A Survey Study on QR Code and Data Matrix Symbols

Mahdi Shahbaba
School of Electrical and Computer Engineering
Ryerson University
mshahbaba@screamingpower.ca

Gary Michor
Screaming Power
147 Dalhousie Street
Toronto Ontario
gmichor@screamingpower.com
www.screamingpower.com

Soosan Beheshti
School of Electrical and Computer Engineering
Ryerson University
soosan@ee.ryerson.ca

***Abstract*—In this survey study, we compare QR code and Data Matrix barcodes from different aspects such as structure of the symbols, scanning conditions, size and data capacity of the code. We also have developed software for reading utility Flat Files, extracting desired information from customer accounts and converting those information into QR code and Data Matrix symbols. The developed software can generate codes in two ways - automatically and manually - for different code sizes along with the option of HEX encryption on the data. Considering the available research papers on QR code and Data Matrix, along with the results of our experiments, it can be concluded that Data Matrix barcode is a better candidate for the purpose of encoding large amounts of data, such as customer data, in a small area. The main advantages of Data Matrix over QR code are its smaller size, flexibility for scanning in different lighting conditions, quick scanning time and less sensitivity to low contrast symbols.**[1]

I. Introduction

Unlike conventional barcodes that store information in a horizontal manner, two dimensional (2D) codes can store information both horizontally and vertically. Therefore, the amount of information that 2D codes can store is much greater than conventional barcodes. The information encoded in the code may be text, a URL, or other data, and can be read by mobile device cameras. With the increased popularity of smartphones, the applications of 2D codes have also expanded. 2D codes are now being used in commercial tracking, entertainment and transport ticketing, product/loyalty marketing, in-store product labeling, among many other usages [12]. To be able to encode multiple languages in 2D barcodes, they should also have the capability of encoding Unicode and ASCII characters.

Data Matrix and QR code are the only open source 2D barcodes which are mainly invented for the industrial needs and then were adapted for use in mobile phone applications. In this survey, we will compare both of the Data Matrix and QR code symbols from different aspects such as data capacity, symbol size, user scanning experience, etc.

The first half of this article deals with some background information on QR code and Data Matrix, and their compared features as is discussed in the literature. The final part of the article is related to our proposed 2D barcode generator software for utility data. This generator can be used as a tool for analyzing the QR code and Data Matrix symbols on real customer data.

[1]Private and Confidential - Screaming Power 2014. No reproduction allowed.

II. Quick Response Code (QR Code)

QR codes are 2-dimensional barcodes for encoding text string data and were developed by Denso Corporation in Japan, 1994. It has been approved as an AIM standard, JIS standard and ISO standard. QR code is widely used in Japan in different industries such as marketing, food industries and Japanese airlines. There are 40 different versions of QR codes with a different number of modules. The highest version (version 40) has the most number of modules (177×177) which leads to a maximum capacity for data encoding. As a result, a version 40 QR code can encode 7089 numeric characters or 4296 alphanumeric characters [1], [3], [4]. However, designed applications for smart phones only work with versions 1-10 to deal with the limitations of phone cameras [10], where a version 10 QR code can encode up to 652 numeric and 395 alphanumeric characters.

Using Reed-Solomon error correction method, QR code has the capability of error correction for restoring data even if the code is damaged or dirty. There are four error correction levels Low (L), Medium (M), Quartile (Q) and High (H) available to users for encoding. The following table shows different levels of error correction and their related capability:

TABLE I. Four different levels of error correction in QR Codes.

QR Code Error Correction Capability	
Level L	Approx 7%
Level M	Approx 15%
Level Q	Approx 25%
Level H	Approx 30%

In Table I, by increasing the level of error checking and correction (ECC) from L to H, the chance of data restoration for damaged data will be increased along with a cost of larger code size (see Fig. 1). For example, if we have 100 codewords for encoding (each codeword in QR code is 8 bits) and we

want 50 of those codewords to be corrected, then we require 100 codewords of Reed-Solomon code (twice the amount of 50 corrected codewords). Therefore, total codewords are 200 which 50 of that can be corrected. The error correction rate for the total codewords is 25% which corresponds to QR Code error correction Level Q. Besides the error correction

Fig. 1. QR code error correction levels.

capability, size of the QR code is another advantage of this type of codes which is almost ten times smaller than the traditional barcodes.

