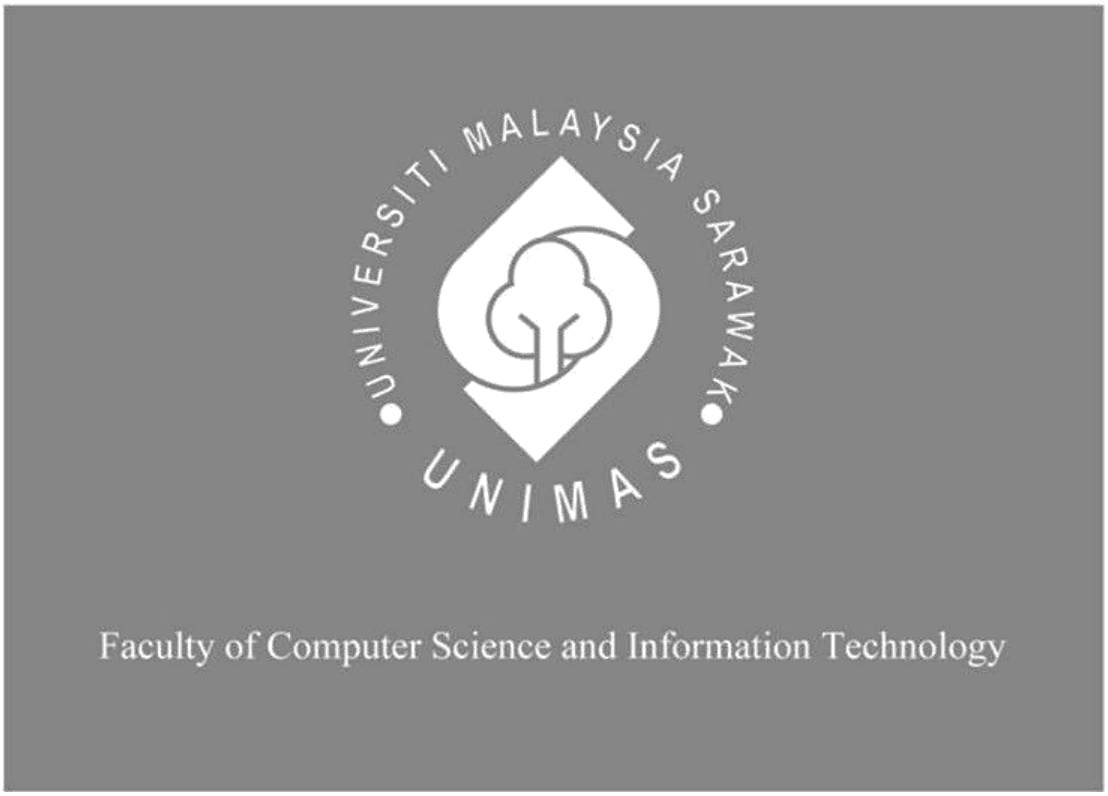

ANDROID APPLICATION QR PHISHING DETECTOR

PHANG CHOW CHING

Bachelor of Computer Science with Honours
(Computational Science)

2020

ANDROID APPLICATION QR PHISHING DETECTOR

PHANG CHOW CHING

This project is submitted in partial fulfilment

of the requirements for the degree of

Bachelor of Computer Science with Honours

Faculty of Computer Science and Information Technology

UNIVERSITI MALAYSIA SARAWAK

2020

UNIVERSITI MALAYSIA SARAWAK

THESIS STATUS ENDORSEMENT FORM

ANDROID APPLICATION QR PHISHING DETECTOR

ACADEMIC SESSION: 2019 / 2020

PHANG CHOW CHING

(CAPITAL LETTERS)

hereby agree that this Thesis* shall be kept at the Centre for Academic Information Services, Universiti Malaysia Sarawak, subject to the following terms and conditions:

