN ANDROID SMARTPHONE USAGE

ABSTRACT

Statistics show there are 6378 million of smartphone users. The usage of mobile applications in smartphones exposes users to privacy risks. This is due to existing works lacking a formalized mathematical model that can quantify both user and system applications risk. There is also no multifaceted data collector tool to monitor user data collection and risk posed by each application. Besides, there is no risk level benchmark that alerts users and distinguishes between acceptable and unacceptable risk levels in smartphone usage. In order to tackle the privacy risk issue, a formalized privacy model called PRiMo is proposed using tree structure and calculus knowledge to quantify the risk in each application, risk posed by each application category, and overall privacy risk faced by the smartphone user. The PRiMo is embedded into an App-sensor Mobile Data Collector (AMoDaC) tool to capture the user data accessed by mobile applications through the permissions granted. Based on the AMoDaC tool analysis, the tools & utility/productivity application posed the highest risk compared to other categories. Furthermore, 29 users faced low and acceptable risk, while two users faced medium risk. The effectiveness and accuracy of the system is 96.8%. Based on the results, a benchmark is proposed in line with the quantification of privacy risk by comparing the proposed PRiMo outcome with the existing available testing metrics. According to the benchmark proposed, users who face risk below 25% are considered safe, while users facing medium, high, and extremely high risk should take further actions accordingly. The AMoDaC helps in alerting the users about the risk they are facing

while using their smartphones and exposing the amount of data being collected by applications. It also recommends mitigation action which would prohibit any risky permission request. In general, it cultivates usage awareness that would protect the user from being hacked and helps the society in reducing the cybercrimes caused by private data leakage.

CHAPTER 1

INTRODUCTION

1.1 Background Study

Internet of Things (IoT) has created a great trend in the maturation of technologies. The IoT is embedded deeply in various domains such as Mobile Services, Smart Home, Enterprise Services, Smart Environments, Futuristic, Personal and Social application, Transportation and Logistic, Healthcare and Utilities (Al Nuaimi & Al Darmaki, 2017). The growth of IoT brings benefits to various aspects. For instance, the advancement in smartphone technologies leads to a higher number of users as the smartphone becomes part of human life on conducting daily basis. In recent years, almost all generations ranging from young to elderly are using smartphones. A statistic portrays that there will be 3.6 billion smartphone users in 2020, and it is expected to increase to 3.8 billion in the coming year, 2021 (S O'Dea, 2020). Furthermore, as of July 2020, the Android operating system holds 74.6% of the mobile operating system market share worldwide (StatCounter, 2020). When it comes to smartphone usage, it is not limited to only completing essential tasks but also used for other purposes such as entertainment, finance, navigation, communication, health and fitness, and more. A smartphone consists of both built-in and externally downloaded applications. As of June 2020, there are 2.96 million applications available in Google Play Store (Clement, 2020). From the statistics provided, it can be concluded that most people use Android smartphones in their daily lives to aid them in completing daily routines.

When owning a smartphone, it is necessary to have built-in and externally downloaded applications. The externally downloadable mobile applications have two main categories, which are free version and paid version. The sources to obtain these mobile applications vary based on the operating system used, such as Google Play Store, iOS App Store, and Huawei AppGallery.

The motives of installing mobile applications are to ease users in many aspects and entertain them by providing the services they need. Mobile application developers deliver these services. Most of the developers want to gain the benefits and profits from the services they provide. Consequently, they started to interfere in users’ information. This is where the users’ privacy is put at risk. The developers require sensitive or non-sensitive information from the users to gain profits by making deals and exposing them to irresponsible third parties and adversaries. Thus, how could developers possibly access and collect users’ data? The developers request the permissions of accessing information in smartphones before downloading the desired applications, which users need to agree and accept all the permissions in order to use the services. The most significant impact of accepting these permissions is exposing and disclosing the smartphone user's sensitive, private, and confidential information. In contrast, if the user disagrees with any of the permissions listed, that user should stop downloading the applications. By doing this, the user will be unable to use the services, but the privacy is preserved.

The privacy preservation model is need because privacy becomes an issue for smartphone users as the collected data might be susceptible and need to be handled with much care. There is a myriad of aspects that contribute to the information privacy issue. For instance, privacy attacks, carelessness and insufficient knowledge on users' privacy, the over-claimed permissions requested by developers, and the amount of user data accessed by applications with or without the consent of users.

All smartphones have built-in sensors that perform data capturing activities. The developers of mobile applications can easily access the information or metadata captured by these sensors to intrude on users' sensitive data. Furthermore, revealing users' personally identifiable information is a typical privacy issue that arises in smartphone usage. Untrusted third parties and developers themselves might access the information stored in the server without users' consent and knowledge. However, some mobile application developers and portable device manufacturers maintain their integrity by protecting users' data from being leaked and exposed.

In order to preserve the privacy of smartphone users in various domains, several solutions were proposed by previous researchers. For instance, blockchain-based system (Azaria *et al.,* 2016; Liang *et al.,* 2018; Liang *et al.,* 2017), attribute aggregation (Priesnitz Filho *et al.,* 2019), data anonymization (Abouelmehdi *et al.,* 2018; Nayahi & Kavitha, 2017), passive authentication (Naseer *et al.,* 2019), and privacy calculus (Enck *et al.,* 2019; Jozani *et al.,* 2020; Wang *et al.,* 2016). Among the solutions mentioned, privacy calculus seems to be a great solution in mitigating privacy issues. The privacy calculus quantifies the risks faced by smartphone users and portrays them as a quantitative result that allows ordinary users to understand clearly. Thus, the main focus of this study is quantifying application risks and privacy exposure levels using privacy calculus.

