The Ecosystem of Public SMS Inboxes: Characteristics, Usage, Threats and Mitigation

Master Thesis in Information Systems, Production and Logistics Management

by

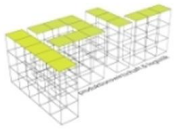
Leonhard Zacharias

submitted to the School of Managment,
Department of Information Systems, Production and Logistics Management

in partial fulfillment of the requirements
for the degree of Master of Science

supervisor: Ass.-Prof. Dr. Svetlana Abramova,
Department of Computer Science

Innsbruck, 27 November 2022



Master Thesis

The Ecosystem of Public SMS Inboxes: Characteristics, Usage, Threats and Mitigation

Leonhard Zacharias (01417092)
leonhard.zacharias@student.uibk.ac.at

27 November 2022

Supervisor: Ass.-Prof. Dr. Svetlana Abramova

Abstract

This thesis examines public SMS inboxes that are free-to-use websites offering phone numbers and enabling users to receive SMS messages sent to those numbers. These services get used for malicious and benevolent actions, which results in various threats to users, internet services and third parties.
By incorporating evidence from third-party data sources and a set of 14 million SMS messages scraped from ten websites, we examined the ecosystem of public SMS inboxes end-to-end. We uncovered that these websites share numbers and, to some extent, rely on illegal SIM boxes. We conclude that websites offering public SMS inboxes are not legitimate businesses. Their services facilitate cybercrime mainly by providing phone numbers as free resources for phone-verified account evasion at internet services while also serving legitimate privacy-related use cases. The threats posed by public SMS inboxes can be mitigated by checking phone number reputation and other countermeasures.

Acknowledgments

I'm extremely grateful to my supervisors Ass.-Prof. Dr. Svetlana Abramova and Dr. Daniel Woods for providing me with guidance, feedback and patience throughout the process of writing my thesis.
Many thanks also go to Univ.-Prof. Dr.-Ing. Rainer Böhme and his working group „Security and Privacy Lab“ for allowing me to work on this exciting topic in their department.
I had the pleasure of experiencing the company of my wonderful study colleagues, who helped while working on my thesis.

Contents

List of Figures

List of Tables

1 Introduction

The first SMS (Short Message Service) message was sent nearly 30 years ago (BBC, 2002). Despite its legacy status, security threats, and limited capabilities (Androulidakis, 2016), SMS A2P (Application-to-Person) messaging traffic is increasing (Mobilesquared, 2021). Even big tech companies like Amazon, Google, Facebook, and Microsoft cannot avoid using SMS messages in their authentication processes due to its easy deployment, ease of use, and consumer habits (Okta, 2020). Nowadays, phone verification is also a deterrent against bulk account creation by bots (Thomas et al., 2015), as it is assumed that individuals only have access to their (single) phone number, and the acquisition of multiple numbers is expensive.

Similar to disposable e-mail services (Hu et al., 2019), websites emerged that provide disposable phone numbers. Legit use cases for these services exist, for example, to hide someone's phone number for privacy reasons or when a local phone number with a distinct prefix is required for an app. However, they also enable malicious users to fake their phone numbers, facilitating phone verified account (PVA) evasion (Thomas et al., 2014). In addition, several other threats, primarily privacy leakage, evolve due to the openness and nature of these services.

To further understand this phenomenon, this study examines the ecosystem of public SMS inboxes end-to-end by measuring how the different parties interact. First, the characteristics of websites that offer public SMS inboxes and the SMS gateways that face the mobile network providing the phone numbers and SMS messages are examined. Second, the usage of phone numbers and the affected services are studied. Third, we discuss possible threats emerging by using public SMS inboxes, and fourth, we evaluate approaches to mitigate these.

