NTNU – Trondheim
Norwegian University of
Science and Technology

Security of QR Codes

Ioannis Kapsalis

Master in Security and Mobile Computing

Submission date:	June 2013
Supervisor:	Danilo Gligoroski, ITEM
Co-supervisor:	Tuomas Aura, Aalto Edgar Weippl, SBA-Research

Norwegian University of Science and Technology
Department of Telematics

Title:	Security of QR Codes
Student:	Ioannis Kapsalis

Problem description:

This master thesis studies the security issues related to QR Codes. The first part of our work includes 3 different attack schemes on QR Codes aiming to retrieve an alternated content when decoded, by repainting parts of the QR Code. This thesis also examines how easy it is to trick users into scanning QR Codes. To answer this question an empirical study is conducted. For the needs of the empirical study we print QR Code stickers and after posting them on different locations in 4 different European cities, we measure how many people scanned them and we finally analyze their behavior while scanning them. Identifying the users' level of security awareness concerning the security issues related to QR Codes through our online survey, is also among our main goals.

Responsible professor:	Danilo Gligoroski, ITEM
Supervisors:	Prof. Tuomas Aura, AALTO University
	Prof. Edgar Weippl, SBA Research
Instructors:	Martin Mulazzani Phd. Student, SBA Research
	Katharina Krombholz Phd. Student, SBA Research

Abstract

The 2-dimensional barcodes known as QR (Quick Response) Codes are increasing their popularity as they appear in more places in the urban environment. QR Codes can be considered as physical hyper-links that give the ability to users to access, through their mobile devices that are able to scan QR Codes, additional information located in a web-page. Apart from marketing, QR Codes have been also adopted in different areas such as the on-line payments. This development along with the trend that some of the users may follow which indicates to scan unauthenticated data, such as QR Codes located in public places, motivated us to investigate how QR Codes can be used as an attack vector. We first developed an implementation which attempts to brute-force QR Codes by attacking directly the modules, aiming to retrieve an alternated URL upon scanning the QR Code and after having applied the module changes. Our implementation showed us that such an attack is unfeasible in a real attack scenario. However, the second approach that we followed, in which we attacked the binary representation of the encoded string, we managed to produce the desired result. Furthermore, we conducted an empirical study aiming to identify the users' level of security awareness concerning the security issues related to QR Codes. The on-line survey that was accessible through our QR Code stickers, was our mean of interaction with the users. We deployed our stickers in 4 European cities (Vienna, Helsinki, Athens and Paris) and we managed to attract 273 individuals that scanned and visited our web pages. Out of these visitors, 83 participants completed our online survey. The results collected indicate that users are motivated mainly by their curiosity and they have serious lack of knowledge on the potential threats and the ways to protect themselves.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Motivation

QR (Quick Response) Code is a 2-Dimensional barcode that can store different kinds of information such as a link, plain text, SMS text message, addresses, URLs, Geolocation, email, phone numbers or contact information. QR Codes were introduced in Japan in order to track automobile parts but they became well known only when they were used as an advertising medium to distribute additional information to the users. When a user scans a QR Code with his/her smartphone camera using the appropriate QR Code software reader, he/she can reach the additional information. Thus QR Codes can be described as paper-based hyperlinks. This novel technology is now used in many new areas and according to latest measurements [13], [19] has been adopted by millions of smartphone users. This explosive growth in the last years indicates that that QR codes are not just a momentary fashion but a very powerful and versatile tool for the future.

This development along with the introduction of QR Codes in payment systems [23], [10], [43], [32] motivated us to explore what are the security issues that they might have. We do not only want to address the security implications that have already been the subject of previous research [38], but we intend to show in practice how a QR Code can be used as an attack vector. We focus our research in the scenario of phishing attacks deployed using QR Codes. However this does not mean that QR Codes have limited use only to this kind of attack. As we explain in this thesis there are many different attack schemes were QR Codes can be a useful tool.

Moreover, exploring the way that people face QR Codes is a task that gave us great motivation. Analyzing and defining people's behavior is usually a difficult assignment. However, it is essential when we examine some aspects of the problem that are affected by the users' behavior. Furthermore, social engineering attacks, which are included in our research, are also based in human behavior. Social engineering attacks try to manipulate users by exploiting some “flaws” in the users' behavior or by exploiting the lack of technological education.

1.2 Framework

At first we want to show in practice that some security issues related to QR Codes can easily be exploited by a malicious attacker. Thus, we developed different version of a program that is attacking the modules of a QR Code, aiming to retrieve an altered content when decoding the QR Code. We implemented three different attack scheme: brute-forcing the modules, attacking the binary representation of the encoded string and attacking the codewords. Our work focuses on the case were QR Codes encode a URL. However the same implementation could be applied in the case where a QR Code encodes other kind of information such as payment information. The brute-force implementation produce an output that indicates which modules of the QR Code have to change color in order to get the altered content. The only tools that a potential attacker would need is a black marker and/or a white correction tape/marker. Having these tools an attacker can change a legitimate existing QR Code in a way that it leads to a phishing web site.

Having the knowledge and proof that such an attack scenario is possible, we take the next step to explore users' behavior towards QR Codes. Our general purpose is to identify the level of security awareness that people have, concerning the security issues of QR Codes. Exploring the motivation of the user that scans a QR Code is our first challenge. However to better comprehend the root of the problem, which is people falling victims of phishing attacks, we have to explore aspects such as the technological experience or the mindset that users have when scanning a QR Code. The tool that we use to collect all this information is an on-line survey. In order for a user to reach this survey, he/she has to scan one of the QR Code stickers deployed by us in one of the selected locations. Finally to further expand our research and get a broader and international overview of the problem, we deploy our empirical study in four different European cities (Vienna, Helsinki, Athens and Paris). The intercultural nature of our study allows us to combine and compare the collected data from the different urban environments.

1.3 Research Questions

In order to focus our work in specific tasks and clearly determine the purpose of our research we state a set of research questions on which our study will be based. This thesis has as main goal to answer the phrased research questions that follow.

- What are the security issues related to QR Codes?
- Is it possible to attack a QR Code?

- Can QR Codes be used as attack vectors, especially in phishing attacks?
- What is the users' level of security awareness against threats related to QR Codes?
- Which countermeasures can be taken against the QR Code security issues?

1.4 Structure of the Thesis

This thesis is structured as follows: In Chapter 2, we present background information that is necessary to understand our work. In the same Chapter, we discuss the related work regarding the different areas were QR Codes are used, the phishing attack, and how this technique can be used in combination with QR Codes. Additionally, we discuss some of the known security issues related to QR Codes. Furthermore, in Chapter 3, the research methods followed are described. Detailed information about our implementation and the results retrieved when attacking a QR Code are included in Chapter 4. In Chapter 5, we present our empirical study as well as the obtained results. Additionally, in Chapter 6, we discuss our work, we outline some of the security implications revealed by our study and we answer the research questions on which this thesis is based. Finally, in Chapter 7, we present the main conclusions that we reached from our research.

Chapter 2
Background and Related Work

In this chapter we present selected papers, articles, books and other state-of-the art research publications that are related to our work. These publications helped use acquire knowledge about QR Codes, the different uses that they have and the security aspects that are related to them. Especially for the part of the security issues related to QR Codes, there are not many papers available in the literature. This thesis aims to contribute to this field and help the future research. The rest of the related work belongs to the field of social engineering and more specifically to phishing.

2.1 What is a QR Code

QR (Quick Response) Codes, are 2-dimensional bar codes that encode text strings and were introduced by the Japanese corporation Denso Wave Incorporated [12]. QR codes are considered as the evolution of the one dimensional barcodes. They are able to encode information in both vertical and horizontal direction, thus able to encode several times more information than the one dimensional barcodes. QR codes consist of black and white modules which represent the encoded data. In order to access the encoded data in a QR code, a built-in smartphone camera is used to capture an image of the QR code and then decode it using QR code reader software. There are 40 different versions of QR codes with different data capacities. Version 1 consists of 21 X 21 modules from which 133 can be used for storing the encoded data. Version 40, which is the largest QR code, has 23,648 modules which can be used for storing data. This practically means that it can hold up to 4296 alphanumeric characters [45], [55]. Here we use as an example version 2, see Figure 2.1, which is the size that is most widely used, and based on [38], [49], [61], [15] analyze the structure of the QR Code.

- (1) Finder Pattern: The three identical structures that are located in the upper corners and in the bottom left corner enable the decoder software to recognize

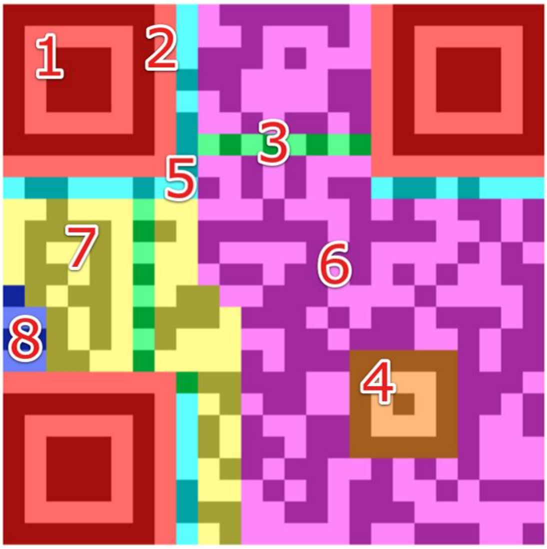


Figure 2.1: Structure of QR Code Version 2 (from [38] with permission)

the QR code and determine the correct orientation. These patterns also allow 360 degree (omni-directional) high-speed reading of the code. These structures consist of a 3 X 3 black square surrounded by white modules that are again surrounded by black modules.

- (2) Separators: The white separators that surround the Finder Patterns have width of one pixel and make it easier to distinguish the patterns.
- (3) Timing pattern: A sequence of black and white modules that help the decoder software to determine the width of a single module.
- (4) Alignment Pattern: This pattern allows the QR reader to correct for distortion when the code is bent or curved. The alignment pattern appears on version 2 and higher and the number of alignment patterns used depends on the version selected fro the encoding.
- (5) Format Information: This section consists of 15 bits and contains the error correction rate and the selected mask pattern of the QR code. The error correction level can be identified from the first two modules of the timing pattern (see figure 2.2). The format information is read first when the QR code is decoded.
- (6) Data: After the data is converted into Reed-Solomon-encoded data bits, it is stored in 8 bit parts (codewords) in the data section.

- (7) Error Correction: The data codewords are used in order to generate the error correction (EC) codewords, which are stored in the error correction section.

- (8) Remainder Bits: This section contains empty bits if the data or the error correction bits cannot be divided into 8 bit codewords without a remainder.

QR codes are able to encode different types of data, such as numeric, alphanumeric, binary, Kanji or control codes. Another big advantage of QR codes is that they are readable from different angles and the data can be decoded successfully even if the code is partially dirty or damaged [12]. This is because of the error correction that QR codes have, which is based on Reed-Salomon Codes [22]. There are four different error correction levels; Low(L), Medium(M), Quartile(Q) and High(H) (see also Figure 2.2) which can tolerate up to 7%, 15%, 25% and 30% damage, respectively [44]. The error correction level along with the type of encoded data influences also the capacity of the QR Code. Higher error correction levels increase the percentage of codewords used for error correction and thus decrease the amount of data that it is possible to be stored in a code. That is the reason why the error correction level L is usually preferred. An additional feature of the QR codes which increases the contrast of the picture and thus helps the reader software to decode the QR code is called masking. With masking, the generated QR codes have an equal distribution between black and white modules. The appropriate mask is automatically chosen by the encoding software while creating the QR code.

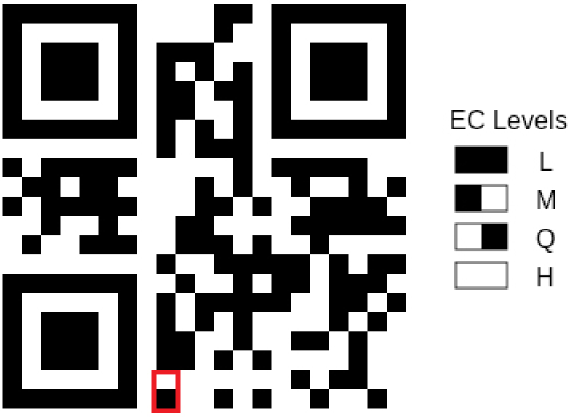


Figure 2.2: Error correction level of QR Code

2.2 Uses of QR Codes

QR codes were initially introduced as an efficient way to track vehicle parts by the manufactures. However, the use of the QR codes has changed and evolved significantly during the last years. QR codes are used to store contact information, geo-location data and text. Nevertheless, the most common usage is encoding URLs. The abilities that QR codes have were discovered very fast and, along with the increased use of smartphone, led to wide use of QR codes [7]. The industry of marketing is widely using QR codes as a supplementary way of advertising. A common practice in marketing campaigns is to add on an advertisement a QR code which takes the user to the company's web page, where additional information about the product is located. It can be used also as a promotion medium for a special offer to the people that will scan the code. In [16] are presented some very clever and innovative uses of QR codes in advertisements. The chain supermarket *Tesco* used QR codes as the main tool to boost on-line shopping and to penetrate further into the South Korean market. Another innovative and cost efficient marketing campaign was launched by a shampoo company, by introducing the QR Code hairstyle. People with these haircuts were actually a moving advertisement for the shampoo, since their "hair tattoo" upon scanning was leading to the company's web site. There are even advertising campaigns that are based only on QR Codes that are presented in oversized posters near central squares or crowded streets [30]. This is very convenient for the users since they do not have to type in the URL to visit the webpage to gather additional information about a product in a trivial way. Customers are able to purchase a product or pay for a service through a QR code since some companies have adopted the so called "one-click" payment [43], [32], [23]. In this case, a customer that wants to buy a product uses a QR code software reader on his smartphone to scan the QR code that is included in the promotional poster of the product. Then he/she is led either to an intermediate payment agent or to the company's web page to purchase the product. PayPal, which is one of the biggest payment companies, has already adopted this payment practice in some countries [10]. Of course, security issues arise especially when someone transfers money using this method. We will discuss the security issues related to QR codes in section 2.4.

Furthermore the versatile use of the QR Codes can be confirmed by the numerous uses that are presented in the literature. In [65] QR codes have been used for physical access control in combination with other security enhancing methods. In their paper Kao et al. proposed a safe authentication system by combining QR codes and the One Time Password technique (OTP). Their design includes a main server which holds the user information, a mobile application that generates QR codes, and a client PC with a camera in order to scan the QR code. In order to be authenticated, the user has to show to the client PC the QR code that the mobile application on his smartphone created. This QR code actually encodes an encrypted password, which was generated by the main server. This authentication scheme offers a good level of security; however, there are still some security issues on this model that remain open.

QR codes have also been used as a means to enhance digital government services in order to effectively distribute valuable information to the public [8]. In this case, QR codes are used to increase citizen participation and provide an enhanced experience and make the information exchange more interactive. Scanning a QR code can help to navigate through park trails by presenting maps to the users. Furthermore they can act as a supplemental material for education since they can be used with games that reward the active users that participate in scientific exploration competitions. QR codes were also proven to be an ideal means when it comes to sharing information between people who participate in the same social event [9] or when the goal is to trivially and effectively share information in order to support the learning process [24], [62]. Furthermore, interesting and creative uses of QR codes are presented in [18] and [26] where QR Codes are used as a surface on which an augmented reality application is deployed and as a result, impressive 3D virtual objects are produced and displayed to the user.

2.3 Phishing

2.3.1 What is Phishing and why it works

In Information Technology (IT) security, social engineering refers the art of manipulating people to reveal to unauthorized people confidential secret information and it is mainly used to steal data or to force one's way into a system that he is not authorized to access. One of the most popular practices in social engineering is phishing. The empirical study that is presented in this thesis is linked to techniques that are used for phishing attacks and, more specifically, to how QR Codes can be used as a tool in phishing attacks. For this reason, we explain what phishing is and why it actually works. Phishing is the practice of directing users to fraudulent web sites, which masquerade as legitimate web sites aiming to steal sensitive personal information such as usernames and passwords or credit card information. One of the main practices in the phishing attacks is the phishing emails that contain links to spoofed web sites or to web sites with malware. For example, an ordinary user might be tricked by an email which spoofs the identity of a popular site where the user already has an account. After clicking the link, the user is directed to a phishing site which in most cases looks exactly like the legitimate site, but it is a fake site. If the users will enter his credentials, like username and password, or even his credit card information, then the data will be forwarded to the scammers. The same scenario is followed if the user clicks on an attractive fake online advertisement. In these cases the user does not pay attention to the address bar or the browser's security indications (lock icon, SSL connection notice), which indicate that the site could be fake. However, as the related work presented below suggests, the average user usually ignores or does not understands the meaning of these cues.

Because of the fact that there are a lot of people that ignore the existence of phishing sites or do not even know what the term phishing means, phishing is a widespread phenomenon and with good profit for the attackers. Gartner Inc [17] conducted a survey in 2007 with the participation of 4,500 web users from U.S.A. Based on the data that Gartner Inc collected from this survey, it was estimated that 3.2 billion dollars were lost due to phishing attacks in 2007. The total number of victims to phishing attacks was estimated to be 3.6 million adults with an average loss of 886 dollars per phishing incident. In this survey, it is also mentioned that Paypal and eBay are the most commonly spoofed sites, but the phishing techniques have expanded to services like electronic greeting cards or charity businesses by using devious social engineering attacks. The same incorporation presented new data in 2008 in a report named “The war on phishing is far from over” [28]. This report states that the average consumer loss due to phishing attacks had dropped 60% comparing to 2007. On the other hand, the number of phishing attacks had increased by 40%. A more recent overview about the economics of phishing is presented at [11]. According to this source, the average loss on the detected phishing attacks for each person successfully phished was 1,200 dollars. It is also worth mentioning that majority of the 28,148 phishing web sites detected were hosted in the U.S.A.

The paper [46] attempts to answer the question why phishing attacks actually work and also which malicious strategies are the most successful at deceiving the users. This paper is based on a set of captured phishing attacks as well as on a usability study with 22 participants. The main findings indicate that the large majority of the well-designed phishing websites (90%) can fool most of the users despite the education level, age, sex or previous experience of the users. The findings also show that a big part of the users, ignore popup warnings about fraudulent certificates, in the cases examined in this paper. The ineffectiveness of the current anti-phishing cues is also revealed, as one out of four users in the study did not take into account the security indicators, the address bar or the status bar.

Accordingly, Downs et al. [25] conducted a study which included interviews with 20 non-expert computer users in order to identify and understand their decision-making process when they encountered possibly suspicious emails. The conclusions of this study reveal that being aware of this type of attack is not enough for a user to protect himself against it. Another valuable finding is that message-specific contextual strategies lead to more effective attacks than sender-specific. This is because in the message-specific contextual strategies the attackers try to exploit the fact that the user has already a valid account to the web site that they are spoofing. Moreover, it is mentioned that most of the strategies that users follow do not effectively protect them. Most of the users did not properly interpret cues like the lock icon or the origin of the phishing email. Finally, it was found that users had difficulties in understanding the warning pop-up messages, especially those that did not required any action.

