QRisk: Think Before You Scan QR codes

Abhishek Kumar Mishra[1], Guillaume Gagnon[2], Mathieu Cunche[1], and Sebastien Gambs[2]

[1] INSA-Lyon, Inria, University of Lyon, CITI Lab., France
{abhishek.mishra, mathieu.cunche}@insa-lyon.fr
[2] Université du Québec à Montréal, Canada
{gagnon.guillaume.5@courrier.uqam.ca, gambs.sebastien@uqam.ca}

Abstract. QR codes are pervasive in modern digital interactions, but despite their convenience, they pose significant privacy risks that are often underestimated. For instance, privacy issues escalate when scanned URLs trigger HTTP redirections involving QR URL shorteners and third-party domains, exposing user data to external entities. However, a comprehensive study of the privacy implications of QR code interactions concerning cookie exploitation and query strings remains lacking in the literature. To address this, we collected a dataset of 860 QR codes over a two-year period from France, China, Austria, India, and Canada, to analyze the privacy risks associated with QR code usage. We find in this paper titled `QRisk`, that 39.2% of redirected URLs set cookies, including tracking, analytics, and advertising cookies, enabling potential cross-session behavioral profiling. Additionally, over 25% of QR URLs embed query strings that not only contain sensitive user identifiers but also carry information such as location data in them, leading to user profiling and social link inference.

Keywords: QR codes, URL shorteners, Privacy, Cookies, URL strings

1 Introduction

Quick-Response (QR) codes have become an integral part of modern digital communication, allowing seamless access to digital information in the physical world [9]. For instance, they are commonly used in various domains, from marketing campaigns to payment systems, and play a pivotal role in bridging physical and digital spaces. In addition, QR codes are closely associated with user smartphones, as they are primarily accessed through mobile devices via the camera app's built-in reader or a dedicated scanning application. This association makes users potentially vulnerable to various privacy breaches.

Indeed, QR codes transform analog interactions, such as ordering at a restaurant, into digital ones, enabling data collection like order histories and contact information [5,21]. More broadly, interacting with QR codes creates *a digital trail from everyday movements* [1], activities, and consumption habits [13]. This has the potential to generate vast amounts of data on previously untracked aspects of our lives, often without explicit consent [14], thereby posing significant privacy threats.

From a more technical point of view, QR codes often contain Uniform Resource Locators (URLs) with query strings (key-value pairs) that enable server-side functionalities like referral tracking and analytics. Potentially, these query strings can carry sensitive information, such as user identifiers, location data, and other contextual information. Additionally, although a standalone QR code URL may appear harmless, its risks are magnified when scanned and subjected to a series of HTTP redirections. These chains frequently involve URL shorteners and potential third-party tracking domains, exposing user data to entities beyond their control. This ecosystem can be a fertile ground for tracking cookies and analytics scripts for gathering detailed user behavior profiles.

The security of QR codes has been studied previously and various attack strategies have been demonstrated [6]. In contrast, the privacy risks associated with specific URL access encoded in QR codes remain largely unexplored in the literature. Indeed, while some studies have explored the risks associated with URL query strings [20] or shorteners [11,12,9] separately, they have not comprehensively examined vulnerabilities specific to QR code interactions from a user-centric perspective. Moreover, they lack privacy threat investigation across diverse geographical locations with various internet regulations.

This paper titled `QRisk` precisely closes this gap by analyzing the privacy implications of query strings, cookies, and redirection chains in QR-scanned URLs across three continents including heavily censored [19] countries such as China. Our findings highlight significant risks, ranging from invasive tracking to the inadvertent exposure of personal information. In addition to unveiling these privacy risks, our objectives are to raise the awareness of stakeholders about the vulnerabilities inherent in QR code practices as well as to encourage users to adopt more secure practices.

In summary, the main contributions of our paper are:

- We perform an exploration of URL shortening services (USS) in QR code workflows, demonstrating their role in obfuscating URLs potentially enabling user profiling and tracking.
- We perform an in-depth examination of cookies set by third parties (USS) during URL redirection chains, categorizing them and identifying highly intrusive ones.
- We investigate query strings in QR-scanned URLs, revealing their widespread use for tracking metrics and location.
- We plan to release the first diverse and large public dataset of QR codes, providing a valuable resource for further research into their privacy and security challenges.

The outline of the paper is as follows. First, we introduce the related works in Section 2 and formulate our research questions. Next, we present the required background on QR codes in Section 3 before detailing the QR code dataset that we have collected in Section 4. Then, we investigate the redirections and USS in Section Section 5) and demonstrate that QR USS does deploy privacy intrusive cookies (Section 6). We further proceed to showcase the privacy implications of

URL paths and query strings (Section 7). Finally, we conclude by discussing how the revealed privacy risks could be mitigated and regulated in Section 8.

2 Related Work and Research questions

2.1 Related Work

The existing literature on the privacy and security challenges of QR codes and their associated workflows can be categorized into three types of studies that we briefly review hereafter.

QR Code Usage and Ecosystem Analysis. To understand the ecosystem of QR code usage, Lerner and collaborators have conducted a large-scale analysis of QR and barcode scans [9,8], uncovering general usage patterns from conventional Web links to emerging applications like Bitcoin wallets. While these studies show the growing ubiquity of QR codes, they have not analyzed the privacy threats posed by URLs and associated metadata.

Security Mechanisms in QR Code Scanners. Securing QR code scanning applications is emphasized by [7], who propose design recommendations to mitigate phishing attacks. SafeQR [22] leverages APIs like Google Safe Browsing and Phishtank for enhanced detection of malicious URLs. Similarly, [18] offer recommendations for creating applications that are secure, privacy-conscious, and user-friendly, demonstrating their approach with BarSec Droid, a prototype Android application adhering to these guidelines. While these works focus on malicious QR codes and scanner vulnerabilities, they do not address privacy risks related to sources like query strings, cookies, or redirection behaviors.

URL Query Strings and Shortening Services. Generic privacy risks of URL query strings and shortening services have been explored by [11,20,12]. Encoding of sensitive data, such as usernames and passwords, within query strings across user-submitted URLs, is highlighted by [20], who ultimately propose a prototype framework called CleanURL to sanitize URLs. Another work focuses on ad-based URL shortening services, and concludes the greater risks posed by it to users compared to traditional shortening services due to the involvement of third-party advertising networks [12]. Finally, [11] has investigated the vulnerabilities of URL shortening services (USS) claiming the leakage of URLs to search engines and also hint at USS setting long-term cookies, without digging deeper into the nature of these cookies.

All of these studies focus on generic URLs and do not delve into the specific in-depth privacy challenges (*i.e.*, intrusive cookies, URL strings exploited in QR URLs, issues with redirections) associated with QR codes and the services that provide shortened URLs for them. Thus, to the best of our knowledge, `QRisk` is the first work to investigate such critical aspects of QR code scanning.