1. The Thesis is solely owned by Universiti Malaysia Sarawak
2. The Centre for Academic Information Services is given full rights to produce copies for educational purposes only
3. The Centre for Academic Information Services is given full rights to do digitization in order to develop local content database
4. The Centre for Academic Information Services is given full rights to produce copies of this Thesis as part of its exchange item program between Higher Learning Institutions [or for the purpose of interlibrary loan between HLI]
5. ** Please tick (√)

☐	CONFIDENTIAL	(Contains classified information bounded by the OFFICIAL SECRETS ACT 1972)
☐	RESTRICTED	(Contains restricted information as dictated by the body or organization where the research was conducted)
☑	UNRESTRICTED	

Phang

(AUTHOR'S SIGNATURE)

Validated by

(SUPERVISOR'S SIGNATURE)

Permanent Address

24-2-08, Dahlia J, JLN S2 G5,
Garden Avenue, Seremban 2

Date: 3 August 2020

Date: 3.8.2020

Note * Thesis refers to PhD, Master, and Bachelor Degree

** For Confidential or Restricted materials, please attach relevant documents from relevant organizations / authorities

ACKNOWLEDGEMENT

I would like to show my deepest gratitude to my project supervisor, Dr. Chiew Kang Leng for his guidance, advice and encouragement throughout the project. Besides, I also want to thank Prof. Wang Yin Chai for providing us lectures with the guideline of doing the final year project. Lastly, I would like to take this opportunity to express my gratitude to my family for supporting me financially and friends who have shared their opinion and suggestion for the project.

ABSTRACT

Phishing has brought many damages to the user such as exploit and expose the sensitive data of the user. Phishing is normally done through calls, SMS and email. However, phishing now attacks through QR code to redirect the user to the intended website to retrieve the sensitive data. Due to the QR code which is not readable for the human eye, many user falls into the trap because not able to distinguish the legitimate and phishing QR code. The main objective of this project is to develop an Android app to detect phishing QR code. Besides, the proposed mobile application can protect the user from been redirected to the phishing website. Rapid application development will be used as the methodology of this project. The heuristic rules was evaluated with a dataset of 100 phishing URLs and 100 legitimate URLs. The analysis and result showed that the proposed mobile application can reduce QR code phishing by 88.5%. The significant of the proposed project is to protect the user from phishing attack which can result in financial loss. It also raises the security awareness of the QR code.

ABSTRAK

Phishing telah membawa banyak kerosakan kepada pengguna seperti mengeksploitasi dan mendedahkan data sensitif pengguna. Phishing biasanya dilakukan melalui panggilan, SMS dan e-mel. Walau bagaimanapun, phishing kini menyerang melalui kod QR untuk mengalihkan pengguna ke laman web yang palsu untuk mengambil data sensitive dari pengguna. Oleh kerana kod QR tidak dapat dibaca untuk mata manusia, ramai pengguna jatuh ke dalam perangkap kerana tidak dapat membezakan kod QR yang sah dan phishing. Objektif utama projek ini adalah untuk membangunkan aplikasi Android untuk mengesan kod QR phishing. Selain itu, aplikasi mudah alih yang dicadangkan dapat melindungi pengguna daripada dialihkan semula ke laman web apabila URL itu mencurigakan dan mengandungi phishing. "Rapid Application Development" akan digunakan sebagai metodologi projek ini. Peraturan heuristik dinilai dengan set data 100 URL pancingan data dan 100 URL sah. Analisis dan hasilnya menunjukkan bahawa aplikasi mudah alih yang dicadangkan dapat mengurangkan phishing kod QR sebanyak 88.5%. Ketara projek yang dicadangkan ini adalah untuk melindungi pengguna daripada mengeksploitasi data sensitif yang boleh mengakibatkan kerugian kewangan. Ia juga meningkatkan kesedaran keselamatan tentang kod QR.

