A DETAILED ANALYSIS OF QR CODE AND A TECHNIQUE TO GENERATE QR CODES FOR CRYTOGRAPHY

PAYAL SHAH[1], RADHIKA SULE[2], TANMAY SAWANT[3] & VIRAJ MANJAREKAR[4]

[1,2,3]Department of Electronics and Telecommunication Engineering, St. Francis Institute of Technology,
Mumbai University, Maharashtra, India

[4]Department of Computer Engineering, St. Francis Institute of Technology, Mumbai University, Maharashtra, India

ABSTRACT

QR codes or Quick Response codes were first created by Denso Wave Company. The QR codes are basically 2-D barcodes and have several advantages over the conventional barcode. A QR code is faster, having more capacity to store information, not direction oriented as well as resistant to distorted systems. Along with numeric and alphanumeric data, it can also encode Kanji and Kana characters These QR codes can be easily scanned by using scanners that can even be available on the Smart phones. QR codes thus not only have wide range of advantages but also can be used by a larger class of people. QR codes can be used to store large information and retrieve the information even if there is a little distortion. Thus, as compared to other data matrices, QR code is a better, newer and more efficient technique. Our paper gives step-by-step procedure that can be implemented to generate a QR codes. The paper also gives an overview of the QR codes, its applications and advantages.

KEYWORDS: Advantages, Error Correction, Generation of QR Codes

INTRODUCTION

The Supply chain management (SCM) is the management of an interconnection or interlink between the networks channels and node businesses involved in providing products and service packages required by the end customers in a supply chain. The integration era of the SCM was highlighted with the development of electronic data interchange, introduction of enterprise resource planning and also expansion of Internet-based collaborative systems. During this era, the most important aspect of supply chain management was the need for facts and figures. Information was believed to hold the key to a successful business venture, and companies lacking in keeping track of information found themselves at a distinct disadvantage. So in order to keep the information of all the products and services, a new technology called the 'Barcode' was introduced. This new 'Barcode technology' helped the vendors and the suppliers to keep a track of the products and services and all the relevant information about it. The use of barcodes ensured detailed and correct information, which ultimately helped to reduce costs. Barcodes provided accuracy and speed that helped in reducing expenses incurred to rectify errors made by manual oversight or faulty data entry. Thus supply chain management became a very easy process with the aid of barcodes. In the 1960's these barcodes were used for automatic tracking of rail road cars. Later in the 1970's they were first introduced in the grocery cartons and for automobile components tracking. Nowadays, barcodes are highly used to keep information of the products like in distribution or parts warehouses, in Pharmaceutical companies, on batch control sheets, and automatic indexing, in libraries and many other fields.

BARCODES

A **barcode** is printed horizontal strips of vertical bars which is an optically machine readable representation of the

data to which it is attached. It is used to hide a data of the product. It is widely used in grocery stores, libraries and many other places.
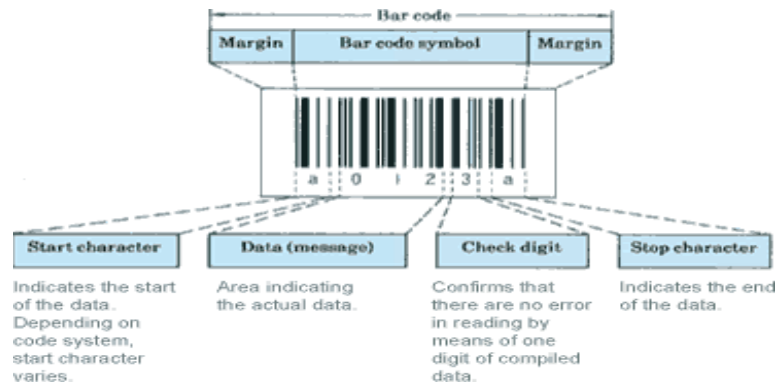

Figure 1

DISADVANTAGES OF A BARCODE

Barcode only holds information in the horizontal direction, thus it takes more space for printing to carry information. Barcodes can only be scanned in one direction thus causing interference and negative effects from the backgrounds. It carries less information.