The key contributions of this thesis are:

- Extensive literature review on temporary phone numbers, particularly public SMS inboxes and related topics.
- Development and Deployment of a scraping system to collect SMS messages from public SMS Inboxes.
- Collection of a dataset of 13 million SMS messages from 10 websites providing public SMS inboxes for two months.
- Uncovering the inner workings, characteristics, and collaboration of websites providing public SMS inboxes.
- Proof of linkage between public SMS inboxes and illegal „SIM Box“ gateways.

- Enumeration of malicious and benevolent use cases, providing evidence with our dataset.
- Study on the primary usage of public SMS inboxes, the evasion of phone verification.
- Enumeration of threats for users, services, and third parties and assessment thereof with the gathered data.
- Assessment of possible counter-measures to mitigate stated threats and risks.

2 Background

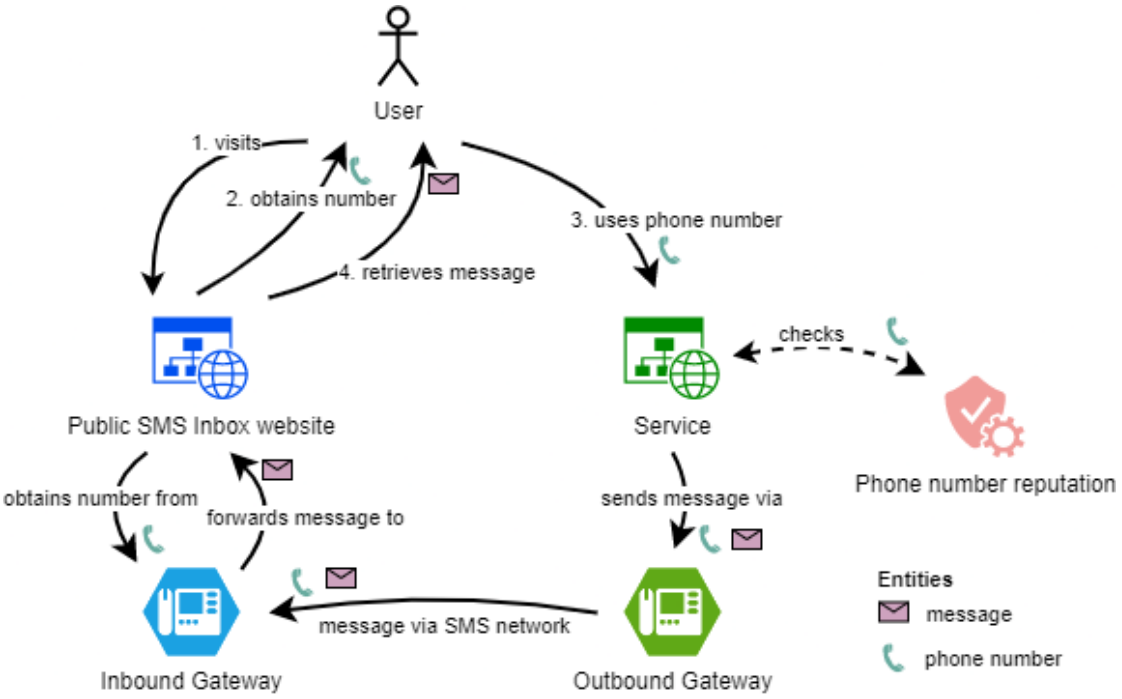


Figure 2.1: Interaction of users, public SMS inbox websites, and internet services.

Public SMS inboxes are services that publish phone numbers that individuals can use anonymously and freely to receive SMS messages on these numbers. The SMS messages sent to these are made available for everyone on the website publishing these phone numbers. Usually, a user can choose from many numbers with different country prefixes on the site. The characteristic is that this service is free for users and can be accessed unrestricted.

For example, User A wants to book a flight online. Because A fears spam misuse of his phone number by the online travel agency, A does not want to hand his phone number out. So A goes to a public SMS inbox website, looks up a number that matches A's country, and uses it instead of A's own. The online travel agency requires to verify the number and thus sends an SMS with a One-Time Password (OTP) to the number provided by A. The public SMS inbox website receives this SMS and displays it on its website. A now retrieves the SMS with the OTP from the public inbox and enters it at the online travel agency website.