The unique structure of QR codes makes it readable from any direction in 360°. As Fig. 2 shows, each QR code is consists of five different parts. In this figure, version information

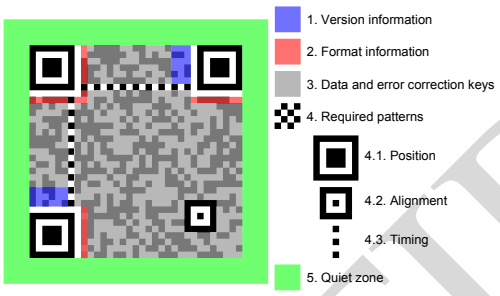

Fig. 2. Different structural parts of a QR code.

(size of the code) is a string stored in the highlighted area. Format information is also another string which gives some information to scanner about the chosen ECC level (L, M, Q, and H) and the masking system. There are eight possible masking systems which are designed to change the color of QR code pixels from dark to light (black to white) or vice versa to have better readability in the scanner. The appropriate masking system is chosen automatically by the encoding software which makes the distribution of light and dark modules equal. This change leads to a better contrast and readability.

The gray shaded area includes the data and error correction keys for reconstruction. The position and alignment marks are necessary for the scanner to find the location and alignment of the QR code. Timing pattern helps the decoder to find the location of each cell in the QR code. Finally, Quiet Zone is a required white margin around the QR matrix which should be at least 4-module wide.

Structured appending feature is another advantage of QR codes. Based on that, each QR code symbol can be divided into up to 16 symbols which allows to print the code in a narrow area. Conversely, information stored in multiple QR symbols can be reconstructed as a single QR code (see Fig. 3).

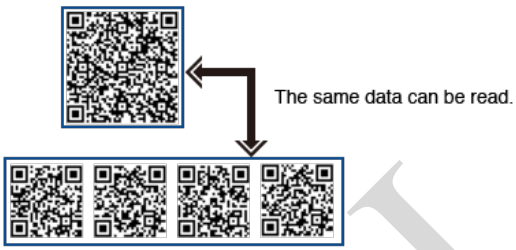

Fig. 3. Structured appending feature of QR codes.

A. QR code and security issues

There are a number of studies on QR codes and how they can be used as an attack vector [6]. As the related studies prove, a malicious QR code not only can have consequences for the victim, it also can affect the back-end system which is designed to respond to the scanning event.

SQL injection is one of the feasible attacks that can be achieved just by adding a semicolon followed by an SQL query on the request that will be processed by the system.

Command injection could be another attack based on this automated system. In this case, the encoded information is used as a command line parameter without sanitation. For example, attacker can run commands and damage the server.

In simple words, if user scans a malicious QR code created by an attacker, then the unwitting user will generate the attack on behalf of the attacker.

Malicious Pixels for using QR codes as attack vectors are discussed in [7]. According to that, an attacker just by using a marker can change the color of white and black modules and create a faked QR code. The faked QR code can be used for phishing and redirecting the user to an untrusted web page. The conducted experiment in [8] has proved the failure of this approach for phishing or at least it will be very difficult. The reason behind that is because of the error correction area. It can reconstruct the damaged data to some extent and if the error correction area is the target of the attack then the code will not be readable by the decoder.

In another experiment in [7], the author has tried to attack the binary and codewords (each codeword is 8 bits of the binary data) of the QR code symbol instead of changing the final modules of the QR code. In this approach, the binary code or the codewords are manipulated before the encoding and generating the QR code. In both cases, the error correction code is constructed based on the attacked data which guarantees the success and readability of the phishing attack. They have reported that by targeting the codewords for the attack instead of the binary code it is possible to minimize the difference between the original QR code and the malicious QR code.