2.2 Research questions

Based on the results of the literature review, in this paper, we aim to explore and address critical questions regarding the privacy implications of QR code usage, which we detail hereafter.

$\mathcal{Q}_1$ - *Information collected by QR URL shorteners without user consent.* URL shorteners play a central role in QR code workflows by deferring the final destination URL. This layer of re-direction raises significant privacy concerns:

- $\mathcal{Q}_{1.1}$ **- Are QR URL shorteners setting cookies?** We investigate if URL shorteners set cookies during the redirection process without consent to gain insights into their role in tracking and profiling users. (*cf.* Section 6.1)
- $\mathcal{Q}_{1.2}$ **- What types of cookies are being set?** Cookies vary in purpose, from functional and session-based to tracking and advertising. We identify and classify the types of cookies set by QR URL shorteners to reveal the extent to which they contribute to privacy risks. (*cf.* Section 6.2)

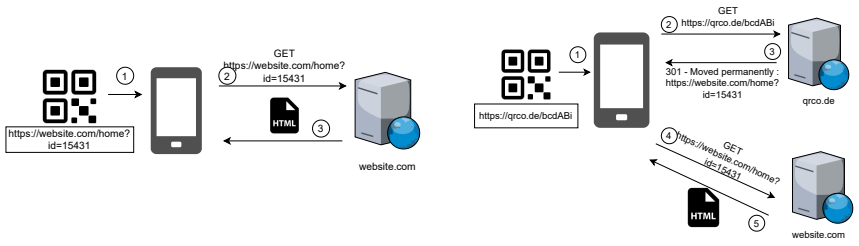
$\mathcal{Q}_2$ - *Information collected by the target party:* The end destination of a QR code scan—the landing page (server)—has multiple opportunities to collect user data. This aspect of the workflow leads to the following questions:

- $\mathcal{Q}_{2.1}$ **- What mechanisms are employed to transfer information?** Information can be transferred through cookies, URL paths, query strings as well as other techniques during the redirection chain. We investigate and understand these possible mechanisms for assessing the scale of sensitive data inference. (*cf.* Section 5)
- $\mathcal{Q}_{2.2}$ **- How prevalent are these mechanisms?** We quantify the frequency of these mechanisms to evaluate how common privacy-intrusive practices are in QR code workflows. (*cf.* Section 6 and 7)
- $\mathcal{Q}_{2.3}$ **- What types of data are being transferred?** Data such as user location, session identifiers, or demographic information can be encoded in query strings. We analyze the nature of this data and identify scenarios in which it can be exploited to compromise user privacy. (*cf.* Section 7)

3 Background on QR codes

Quick-Response (QR) codes are 2-dimension codes that can be read via the camera, typically from a smartphone. A QR code can encode various types of data, such as digital identifiers, emails, phone numbers, or URLs. There exist multiple *versions* of QR codes, corresponding to capacity ranging from 25 up to 4296 characters in version 40 [15].

URL shortening. QR codes have a limited capacity that cannot always accommodate the needs of the application. Hence, URL Shortening Services (USS) are regularly used to convert a long URL into a short one that can easily fit in a QR code. For instance, the service QR-Code-Generator (affiliated to `bit.ly`) uses the domain `qrco.de` for URL shortening purposes. Other USS, not related to QR code services, may also be used. QR code services are offered by a number of companies through both free and paid plans. In addition to the generation and customization, they usually offer additional services, including dedicated URL shortening, analytics and reporting as well as dynamic link management.



(a) Without URL shortener. (b) With URL shortener.

Fig. 1: Process of scanning a QR code to access a URL, with and without a URL shortener.

In the situation in which a URL is encoded, QR codes are used according to the following model described in Figure 1. First, a QR code is scanned by the smartphone of a user, the browser on the smartphone then opens the URL and retrieves the landing page. A URL shortening may be involved, in which case the browser will first contact the USS that will redirect the browser to the landing page. In practice, additional redirection may occur on the website as well as on the USS.

4 QR code dataset

Collection. We have collected the dataset from QR codes displayed in the wild in public places across France, China, Austria, India, and Canada over a two-year period (January 2023 – December 2024). It encompasses diverse real-world settings such as restaurants, bus stops, malls, public transport hubs, academic complexes and supermarkets as shown in Figure 2.

After de-duplication and filtering out invalid entries, non-functional URLs, and corrupted codes, the curated resulting dataset contains 860 unique QR codes. Due to the diversity of locations used for the collection, we believe that this dataset captures diverse user experiences and privacy implications.

(a) Fast food chain in `France`. (b) University campus in `India`. (c) Hotel premise in `China`. (d) Restaurant in `Canada`.

Fig. 2: Illustration of collected QR codes in various contextual scenarios across diverse geographic locations.

Characteristics. Table 1 outlines the key features of our dataset, highlighting how QR codes predominantly contain URLs ($> 93\%$). First, we can also observe a strong adoption of secure protocols, with 81% of URLs using `https`, and a wide range of URL lengths, ranging from a minimum of 12 characters to a maximum of 203, with an average length of 45 characters. To analyze the diversity of remote content types when accessing a URL, we extract the `Content-Type` header from HTTP responses. Most URLs point to web content while static files like PDFs are also present, reflecting a broad mix of applications. The dataset further showcases diverse top-level domains (TLDs), with `.com`, `.fr`, `.de` and `.cn` being the most common among the top 10, underscoring the breadth of QR code use cases across various contexts.

Table 1: Characteristics of the QR Dataset

Feature	**Description/Statistics**
Number of QRs	860
Countries	`France` (721), `China` (60), `Austria` (58), `India` (15), and, `Canada` (6)
Collection Period	Jan. 2023 – Dec. 2024
HTTPS Usage	81%
Average URL Length	45 characters
Shortest URL Length	12 characters
Longest URL Length	203 characters
Content Types	web, plain-text, PDFs
Top 10 TLDs	`.com`, `.fr`, `.de`, `.cn`, `.co`, `.link`, `.org`, `.to`, `.me`, `.gouv.fr`, `.eu`

QR code capacity utilization. Figure 3 presents the distribution of URL lengths in bits, with the objective of illustrating the fraction of each QR code's total capacity that is used along with the necessity of using USS. Vertical lines denote the capacity of various QR code versions, annotated with the percentage of QR codes in our dataset that use each version. The plot reveals that around 85% of the QR codes are enough to use just version `V4` or lower, though it represents just 2.1% of the seen versions. This suggests that QR codes are often not optimized for storage efficiency, as even higher-capacity versions—such as `V29` (5.53%), `V27` (4.39%), and `V23`, `V31` and `V33` (4.20%)—are either completely unused or significantly underutilized. This observation contradicts the assumption that USS are required due to URL length constraints. While the use of USS may be motivated by other reasons (analytics, dynamic link, etc.), due to the privacy and security risks associated with USS [12], URL shortening only for capacity reasons should be avoided.

