JCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-2, Issue-11 E-ISSN: 2347-2693

QR Code and its Security Issues

Dhwanish Shah, Yash Shah

[1*,2] *Dept. of Computer Science and Engineering, Nirma University ,India*
www.ijcaonline.org

Received: 01 Nov 2014 Revised: 12 Nov 2014 Accepted: 20 Nov 2014 Published: 30 Nov 2014

Abstract— With the wireless media becoming the most accessible commodity of the time to the people, barcodes are used as the message carriers with significant information transferred through at both ends. This paper aims to provide details about the most commonly 2D barcode, QR code in respect to its structure, symbology. Also it gives overview of different schemes that the attacker can use to deceive the people by directing them to malicious websites almost similar in domain name to the source.

Keywords—QR code, Error Correction, Reed Solomon, Malicious Code, Alteration, Phishing

I. Introduction

Quick Response (QR) is a two-dimensional barcode symbology developed first by the Toyota company's daughter company, Denso Wave[6] in 1994. It is a Matrix code capable of overpowering the 1D barcodes. Developed initially for the purpose of tracking automobile parts is now used worldwide commercially for promotional purposes, faster product processing and inventory management. QR codes provide an edge over the 1D symbology for their difficult interpretation by a naked eye.

QR codes can store on an average large amount of data on a small size than its predecessor largely for its matrix structure.1D barcodes can hold a maximum of [1] 20 alphanumeric digits, while QR codes on contrary hold around 7089 numeric characters and around 4296 alphanumeric characters.

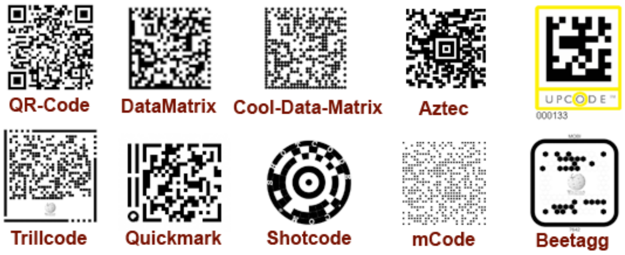

Figure 1. Different 2D codes[1]

Also in support to that feature is its ability to be scanned from any direction(360°) because of a unique structural property of the symbol, unlike the 1D ones which are unidirectional.

With the ease of use, they provide wide variety of applications. The user when scans the symbol using a scanner it connects the web link using the browser. Widely used to provide contact details on a business card, link to the websites, restaurant menus, ticket processing; with usage even including the provision for WiFi access keys.

In the past few years government agencies and big corporate giants also have started using them for speedy processing.[2]National Bureau of Investigation in Philippines, India's most recent Aadhar card includes QR code implementation. Even big airline giants in India have started using it for ticket processing.

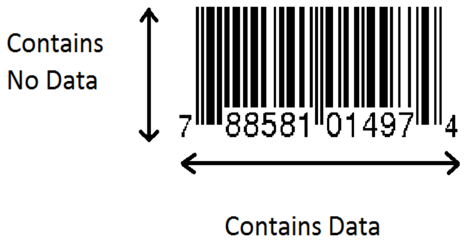

Figure 2.Unidirectional Code[3]

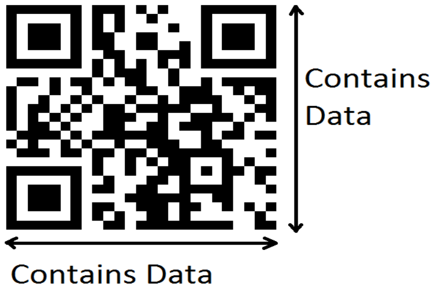

Figure 3. QR code with dual side data[3]

Corresponding Author: *Dhwanish Shah*
Dept. of Computer Science and Engineering, Nirma University ,India

II. DEVELOPMENT

With each barcode having its own functionality and its effective properties, some of them can be integrated to provide the individual characteristics of each part in one single symbol. QR code is one such example with integrated features of 3 different codes in one .[4]The high capacity of PDF 417 clubbed with speedy reading of the MAXI code and smaller size of Data matrix to increase the strength of Quick Response codes.