These services are also called temporary phone numbers. Prior research denominates them as “Public Gateways” (Reaves et al., 2016), “Disposable Phone Numbers” (Cheng et al., 2020) or “Online SMS receiving websites” (Berenjestanaki et al., 2019). We see the terms as not precisely describing the observed phenomena, so we conceptualize the term “Public SMS Inboxes”. “Public” refers to unrestricted access without registration and free and anonymous usage. “SMS” refers to the exclusive protocol in these services, which can receive only SMS, not MMS, phone calls, or messengers. “Inbox” is the interface describing that messages are read-only and can not be sent through these services.

Other sources for temporary phone numbers exist that do not correspond to the definition given above. So-called “Burner phone number” apps[1] exist that offer a paid service where registration is needed and thereby are not truly anonymous. Similar to public SMS inboxes, websites[2] exist that rent out phone numbers for a short time, specially tailored for phone verification at specific websites. Recent studies found that some of these services are linked to Android malware partially intercepting SMS messages (Dong et al., 2022).

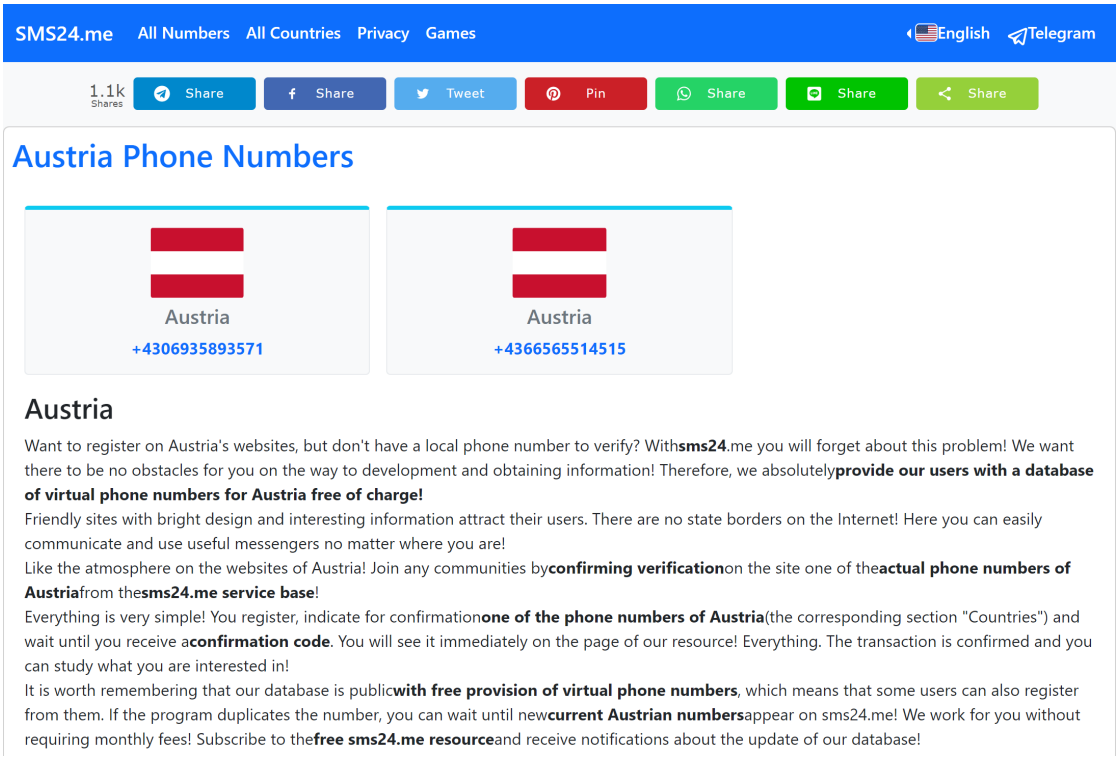


Figure 2.2: Example of a website advertising public SMS inboxes with Austrian phone numbers.