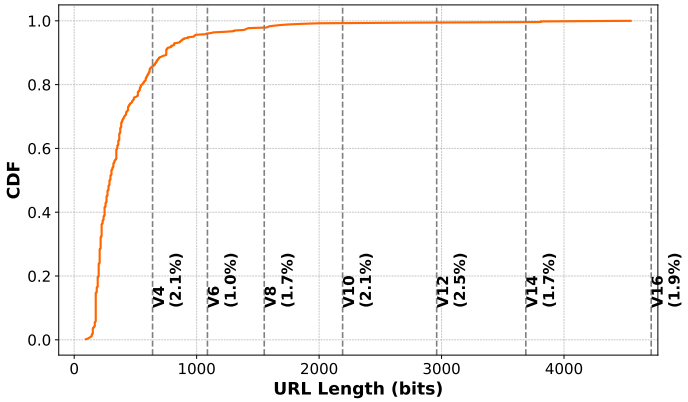


Fig. 3: Distribution of the length of the URL encoded in the QR code, along with the capacities of main QR code versions (vertical lines).

Unresolvable QR codes. Figure 4 presents an analysis of various failure modes encountered when attempting to resolve URLs encoded within QR codes. These categories include *Connection Timeout*, *Failed to Establish Connection*, *SSL Certificate Failure* and *Redirection Issues*. Among these, connection failures and timeouts are observed to be the most frequent errors, suggesting that a notable fraction of QR codes point to unavailable or misconfigured web resources, at the time of our access. SSL certificate failures indicate the presence of expired or untrusted certificates, which may compromise the security of interactions initiated through QR scans. The occurrence of redirection issues highlights cases in which excessive or broken redirections hinder access to the intended content. These results underscore potential usability and security concerns associated with QR

code scanning, reinforcing the need for caution when interacting with encoded URLs.

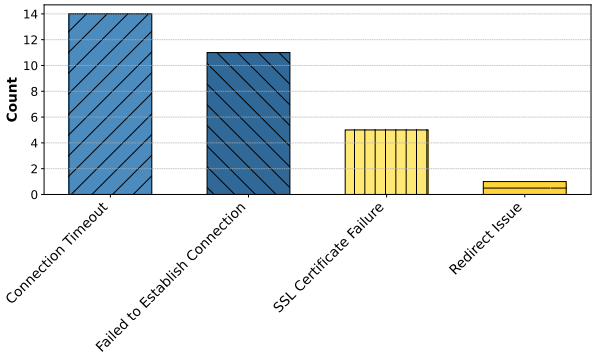


Fig. 4: Error in fetching encoded URLs

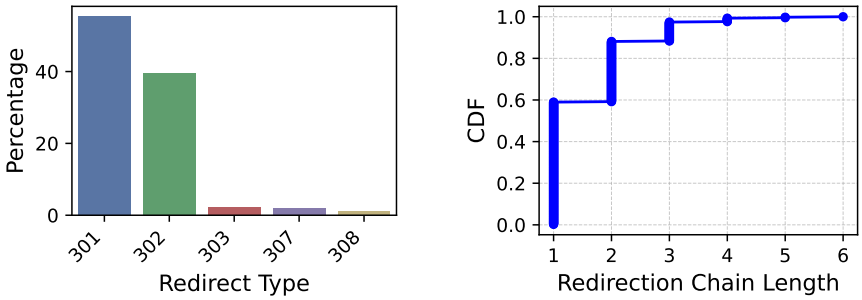
5 Redirections and USS

5.1 Redirection

Redirection is a common mechanism in USS, in which a user is directed from a shortened URL to its destination URL. Analyzing redirections provides insights into user navigation patterns and potential privacy concerns. We explore three key metrics related to redirection: the proportion of URLs that involve redirection, the distribution of redirection type, and the cumulative distribution function (CDF) of redirection chain lengths.

First, approximately 55% of the URLs exhibit at least one redirection, indicating that a majority of the shortened URLs require additional hops to reach their final destinations. Second, we have analyzed the distribution of redirection types, as shown in Figure 5a. The majority of redirections are HTTP 301 (*i.e.*, Moved Permanently) responses, accounting for close to 60% of the total while HTTP 302 (Found) responses follow at around 40%. This distribution suggests that most redirections (301) are designed to be persistent.

Finally, the CDF of redirection chain lengths, depicted in Figure 5b, provides insight into the depth of redirection paths. The result shows that 60% of URLs involve only one redirection, while 40% require two or more redirections. Notably, 10% of URLs exhibit three or more redirections, which could amplify privacy risks as users' navigation paths traverse multiple intermediate domains, specifically if it involves new parties (servers).



(a) Redirection types. (b) Lengths of redirection chains.

Fig. 5: HTTP redirections in QR URLs.

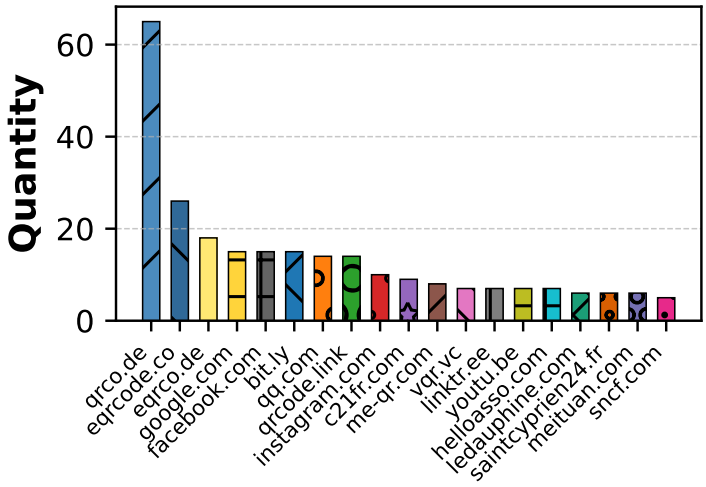


Fig. 6: Reuse of domain names across different QR codes.

5.2 Top Providers

Identifying popular QR code generators and providers is crucial to understanding the USS ecosystem within QR codes. One of the key approaches to this analysis is examining the reuse of domain names across different QR codes, as illustrated in Figure 6.

Our findings highlight a diverse set of top domains frequently associated with QR codes, indicating their widespread adoption and reuse. These include dedicated QR code providers like `qrco.de`, `eqrcode.co`, `qrcode.link`, `me-qr.com` and `vqr.vc`, as well as general-purpose URL shorteners such as `bit.ly`. These domains dominate the QR USS landscape, with significant reuse across different QR code deployments.