2.3.2 The Battle against Phishing

In [47], Dhamija and Tygar examine how users can be protected from phishing attacks by using simple means that an average Internet user can easily distinguish. The security scheme that the paper introduces includes two techniques to prevent spoofing of legitimate web sites. Following this scheme the remote web server would be able to prove its identity to the users in a trivial way and, at the same time, it would make it really difficult for the attackers to spoof its identity. The first technique that the scheme uses is adding an image from the user database as a background to a trusted window. This can be combined with a visual hash generated by the browser and the server. These two elements form a background that help the user to identify that the window with the password input is trusted. However, as the paper notes, this technique might improve the security level, but is still vulnerable to some kinds of attacks like malware, leakage of the images, or spoofing of the visual hashes. In addition, this scheme presupposes that users are familiar with this scheme and are trained in a way that they will be able to recognize the trusted window. In most cases, training the users to behave in the proper and safe way is the most difficult mission but also a key element to the success of these kinds of schemes.

Similarly, the paper [50], through an empirical study, examines which security warnings are more effective and can prevent users from giving away sensitive personal information to phishing web sites. Two types of warnings are included in this paper: the active browser phishing warnings and the passive warnings. The active warnings interrupt the user's task and require form him to take some certain action, while the passive warnings do not interrupt user's task but appear just as notifications. The results of the study clearly show that warning indicators are definitely necessary since almost all the users that participated in the study were tricked by the phishing emails. The results also indicate that in general Internet users tend to ignore passive warnings when in the opposite case, they follow the active browser warnings which interrupt the user's task.

CANTINA is a content-based approach that was introduced by Zhang et al. [66] aiming to detect phishing web sites. The common way to identify if a web page is a phishing page is by looking at the surface characteristics like the URL and the domain name. CANTINA differs from these approaches because it uses the TF-IDF (term frequency/inverse document frequency) algorithm. By using this algorithm, CANTINA calculates the frequency of appearance of each term on the web page and then creates a signature by taking the five terms with the higher TF-IDF weights. This signature is given as input to Google search engine and, if the web pages that are in the top search results include the domain name of the web page which is under inspection, then this web page is considered as legitimate. The evaluation of this approach shows that CANTINA is very effective. The results show that the pure TF-IDF approach reaches up to 97% accuracy on phishing sites with 6% false positives and when it is combined with heuristics can score 90% accuracy with only 1% false positives. However, as [66] states, educating people about phishing attacks is a necessary countermeasure. For this reason we will analyze in the next section how users can be educated.

2.3.3 Educating people against Phishing

It is common belief that a good way to protect internet users against phishing is by educating them. Even though the automated tools are necessary and protect users from phishing attacks, education is believed to be a quite effective “weapon” against electronic scams. In [51] Sheng et al built an educating game in order to support users so that they can make better trust decisions. In that game Anti-Phishing Phil is a fish that eats worms which represent URLs. The users have to distinguish which worms are “healthy” and feed them to Phil. The game includes tutorials which are presented as stories and aim on teaching people how to discriminate legitimate and phishing URLs. Users receive additional feedback while they are playing based on their decisions. Moreover, between stages the users are able to review their decisions and get a short explanation or tips on the URLs they interacted with. All these learning processes, which aim to teach users how to identify phishing URLs, where to look for cues in web browsers and how to use search engines to find legitimate sites, proved to be beneficial for the users. The results suggest that interactive games like Phil can improve the ability of users to identify phishing web sites.

In a more recent study Sheng et al [39] are using an embedded training system, called PhishGuru, that teaches users about phishing while they use their email account. In the scenarios that this study takes into account, users receive training emails, which are actually simulated phishing emails. These training emails are not actually phishing emails, meaning that the users are not forwarded to a phishing site, but they look identical to real phishing emails. If a user falls for the phishing emails, and clicks on the link included in them, then a message appears explaining the risk is such situations while giving some advices to avoid such mistakes. This learning process, which is based mainly on the principal of “learning by doing”, proved to be very beneficial to the users. After the training phase, the users that participated in the study gained significant knowledge on what phishing is and how to not to fall for phishing. The paper concludes that a combination of automated tools for phishing detection combined with user education is the most efficient approach.

2.4 QR Code Security Issues

Despite the fact that the use of QR Codes is extended every day and they are adopted in more and more areas, the security issues that arise have not been examined in depth. There are only a few articles that try to warn people concerning the safety of QR codes [6], [67]. These articles describe how attackers can place QR Code stickers over a legitimate QR code to lead the users to a malicious web page, or even create a new fake advertisement masquerading a legitimate source. Scammers choose traffic-heavy public places like city centers or airports to deploy phishing or other variants of social engineering. Most people are unaware of these threats and what is even more alarming is that, as we show with our work, most users are unaware of simple ways that can protect them against social engineering attacks using QR codes.

Even in technologies like the RFID chips, which nowadays are adopted in a wide variety of services, security issues are not always considered seriously. RFID chips are used like access cards, bus tickets, ski lift passes or to detect domestic pets. The security issues of RFID chips and the attacks that can be deployed on them were examined in [31]. In this paper, Rieback et al. showed that it is feasible to generate an SQL injection attack or a buffer overflow using an RFID tag. More precisely, they developed a self-replicating virus which was using RFID tags as a vector to compromise back-end software systems via an SQL injection attack.

Concerning the security issues of QR codes, the first approach to this matter that can be found in the literature examines how a QR Code can be used as an attack vector [38]. More specifically the paper takes into account many different attack strategies that an attacker can follow. It also addresses the possible consequences to the victims who scan a malicious QR Code. The consequences can also affect the back end system that will try to serve the request generated upon scanning the QR code. For example, the SQL injection is believed to be feasible just by appending a semicolon followed by an SQL query on the request that will be processed by the back-end automated system. Another attack based on automated processes is the command injection. If the encoded information is used as a command line parameter without sanitation, this could be exploited to run commands on behalf of the attacker with severe damage to the end system. In simple words, if someone just scans a random QR code which an attacker has created and which includes one of the attack methods, that we just described then the unwitting user will generate the attack on behalf on the attacker. However the attack can target the device of the user that will scan the QR code. Depending on the way that the software reader works many attack scenarios can be deployed such as Cross-Site Request Forgery attack (CSRF) [36] or Cross-Site Scripting (XSS) attack [37]. Moreover the paper describes how through a QR code someone can deploy social engineering attacks, frauds or phishing and pharming attacks just by changing the color of some modules. In the book chapter called “Malicious Pixels - Using QR Codes as Attack Vector” [2], which is actually an extended version of the paper [38], a practical example of an attack is presented. The attack is actually deployed by changing white modules to black, only by using a marker. By this way the altered QR code now contains the URL of a phishing web page which has a similar URL with the original one. The example presented in the book shows how from a legitimate QR code which leads to “yahoo.at” it is possible to produce another phishing URL, which in this case is “yghqo.at”. This URL is very similar to the original one and this results can be achieved just by changing white modules to black. The work presented by Kieseberg et al. is the basis for the brute-force attack implementation presented in this thesis. However, the work presented in in this thesis differs in the sense that we attempt a brute-force attack on the QR code which is totally automated and the attacker does not have to produce the target QR code beforehand. More information about the implementation will be presented in Chapter 4.

In [57] Moore and Edelman present a method to identify typosquatting. The very interesting information presented in their work, which is related our work, is the economic viability of typosquatting. Typosquatting is the intentional registration of misspellings of popular website addresses. In 2010, it was estimated that there are at least 938,000 typosquatting domains targeting the top 3,264 “.com” sites, and most of these sites supported the pay-per-click ads. This information is fundamental proof that attacking a QR code in order to produce a misspelled and misleading domain name can be the basis of an effective phishing attack. The empirical study presented in this thesis attempts to identify the level of awareness that people have especially on this kind of phishing attack.

2.4.1 QR Codes and Phishing

Scanning a QR code in the wild is not a safe practice because, as mentioned before it can generate attacks to the back-end system that will serve the request or to the user’s device. Currently the only way to avoid phishing attacks based QR codes rely on the awareness and the ability of the user to identify malicious URLs or involve cues from external tools such as blacklisted domain services. These cues, as mentioned before are not always interpreted correctly by the users and are not so effective. Through the empirical study that we conducted, we attempt to identify how QR codes can be used in phishing attacks and how educated users are about this threat.

A very interesting project that aims to identify the user behavior concerning the QR codes is presented in [27]. This project investigates how easily stickers with QR Codes (*PlaceTagz*) are scanned by urban dwellers in a real world environment. *PlaceTagz* were deployed in different locations such us cafeterias, libraries, toilets or other places where someone would spend some time. When a dweller scans a *PlaceTagz*, he is taken to a dialog box where he can read comments from previous visitors but also leave his own comments. The findings indicate that curiosity is the main motive for dwellers that want to discover where the non-contextual QR Codes lead. Thus, with curiosity been a strong motivation to interact with an unknown source, which in this case involves a QR code, users ignore the security threats that might be hidden.

QRishing is the term introduced in [58] which describes the QR Code-initiated phishing attacks. Vidas et al. deployed for their study a QRishing experiment, but they did not use only flyers with QR codes but also rip-off flyers which they posted in different places around the city of Pittsburgh. They used 3 different kind of posters containing QR codes: plain QR code, QR code with instructions on how to scan the QR code, and QR code with information about their study. The locations selected for the study were mainly restaurants, bus stops and cafes. The people that would scan a QR code from a poster would be taken to a web page where they were asked to participate in an optional survey. This survey contained a questionnaire with questions that were trying to identify the initiatives and the behavior of the people that scanned the QR codes. In the same study, they also conducted another experiment which included visual monitoring of the user interaction with QR codes. A camera was set over a QR code poster with the primary goal to observe how many users scanned the QR code but decided not to visit the URL. Even though the participation was not enough to lead to safe conclusions, the big majority (85%) of the people that scanned the QR code visited also the web page. The QRishing experiment is similar to the empirical study that we conducted; however, our study differs in many aspects as described in Section 4 . The main differences are that our study was conducted under circumstances that are close to those that a real attacker would face and more importantly approaches the issue in an international setting which enables us to identify cultural differences between the users from different countries.

The different ways that QR codes can be used as an attack vector can be also verified by the attack that Jester, an anti-Anonymous hacker, claimed to have achieved [41]. He changed his profile picture on his twitter account to a QR code that encoded a shortened URL. The QR Code led the victims to a webpage which hosted hidden code exploiting a known browser vulnerability on iOs and Android. Apart from the WebKit [4] there were also secondary exploits that were exposing the device to the attacker. The attacker claims that he successfully trapped 500 victims that executed the OS-specific payload. Even though researchers and security specialists are questioning the success of this attack, they confirm that this king of attack is feasible.

Chapter 3 Methods

In this chapter we provide information concerning the research methods that were followed in order to complete this thesis. Both for attacking QR Codes and the empirical study, we followed well established research methods.

3.1 Attacking QR Codes

The first phase of this thesis included the implementation of attacking QR Codes in order to get different content when scanning the altered QR Code. Therefore the method we followed was "*Academic programming*" [29]. There are some special requirements and features that we must meet when following *Academic programming* :

- There are no specifications that have to be followed.
- Focus on back-end, instead of front-end(user interface).
- Very small iterations between introducing new features and testing them.
- Develop individually.
- Very fast development of the application.

Having these aspects in mind, we decided to search for an open-source project for encoding and decoding QR Codes. The decision to base our work on an open-source project was made because developing a project from scratch would not serve any of our goals and would violate some of the requirements. Furthermore having in mind the limited time that we had available, working on an existing project and developing our solution by expanding it, was the optimal choice. The required software did not need to have a friendly user interface since we focus on the low level of the implementation and especially on the part of encoding and decoding the QR Code. It turned out that there are enough developers that have implemented their own version of a QR Code encoder and decoder. There are 3 different versions of encode/decode libraries that qualified for our needs. The first open-source QR Code library was found in [53]. This library included an implementation of a QR Code decoder in *Java* and had received very positive feedback from the community of *Sourceforge*. There is also a project for encoding and decoding QR Codes written in *python*[1] [54]. This project seemed to be very well structured; however due to the lack of personal experience in the programming language *python* it was rejected. The last available library, which is probably the best and most popular open-source project, is the *ZXing* ("Zebra Crossing") library [68] which was also implemented in Java. The *ZXing* project, which is supported by Google is one of the most reliable and well structured 1D/2D barcode image processing libraries. It focuses on using the built-in camera on mobile phones to scan and decode barcodes but includes also an implementation for encoding and decoding QR Codes on desktops, which would cover our needs. The initial decision was to use the project from *Sourceforge* because the implementation was simpler than in the *ZXing* project and was also more easily adjustable. This was verified in the first test that we conducted, where it seemed to respond better to our changes than the *ZXing* implementation. Moreover, Java was the best available option to develop our solution, since it is a programming language on which there is sufficient previous experience. However, the final solution presented in this thesis required to combine the open-source project from *Sourceforge* with the *ZXing* library. More details are presented in Chapter 4 of this thesis.

3.2 Empirical Study

The empirical study we conducted aims to identify the level of security awareness that people have, concerning the security issues of QR Codes. For the needs of this empirical study we decided to use a survey. A survey is a non-experimental, descriptive research method which is very useful to researchers when collecting the desired data on phenomena cannot be done directly through observation [60]. In our case the survey is serving an additional purpose which is to gather spontaneous information from the users. The data was collected through questionnaires that users were asked to answer. In order for a user to reach a questionnaire he/she had to scan with his/her smartphone or tablet one of our stickers that contained a QR Code. Upon scanning a QR Code he/she was directed to our web site where the questionnaire was located. According to Babbie's terms [1] our survey is cross-sectional because we gather information at a single point in time and quantitative since all the questions asked were forced-choice questions (multiple choice). Furthermore, as Babbie suggests in his book "Survey Research Methods" [1] there are some main factors to consider when designing a questionnaire:

[1] http://www.python.org/

- State the questions clearly and make them understandable to the respondents.
- The questions have to be relevant to the topic.
- Short questions are easily readable and can be answered quickly.
- Avoid negative questions.
- Avoid biased items and terms.

Following these rules, we designed a questionnaire that contained only 7 questions which however are targeted and are able to extract the desirable information from the respondents. One of the main goals was to have a questionnaire that is not boring and can be answered in a few minutes. Moreover we used a simple format, normal fonts and neutral colors so that the look of the questionnaire is pleasant to the users.

Initially, the empirical survey was planned to take place only in Vienna, Austria, which was our current location, and we also had the necessary resources to deploy the survey. However, our will to establish an intercultural study, which will show the differences on the security awareness between people who have different cultural and educational background, led us to expand the study to four countries in total. More specifically, the empirical study was deployed in Athens-Greece, Paris-France, Helsinki-Finland and Vienna-Austria. We managed to ensure that we have the resources needed to successfully deploy our study. The main workload for the survey in the cities apart from Vienna was assigned to fellow colleagues who volunteered to help in our study by posting the stickers in different places. Specific guidance and requirements were given to our fellow colleagues on how they should proceed. The progress and the correctness of the whole procedure was also ensured by physically visiting these cities during the deployment. More details on how exactly we conducted the study are presented in Chapter 4. Finally, our decision to expand the survey gave us an insight on the variation of the technology adaption in these different countries and on how familiar people are with relatively new technological means like the QR Codes. As presented in section 5.2, the behavior of people towards QR Codes varies depending on the country of origin.

Chapter 4
Attacking QR Codes

4.1 Implementation

4.1.1 Brute-Forcing implementation

The first part of this thesis is focused on how a brute-forcing implementation on QR Codes can be developed. The main idea of brute-forcing a QR Code is that just by changing the color of some modules on the QR Code we will be able to generate an altered URL to which the user scanning the QR Code will be directed. In previous research [38], a similar solution is presented. However it is not based on brute-forcing or on targeting specific codewords, and is only based on changing white modules to black by comparing the target QR Code to the original. We decided to attack both black and white modules because we believed that it will lead to better results while keeping the simplicity of the attack on the same level. The only additional "tool" that an attacker needs is a white roll-on correction tape or a white marker. All the source code written for the brute-forcing of the QR Code was *Java* code and the environment in which it was developed was *Eclipse* Indigo IDE [59].

Attacking the modules

The first approach to our problem was to attack the modules of the QR Code directly. This is translated in changing randomly the color of some modules. The first plan included attacking the whole area of the QR code apart from the finder patterns which are the three squares located at the two upper corners and at the lower left corner, the separators which are located around the finder patterns, and the timing pattern. The error correction area had excluded since we wanted to test if just by changing the modules that represent the data the goals can be achieved. The targeted area is marked in red color in figure 4.1. All the elements that were excluded are essentials so that the decoder software will recognize the QR code and determine the correct orientation. The alignment pattern is also used for detection reasons and is excluded from our target area. The number of modules to be changed is determined according to the error correction level and the size of the QR code. In most of the cases the error correction level that is used is LOW = 7%. The goal was to get an altered URL from the QR Code while changing the minimum number of modules. Therefore the number of modules to be changed was a bit more than 7% of the total number of modules that correspond to data. For Version 2, this is translated to 28 modules.

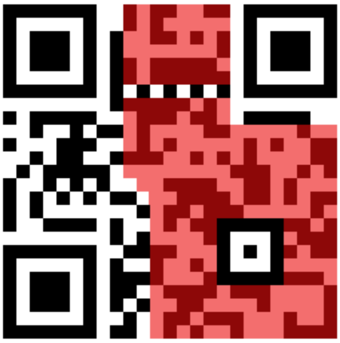


Figure 4.1: Target area on QR Code

The project initially used to implement this concept was the open-source project from *Sourceforge*. Our process is the following: the QR code image is given as a parameter to the program and initially the decoder function is called to decode the given QR code. While the QR code is decoded, we store in a boolean 2D array the representation of the legitimate QR code. This representation contains the values true which corresponds to black and false which corresponds to white. We have to decode once the QR code in order to get the error correction level, which is essential for the brute-forcing procedure as described before. After the decoding is completed and the legitimate URL is retrieved, we call again the decoder but this time with another parameter which enables the brute forcing function. The brute forcing function, having the error correction level and the dimensions of the QR code, is calculating how many modules should be changed. Having this information, and after selecting randomly the module positions but following the restrictions that are described above, we change the corresponding values on the array that represent the "legitimate" QR code. We also store the changed position in a different 2D array in order to present the positions to the user. Then, the altered boolean array, which represents now the new QR code, is fed to the reader for parsing. This procedure was testing for both cases: changing modules only from white to black and changing both black and white modules to the opposite color. Apart from experimental reasons, the decision to test changing both colors was made because the first test runs showed us that we get better and faster results when trying both colors.

The results generated from this first try using the open-source project from *Sourceforge* seemed to be the desirable. The generated output included modified URLs that were identical with the original one with only a few changes. For example, for "http://www.lufthansa.com", the generated output was "http://www.luftlansa.com" which is an ideal result that can be used for phishing attacks. However, when testing in practice to scan the altered QR Code by applying the changes indicated by the output, the QR Code was either unreadable or it was interpreted by the QR Code software reader as the original URL. The result of having unreadable QR Code was because of the fact that the changes applied were more than the maximum that the error correction can tolerate and in the same time the error correction modules were not changed properly so they would allow the QR Code to be decoded in an modified URL. In the case where the QR Code was decoded in the original URL, this was happening simply because the module changes indicated were not enough to produce a new result and the error correction offered by Reed-Solomon code was able to recover the "damage". This led us to the conclusion that the open-source project was either unreliable or it was working in a way that is not suitable for our task. Thus we decided to change our way of working and combine this open-source project with the *ZXing* library. The reason that we kept working on the open-source project from *Sourceforge* was because it offer some functionalities that the *ZXing* library does not offer. For example the *ZXing* library does not tolerate the changes on the boolean 2D array which represents the QR Code and returns exemption in almost every try.

