Journal on Big Data
DOI:10.32604/jbd.2021.019236
Article

Tech Science Press

A QR Data Hiding Method Based on Redundant Region and BCH

Ying Zhou[*] and Weiwei Luo

CETC Cloud (Beijing) Technology Co., Ltd., Beijing, 100041, China
[*]Corresponding Author: Ying Zhou. Email: zy9895axy@sina.com

Received: 06 April 2021; Accepted: 28 June 2021

Abstract: In recent years, QR code has been widely used in the Internet and mobile devices. It is based on open standards and easy to generate a code, which lead to that anyone can generate their own QR code. Because the QR code does not have the ability of information hiding, any device can access the content in QR code. Thus, hiding the secret data in QR code becomes a hot topic. Previously, the information hiding methods based on QR code all use the way of information hiding based on image, mostly using digital watermarking technology, and not using the coding rules of QR code to hide information. Therefore, we propose a distributed information sharing method based on QR code information hiding. On the one hand, we can start from the coding rules of QR code and use the information hiding method to maximize the capacity of secret information embedded in QR code. On the other hand, the distributed information sharing method is used to decentralize. At the same time, experimental results and analysis show that the proposed scheme can prevent the attack of camouflage attacker, and finally recover hidden information by secret reconstruction.

Keywords: QR code; image encryption; information hiding; secret sharing

1 Introduction

Two-dimensional codes are formed by expanding the data dimension in the vertical direction on the basis of one-dimensional codes. There are many types of QR codes, the most common of which is QR Code (hereinafter referred to as QR code) which is widely used now. Compared with other types of barcodes, QR codes can store more information, and many convenient operations can be achieved through the jump or response of different protocols by the scanner, such as quick access to URL addresses, quick access to business card information, and speed dialing, quickly connect to WALN, etc. Since QR code encoding is an open standard, almost everyone can easily generate their own QR code with the help of tools. This has led to the rampant forgery of QR codes, which poses greater security risks, and due to QR code books Health does not have the function of hiding information, and there is no design protection mechanism, so any device can read the same information, including private information and confidential information, so it is not safe. This article will introduce a QR code-based information hiding method, which embeds the cipher text into the QR code without changing the content of the ordinary user's recognition of the QR code, so that only certain people can read specific information. QR code. In some scenarios that require the interaction of information from all parties to form a result (such as electronic coupons, authorization in e-commerce, etc.), if the QR code is used as the carrier of information interaction and the security in the process can be guaranteed, it will make the system easier to use.

2 Related Works

2.1 QR Coding

Two-dimensional bar code has been widely used on the Internet and mobile devices in recent years. Quick Response Code is a widely used two-dimensional matrix notation. It was developed by Denso

Wave in Japan in 1994 [1–2]. It uses a certain geometric figure on a plane according to a certain rule. Distributed black and white graphics used to record data symbol information. It is mainly composed of 4 major functional modules: detection graph, positioning graph, data area and correction graph.

The QR code encoding process is roughly divided into data analysis, data encoding, error correction encoding, organizing data, applying data mask, filling format and version information. The QR code standard provides 40 QR code versions. The larger the version, the more data can be filled. Another feature of QR code is its reliability. QR code can be set with error correction level, divided into L (Low), M (Medium), Q (Quartile) and H (High), with 7% and 15%, respectively 5%, 30% error correction rate.

2.2 Data Hiding Based QR

After the popularity of QR codes, the security of QR codes has always been a hot topic of research. Nowadays, most of the application schemes use encrypted information in the content of QR codes. Determine whether it is correct and perform subsequent operations. In recent years, under domestic and foreign research, some new research programs have also emerged.

The specific solution is to embed the QR code image as a watermark into the carrier after the QR code is generated [3–6]. When reading the QR code, first extract the QR code from the carrier. The extracted QR code will inevitably have a certain distortion in content, but the QR code itself has a certain error correction mechanism, so that the extracted watermark can also be read normally by the identification device. This kind of scheme has certain security and robustness, but because the generation and recognition are too cumbersome, it may bring some inconvenience in life.