TABLE OF CONTENTS

ACKNOWLEDGEMENT i
ABSTRACT ii
ABSTRAK iii
TABLE OF CONTENTS iv
LIST OF FIGURES vii
LIST OF TABLES viii
CHAPTER 1: INTRODUCTION 1
1.1 Introduction 1
1.2 Problem Statement 2
1.3 Scope 2
1.4 Aims and Objectives 2
1.5 Brief methodology 2
1.6 Significant of Project 4
1.7 Project Schedule 4
1.8 Expected outcome 5
1.9 Thesis Outline 5
CHAPTER 2: LITERATURE REVIEW 7
2.1 Introduction 7
2.2 Review on the Features and Functions of Related Applications 7
2.2.1 QR Scanning 7
2.2.2 User Interface Design 9
2.2.3 Process after QR Scanning 11
2.3 Review of Phishing Detection Technique 12
2.3.1 Blacklist-based detection 12
2.3.2 Whitelist-based detection 13
2.3.3 Machine Learning-based detection 13
2.3.4 Heuristic-based detection 13
2.4 Comparison between the Existing System and Proposed Application 14
2.5 Review on the Tools and Technologies used 15
2.5.1 Android Studio 15
2.5.2 Balsamiq 15
2.6 Summary 16
CHAPTER 3: Requirement and Analysis Design 17

3.1 Introduction 17
3.2 Rapid Application Development 17
3.3 Requirement Planning 18
3.3.1 Software Requirement 18
3.3.2 Hardware Requirement 19
3.3.3 Functional Requirement 19
3.4 System Design and Development 22
3.4.1 Use case diagram 23
3.4.2 Use case description 24
3.4.3 Activity diagram 25
3.4.4 Sequence diagram 27
3.4.5 Class diagram 28
3.4.6 User Interface 29
3.5 Construction 31
3.6 Cutover 32
3.7 Summary 32
CHAPTER 4: Implementation 33
4.1 Introduction 33
4.2 Implementation of QR Phishing Detector 33
4.2.1 Camera Activity 34
4.2.2 Result Activity 35
4.2.3 History Activity 36
4.3 Implementation of Phishing Detection Technique 36
4.4 Summary 39
CHAPTER 5: Testing and Evaluation 40
5.1 Introduction 40
5.2 Functional Testing 40
5.2.1 Evaluation Results 40
5.2.2 Comparison of threshold and result 43
5.3 System Testing 44
5.4 Summary 44
CHAPTER 6: Conclusion and Future Work 45
6.1 Introduction 45
6.2 Project Achievement 45
6.3 Limitation 45

6.4 Future Work 46
6.5 Conclusion.....46
REFERENCES 47

LIST OF FIGURES

Figure 1.1. Modified Graphical Representation of Rapid Application Development (LucidChart, 2018).....3
Figure 1.2. Gantt Chart 4
Figure 2.1. Kaspersky QR Scanner.....9
Figure 2.2. G DATA QR Scanner.....9
Figure 2.3. Free Scanner by Avira.....10
Figure 2.4. QR Droid.....11
Figure 3.1. Modified Graphical Representation of Rapid Application Development (LucidChart, 2018).....17
Figure 3.2. Use Case Diagram.....23
Figure 3.3. Activity diagram of Scan QR code.....25
Figure 3.4. Activity diagram of view history.....26
Figure 3.5. Sequence diagram of scan QR code.....27
Figure 3.6. Sequence diagram of view history.....28
Figure 3.7. Class diagram.....28
Figure 3.8. Main Page.....29
Figure 3.9. Safe Message Page.....30
Figure 3.10. Warning Message Page.....30
Figure 3.11. History Page.....31
Figure 4.1. dependencies of the application.....33
Figure 4.2. Camera Activity.....34
Figure 4.3. Result of legitimate link.....35
Figure 4.4. Result of phishing link.....35
Figure 4.5. History Activity.....36
Figure 4.6. Age of Domain.....37
Figure 4.7. HTTP Protocol.....38
Figure 4.8. Dots and slashes.....38
Figure 4.9. IP Address in URL.....38
Figure 4.10. Length of URL.....39
Figure 4.11. Special Characters.....39