In addition to dedicated providers, search engines and social networks also play a role in this space. For instance, we also notice in our dataset that `google.com` is occasionally used for URL shortening. Social network platforms like `facebook.com`, `youtu.be`, and, `instagram.com` sometimes provide short URLs and QR codes to facilitate content sharing. These cases complement the activities of specialized QR code generators and emphasize the multi-faceted nature of URL shortening in the QR code ecosystems. In the next section, we will be specifically focusing on dedicated QR USS.

6 Cookie exploitation by QR USS

The use of cookies in USS linked to QR codes poses significant privacy risks, particularly due to their involvement as a third party. These services can set cookies during redirection, often collecting user data without explicit consent. Domains frequently involved in redirections could use tracking and analytics cookies to monitor user behavior, including link clicks, browsing patterns, and locations.

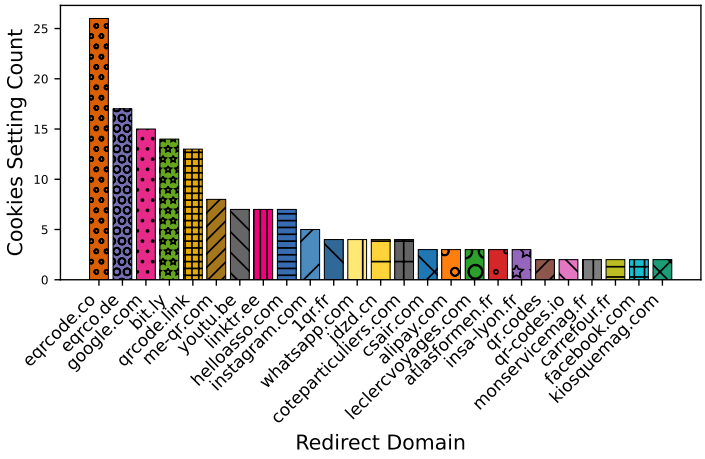


Fig. 7: Number of instances that top redirectors set cookies.

6.1 Cookies in Redirections

Cookies' potential for misuse raises significant privacy concerns, especially when they collect data without users' explicit consent. Our analysis reveals that 39.2% of URLs with external redirections set cookies, potentially involving third-party domains that engage in tracking and analytics. Top redirectors using cookies (see Figure 7) include prominent domains like `qrcode.co`, `eqrco.de`, `google.com`,

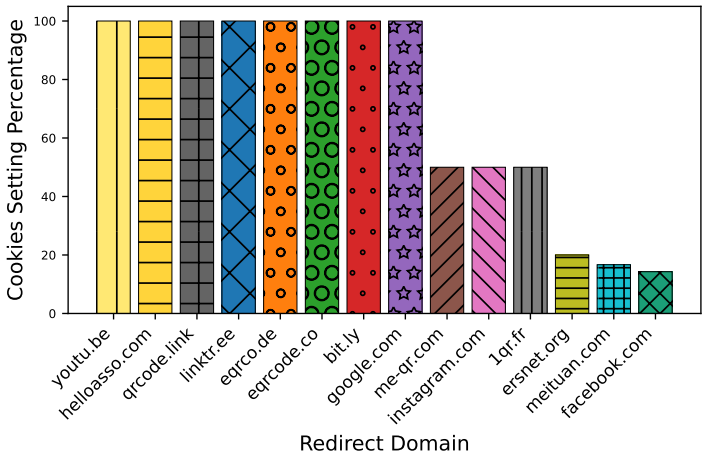


Fig. 8: Individual cookie usage by top redirectors.

`bit.ly`, and `qrcode .link`, which leverage cookies for a variety of purposes which we investigate in Section 6.2. Social networks and content-sharing platforms such as `instagram.com`, `whatsapp.com`, and `linktr.ee` also integrate cookies into redirection chains.

An analysis of individual cookie usage by top redirectors (Figure 8) reveals that certain redirection domains consistently set cookies in 100% of their QR services. Domains such as `qrcode.link`, `eqrco.de`, `eqrcode.co`, `linktr.ee`, `bit.ly`, `helloasso.com`, `youtu.be` and `google.com` exemplify this trend. This uniform behavior underscores their reliance on cookies as a mechanism for enabling functionalities like session tracking, personalization, and data collection.

6.2 Classifying Cookies Set by QR USS

Inspired by prior work [2][10], we have used `Cookiepedia` [16] as a key resource for classifying cookies. `Cookiepedia` is a comprehensive database of cookies maintained by the Consent Management Platform (CMP) OneTrust [4,2]. Table 2 lists potentially intrusive cookies (analytics/advertising/tracking) classified by `Cookiepedia`.

Table 2: Intrusive (tracking/advertising/analytics) Cookies collected by QR USS [16].

Cookie Names
NID, dwanonymous_7ad186a87c92467d3c0495460d6555fc, _fbp, mgrefby, dwac_10e76b50b2189fd2b0a7ed1e4f, cquid, VISITOR_INFO1_LIVE, mgref, __wpdm_client, G, SS, AS, SP

Targeting/Advertising/Analytics Cookies. We found 13 targeting/advertising cookies set by QR USS. By tracking activities across multiple QR scans, these cookies may enable advertising networks to understand shopping habits, interests, and preferences. In the following, we highlight some of the notable cookies and their purposes mentioned in `Cookiepedia`:

- `VISITOR_INFO1_LIVE`, monitors YouTube video interactions on external sites, influencing video recommendations and optimizing ad placements tailored to user engagement.
- `SP`, associated with event-booking, collects click data to enhance advertising campaigns across multiple platforms, ensuring the effectiveness of ad delivery strategies.
- `dwanonymous_7ad186a87c92467d3c0495460d6555fc`, tracks anonymous interactions for dynamic ad targeting, analyzing patterns without directly identifying users.
- `dwac_10e76b50b2189fd2b0a7ed1e4f`, collects data on user navigation and interaction to fine-tune real-time ad targeting and personalization.
- `cquid`, supports advanced ad delivery by identifying user sessions and linking them to ad content most likely to drive engagement.
- `__wpdm_client`, facilitates tailored advertising by tracking interactions with dynamic web content.
- `mgref`, tracks referrer information to analyze the origin of user traffic, such as from websites, ads or social media campaigns, enabling precise attribution of ad performance.
- `mgrefby`, captures intermediate interactions to provide a detailed view of multi-step browsing paths, enhancing the granularity of campaign evaluations.
- `SS` (Session Source), gathers session data in real-time, including clicks and page visits, allowing advertisers to dynamically adjust campaigns to reflect user behavior.

The integration of these cookies within QR USS poses serious privacy challenges, as they can aggregate data across platforms to build detailed user profiles. Furthermore, there is no consent collection nor information of the user, which prevents the exercise of user rights with respect to privacy, in contrast to other online situations.