The ability of decoder for showing the URL of the code before directing the user to any website will be helpful for verification of the contents of the QR code. It is also possible for the reader application to notify the user about a decoded

URL which has no valid certificate and is attempting an insecure connection.

A possible way of recognizing the certified QR codes is adding a digital signature within encoded data [2]. The digital signature can be decoded in the scanner and it will be compared with the original publisher on a verified server database. If the signature is not matching with any of the known publishers, the user will be notified by a security message.

Adding an extra layer of masking data to the original data also could make it possible to guarantee the data security. By this approach, only users with the specific decoder (mobile application) are able to read and decode the data.

The above risks and solutions can be also extended to Data Matrix barcodes, as they are very similar to QR codes in terms of physical representation and decoding procedure.

III. Data Matrix Code

Data Matrix is also a 2-dimensional symbology which can encode a large amount of different types of data such as numerical and alphanumerical in a very small area. It is a representation of binary code with two solid lines at the left and bottom of the code for recognizing the location, orientation and physical size of the symbol which is called Finder Pattern. Two alternating dark and light lines on the top and right side of the code is the Timing or Synchronization Pattern for recognizing the location of modules and cell structure. Also for the readability, there should be a white Quiet Zone around the symbol with a width of two modules or more (see Fig. 4). Data Matrix is mainly used for storing the serial number of products in manufacturing industry. It was initially invented for the Space Shuttle Program for item tracking.

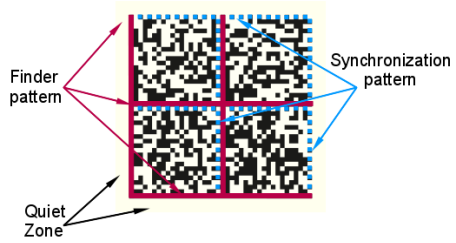

Fig. 4. Structure of Data Matrix.

Based on the employed error checking and correction level, Data Matrix has two types of error correction algorithms. There are five different error correction levels in ECC 000 to ECC 140 which are constructed based on convolution code-error correction. However, ECC 200 is the widely used algorithm for error checking and correction which uses the Reed-Solomon method.

In Data Matrix, the correction level of the ECC 200 is automatically chosen based on the size of the code. Data Matrix ECC 200 has an even number of blocks on each side, and each block is limited to 24 rows $\times$ 24 columns to prevent the symbol distortion. If the size of data is getting larger than the block capacity, the data will be split among more than a block. This type of the Data Matrix is the only version which has a white module in the top right corner of the symbol. The mentioned error correction method can reconstruct a symbol which is damaged up to 60%.

Data Matrix is highly supported in different industries, and it has specific standards and specs in industries such as Society of Automotive Engineers (SAE) parts marking, Air Transportation Association's (ATA) Spec 2000, Electronics Industry Marking Standard, NASA-STD-6002 - aerospace parts marking and many other similar industries.

The 2D Data matrix is also a recommended choice for sending barcodes over faxed documents because Data Matrix can withstand poor resolution and scanning conditions.

A. Comparison between QR Code and Data Matrix

There are a lot of comparison studies for QR code and Data Matrix in the literature. Here, we will highlight some of those conclusions.

TABLE II. Data capacity and features of Data Matrix and QR Code.

	Data Matrix	QR Code
Developer	RVSI Acuity CiMatrix (USA)	DENSO (Japan)
Numeric capacity	3116	7089
Alphanumeric capacity	2355	4296
Binary capacity (bytes)	1556	2953
Main features	Large capacity Small printout size High speed scan	Large capacity Small printout size High speed scan
Standards	AIM International ISO	AIM International JIS ISO

It is mentioned in [9] that compared to QR code, Data Matrix has the following advantages:

- Data Matrix is 30 to 60 percent more spatially efficient for encoding and it fits more easily on the page or screen.
- Data Matrix has more industrial support for encoding and decoding.
- The minimum size of Data Matrix is 77 percent smaller than QR code.