LIST OF TABLES

Table 2.1. Comparison of the existing and proposed mobile scanner QR code application 14
Table 3.1. Hardware requirement 19
Table 3.2. Heuristic-based for detect phishing URL 21
Table 3.3. Scan QR code 24
Table 3.4. Check history 24
Table 5.1. Confusion Matrix 40
Table 5.2. Result of phishing URLs 41
Table 5.3. Confusion matrix for threshold value 3 41
Table 5.4. Comparison of threshold and result 43
Table 5.5. Test Case 44

CHAPTER 1: INTRODUCTION

1.1 Introduction

Quick Respond (QR) code is a two-dimensional barcode that was first created by Denso Wave to track the vehicle part during manufacturing in Japan (History of QR Code , n.d.). It also can store more data compared to the one-dimensional barcode. Nowadays, QR code is getting more popular and being used for marketing and for cashless payment, pre-book the event, information about the company or the event and even used as smart attendance. This is because the QR code has fast readability and very convenient. However, the security of the QR code is a concern as the attacker found the vulnerability of QR code to launch the QR phishing or also known as QRishing. A lot of victims fall into it because of QR code look very similar visually. According to Petrova (2016), QR codes are very hard to identify using the human eye and the addresses stored in the QR codes use URL shortening services to hide the true URL.

Phishing is a criminal activity that redirects the user to a cloned website which looks similar to the original website and prompt user to enter their privacy data illegally. Phishing is commonly used in email platform. QRishing happens when an attacker can simply replace the QR code with a fake QR code of an official website and using the URL shortening service to redirect the user to a phishing website and retrieve their personal data such as username, password and credit card number.

Currently, there are several Android apps that can prevent from phishing attack. Kaspersky Lab (2015) has released a QR Scanner, which is able to scan then detect phishing and malicious website that encoded in QR code. If QR code direct user to phishing or malicious website, it will warn the user. This significantly raises the security awareness of the QR code.

1.2 Problem Statement

Majority of the mobile QR code Scanner are using blacklist-based method to check the phishing website. Expose and exploit user data after redirect to the phishing website by scanning the fake QR code often occur. User cannot identify the QR code phishing even though showing the full URL that embedded to the QR code.

1.3 Scope

This project will be developed in Android OS platform. The Android application will have camera function to scan QR code. The application is able to detect phishing and legitimate QR code.

1.4 Aims and Objectives

The aims and objectives of the projects are:

- To create an Android app to detect phishing QR code.
- To protect a user from redirected to phishing website.
- To use the heuristic method to determine the phishing website and legitimate website.

1.5 Brief methodology

The selected methodology for this Android application QR phishing detector is rapid application development (RAD) which is an agile project management strategy. As the mobile application has a shorter development life cycle compared to the desktop application, RAD is

the best option. RAD has emphasised more on the rapid prototyping over planning and requirement analysis. Besides, RAD modal also uses a shorter time to finish and able to produce a more quality project compare to waterfall modal.

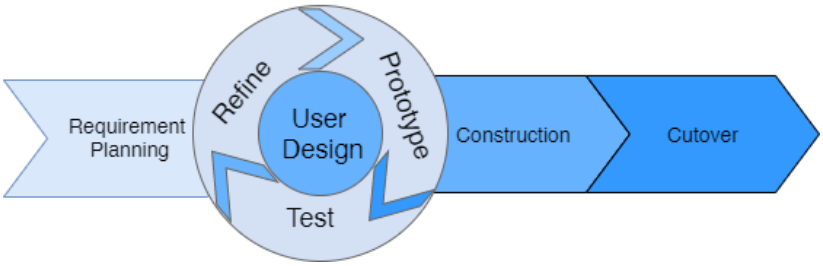

Figure 1.1. Modified Graphical Representation of Rapid Application Development (LucidChart, 2018)

1. Requirement Planning

   During this phase, all the system requirement will be collected to solve the problem of the project. Besides, similar mobile applications that are available in the Google Play store will be download and tested to collect the necessary information.