Although existing works have been proposed, privacy issues persist because existing solutions are ineffective. This shows that the current works overlooked some significant elements in their proposed works. Their models are single-faceted, non-extensive, and the models calculate risk for uninstalled applications instead of pre-installed and downloaded applications that are running at the background with and without the consent of users (Alshehri *et al.*, 2019; Lo *et al.*, 2016; Liu & Terzi, 2010; Khatoon & Corcoran, 2017). Apart from that, privacy attacks are also one of the reasons that lead to privacy breaches. Some current works focus on how to mitigate privacy attacks in smart environments and IoT (Al-Turjman, 2019; Lin & Bergmann, 2016). Still, they overlooked the first risk-leading element that became a loophole for the privacy breaches: user behaviour in smartphone usage. If preventive measures are applied in smartphone usage, especially during installation and use of mobile applications, then the chances of privacy attacks are low. Besides, the privacy of smartphone users must be preserved as a whole instead of focusing on particular elements, environments or attributes. Smartphones are capable of running a myriad of applications simultaneously. Due to this feature, the risks faced by users are continuous and should be monitored in real-time. Thus, a continuous risk assessment needs to protect and preserve the privacy of smartphone users fully.

The quantification of user privacy in the IoT environment is strongly related to two vital elements, which are Continuous Risk Monitoring and Assessment (CRMA) (Moon, 2016) and Continuous Adaptive Risk and Trust Assessment (CARTA) (MAR, 2018). According to (Moon, 2016), CRMA is a concept that “monitors and assesses an entity’s risk exposure levels and prioritizing audit and risk management procedures focusing on the entity’s high-risk areas in a more real-time manner”. CARTA is proposed by Gartner Inc. and is considered as an alternative concept to cybersecurity. According to (MAR, 2018), CARTA is an approach that urges for real-time risk assessment and making trust-based decisions. This research aims to quantify the continuous risk posed by applications and the privacy exposure level a user faces in smartphone usage. Thus, the CRMA concept is considered more suitable in quantifying continuous risk in smartphone usage compared to CARTA.

1.2 Research Motivations

Specific privacy models have been developed in mitigating privacy issues, although they have unforeseen privacy holes. The rapid maturation of technologies mainly causes these privacy holes. Although there are several types of research done on mitigating privacy issues, yet the problems persist (Azaria *et al.*, 2016; Dini *et al.*, 2018; Jozani *et al.*, 2020; Nayahi & Kavitha, 2017; Priesnitz Filho *et al.*, 2019; Wang *et al.*, 2016; Wottrich *et al.*, 2019). There is still a need to develop effective solutions that can preserve the privacy of smartphone users.

The act of protecting sensitive information is vital in preserving privacy. Smartphones that have built-in sensors are making way for intruders to breach privacy. Information or data collected from smartphones have a different level of sensitivity and should be protected to avoid any harm on the privacy rights (Lima *et al.*, 2018; Mayer *et al.*, 2016). Thus, a privacy-preserving system that is visible to the user to be aware of the level of risk they face in a smartphone environment is in need (Wang *et al.*, 2016).

The application of privacy calculus in smartphone usage is an excellent method and worth further researched and explored as it quantifies sensitive information to protect and preserve users' privacy (Wottrich *et al.*, 2019). The mathematical model consists of attributes based on the permission levels of mobile applications and the data types collected by the sensors. This encourages the study of privacy and risk quantification model in the usage of the smartphone. Thus, this research concentrates on the permission level types and user data size that contribute to quantifying risk in applications and privacy exposure level in the smartphone environment.

Previous researchers develop risk management models to mitigate risk. For instances, Continuous Risk Monitoring and Assessment (CRMA) (Moon, 2016), privacy calculus model (Wang *et al.,* 2016), privacy risk assessment for sensitive data (Senarath *et al.,* 2018), and Privacy Risk Mitigation (PRMM) and Privacy Risk Indicator (PRI) for open data (Ali-Eldin *et al.,* 2018). All the privacy risk assessment models evaluate the risks faced by the users in various environments. The usage and implementation of attributes in developing the models to quantify the risk are different and insufficient to protect the user's privacy in the full use of the smartphone.

1.3 Research Gaps

There are two elements involved in describing information which is content and metadata. Often, communication contents are given attention, whereas communication metadata are being ignored or overlooked in protecting users' privacy (Mayer *et al.,* 2016). Some of the metadata are sensitive and confidential that can harm users' privacy (Jin *et al.,* 2018). For instance, time of call or Short Message Service (SMS), call details, phone number of caller and recipient, duration of the call, length of SMS, age, gender, location, etc. (Mayer *et al.,* 2016).

Furthermore, two challenges can occur in preserving the privacy of smartphone users: privacy challenges and security challenges (Virat *et al.,* 2018). For instance, a tremendous amount of data is being generated by IoT devices. This will make the information vulnerable and create a way for the intruders to interfere by eavesdropping. Furthermore, the most common security challenges in the IoT environment are phishing attacks, malicious worm/virus attacks, and malicious scripts attacks (Virat *et al.,* 2018). Consequently, users lose confidence and trust in

IoT services. These issues and challenges need some attention to mitigate privacy concerns. The gaps that lead to these issues are found and described further below.

