International Journal of Cyber-Security and Digital Forensics (IJCSDF) 8(2): 134-143
The Society of Digital Information and Wireless Communications (SDIWC), 2019 ISSN: 2305-001

Anti-Qrishing Real-Time Technique on the QR Code Using the Address Bar-Based and Domain-Based Approach on Smartphone

Kurniadin Abd. Latif [1], Bambang Sugiantoro [2], Yudi Prayudi [3]
[1, 3]Department of Informatics Engineering, Islamic University of Indonesia
[2]Department of Informatics Engineering, Sunan Kalijaga State Islamic University
[1]kimkurniadin@gmail.com, [2]bambang.sugiantoro@uin-suka.ac.id, [3]prayudi@uii.ac.id

ABSTRACT
Phishing is a technique of cybercrime that commits fraud by masquerading as trusted entity. The focus of phishing is to get sensitive information. At present, the development of phishing has used QR code as phishing media. The aim of phishers is to cheat the target by visiting fake webs via QR code. This research uses the Address Bar Based and Domain-based approach. The Address bar-based approach is an approach to examine some unique criteria from a URL. A domain-based approach is a third-party site that provides real-time domain checks (Website traffic and age of the domain). Additional blacklist features from Phistank are complementary features. This research aims to design an Anti-Qrishing technique to detect Qrishing (QR Code Phishing) in Real-Time via Smartphone. The test was conducted with 500 QR Code, consisting of 300 QR Code phishing with various manipulation techniques, 120 safe QR Code and 80 suspicious QR Code. This test gets 98,4% accuracy, and 1,6% error detection.

KEYWORD
QR Code, Phishing, Qrishing, Address Bar-Based, Domain-Based.

1. INTRODUCTION
Phishing is a type of psychological manipulation attack that focuses on getting sensitive information by masquerading as a trusted entity [1]. Based on the Oxford dictionary, Phishing is an alternative to the word "fishing". The point is to provoke the target to provide personal information through a website that has been manipulated by the perpetrator or called a phisher. Based on quarterly survey reports conducted by Infosec professionals in 2018, State of the Phish Report, the impact of phishing in 2017 was 49% malware infection, 38% Compromised Accounts, and 13% Loss of data [2].

The cases of phishing have developed rapidly and increased in numbers significantly. Phishing is not only through messages and e-mail, but also uses the QR Code technique to trick victims. The QR Code generally contains URLs, SMS, Contacts, Addresses, Plain Text, Geolocation [3]. The QR Code is used by several companies and institutions in various activities such as payment, product information, news, advertisements, coupons, tickets, app downloads, loyalty programs, social media and others. Growth in the use of QR Code has increased which has led to increased crime growth. Phishers embed phishing or malicious URLs on QR Code called Qrishing (QR Code Phishing).

The QR Code is difficult to recognize compared to other techniques. This scheme becomes a profitable opportunity for phishers. This research proposes the Anti-Qrishing technique as Qrishing prevention step. This Anti-Qrishing technique uses the Address Bar Based and Domain-based approach as parameter to detect phishing. In addition, the blacklist feature was used as a complement to this research. Anti-Qrishing technique aims to detect Qrishing in real-time.

2. RELATED WORK
Research on phishing has been done previously by several researchers. Research [4] proposed a string matching method for detecting phishing attacks. This method determines the rate of URL similarity with the blacklist URL. Blacklist detection is divided into several parts to match, such as: IP, HostName, Directory Structure, Brand Name. The results of this research are effective in detecting phishing attacks with very low false negative and false positives. Two string matching methods are implemented. The Longest Common Subsequence provides an accuracy rate of 99.1% and Edit Distance the accuracy rate is 99.5%. However, this research has not been able to detect phishing 0 day or phishing new models. This method relies on blacklisting so that additional methods are needed to help phishing detection of new models.

