

International Journal of
Engineering Research and Science & Technology



ISSN:2319-5991

www.ijerst.org

E-mail: editor@ijerst.org or ijerst.editor@gmail.com

SECURE QR CODE SCANNER TO DETECT MALICIOUS URL USING MACHINE LEARNING

1 Mrs. J. S. RADHIKA, 2 K. VARSHA, 3 CH. VRUTHVIK, 4 L. GAYATHRI

5 M. D. ABDUL MANNAN, 6 M. D. ABRAR AHMED

1Assistant Professor; Department of IT, Sri Indu College of Engineering and Technology-Hyderabad

23456Under Graduate, Department of CSIT, Sri Indu College of Engineering and Technology-Hyderabad

ABSTRACT

Q-R codes are utilised for a variety of purposes, including accessing online web-pages and making a settlement. The Internet facilitates a wide range of illegal acts, including unsolicited e-marketing, financial embezzlement, and malicious distribution. Even though all the users identify the presence of Q-R codes visually, the information stored in those codes can only be accessed through an allocated Q-R code decoder. Q-R codes have also been shown to be used as an effective attack vector, For example techniques include social engineering, phishing, pharming, etc. Harmful codes are distributed under false pretences in congested areas, or malicious Q-R codes are pasted over current ones on billboards. Finally, consumers rely on decoder operating system to determine a random Q-R code is whether malicious or benign.

For the purpose of this report, we consider the identification of malicious Q-R codes as a two-way classification problem in this research, and we test the effectiveness of many well-known M-L algorithms, including namely K-Nearest Neighbour, Random Forest, Binary LSTM and Support Vector Machine. This implies that the proposed method might be deemed an optimal and user- friendly QR code security solution. We created a prototype to test our recommendations and found it to

be secure and usable in protecting users from harmful QR Codes.

Keywords: Cloud storage, Data deletion, Data transfer, Counting Bloom filter, Public verifiability

I. INTRODUCTION

Quick Response codes have found their way into our daily lives as a result of the Internet's undeniable omnipresence. QR codes are data-encoding two-dimensional matrix barcodes. They were originally designed to track automobile parts, but they have since been adapted for a wide range of applications. The most typical application is to encode a link or other textual information so that it can be accessed quickly without requiring the user to write the URL manually. Covid-19 has impacted the expansion of internet business like e-commerce and e-banking, primarily because QR-code payment streamlines and secures the payment process's correct execution, saves time, and eliminates data entry problems. Unfortunately, technical breakthroughs are sometimes accompanied by state-of-the-art methods for exploiting users.

Phishing website that heist any type of personal data, a spammer can use are prominent instances of such attacks. QR Codes have been widely exploited and abused to take advantage of consumers' vulnerabilities. According to data on the rise in the number of dangerous QR Code

distributions over time, indicates an clear necessity to research, implement procedures or ways to prevent and detect them.

In a survey Katharina Krombholz et. al. provides a detailed summary of current studies on QR code usability and security in a survey. They determined the most significant use case as well as the corresponding attack routes. According to the paper, in addition to their many benefits, Social engineers have used QR codes as an attack vector. Malicious links are encoded by attackers in QR code, for example, redirect to run fraudulent code or phishing sites. Social engineering is the most commonly reported assault scenario. Social engineering is a strategy used in information technology (IT) security to manipulate people into revealing secret information to the social engineer.

Katharina Krombholz et al. recognise the importance of QR code security and give a thorough assessment of quick response code reader vulnerability, with an emphasis on practical security techniques. The initial portion of this research concentrated on the decoder software, while another phase is focused upon a user. Their findings reveal and underline the need for security enhancements to ensure a secure user experience when scanning QR codes. Based on an examination of privacy and security problems in mobile software, they suggested a design set of guidelines for an application that takes security, privacy, and usability into account. They demonstrated that when security, privacy, and usability are all taken into account, mobile-phone software can efficiently safe guard consumers from harmful quick response code.

Heider and Flaminia [3] have done a study that gives a thorough review of QR code scanning applications in terms of security, usability, and privacy. They identified the shortcomings and found that: Many of these QR code scanners does not meet the consumer privacy and security needs.

Several papers in the literature address this issue from the perspective of Machine Learning. In other words, they construct a list of URLs categorised as harmful or safe along with characterise each uniform resource locator using fixed group of criteria. Classification algorithms are then supposed to learn the border between the decision classes.

The extensive literature on classification models provides a variety of alternative solutions and approaches to classification challenges. Among the most common classifiers are K-nearest neighbors, Random Forest, Support Vector Machine, and Bidirectional LSTM, each with its own set of pros and disadvantages. Scammers use a variety of cyber techniques, such as luring consumers to press on phished link, that can causes the system to be compromised.

QR codes make it simple to accomplish this. To assist in identifying malicious websites, the online security community has launched blacklisting services. While blacklisting a URL has been shown results effectively in some situations, an scammer can readily trick the system by modifying more or one components of the link. Many harmful websites are never banned, either they were never or incorrectly assessed or they are too new.