Extended Cookie Analysis. We identify 22 out of 89 cookies that are not included in the `Cookiepedia` classification in Table 2 and are labeled as *unknown*. We start with still using `Cookiepedia` insights to dive into four key aspects of such cookies to know their prevalence: i) the host domains setting such cookies, ii) the websites reporting them as third-party cookies, iii) the websites reporting them as persistent cookies, and finally iv) the websites reporting them as session cookies.

Figure 9 provides a distribution of the above four key attributes. The figure reveals that these unknown cookies exhibit diverse behaviors in terms of their presence across different domains and classifications. The first plot illustrates

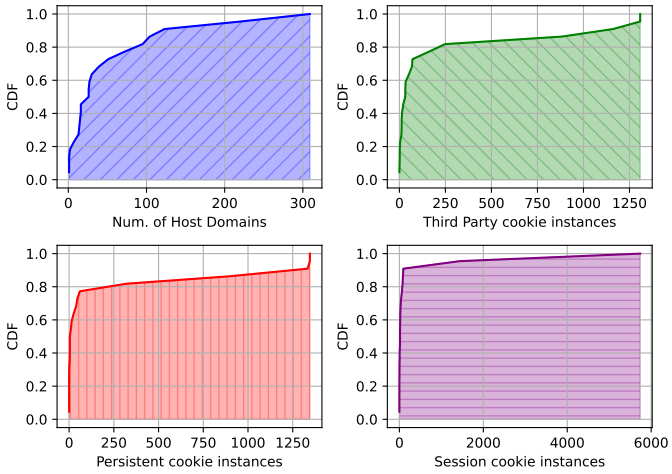


Fig. 9: Characteristics of unclassified cookies.

that while most cookies are confined to a small number of host domains, a few are widely distributed, spanning nearly 300 domains. The second and third plots highlight that these unclassified cookies are frequently observed as third-party and persistent cookies, with some instances exceeding 1000 occurrences. The final plot shows that session cookies, which typically exist only for a single browsing session, have a significantly higher number of occurrences, with some appearing in over 6000 instances. This suggests that these unclassified cookies may be extensively used for tracking purposes, particularly through session-based mechanisms, reinforcing the need for further investigation into their role and privacy implications.

Moreover, an extended investigation revealed that among these *unknown* cookies are the tracking cookies `_bit` [1]from Bitly which stands out for monitoring interactions with shortened URLs, such as clicks and timestamps, potentially uncovering behavioral patterns. On the advertising front, cookies like `adsStatData` [2] and `trackingId` [3] enable targeted advertising by assigning unique user identifiers and tracking ad performance. Given the large lifetimes of several other cookies (see next section) like `qB`, `msv4_idCookieUser`, `qrm` and `QoSID`, their classification as an intrusive cookie looms large.

6.3 Cookie Lifetimes

We analyze the lifetimes of all the above-identified potentially intrusive cookies to understand the duration of data collection. Figure 10 illustrates the distribution

[1] `https://bitly.com/pages/privacy`
[2] `https://me-qr.com/privacy-policy`
[3] `https://www.facebook.com/privacy/policy`

of privacy-intrusive cookies' lifetimes in days. Results show that intrusive cookies persist way beyond a single browsing session, with more than 95% of cookies lasting between 30 and 7300 days. This indicates that QR USS set cookies that could potentially construct detailed user profiles over extended periods. Among all cookies, `COMPASS` stands out with a lifetime of less than a day, suggesting that it is primarily used for session-based tracking. In contrast, `QoSID` cookies persist for 2315 days, while `qrm`, `msv4_idCookieUser` and `qB` last respectively for 3650, 4166 and 7300 days, indicating long-term tracking capabilities. In comparison, we observed that more than 27% of cookies that are not classified as privacy-intrusive by `Cookiepedia` expire when the browser is closed or have a lifetime of less than a day.

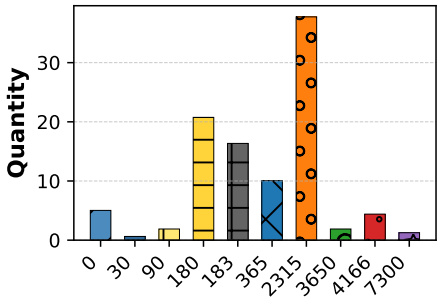


Fig. 10: Lifetimes of intrusive cookies (in days).

6.4 Set-Cookie Prevalence Across Geographic Locations

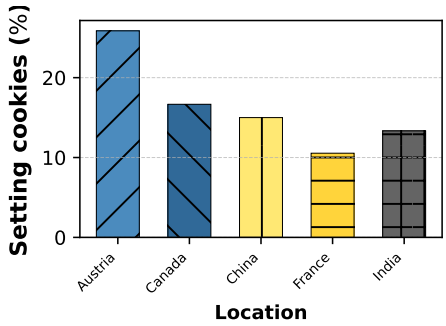


Fig. 11: Variance of set-cookies across geographic locations

Figure 13 illustrates the distribution of websites that actively set cookies across different geographic locations of collection. The results reveal notable

differences in cookie-setting behaviors among countries. Austria and Canada exhibit a relatively higher prevalence of websites that set cookies, whereas Canada, China, and India also show higher percentages. These variations can be attributed to differences in privacy regulations, enforcement levels, and common industry practices. The elevated presence of set cookies may be influenced by factors such as the widespread use of analytics and tracking services. Conversely, the relatively lower prevalence in France reflects stricter implementation of web-related privacy laws such as the GDPR, which influence how websites manage user data. The results indicate that user privacy experiences may vary substantially depending on geographic location, reinforcing the need for a global perspective on cookie regulation and enforcement.

7 Privacy Implications of URL Path and Query Strings

Independently of the resource they point to, QR codes are often associated with a specific context. Information related to this context can be transmitted to web servers through the URL associated with the QR code. The USS may also have access to this information, though it is unclear whether they process it. Such contextual information is typically encoded via the Query String part of the URL (*e.g.*, the URL path can also be used for this purpose). In our analysis, we found that 25.7% of the URLs contained query parameters, increasing even more when also taking into account those hidden behind URL shorteners and redirections, indicating their significant presence in QR code usage.

Table 3 provides an overview of the most frequent query parameters extracted from URLs embedded in QR codes. These query strings often encode metadata related to marketing campaigns (`utm_source`, `utm_campaign`, `utm_medium`), user identifiers (`id`, `token`, `cid`), payment information (`pa`, `pn`, `mc`), and other contextual data such as session details, venue and timestamps. The presence of structured tracking parameters suggests a significant potential for user profiling and behavioral tracking across multiple web domains. Notably, the high occurrence of payment-related identifiers indicates that QR codes are increasingly used in financial transactions, raising concerns regarding potential data exposure and security risks. In the following, we perform a detailed analysis of the privacy threats posed by these query strings, investigating their implications for user tracking, data retention, and third-party sharing mechanisms.