Research doing by Wardman [5] present a set of methods that demonstrate a relationship

between phishers and defacers. Highlighting this relationship assists in building substantial defenses and law enforcement cases against this threat and shows that the proposed strategy can be used to predict when and where new phishing websites and related attacks will surface next. This research is not optimal for detecting phishing new models, because this study predicts based on the relationship between phishers and defacers.

Research doing by Ahmed [6], try to help the user distinguish between legitimate web pages and phishing in real time by using URLs as indicators. This study uses the Address Bar Based approach as a decision maker and added to the blacklist feature. However, this research cannot detect 0-day phishing or new phishing models. This method relies on blacklisting and limited phishing features, so additional methods are needed to help phishing detection of new models.

Meanwhile, research from [7] is to build Qrphish or QR Phishing, which is an effective mechanism for detecting phishing URLs in a QR Code. This methodology does not only exploit basic phishing detection features that depend on the URL and features of suspicious web pages. Specific QR Code and host-based features are used in this study. Machine learning is also used to make classifications. However, this study has false positives that are not maximal, because the number of datasets is less maximal and limited.

Another research from [8], proposes authentication simulation procedures for users through the hashcode comparison technique. UnPhishMe intercepts a login page opened by a user and simulates the login procedure with fake credentials. Technically, an authentication attempt to a login webpage with incorrect login credentials tests the trustworthiness of that page. However, a user needs to have a piece of prior knowledge and remembers to do so every time she encounters a suspicious page.

Solution using a parse tree validation approach from [9], proposes a parse tree validation approach to determine whether a web page is legitimate or phishing. This is a new approach to detect phishing websites by intercepting all hyperlinks from the current page through the Google API, and building parses trees with intercepted hyperlinks. But in this research, there were false positives and false negatives that were not maximal. Lexical analysis can be used to find Phishing websites.

3. RESEARCH METHODOLOGY

This research is the development of Anti-Qrishing techniques to prevent Qrishing. Anti-Qrishing technique consists of several features which are parameters in determining whether the URL is phishing or not. In this research using the Address Bar Based and Domain-Based approaches with additional Blacklist features, including:

3.1 Address Bar Based

The term address bar refers to a text field in a web browser that identifies the user's location on the web and allows them to access different websites. This address bar is where the URL input refers to a website. This address bar-based is used to check phishing characters in the URL. The features that are used in this approach include:

3.1.1 URL Shortening Services

This is accomplished by means of an “HTTP Redirect” on a domain name that is short, which links to the webpage that has a long URL. For example, the URL “http://portal.hud.ac.uk/” can be shortened to “bit.ly/19DXSk4” [10].

Rule: IF

$$\begin{Bmatrix} \text{URL Shortening} \rightarrow \text{URL Redirect Analysis} \\ \text{Otherwise} \rightarrow \text{URL Analysis} \end{Bmatrix}$$

3.1.2 Using Protocol “http” on URL

Legitimate websites use secure domain names whenever sensitive information is transferred. The existence of https is important in giving the legitimacy of the website [11]. A site that is secure uses protocol that is secure in conducting transactions. The safe protocol to use is "https". Sites that use the "https" protocol do not mean that the site is truly safe. However, safe sites use the https protocol.

$$\text{Rule : IF} \begin{Bmatrix} \text{URL using http} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{Bmatrix}$$

3.1.3 Using Symbol “@” on URL

Using “@” symbol in the URL leads the browser to ignore everything preceding the “@” symbol

and the real address often follows the "@" symbol [10].

$$\text{Rule: IF} \begin{Bmatrix} \text{Url using @ Symbol} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{Bmatrix}$$

3.1.4 URL contains ".exe" or "apk"

Phishers usually don't just direct users to their site through fake URLs to get credential information. Phishers also embed malware files such as ".exe" or ".apk" files on their fake URLs, so that phishers can do some malicious actions against the user.

$$\text{Rule: IF} \begin{Bmatrix} \text{URL contain.exe / apk} \rightarrow \text{Phishing} \\ \text{else} \rightarrow \text{Legitimate} \end{Bmatrix}$$