The first gap is no development of a privacy preservation model to preserve user privacy in the smartphone environment. A smartphone consists of two types of leading applications, which are system applications and user applications. Both applications need to be quantified to preserve privacy as system applications also pose risks to users. Smartphone users might have diverse categories of mobile applications downloaded and installed on their smartphones in terms of user applications. Each application poses a different level of risk, and each different category of the mobile application collects various types of content and metadata about users. The risk levels posed by different categories of applications determine a user's privacy exposure level in smartphone usage. However, previous researchers only focus on developing a privacy model to quantify the user's privacy in selective mobile applications and the category of applications. For instance, researchers (Min, 2016; Yin *et al.*, 2017) focus on privacy preservation in social networks, researchers (Chopdar *et al.*, 2018) discuss mobile shopping risk, researchers (Sampat & Prabhakar, 2017; Yasaka *et al.*, 2020) discuss the privacy preservation in health app category, researchers (Rastogi & Hendler, 2017) discuss on the messaging applications, and researchers (Russell *et al.*, 2018) discuss privacy in gaming applications. Thus, a privacy model that quantifies multiple applications risks needs to determine the overall privacy exposure level faced by the user in smartphone usage.

The second gap is the incompatibility of existing privacy models for tackling confidentiality. When there are new technologies, there will be new privacy issues and challenges. The new technologies could be new sensors, new applications, usage of new smart devices, etc. The current privacy models cannot mitigate rising privacy issues in line with the advancement of technologies as the features used in the models are non-extensive (Enck *et al.,* 2019; Lo *et al.,* 2016). Therefore, a novel privacy risk model with an extensive feature is needed to preserve privacy in emerging technologies.

The third gap is privacy is assumed to be a single facet. This is closely related to the determination of attributes in defining privacy breaches. Previous researchers focused on developing single faceted systems when conducting their research (Chopdar *et al.,* 2018; Min, 2016; Sampat & Prabhakar, 2017; Yasaka *et al.,* 2020; Yin *et al.,* 2017). However, the single faceted system is insufficient to mitigate privacy breaches and portray users' overall risk. Multi attributes should be investigated that may lead to privacy breaches of users and put them at risk. Thus, using a multifaceted system in defining one's privacy risk is encouraged as it gives a more accurate quantitative value of risk posed by applications and the privacy exposure level faced by each user.

The final gap is there is no benchmark of privacy risk for smartphone users. The benchmark helps users to locate themselves the level of risk they are facing. Some risks are acceptable, and users can live with them. Most previous researchers developed privacy-preserving models without proposing a benchmark (Alshehri *et al.,* 2019; Lo *et al.,* 2016; Sampat & Prabhakar, 2017; Wottrich *et al.,* 2019). Thus, a benchmark is in need for users to know the level of risk they are facing.

1.4 Research Problems

Smartphone sensors collect all the possible data related to user activities, including sensitive information (Köping *et al.,* 2018). When this information is disclosed to irresponsible parties, data leakage and privacy breaches occur. This data leakage happens due to the non-existence of device-user mapping and privacy that is not enforced properly.

Apart from that, current privacy models cannot preserve user privacy in smartphone usage due to the lack of a formalized mathematical model that can quantify risk posed by user applications and system applications. Besides, existing works are incompatible in tackling the confidential issues due to the non-extensive models (Enck *et al.,* 2019; Lo *et al.,* 2016).

Furthermore, there is no multifaceted privacy quantification and privacy-preserving system (Chopdar *et al.,* 2018; Min, 2016; Sampat & Prabhakar, 2017; Yasaka *et al.,* 2020; Yin *et al.,* 2017). This leads to the massive collection of sensitive and non-sensitive information by applications without the knowledge and consent of users. This system is very significant in monitoring the real-time user behaviour and application behaviour to determine the risk posed by applications and the privacy exposure level of users. The developed system exposes the permissions requested by applications and the size of user data collected by them.

Lastly, there is no benchmark of risk level in mobile application usage that users can refer to accept the risk and live with it (Alshehri *et al.,* 2019; Lo *et al.,* 2016; Sampat & Prabhakar, 2017; Wottrich *et al.,* 2019). The smartphone users are facing risks in smartphone usage. Therefore, the need to benchmark the risk level is important to understand their privacy exposure level.

1.5 Research Questions

In this research, there are three main research questions:

1. What is the solution to quantify and preserve the privacy of users in smartphone usage?

2. What is the system that can perform real-time monitoring of user behaviour and applications behaviour in smartphones?

3. What is the effectiveness of the proposed model compared to other privacy risk models in benchmarking the privacy exposure level of a user?

1.6 Research Objectives

Based on the problem stated, the aim of this research is to propose a mathematical model to quantify the privacy risk in Android smartphone usage. The following are the objectives of this research:

1. To formalize a mathematical model using tree structure and propose a mathematical model designed using privacy calculus solution that will preserve users' privacy in the smartphone environment.

2. To design a multifaceted system that can perform real-time monitoring and collecting information on user behaviour and applications behaviour in a smartphone environment.

3. To benchmark the proposed privacy risk model outcome with the existing available testing metrics.

1.7 Research Contributions

The contributions of this research are:

1. A novel mathematical model known as PRiMo to quantify privacy and risk of the user in smartphone environment is proposed.
    - The privacy calculus solution is used to develop the privacy risk model to quantify the risk posed by applications and the privacy exposure level faced by the smartphone user.
2. Enhances tree structure model that fits smartphone environment to tackle privacy risk problems.
    - The related attributes used to develop the model are permission level, sensor, and personal data that fit smartphone environment to tackle privacy issues.
3. A multifaceted system, AMoDaC is designed to monitor the collection of user data size by accessing the granted permissions and portraying the risks posed by smartphone applications.
    - AMoDaC system is embedded with PRiMo to monitor the collection of user data size and risk posed by the applications for each user that is real time. By having this system, users can evaluate the list of permissions requested, the amount of data accessed and collected by an application, and the risk posed by each application.