7.1 Privacy Implications

A manual analysis of the dataset uncovered a number of data items transferred via query string parameters. Representative examples are presented in Table 4, 5, and 6. In the following, we describe those items, as well as the personal information that can be inferred from them.

Generic User Profiling. Query parameters embedded in QR codes can reveal various types of personal information that could be exploited for user profiling.

Table 3: Most frequent query strings

Query	Count	Examples of values
utm_source	35	seatbackad, AfficheVitrine, INSA, akilux, QR code, QR-code, qr, VersionFemina, affichage, lbpcatcadeau, QR_Code, qr_code, affiche, tag2d, maline, print, Print
utm_campaign	25	Bien, fermeture-covid, op12-aout-2023, carrefour, 2023_03_01_Aidants_VersionFemina, work_info, b.nouveau_programme_de_fidelite.2024-01-31 Epigenetic, ZT, QR-code-revetsens, akilux-21-22, vuelingsyteng, store4574, INSA_Partnership_FRA_TutRecruit, francais-parlent-anglais jeu-rugby
utm_medium	25	ooh, qrcode, akilux, QR code, qr, Vitrophanie, Display_Branding, informative_sign, affichage, Print, metro, QRcode, stopper, Partnership
id	14	com.freetnessenergy, 100057276736505, CNU85387, 5d985bd8-1a52-49c4-b177-a3186016aa0a, 1394140769041321986, CNU66119, 100087057884876, 192189 com.keolis.ruban, 61153, prod.maasify.evian, 61553441277844
r	10	qr
pa	9	Q751854265@ybl, gpay-11240877578@okbizaxis, paytmqr1w14125ijv@paytm
pn	9	PhonePeMerchant, Paytm, Google Pay Merchant
mc	5	5399, 0000
utm_content	5	Presse, iledefrance, 2557, NW_zt, revetsens
purpose	4	00
token	4	q1ZKLUvNK4nPTFGy...
cid	4	B2010120080300323, B64105231128C2999, B72106230517A0809
table	4	438, B02, 506, T5
COUNTRY	4	1
questlist	4	COUNTRY;CDPF;SOURCE
CDPF	4	31122, 32048, 17501
mode	4	02
starts_at_from	3	2024-09-08 08:00:00, 2024-09-10 08:00:00, 2024-09-10 12:30:00
day	3	2024-09-07
days	3	2024-09-10, 2024-09-08
cu	3	INR
contentsessiontype_ids	3	7585
order	3	room_title
filterbox	3	true
ends_at_to	3	2024-09-10 09:30:00, 2024-09-08 09:30:00, 2024-09-10 14:00:00
timezone	3	false
room_ids	3	8374,8373

Beyond consumption habits, QR codes can also be linked to medical contexts. For instance, QR codes associated with clinical trials may contain parameters disclosing the *medical center* conducting the research and the *trial date* (see Table 4). Additionally, event-related QR codes often include details about *time, venue, and session types*, allowing organizations to track attendance patterns and user interests in specific topics. Even seemingly minor details, such as a *flyer ID*, can be leveraged to infer which promotional materials a user has interacted with, along with the location.

QR codes found on product packaging frequently contain *product identifiers* or *descriptive details*, revealing specific consumption behaviors (see toothpaste and liquor entries in Table 4). Similarly, QR codes in print advertisements may

Table 4: Generic profiling through query string parameters from QR codes.

Information	Query String	Notes
Medical information	c i d = q r % 3 A **Nucleus_Global%3Arespiratory%2Fongoing-clinical-trials%3A29_07_2024** % 3 A L a n d i n g _ Page	*Center* and *date* of the clinical trials
Event tracking	'contentsessiontype_ids': ['7585'], **'starts_at_from': ['2024-09-10 08:00:00'], 'ends_at_to': ['2024-09-10 09:30:00'], 'room_ids': ['8374,8373]**, 'grade': ['1'], 'filterbox': ['true'], 'legendbox': ['false'], 'timezone': ['false'], **'days': ['2024-09-10'], 'day': ['2024-09-07']**, 'viewType': ['list'], 'segment': ['poster'], 'segmentname': ['Abstract Sessions'], 'order': ['room_title']	*Time* and *venue* of the event.
Flyer ID	**'flyerCode': ['bonapp106ja983']**, 'redirectURL': ['https://bit.ly/3A3SIG6']	*ID* of the information flyers.
Toothpaste	qrcode=onpack&utm_source=PackSecondary&utm_medium=QR&utm_campaign=Alma-BH0437-BE-FR&utm_content=Sustainability&Purpose-BOP&utm_term=**I8_White**&**evt_product_id=VyeH9t7wKCVgaUxTdnYAwbab**&evt_implicit_scan_id=VET5myhUd9CPr2e2GBYAUkhr	Name (*Signal Integral 8 white*) and identifier of the toothpaste
Liquor	utm_source=onpack&utm_medium=qr&utm_campaign=**get-27**&utm_content=**70cl-SEU**	Name of the liquor and size of the bottle
Cheese	u t m _ s o u r c e = p a c k & u t m _ c a m p a i g n = **camembert**&utm-source=QR_CODE	Type of cheese
Name of the magazine and issue	utm_source=Presse&utm_medium=QRCODE&utm_campaign=LEGS24&utm_content=**Diverto1+S2+24**	Name of the magazine and issue (2nd semester 2024)
Name of the magazine and issue	utm_source=**FEMINA-OCT**&utm_medium=QRCODE&utm_campaign=24LEGS&utm_content=PRESSE	Name of the magazine (*Femina*) and issue (*October*)

include query parameters identifying *the publication source and issue number*, thereby exposing *reading habits and preferences* (see magazine entries in Table 4). These findings highlight how query parameters in QR codes serve as *silent carriers of personal information*, emphasizing the need for stricter privacy considerations when embedding tracking elements in consumer-facing QR campaigns.

Location Data. Query string parameters embedded in QR codes often contain location-related information, allowing third parties to infer a user's whereabouts. In some cases, these parameters provide *coarse-grained geographical details*, such

Table 5: Location inference through query string parameters from QR codes.