[1] See `https://www.burnerapp.com/`
[2] See `https://sms-rent.com/`

3 Related work

We use the literature review to determine what is already known about the topic and to identify research gaps. It is a process to gather information from numerous sources related to that topic (Webster and Watson, 2002).

3.1 Sources

We use a three-step approach according to Webster and Watson (2002) to identify the sources for our literature review.

3.1.1 Catalog search

Our topic is the ecosystem of public SMS inboxes. To exploratively look at it, we broaden the search by dividing the topic into three main concepts: „Public“ refers to access to the service for everyone, „SMS“ is the protocol used, and „inbox“ scopes to the inbound messages. We link these concepts by AND operators in our search string. Prior research might use different terms, so we also search for synonyms. We link synonyms for each concept block by OR operators. The search terms might have different endings, so we truncate them to root words.

Concept 1	**Concept 2**	**Concept 3**
public	SMS	inbox
free	phone	gateway
disposable	number	service
temporary		receive
		website

Table 3.1: Concepts and their synonyms used for catalogue search.

The resulting search string „public OR free OR disposable OR temporary AND SMS OR phone OR number AND inbox AND gateway AND service AND receive AND website“ was applied to the four catalogues ACM DL[1], BibSearch Universität Innsbruck[2], Google Scholar[3] and Web of Science[4].

[1]See `https://dl.acm.org/`
[2]See `https://bibsearch.uibk.ac.at/primo-explore/search?vid=UIB`
[3]See `https://scholar.google.com/`
[4]See `https://www.webofscience.com/`

Of each catalogue, we skimmed the first 100 results by title. If the title seems relevant, we skimmed the abstract, and if that seems relevant, the paper. The search yielded four papers that are related to the topic of public SMS inboxes or deal with aspects of it:

- „Characterizing the security of the SMS ecosystem with public gateways“ (Reaves et al., 2018)
- „Sending out an SMS: Characterizing the Security of the SMS Ecosystem with Public Gateways“ (Reaves et al., 2016)
- „Characterizing the Security Threats of Disposable Phone Numbers“ (Cheng et al., 2020)
- „On the Exploitation of Online SMS Receiving Services to Forge ID Verification“ (Berenjestanaki et al., 2019)

3.1.2 Backward citation review

In the second step, we review the citations of the articles identified in 3.1.1 to determine possible related articles that were published prior. Unfortunately, this review of citations yielded no further relevant articles for our topic.

3.1.3 Forward citation review

We identify publications that cite the papers defined in step 3.1.1. To find these citations, we review only the catalogues „Web of Science“, and „Google Scholar“, as they are the only ones to provide this functionality. This review resulted in no additional papers related to public SMS inboxes.

3.2 Concepts

To synthesize the literature, we use a concept matrix (table 3.2) according to Webster and Watson (2002). A concept matrix helps to identify the research conducted and the coverage of topics in each paper. Further, it enables us to compare the documents on a concept basis.

Concepts	Papers			
	(Reaves et al., 2018)	(Reaves et al., 2016)	(Cheng et al., 2020)	(Berenjestanaki et al., 2019)
Methodology				
Web Scraping SMS messages	✓	✓	✓	✓
Dataset description	✓	✓	✓	✓
Third-party data sources	✓	✓	✓	
Ethical considerations	✓	✓	✓	
Algorithmic classification	✓	✓		✓
Limitations	✓	✓		
Analysis				
Characteristics				
Domain location and registration date	✓	✓	✓	
Monetization	✓	✓		
Shared numbers	✓	✓	✓	
Number similarity	✓	✓		
Networks	✓	✓	✓	
Geography	✓	✓	✓	
Activity	✓	✓	✓	
Lifetime	✓	✓	✓	
Usage				
Intent	✓	✓		✓
Senders			✓	
Language	✓	✓		✓
Abuse in SMS	✓	✓		
Use cases	✓	✓	✓	✓
Longitudinal study	✓			
Privacy leakage				
PII & sensitive information	✓	✓		
PII leakage via connected services	✓	✓	✓	
One-time passwords	✓	✓		✓
Prevention				
Probing detection			✓	
Possible countermeasures	✓	✓	✓	✓