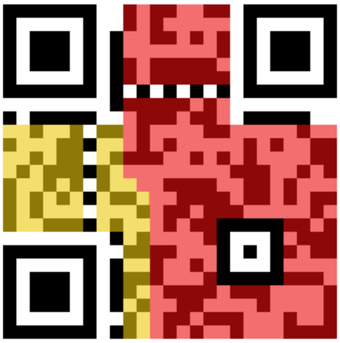


Figure 4.2: Target areas on QR Code

The new testing scheme is the following: the boolean 2D array generated by the qrcode project from *Sourceforge*, which represents the brute-forced QR Code, is used to create the new QR Code. To create the new QR Code image we used the encoder provided by the the *ZXing* library and the source code from [40] after applying the necessary changes. This image of the new QR Code including all the module changes is then decoded by the decoder of the *ZXing* library. This procedure verifies whether the changes applied on original QR Code resulted in the desired result. In this case, we do not only test brute-forcing the area presented in figure 4.1 but we extend our target area by including the error correction area as shown in figure 4.2 (yellow area).

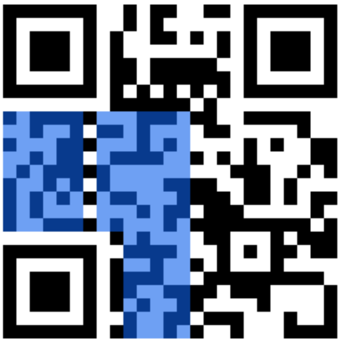


Figure 4.3: Target area on QR Code

Unfortunately, neither this method gave us the expected results. The produced QR Code was either unreadable from the QR Code software or was decoded to junk data. Again, the reason fro having unreadable QR Codes is the same as described before. The few cases where we managed to decode the QR Code into junk, were because the module changes in the error correction area and the data area manage to match in a valid decoding results. However, the produced result was not close at all to the original URL but contained garbled data. We also tried to target to smaller and more specific areas of the QR Code in order to succeed our goal. The idea to target to these areas was born by observing different QR Codes that were encoding similar URLs. We found that in some cases, URLs that have only one different letter, differ each other only in the area that is marked in figure 4.3. Again our try was unsuccessful since no valid QR Code was created. More details concerning the results are presented in the section 4.2.1.

4.1.2 Attacking the binary representation

After the failure to produce the desirable result with our initial planning, we decided that we have to change our “strategy”. Therefore, we tried a new approach which in fact was to attack the binary representation of the encoded URL in the QR Code.

Before a string is encoded in a QR Code it is transformed into a binary representation. In this way, the original message bits are included directly in the message’s Reed-Solomon encoding. Thus, each bit in the original message corresponds to a pixel in the QR code. Since our efforts to brute-force the QR Code by attacking directly the pixels-modules, we decided to randomly change a small number of bits in the binary representation of the encoded URL.

More precisely, we use the open-source project from *Sourceforge* to read the QR Code image which is given as a parameter to the program. The decoder is able to return us a boolean 2-dimensional matrix that represents the QR Code. The matrix consists of true and false, where false corresponds to white and true to black. This matrix is useful as described later to locate the modules that have to be changed. Moreover, while reading the QR Code image, we have the ability to retrieve the binary representation of the encoded URL. The binary representation of the string is actually included in the 2-dimensional boolean array that represents the QR Code. By isolating the area where the data is located we can target the bits corresponding to the string. The order of the bits is not the same as when directly converting the string to binary format, but this does not affect our task since, when changing randomly the bits, the order is not of any importance. While having access to the binary representation of the URL, we change randomly at least 1 bit. As we explain in Chapter 4.2.2, when having small Hamming distance between the original and the new representation, we achieve better results, meaning that the new altered QR Code has only a few differences with the original one. Afterwards, we get the string in which the binary sequence corresponds and we create a new QR Code image that encodes the altered URL. For this task we use the encoder of the *ZXing* library and the source code found in [40] that was adjusted properly to create the QR Code image as desired. We have to mention that we apply proper filters in order to select for encoding only the new URLs that have small Hamming distance [1]. The new URLs that have small Hamming distance, meaning that they differ from the original one only in a few positions, are more likely to be encoded to almost identical QR Codes with the original QR Code.

Then, the newly created QR Code image is fed back to the decoder of the *Sourceforge* project and a 2-dimensional boolean matrix representing the QR Code is retrieved. This matrix is compared with the boolean matrix that represents the original QR Code. Every position in each one of these 2 matrices represents a module of each QR Code. Every position on the new matrix that has a different value than the corresponding position of the matrix representing the original QR Code means that the module in this position has to change color. In our implementation we filter the results in a way that we keep only the QR Codes that were created by a URL that has small Hamming distance with the original URL and thus the number of modules that should be changed from the original QR Code image is small. In order to have a better visual representation of our results, we create an image that presents the modules that have to be changed. In this image the positions that have a black square correspond to the white modules of the original QR Code that should be

[1] http://en.wikipedia.org/wiki/Hamming_distance

changed to black and the positions with red square represent the black modules that should be changed to white. More detailed and comprehensive results accompanied with images of our produced results are presented in the Chapter 4.2.2.

4.1.3 Attacking the codewords

Although we had already achieved our goal to retrieve an altered result by changing some modules of the QR Code, we wanted to test another attack scheme. In this approach, we target the codewords that a QR Code consists of. The first step in order to encode data in a QR Code is to transform it in binary format. The binary sequence of the data along with the information about the encoding mode and the character count indicator, which are also in binary format, form the message polynomial. The message polynomial is divided by a generator polynomial in order to generate the error correction (EC) codewords. This procedure involves multiple steps of polynomial long division and is explained in details in [61] where there is also a detailed tutorial on how QR Codes work. In the final stage of the procedure, a final message in binary format is produced that includes the data and EC bits. In lower versions such as version 2 we use the final message as-is, when in larger versions we have to split the message into blocks. The bits of the final message are distributed to the codewords of the QR Code. Figure 4.4 shows how the codewords are distributed in a version 3 QR Code, along with the way that bits are filled inside a codeword. In our work, we mainly take into account version 2 QR Codes which are the most widely used, but the way that the codewords are distributed is identical in all the versions. As we can see codewords are 8 bits long and are used to store the data bits as well as the EC bits.

Following this information, we designed an attack scheme were we target one or more data codewords that we will change so that we will produce a different URL. In version 2 QR Codes with L (Low) error correction, there are 34 data codewords and 10 EC codewords. In our design we change one targeted data codeword in a value that we will produce the desired result, meaning that the URL will change as desired. Since we want to change the encoded URL only in one or two positions, changing the value of one codeword is enough. The selection of the correct codeword that corresponds to the exact part of the URL that we want to change can be “tricky”. This is because the binary sequence of the message is broken up into 8-bit binary bytes that do not necessary correspond to only one character, but contain information also for the neighboring characters. We also have to consider that the first codewords of the data part are used to store information about the encoding and the length of the message. Nevertheless, after the proper analysis it is not difficult to locate the exact position of the codewords corresponding to the URL and select the codewords representing the character that we want to change. However, after changing the data codeword we have to re-calculate the new EC codewords. Following the EC

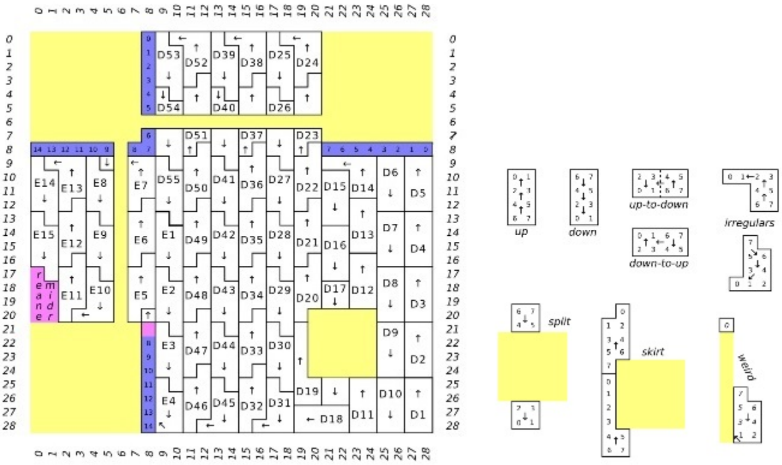


Figure 4.4: Codewords distribution and bit order inside codewords (from [64])

code properties, see [63] and [21], we know that we do not have to change all the EC codewords, but as we show in our results in section 4.2.3, if we change at most half of the EC codewords and leave the rest to the original value, we can still decode correctly the QR Code. To implement this attack, we based our implementation on the *ZXing* library. The *ZXing* implementation allows us to retrieve the values of the codewords both for the data and the error correction part. In more details, we use the encoder of the *ZXing* library to get this information for the original URL encoded in the QR Code as well as for the new URL that we want to create. The data codewords are the same for both URLs apart from the one codeword that we selected to change. Having the error codewords both for the original message and the new altered one, we compare each "new" codeword with the corresponding original one. In this way we can choose to update the EC codewords that will lead us to change the least number of pixels. Finally, we use the new EC and data codewords to produce the new altered QR Code. Then we compare the newly create QR Code to the original one to determine the number and the position of the different modules/pixels.

After applying this methodology, we wanted to automate a bit more our implementation and perform more tests on other QR Codes. This time we did not change the data codeword manually but used the encoder from the *ZXing* library to produce the data codewords for many different strings that had a small Hamming distance to the original URL. We then used the data codewords to produce the new EC codewords that corresponds to the new URL. Again, we were choosing the EC codewords that that will require the least number of pixel changes.

This last approach is not only the one that would allow to deploy a more targeted attack but is the most promising one to produce ideal results that can be applied in a real attack scenario.

4.2 Results on Attacking QR Codes

4.2.1 Brute-forcing the modules

The first approach to the problem was to attack directly the modules of the QR Codes and brute-force it by changing randomly the color to some of the modules. The results of this approach were not the expected as we did not manage to produce a QR Code that would decode in an altered URL that would be almost the same with the original. In more details we run our implementation as described in section 4.1.1 on a personal laptop as well as in a server offered by the Vienna University of Technology (TU). The laptop that we used has a 2-core processor tuned at 2.3 GHz and 4 GB of RAM. Our program was running in this machine for several hours per day without producing any valid result. TU's server has a processor with 6 cores tuned at 3 GHZ and 8 GB of RAM. Our brute-forcing program was running on the server for 64 hours and the only result that we retrieved was a readable QR Code which however was decoded in string that had no relation with the original one and could be classified as junk.

4.2.2 Attacking the binary representation

Our new approach was based on attacking the binary representation of the encoded string, produced some notably results. To test our implementation we selected some URLs from the top 500 global web sites. The first domain that could successfully be used in a potential phishing attack using QR Codes is the URL: *http://ebay.com*. As figure 4.5 shows, with only 24 module changes we can change the content of the QR Code to the phishing URL *http://gbay.com/*. The first square presents the original

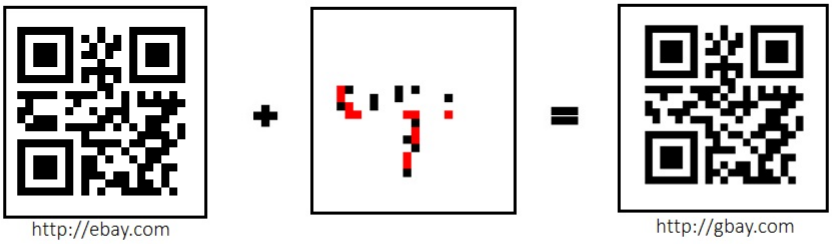


Figure 4.5: Attack scenario on QR Code encoding "http://ebay.com"

QR Code that is decoded to *http://ebay.com*. The positions of the black squares in the second frame, indicate that the corresponding modules have to change from white to black whereas the position with the red square indicate that the corresponding modules have to change from black to white. We can see that the altered URL differs by 1 character from the original one. However, the users that do not check the URL when scanning a QR Code would fall for this phishing attack. If we also consider the trust that people have towards automation our scenario is very probable to be successful. The results of our empirical study verify this conclusion as many users, trust a QR Code reader that visits automatically the link without showing them the URL. We also have to keep in mind that it is possible to achieve the same result by changing less modules than those indicated in the second frame since the error correction would tolerate some mistakes. Mistakes in this cases mean less module changes. In our implementation we included in the output the exact position where the changes had to be made, but here we only include the visual representation.

Another domain which is very popular and is used by millions of users mainly to store their data is taken into account in our second phishing attack scenario, is *http://dropbox.com*. Figure 4.6 shows again how from the original QR Code in the first frame that is decoded to *http://dropbox.com*, by applying the changes in the second frame we can get the altered URL *http://droqbnx.com*.

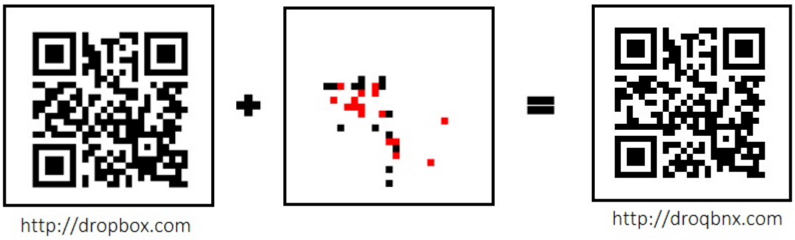


Figure 4.6: Attack scenario on QR Code encoding “http://dropbox.com”

The version of the QR Code used in this example is version 2. Thus it is anticipated that we will have more module changes than before. In this case the number of modules to be changed is 30, but as we explained before it is possible to get the same result with fewer module changes. The domain name *http://droqbnx.com* is a domain that was not registered at the time that we conducted our experiment and could be used by an attacker to build his phishing web page. The Hamming distance between the original and the altered URL is two since there are two different characters between them; ’q’ and ’n’ instead of ’p’ and ’o’. Again a user that automatically visits the URL or does a perfunctory check before visiting the link is very likely to become a victim of such an attack.

The last attack scenario that we present in this thesis is when using the domain name *https://www.americanexpress.com*. This domain name might be at the last positions of the top 500 domain names but is a web page that contains sensitive user information about their credit card. A phishing attack based on this domain could have a huge impact on the safety of electronic transactions and, as a result, many users would lose money. Figure 4.7 presents the results that our program produces. The total number of modules that an attacker would have to change in this case to get the altered URL *https://www.aoericanexpsers.com* is 28. Again in this case the version of the QR Code is version 2. Considering that version 2 has in total 625 modules, the 28 that have to be changed is a very low number.

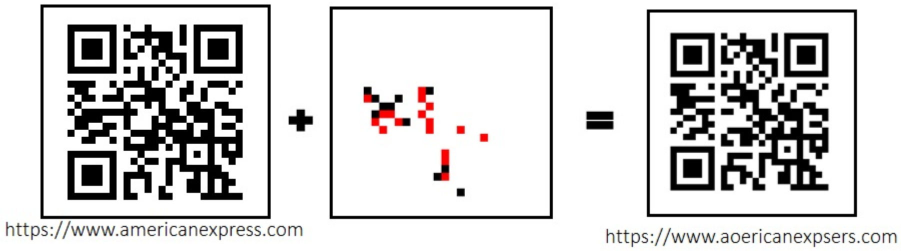

https://www.americanexpress.com
+
=
https://www.aoericanexpsers.com

Figure 4.7: Attack scenario on QR Code encoding "https://www.americanexpress.com/"

Moreover, the altered URL that is produced after the changes has Hamming distance 3 compared to the original one. This means that three letters are different and thus could it be used as a phishing site. A user performing a superficial check on the URL before visiting it could fall for this phishing attack.

The execution time in all the above examples presented was significant low. The time needed to get the desired results depends on many factors like the version of the QR Code and the Hamming distance that we want to achieve, but most importantly it depends on the number of modules that we will select as upper limit. When the limit of the number of modules to be changed is set to less than 24 we did not achieved to get any results. This result was expected as the error correction limit for the QR Codes that we tested was 7%. If we consider only the data and error correction modules for QR Codes of version two, we can verify that this 7% corresponds to 19 modules. This means that our implementation failed to reach the optimal performance which would be to have a valid result with only 19 modules instead of 24. When the limit was between 24 and 30 modules, the execution time for most of the cases was in average approximately 20s. However, despite the fact that we retrieved the results presented before, our program failed to produce any result meeting our requirements for other popular domain names such as *http://paypal.com* or *http://amazon.com.*

The retrieved results helped us to make some useful observations related to the behavior of QR Codes. The first interesting observation was that a URL which uses almost the full capacity of the respective version of a QR Codes that is encoded in is more easily changed to another URL with small Hamming distance. Especially for Version 2 QR Codes, we noticed that URLs with small Hamming distance are much more likely to differ only in a small number of modules when encoded to a QR Code. However, this observation is not enough to form a rule since in many cases and especially when using version 1 of QR Codes, this assumption does not hold. Furthermore, it was easy to notice that in general URLs with Hamming distance greater than three are encoded in QR Codes that differ a lot and it is very unlike to find module collisions.

4.2.3 Attacking the codewords

As we explained before in 4.1.3, in this approach we target codewords in order to get the desired altered result. We applied this approach for two different QR Codes aiming to get an altered result with the least number of pixels changed. The first QR Code that we decided to attack was the one encoding the URL *http://www.lufthansa.com.* We targeted our attack on the characters “uf” which according to the encoding scheme which groups the character 2 by 2, they should be encoded in the same codeword. Due to the way that the binary representation is split into 8 bits long sequences (see [61]), it is possible that some bits corresponding to one of the character are located to another codeword. Nevertheless, by changing the value of one codeword, it is still possible to change 1 or 2 characters. According to our analysis, the codeword where the characters “uf” are located in the codeword 15, which is located in the same position as the codeword 20 in figure 4.4. By changing the value of the codeword from “00111110” to “10111110” and accordingly changing 5 out of 10 codewords that this QR Code has in total we manage to produce the URL *http://www.lunthansa.com.* In figure 4.8, we present the necessary pixel changes that have to be applied on the original QR Code to get the altered URL. In total, we need to change 23 modules/pixels which is a fairly reasonable number for a version 2 QR Code. We have to mention here that if we encode the URL *http://www.lunthansa.com* following the normal procedure and then compare this QR Code with the QR Code representing the URL *http://www.lufthansa.com* we will find out that there are 36 different modules between the 2 QR Codes. So we can see that our implementation reduces significantly the number of modules that have to be changed. Moreover, the number of module changes that we determine with our implementation is the minimum number of modules that should be changed in order to retrieve the altered result.