Information	Query String	Notes
Region	v=iledefrance-affichage&utm_source=affichage&utm_medium=metro&utm_campaign=francais-parlent-anglais&utm_content=**iledefrance**	Region *Ile de France*
City	CD_ACCESS=NF8E4X&**CP=35300**&QR=1&utm_source=lbpcatcadeau&utm_medium=qrcode	Postal code
Train station	**uicDeparture=8738249**	Train station identifier
Store	https://one.asics.com/instore/?**store=store4574**&utm_campaign=store4574	Store identifier
Restaurant venue and table	**restaurant=K0076&table=T5**&terrace=false&token=6A5D2166CB73A75F34141DC991350B2C	Restaurant identifier and table identifier.
Points of interest	'continue': [**placeid=ChIJo82WsnB9EcRjog UKxUEHhg**&source=g.page.m.rc._&laa=merchant-web-dashboard-card]	Unique location ID
GPS coordinates	country=FR&articleId=IN7017& **lat=48.73085021972656** & **lon=2.174370050430298**	Latitude and longitude

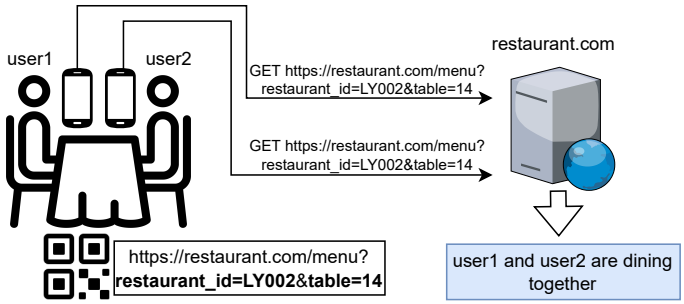


Fig. 12: Scanning a QR code to access the menu at a restaurant can reveal that two persons are dining together.

as a *region* or a *postal code*, revealing the general area in which the QR code was scanned (see Table 5). Additionally, store identifiers and train station references can be leveraged to infer visits to specific commercial venues or transportation hubs. These identifiers serve as digital breadcrumbs that, when combined with other contextual data, can contribute to behavioral profiling.

Beyond regional and commercial location data, query parameters may also expose *precise location details*. This includes parameters containing *GPS coordinates*, allowing the tracking of exact user movements or *unique place identifiers* associated with points of interest (*e.g.*, restaurants, and review pages). The presence of restaurant-specific table numbers further suggests the capability to monitor *in-store seating preferences*. Consequently, if users open these QR codes on-site (as is typically the case), external parties can correlate location data with *timestamps* to reconstruct detailed visit timelines, raising significant privacy concerns. Hereafter, we discuss in detail the extended version of this threat.

Social Link Inference. Location-specific QR codes could potentially be leveraged to infer social links between users, as spatio-temporal coincidence [3] (being in the same place at the same time) is a strong indicator of social ties. For instance, in some restaurants, QR codes provided to access the menu and place orders are specific to the venue and the table: https://www.mcdonalds.fr/restaurants/564/commander/eat_in?table=506.

A group of people dining together would access this QR code at approximately the same time, revealing that they are all seated at the same table and, therefore, part of the same social group (see Figure 12). Although we found no indication that such an inference is currently being performed, we observed that the underlying information is readily available in many cases, as these table-specific QR codes are commonly used by major fast-food restaurant chains.

Other Leaks. Beyond user profiling and location inference, QR code query strings may also expose *sensitive security-related information*. Notably, some URLs contain *API keys* or *authentication tokens* embedded directly within the request parameters (see Table 6). These keys, if leaked or intercepted, could allow unauthorized access to protected resources, potentially leading to data breaches or system compromises. The presence of authentication keys further exacerbates security risks, as malicious actors could exploit them to impersonate legitimate users or access restricted services. Additionally, query strings reveal *application identifiers* used for tracking and analytics. These identifiers allow third parties to monitor user interactions across different digital platforms, facilitating targeted marketing and behavioral profiling.

7.2 Obfuscated and Encoded Parameters

The previous cases were identified because the data was in plaintext and understandable. However, other parameters may carry similar information in an obfuscated or encoded format, which can be associated with context information by the website owner. Our dataset includes 38 QR codes in which the URL contains an alphanumerical identifier of at least 6 characters[4]. However, we were

[4] Examples of identifiers: `utm_campaign=042022_media_mdlz_ritz`, `token=cd0d80288ae42eef94b4132d982064e9d7f37834`, `products_id=51658860`,

Table 6: Other leaks.

Information	Query String	Notes
API key	**'apiKey': ['gQpFZIioRDSOCYMoeFLY5Kbo3Dl19Quz']**	Security threat
Auth key	'fiche_type': ['publique'], 'lang': ['fr'], 'idann': ['26472'], 'internal': ['1'], 'idag': ['1'], **'authKey': ['ab7127222e4ac2f3d0fdc7a8e8486f4f']**	Security threat
Application ID	**'appId': ['3']**, 'campaignId': ['1699281368256960'], 'utm_source': ['cobranding'], 'utm_campaign': ['vpfr-migration-1223'], 'utm_medium': ['cobrandingnoel']	user-analytics

unable to determine whether these were used to convey context information. The cases presented in Section 7.1 should, therefore, be considered as the visible part of the threat.

7.3 Query Strings across Countries

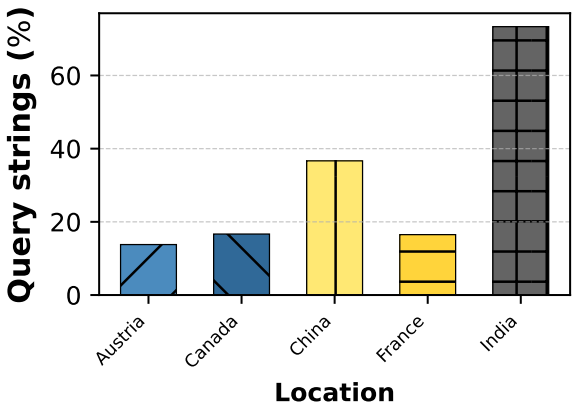


Fig. 13: Variance of present query strings across geographic locations

Figure 13 presents the distribution of query strings embedded in QR code URLs across different geographic locations. The variation in the presence of query strings is indicative of differences in how QR codes are generated and utilized globally. Notably, India exhibits the highest proportion of URLs containing

`id=1730457376655-ddaf88af-1fe6-4f5b-b657-6426c82d4ce3`, `scenario_uid=11f23a77f5a8a186f444`

query strings, surpassing 60%, highlighting a large potential for embedding tracking or identification. China also shows high levels, with approximately 40% of QR codes embedding such parameters. In contrast, France, Canada, and Austria have significantly lower occurrences, suggesting a lesser reliance on query-based tracking in QR-linked services. These discrepancies may be attributed to regional regulations, business practices, or differences in QR code deployment strategies. The presence of query strings in QR codes has already raised privacy concerns underscoring the necessity for location-specific privacy considerations when assessing QR code-related risks.