2. User Design

   In this phase, the user interface will be designed to meet the user experience. Then, it will consecutively test and refine based on the requirement to improve the prototype until all the requirement is achieved.

3. Construction

   All the design, feature and requirement of the application collected will be turned into source code. At the same time, the working prototype of the application will be built and improve.

4. Cutover

   The completed and refined mobile application that can solve the problem will be released.

1.6 Significant of Project

This project can protect the user from lose sensitive data such as username, passwords and credit card which can result in financial loss. User will not be redirected to the website and will be able to check the QR code whether phishing or legitimate effectively by scanning it. Besides, it also raises the security awareness of QR code.

1.7 Project Schedule

ID	Task Mode	Task Name	Duration	Start	Finish	Predecessors	Resource Names
1		**Android Application QR Phishing Detector**	**265 days**	**Wed 11/9/19**	**Mon 1/6/20**		
2		**Brief Proposal**	**19 days**	**Wed 11/9/19**	**Sun 29/9/19**		
3		Prepare brief proposal	2 days	Wed 11/9/19	Thu 12/9/19		
4		Discussion with supervisor	1 day	Fri 13/9/19	Fri 13/9/19	3	
5		Research on QR code phishing	15 days	Sat 14/9/19	Sat 28/9/19	4	
6		Submit brief proposal	1 day	Sun 29/9/19	Sun 29/9/19	5	
7		**Final Proposal**	**20 days**	**Mon 30/9/19**	**Sat 19/10/19**	2	
8		Research on methodology	4 days	Mon 30/9/19	Thu 3/10/19		
9		Amendment after meeting supervisor	15 days	Fri 4/10/19	Fri 18/10/19	8	
10		Finalize and submit final proposal	1 day	Sat 19/10/19	Sat 19/10/19	9	
11		**Chapter 1 : Introduction**	**7 days**	**Sun 20/10/19**	**Sat 26/10/19**	7	
12		Amendment from proposal	4 days	Sun 20/10/19	Wed 23/10/19		
13		Compile and submit Chapter 1	3 days	Thu 24/10/19	Sat 26/10/19	12	
14		**Chapter 2: Literature Review**	**21 days**	**Sun 27/10/19**	**Sat 16/11/19**	11	
15		Research on Existing QR apps	19 days	Sun 27/10/19	Thu 14/11/19		
16		Compile and submit chapter 2	2 days	Fri 15/11/19	Sat 16/11/19	15	
17		**Chapter 3: Requirement Analysis and Desi**	**19 days**	**Sun 17/11/19**	**Thu 5/12/19**	14	
18		Start on System Design	3 days	Sun 17/11/19	Tue 19/11/19		
19		Compile and submit Chapter 3	16 days	Wed 20/11/19	Thu 5/12/19	18	
20		**FYP Final Report**	**13 days**	**Fri 6/12/19**	**Wed 18/12/19**	17	
21		Finalize final report	6 days	Fri 6/12/19	Wed 11/12/19		
22		Submit final report	1 day	Thu 12/12/19	Thu 12/12/19	21	
23		prepare presentation slide	4 days	Fri 13/12/19	Mon 16/12/19	22	
24		Final Year Project Symposium	2 days	Tue 17/12/19	Wed 18/12/19	23	
25		**FYP 1 Report**	**24 days**	**Thu 19/12/19**	**Sat 11/1/20**	20	
26		Discussion with supervisor	1 day	Thu 19/12/19	Thu 19/12/19		
27		Amendment and Modification	22 days	Fri 20/12/19	Fri 10/1/20	26	
28		Submit FYP 1 final report	1 day	Sat 11/1/20	Sat 11/1/20	27	
29		**Chapter 4: Implementation and Testing**	**50 days**	**Mon 27/1/20**	**Mon 16/3/20**	25	
30		Implement Solution	9 days	Mon 27/1/20	Tue 4/2/20		
31		Construct Prototype	30 days	Wed 5/2/20	Thu 5/3/20	30	
32		Tesing and evaluation	10 days	Fri 6/3/20	Sun 15/3/20	31	
33		Compile and submit chapter 4	1 day	Mon 16/3/20	Mon 16/3/20	32	
34		**Chapter 5: Conclusion and Future Work**	**25 days**	**Tue 17/3/20**	**Fri 10/4/20**	29	
35		Analyze and Discuss Result	13 days	Tue 17/3/20	Sun 29/3/20		
36		Writeup Project report	5 days	Mon 30/3/20	Fri 3/4/20	35	
37		Writeup Project limitation and improvemer	4 days	Sat 4/4/20	Tue 7/4/20	36	
38		FYP 2 Symposium	2 days	Wed 8/4/20	Thu 9/4/20	37	
39		Compile and submit FYP 2 report	1 day	Fri 10/4/20	Fri 10/4/20	38	