QR CODES

QR codes are 2-D barcodes, with one distinct added benefit—they have a matrix format that allows them to store a large amount of unique data. Standard linear barcodes can only hold up to 20 alphanumeric digits, but QR codes can hold up to 7,089 numeric characters and up to 4,296 alphanumeric characters worth of data. It holds information in both horizontal as well as vertical direction thus it occupies less printing space compared to traditional barcodes. They can be scanned in all directions i.e. 360 degrees, thus eliminating any interference and negative effects from backgrounds.
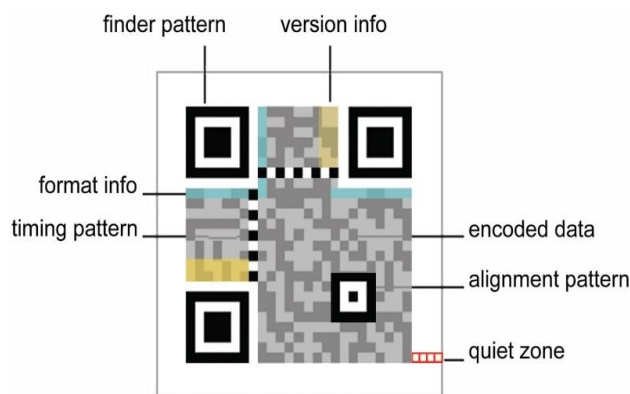

Figure 2

GENERATION OF A QR CODE

Generation of QR codes consists of various steps. These steps are dependent on the type of information to be encoded in the QR code. The steps are as follows:

Data Analysis

This data analysis step deals with various standards of QR code. This code is basically used for encoding string of text i.e. converting text into 0's and 1's. The QR code consists of four standard encoding modes which are as follows:

- Numeric mode

- Alphanumeric mode

- Byte mode

- Kanji mode

Although the main task of these modes to convert string from input text, each mode uses different methods for converting the same depending upon the situation and requirement. Finally all these modes generate optimized shortest possible bits of string which is convenient for encoding.

These modes use specific characters for encoding. The brief explanation about characters used in each mode has mentioned below.

A **numeric mode** is a simple encoding mode which is used for decimal digits 0 through 9. Also in this mode 3 characters are encoded to 10 bit length. The capacity of this mode is up to 7089 characters.

An **alphanumeric mode** is for numeric i.e. for decimal digits 0 through 9 as well as uppercase letters i.e. A through Z but not for lowercase letters. Also it is used for encoding special symbols such as +, -, *, %, $, ., /, : and space as well. For creating QR code for this mode, we have to use specific alphanumeric values of the characters during encoding process. These alphanumeric values are referred from predefined table [1]. Here unlike numeric mode, 2 characters are encoded to 11 bit length. The capacity of this mode is up to 4296 characters.

A **byte mode** is by default for characters from the ISO-8859-1 character set [2]. QR code scanner automatically detects when some code of other format is inserted i.e. UTF-8 [3]. Generally this mode uses 8 bits data having capacity up to 2953 characters.

A **kanji mode** is for double byte character from the JIS character set [4] having bit length of 13.

For UTF-8 to encode kanji characters it should use 3 or 4 bytes; but shift JIS uses only 2 bytes for encoding kanji characters that increases efficiency. Generally it has capacity up to 1817 characters. Also there are three different modes other than these described modes which use specific character set for encoding.

An **Extended Channel Interpretation (ECI)** also uses UTF-8 character set. This mode is not broadly used as many QR code readers do not understand this ECI format.

A **structured appended** mode is used for encoding data from multiple QR codes (up to 16 QR codes).

A **FNC1** helps to allow QR code to function as GS1 barcode.