First read rate (FRR) is another feature for comparison. It is the number of successful reads divided by the number attempts. In [10], this feature is studied and compared for QR code and Data Matrix symbol using different phone cameras, distances and lighting conditions. The following table shows the results of this comparison:

The values in TABLE III are based on captured images from a distance between 5 to 25 cm using a Cold Cathode Fluorescent light. Kaywa and Semacode are two available online 2D code

TABLE III. FIRST READ RATE COMPARISON FOR DATA MATRIX AND QR CODE.

Percentage of success (%)			
	Full light	Half light	No light
Data Matrix Kaywa	100	100	100
Data Matrix Semacode	83.8	83.8	80.8
QR code Kaywa	75	75	75
QR code Quick Mark	100	100	100

generators. It is concluded that symbol size and data quantity are two important factors in FRR. In general, larger symbols with lower density and more sparsity have higher chance for being detected. The resolution of camera is not the most important requirement for a successful scan and decoding, for example a VGA camera could perform as well as a 1.3-megapixel camera. The auto-focus feature of a camera is also considered as an advantage for scanning. In the same experiment, the average measure distance for a successful scanning was 21.7 cm. They reported that reading distance doubled when the symbol size doubled. However, for a doubled dense symbol the distance was tripled. In summary, the key factors for the robustness of 2D-barcode reading are hardware capability of the reader, decoding algorithm of the reader and the cell size of the symbol.

In another study [11], QR code and Data Matrix are compared based on having some of the necessary scanning features. The following table is showing this comparison:

TABLE IV. COMPARISON OF QR CODE AND DATA MATRIX FOR SCANNING REQUIREMENTS.

	QR code	Data Matrix
Omni-Directional	√	√
Error Correction Ability	√	√
Work with Low light/Contrast	×	√
Clearly legible at 10cm away	×	√
Support VGA resolution	√	√
Data Capacity	√	√
Support Kanji characters	√	×
Additional Features	Scalable/can be divided	Scalable

Both of the 2D symbols have satisfactory data capacity, for example, a symbol with an approximate size of a version 10 QR code with 15 percent of data correction level will have a maximum capacity of 513 numeric and 311 Alphanumeric characters. This number of characters would be enough for most of the applications, but Data Matrix can offer a smaller symbol for the same amount of data.

IV. SCREAMING POWER FLAT FILE CONVERTER

We have developed barcode generator software to automatically search inside the utility batch files, extract the customer account information and convert them into 2D barcodes (QR code and Data Matrix). The Fig. 7 shows a snapshot of the Graphical User Interface (GUI) of the software. As the figure shows, user can select size of output barcodes, select the type of 2D barcodes, and will have the choice of HEX conversion. HEX conversion can be used for security purposes and also for encoding different types of data (or languages) which are not recognizable by the available decoder in mobile phone.

The summary part of the software shows some statistics about the different types of services and the distribution of that over the customers. The header of the summary part also gives the date, time and the barcode conversion settings. An example of the log file (summary text file) of the 2D barcode generator is presented in Fig. 8.

The input of the software could be a single Flat File that consists of multiple accounts or multiple Flat Files with different number of accounts in each file. The default input format of the Flat Files are ”*.dat” ,”*.pri” and ”*.txt” files, but it could be changed to any other format which supports text files. The input and output path could be easily set by the user in the GUI.

There are two ways of converting Flat Files to codes in the GUI. In the manual mode, the user can press the ”Run” button to browse the output folder and the input files. In the automatic mode, the user can press the ”Automatic” toggle button and the software will constantly check the input folder for the available Flat Files in it or any newly added files to the folder. Consequently, any file in the folder will be converted to barcodes and will be placed in the output folder in real-time. To have a lossless output image of the barcodes, all of the Data Matrix and QR code symbols are saved in the format of PNG files. Fig. 5 and Fig. 6 are examples of Data Matrix and QR code symbols generated by the software for the same customer. These barcodes hold most of the information that is on the utility bill. As is obvious in the figures below, the Data Matrix barcode requires a smaller area for this amount of data.