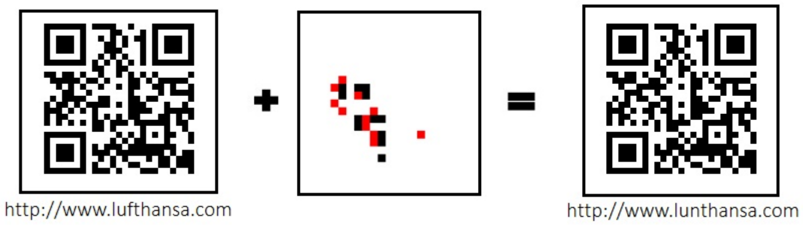


Figure 4.8: Attack scenario on QR Code encoding "http://www.lufthansa.com"

After retrieving this result, we performed some tests with the automated procedure that slightly differs with the one used in the above example. We tested our implementation on many different QR Codes with success, but we present the case where *https://www.americanexpress.com* is the encoded URL mainly to show clearly how much more efficient this attack scheme is, compared to the one attacking the binary representation. In this case we choose randomly a URL with small Hamming distance to the original one and we generate the data and EC codewords. Afterwards, we create the new QR Code and we measure the different pixels between the new and the original QR Code. To produce the result presented in figure 4.9 we set the upper limit of the different modules/pixels to 20. We actually managed to get even better results, since as figure 4.9 presents by applying only 18 module/pixel changes on the original QR Code that encodes the URL *https://www.americanexpress.com*, we get the altered URL *https://www.americanexhress.com*. The new URL has only one character different than the original one and can it be an ideal example for a phishing attack scenario.

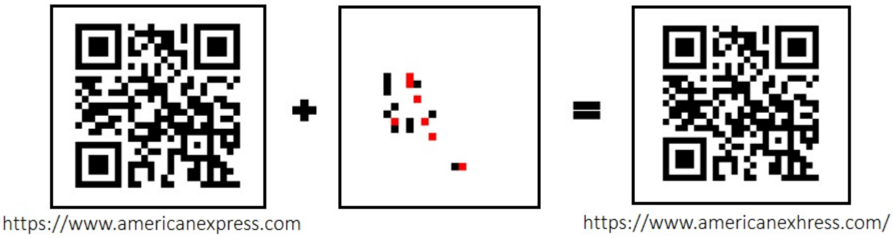

https://www.americanexpress.com
https://www.americanexhress.com/

Figure 4.9: Attack scenario on QR Code encoding "https://www.americanexpress.com"

As we can see, attacking the codewords is producing significantly better results than the previous approaches. Apart from that, attacking the codewords and applying the methodology described in 4.1.3, is the only way to retrieve the minimum number of required module/pixel changes. Additionally with this attack scheme we are able to deploy either targeted or generic attacks on a QR Code. In a real attack scenario it would be much more convenient for an attacker to target a specific part of the URL and produce the altered URL as desired. It might be a bit “tricky” to distinguish the exact area where the targeting character is located but with the appropriate analysis on the codewords it becomes trivial.

Chapter 5 Empirical Study

Having in mind that QR Codes could be used as attack vector in phishing attacks, we wanted to explore users' behavior against such an attack scenario. One of our main goals in this thesis was to identify how educated users are about the security issues related to QR Codes. We aimed to determine if users realize that scanning a QR Code with their device is not always a risk-free action. Additionally we wanted to understand how users interact with QR Codes in public locations and what their motivation on scanning them is. Our last goal was to prove how effective a phishing attack is while using QR Codes as the main means to deploy such an attack. Thus, we based our empirical study on the on-line survey that we created and deployed in Vienna, Helsinki, Athens and Paris. In order a user to participate in our survey, all he/she had to do was to scan a QR Code located on one of our stickers, that could be found in different location in each city. Upon scanning a QR Code, the web page where our on-line survey was hosted would open on his/her browser. As mentioned earlier we intended to deploy the survey also in Tokyo. However due to reasons that we explain in the section 5.1.3 we had to change our initial plans and abort the Japanese version of our survey.

5.1 On-line Survey

The first step for the empirical study was to create the survey which includes a questionnaire. Since the questionnaire was the only means that we had available to interact with the users and extract from them the desired information, it had to be very carefully designed. Fortunately, Google docs offer a service where you can create forms that can be used as on-line survey. This saved us time and helped us to avoid creating an unsightly survey. Moreover the fact that the created questionnaire was automatically connected with a spreadsheet was convenient to keep track of the answers. Thus, we used Google docs to create 5 different questionnaires in 5 different languages. The language used are: German for Vienna, French for Paris, Greek for Athens and Finnish for Helsinki. Apart from the official languages that each of the four cities has, we had an English version of our questionnaire which was easily accessible for all the user regardless their origin. We always had in mind that we have to create a questionnaire that is attractive to the user, has a neutral look and the most important, is not boring. Therefore we created a questionnaire with only 7 questions that were very targeted to the information of our interest. It was also very important for us that the users are able to answer the questionnaire in a few minutes. This was crucial to avoid having users that despite being willing to participate in our survey, when facing the questionnaire, they just abort it.

The questionnaire itself can be found in Appendix A; however, it is worth outlining the questions in order to give a better overview on the information that we are targeting. The first question that we asked users was "Why did you scan this QR code?". Revealing the motivation of the users will give us a better insight on how users face QR Codes and how someone can draw their attention. As the paper "No Cure for Curiosity: Linking Physical and Digital Urban Layers" [27] indicates, curiosity is in most cases the reason why people scan a QR Code. This was an aspect that we took into account when creating the stickers, and with this question we try to verify that our goal to pick their interest and make them question themselves what is behind this QR Code succeed. Another very important aspects in our study was the level of suspiciousness that users have against QR Codes. For this reason we asked them if they had any doubts or if they were expecting any malicious content before scanning the QR Code. Until these days, QR Codes are considered as an "innocent" and fast way to access a URL or to share contact information in an easy way. As described in section 2.4, the security issues related to QR Codes are barely addressed in the literature. Thus we wanted to identify if users have in mind that they might access malicious content through a QR Code, since they do not know where this QR Code leads. In close relation to users' suspiciousness, the next question that we addressed to the users was if they check the web address to which they are directed by the QR Code. The manual control that users can perform in order to verify that the URL of the web page that they are about to visit is a legitimate one, is the most basic security measure. However, this security check can be performed only when the poster containing the QR Code is a contextual poster which probably advertises a product or a service. In the case that there is not related information with the QR Code, then the user could only check if the URL belongs to a well-known domain. The task to distinguish a malicious URL just by optically checking it is almost impossible even for experienced users or IT professionals. The next question in our questionnaire aims to identify what is the users' background concerning phishing. Having been a victim of a phishing attack usually means that a user has gained some knowledge on how to protect himself/herself against a new phishing attack. The information from this question in addition to the previous answer can give us a good overview of the user's background on IT security. It is also useful to know how familiar users are with QR Codes in general. That is why we asked them how often they scan QR Codes. The data from this question can help us to verify if there is a correlation between the familiarity that users have with QR Codes and their level of security awareness. The information collected about the scanning frequency also provides an overview on the adoption of QR Codes. The last two questions were used to gather information about the age and the gender of the users. This information would help us to form some pattern on how security awareness is affected depending on age and sex. Our expectation were that younger users would be more active in this survey; however it was left to discover how well educated they are. As we can see in Appendix A, in some of the questions users have the ability to skip a question by choosing “I prefer not to answer”. We gave this ability to the users in case that they believe that a question is too sensitive or inappropriate for them and they prefer not to give any information.

At this point, we have to mention that our intention was that when a user reaches our on-line survey he would clearly understand what this questionnaire is about and what the purpose is. Thus, we added a header above the questionnaire where we explain why a user is visiting our survey. We clearly state that we are researchers aiming to identify the level of awareness that people have, concerning the security issues of QR Codes. Additionally, as we explain in the next section 5.1.1 we informed our visitors that we do not collect personal or sensitive information and the answers that they give us are anonymous. These clarifications were essential in order to gain users’ trust.

5.1.1 Survey Environment

The next step after creating the questionnaire was to create and configure the environment where we will host our survey. The Vienna University of Technology offered us generously a server to host our on-line survey. We created 5 different web pages, one for each language (English, German, Finnish, Greek and French). Additionally we created a front page which included links to the on-line surveys with different languages. This front page normal would not be accessed by the users, since the QR Codes lead directly to the questionnaires. However, if a user would manually change the URL on his device, then he would be able to visit the front page. In every web page we added a link to the English version of the survey. This was done to give the ability to foreign people living in one of the four cities to participate in our survey. This could be the case especially in the universities where there are a lot of exchange students. The different versions of our on-line survey can be found in Appendix A. Since on our web pages we mainly wanted to embed the Google form, the source code was simple *php* and *html*. The complete source code is also included in Appendix A.

Monitoring the traffic of our web pages was critical. Thus we included in the source code a script from Google Analytics which would help us to keep track of the visitors. Google Analytics provides some useful features like identifying the city of the visitor, the operating system of his/her mobile device or the web browser that the visitor used. Moreover, the total number of visitors per day which can be distinguished to unique visitors and returning visitors is easily obtained using Google Analytics. However, in order to have a redundant system and because we wanted to store some unique values for every visit, we implemented a piece of code which logs every visit. As mentioned before, we ensured that no private or sensitive information is collected from the users. That is the reason why we did not store the users' IP address or any other sensitive information. The exact values that were stored for each visit, along with the way that our logging system works are described in detail in subsection 5.1.2 were we also analyze the data that each QR Code encodes. Of course the necessary security measures were taken and the file containing the logs was not accessible by the users or any other external entity.

We explained before that the manual check of the URL to which a QR Code directs a user is one of the countermeasures against phisihng attacks. The server offered by the TU has a verified SSL certificate which means that the web browsers recognize that it is a trustworthy source. Nevertheless the domain name of the server was irrelevant with respect to our survey and it may give the impression to the users that it was not reliable. Therefore we bought the domain "qrcodesurvey.org" that gives a much better impression to the users that will check the web address. However, since we had already built the survey environment in the local server offers by SBA, which also offered us more versatility, we were just redirecting from the domain "qrcodesurvey.org" to the domain of the local server. In more details, we had 5 sub-domains under the qrcodesurvey.org with each one of them, redirecting to the equivalent location on the local server:

- en.qrcodesurvey.org → https://crunch0r.ifs.tuwien.ac.at/qrsurvey/en/
- at.qrcodesurvey.org → https://crunch0r.ifs.tuwien.ac.at/qrsurvey/at/
- fi.qrcodesurvey.org → https://crunch0r.ifs.tuwien.ac.at/qrsurvey/fi/
- gr.qrcodesurvey.org → https://crunch0r.ifs.tuwien.ac.at/qrsurvey/gr/
- fr.qrcodesurvey.org → https://crunch0r.ifs.tuwien.ac.at/qrsurvey/fr/

5.1.2 QR Code Stickers

To deploy our survey and reach out the users, we needed a medium that will pick users' attention. Thus we decided to create stickers that include QR Codes. In similar previous research papers like [27] and [58], different kind of stickers or posters were used. In [27] stickers accompanied by a small grey figure were used, when in [58] they used a combination of posters with plain QR Code, with QR Code followed by instructions how to scan the QR Code and posters with QR Code followed by text advertising the survey. In our study we decided to use 3 different kind of stickers. To create the QR Codes that we needed for our stickers we could have used one of the many on-line QR Code generators like the ones offered by Zxing team [69] or by QRGO [20]. However, this would be extremely time consuming since the total number of the stickers deployed is close to one thousand. Thus we created a small *Java* program, based in the source code found in [40], which uses the Zxing implementation. After the necessary changes, we created the *Java* program that generated the QR Codes used in our stickers.

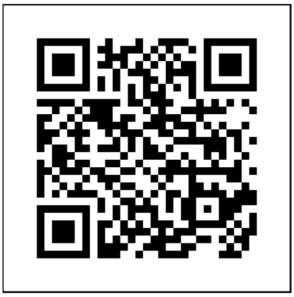


Figure 5.1: Plain QR Code sticker

The first type of stickers is the plain QR Code stickers, see figure 5.1. These stickers are very subtle and small in size and thus may not be easily noticed by the users. However, a plain QR Code sticker, when noticed, is also probable to pick users' interest since they might be curious to discover what is this "strange" QR code about. Exploring something totally unknown for some people is more intriguing than a contextual poster.

The second kind of stickers are those that, apart from the QR Code, included contextual information about our survey. In this case the user that face this sticker will know what the purpose of this QR Code is and what to expect upon scanning it. The related information may have different effects; it may simply partly satisfy user's curiosity and make him ignore the QR Code or it may create an additional interest to the user to participate in the study. Some users may even feel more confident to scan a QR Code that seems to be trustworthy since it provides information about the intentions behind it. Figure 5.2 shows the design of a QR Code sticker with description.

The last type of stickers that we produced was the one that included an image. We believed that this type of stickers would be the most successful because it had either a nice or a funny picture on it that would create a pleasant mood to the user.

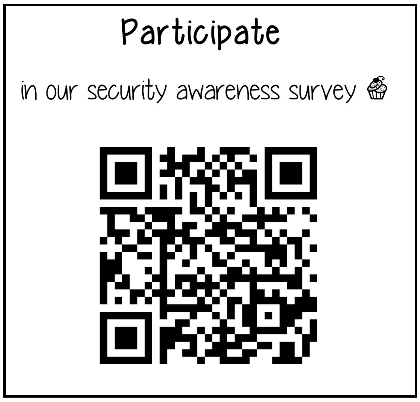


Figure 5.2: QR Code sticker with description

A user that has a positive predisposition is more likely to scan this sticker. Only the fact that someone laughs when he/she sees the image will motivate him/her to discover what is this QR Code about. He/She might be disappointed when he/she realize that the QR Code takes to in a not so funny content; however, we will have achieved one of our goals which is to make the user scan the QR Code, even if he will not leave his/her answers to our questionnaire. A sample of 2 out of the 12 different created versions of stickers with images and QR Codes is shown in Figure 5.3.

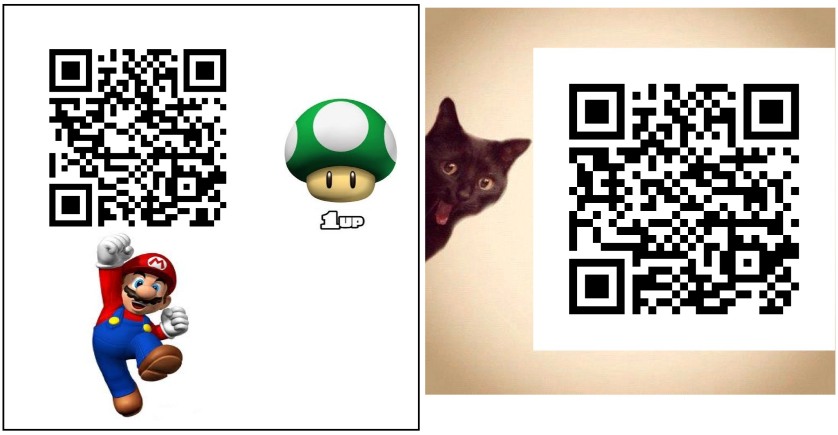


Figure 5.3: QR Code stickers with image

Locations and Target Groups

One of the decision that we had to take was to determine in which locations we will post the stickers. This was very essential because if we selected unsuitable locations then we will not have the desirable participation since the users will not notice our QR Code stickers. The main fact that determined our decision was to find locations where people spend some time either waiting or having fun or even working. For example places where people go to smoke a cigarette, talk with a friend or wait some minutes for a service to be delivered. Concerning the target groups, we made the assumption that, since younger people like student are more familiar to new technologies, they would be also more familiar to QR Codes. Of course this does not mean that we excluded other age groups. Thus we decided to focus on 3 main locations (and only a few stickers destined for random places):

- Transportation means
- Universities
- Toilets
- Random places

In all the big cities people spend a share of their daytime for their transportation. Metro, buses or trams are the most popular transportation means and thus the ones that we focused on. While waiting for the bus or the tram to arrive, people usually get bored and try to spend their time in a more productive way. We believe that people would have interest in exploring our QR Code stickers and they would be willing to spend these few minutes to participate in our survey. Therefore, we posted QR Code stickers in bus or tram stops, metro stations and in a few cases inside buses. An additional fact that led us to focus on the transportation means is that in all cities a large amount of people use them and thus we can address to a lot of people. In figure 5.4 we can see one of our stickers deployed in a central tram station after 2 weeks of posting it.

The universities located in the different cities where we deployed our study were one of our prime targets. As we mentioned before, young people were expected to be more familiar to new technologies like QR Codes. Additionally, as students ourselves, we know that students can easily get bored and they may search for a more pleasant way to spend their time. It is also common knowledge that young people are more curious about things that surround them. Our stickers can pick their interest and can create the will to the students to explore where these QR Codes actually lead. Moreover, especially the stickers including the description that specify that this is about a security research study, may be more successful in the university environment

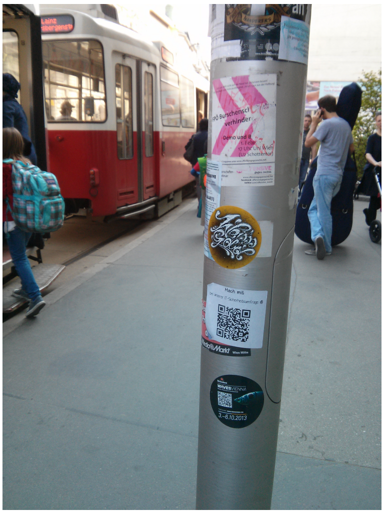


Figure 5.4: Sticker close to a tram station in Vienna

since there are a lot of students that conduct their own research projects and may want to participate in our survey out of solidarity.

The third selected target location was toilets. When someone goes to the toilet, apart from the fact that he/she is alone and relaxed, he/she usually observes the room around him/her. The average time that someone needs in the toilets is very close to the time needed to complete our questionnaire which was calculated to be around 2 minutes. A sticker which for example is placed against or over the toilet seat will most probably be noticed by the users. Having nothing else to disturb them, exploring a strange QR Code may make their time in the toilet more interesting.

Our last option for posting our stickers included some random places. We added this option in case that we would find a place that seems promising to get good results and is not included in our initial plan. This option was added only to give us a bit more versatility. However the stickers produced for these random locations were significantly fewer than in the other types of locations. This was because we did not want to lose our focus from our prime locations.

The total number of stickers that we created was 784. These stickers were equally distributed between the four different cities. For each city we had to distribute 196 stickers in a way that we will get the optimal results. However, in the same time we wanted to keep a balance between the locations so that we will have unbiased and comparable results. Only because of the fact that we anticipated young people to

Stickers/Location	**Transportation means**	**Universities**	**Toilets**	**Random places**
Plain QR code	25	25	25	12
QR Code with Image	12	12	12	6
QR Code with description	18	25	18	6

Table 5.1: Stickers Distribution

be more interested in our study, we gave a slight advantage to the universities. The exact distribution of the stickers is shown in table 5.1.