3.1.5 Length of URL

Phishers can use long URLs to hide doubtful parts in the address bar. Long URLs are usually used by phishers to hide suspicious parts. There is no definite length to show phishing sites [10] however, the authors report that the normal URL length does not exceed 75 characters.

$$\text{Rule: IF} \begin{Bmatrix} \text{If Length of URL} > 75 \rightarrow \text{feature=Phishing} \\ \text{else} \rightarrow \text{feature=Legitimate} \end{Bmatrix}$$

3.1.6 Using the IP Address

If an IP address is used as an alternative of the domain name in the URL, such as "http://125.98.3.123/fake.html", users can be sure that someone is trying to steal their personal information [9].

$$\underline{\text{Rule}}\text{: IF} \begin{Bmatrix} \text{IP Address Exist} \rightarrow \text{Phishing} \\ \text{Else} \rightarrow \text{Legitimate} \end{Bmatrix}$$

3.1.7 Redirecting using "//"

The existence of "//" within the URL path means that the user will be redirected to another website. An example of such URL's is: "http://www.legitimate.com//http://www.phishing.com" [9].

$$\text{Rule: IF} \begin{Bmatrix} \text{The Position of '// ' the URL} > 7 \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{Bmatrix}$$

3.1.8 Using URL Academy

The "ac" URL Academy is the URL of an educational institution at the university. This domain is officially registered with certain administrative requirements. It can be ascertained that this URL academy is safe from fraudulent purposes such as phishing.

$$\text{Rule: IF} \begin{Bmatrix} \text{Using URL Academy} \rightarrow \text{Legitimate} \\ \text{Otherwise} \rightarrow \text{Checking URL} \end{Bmatrix}$$

3.2 Domain Based

The domain is another name for "domain name" or the address of a website. The domain is used instead of the IP Address in searching for an address. Domain makes it easy to remember a web address. The domain-based approach is used to find information about a website. The feature used in this approach includes:

3.2.1 Age of Domain

Based on the fact that a phishing website lives for a short period of time, we believe that trustworthy domains are regularly paid for several years in advance. In our dataset, we find that the longest fraudulent domains have been used for one year only [10]. This research uses WHOIS [12] to check the age of the domain.

$$\text{Rule: IF} \begin{Bmatrix} \text{Age of Domains} \leq 1 \text{ years} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{Bmatrix}$$

3.2.2 Website Traffic

However, since phishing websites live for a short period of time, they may not be recognized by the Alexa database [13]. The database measures average number of daily users visiting a website and its page views in the last 3 months to determine its ranking position [14]. Legitimate websites ranked among the top 100,000 [10].

$$\text{Rule: IF} \begin{Bmatrix} \text{Website Rank} < 100{,}000 \rightarrow \text{Legitimate} \\ \text{Else if Website Rank} > 100{,}000 \rightarrow \text{Suspicious} \\ \text{Otherwise} \rightarrow \text{Phishing} \end{Bmatrix}$$

3.3 Blacklist Feature

The Blacklist feature or approach is an approach where URLs are matched with phishing URLs in the database [15]. This study uses Blacklist that has been provided by Phistank [16] through the API.

$$\text{Rule: IF} \begin{cases} \text{Databaseis false} \rightarrow \text{Legitimate} \\ \text{Else if Databaseis True} \rightarrow \text{Suspicious} \\ \text{Validationis true} \rightarrow \text{Phishing} \end{cases}$$

Based on some of these features, then it was designed in the form of a smartphone application. This technique is designed to detect Phishing in a QR Code. The design can be seen in Figure 1.