1.8 Research Summary

The research summary consisting of research gaps, problem statements, research questions, objectives, and research contributions is shown in Table 1.1.

Table 1.1 Research Summary

Research Gaps	Problem Statements	Research Questions	Research Objectives	Research Contributions
No development of privacy preservation model to preserve the privacy of users based on their usage behaviour in using different categories of mobile applications (Min, 2016; Yin *et al.*, 2017; Chopdar *et al.*, 2018; Sampat and Prabhakar, 2017; Yasaka *et al.*, 2020).	Lack of formalized mathematical model that can quantify the risk posed by both user applications and system applications (Enck *et al.*, 2019; Lo *et al.*, 2016).	What is the solution to quantify and preserve the privacy of users in smartphone usage?	To formalize mathematical equation using tree structure and propose a mathematical model designed using privacy calculus solution that will preserve the privacy of users in the smartphone environment.	A novel mathematical model called PRiMo to quantify privacy and risk of the user in smartphone environment is developed.
				The related attributes used to create the model are permission level, sensor, and personal data. This is shown using a tree structure.
Incompatibility of existing privacy models for tackling confidentiality (Enck *et al.*, 2019; Lo *et al.*, 2016).				
No multifaceted system to monitor smartphone usage (Min, 2016; Yin *et al.*, 2017; Chopdar *et al.*, 2018; Sampat and Prabhakar, 2017; Yasaka *et al.*, 2020).	No multifaceted system to monitor the collection of user data and risk posed by the applications for each user (Min, 2016; Yin *et al.*, 2017; Chopdar *et al.*, 2018; Sampat and Prabhakar, 2017; Yasaka *et al.*, 2020).	What is the system that can perform real-time monitoring of user behaviour and applications behaviour in the smartphone?	To design a multifaceted system that can perform real-time monitoring and collecting information on user behaviour and applications behaviour in a smartphone environment.	A system combining PRiMo and multifaceted features known as AMoDaC tool is designed.
No benchmark of privacy risk for smartphone users (Lo *et al.*, 2016; Sampat and Prabhakar, 2017; Alshehri *et al.*, 2019; Wottrich *et al.*, 2019)	No benchmark of risk level in mobile application usage that user can accept and live with it (Lo *et al.*, 2016; Sampat and Prabhakar, 2017; Alshehri *et al.*, 2019; Wottrich *et al.*, 2019).	What is the effectiveness of the proposed model compared to other privacy risk models in benchmarking the privacy exposure level of a user?	To benchmark the proposed privacy risk model outcome with the existing available testing metrics.	Proved that PRiMo and AMoDaC provide analysis of privacy risk.

1.9 Research Scope

The scope of the research is narrowed down by clustering the diverse categories of applications into 11 main categories. The study is done on these elements and attributes: mobile applications, permission levels and user data size. Application risk and privacy exposure level are the main theme in this research. Privacy calculus has been chosen as a method to quantify privacy exposure level, risk of an individual application, and risk of each category of application. The privacy score is calculated for an individual user. The operating system that will be focused on in this study is Android Operating System which is widely used globally. The risk is quantified by referring to the sensor data and personal data accessed by the existing sensors in smartphones. The quantification of risk is done on both user applications and system applications.

1.10 Research Methodology

Figure 1 illustrates the methodology for the proposed work. This methodology will be used throughout the research. The methodology starts with studying previous literature and ends with the conclusion. Figure 1 also highlights the objectives that are achieved throughout the research.