Data encoded and Logging System

As described in Chapter 2 QR Codes give the ability to encode a sufficient amount of data. Of course the capacity depends on the version of the QR Code that is used. In our study, we used version 2 of QR Codes because the amount of data needed to be encoded was relatively small. At the same time, version 2 is easy readable by the QR Code software readers and can be decoded very fast. The error correction level that was selected for all the QR Codes generated was L (Low) which is the level used in most cases. These settings give us the ability to encode up to 47 alphanumeric characters. In our study we had to encode the URLs that led to our on-line survey. Each URL was similar to this example: “http://fi.qrcodesurvey.org/?c=h&l=u&k=412815601”. The first part of the URL is the domain name, which in this case start with “fi” since it is for Helsinki (Finland). The domain name is followed by the parameters that we selected to have. This parameters were used in order to be able to keep logs and thus track more efficiently the traffic in our survey. The first parameter that we were logging is the city “c”, which for our example is Helsinki “h”. After the city, the location where this specific sticker is located is represented by the second parameter “l”. The last parameter is the unique key “k” which each QR Code sticker had. Having this unique key we could track the traffic that each sticker produced. These 3 parameters along with the date and time of each visit to our web page were logged in a secure file. The first use of this log file was to keep track of the traffic which was compared also with the results produced by Google Analytics. Moreover the log file helped us to measure how much time on average a user needed to complete our survey. Detailed information about these values is presented in Chapter 5.2.

5.1.3 Limitations

Our empirical study was designed in a way that it will approach a real scenario of a phishing attack using QR Codes. In this way we try to simulate how a real attacker would work, which however means that we might not achieve the optimal results that our research survey could have. Thus, we deployed our stickers in all the cities and all the locations mentioned earlier, without having a special license. This means that while deploying our stickers we could have experienced problems with the security staff of the buildings or the police when posting stickers in public. Fortunately we did not face any problem and we managed to deploy our stickers as we planned. However, the lifetime of each sticker was really uncertain. In many cases we found out that our stickers were removed, in most cases by cleaners, during the next few days after posting them. Especially the stickers located in university toilets or in bus stops were removed even on the same day. Stickers were not only removed by cleaners but also by random people that for some reason did not like their view. Another factor that affected the lifetime of the stickers was the weather. The empirical study was conducted mainly during the months of March and April, when in most cities of our study the weather conditions were not the optimal. Rain, snow or sun may destroy some of our stickers, especially those on the bus and tram stops. All these factors made our work particularly difficult in some cases and prevented us from achieving the optimal result in our empirical study.

As we described before, the initial plan included Tokyo as a location that we would conduct our survey. It would be a unique opportunity to deploy our study in Tokyo and to collect very interesting intercultural results since QR Codes are much more popular in Japan than in Europe. However, we did not managed to meet the requirements that the ethical department of the National Institute of Informatics in Tokyo had set to us. The fact that we designed our study to be close to a real attack scenario and the fact that the users were not totally aware where the QR Code stickers were directing them made it impossible to deploy our study in Tokyo.

5.2 On-line Survey Results

Having the fact that QR Codes can be altered and thus may be used as an attack vector in a phishing attack, we deployed our on-line survey so we can combined this result with the information collected by the users. Our survey was active in each city that we conducted our study for 40 days. The starting and ending date was not the same for all cites since we did not have the resources to deploy our study simultaneously in all cities. However, in general the study started on the 1st of March and was completed in mid-May of 2013 for all the cities. As mentioned earlier we used Google forms to collect the answers from the users that were scanning our QR Codes. Additionally we were logging the traffic of our web pages into a file with our own implementation as described in section 5.1.1. For the same purpose a Google analytics script was used as an additional way to track the traffic. The results presented in this section are extracted from these 2 sources. To achieve better accuracy we combined the data from both sources and we cross-checked the logging information. As mentioned earlier in section 5.1.2 for the needs of our study we posted in total, in all the cities that our study was conducted, 784 stickers with QR Codes. The total number of unique hits that our stickers gathered was 273.

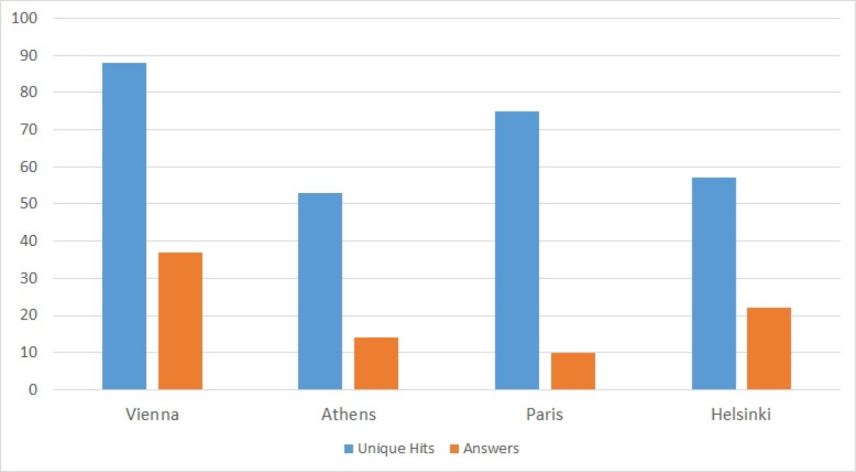


Figure 5.5: Unique hits and survey answers among the 4 cities

The distribution of these hits among the different cities is presented in figure 5.5.Out of the 784 stickers produced for all the cities, 113 (14,4%) were utilized by participants at least once, totaling the 273 hits. The percentage of the stickers' utilization is significant low, but in this Chapter we analyze why we had this result. Out of these visits, 83 participants completed our on-line survey. More detailed results related to each city where we conducted our study, follow.

5.2.1 Vienna

Vienna was the city where our study was most successful since our QR Code stickers gathered 88 hits and 37 survey answer. These hits were gathered from 40 different stickers that were scanned by users. In the on-line survey to which QR Codes took the users upon scanning them, participants were at first asked “Why did you scan this QR Code?” (the full survey is in the Appendix A). The majority (73%) of the users that responded to our survey, scanned the QR Code out of curiosity. Additionally 13% of the participants did it because they were bored and only another 13% was attracted by the related information. Figure 5.6 shows the distribution of survey responses from participants. We also included the option “I don't know what a QR Code is” and “I prefer not to answer” that received no responses from the participants. The choice “I don't know what a QR Code is” was added to identify that users actually know that this 2 dimensional barcode is a QR Code and that they have some familiarity with the technology. It was expected that users would scan the QR Code to satisfy their curiosity as our stickers, except those followed by a description, where not giving any hint where they lead to. However, following your curiosity might be harmful since scanning a random non-contextual QR Code means that you actually visit a totally unknown web site with unpredictable content.

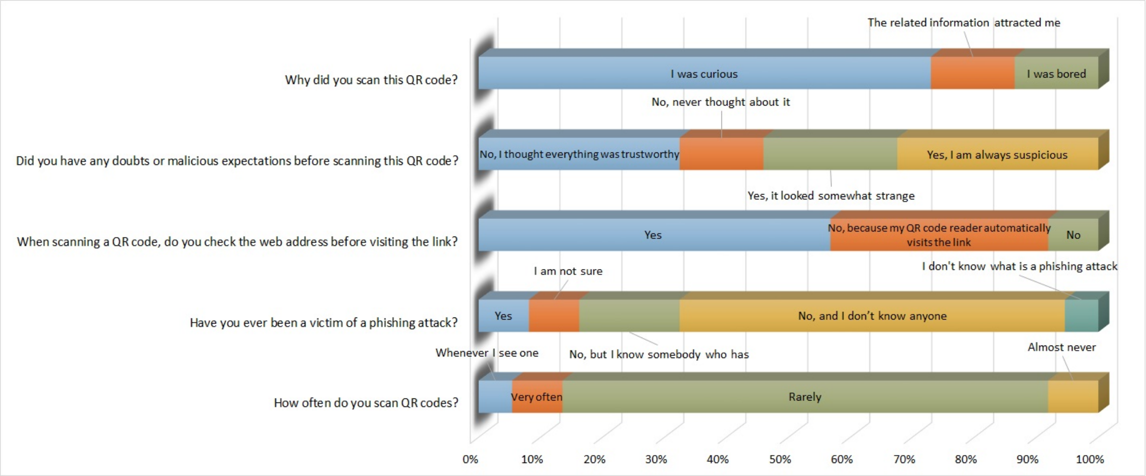


Figure 5.6: Survey responses in Vienna

To further explore users' familiarity with QR Codes, we asked them how often they scan QR Codes in general. The question was also aiming to identify how much the QR Code has managed to establish itself as a new technology. The vast majority (78%) of the users replied that they scan QR Codes rarely and another 8% replied "Almost never". This indicated that QR Codes have a long way to reach a satisfying level of adoption among the users and to become more popular to a wider range of people. This conclusion particularly holds for the users in Austria since they reported to scan QR Codes less often than other users from the other countries.

We also asked the participants if they had any doubts when scanning the QR Code or they were expecting malicious content. It is quite perturbing that 46% of the participants felt either safe to scan a random QR Code or, even worse, never thought about any danger that might be hiding. However, it is encouraging that 56% of the participants checks the web address before visiting the web page. This manual check is the primary and unfortunately the only way that a user can protect himself/herself from visiting a malicious web page and only in the case where the QR Code is accompanied by some related information. The other case when a user can verify that the URL is a legitimate URL, is when it is a well-known and trusted domain. Distinguishing a potentially harmfull web site is non-trivial even for

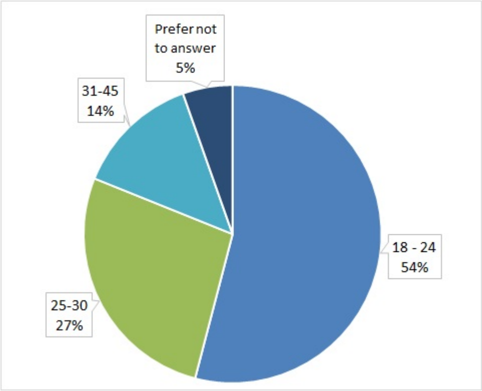


Figure 5.7: Age distribution, Vienna

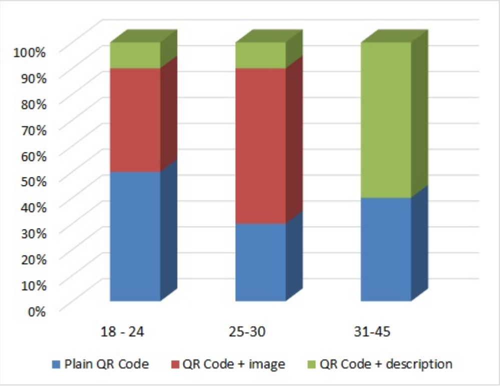


Figure 5.8: Stickers' preference depending on the age, Vienna

experienced users, as distinguishing a malicious URL just by looking at it, is in most cases unfeasible. Nevertheless, even for the participants that checked the URL before visiting the link, the fact that they actually visited an obscure URL to an unknown domain is still problematic.

Additionally we can see that phishing attacks are not widespread, as 62% has never been affected by a phishing attack. However, this is only based on what users reported. It might be the case that users are embarrassed to admit that they have been victims or maybe they never realized it. Nevertheless the 8% of the users that reported to have been victims of phishing attacks along with the 16% that had someone in his/her environment that had been a victim of phishing attack indicates that phishing is an actual threat and affects a part of the users. Moreover, there is a small percentage of users which is either not familiar to the term phishing attacks, which might not necessarily mean that they aware of the treats related to it.

As figure 5.7 shows more than half (54%) of our participants were in the range of 18-24 years old. Together with the 27% of the participants that reported to be in the range of 25-30 years old, they account the 81% of our participants in total. This does not align with the two age groups that have most adopted smartphones [35]. However, this result can be explained as our target group was mainly the young people. Among this ages we can also see (see Figure 5.8) that plain QR Code stickers and the stickers including an image are those that were mostly preferred. Opposite to that behavior, the participants in the range of 31-45 prefer the stickers including description and the plain QR Codes. The result that was quite unexpected was the fact that the participation of women in our survey was significantly low. In Vienna only 3 out of the 37 in total answers were given by female participants. Our expectation were much higher than that, since women in Europe are using smartphones at the same level or even more than men. According to [35], male users account for the 49.5% of the total mobile users while female users are the remaining 50.5%. However, it seems that female users are not so familiar with QR Codes or that they are not so “attractive” to them.

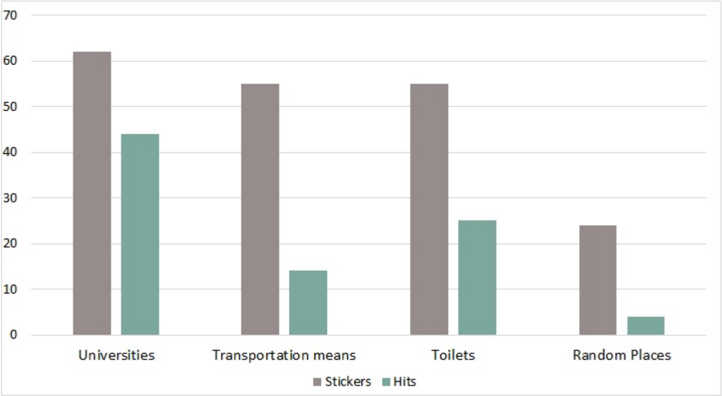


Figure 5.9: Stickers’ performance by location, Vienna

Figure 5.9 shows the distribution of the hits that each group of stickers gathered, depending on the location. We can notice that the stickers posted at the universities outperformed all the other locations. Most of our stickers were deployed in the premises of the Vienna University of Technology. The performance of these stickers is justifiable since a large number of people is visiting the university premises every day. Moreover, the curiosity that students have along with their familiarity with new technologies can produce such a result. As the collected answers suggest, quite enough students are bored while they are in the university and they are looking for ways to entertain themselves. We were expecting many hits from the stickers located in the bus stops or the other spots related to the transportation means since people which for example are waiting for the bus are likely get bored and would be pleased to find a way to fill somehow this idle time. However, our expectation were not met and this is because our stickers were destroyed in a very short time after deploying either due to weather conditions or because the cleaning personnel removed them.

The fact that each sticker had a unique key helped us also to identify which kind of sticker was the most popular. As we can see in the figure 5.10, the stickers that included an image along with the QR Code were the ones mostly preferred by the users. With respect to the total number of stickers, those with an image gathered not only the largest number of visits but also most of the answers. Apparently people are more attracted by a funny image and feel the need to explore what is behind

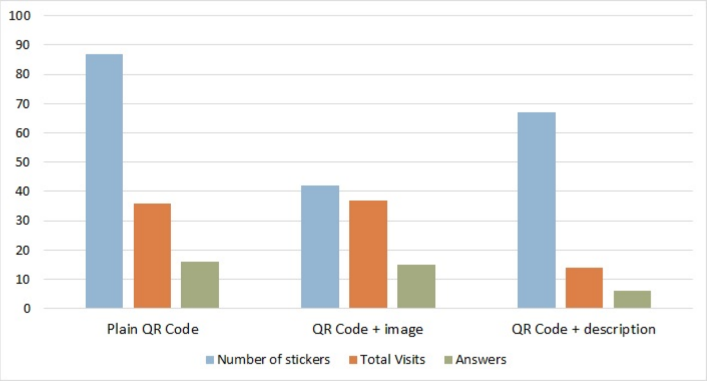


Figure 5.10: Stickers' performance by type of sticker, Vienna

it. This is an interesting finding and should be taken into account when analyzing phishing techniques or other kind of attack. Certainly, a funny image is not the most appropriate tool when deploying phishing attacks based on a bank's spoofed web site, but they can be used to trap the user in revealing less valuable but still sensitive information such as a personal identification number.

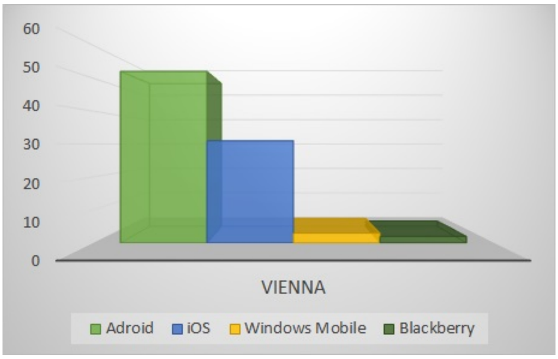


Figure 5.11: Distribution of operation systems, Vienna

Finally, we examined some other metrics such as the users' devices operating system. Out of the 88 visits that we had from Austria, 47 (53%) were from users possessing an Android device and 28 (32%) from users using an Apple iOS device. The rest 15% was a mix of Windows Phone and Blackberry mobile devices (see figure 5.11). Our results on the distribution of the operating systems follow the data presented in [34] about the market share on smartphones that every operating system has. The percentage of the Android devices presented in [34] is even bigger than the one that we noticed. However, our sample is small and very targeted, and thus it can be expected to deviate to some degree from the general statistics.

5.2.2 Helsinki

Helsinki is the next city that is included in our study where we managed to collect 23 answers out of the 57 hits that we had in total. It is well known that stickers and other kind of posters in the Scandinavian countries are not so commonly and widely used. They are usually considered as vandalism if they are not placed in some specific areas that it is allowed to be posted. Moreover neat and clear bus stops or walls of public buildings is a very common image in Scandinavia. These are some of the factors that only 15 of our stickers were used from the users. Again, as in Vienna curiosity was the main reason for scanning the QR Code as 64% of the users reported. In Helsinki we also received some interesting free text answers such as: "Because haters are going to hate cats" . This interaction give us the feeling that some of the users were feeling quite comfortable and safe in our survey environment and this is an aspect what an attacker would like for his victims.

As figure 5.12 shows users from Finland are more suspicious and do not trust easily an unknown source as 72% of the participants responded either "Yes, it looked somewhat strange" or "Yes, I am always suspicious" when we asked them if they had any doubts or malicious expectations before scanning the QR code. This behavior matches also with the 64% of the user that reported to check the URL of the web page before visiting it. However, the fact that 22% of the participants use a QR Code reader that visits automatically the link prevents them from doing the most basic and simple security check, which to check the URL before visiting it.

The question "Have you ever been a victim of a phishing attack?" was also given to the users. We can see that only around 10% of the participants has been victim of a phishing attack and that almost 50% of the participants was not affected directly by a phishing attack. However, 10% of the participants still do not know what a phishing attack is. This group of users might actually know what are the related threats to a phishing attack but they are unaware of the term. However, they might also be unaware of the related threats, thus it is very likely to become victims of such an attack. Comparing this result with the corresponding data collected from the other cities, we can conclude that phishing attacks are a bit more widespread in Finland than in the other countries. This might also be the result of the honesty of the participants from Finland have.