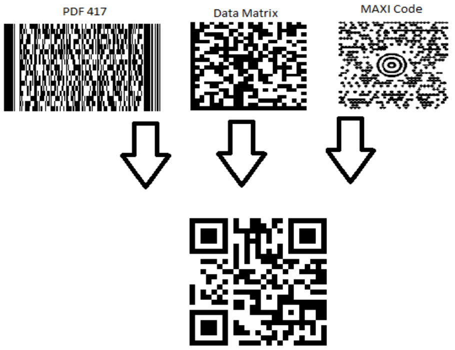

Figure 4. Integration of different codes

III. STRUCTURE OF QR CODES

Structurally the QR code has been divided into various segments depending upon the purpose served. As it is an encrypted and compressed code its requires significant amount of error correction bits to support information authenticity. Also keeping in mind the feature of any angle capturing ,it needs to have a structure to identify the data and error control part. For each symbol the smallest unit is called as module.

Comes in various versions right from Version 1 with 21x21 modules to largest one with 177x177 modules supporting higher data inclusion.

The following are the various sections of the QR symbol as shown in figure 5.

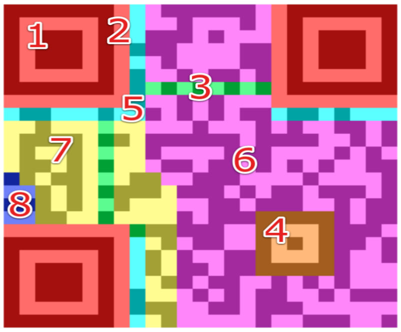

Figure 5. Structure with sectional divisions[3]

There is a white region surrounding the code to include quiet zone for improved code recognition by the decoder.[3]

1. Finder Pattern - The decoder detects the orientation of the symbol from these 3 concentric squares. Consists of 3 symmetrical structures at 3 ends with one missing at bottom right. Depending on the size of matrix has concentric square rings of alternate black and white pixels. Example for NxN matrix of finder structure ,it has a outer matrix of NxN(black),inner of N-2xN-2(white) and innermost N-4xN-4(black)
2. Separators - For improved recognition of finder patterns ,white separators are used.
3. Timing Pattern - To indicate module width(smallest element of symbol) alternating white and black pixels are used.
4. Alternate patterns - To compensate for finder patterns due to image noise. Generally available with version 7 and above.
5. Format Details - Details regarding the error correction level and used masking pattern is included.
6.Actual Data - stored as 8-bit codewords after converting to a data stream.
7. Error Correction - includes 8-bit codewords for error control.
8.Remainder Bits - contains extra bits if information and error control bits can't be sectioned as 8 bit words without a remainder.

IV. GENERATION OF QR CODES

1. Initially the data source is converted to data stream.
2. Data information length details added to the header part.
3. 4-bit blocks of data are encoded and corresponding error correction bit is generated using the Reed Solomon(RS) approach.
4. Error correcting codewords are appended at the end of the information codewords.
5. Corresponding symbols for each codewords are generated.
6. All symbols are joined to form the final QR code.

Also RS approach for Data encoding and BCH approach for Format and Version details encoding are used. There are 4 possible character modes possible: Numerals(10 bit coding per 3 number digits),Alphanumeric(11 bit coding per 2 characters), Kanji(13 bit coding per 2 characters) and 8-bit words.

Error correction has various levels in the code which as per the increased level provides higher and improved correction. This provides a correct decoded sequence from the dirty symbol or transmission error.

The 4 variants available are:

- L - 7%
- M - 15%
- Q - 25%
- H - 30%

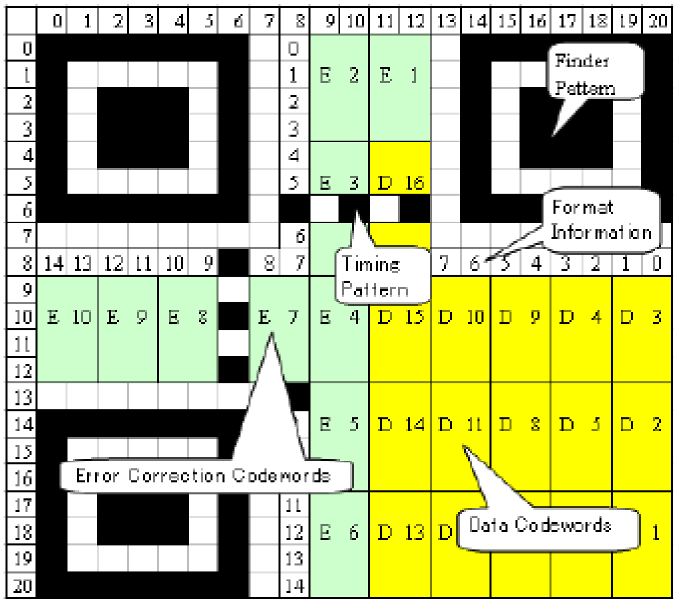
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Finder Pattern
Format Information
Timing Pattern
Error Correction Codewords
Data Codewords
E 1 E 2 E 3 E 4 E 5 E 6 E 7 E 8 E 9 E 10
D 1 D 2 D 3 D 4 D 8 D 9 D 10 D 11 D 13 D 14 D 15 D 16

Figure 6[5]. pixel wise division of QR symbol with data and error correction codewords

V. SECURITY ISSUES WITH QR CODES

Concerning factor irrespective of the widespread use of the QR code in the world of multimedia is how there can be a malicious use of the stickers in market. Appearing as very appealing while scanning in the mobile scanners but coming along with them are the attack vectors from the QR code itself which direct the legitimate intended page or media to a malicious content online to an extent of having a very identical name just differing by one or two characters. Example www.nirmauni.ac.in will be directed to a very similar domain www.nimrauni.ac.in. and the victim will suffer an advertising attack bombarded one after the other. Also such phishing attacks are a common trait experienced if a person randomly scans a QR code in public places like cafes, markets, malls, toilets, etc. Hence, the people need to be careful while scanning from wild to avoid the worst.

The victims are not having any idea about the dirty bits set in the QR code because it is not possible to identify the error by naked eye. Masking of certain bits are very much difficult to identify. Kaspersky Lab detected first of its type of dirty QR code in 2011. Also they found a loophole by finding some websites which possess Trojans which can send text messages to short rate numbers[2]. Also the dirty codes are capable of using the QR codes as attack vectors to cause harm to the back-end servers and systems by a simple SQL injection query. Besides, many online websites host net payment and banking. With an altered QR code it is possible for the attacker to direct the payments or donations made in name of some NGO to his/her account. This is the extent to which the phishing attacks can cause harm to the common man.

Typosquatting is an approach many attackers use to direct to malicious websites. This approach is basically the registering of the similarly named popular websites with intentional misspell of some characters. In 2010 almost 95000 such websites have been found to replace around 3200 popular .com websites[4].

VI. DIFFERENT APPROACHES TO ATTACKING QR CODE

In this section we discuss the various approaches which can be considered by the attacker based on the level of QR code where he intends to alter the code or mask the code. *Ioannis Kapsalis* in [4] has widely discussed the

- Altering modules by Brute -force approach
- Altering at binary level
- Altering at codeword level

for altering the QR code.

A. Altering modules by Brute- Force Approach

In this level of alteration, the most conventional way is considered by swapping some modules in either of the way. One by changing only one type of modules, example white to black or black to white; other in which both color modules are swapped. It requires very simple tools like a black marker to swap values or a tape to hide black modules.

While attacking the QR code at those module level, the attacker makes sure that he doesn't make module alterations to the finder, alignment and timing pattern else it will make the QR Code recognition difficult. So the attacker considers the target area from either data section or data and error correction as a whole. Knowing that the original source is capable of correction on its own with error correction capacity based on the level of correction (L,M,Q,H) ,the attacker has to alter the modules in such a way that they aren't less than the correction capabilities or else the original code will be generated. It is required to first scan the QR code and know about its error correction level and size to generate a 2D Boolean array to replace the original data array. But studies in [4] suggest that the results obtained aren't significant to generate a malicious code either because of the less module change than the error correction capacity (7 %,15 %,25 %,30 %) or its unreadable as its error is way beyond the limit to generate data capable similar to original source websites. Hence, the brute force approach doesn't make sense in usage due to its insufficiency.