Fig. 5. An example of Data Matrix symbol.

Fig. 6. An example of QR code symbol.

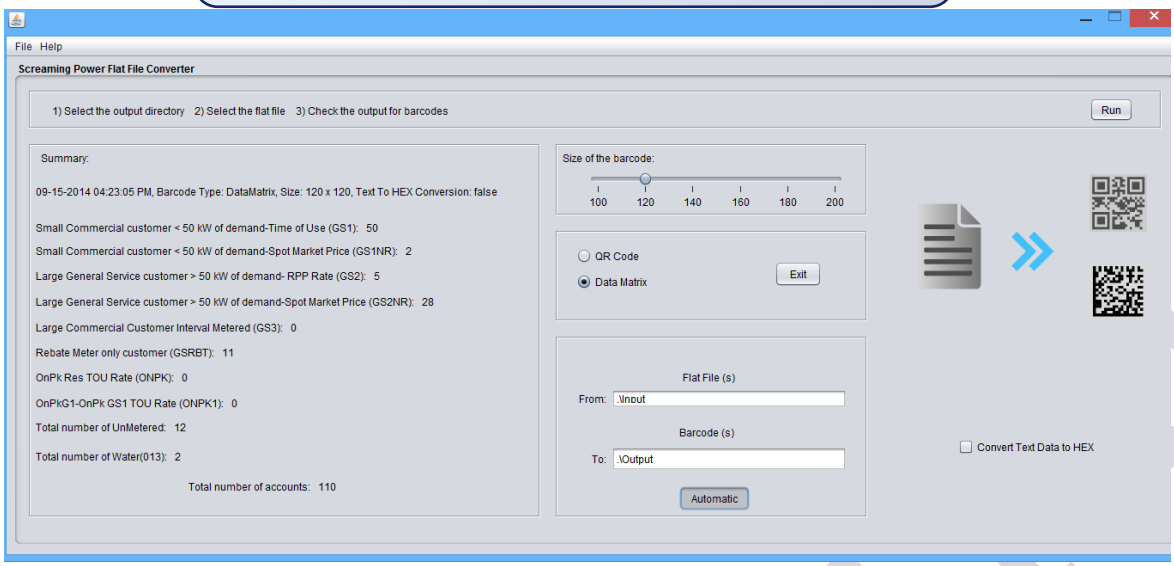

Fig. 7. Screaming Power Flat File Converter.