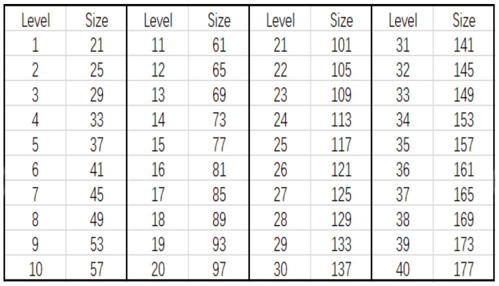
The specific solution is to add the secret information as a watermark to the QR code image after using the normal content to generate the QR code image, and directly read the watermark information in the QR code image when extracting [7–10]. But this scheme has relatively big limitations. Because the external noise interference will affect the extraction effect of the watermark, this leads to certain difficulties in the extraction of the watermark. Especially under the current conditions of use, nearly half of the application scenarios are to share QR codes through social platforms or scan and print QR codes. Sharing pictures on social platforms will compress the pictures to reduce traffic. The scanned and printed QR code will directly determine the recognition effect by the pixels of the scanning device. In these two scenarios, a large amount of noise will be artificially introduced, which will have a huge impact on practical applications. In China, Li et al. proposed a digital watermarking algorithm that can be used on QR codes and optimized the recognition process [11]. First, blur and add noise, and pay attention to comparison during embedding. The relationship between DCT coefficients. In order to resist the distortion problem caused by scanning and printing QR code, they used multiple times to embed the watermark information into the intermediate frequency component of the QR code image DCT coefficient. Later, in the process of extracting the watermark, they used the principle of maximum membership under two pattern recognition, and no longer referred to the original image, which can greatly improve the extraction effect of the watermark when scanning the printed QR code.

In recent years, there have also been schemes to hide information in QR codes using “deformation” technology [12–13]. The specific method is: appropriately increase or decrease the width and height of the “lattice”, and the specification of increase and decrease corresponds to the sequence of hidden information 0 and 1, for example, increase means 1, decrease means 0, or change means 1, and unchanged means 0. So as to realize the embedding of hidden information. At the time of reading, the inherent error correction properties of the QR code can read the content of the QR code, and at the same time record the “lattice” change sequence, which corresponds to a binary stream, which is restored to hidden information. However, this type of algorithm has an obvious flaw. The ratio of “lattice” increase or decrease is limited, that is, the range of change can only be within the general size or even less, otherwise it will affect the recognition of the QR code itself. However, if the ratio of changing the size of the “grid point” is too small, the hidden sequence cannot be extracted normally, and the sequence itself is wrong and the hidden information cannot be restored.

3 Our Proposed Method

3.1 Design Concept

There are 40 versions of QR code. It uses the binary format of numbers 1–40 to identify the version number in the version information area. The size of the QR code generated by different version numbers is different. The generated QR code side the formula for calculating the number of long grid points N is N = version * 4 + 17, and the specific size is shown in Fig. 1.



Level	Size	Level	Size	Level	Size	Level	Size
1	21	11	61	21	101	31	141
2	25	12	65	22	105	32	145
3	29	13	69	23	109	33	149
4	33	14	73	24	113	34	153
5	37	15	77	25	117	35	157
6	41	16	81	26	121	36	161
7	45	17	85	27	125	37	165
8	49	18	89	28	129	38	169
9	53	19	93	29	133	39	173
10	57	20	97	30	137	40	177

Figure 1: The size of different QR versions

Each version also has 4 error correction levels, namely L, M, Q, H, which can correct about 7%, 15%, 25%, and 30% of errors in the data area respectively. As the level of error correction increases, the requirements for error correction increase, and the proportion of error correction codes in the entire data area will increase, which leads to a reduction in the codeword space of the content.

3.2 Embedding Based on QR Code

In the experiment, it was found that the 2-M version of QR code with different content interfered with 8 data blocks (64 bits) which is the maximum value that the existing built-in scanner can accept. In the experiment, the contents of 8 consecutive data codeword blocks (D20~D27 are used here) are reversed (that is, all errors), and the data contents can still be read normally. This part of the contents has a total of 64 bits. And no one more error can continue to be recognized. It can be seen that the maximum error correction capacity is 64 bits. In addition, in the case of random rewriting, the data cannot be read after only modifying more than 30 grid points, indicating that there are certain requirements for error correction, and it is not possible to arbitrarily specify 15% for error correction. Specifically, more in-depth research on the error correction mechanism is required.

First, we can calculate the size of the data area of the QR code with the version number of 2 and the error correction level of M. After calculation, there are 28 data code blocks and 16 error correction code blocks, a total of 224-bit data code words and 128 bits error correction codeword. The error correction codeword can theoretically correct the same number of refusal errors, or half of the number of rewrite errors, the specific formula is $e + 2t \le dp$, e represents the number of refusal error blocks, t represents the number of rewrite error blocks, and d represents error The number of wrong codeword blocks, p represents the number of error detection codeword blocks (in the 1L version, p is 3; in the 1M and 2L versions, p is 2; in the 1H, 1Q, and 3L versions, p is 1; all others are 0). For example, here, up to 64 bits rewriting errors can be corrected, which is consistent with the experimental data, that is, the maximum error correction rate is about 18%. However, the error correction mechanism of the QR code is to convert each codeword block into a number between 0 and 255, and then use the BCH algorithm for error correction for D0~D27 and E0~E15, rather than for individual bits. Perform error correction. In addition, there is a certain "division of labor" in the error correction code under certain versions, that is, a part of the error correction code is responsible for part of the data code word error correction. In this case, the error correction code word and the data code word need to be grouped, and the same group The number of error blocks within cannot exceed the maximum value that can be corrected for this group, and the maximum number of error corrections for each group is calculated using the formula $e + 2t \le dp$.