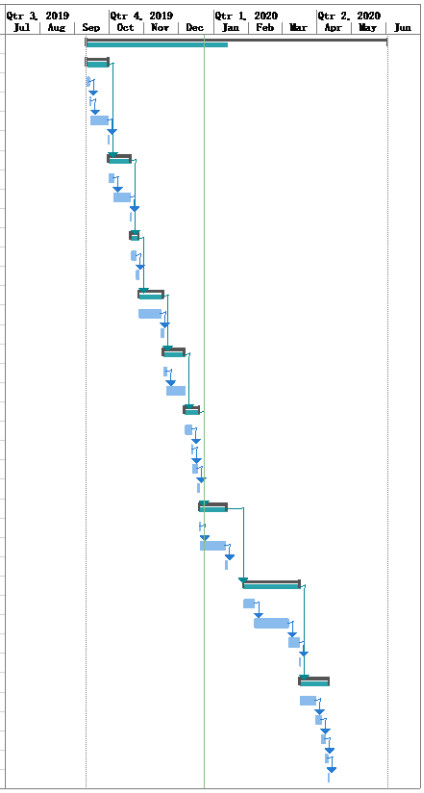

Figure 1.2. Gantt Chart

1.8 Expected outcome

The expected outcome of this project is an Android application that can detect the phishing website encoded in the QR code and protect user from redirected to the phishing website. The application can differentiate the QR code phishing and legitimate.

1.9 Thesis Outline

The project outline consists of these chapters:

Chapter 2: Literature Review

In this chapter, the existing android mobile QR scanners that are available on Google play will be download for review purpose. The comparison of the feature and function of the existing mobile QR scanner will be documented. Besides, the technology and tools used will be reviewed in this chapter.

Chapter 3: Requirement Analysis and Design

This chapter will discuss the requirement analysis of rapid application development methodology in detail. The UML diagram of the system design will be presented in this chapter. The concept of the User Interface of the proposed mobile application will be included in this chapter.

Chapter 4: Implementation and Testing

The implementation of the proposed project will start in this chapter. Coding and debugging will be done to make sure it meets the objective of the project. Testing will be done to verify the proposed mobile application. Besides, the result of the testing data will be recorded in this chapter.

Chapter 5: Conclusion and Future Work

In this chapter, conclusion and lesson learned throughout the project will be stated here. Next, the limitation and future improvement of the project will be recorded to improve the quality of the project.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The QR code is now universally used for advertising, transaction, website, and name card because of the QR code is convenient by just scanning the QR code to get the information. Now, all types of QR scanner application is all over in the Google Play Store and App Store. As the QR code is not readability for the human eye, the attackers take advantage of it to do QR code phishing to trick user such as to replace the existing QR code and alter individual modules of the QR code (Kieseberg, et al., 2010). Now, there are only a few QR scanner applications can validate the URL embedded in the QR code. Vidas, et al., (2013) state that there is some QR code reader application instantly redirect the user to the URL without showing the content of the QR code. Besides, the user also not possible to view or verify the full address of the URL in the mobile phone due to the small screen size. As a result, plenty of victims tend to fall into the trap and sensitive data get divulge.