Finally the responses about the frequency of scanning QR Codes are a bit different to those from Vienna. The 32% of the participants reported that they scan QR Codes either whenever they see them or very often. This indicated that users in Finland are a bit more familiar with this new technology. However, QR Codes still

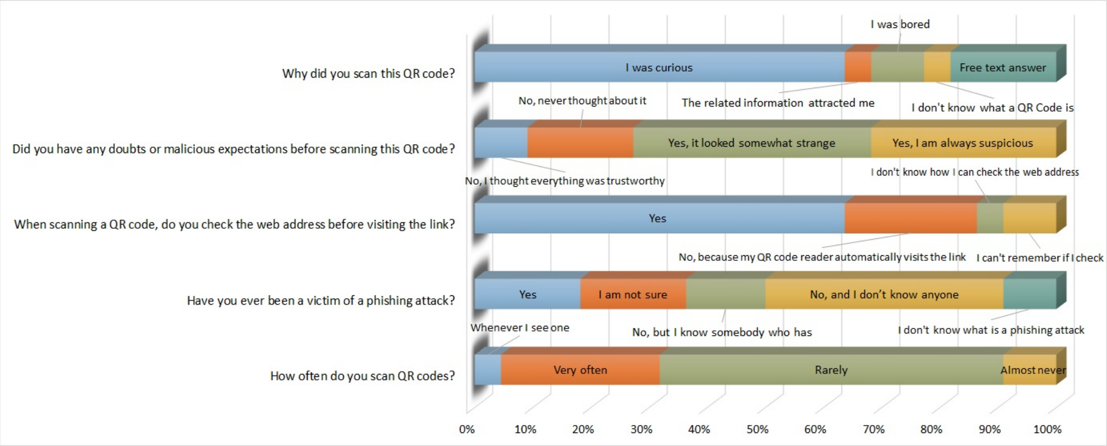


Figure 5.12: Survey responses in Helsinki

need additional time to penetrate more in the mobile market and gain a larger group of users.

In figure 5.13, we present the age groups that participated in our survey. As expected the biggest majority (71%) of users consists of users in the age range of 18-24. Another 19% of participants consist of users in the age range 25-30. These two groups sum up to 90% of our total participants and surpass by far the corresponding percentage of mobile use that these age group posses according to [35].

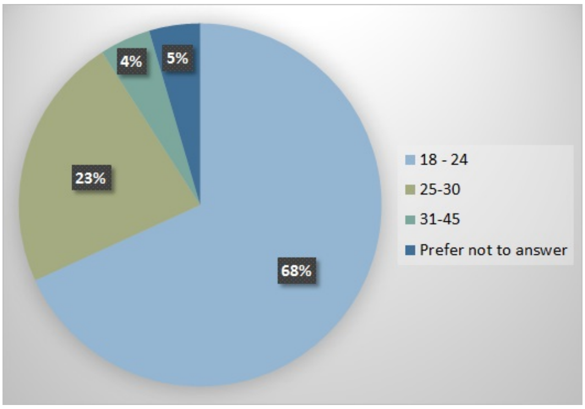


Figure 5.13: Age distribution, Helsinki

As we explained before, the reason for the age distribution noticed is that our target group was mainly the young ages and because younger users tend to be more familiar with new technologies. Concerning the sticker preferences that the participants in our survey had, we can say that stickers including an image dominated users' selection. Almost all the answers in all the age groups were gathered by stickers including an image. Furthermore, male participants were again the vast majority of our participants. Only 10% of our participants were female users which does not align with the smartphone usage for women in Europe [35].

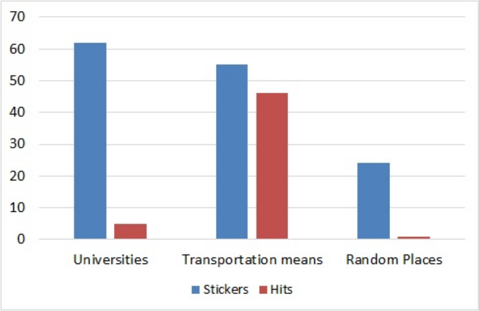


Figure 5.14: Stickers' performance by location, Helsinki

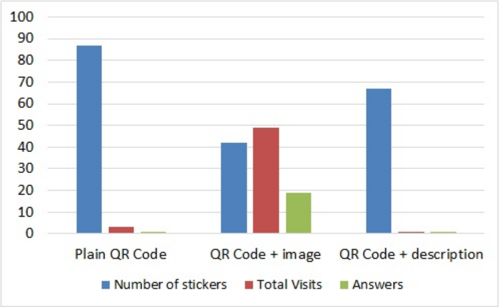


Figure 5.15: Stickers' preference by type of stickers, Helsinki

In the figure 5.14, we present the performance that stickers had, depending on the location. It is clear that only the stickers posted in the different spots related to the transportation means were quite successful as they manage to gather most of our visits. People in Finland use the public transportation much more than the average citizen in central Europe so that is why we had the chance to address our survey to a big part of the population. Unfortunately, the stickers deployed in the universities did not meet our expectations. Student participation was very low and this was probably because our stickers were removed in a short period or because students were too busy to notice our stickers. Moreover we notice that there were no registered visits from the stickers deployed in toilets. Apart from the fact that toilets are cleaned in a daily basis we are unable to explain why our stickers did not manage to gather any hit. Additionally in figure 5.15 we show how successful were the stickers including a QR Code and an image. Comparing to the total number of stickers including an image, to the total number of visits, proves that this kind of stickers is the most appropriate to draw users' attention.

Unsurprisingly, the majority (51%) of users from Helsinki that visited our on-line survey used an Android mobile device. However, as figure 5.16 shows the second biggest group of users that visited our web site were those using a Windows phone mobile device. Especially for Finland, it is not very unusual to see a large group

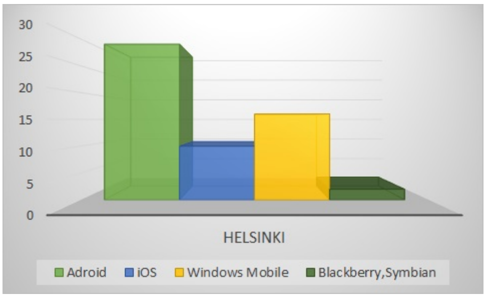


Figure 5.16: Distribution of operating systems, Helsinki

of users with Windows mobile devices since Finland is Nokia's home country and its market share is significant. What is surprising is the fact that the number of Windows phone users was even higher than the number of those with an Apple iOS device. It is also notable that there are still users with Symbian devices which is an operation system that is obsolete and there are not many new devices supporting it.

5.2.3 Athens

Athens is another European city where we conducted our empirical study. The population and the city area of Athens is significantly large and this made it especially difficult to find the appropriate places where we would deploy our stickers. Even with the four specific categories that we have already distinguish it was difficult to decide where to deploy the stickers. The final decision was to mainly target some central metro and bus stations and the biggest university of Athens (National and Kapodistrian University of Athens). As outcome of our efforts we collected 15 survey answers out of the 53 visits that we had in total. The sum of the hits were gathered from 30 different stickers.

As we can see in figure 5.17 all the participants in our survey reported to scan our stickers out of curiosity. This result not only aligns with the answers in the same question from Vienna and Helsinki but verifies that triggering users' curiosity is the most effective way to attract "victims". Moreover, it is clear that the majority of our participants in Athens are more unsuspicious since more than 60% of them responded that they had no doubts or were not expected any malicious content before visiting our web page which was totally unknown to them. These results could indicated that there is a lack of technological education in many users. However, the answers in the next question contradict to this conclusion as almost 70% of the participants are actually performing the manual URL check before visiting the web page. Identical to the results from the previous cities there is a group of users, which in this case is 14%, that use a QR Code software that automatically visits the URL upon scanning the

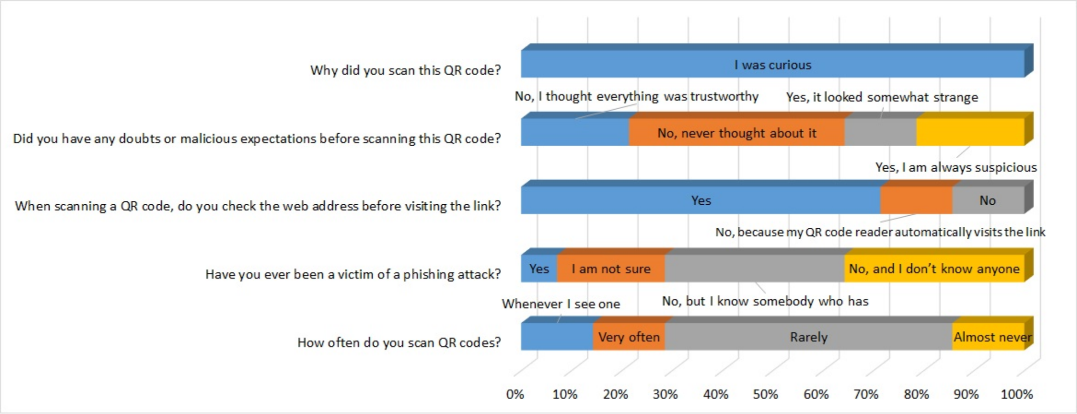


Figure 5.17: Survey responses in Athens

QR Code. In the question "Have you ever been a victim of a phishing attack?" the majority of the users reported that they were not affected from this kind of attack. Nevertheless, the 7% that had been victims of phishing attack along with the 21% that might have been victims without realizing it, indicates that the phishing is an actual threat for the users also in Greece. The answers on the question about the scanning frequency of QR Codes are similar to those from Vienna and Helsinki and yields the same conclusion that QR Codes are not very popular and they are not widely used by the smartphone users in Greece.

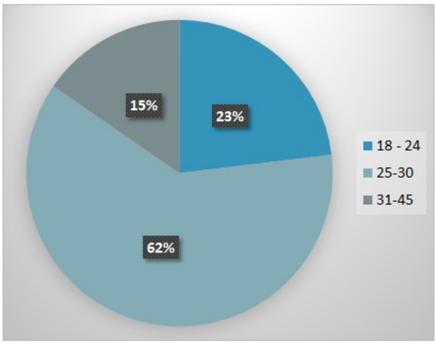


Figure 5.18: Age distribution, Athens

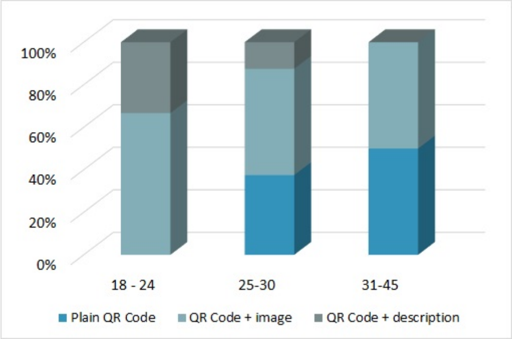


Figure 5.19: Stickers' preference depending on the age, Athens

The age distribution of our participants is presented in figure 5.18. In contrast to the age distribution recorded in Vienna and Helsinki, in Athens the majority (62%) of the participant were in the range 25-30. Nevertheless, this group along with the 23% of the participants that belongs to the age range 18-24, consist more than the 90% of our total participant and confirms our conclusion that younger users are more familiar with QR Codes and curious enough to explore where they lead. Additionally, as we can see in figure 5.19 stickers with QR codes accompanied by an image are mostly preferred by all the age groups. However, plain QR Codes are also make up an important share in the age groups 18-24 and 31-45. Finally the stickers including related information are the least attractive for all the age groups.

Figures 5.20 and 5.21 present respectively the hits gathered depending on the location and the different kind of sticker. Universities are again the most popular place for our stickers. Despite the fact that in Greek universities there are a lot of posters and our stickers can be easily removed, covered, or just ignored by the students, they managed to score a good performance and gather most visits. Stickers in transportation means and toilets gathered almost the same hits. Taking into account the "chaotic" situation in the center of Athens and the fast paced life that citizen of Athens follow, we were expecting that our stickers would not be noticed that much. Additionally the total number of visits is not that high and this indicates that Greek users are not so familiar with QR Codes. It is true that QR Codes are not yet widely used in Greece, at least not as much as in other European countries. Concerning the sticker performance, once more the stickers with including an image were those preferred the most by the users.

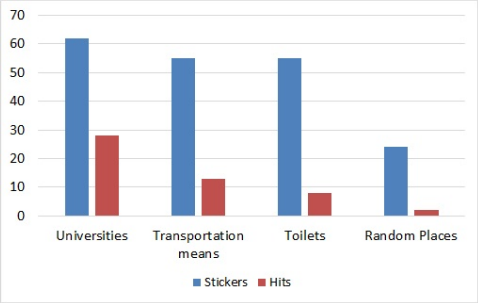


Figure 5.20: Stickers' performance by location, Athens

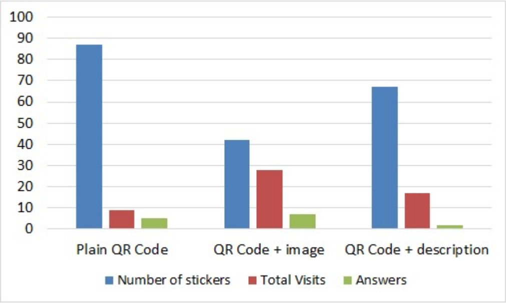


Figure 5.21: Stickers' preference by type of stickers, Athens

Finally, we tracked the operating system of users' devices, that were used to visit our on-line survey. Similarly to Vienna and Helsinki Android was the operating system running in the majority of the mobile devices. More specifically, out of the 53 visits that we had in total, 37 were from users possessing an Android device. Users with Apple iOS operating system were the next biggest group with 19% of the total visits. The remaining 11% of visit was made by users that had either a Windows mobile smartphones or some other operating system.

5.2.4 Paris

The fourth and last city where we conducted our survey is Paris. Apart from the fact that Paris is one of the biggest cities in Europe, according to the Google statistics [19], Paris is a place where QR Codes are quite popular. The total number of visits registered from Paris was 75. These visits were generated by 30 different QR Code stickers. Out of these visits we collected 10 answers. This result is not exactly what we were expecting; however we realized that especially the stickers posted in the bus stop were removed in a very short time period. It is remarkable that only 13% of our visitors from Paris participated in our survey. The corresponding percentage in the other cities is at least the double. This is probably related to the culture that French people have which includes a certain unwillingness to participate in an unknown survey. Additionally the "chaotic" situation in Paris due to its size did not allow us to focus on specific locations where we probably could get better results. In our first question about the reason that the users scanned our stickers, again curiosity is the main motivation. However, some of the free text answers reveal that users were especially attracted by the images that we used. For example one user reported as his motivation to scan the stickers "The photo of the cat that yawns". The unwillingness that we mentioned before from French people can be correlated also to the suspiciousness that they have. Our second question supprots that, since more than 80% of our participants reported to have doubts before scanning our QR Code sticker.

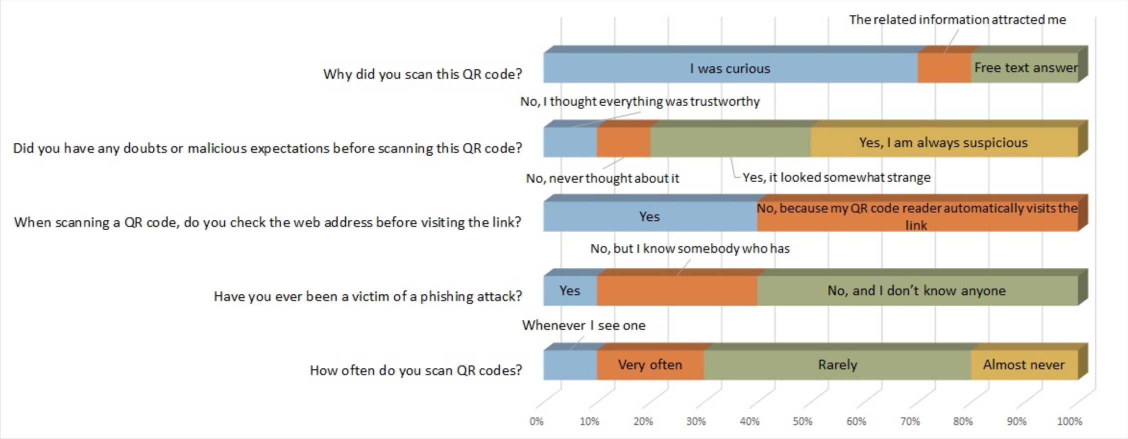


Figure 5.22: Survey responses in Paris

Despite being suspicious, the majority (60%) of the survey participants do the mistake of using QR Code software readers that automatically visit the link without asking prior asking the user. This particular data and in comparison to the answers from the other cities leads us also to the conclusion that French users tend to trust more the automated procedure. We notice that the answers collected by the users are controversial since the same users reported to be suspicious and in the same time they use automated readers to scan the QR Codes. This reveals that users have a significant lack of knowledge on the security threats and on how they can be protected. Moreover according to our participants, how most of them reported not to be affected by such an attack, it seems that phishing attacks are not a serious threat for French users. However due to the small sample this is not a very safe conclusion. The answers in our last question verify once more the conclusion made before based on the answers from the other cities, that QR Codes are not very popular among the users and their not used very often by them.

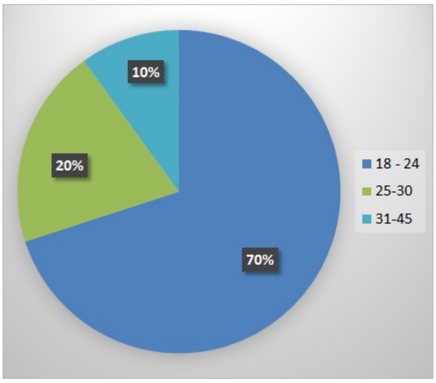


Figure 5.23: Age distribution, Paris

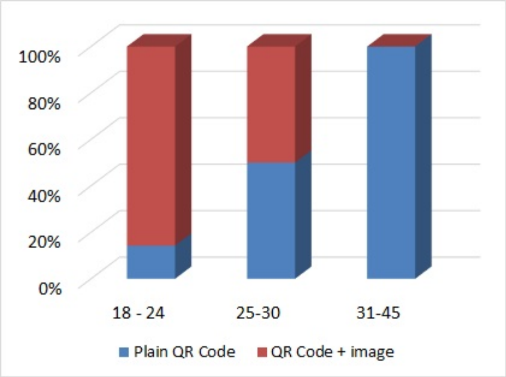


Figure 5.24: Stickers' preference depending on the age, Paris

Figure 5.23 presents the age distribution of the participants in our survey. As in the Vienna and Helsinki the majority (70%) of the users belongs in the range 18-24. The remaining 30% is covered mostly by the age range 25-30 which holds the 20%. The age distribution meets our expectations, that young users are attracted more than the older users by the QR Codes. It is usual that the new technologies are adapted faster by younger users. Additionally, figure 5.24 shows which stickers these age groups prefer. As a general conclusion we can say that younger users are more attracted by the stickers including a QR Code and an image while the age group 31-45 prefers plain QR Codes.

As in the previous cities we present in figures 5.25 and 5.26 the performance of each group of stickers depending on the location and the performance of each type of stickers independently of locations. The first comment on the stickers' performance is that our stickers located in the transportation means had a very poor score. As we explained before, we identified our stickers in these locations get removed in a very short time. The stickers deployed in the universities and more specifically those deployed in the "University Pierre and Marie Curie" outperformed all the other locations. Stickers located in the toilets managed to collect a large portion of the total visits.