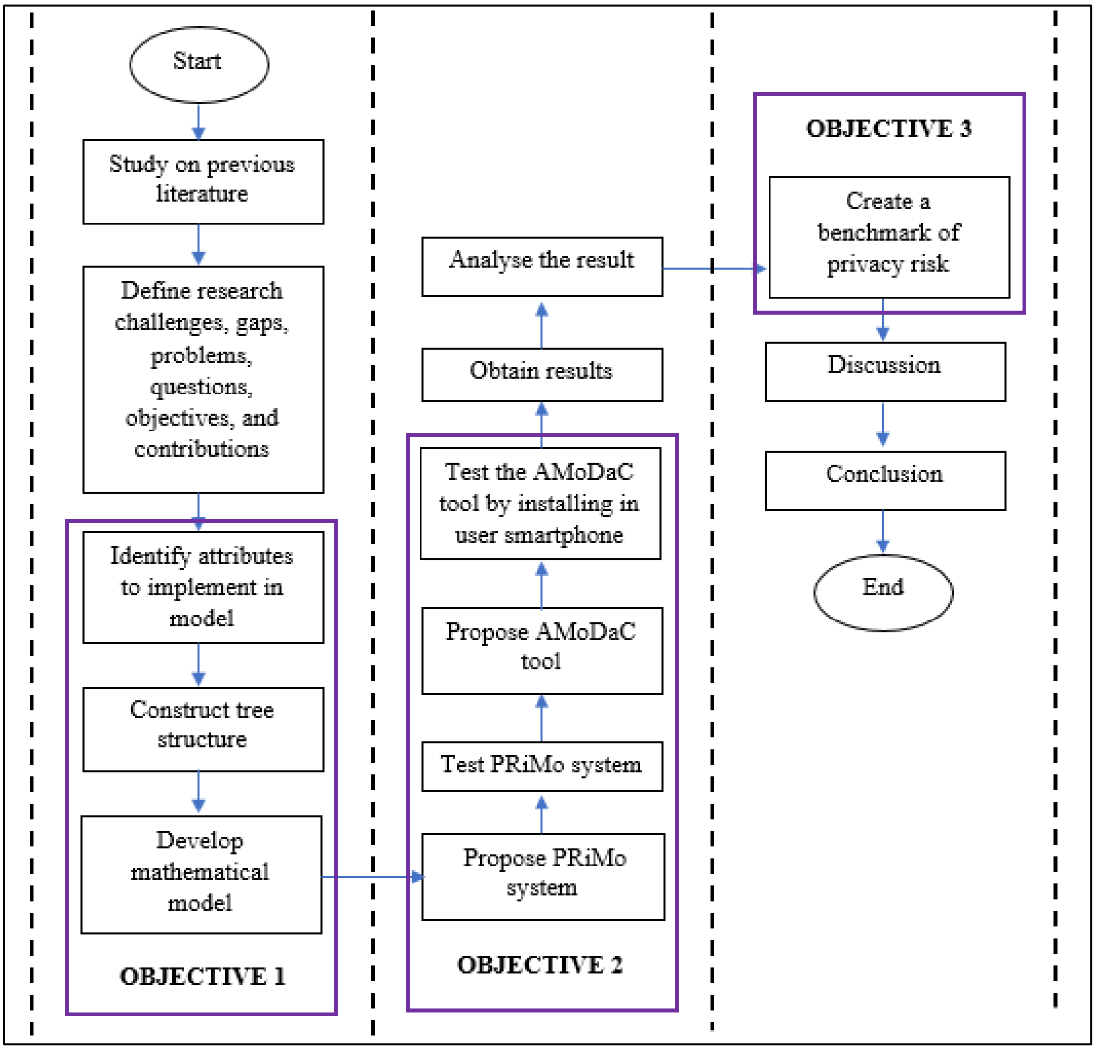
Start
Study on previous literature
Define research challenges, gaps, problems, questions, objectives, and contributions
Identify attributes to implement in model
Construct tree structure
Develop mathematical model
OBJECTIVE 1
Propose PRiMo system
Test PRiMo system
Propose AMoDaC tool
Test the AMoDaC tool by installing in user smartphone
OBJECTIVE 2
Obtain results
Analyse the result
OBJECTIVE 3
Create a benchmark of privacy risk
Discussion
Conclusion
End

Figure 1.1 Research Methodology

The research starts with studying previous literature to understand the challenges and issues that occur in privacy-preserving. Then, research is continued by defining the gaps, problems, questions, objectives, and contributions. Next, the elements and attributes to be implemented in the model are identified. Then, the tree structure is constructed to clarify the risk-leading elements. Finally, the proposed mathematical model is developed. By doing this, the first objective is achieved.

The next step is to propose a PRiMo system, test the efficiency of the PRiMo system. Then, AMoDaC is proposed as part of PRiMo system as a data collector tool and is tested by installing on users' smartphone. By completing this stage, the second objective is achieved. Later the results are obtained through actual experiments and further analysed.

Using the outcome of the PRiMo system, a benchmark is created by comparing it to the existing testing metrics. This completes the third objective. Finally, a discussion on the research is done, and the study ends with a conclusion.

1.11 Structure of Report

The report is divided into seven chapters. Below are the brief descriptions regarding the following four chapters in this report:

Chapter 2: This chapter provides a background and literature of previous works related to privacy risk. This chapter also discusses the risk-leading elements that lead to privacy breaches. The techniques and solutions used to quantify privacy are also explained in this chapter.

Chapter 3: This chapter discusses the proposed techniques and model for quantifying the privacy risk of smartphone users. This chapter also provides the research strategies and the environment used to implement the proposed methods.

Chapter 4: This chapter elaborates the technical details on how the proposed works are developed. The proposed model to quantify the privacy risk will be designed and evaluated.

Chapter 5: This chapter portrays the results and findings obtained from the research done. The discussion on research is also done in this chapter.

Chapter 6: This chapter concludes the research, discusses the limitation of this research and the future works that can be done to overcome the issues.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Chapter 2 mainly consists of literature review. In this chapter, the background information on information privacy in smartphone usage are discussed. The discussion starts with a background of study and ends with a summary on privacy risks in smartphone usage.

In this chapter, the research provides background information on IoT and smartphone. The discussion consists of a synthesis of IoT and smartphones, privacy challenges and threats that could occur in both IoT and smartphone environments, related work, factors that impacted the effort to preserve privacy, and the risk-leading elements in smartphone usage. The research also provides literature review related to the proposed work. The topic touched here is privacy calculus and risk-leading elements. This chapter ends with a summary of the privacy risk in smartphone usage. Figure 2.1 highlight the topic of discussion in this chapter. Next, the research proceeds with a discussion about the background study of this research.