2.2 Review on the Features and Functions of Related Applications

2.2.1 QR Scanning

2.2.1.1 Kaspersky QR Scanner

Kaspersky QR Scanner will scan and decode the QR code directly at any angle when the camera is pointed at the QR code (Kaspersky QR Scanner, n.d.). It can scan 1D and 2D format of the code. Kaspersky QR code has a built-in web browser to open the URL embedded in the QR code after scanning. It will scan the redirected URL instead of validating the URL encoded in a QR code (Dudheria, 2017). It will also store the information scanned in the device. Kaspersky QR code requires an internet connection to scan the link in the QR code.

2.2.1.2 Avira Insight & QR Scanner

Avira Insight & QR Scanner using Avira URL Cloud technology to validate the URL that is encoded in the QR code (Avira QR Scanner, n.d.). It requires an internet connection to check the content in the QR code else it will show error after scanning. The downside of Avira QR Scanner is the speed of scanning QR which is slower than Kaspersky QR Scanner as the QR code must point directly to the camera to decode.

2.2.1.3 G DATA QR Scanner

G DATA QR Scanner will scan the QR code and show the result of the URL link to for the user to open in browser or report the URL to G DATA if it is suspicious (G DATA, n.d.). It will save the history of the QR code scanned in the application. G DATA QR Code Scanner also scan the final redirected URL in the QR code. It requires an internet connection to scan and check the URL in the QR code else it will assume the scanned QR code is a phishing site.

2.2.1.4 QR Droid

In QR Droid, it will scan the QR code and use the Google browser to open the link contained in the QR code (QR Droid Zapper, n.d.). Dudheria (2017) found that QR Droid using Google Safe Browsing database to check the URL in the QR code before it redirects the user to the link. It requires an internet connection to check and validate the URL since it opens the URL in Google browser.

2.2.2 User Interface Design

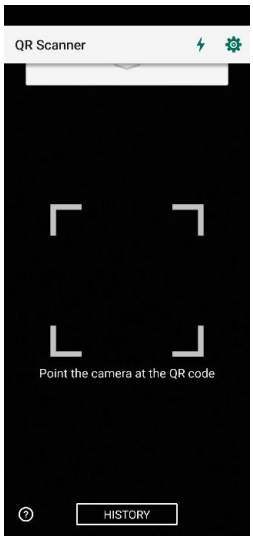

Figure 2.1. Kaspersky QR Scanner

Figure 2.1 shows the Kaspersky QR Scanner has a simple user interface design. It is very convenient to scan as the main screen of the app is the camera with the crosshair to guide the user to point at the QR code. It also provides history function to allow the user to search the scanned QR code and flashlight function for user to scan QR in the dark place.

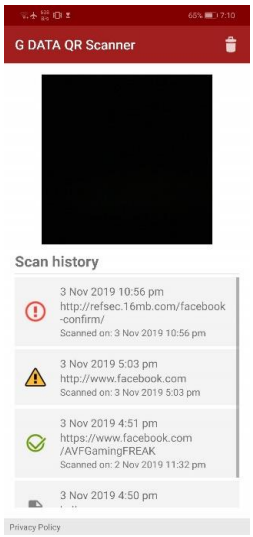

Figure 2.2. G DATA QR Scanner

G DATA QR Code Scanner has a unique and simple user interface design where the user can see the small size of the camera view and history of scanned QR code at the main screen. Figure 2.2 shows the interface of G DATA QR Scanner. This is very convenient for a new user as they can watch everything on the first page.

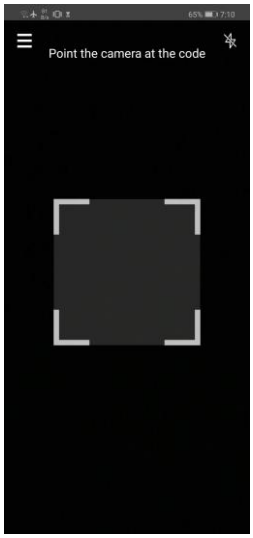

Figure 2.3. Free Scanner by Avira

Figure 2.3 shows the interface of the Free QR Scanner by Avira. Free QR Scanner by Avira also has almost the same user interface design as Kaspersky QR Scanner. The difference is the camera function and flash feature are on the main screen only, the other function is on the menu lists.