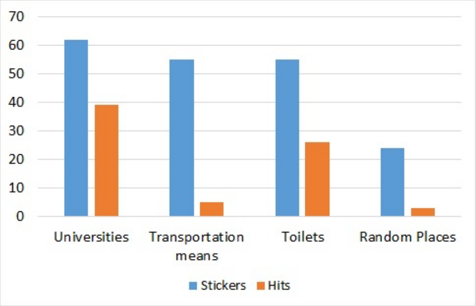


Figure 5.25: Stickers' performance by location, Paris

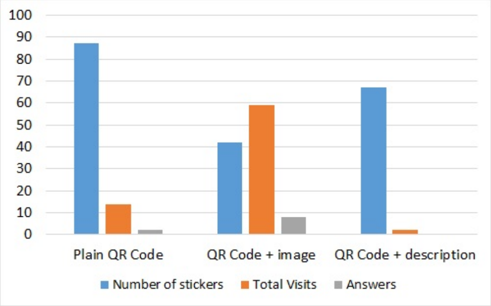


Figure 5.26: Stickers' preference by type of stickers, Paris

Finally, in figure 5.26, we can clearly see that stickers with QR Code accompanied by an image collected the most of the visits in total indicating that users in general are more attracted by them. However, the total number of answers that we collected is low so we cannot make any safe conclusion concerning the sticker preference of our survey participants.

As we did for all the other cities of our study, we also stored for Paris the operating systems that users' devices were running. Android was again the operating system that most of the users had on their devices. More specifically 39 of the 75 visitors were using an Android device. The second largest group of users was those using an Apple iOS device covering the 35% of the total visits. This is translated to 26 visitors with Apple iOS devices. The remaining 12% of our visitors were using either Windows Phone or Blackberry devices.

5.2.5 Limitations

In Chapter 4, we already mentioned that our initial goal was to approach as much as possible a real attack scenario. However, we are bound to ethical and legal limitations that restrain us from deploying our study identically to a real attack scenario. For example we could have posted our stickers over legitimate QR Codes or we could have included a fake advertisement on our stickers. False pretenses such as an ideal work opportunity or a discount in a local store could make our stickers much more attractive for the users. Nevertheless, we did not want to trick users in an immoral way just to get more participants in our study. Moreover, we had limited time to conduct our study and in some cases our resources needed to deploy our stickers were also limited. Every sticker that was either destroyed by a human or by the weather condition was not replaced with a new one. All these limitations along with the fact that our study was targeted to specific locations and as a result to the respective populations, reflect to the results of our study.

Chapter 6 Discussion

In this Chapter we first discuss the security implications related to QR Codes that our study revealed and in this way answer our first research question. The discussion that follows on the issues that our research revealed help us to answer the rest of the research questions phrased in the *Introduction* based on the obtained results.

6.1 What are the security issues related to QR Codes?

The first part of this thesis shows how a QR Code can be easily altered and as a result be used as a tool to perform a phishing attack. The main security implication that we have to focus on is the case where a QR Code is used as a medium for payments. We discussed earlier in section 2.2 how Paypal has built online stores based on QR Codes. Especially payments via Paypal usually only require to enter the username of the recipient of the payment. A potential attack could involves an implementation similar to ours, where the attacker changes the QR Code in a way that the name of the recipient is altered. Using a username very close to the legitimate one would make it easier for the attacker to redirect the payments to his account without the customer even noticing. Similarly, the system of payments that Erste Bank [14] has introduced shows some vulnerabilities to the same kind of attacks. More specifically, when the system “Scan and Pay” is used in public campaigns such as charity campaigns or some contests where poster are used to advertise the event, a phishing attack can be deployed. A malicious attacker could change the QR Codes located on the posters, and in the same way as we described before, redirect the donations to his account. This system was proposed by STUZZA [56], which is an existing collaboration platform between the largest Austrian banks, is on its way to become a European standard. However, we strongly believe that such security implications have to be taken into account before standardization.

Security implications also arise while analyzing the results of our survey. As we presented before in section 5.2, most of the users in all the cities were using

Android devices while users with Apple iOS devices were the second largest group. In all the four cities that we conducted our survey, 57% (157) of the participants had an Android device and another 31% (85) of the participants had an Apple iOS device. Another 8% (22) of the participants were holding Windows Phone devices (the remaining 4% was a mix of Blackberry and Symbian devices). Both for Android and iOS devices, the majority of the browsers were WebKit [4] based. At the time that we conducted our research, there were known vulnerabilities and public exploits that target the mobile browsers or content handlers. In [3] and [5] are listed some of the well-known exploits that researchers have discovered, related to the Webkit browsers. These vulnerabilities can be exploited by an attacker that "tricks" a user to visit a malicious web page. QR code are a very convenient tool in this case, as they can be the medium to attract the users. When a user visits such a webpage, his browser will execute the malicious content located in the webpage and based on the vulnerability that the browser has will manage to complete successfully the attack. A successful attack may result in the attacker having access to the resources that the browse also has. There are many types of attacks that can be deployed in this way such as cookies stealing, session hijacking or even Cross Site Scripting (XSS) [37] which may result in the attacker having control of the device. However, we have to mention that most of the users having Android (85) devices were using the latest version of the mobile browser, which is secured against many of these attacks. For users having Apple iOS devices we noticed a wide variety of web browser versions and only 19 of the users had the latest version of Safari. Another attack scenario is based on rooting exploits. In this case, the attack is targeting exploits related to the operating system. In our study, we noticed 10 different versions of Android operation system and 10 different Aplle iOS versions. The majority of the Android users (45) were using version 4.1.2 which is not the latest version. Correspondingly 46 participants having Apple devices were using one of the latest versions 6.1.x. In the same time there were participants that had devices that used an operating system that is more than two years old and have major security issues. In [33] there is a short list of some of the known root exploits that affect many of the operating systems observed in our study.

6.2 Is it possible to attack a QR Code?

In this thesis we presented the three attack schemes that we used in order to test whether it is possible to attack a QR Code. As we discussed earlier in Chapter 4.1.1 attacking directly the modules of the QR Code by changing randomly the color is not feasible. However, our second approach which is attacking the binary representation of the encoded string shows that it is feasible to attack a QR Code and produce similar results as a brute-force attack would. The time needed for this kind of attack was significantly low and the results were quite satisfying since we managed to produced results (see section 4.2.2) where only a few modules have to be changed in order to retrieve an altered content from the QR Code. Nevertheless, our third attack scheme was the most successful one. By attacking the codewords of a QR Code we showed that is not only possible to get an altered result but you are actually able to target specific areas that correspond to a certain part of the URL. In this way an attacker is able to deploy a more targeted attack and in the same time change the minimum number of module needed to successfully change the content.

6.3 Can QR Codes be used as attack vectors, especially in phishing attacks?

Our empirical study was designed and deployed in a way to determine the feasibility of a phishing attack using QR Codes. Our results suggest that such an attack is feasible and most likely to be successfully performed. Only the fact that 273 users scanned our QR Code stickers and visited our web page could be a proof of such an attack scenario. All these users scanned a random QR Code which in most cases was giving no information about its content, and after that visited an unknown and obscured domain. Our participants in other circumstances could be victims of a phishing attack deployed using QR Codes. It is uncertain how successful such a phishing attack would be in terms of how many users would proceed far enough so that the attacker could steal valuable information. However, it is certain that enough users would visit the phishing web page while the success of the attack depends on how "professional" the phishing web site is. If we consider other techniques that an attacker can follow, such as covering existing QR Codes or using fake advertising, the effectiveness and the success of a phishing attack can improved significantly. Covering a QR Code with a new one is a more neat way of deploy the same attack as we did and with less effort. However, someone can argue that a user may realize that there is a sticker over the legitimate QR Code and avoid scanning it. In any case these attack techniques can be the subject of a future research.

6.4 What is the users' security level of awareness against threats related to QR Codes?

To determine the users' level of security awareness we rely on the answers retrieved from the online survey and users' general behavior noticed. In all the cities that we conducted our empirical study most of the users reported to know what a phishing attack is. However, that does not mean that they know how to protect themselves. Actually some of them, despite having already been victims of phishing attacks, they did not gain valuable knowledge from this incident on how to avoid such threats. Moreover, the majority of the users reported being suspicious before visiting the link from the QR Code and also checking the URL before visiting it. Nevertheless, they finally visit an unknown and potentially malicious web page. These facts lead us to the conclusion that users may have a theoretical understanding of what phishing attack is and that in general they could be target of such an attack. However, they are unable to apply their knowledge practically and protect themselves. It is also possible that even though they are aware of possible threats, they do not have the appropriate skills to tackle them. The lack of technical means such as effective mobile security software that provides the appropriate security indications, makes this task even more difficult. Moreover, the trust that users have to the automation is a serious obstacle in building security awareness. We are led to this conclusion from the fact that a large group of users, uses a QR Code reader which automatically visits the link without showing the URL.

6.5 What countermeasures can be taken against QR Code security issues?

Along with the security implications that we presented, we also consider that, for mobile devices, there are not available effective technical controls, such as anti-virus, anti-spyware and firewall software, which are available in desktop computers. In the last years quite many mobile security software products have been published, but there are serious doubts about their effectiveness [48], [52], [42]. This leads the mobile users to a more unsafe environment since even simple indications that help users to make a security-conscious decision are not available.

In order to tackle these security issues we propose some countermeasures that can be applied. The first fundamental security measure is to introduce a security framework that all the QR Code readers should follow. Manually checking the URL before visiting the link, is the simplest but also the most basic security check that a user can perform. As mentioned before, having a QR Code software that allows to check the URL before visiting gives to the user the ability to verify that the URL that he/she is about to visit is the same as the one advertised. In non-contextual QR Codes this security check is unfeasible. Thus, all the QR Code readers, before visiting the link, should show to the user the decoded URL and ask him if he/she want to visit the link. Moreover, some of the security indications that exist in the browsers can be embedded in the QR Code reader. For example, when a user scans a QR Code which leads him/her to a web page that has no valid certificate or attempts to establish an insecure connection, the reader should notify the user. Additionally, we suggest the adoption of security countermeasures that ensure that a QR Code contains the original content. As we showed in this thesis, QR Codes can link to phishing sites or to web pages that the user had no intention to visit. A possible way to face this security issue, as proposed in [8], is to include a digital signature within the QR Code. This digital signature created by the publisher of the QR Code, which accompanies the content encoded, could verify the integrity of the encoded information. To accomplish that, the QR Code reader should have access to a verified database where it can match the digital signature with the original publisher. If the signature is not matching with any of the known publishers then the QR Code reader could show a security notification to the user and leave him the final decision whether to visit the link or to quit. This solution might complicate the way that QR Codes work and will also add an overhead since the digital signature will consume a part of the QR Code's capacity. Centralized solutions are difficult to scale up since many vectors have to agree on the proposed solution. Moreover, as we already know, the fact that a web site is certified, with an SSL for example, does not necessary means that they are completely safe. However, such a solution would secure users from many malicious attacks without adding significant complexity to the user's side.

Finally as mentioned earlier, educating the users is the most effective but also the most difficult way to tackle the security issues. First of all, users should be educated so that they always check the URL before visiting the link. Realizing that scanning a random QR Code and visiting the corresponding URL is exactly the same as visiting an obscure and unknown domain is the fundamental objective. Users should be educated to check that the decoded URL is the same as the one that the related poster advertise to be. If the QR Code is not accompanied by any information about where it links, then the user should be extra cautious and aware that proceeding might mean visiting a malicious web page. Nevertheless, educating users is always an expensive and time consuming process. Users tend to trust more the automated features and are unwilling to spend their time in acquiring proper technical education. Moreover, as mentioned earlier, verifying that a URL is malicious just by looking at it is a rather difficult task. Thus, the users' education should be focused on avoiding visiting obscure URLs and not giving away sensitive and personal information to unverified sources.

Chapter 7
Conclusion

QR Code are an upcoming and fast-growing technology that is used in many different fields. QR Codes offer a range of benefits that stakeholders from different fields are exploring and adopting to fulfill their requirements. Even though advertising is the area where QR Codes mostly used, new services like payment using QR Codes have been introduced over the last years. Along with this development security concerns arise. This thesis examines a specific security issue, which is attacking QR Codes in different ways, and in the same time explores users' level of security awareness concerning the security issues related to QR Codes. Our empirical study is mainly focused on phishing attacks where QR Codes are used as attack vectors.

Our work on attacking QR Codes explores a security issue that has not been within the scope of recent research projects and thus open new aspects in this area. We developed an implementation that attempts brute-forcing a QR Code. However, our results showed us that this task is not feasible in a short period. The execution time of our implementation indicates us that such an attack cannot be deployed in a real attack scenario. However, our second approach which targets on the binary representation of the encoded string was successful. We managed to produce the same result as a successful brute-forcing implementation would have and we showed that it is possible to alternate a QR Code so that it will lead to a new URL. This new URL may lead to a phishing web site and the QR Code would be in this case the attack vector. The time and the resources needed for such an attack are very limited. Even better results were obtained by following the third attack scheme that we introduce. More targeted results and the minimum number of modules changes are achieved when attacking the codewords of a QR Code.

To further explore possible uses of QR Codes involving malicious scenarios like a phishing attack, we deployed our empirical study. Our on-line survey was addressed to the users of four different European cities, allowing us in this way to work in an intercultural environment. Our research that is comparable to a social engineering experiment, achieved at least on a certain level its goal, which was to identify users' level of security awareness concerning the security issues related to QR Codes. We managed to attract 273 visitors through the QR Code stickers that we deployed in different places. Only by this result, we conclude that a potential phishing attack will be successful. In the same time we saw that most of the users are curious to explore where a strange QR Code leads, forgetting that this behavior may harm them. Not even the suspiciousness that the majority of the users reported to have prevented them from visiting an unknown and obscure URL. As our results indicate, phishing attacks are not very widespread in this domain. However, they affect a small part of the users with unexpected cost for them. The differences in the behavior between the users from the different cities that we conducted our survey were not highly significant. Nevertheless there are some behavior patterns such the increased suspiciousness of the Finnish users or the trust to automation from the French users that can be derived from our results.

Finally we propose a set of security countermeasures that can be adopted in order to tackle the security issues addressed in our research. Automated procedures along with proposals for proper user education are included in our set of countermeasures. These countermeasures have not been examined in depth yet, but they are able to form a solid base where further research can be deployed.

References

[1] *Survey Research Methods*, 2nd ed. Wadsworth Publishing, 1990.

[2] *Trustworthy Ubiquitous Computing, Volume 6 of Atlantis Ambient and Pervasive Intelligence Series*, illustrated ed. Springer, 2012, pp. 21–38.

[3] Apple Webkit Security Vulnerabilities, 2013. http://www.cvedetails.com/vulnerability-list/vendor_id-49/product_id-10007/cvssscoremin-5/cvssscoremax-5.99/Apple-Webkit.html. Accessed 27 May 2013.

[4] The WebKit Open Source Project, 2013. http://www.webkit.org/. Accessed 27 May 2013.

[5] Vulnerability for WebKit in Google Chrome, 2013. http://www.cvedetails.com/cve/CVE-2013-0912/. Accessed 27 May 2013.

[6] AVG. AVG (AU/NZ) Cautions: Beware of Malicious QR Codes, 2011. http://www.pcworld.idg.com.au/mediareleases/12655/avg-aunz-cautions-beware-of-malicious-qr-codes/. Accessed 29 Mar 2013.

[7] C. Dow, Y. Lee, H. Yang, W. Koo, J. Liao . A location-based mobile advertisement publishing system for vendors. In *Eighth International Conference on Information Technology: New Generations* (2011), pp. 24–29.

[8] 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. 21–29.

[9] D. Pirrone, S. Andolina, A. Santangelo, A. Gentile, M. Takizava. Platforms for human-human interaction in large social events. In *Seventh International Conference on Broadband, Wireless Computing, Communication and Applications* (2012), pp. 545–551.

[10] David Moth. PayPal trials QR code shop in Singapore subway, 2012. http://econsultancy.com/at/blog/8983-paypal-trials-qr-code-shop-in-singapore-subway. Accessed 10 Mar 2013.

[11] Dell SonicWALL. SonicWALL Phishing IQ Test - Phishing Facts, 2013. http://www.sonicwall.com/furl/phishing/. Accessed 7 March 2013.

[12] DENSO Wave Incorporated. What is a QR Code?, 2013. http://www.qrcode.com/en/. Accessed 10 Feb 2013.

[13] eMarketer. US Ahead of Western Europe in QR Code Usage, 2013. http://www.emarketer.com/Article/US-Ahead-of-Western-Europe-QR-Code-Usage/1009631. Accessed 23 May 2013.

[14] Erste Bank. Erste Bank und Sparkassen: Surprising Customers with Innovation, 2013. http://www.erstegroup.com/en/Press/Press-Releases/Archive/2013/4/15/Surprising-Customers-with-Innovation~Produkte;GPJSESSIONID=Y7vwRvxTLNBRWMSNscXT1xkkSyFQMvmDcGBjvV4pqPCHHfT6G25n!837641963. Accessed 25 May 2013.

[15] Esponse. Innovative QR Code campaigns (About QR codes), 2013. http://www.esponce.com/about-qr-codes. Accessed 23 Mar 2013.

[16] Esponse. Innovative QR Code campaigns (Real world case studies), 2013. http://www.esponce.com/case-studies. Accessed 23 Mar 2013.

[17] Gartner. Gartner Survey Shows Phishing Attacks Escalated in 2007; More than $3 Billion Lost to These Attacks, 2013. http://www.gartner.com/newsroom/id/565125. Accessed 24 Mar 2013.

[18] Gia M. Agusta, K. Hulliyah, Arini, R. B. Bahaweres. QR code augmented reality tracking with merging on conventional marker based backpropagation neural network. In *International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (2012), pp. 245–248.

[19] Google. QR Code interest, 2013. http://www.google.com/trends/explore#geo=FR-J&q=qr+codes. Accessed 23 May 2013.

[20] GOQR.me, 2013. http://goqr.me/. Accessed 10 Apr 2013.

[21] Huffman, W., and Pless, V. *Fundamentals of Error-Correcting Codes.* Cambridge, Ma University Press, 2003.

[22] I. Reed and G. Solomon. Polynomial Codes Over Certain Finite Fields. *Journal of the Society for Industrial and Applied Mathematics 8*, 2 (1960), 300–304.

[23] J. Gao, V. Kulkarni, H. Ranavat, Lee Chang Hsing Mei. A 2D barcode-based mobile payment system. In *Third International Conference on Multimedia and Ubiquitous Engineering* (2009), pp. 320–329.

[24] J. Rouillard, M. Laroussi. Perzoovasive: contextual pervasive QR codes as tool to provide an adaptive learning support. In *Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology, CSTST '08* (2008), pp. 542–548.

[25] J. S. Downs, M. B. Holbrook, L. F. Cranor. Decision strategies and susceptibility to phishing. In *Symposium On Usable Privacy and Security (SOUPS)* (2006), pp. 79–90.

[26] J. Wang, C. Shyi, T.-W. Hou, C.P. Fong. Design and implementation of augmented reality system collaborating with QR code. In *International Computer Symposium (ICS)* (2010), pp. 414–418.

[27] Jan Seeburger. No cure for curiosity:linking physical and digital urban layers. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (2012), pp. 247–256.

[28] Kevin Joy. The War on Phishing is far from over, 2009. http://info.brandprotect.com/Blog/bid/15103/The-War-on-Phishing-is-far-from-over. Accessed 24 Mar 2013.