After using this property, the maximum concealable capacity of each version can be calculated quantitatively and there is no need to worry about whether it will be unreadable.

In this example, the error can only appear in 8 codeword blocks, otherwise the code cannot be scanned. In actual experiments, there may be cases where the modification in 9 codeword blocks can still be read. For example, in my experiment, I also tried to select 9 codeword blocks to modify one codeword, but it was still able to read. The reason for this situation is not an error in the analysis here, but it may come from the error when the barcode scanner reads the matrix grid points. You can directly set the value of the matrix through the code and then read it, and it can display normally. Tips for error correction. In fact, after several subsequent tests, the data area of each position of the two-dimensional code was reversed. When the change appears in 8 or less codeword blocks, the two-dimensional code can be read normally. However, when an error occurs in more than 9 data code words, even if the amount of error is small, error correction cannot be performed (the accidental event of normal decoding while ignoring the accuracy error of the scanning device itself). Therefore, in this example, 8 data codeword blocks can be selected for information hiding.

3.3 Embedding in QR Code Redundant Part

First, we should first analyze whether the redundant part can be used as an information hiding area. After analyzing the composition structure and information capacity of the QR code in the previous article, it can be seen that the redundant parts in the QR code are mainly: 1) The format information area has two copies of the same information; 2) The version information area has two copies of the same information. And the version number can be calculated by passing through the side length N; 3) When the data code word is too large to produce filler words, the content after the terminator belongs to the redundant part. The following experiments will analyze the feasibility of these three parts as areas of hidden information.

1) Format information area

After analysis, it is learned that the mask number in the format area determines the way to reverse the mask of the QR code. If this part of the content is missing, the scanning device will not be able to decode it. After experimental testing, after erasing the two sets of format information areas in the generated QR code (as shown in Fig. 2), only 31 grid points were erased, which is far less than the error correction level of the QR code, but it has been Unrecognized. It can be seen that the format information area is not a redundant part and is not suitable as a space for information hiding.

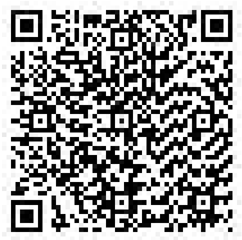


Figure 2: The erased format information area in QR code

2) Version information area

After analysis, it is known that the information in the version information area is the binary format of the version number of the QR code calculated by BCH (18,6), and only the version number is greater than 5 to have the version information area. However, the version number itself can be calculated from the size of the QR code, which is indeed redundant data. However, through experiments, it was found that the 36 grid points of the two version information areas could not be read normally after erasing the two blocks, which is shown in Fig. 3. It can be seen that although the version information area is redundant, it is indispensable.

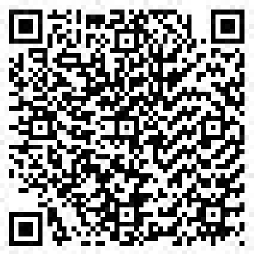


Figure 3: The erased version information area in QR code

3) Filler area in data codeword

The filler in the data code is the data area obtained by alternately filling the unused data area with 0 × EC and 0 × 11. It has no practical meaning in itself, and it can be determined that this part is a redundant part.

Next, verify whether the redundant part can be used for information hiding. The QR code version number shown in the figure below is 2, the error correction level is M, and the content is a numeric character 0. The reason for choosing this level is that the capacity of this level is sufficient for experimental use, and this version has no version information area, follow-up experiments and the plan design will also be explained based on this two-dimensional code. First of all, the QR code with the version number of 2 and the error correction level of M can store up to 63 numeric characters, which means that there is a large amount of filler space that can be used to hide information. After calculation, first the data 0 will be converted into 4-bit data, then add the character count indicator (10 bits in total) in front, then add the mode indicator 0001 in front, and finally complete it with 0 and add the terminator 0000. A total of 24 bits of data. The breakpoint test of the code also proved that starting from the 25th bit is filled in the word area (as shown in the gray part of Fig. 4).