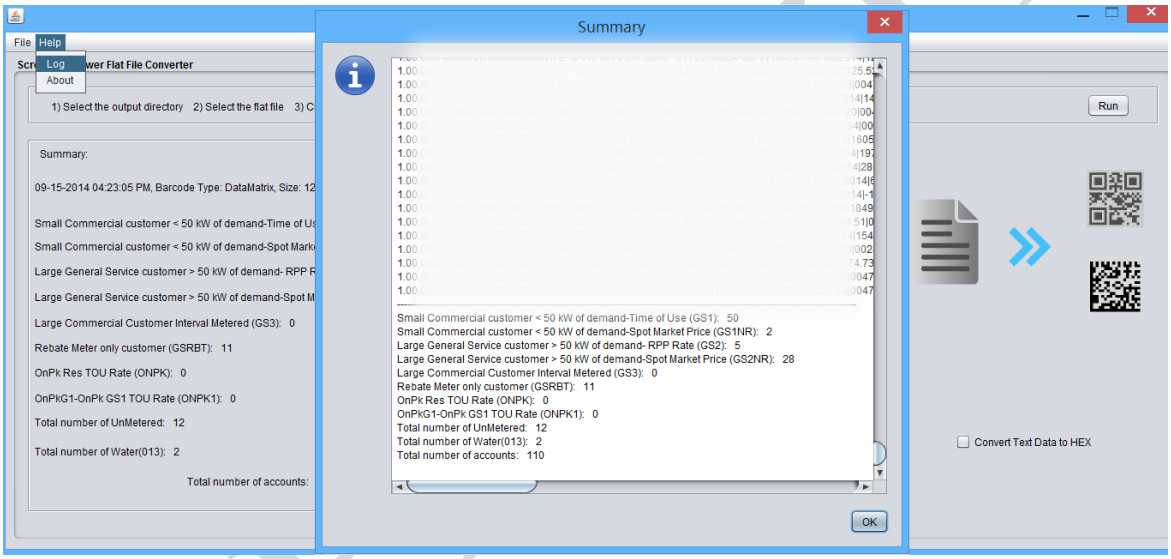

Fig. 8. Summary and log file of the customer accounts after generating 2d barcodes.

V. Conclusion

Data Matrix and QR code are two of the best and most widely used open source 2D barcodes. QR codes are traditionally used as a tool for marketing, advertisement and redirecting users to the website of a specific service or product. In contrast, Data Matrix is highly supported by industrial parties for storing and recovering manufacturing data in an efficient manner. This is the main reason that majority of people are more familiar with QR code rather than Data Matrix. Both of the 2D barcodes have the advantage of being able to encode a large amount of data in a small area. They are supported by a majority of smartphone scanners and can be used for encoding utility data. However, the Data Matrix compared to the QR code can store the same amount of data in an even smaller area. In general, Data Matrix is less sensitive to the lighting condition and the scanning experience is faster and easier than QR code. In our experiments, we found it easier and faster for untrained users to scan Data Matrix compared to QR code. On average, Data Matrix symbols could be scanned in a second, while QR codes needed a little longer scanning time.

References

[1] Yue Liu; Ju Yang; Mingjun Liu, "Recognition of QR Code with mobile phones," Control and Decision Conference, 2008. CCDC 2008. Chinese , vol., no., pp.203,206, 2-4 July 2008 doi: 10.1109/CCDC.2008.4597299

[2] D. Lorenzi, B Shafiq, J. Vaidya, G. Nabi, S. Chun, V. Atluri. Using QR codes for enhancing the scope of digital government services. In Proceedings of the 13th Annual International Conference on Digital Government Research (2012), pp. 2129.

[3] QRStuff. Whats a QR Code?, 2011. *http* : *//www.qrstuff.com/qr*$_c$*odes.html*.

[4] Steeman, J. QR code data capacity, 2004. QR4 QR Codes blog: http://blog.qr4. nl/page/QR-Code-Data-Capacity.aspx.

[5] D. Lorenzi, B Shafiq, J. Vaidya, G. Nabi, S. Chun, V. Atluri. Using QR codes for enhancing the scope of digital government services. In Proceedings of the 13th Annual International Conference on Digital Government Research (2012), pp. 2129.

[6] P. Kieseberg, M. Leithner, M. Mulazzani, L. Munroe, S. Schrittwieser, M. Sinha, E. Weippl. Qr code security. In Proceedings of the 8th International Conference on Advances in Mobile Computing and Multimedia, MoMM 10 (2010), pp. 430435.

[7] Trustworthy Ubiquitous Computing, Volume 6 of Atlantis Ambient and Pervasive Intelligence Series, illustrated ed. Springer, 2012, pp. 2138.

[8] Kapsalis, Ioannis. "Security of QR Codes." (2013).

[9] Arendarenko, Ernest. "A study of comparing RFID and 2D barcode tag technologies for pervasive mobile applications." Joensuu: University of Joensuu (2009).

[10] Kato, Hiroko, and Keng T. Tan. "Pervasive 2D barcodes for camera phone applications." Pervasive Computing, IEEE 6, no. 4 (2007): 76-85.

[11] Kato, Hiroko, and K. T. Tan. "2D barcodes for mobile phones." In Mobile Technology, Applications and Systems, 2005 2nd International Conference on, pp. 8-pp. IEEE, 2005.

[12] Lo, Leo, Jason Coleman, and Danielle Theiss. "Putting QR codes to the test." New Library World 114.11/12 (2013): 459-477.