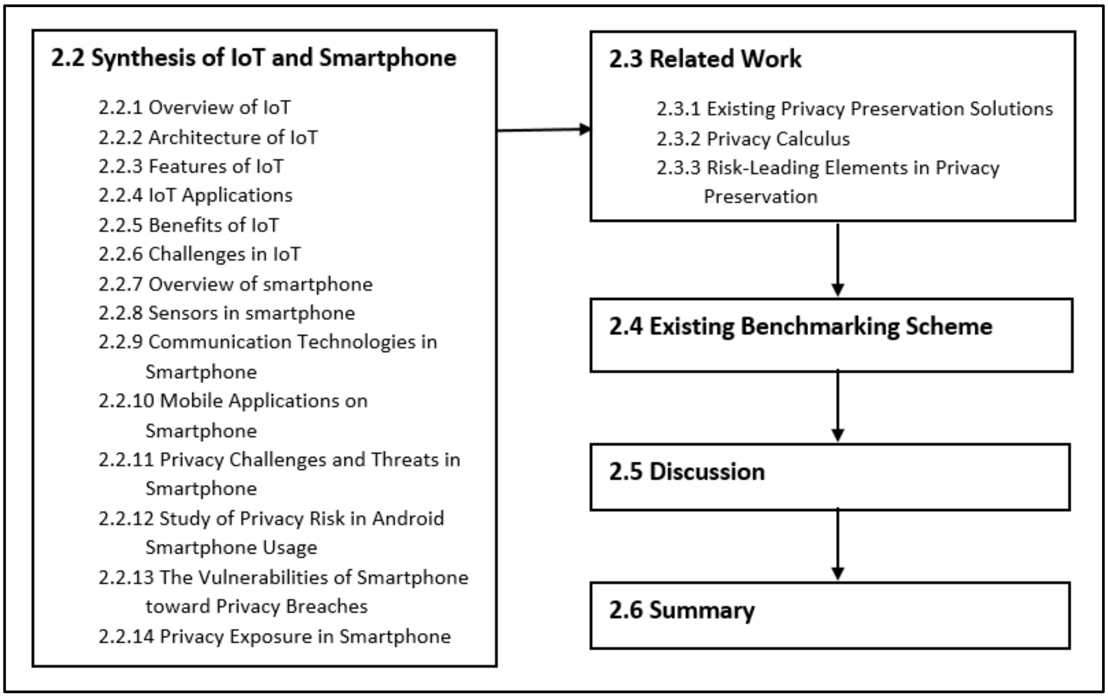
2.2 Synthesis of IoT and Smartphone
2.2.1 Overview of IoT
2.2.2 Architecture of IoT
2.2.3 Features of IoT
2.2.4 IoT Applications
2.2.5 Benefits of IoT
2.2.6 Challenges in IoT
2.2.7 Overview of smartphone
2.2.8 Sensors in smartphone
2.2.9 Communication Technologies in Smartphone
2.2.10 Mobile Applications on Smartphone
2.2.11 Privacy Challenges and Threats in Smartphone
2.2.12 Study of Privacy Risk in Android Smartphone Usage
2.2.13 The Vulnerabilities of Smartphone toward Privacy Breaches
2.2.14 Privacy Exposure in Smartphone
2.3 Related Work
2.3.1 Existing Privacy Preservation Solutions
2.3.2 Privacy Calculus
2.3.3 Risk-Leading Elements in Privacy Preservation
2.4 Existing Benchmarking Scheme
2.5 Discussion
2.6 Summary

Figure 2.1 Overview of Chapter 2

2.2 Synthesis of IoT and Smartphone

The Internet of Things (IoT) has matured tremendously because of its pervasive feature. In this section, the synthesis of IoT and smartphones, privacy challenges, and threats in smartphones are discussed further.

2.2.1 Overview of IoT

Internet of Things (IoT) emerges and dramatically changes the environment, human beings, and lifestyles. IoT has created a significant trend in the maturation of technologies. There are several definitions of IoT defined by previous researchers (Patel *et al.*, 2016; Alaba *et al.*, 2017; Lin *et al.*, 2017; Hassan, 2019). According to (Gregorio *et al.*, 2020), IoT is defined as “the total number of devices interconnected through the Internet, capable of collecting data to monitor and control everyday things, remotely, without the need for continuous interaction between things and people". The world is undergoing an evolution of networks where interconnected computers are now evolving into interconnected objects. This phenomenon is due to the ease of communication, ability to control and automate remotely, and cost-saving. Figure 2.2 shows the statistic of IoT connected devices worldwide in 2018, 2025, and 2030.

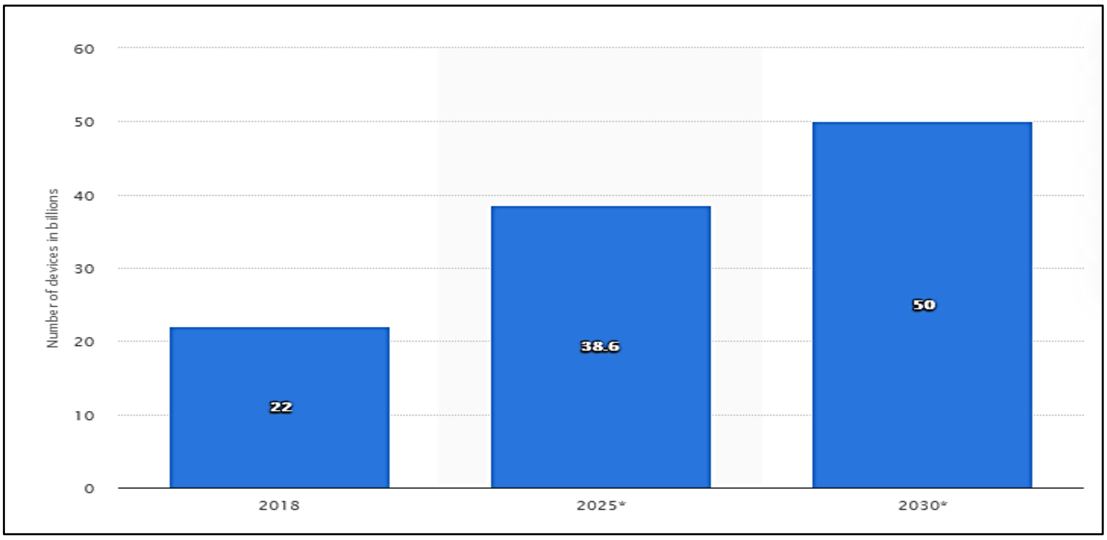

Figure 2.2 Number of IoT Connected Devices Worldwide (Statista, 2020)