It is import to choose appropriate mode so as to increase the efficiency and accuracy of the scanner. The most optimum mode is selected depending on characters used in the input string and by checking the following stated conditions:

- Use numeric mode when input string consists of only digits (0 through 9).

- Use alphanumeric mode when input string consists of digits along with symbols such as $, +, - etc and also when it has uppercase letters (A through Z). Refer table.

- Use byte mode when characters in input string are not present in above two modes but are from ISO 8859-1. Also QR readers can recognize UTF-8 in byte mode.

- Use kanji mode when all characters from input are in shift JIS character set. Shift JIS characters can be encoded in UTF-8 hence it is possible to use byte mode also for kanji mode as it is more efficient.

- Sometimes it is beneficial to multiple modes in a single QR code by including the mode indicator before each section of bytes that uses that mode. Also by examining the characters in the input text it is possible to choose most optimum mode for encoding.

Data Encoding

The main goal of any encoding mode is to design the shortest possible string of bits for the characters that are used in that mode. Depending on the situation each mode uses different method for converting input text to string of bits (0's and 1's).

There are several steps which are to be followed for data encoding.

Step 1: Choose the Error Correction Level

This is the first step in encoding where the correction level is selected. For that purpose Reed-Solomon error correction method is used. It creates byte stream of error correction code-words based on encoded data. These error correction code-words used by the QR scanner if it did not read data correctly and used for correcting those errors.

Table 1

Error Correction Level	Error Correction Capability
L	Recovers 7 % of data
M	Recovers 15 % of data
Q	Recovers 25 % of data
H	Recovers 30 % of data

Step 2: Determine the Smallest Version of Data

In this step character count from the input text is maintained, this count is nothing but the size of the data which is referred as a *version*. There are 40 *versions* in total and each version is 4 pixels larger than the previous *version*. Depending on the capacity of the characters in the individual mode as mentioned in data analysis, suitable version is selected.

Step 3: Mode Indicator

Every encoding mode is identified by the four bit mode indicator. The encoded data should start with appropriate mode indicator that specifies the mode being used for the bits that come after it. Refer to following table for each mode indicator used for each mode.

Table 2

Mode Name	Mode Indicator
Numeric Mode	0001
Alphanumeric Mode	0010
Byte Mode	0100
Kanji Mode	1000
ECI	0111

Step 4: Add the Character Count Indicator

This indicator must be placed after mode indicator which co   ntains the total count of the characters present in the input. Depending on the count, the string with the binary numbers is created.

For example consider a message to be encoded in numeric mode **1234.**

Since it is numeric mode, the value of mode indic   ator will be **0001.**

We use **H-level** error correction and version **1**.

123 4

123=0001111011 4=0100

00011110110100

Using 1-H

9*8=72 bits

Now padding 4 zero bits

00011110110100 0000

Now the above length is 18bits, now we need to add more 6 zeros to make it 24 bits which will be divisible by 8

00011110110100 0000 000000

Count of the above code is 24 bits

We are using 1-H

9*8=72 bits

We need to extend 24 bits into 72 bits by appending ALTERNATIVELY the following two 8 bit numbers. These are known as pad bytes.

1)   11101100 2) 00010001

Now the number of bits to be added = 72-24 =48

So we need to add 48/8=6 sets of pad bytes

So the output will be:

00011110110100 0000 000000 11101100 00010001 11101100 00010001 11101100 00010001

Error Correction

Error correction is done so that the QR code scanners can detect any errors present in the QR code. It is also used to correct the errors that are present. The error correction technique generally used is Reed Solomon (RS) error correction. The results that are obtained by data encoding need to be delimited based upon a standard table defined for QR codes [5]. Create a polynomial depending on the number of error correcting code words and name it g(x) [5]. Convert the delimited result obtained into its corresponding polynomial based on predefined table and name it f(x) [5]. Now by using polynomial long division, we divide f(x) by g(x). We repeat same logic until this divide calculation is over. Finally we get a remainder whose degree is one less than the degree of polynomial g(x).