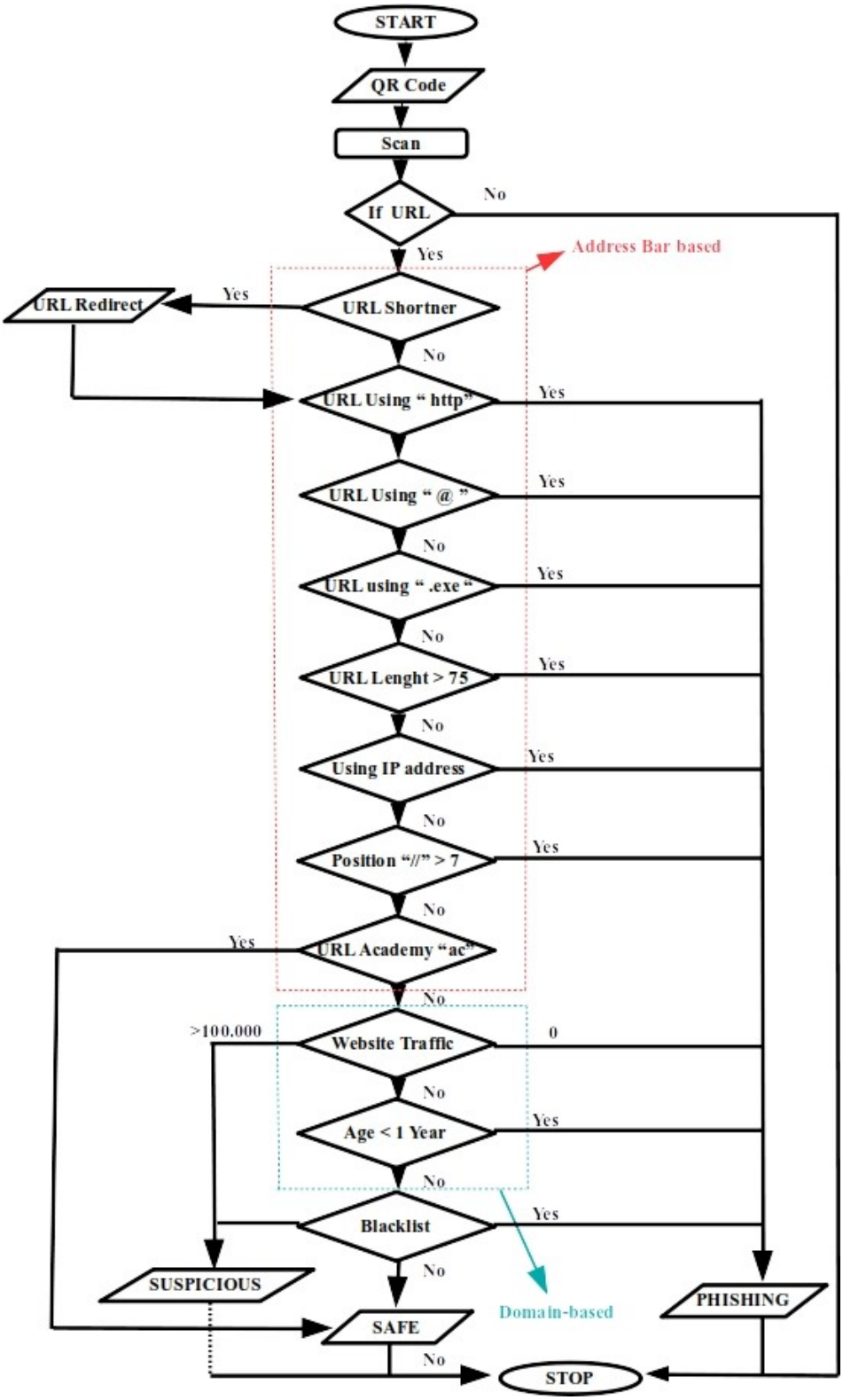

Figure 1. Flowchart of Anti-Qrishing Detection Technique

The above flowchart is a collection of features. Flowcharts are designed with a form of condition validation techniques. Each feature has a condition. The conditions of the features used are different, some have 3 conditions and there are those that have 2 conditions. The red box line is a sign of the elements inside representing part of the bar-based address. Whereas, the blue box line is part of domain-based. Each element of the feature is explained before Figure 1.