Based on the statistic shown in Figure 2.2, 22 billion IoT connected devices worldwide were recorded in 2018. The number is expected to increase to 38.6 billion in 2025 and will achieve 50 billion in 2030. Thus, this shows that the adaptation of IoT is being practised and accepted by many people. The architecture of IoT is discussed further in the next section.

2.2.2 Architecture of IoT

The IoT architecture is divided into four layers: object sensing layer, data exchange layer, information integration layer, and application service layer (Shahid and Aneja, 2017). Figure 2.3 shows the four-layered architecture of IoT.

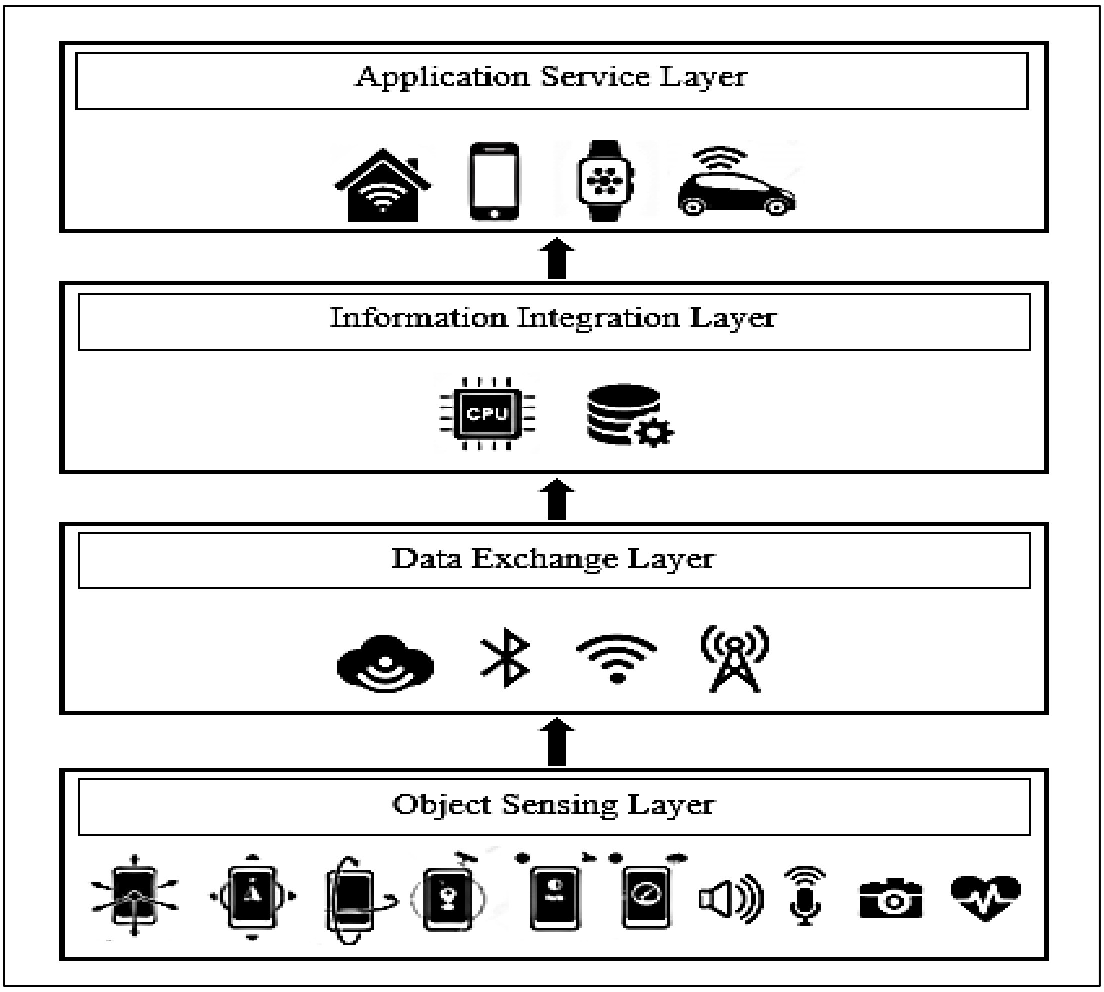

Figure 2.3 Architecture of IoT

The architecture starts from the bottom layer, the object sensing layer, to the top layer, the application service layer. First, the object sensing layer detects and gains the data from the physical objects. The obtained data are then transmitted to the data exchange layer that handles the communication of data. The information integration layer then processes the data by merging, recombining, and cleaning the undetermined data size obtained from the network. Besides, the conversion of anonymous information into usable knowledge occurs in this layer. Lastly, a diverse of content services are provided to the users through the application service layer.