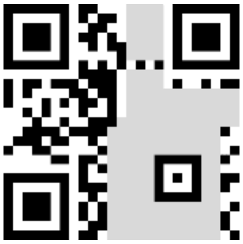


Figure 4: Filler area in data codeword of QR

In order to prove that the filler part can be modified and will not affect the reading of the content of the QR code itself. In the experiment, the filler was replaced with other meaningless characters, and a new QR code was generated. Fig. 5 shows a comparison between the two-dimensional code generated by modifying the filler to 0 × 00 and 0 × FF and the original two-dimensional code. These three can be read, indicating that the filler is indeed a redundant part and will be ignored by the scanner, and will not affect the scanning and decoding process. It can be seen that the filler area can indeed be used as an area for information hiding.

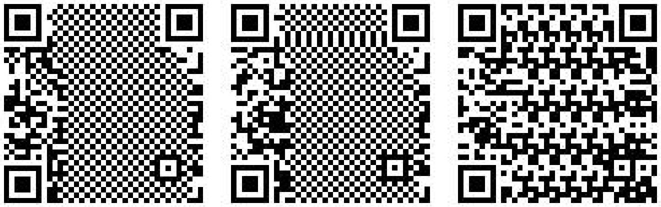


Figure 5: The filler words are 0 × 00 (left), 0 × FF (middle), 0 × EC and 0 × 11 alternately (right)

3.4 An Example of Information Hiding Method Based on QR

According to the above-mentioned two-dimensional code information hiding scheme, and the two-dimensional code is used as the carrier for the information required by each participant, and hide it in the two-dimensional code, which can only be recognized by a specific device, or in the two-dimensional code, adding parameters in the embedding process so that only specific users can get the message. The specific plan is as follows:

Embedding process

1) choose a redundant part and embedding extracted data by bit replacement.

2) Form a standard QR code, the content of the QR code can be the download address of the scanner, promotion page, etc.

Recovery process

1) The hidden information of the QR code is scanned by the participant's scanner.

2) Execute the above secret reconstruction algorithm to recover the information.

4 Conclusion

In this paper, the use of two-dimensional code for information hiding is different from the traditional use of pictures for information hiding. This article relies on the principle of two-dimensional code encoding design and error correction mechanism, and obtains hidden information through exclusive OR operation after error correction. The distributed key sharing design proposed in this paper can be checked before the secret is reconstructed, and the attacker can be excluded. With the two-dimensional code information hiding technology, it will meet many specific needs, such as multi-party authorization of e-commerce, identification of electronic coupons, etc. It is worth mentioning that the secret sharing scheme proposed in this paper can be combined with other information hiding schemes, and the specific hiding scheme can be adjusted according to actual needs.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

[1] Denso-Wave, QR Code Standardization. 2003. [Online]. Available: http://www.qrcode. com/en/index. Html.

[2] Psytec, QR Code Editor Software. 2013. [Online]. Available: http://www.psytec.co.jp/ docomo.html.

[3] Z. Wang, L. Sun and M. Li, "QR code watermark technology with strong robustness," *Packaging Engineering,* vol. 33, no. 15, pp. 84–87, 2012.

[4] H. Lee, C. Dong and T. Lin, "Digital watermarking based on JND model and QR code features," *Advances in Intelligent Systems and Applications*, vol. 2, pp. 141–148, 2013.

[5] W. Wu, Z. Lin and W. Wong, "Application of QR-code steganography using data embedding technique," *Information Technology Convergence*, pp. 597–605, 2013.

[6] A. Gaikwad and K. Singh, "Embedding QR code in color images using halftoning technique," in *Int. Conf. on Innovations in Information, Embedded and Communication Systems*, pp. 1–6, 2015.

[7] R. Xie, H. Zhao and K. Wu, "QR two-dimensional barcode security technology based on discrete wavelet transform," *Computer Engineering*, vol. 39, no. 12, pp. 126–129, 2013.

[8] B. Sun and M. Gao, "Research on digital watermarking algorithm based on QR code," *Computer and Modernization*, vol. 11, pp. 74–77, 2011.

[9] Z. Li, "Research on text-based 2D barcode digital watermarking algorithm," in *Automation and Instrumentation*, no. 4, pp. 14–17, 2014.

[10] Y. Gao and C. Xu, "Research and implementation of secure and practical QR code," *Information Network Security*, no. 10, pp. 47–50, 2012.

[11] L. Li and R. Wang, “A digital watermarking method for QR code,” *Journal of Hangzhou University of Electronic Science and Technology*, vol. 31, no. 2, pp. 46–49, 2011.

[12] J. Shen, “Research and implementation of two-dimensional code anti-counterfeiting system based on information hiding,” M. S. thesis, University of Electronic Science and Technology of China, 2019.

[13] P. Lin and Y. Chen, “High payload secret hiding technology for QR codes,” *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, 2017.