[29] Lukyanenko, A. Methodology for Computer Science Research, 2012. https://noppa.aalto.fi/noppa/kurssi/t-110.6130/luennot/T-110_6130_lecture_5_slides.pdf. Accessed 10 Apr 2013.

[30] M. Ebling, R. Cres. Bar Codes Everywhere You Look. *PERVASIVE computing, IEEE 9*, 2 (2010), 4–5.

[31] M. R. Rieback, B. Crispo, A. S. Tanenbaum . Is your cat infected with a computer virus? In *Proceedings of the Fourth Annual IEEE International Conference on Pervasive Computing and Communications, PERCOM '06* (2007), pp. 169–179.

[32] Matthew Talbot. QR Codes: Scanning For Loyalty And Payment, 2013. SAP blog: http://blogs.sap.com/innovation/industries/qr-codes-scanning-for-loyalty-and-payment-3-025064. Accessed 26 Mar 2013.

[33] metall0id. Root exploits, 2013. https://github.com/mwrlabs/mercury/issues/56. Accessed 27 May 2013.

[34] Michael Oleaga. iOS vs. Android Market Share, 2013. http://www.latinospost.com/articles/19393/20130517/ios-vs-android-market-share-apple-google-mobile-operating-systems.htm. Accessed 22 May 2013.

[35] New Media Trend Watch. Mobile Devices, 2013. http://www.newmediatrendwatch.com/regional-overview/103-europe?start=4. Accessed 16 May 2013.

[36] OWASP. Cross-Site Request Forgery (CSRF), 2013. https://www.owasp.org/index.php/Cross-Site_Request_Forgery_(CSRF). Accessed 27 May 2013.

[37] OWASP. Cross-site Scripting (XSS), 2013. https://www.owasp.org/index.php/Cross-site_Scripting_(XSS). Accessed 27 May 2013.

[38] 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. 430–435.

[39] P. KUMARAGURU, S. SHENG, A. ACQUISTI, L. F. CRANOR, J. HONG. Teaching Johnny Not to Fall for Phish. *ACM Transactions on Internet Technology (TOIT) 10*, 2 (2010).

[40] Pankaj Kumar. Generate QR Code image from Java Program, 2013. http://www.journaldev.com/470/generate-qr-code-image-from-java-program. Accessed 10 Apr 2013.

[41] Paul Wagenseil. Anti-Anonymous hacker threatens to expose them, 2012. http://www.nbcnews.com/id/46716942/ns/technology_and_science-security/#.UWV6l7VkPL-. Accessed 27 Mar 2013.

[42] Paul Wagenseil. Study Faulting Anti-Virus Effectiveness May Itself Be Flawed, 2013. http://www.technewsdaily.com/16177-imperva-malware-study-flaws.html. Accessed 27 May 2013.

[43] QR Pay Limited. QRpay, 2013. http://qrpay.com/. Accessed 26 Mar 2013.

[44] QRStuff. QR Code Error Correction, 2011. QRStuff blog: http://www.qrstuff.com/blog/2011/12/14/qr-code-error-correction. Accessed 25 Jan 2013.

[45] QRStuff. What's a QR Code?, 2011. http://www.qrstuff.com/qr_codes.html. Accessed 25 Jan 2013.

[46] R. Dhamija, J. D. Tygar, M. Hearst. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06* (2006), pp. 581–590.

[47] R. Dhamija, J.D.Tygar. The battle against phishing:dynamic security skins. In *Symposium on Usable Privacy and Security (SOUPS)* (2005), pp. 77–88.

[48] R. Ramachandran, Tae Oh, W. Stackpole. Android Anti-Virus Analysis. In *ANNUAL SYMPOSIUM ON INFORMATION ASSURANCE SECURE KNOWLEDGE MANAGEMENT, ASIA SKM 12* (2012), pp. 35–40.

[49] Russ Cox. QArt Codes, 2012. http://research.swtch.com/qart. Accessed 05 Mar 2013.

[50] S. Egelman, L. F. Cranor, J. Hong. You's been warned: An empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08* (2008), pp. 1065–1074.

[51] S. Sheng, B. Magnien, P. Kumaraguru,A. Acquisti, L. F. Cranor, J. Hong, E. Nunge. Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish. In *Symposium On Usable Privacy and Security, SOUPS 07* (2007), pp. 88–99.

[52] Seth Rosenblatt. Android antivirus apps improve their grades–just not very much, 2012. http://download.cnet.com/8301-2007_4-57398501-12/android-antivirus-apps-improve-their-grades-just-not-very-much/. Accessed 27 May 2013.

[53] SourceForge.jp. Open Source QR Code Library, 2007. http://sourceforge.jp/projects/qrcode/. Accessed 20 Jan 2013.

[54] Sourceforge.net. pyqrcode, 2007. http://pyqrcode.sourceforge.net/. Accessed 25 Jan 2013.

[55] Steeman, J. QR code data capacity, 2004. QR4 QR Codes blog: http://blog.qr4.nl/page/QR-Code-Data-Capacity.aspx. Accessed 3 Feb 2013.

[56] STUZZA. QR Code payment, 2013. http://www.stuzza.at/11252_DE. Accessed 25 May 2013.

[57] T. Moore, B. Edelman. Measuring the perpetrators and funders of typosquatting. In *Proceedings of the 14th international conference on Financial Cryptography and Data Security, FC'10* (2010), pp. 175–191.

[58] T. Vidas, E. Owusu, S. Wang, C. Zeng, L. Cranor. QRishing: The susceptibility of smartphone users to QR code phishing attacks. In *CMU-CyLab-12* (2012), pp. 1–12.

[59] The Eclipse Foundation. Eclipse, 2013. http://www.eclipse.org/. Accessed 8 May 2013.

[60] The University of Texas as Austin, School of Information. Survey Methods, 2013. https://www.ischool.utexas.edu/~palmquis/courses/survey.html. Accessed 10 Apr 2013.

[61] Thonky.com. QR Code Tutorial, 2012. http://www.thonky.com/qr-code-tutorial/. Accessed 10 Feb 2013.

[62] Ugo B. Ceipidor, Carlo M. Medaglia, A. Perrone, M. De Marsico, G. Di Romano. A museum mobile game for children using QR-codes. In *Proceedings of the 8th International Conference on Interaction Design and Children, IDC '09* (2009), pp. 282–283.

[63] Wikipedia. Decoding beyond the error-correction bound, 2013. http://en.wikipedia.org/wiki/Reed%E2%80%93Solomon_code#Decoding_beyond_the_error-correction_bound. Accessed 15 June 2013.

[64] Wikipedia. QRCode-3-Layout, 2013. http://en.wikipedia.org/wiki/File:QRCode-3-Layout,Encoding.png. Accessed 31 May 2013.

[65] Y. Kao, G. Luo, H. Lin, Y. Huang, S. Yuani. Physical access control based on QR code. In *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery* (2011), pp. 285–288.

[66] Y. Zhang, J. Hong, L. Cranor. Cantina: A content-based approach to detecting phishing web sites. In *Proceedings of the 16th international conference on World Wide Web, WWW '07* (2007), pp. 639–648.

[67] Zeljka Zors. Malicious QR codes pop up on traffic-heavy locations, 2012. http://www.net-security.org/secworld.php?id=14099. Accessed 29 Mar 2013.

[68] ZXing. Malicious QR codes pop up on traffic-heavy locations, 2012. https://code.google.com/p/zxing/downloads/detail?name=ZXing-2.1.zip&can=2&q=label%3AType-Source. Accessed 29 Feb 2013.

[69] ZXing. QR Code Generator, 2013. http://zxing.appspot.com/generator/. Accessed 10 Apr 2013.

Chapter A
First appendix

Questionnaire

Why did you scan this QR code?
() I was curious
() The related information attracted me
() I was bored
() I don't know what a QR Code is
() I prefer not to answer
() Other:

Did you have any doubts or malicious expectations before scanning this QR code?
() No, I thought everything was trustworthy
() No, never thought about it
() Yes, it looked somewhat strange
() Yes, I am always suspicious

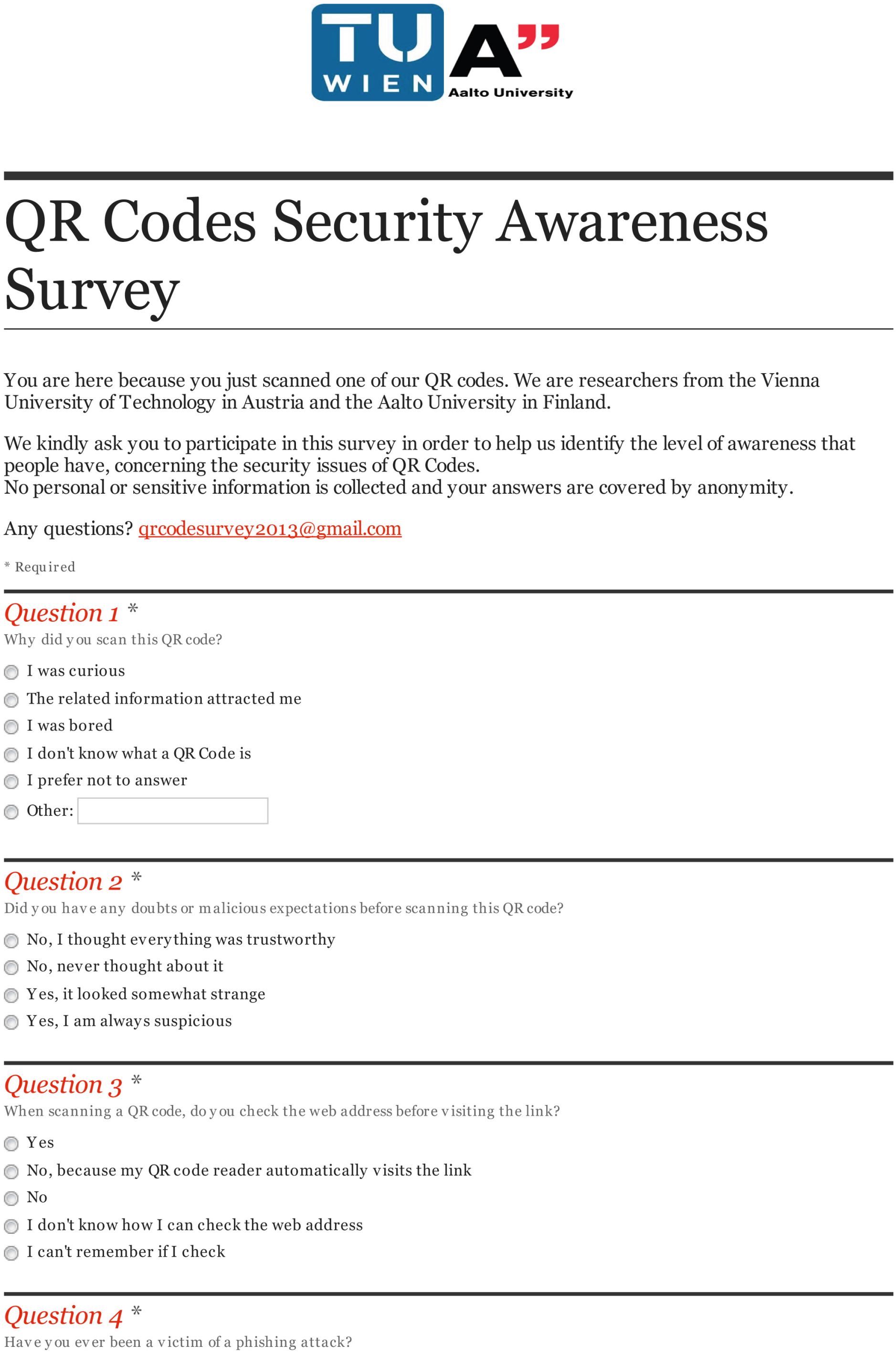
When scanning a QR code, do you check the web address before visiting the link?
() Yes
() No, because my QR code reader automatically visits the link
() No
() I don't know how I can check the web address
() I can't remember if I check

Have you ever been a victim of a phishing attack?
() Yes
() I am not sure
() No, but I know somebody who has
() No, and I don't know anyone
() I don't know what is a phishing attack
() I prefer not to answer

How often do you scan QR codes?
() Whenever I see one
() Very often
() Rarely
() Almost never

What is your gender?
() Male
() Female
() I prefer not to answer

What is your age?
() Under 18
() 18 - 24
() 25 - 30
() 31 - 45
() 46 - 60
() 61 and above
() I prefer not to answer



QR Codes Security Awareness Survey

You are here because you just scanned one of our QR codes. We are researchers from the Vienna University of Technology in Austria and the Aalto University in Finland.

We kindly ask you to participate in this survey in order to help us identify the level of awareness that people have, concerning the security issues of QR Codes.
No personal or sensitive information is collected and your answers are covered by anonymity.

Any questions? qrcodesurvey2013@gmail.com

* Required

Question 1 *

Why did you scan this QR code?

○ I was curious

○ The related information attracted me

○ I was bored

○ I don't know what a QR Code is

○ I prefer not to answer

○ Other: ______

Question 2 *

Did you have any doubts or malicious expectations before scanning this QR code?

○ No, I thought everything was trustworthy

○ No, never thought about it

○ Yes, it looked somewhat strange

○ Yes, I am always suspicious

Question 3 *

When scanning a QR code, do you check the web address before visiting the link?

○ Yes

○ No, because my QR code reader automatically visits the link

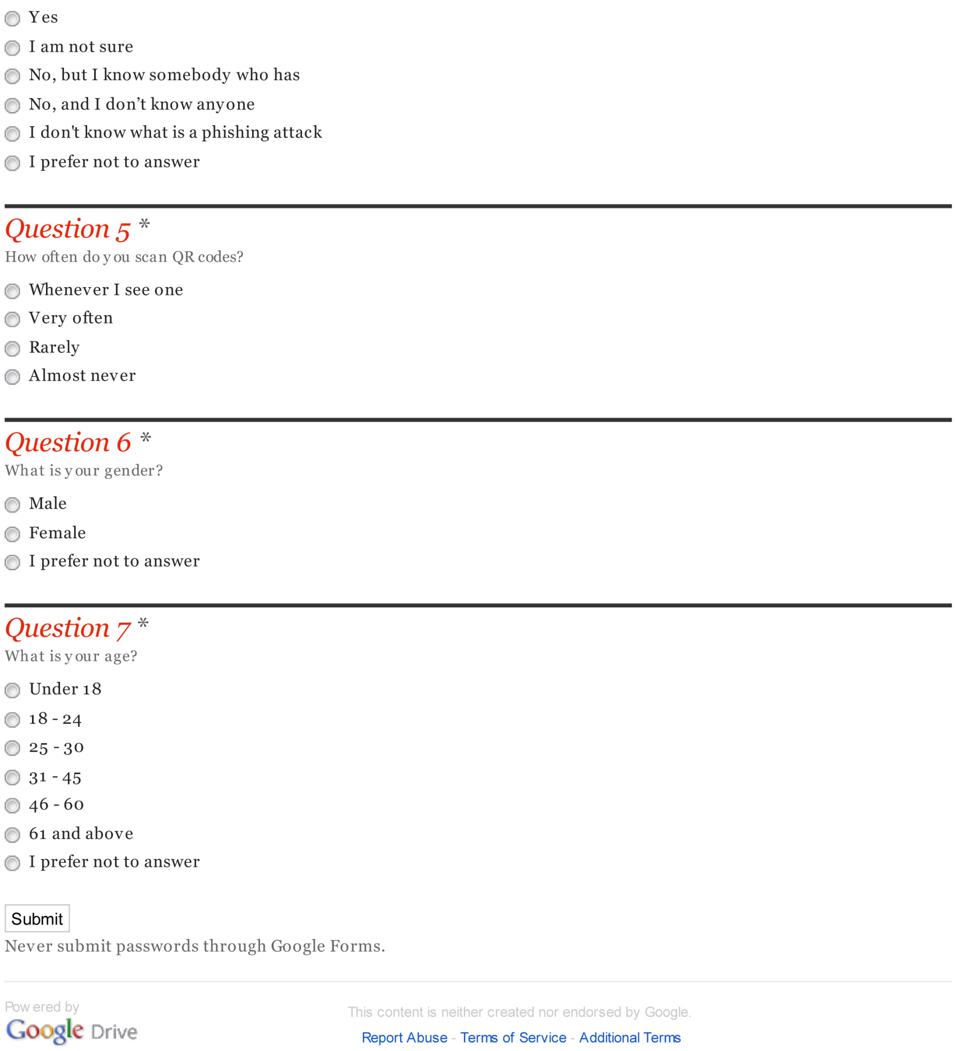
○ No

○ I don't know how I can check the web address

○ I can't remember if I check

Question 4 *

Have you ever been a victim of a phishing attack?



○ Yes
○ I am not sure
○ No, but I know somebody who has
○ No, and I don’t know anyone
○ I don't know what is a phishing attack
○ I prefer not to answer

Question 5 *

How often do you scan QR codes?

○ Whenever I see one
○ Very often
○ Rarely
○ Almost never

Question 6 *

What is your gender?

○ Male
○ Female
○ I prefer not to answer

Question 7 *

What is your age?

○ Under 18
○ 18 - 24
○ 25 - 30
○ 31 - 45
○ 46 - 60
○ 61 and above
○ I prefer not to answer

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google.

Figure A.1: On-line Survey in English

```
<script type="text/javascript">

  var _gaq = _gaq || [];
  _gaq.push([’_setAccount’, ’UA-38729606-1’]);
  _gaq.push([’_setDomainName’, ’ifs.tuwien.ac.at’]);
  _gaq.push([’_setAllowLinker’, true]);
  _gaq.push([’_trackPageview’]);

  (function() {
    var ga = document.createElement(’script’); ga.type = ’text/javascript’;
    ga.async = true; ga.src = (’https:’ == document.location.protocol ?
    ’https://ssl’ : ’http://www’) + ’.google-analytics.com/ga.js’;
    var s = document.getElementsByTagName(’script’)[0];
    s.parentNode.insertBefore(ga, s);
  })();

</script>
<html>
<head>
<title>QR Code Survey</title>
</head>
<body>
<p><center><img src="../tu.jpg" alt="TU_Wien" width="110" height="80"<center>
<img src="../aalto_big.jpg" alt="Aalto" width="110" height="70"></center></p>

<center><iframe src="https://docs.google.com/forms/d/
1mhi2X7_gnt6a_XAkqn_H0wJE3LBXUsEoSUphbhrJ8WY/viewform?embedded=true"
width="760" height="1870" frameborder="0" marginheight="0" marginwidth="0">
Loading...</iframe></center>

<?php
        $timestamp = strftime(’%c’);

        $file = ’/var/www/qrsurvey/logs/logs’;
        file_put_contents($file, "redirected,", FILE_APPEND | LOCK_EX);
        foreach($_GET as $key => $value)
        {
          file_put_contents($file, $value, FILE_APPEND | LOCK_EX);
            file_put_contents($file, ",", FILE_APPEND | LOCK_EX);
        }
        file_put_contents($file, $timestamp, FILE_APPEND | LOCK_EX);
        file_put_contents($file, "\n", FILE_APPEND | LOCK_EX);
?>
</body>
</html>
```

File: index.php