B. *Altering at binary level*

Keeping in mind that encoding of the QR code data is done in Reed Solomon encoding with a binary level representation, the attackers can exploit this level to the extent that they can alter the binary values in small amounts that are not significant enough to be detected unless checked properly. Source URL is encoded with new values in such a way that it directs to a malicious one. As seen in above method QR code is converted to a Boolean 2D matrix with true as black and white as false by the decoder. Using that matrix the module values can be changed to generate a code for the altered URL. By separating the information part from the 2D matrix we can access the original source URL in binary form. As the change of bits will be random it won't affect the procedure as the data bit order is not direct. By considering smaller Hamming distances between the source and new codes ,improved results can be obtained in such a way that both codes are only differing by minimal characters[4].

On comparing the two codes from the decoder we can identify the modules which need to be swapped for the small number of binary values which vary from each other.

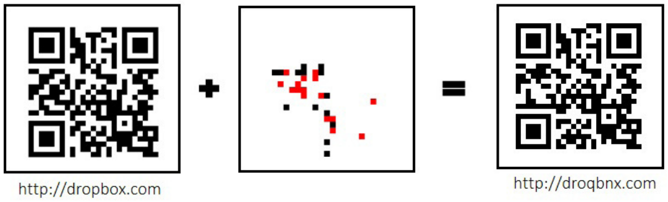

Figure 7[4]. Example showing the bit value change in modules

On comparing the values as mentioned above, we can count the number of values of black and white pixels which will be changed. In second block the red pixels indicate the number of black ones which change to white and black pixels indicate which change from white to black. The difference of that number gives the actual number of characters which will be changed in the URL.

In this way other online transaction sites can are diverted to the malicious ones with very little change in source URL characters.

C. *Altering at codeword level*

In this approach we attack the codeword part of the binary stream of data. The original URL data is encoded with the Reed Solomon code and corresponding error correction codewords are generated. On generating a stream of bits , the final stream in higher versions of QR code (2 and higher) are divided into codewords.

With changes in required data codeword , the corresponding error correction codeword is regenerated and appended accordingly. But it is a difficult task to identify the required codeword as the message is broken in 8-bit words such that the bits of a required codeword are spread in two or more different codewords. On deciding the codeword to be changed with appropriate study, we can separate that codeword from the rest and calculate its error correction codeword. By this separation we can easily know which modules values are expected to change.

This process can be improved by automatically generation codeword that have a small hamming distances instead of a manual change. These datawords are then used to generate the corresponding error codewords with minimum pixel / module changes making this approach above the other two discussed. Also it is the only approach which gives reduced pixels.

VII. CONCLUSION

After analyzing the various approaches mentioned it can be inferred that the codeword altering approach provides better results. There is a significant reduction in the number of modules changed compared to the conventional methods. Also, this paper gives the people an overview that how easily can the QR codes be changed making them suffer a loss based on their usage and the level and sophistication of alteration.

ACKNOWLEDGMENT

We would like to thank for the invaluable support we have received directly and indirectly during paper preparation. We express our gratitude to our Guide Prof. Purvi Kansara for her valuable guidance.

REFERENCES

[1]. Damon Gura, Kevin O'Shea, Arjuna Reddy ,Micheline (Mich) Sabatté, "QR Codes",Kellogg School of Management, Friday ,March 11, 2011

[2]. A. Sankara Narayanan, "QR Codes and Security Solutions", *International Journal of Computer Science and Telecommunications*,Volume 3, Issue 7, July 2012

[3]. Peter Kieseberg, Manuel Leithner, Martin Mulazzani, Lindsay Munroe, Sebastian Schrittwieser, Mayank Sinha, Edgar Weippl, "QR Code Security", SBA Research, Favoritenstrasse 16 AT-1040,Vienna

[4]. Ioannis Kapsalis, "Security of QR Codes", Norwegian University of Science andTechnology, June 2013

[5]. Sona Kaushik, "Strength of Quick Response Barcodes and Design of Secure Data Sharing System ", *(IJACSA) International Journal of Advanced Computer Science and Applications,* Volume 2, Issue 11, 2011

[6]. Denso Wave. To two-dimensional code from the barcode, http://www.qrcode.com/aboutqr.html, November 5,2014