Anti-Qrishing is a phishing prevention technique on Qr Code. Anti_Qrishing focuses on Qr Code that has URL content. Anti-Qrishing Uses an Address bar-based and domain-based approach with additional blacklist from phistank. From these two approaches, there are 10 features and additional features blacklist. From some of the features, the URL Academy feature and URL shortener redirect analysis are new features proposed in this study. The detection technique used in this study is to use the technique of validating phishing character or conditions and online domain validation. The results of this detection process will produce results like "Safe, phishing and suspicious". Detection techniques can be seen in Figure 1.

4. RESULT AND ANALYSIS

4.1 Result

Based on the proposed Anti-Qrishing Technique, a case simulation will be carried out to test the design results. This simulation aims to determine whether the QR Code contains phishing or not. The simulation flow can be seen in Figure 2.

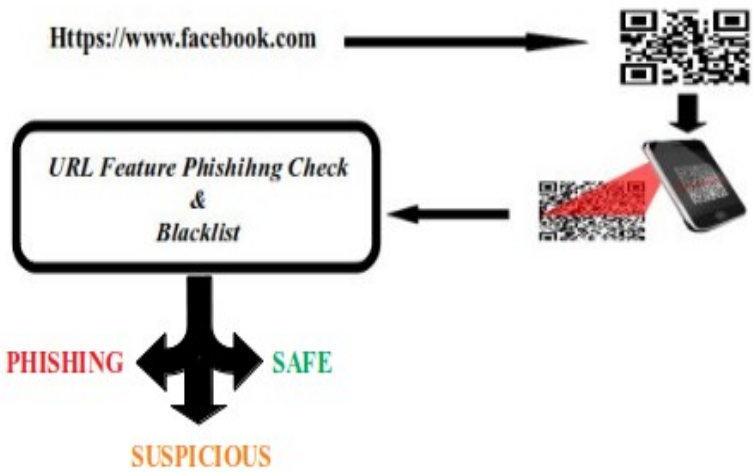

Figure 2. Qrishing Case Simulation

The stages in this case simulation include:

a. Create a QR Code by embedding a Phishing URL or a valid URL.
b. Perform scanning using the App that has been made.
c. Results that appear in the form of information such as safe, suspicious and phishing.

Scanning Process

The display of scanning results can be seen in Figure 3.

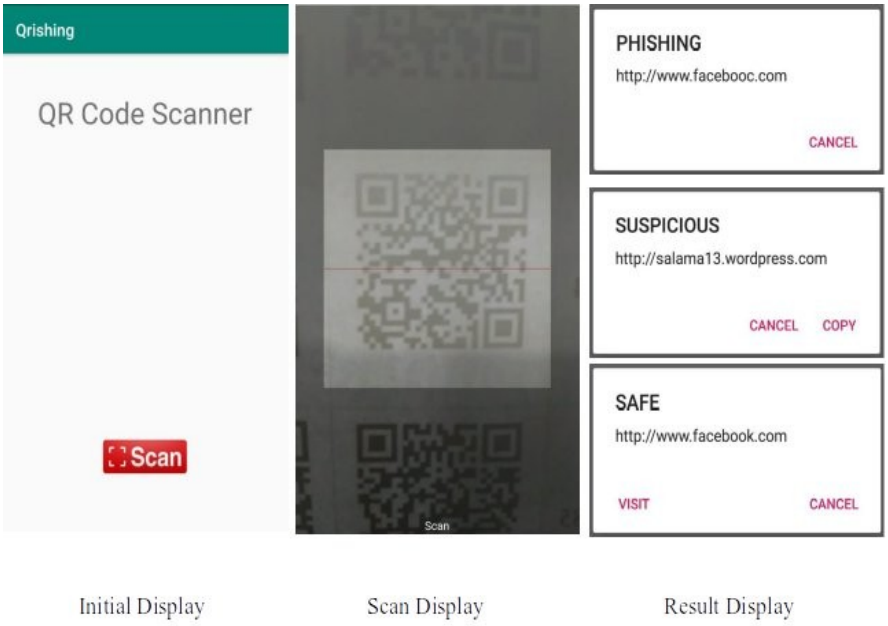

Figure 3. Application Interface

Result of Test

After going through a series of tests, then analyzed to find out the results of accuracy, false positives, and false negatives. This study scanned 500 QR Code consisting of 300 QR Code phishing, 120 safe QR Code and 80 suspicious QR Code. The results of testing can be seen in table 1.