II. LITERATURE SURVEY

QR (Quick Response) codes are two-dimensional barcodes with the ability to

encode different types of information. Because of their high information density and robustness, QR codes have gained popularity in various fields of application. Even though they offer a broad range of advantages, QR codes pose significant security risks. Attackers can encode malicious links that lead e.g. to phishing sites. Such malicious QR codes can be printed on small stickers and replace benign ones on billboard advertisements. Although many real world examples of QR code based attacks have been reported in the media, only little research has been conducted in this field and almost no attention has been paid on the interplay of security and human-computer interaction. In this work, we describe the manifold use cases of QR codes. Furthermore, we analyze the most significant attack scenarios with respect to the specific use cases. Additionally, we systemize the research that has already been conducted and identified usable security and security awareness as the main research challenges. Finally we propose design requirements with respect to the QR code itself, the reader application and usability aspects in order to support further research into making QR code processing both secure and usable.

The widespread use of smartphones is boosting the market take-up of dedicated applications and among them, barcode scanning applications. Several barcode scanners are available but show security and privacy weaknesses. In this paper, we provide a comprehensive security and privacy analysis of 100 barcode scanner applications. According to our analysis, there are some apps that provide security services including checking URLs and

adopting cryptographic solutions, and other apps that guarantee user privacy by supporting least privilege permission lists. However, there are also apps that deceive the users by providing security and privacy protections that are weaker than what is claimed. We analyzed 100 barcode scanner applications and we categorized them based on the real security features they provide, or on their popularity. From the analysis, we extracted a set of recommendations that developers should follow in order to build usable, secure and privacy-friendly barcode scanning applications. Based on them, we also implemented BarSec Droid, a proof of concept Android application for barcode scanning. We then conducted a user experience test on our app and we compared it with DroidLa, the most popular/secure QR code reader app. The results show that our app has nice features, such as ease of use, provides security trust, is effective and efficient.

The World Wide Web supports a wide range of criminal activities such as spam-advertised e-commerce, financial fraud and malware dissemination. Although the precise motivations behind these schemes may differ, the common denominator lies in the fact that unsuspecting users visit their sites.

These blacklists are in turn constructed by an array of techniques including manual reporting, honeypots, and web crawlers combined with site analysis heuristics. Inevitably, many malicious sites are not blacklisted either because they are too recent or were never or incorrectly evaluated. In this paper, we address the detection of malicious URLs as a binary classification problem and study the

performance of several well-known classifiers, namely Naïve Bayes, Support Vector Machines, Multi-Layer Perceptron, Decision Trees, Random Forest and k-Nearest Neighbors. Furthermore, we adopted a public dataset comprising 2.4 million URLs (instances) and 3.2 million features.

III. SYSTEM ANALYSIS

EXISTING SYSTEM

In general, QR codes are divided into sections that are dedicated for specific uses. Some areas of the Quick Response code are working and by error correction it cannot be recovered. Black and white modules are used to encode the data. If you use the QR code as an attack vector, you can change the QR code partially or completely. Partial changes occur when individual modules being flipped from black to white and conversely. Malicious mods are generally of two different type. Primary method modifies both black white pixels, while the other method only allows white pixel to black pixel changes. limitations in 5 point.

LIMITATIONS

- The effectiveness of machine learning algorithms in distinguishing between benign and malicious URLs heavily relies on the quality and representativeness of the training dataset. If the training data is biased or lacks diversity, the system may struggle to accurately identify new and evolving threats.
- Malicious actors constantly evolve their techniques, making it challenging for any static system to keep up with emerging threats. The system may face limitations in detecting sophisticated and novel

attack vectors that were not present in the training data, leading to potential false negatives.

- Striking a balance between sensitivity and specificity is a challenge. A system with high sensitivity might detect more threats but may also produce more false positives, causing inconvenience to users. Balancing this trade-off is crucial to maintaining user trust and preventing unnecessary rejection of legitimate QR codes.

PROPOSED SYSTEM

The proposed Secure QR Code Scanner system to detect malicious URLs using Machine Learning aims to address the inherent vulnerabilities associated with QR code usage in a dynamic and evolving digital landscape. This system leverages cutting-edge machine learning algorithms, including computer vision and natural language processing, to conduct real-time analysis of QR codes. User education features will be integrated to inform individuals about potential risks associated with QR codes, fostering a more informed and secure user experience. Continuous updates and adaptation to emerging threats will be key components of the proposed system to ensure robust and effective security measures.

ADVANTAGES

- The system operates in real-time, enabling immediate analysis of QR codes as they are scanned. This rapid response time enhances the ability to detect and prevent users from interacting with potentially malicious content, reducing the window of exposure to cyber threats.
- Leveraging machine learning algorithms, the system continuously

learns and adapts to evolving threat landscapes. This adaptability allows the system to stay ahead of emerging attack vectors and improve its accuracy over time, providing robust protection against new and sophisticated threats.

- Integration of computer vision techniques allows the system to extract and comprehensively analyze visual and textual information within QR codes. This multi-faceted approach enhances the system's capability to identify malicious content, even when hidden through various obfuscation techniques.