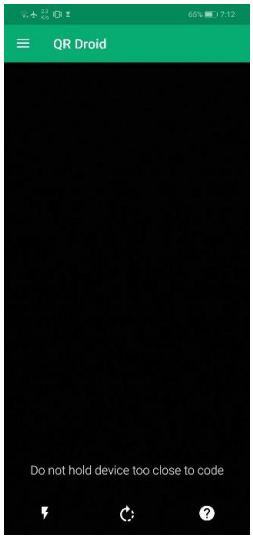

Figure 2.4. QR Droid

Figure 2.4 shows the interface of QR Droid. QR Droid also has the simplest user interface of QR reader. Moreover, it also has the most complex QR reader among the other four scanners reviewed. User can manually set the setting for display, camera, and action after scanning.

2.2.3 Process after QR Scanning

Kaspersky QR Scanner does not display the full URL. It asks the user for confirmation if it detected the link is suspicious or containing phishing and malware encoded in the QR code after scanning. It allows the user to open the phishing website after it detects the QR code is not safe. Kaspersky QR Scanner will automatically open the link of the website when it checked the link is safe.

Avira QR Scanner will show the user the full URL of the scanned QR code and allow the user to open in browser when it detected the URL is safe. It will warn and restricted the user to open the URL if it detected the URL in the QR code as unsafe.

G DATA QR scanner will show the result of the QR code and full address of the URL and ask the user for confirmation before the user opens in the browser directly. It will stop the user to enter the website if the QR code is dangerous.

QR Droid allows the user to set the default action after scanning the QR code. If the default action is set to open URL, then it will not ask user confirmation and open the URL in a browser after scanning the QR code instantly. QR Droid has a special feature which is the preview function after scanning the QR code. This allows the user to preview the content and final URL of the website before opening it in the browser.

2.3 Review of Phishing Detection Technique

2.3.1 Blacklist-based detection

Blacklist method detection is a method that uses a database containing the list of phishing and malware website to validate the URL. The advantage of blacklisting is it is easy to implement. Khonji, Iraqi, & Jones (2013) found that the blacklist method has a high rate of detection and a low false-positive rate than Heuristic. The disadvantage of it is the list of phishing or malware website must constantly update from time to time in order to be effective (Lee, Kim, & Lee, 2015). On the other hand, the downside of using blacklisting is it cannot detect the zero-day phishing website which is the lifetime of the phishing website is less than 24 hours (Sheng, et al., 2009).

2.3.2 Whitelist-based detection

Whitelist method detection is a database only containing the list of a genuine or legitimate website. If the examined website is not in the whitelist, it will consider as suspicious (Sahingoz, Buber, Demir, & Diri, 2019). This method can validate and differentiate the suspicious and phishing website effectively. It has the same advantage and disadvantage as blacklist-based detection (Varshney, Misra, & Atrey, 2016).

2.3.3 Machine Learning-based detection

According to Khonji, Iraqi, & Jones (2013) found that Machine Learning-based detection has high accurate detection among the existing approach where the false positive rate is not more than 1%. It is also able to detect the zero-day attack. In learning-based detection system, it uses a learning algorithm to detect the hidden or uncategorised URL through a noisy dataset that contains various features of the phishing and legitimate website (Sahingoz, Buber, Demir, & Diri, 2019). However, machine learning-based detection requires high test and training time to learn plenty of features to determine the phishing website (Sahoo, Liu, & Hoi, 2017).

2.3.4 Heuristic-based detection

Heuristic-based detection is the most common detection approach as well as the blacklist approach. In order to detect the phishing website effectively, it will analyse and extract the feature of the phishing website. For example, the domain name of the website, the age of domain, character contain in a URL and so on. This can detect the new phishing site and