Structure Final Message

When the QR codes are large it is necessary to break the data code words. Thus, the data code words are basically divided into blocks. Error correction code words for every block are generated. The blocks and the error correction code words are then interleaved together as per predefined QR code standards. As mentioned earlier this step is only for large QR codes, smaller QR codes have only one block and therefore do not require interleaving. The error correction code words simply follow the data code words.

Firstly the data code words and error correction code words are generated based on the version and error correction level of the QR code. Based on predefined standards these are distributed in blocks and groups.

To interleave the blocks the pattern followed is first data codeword from each block followed by second data codeword from each block and so on. Similarly for error correction code words the pattern is first error correction codeword from each block followed by second error correction codeword from each block and so on. The interleaved error correction code words are the appended at the end of the interleaved data code words. It is then converted to 8 bit binary format. Extra bits known as remainder bits need to be added at end of certain versions of QR codes based on the following table.

Module Placement

Module is not dependent on the pixels of the computer screen. Various function patterns are placed in th QR code which are necessary for decoding purpose. The function patterns are finder patterns, separators, alignment patterns, timing pattern and dark module.

Following steps must be followed

- The first step is to set the locations of the finder pattern. The finder pattern has an outer black square of 7x7. The next layer is inside it which is 5x5 white square pattern and a innermost 3x3 black square pattern. Using the finder pattern in the QR code the scanners can orient it properly so as to decode it. The finder patterns have predefined locations which are at top left, top right and bottom left corners of the QR code. These places remain same for all the QR codes and are not dependent on the version of the version of the QR code.

- Separators are then added at the inner edges of the finder patterns. The separators are a white square border. This are used to uniquely identify the finder pattern.

- Alignment Patterns are 5x5 module of Black Square; inside it are a 3x3 module of White Square and a black module in the centre. These are added only on QR codes of version 2 and higher. The location of the alignment pattern varies as per the version of the QR code. It must be noted that these must not overlap the finder patterns and the separators.

- Next is adding of Timing Patterns. These are a set of horizontal and a vertical line with alternate black and white modules. They are located on the sixth row and the sixth column. It is mandatory that the timing patterns start and end at a black module. A dark module is then added next to the finder pattern at the bottom left.

- The format area information is then placed along the separators. This is irrespective of the version set for the QR code. A one module strip is reserved below and right of the separator of the top left finder pattern, below the separator of top right pattern and right of the separator of the bottom left finder pattern. The version information

bits should be be added in a area of 6x3 module above the left bottom finder pattern and 3x6 module to the left of the top right finder pattern.

- Next the data bits are placed. The positioning of data bits starts from bottom right corner and continue upwards occupying adjacent two modules. After reaching the top it moves to the next left columns again consisting of two modules and continues downward. This continues till reaches left bottom corner. The data bots are placed only in the unused modules it skips the already placed function patterns and reserved areas.

The data bits are placed in a zigzag pattern over the two modules of a column. During placement the bits are placed in as follows right module, left module, and right module and so on thus basically forming a zigzag pattern. This pattern is ruled out only when the data bits encounter a vertical timing pattern, no overlap must take place on the vertical timing pattern.

Data Masking

This step is done to ease the scanning of the QR code. This basically involves reversing the color of the modules. The white modules are converted to black and vice versa. The masking is not applied to the function patterns and the reserved area. There are in all eight mask predefined mask patterns each of these is applied to the QR code. This is followed by the evaluation of each pattern. Penalties are given on basis of the rules specified below. The mask pattern with lowest penalty score is chosen. Penalties are given when a 2x2 module matrix is of same color, there are adjacent five modules of same color, existence of patterns similar to finder patterns and finally when more than half modules are black or white. Greater penalty is given for greater difference between the numbers of black and white modules.