8 Discussion and Conclusion

Discussion. Tech-savvy users might attempt to assess potential privacy risks by inspecting the URL embedded in a QR code before scanning it. However, problematic parameters may be concealed through obfuscation or only become visible further in the redirection chain (*e.g.*, after passing through a USS). Thus, users are often unaware of the usage of information. Addressing this issue requires better disclosures and mechanisms for informed user consent.

Another challenge is due to the role of USS providers as third parties in the data collection process. These intermediaries are often not explicitly mentioned in privacy policies, leaving users unaware of their involvement. This omission complicates accountability and it has already been shown that many privacy policies fail to account for such third-party roles [17]. Regulations need to ensure that USS providers are clearly identified in policies. While laws like the General Data Protection Regulation (GDPR) emphasize transparency and consent, they do not specifically address the unique privacy risks associated with QR codes and USS. To mitigate the risks of unauthorized data collection and profiling, policymakers should consider updating existing frameworks to encompass emerging technologies, and service providers must proactively inform and protect users.

Conclusion. This paper provides a comprehensive analysis of the privacy risks associated with QR code usage and URL shortening services (USS). Our study leverages a unique and diverse dataset of 860 QR codes collected over two years, highlighting real-world practices across multiple domains, regions, and internet regulations. Unlike prior work, we examine the interplay between QR codes, USS, and tracking mechanisms, uncovering significant risks such as invasive behavioral profiling and geolocation tracking. Our findings reveal the significant presence of tracking/advertising/analytics cookies set in redirected URLs. Moreover, we find that URLs embed sensitive data like user identifiers and location data in the query strings enabling user profiling and social link inference.

References

1. Bernholz, L.: Philanthropy and Digital Civil Society: Blueprint 2022 (Dec 2021), https://pacscenter.stanford.edu/publication/philanthropy-and-digital-civil-society-blueprint-2022/
2. Bollinger, D., Kubicek, K., Cotrini, C., Basin, D.: Automating cookie consent and {GDPR} violation detection. In: 31st USENIX Security Symposium (USENIX Security 22). pp. 2893–2910 (2022)
3. Crandall, D.J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., Kleinberg, J.: Inferring social ties from geographic coincidences. Proceedings of the National Academy of Sciences **107**(52), 22436–22441 (Dec 2010). https://doi.org/10.1073/pnas.1006155107, https://www.pnas.org/doi/10.1073/pnas.1006155107, publisher: Proceedings of the National Academy of Sciences
4. Hils, M., Woods, D.W., Böhme, R.: Measuring the emergence of consent management on the web. In: Proceedings of the ACM Internet Measurement Conference. pp. 317–332 (2020)
5. Hunter, T.: QR codes are a privacy problem — but not for the reasons you've heard. Washington Post (Oct 2021), https://www.washingtonpost.com/technology/2021/10/07/are-qr-codes-safe/
6. Kieseberg, P., Leithner, M., Mulazzani, M., Munroe, L., Schrittwieser, S., Sinha, M., Weippl, E.: Qr code security. In: Proceedings of the 8th International Conference on Advances in Mobile Computing and Multimedia. pp. 430–435 (2010)
7. Krombholz, K., Frühwirt, P., Rieder, T., Kapsalis, I., Ullrich, J., Weippl, E.: QR Code Security – How Secure and Usable Apps Can Protect Users Against Malicious QR Codes. In: 2015 10th International Conference on Availability, Reliability and Security. pp. 230–237 (Aug 2015). https://doi.org/10.1109/ARES.2015.84
8. Lerner, A.: Measuring and Improving Security and Privacy on the Web: Case Studies with QR Codes, Third-Party Tracking, and Archives. Ph.D. thesis, University of Washington (Aug 2017)
9. Lerner, A., Saxena, A., Ouimet, K., Turley, B., Vance, A., Kohno, T., Roesner, F.: Analyzing the use of quick response codes in the wild. In: Proceedings of the 13th annual international conference on mobile systems, applications, and services. pp. 359–374 (2015)
10. Munir, S., Siby, S., Iqbal, U., Englehardt, S., Shafiq, Z., Troncoso, C.: Cookiegraph: Understanding and detecting first-party tracking cookies. In: Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security. pp. 3490–3504 (2023)
11. Neumann, A., Barnickel, J., Meyer, U.: Security and privacy implications of url shortening services. In: Proceedings of the Workshop on Web 2.0 Security and Privacy (2010)
12. Nikiforakis, N., Maggi, F., Stringhini, G., Rafique, M.Z., Joosen, W., Kruegel, C., Piessens, F., Vigna, G., Zanero, S.: Stranger danger: exploring the ecosystem of ad-based url shortening services. In: Proceedings of the 23rd international conference on World wide web. pp. 51–62 (2014)
13. Rotsios, K., Konstantoglou, A., Folinas, D., Fotiadis, T., Hatzithomas, L., Boutsouki, C.: Evaluating the Use of QR Codes on Food Products. Sustainability **14**(8), 4437 (Apr 2022). https://doi.org/10.3390/su14084437, https://www.mdpi.com/2071-1050/14/8/4437
14. Smith, J.J.: Corporate tips: Are QR codes problematic from a Privacy Law standpoint? (Oct 2021), https://www.lexology.com/library/detail.aspx?g=fe9e7b45-fbbf-4906-972f-9d541f5516d1

15. Tiwari, S.: An introduction to qr code technology. In: 2016 International Conference on Information Technology (ICIT). pp. 39–44 (2016). https://doi.org/10.1109/ICIT.2016.021
16. Trust, O.: Cookiepedia. https://cookiepedia.co.uk (2024), accessed: 2024-11-13
17. Utz, C., Amft, S., Degeling, M., Holz, T., Fahl, S., Schaub, F.: Privacy rarely considered: Exploring considerations in the adoption of third-party services by websites. arXiv preprint arXiv:2203.11387 (2022)
18. Wahsheh, H.A.M., Luccio, F.L.: Security and Privacy of QR Code Applications: A Comprehensive Study, General Guidelines and Solutions. Information **11**(4), 217 (Apr 2020). https://doi.org/10.3390/info11040217
19. Warf, B.: Geographies of global internet censorship. GeoJournal **76**, 1–23 (2011)
20. West, A.G., Aviv, A.J.: On the privacy concerns of url query strings (2014)
21. Woo, E.: QR Codes Are Here to Stay. So Is the Tracking They Allow. The New York Times (Jul 2021), https://www.nytimes.com/2021/07/26/technology/qr-codes-tracking.html
22. Yao, H., Shin, D.: Towards preventing QR code based attacks on android phone using security warnings. In: Proceedings of the 8th ACM SIGSAC Symposium on Information, Computer and Communications Security. pp. 341–346. ASIA CCS '13, Association for Computing Machinery, New York, NY, USA (May 2013). https://doi.org/10.1145/2484313.2484357