Table 1. Result of Test

No.	Jenis	Technique of Manipulation	Number of QR Code	Safe	Suspecious	Phishing
1.	**Phishing**	URL Shortener	40	-	-	40
2.		Using IP	20	-	-	20
3.		Using Symbol "@"	40	-	-	40
4.		Using "//" > 1	40	-	-	40
5.		Using Unicode	40	-	-	40
6.		Using "exe"/ "apk"	40	-	-	40
7.		Web page manipulation	80	-	-	80
8.		Not Manipulation (Safe)	120	117	-	3
9.		Not Manipulation (Suspecious)	80	1	75	4
Total			**500**	**118**	**75**	**307**

Table 1 is the result of a scan that has been done using various kinds of manipulations on the Qr Code content. Manipulation carried out in this study consisted of the manipulation of the phishing character of the URL. The scan results get an error rate of 8 times. The error results obtained are a scanning process that is done in a hurry or the time lag between scanning with one another is too short, and is also done in a weak signal condition.

From the data in table 1, it can be made into the confusion matrix to facilitate the performance evaluation. The form of the confusion matrix can be seen in the table 1.

Table 2. Confusion Matrix

Type of Classification		Detection		
		Phishing	**Suspecious**	**Safe**
Actual	***Phishing***	300	0	0
	Suspecious	4	75	1
	Safe	3	0	117
Total		**307**	**75**	**118**

From confusion matrix, we can find the scanning results that have errors. The confusion matrix is one method that can be used to measure the performance of a classification method. Basically, the confusion matrix contains information that compares the results of the classification carried out by the system with the results of the classification that should be [17]. The blue table is the result of matching scanning with the data. While others are errors in detecting. From this confusion matrix, calculations are made in determining the level of accuracy and errors rate that occur during testing [18]. This performance evaluation uses equations (1) & (2) :

$$\text{Accuracy} = \frac{\text{N correct}}{\text{N}} x\,100 \qquad (1)$$

$$= \frac{492}{500} x\,100\% = 98{,}4\%$$

$$\text{Error Detection} = \frac{\text{N incorrect}}{\text{N}} x\,100 \qquad (2)$$

$$= \frac{8}{500} x\,100\% = 1{,}6\%$$

4.2 Analysis

This stage describes the proposed solution and comparison of existing research so that it gets the appropriate results. This research uses several features used by previous research, besides that it is added with proposed features. The proposed Anti-Qrishing technique is the URL Academy "ac" and the Redirect URL Shortener found in the Address-Based Bar Approach. The Academy URL is the URL of an educational institution at the university. This domain is officially registered with certain administrative requirements. It can be ascertained that this URL of academics is safe from fraudulent purposes phishing. URL Shortener is a condition of "HTTP Redirect" in the changed domain name to be short, so the results of the URL Redirect can be further analyzed.

The Domain Based approach there is Alexa's Web traffic. Alexa's Web traffic is used to determine the popularity of a domain. Besides that additional blacklist features of Phistank's is a database of phishing URLs collected by phistank. From these two features checks are carried out using the API. The comparison of proposed features with previous research can be seen in table 2.