In the real world, end-users only get involved at the application service layer to obtain finalised information and use services provided by the manufacturer or developer. For instance, in smartphone usage, the user utilises the services through the device. However, before the services are ready to reach the user, there are few processes that each layer will undergo. For example, a smartphone user needs the information of the location. The data are captured via the Global Positioning System (GPS) sensor available in a smartphone (object sensing layer). These undetermined data are then transmitted to the processing unit (information integration layer) to convert them into usable knowledge via Wi-Fi, Bluetooth, and any other available communication mediums (data exchange layer). After data are converted into understandable information, they are delivered to end-users through the smartphone (application service layer).

The architecture of IoT is explained in detail to provide a clearer view of IoT. The following section describes the features of IoT that contribute to the tremendous growth of IoT implementation.

2.2.3 Features of IoT

IoT grows tremendously due to its features. According to (Gregorio *et al.*, 2020), the significant features of IoT are vast and extensive scale, interconnectivity, dynamic changes, and heterogeneity.

The first feature of IoT is vast and extensive scale. The number of appliances or devices that transmit is gigantic compared to the number of appliances or devices currently connected to the Internet.

The second feature of IoT is interconnectivity. The worldwide information and communication infrastructure can be interconnected with almost anything, anytime, any service, any path, and anywhere.

The next feature is dynamic changes. For example, the state of the connected or disconnected devices and the location, speed, and number of devices might change dynamically.

The heterogeneity is one of the fundamental features of IoT. This phenomenon is closely related to IoT devices being heterogeneous as they can interact with service platforms and other devices through divergent networks.

The features of IoT are described briefly to portray the characteristics of IoT. The following section discusses the applications of IoT in different domains in real life.

2.2.4 IoT Applications

The emergence of IoT brings benefits to several application domains. For instance, the IoT is embedded deeply in various environments such as Mobile Services, Smart Home, Enterprise Services, Smart Environments, Futuristic, Personal and Social application, Transportation and Logistics, Healthcare and Utilities (Al Nuaimi & Al Darmaki, 2017). Besides, Smart City, Smart Living, and Smart Energy are also included in IoT applications (Gregorio *et al.*, 2020). From the mentioned domains, several IoT applications are identified for potential growth, such as smart health, smart transport, smart industry, smart city, smart home, smartphone, etc. The IoT applications in specific domains are discussed further in the next section.

2.2.4(a) Smart Environment

Smart environment responsible for providing pleasant, cosy, and effortless surroundings. For instance, a smart home offers a comfortable and enjoyable environment. This smart home is created by implementing sensors that aid in providing comfortable life through several scenarios: (1) room temperature is adjusted automatically according to the weather conditions; (2) lighting can be adjusted by sensing the brightness or darkness in a room; (3) energy could be saved by turning on or off appliances by sensing the presence and absence of human (Singh *et al.*, 2019). This could save power consumption costs and become an environmentally friendly product. Several examples of sensors that might be considered to be implemented in a smart home are motion sensor, touch sensor, emotion detection camera sensor, temperature sensor, humidity sensor, fall detection sensor, and RFID tags (Singh *et al.*, 2019; Wai Soon *et al.*, 2015).

2.2.4(b) Smart Health

There are several IoT applications in the healthcare domain. The most common technology in current real life is the usage of the smartwatch. In term of personal use, the smartwatch acts as a wearable device that aid in monitoring the health conditions of users including blood pressure, breathing activities, calories burned, and body temperature (Shahid & Aneja, 2017; Papa *et al.*, 2020). In terms of hospital use, wireless devices and cloud storage store health-related information and records of patients by automatically analysing their behaviour (Mshali *et al.*, 2018). By doing this, it can save a patient’s life in a state of emergency.

2.2.4(c) Smartphone

IoT and smartphones correlate in providing access to any information or records obtained in any smart things. Smartphone has become a vital device in conducting daily routines. It can communicate with any other devices and provide access to any content, anytime, anywhere to anyone. This is due to smartphone features that consist of multiple sensors such as accelerometer, gyroscope, GPS sensor, camera, microphone, etc. (Köping *et al.*, 2018). The smartphone consists of user applications and system applications that provide the preferred services to complete tasks. The smartphone can be ubiquitous as it can communicate through the technologies embedded in a smartphone, such as Wi-Fi and mobile networks. It is being used to conduct tasks in diverse fields such as finance, health and fitness, entertainment, augmented reality, productivity, education, etc. Due to its capabilities, the smartphone has become a powerful device despite its size.

As discussed previously, IoT connected devices and IoT implemented environments are undergoing tremendous growth because of their benefits. Therefore, the next section confers the benefits of IoT.

2.2.5 Benefits of IoT

The maturation of IoT brings benefits to many aspects. It provides convenience to individuals, organizations, important domains, and society on a daily basis. Several benefits of IoT are discussed in this section as follows.

One of the benefits of IoT is cost saving. Costs for power consumption can be reduced by implementing sensors. The sensors can help individuals, business sectors, and society by automating the turning on and off of any appliances such as air conditioners, televisions, lights, machines as needed (Gregorio *et al.*, 2020).

The IoT also provides personalized eHealth and mHealth services. In the healthcare domain, personalized eHealth and mHealth services are significant to monitoring the patients' condition or normal health-conscious users (Papa *et al.*, 2020). This is useful in recording and updating the real-time condition of the users. Besides, it can save a patient’s life in any unexpected situation as soon as the caregiver or hospital is alerted. Moreover, the information of patients can be retrieved from the storage in a short time.