IV. IMPLEMENTATION MODULE DESCRIPTION

User Interaction Module:

This module facilitates seamless communication between the user and the medical chatbot. It includes Natural Language Processing (NLP) algorithms to interpret and understand user inputs, enabling a user-friendly and intuitive interaction for symptom reporting and inquiry.

Admin Module:

we will find out software metrics for Django using Radon that provides cyclomatic complexity raw metrics that consists SLOC, comment lines, number blank lines, Maintainability Index and Halstead metrics, also use pylint Pylint is a source code analyzer that finds for errors in programming, assists to use a coding standard strictly.

QR Code Parsing Module:

This module is responsible for extracting data from scanned QR codes. It involves the implementation of computer vision techniques to interpret visual information

and retrieve the encoded content. The parsed data is then passed on to subsequent modules for analysis.

Machine Learning Analysis Module:

The heart of the system, this module employs machine learning algorithms to analyze the parsed QR code data. It categorizes URLs as either benign or potentially malicious based on the learned patterns from a diverse dataset. The module should be adaptive, continuously updating its models to address evolving cybersecurity threats.

Anomaly Detection Module:

Focused on identifying irregularities or unexpected patterns in the QR code content, this module enhances the system's capability to detect sophisticated attack vectors. By flagging anomalies, the system can raise alerts and take preventive measures against potential threats that may not conform to known malicious patterns.

V. RESULTS



HOME SCREEN



ADMIN HOME PAGE



LOGISTICS PAGE

VI. CONCLUSION

To summarise, phishing is an online platform used to steal personal details from internet users, including personal information as well as details related to online transactions, such as credit card information and net banking passwords. QR Codes are being used to carry out phishing attacks. It's because QR codes are more adaptable, convenient, and visually appealing. However, a lack of focus on QR Code security leads to issues like confidential data loss and inadequate prevention against malicious assaults. The development of QR code phishing detection tools raises the security of quick response codes to a higher level and prevent attackers from abusing QR Codes. This research is essentially for users who scan QR codes into their day to day activity, such as employees or students, and this tool will ensure that their personal details is not compromised or is not leaked while scanning a quick response code. The present detection methodology has the advantage of detecting aberrant and harmful links embedded in QR codes that can be utilised as a vector attack. Before being presented, this phishing detection tool for QR code was built and tested. The protocols that have been utilised to ensure that the scanning is done precisely and that the phishing and malicious page is detected, that are Telnet, NNTP, HTTPs, HTTP, RTSP, SFTP, FTP and Gopher. Support Vector Machines, Random Forest,

Binary LSTM, and k-Nearest Neighbor algorithms were used. The project was also built to check domain, subdomain, IP addresses, prefix and if any symbol is present is also scanned before producing results. This solution has aided in the detection and security of the application, allowing for safer scanning. This programme will allow users to read QR codes more quickly and safeguard them from fraudulent URLs. QR Codes can be used by an attacker to commit fraud. Whenever a user scan the quick response code, they may unintentionally be directed towards a bogus website. With the development of this phishing detection tool, users will be able to perform secure scans and improve the security of their QR codes

REFERENCES

- [1] Katharina Krombholz, Peter Fr"uhwirt, Peter Kieseberg et. al. "QR Code Security: A Survey of Attacks and Challenges for Usable Security", In International Conference on Human Aspects of Information Security, Privacy, and Trust.
- [2] Katharina Krombholz, Peter Fr"uhwirt, Thomas Rieder et.al "QR Code Security - How Secure and Usable Apps Can Protect Users Against Malicious QR Code", In 10th International Conference on Availability, Reliability and Security
- [3] Heider A. M. Wahsheh and Flaminia L. Luccio, "Evaluating Security, Privacy and Usability Features of QR Code Readers", In ICISSP 2019 - 5th International Conference on Information Systems Security and Privacy.
- [4] Frank Vanhoenshoven, Gonzalo N'apoles et. al. "Detecting Malicious URLs using Machine Learning

Techniques”, In 978-1- 5090-4240-1/16/©2016 IEEE

[5] Cho Do Xuan¹ , Hoa Dinh Nguyen et. al. “Malicious URL Detection based on Machine Learning”, In (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 1, 2020.

[6] Shantanu et. al. “Malicious URL Detection: A Comparative Study”, In Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS- 2021) IEEE Xplore Part Number: CFP21OAB-ART; ISBN: 978-1-7281-9537-7

[7] Andrey Averin, Natalya Zyulyarkina, “Malicious Qr-Code Threats and Vulnerability of Blockchain”, In 978-1-7281-8075-5/20/\$31.00 ©2020 IEEE.

[8] Young-Sae Kim et. al. “Design of an efficient image protection method based on QR code”, In 978-1-7281-6758-9/20/©2020 IEEE

[9] Safwati Ismail, Alvin Ebenazer Kumar et. al. “Quick Response Code Validation and Phishing Detection Tool”, In 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE) | 978-1-6654-0338-2/21/©2021 IEEE.

[10] Makoto Takita, Hiroya Okuma, et. al. “A Construction of Fake QR Codes Based on Error- Correcting Codes”, In 2018 Sixth International Symposium on Computing and Networking (CANDAR) IEEE.