Table 3. Comparison of Feature

Approach	Features	Research		
		Old		New (Proposal)
		(Alnajjar et al.,)	(Ahmed & Abdullah)	
Address Bar	Using IP	✓	✓	✓
	Using Symbol ‘@’	✓	✓	✓
	Position ‘//’ > 7 on URL	✓	✓	✓
	Length URL	✓	✓	✓
	Prefix Suffix “-” pada URL	✓	✓	
	The number of the dot “.” > URL	✓	✓	✓
	Sensitive Word	✓		

	Using Protocol “http://”			✓
	URL Academy “ac”			✓
	URL Shortener Redirect			✓
Domain	Page rank	✓		
	Age of Domain (WHOIS)	✓		✓
	Domain confidence level	✓		
	Country	✓		
	Update date	✓		
	Web traffic Alexa			✓
Blacklist	Blacklist		✓	✓

From table 2, it can be seen that this study combines features that have been used by previous research. From the combination, there is also a reduction or selection of features that can affect the results of detection. In addition to increasing the results, additional features are also made that can affect the results of detection. There are several feature proposals that can influence the results of detection, including the use of the "HTTP" protocol, the use of domain academy, URL Shortener Redirect, and improvements to the phistank blacklist.

Besides feature comparison analysis, the comparative analysis of phishing detection techniques was also carried out. This research uses a combination of condition validation techniques and online validation from some existing researches. The comparison of detection techniques can be seen in table 3.

Table 4. Comparison of detection techniques

Technique of Detection	Research		
	Old		New (proposal)
	(Alnajjar et al.,)	(Ahmed & Abdullah)	
Validation of Conditions		✓	✓
Classification of Machine Learning	✓		
Online Validations	✓		✓

The condition validation technique is a detection technique by checking the condition of the content from the QR code. This technique checks for features or characteristics that are similar to features that have been determined as phishing or not.

Then another technique is the detection technique with machine learning methods. this is the most recent detection technique. This technique must be supported by deceptive data in classifying whether the QR Code content is safe or not. This technique classifies based on the data that has been collected, so the results depend on data and algorithms.

Whereas, online validation is checking the domain on third-party services online, whether the domain can be trusted as safe or not.

In this study, combined condition validation and online validation to express better results, because it is not very dependent on data. This technique is useful for detecting new phishing models.

Based on the testing with 500 QR Code, consisting of 300 QR Code phishing with various manipulation techniques and 80 combined safe and suspicious QR Code. The results of this test get accuracy 98,4%, and error detection 1,6%. From these results, a comparison of the results of the accuracy of this study was carried out with related research. The comparison can be seen in table 4.

Table 5. Comparison of the detection result.

Research		Accuracy	Information
Old	(Alnajjar et al.,)	93,34 %	This research can detect phishing new models, but the level of accuracy is not maximal.
	(Ahmed & Abdullah)	96 %	This result is a string and database match, so it gets 96 % results, but this result cannot detect phishing in new models.
New	Anti-Qrishing	98,4 %	This study can do phishing detection of new models, but the level of accuracy is not maximal. The results are not optimal due to bad signals and scanning in a hurry.

From the data in table 4. the results of accuracy are different in each study. From these data, it was found that the proposed anti-Qrishing technique was effective in making detection on the QR Code. By combining, selecting and adding features, anti-Qrishing techniques can detect Phishing with a satisfactory level of accuracy. In addition, this technique can detect phishing new models.

The use of Anti-Qrishing technique is only applied to QR Codes that have URL content, not API keys from specific applications, such as payment applications, Whatsapp and others. QR Code payment is a form of the digital transaction using a QR code as an authentication media. QR Code on WhatsApp also used as authentication. The application used for scanning is a specific application that has been integrated with the system from the service provider. Generally, content generated to form a QR Code is API Keys or specific code known by the system. In Anti-Qrishing Technique, the API Keys are recognized as plain text, causing no further analysis. The development of the Anti-Qrishing technique in the future is to add features that can recognize phishing or hijacking from QR Code payment, WhatsApp and others.

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

The QR Code detection process uses the Address Bar Based and Domain-Based Approaches. The use of the Address Bar Based Approach in this study is to examine the unique string criteria of phishing contained in the URL so that it can be easily detected as phishing. The use of the Domain Based approach is to check domains to third parties online and real time by utilizing the API. So this technique can detect phishing codes that detect better.