Format and Version Information
Adding Format Information

The QR codes must be correctly decoded by the scanner. For correct decoding it is important to add the format information string and the version information string to the QR code. The format information string is indicative of the error correction level and the mask pattern that is used in the particular QR code. The format information string is a 15 bit string. It is generated as follows.

- The first two bits depend on the error correction level that is employed. The bits are 01, 00, 11 and 10 for error correction level L, M, Q and H respectively.

- After this the mask number needs to be determined for the QR code. It is the converted to binary and appended at the end of the two bits obtained earlier.

- By using Reed-Solomon Error Correction a 10 bit long sequence of error correction bits is obtained. The generator polynomial is obtained, error correction bits are generated by using long division method. For this the generator polynomial bits are divided by the format bits The result obtained must be of 10 bits or less. If it is less than 10 bits zeros must be padded at the left to make it 10 bit long.

These 10 bits are appended to the earlier obtained 5 bits and XORed with 101010000010010. The string thus obtained is placed below the top finder patterns and right to the left finder pattern.

ADDING VERSION INFORMATION

The version information string needs to be added to QR codes having version higher than 7. It is used to specify the version that is used for the QR code. It is a standard to use (18, 6) Golay code for adding the version information string.

The first bits are for version and the nest 12 bits for error coding. The generator polynomial and the generator polynomial bits are obtained. This is followed by division similar to the ones done in adding format information. The initial strings must be 18 bit long. The result obtained must be of 12 bits or less. If it is less than 12 bits zeros must be padded at the left to make it 12 bit long.

A 3x6 version block is added above the bottom left finder pattern and a 6x3 version clock is added at the left of the top right finder pattern.

Applications of QR Codes

QR code is latest emerging technology which is used in several fields, also marketers have found out creative ways to integrate these codes into their campaigns. Following are some domains where these QR codes are used:

- **Lifestyles**

Many departmental stores have started using QR codes on clothing signage that allows customer to get fashion tips and clothing information. Some departmental stores have this facility that is used for payment.

- **Leisure and Entertainment**

Some art museums use this code to direct visitors online by giving in-depth information. Few applications using QR codes are used for sending greetings and sharing playlist. Video games have this code to unlock exclusive content. Also it is used for advertising trailer or songs of movies.

- **Food and Drink**

Many food products have this QR code, by scanning this we come to know every information about the product i.e. price, manufacture date, expiry date etc. Restaurants use this code to bring transparency in their menus and help us to place order via phone only.

- **Promotions**

This code is used for promotion purpose of newly launched product in the market where customers has to scan the code and they get gift hampers of that new product. Promotion can be of any campaign such as movie, music, product etc.

- **Events**

With the help of QR codes it is easy to reach public and provide information about events such as festivals, concerts, plays, parties etc.

- **Information**

It allows us to know about business, stock market, news, transport facilities and also medical and educational facilities pertaining to that region.

CONCLUSIONS

QR codes have high capacity and high security; these attributes contribute to their wide range of applications. They also have a significant advantage over the conventional barcode. The generation of QR codes is thus important. It includes several steps which take into consideration several factors. The generation of QR codes must thus be done taking into consideration all the factors stated in the paper in step-by-step manner so as to get fully functional and correct QR code.

REFERENCES

1. http://www.thonky.com/qr-code-tutorial/

2. http://www.ic.unicamp.br/~stolfi/EXPORT/www/ISO-8859-1-Encoding.html

3. http://www.utf8-chartable.de/

4. http://www.itscj.ipsj.or.jp/ISO-IR/159.pdf

5. http://www.swetake.com/qrcode/qr1_en.html

6. "QR Code, Wikipedia", http://en.wikipedia.org/wiki/QR_code [Retrieved 2012-02-09]

7. http://www.qrcode.com/en/

8. http://www.denso-wave.com/qrcode/index-e.html