The test was conducted with 500 QR Code, consisting of 300 QR Code phishing with various manipulation techniques and 120 safe QR Code and 80 suspicious QR Code. This test gets 98,4% accuracy, and 1,6% error detection. Anti-Qrishing technique is a solution for detecting QR Code Phishing or Qrishing on smartphones in real-time. This technique can detect 0-day phishing or new model of phishing.

5.2 Future Work

This research focuses on a QR code that contains URL content. Further research can be developed for phishing QR Code detection techniques in special applications such as Payment and others.

REFERENCES

1. T. Peng, I. Harris, and Y. Sawa, "Detecting Phishing Attacks Using Natural Language Processing and Machine Learning," *2018 IEEE 12th Int. Conf. Semant. Comput.*, pp. 300–301, 2018.
2. Wombatsecurity.com. https://www.wombatsecurity.com/blog/2018-state-of-the-phish-phishing-data-insights-and-advice. accessed on December 20, 2018.
3. T. T. (WSO2) Dayaratne, "A Framework to Prevent QR Code Based Phishing Attacks," *Cryptogr. Secur.*, Jan. 2016.
4. D. Abraham, "Approximate String Matching Algorithm for Phishing Detection," 2014.
5. J. W. Brad Wardman, Dan Clemens, "Phorecasting Phishing Attacks: A New Approach for Predicting the Appearance of Phishing Websites," *Int. J. Cyber-Security Digit. Forensics 5(3) 142-154*, vol. 5, no. 3, pp. 142–154, 2016.
6. A.A. Ahmed and N. A. Abdullah, "Real time detection of phishing websites," *7th IEEE Annu. Inf. Technol. Electron. Mob. Commun. Conf. IEEE IEMCON 2016*, 2016.

7. A. Y. Alnajjar, M. Anbar, S. Manickam, O. E. El-Taj, and Homam, "QRphish: An Automated QR Code Phishing Detection Approach," *J. Eng. Appl. Sci.*, vol. 11, no. 3, pp. 553–560, 2016.
8. J. D. Ndibwile, Y. Kadobayashi, and D. Fall, "UnPhishMe: Phishing Attack Detection by Deceptive Login Simulation through an Android Mobile App," *Proc. - 12th Asia Jt. Conf. Inf. Secur. AsiaJCIS 2017*, pp. 38–47, 2017.
9. C. E. Shyni and G. S. E. Ebby, "Phishing Detection in Websites using Parse Tree Validation," *2018 Recent Adv. Eng. Technol. Comput. Sci.*, pp. 1–4, 2018.
10. R. M. A. Mohammad, "Phishing Websites Features," no. March, 2015.
11. R. M. Mohammad, F. Thabtah, and L. Mccluskey, "Intelligent rule-based phishing websites classification," *IET Inf. Secur.*, no. January, 2013.
12. WHOIS. https://whois.icann.org/en
13. Alexa Web Traffic. https://www.alexa.com/siteinfo
14. T. Nagunwa, "Complementing Blacklists : An Enhanced Technique to Learn Detection of Zero-Hour Phishing URLs," *Int. J. Cyber-Security Digit. Forensics 4(4) 508-520*, vol. 4, no. 4, pp. 508–520, 2015.
15. R. M. Mohammad, F. Thabtah, and L. McCluskey, "Tutorial and critical analysis of phishing websites methods," *Comput. Sci. Rev.*, vol. 17, pp. 1–24, 2015.
16. Phistank. https://www.phishtank.com/. Accessed 2019
17. E. Prasetyo, *Data Mining: Konsep dan Aplikasi menggunakan Matlab*, 1 ed. Yogyakarta: Andi Offset, 2012.
18. A. H. Al Kabir, "Sentiment Analysis of Critical Data and Suggestions Training Information Technology Applications (TITA) Using Algorithms Support Vector Machine,," 2017.