Table 3.2: Literature review - concept matrix.

3.3 Review

We identified a total of 4 papers, whereby the paper from Reaves et al. (2018) is merely the same as Reaves et al. (2016), except for the analysis on a dataset that is double the size of the study published in Reaves et al. (2016). Therefore we refer in the following review only to the Reaves et al. (2018) paper since it includes all findings from the Reaves et al. (2016) article, leading us to a total set of 3 papers that we review in this part.

The authors framed their work differently and interpreted the results in different contexts. Reaves et al. (2018) use the study of public SMS inboxes as a proxy to generalize the findings for the whole SMS system. On the other hand, Cheng et al. (2020) interprets the results only in the context of public SMS inboxes. Berenjestanaki et al. (2019) uses the gathered data only to study to what extent public SMS inboxes are used to bypass ID verification.

Noteworthy is that no standard term for the phenomenon of public SMS inboxes exists yet. Reaves et al. (2018) denotes them as „Public SMS Gateways“, Cheng et al. (2020) as „Temporary phone numbers“, and Berenjestanaki et al. (2019) as „SMS Receiving Services“. Nevertheless, all of them describe the public SMS inboxes in a way we defined them in the introduction, so we can conclude that they are referring to the same research subject.

3.3.1 Methodology

Web-scraping SMS messages All three studies scrape a list of websites for phone numbers and messages in public SMS inboxes at regular intervals over a given period. All three studies compared the messages or the hash of the message to prior scraped messages to avoid duplicates. Reaves et al. (2018), and Cheng et al. (2020) mention the challenge of converting relative timestamps and precision loss. Berenjestanaki et al. (2019) informs that senders' phone numbers or alphanumeric identifiers are partially masked (e.g. +491523XXXXX). Cheng et al. (2020) limit their scraping infrastructure to obtaining two sites from the same domain to 5 to 15 seconds to limit the load on the web server for ethical reasons.

Dataset description All four papers ground their study in a large set of actual SMS messages obtained from public SMS inboxes. Berenjestanaki et al. (2019) collected a dataset of 905,931 SMS messages related to 1,010 phone numbers by scraping 18 websites over three months. Reaves et al. (2018) collected a dataset of 900,655 messages related to 600 phone numbers over 28 months by scraping eight websites. Cheng et al. (2020) gathered a set of 30 million messages and 4,669 phone numbers from 9 different websites, mainly from China, over 12 months.

	Papers			
	(Reaves et al., 2018)	(Reaves et al., 2016)	(Cheng et al., 2020)	(Berenjestanaki et al., 2019)
Period	28 months	14 months	12 months	3 months
Years	[2015-2018]	[2015-2016]	[2019-2020]	[2018-2019]
Websites	8	8	9	18
Phone numbers	625	400	4,669 (2.782 distinct)	1010
Messages	900,655	386,327	30,000,000	905,931

Table 3.3: Comparison of datasets used in previous studies.

Third-party datasources For analyzing the data, Reaves et al. (2018) use APIs from Twilio for phone number lookup on the type (landline, VoIP, mobile), origin country, and network. Additionally, they use OpenCNAM for mapping phone numbers to caller IDs. They analyzed URLs in messages with the help of VirusTotal. Cheng et al. (2020) use Google's open-source phone number database for the location of phone numbers and ip138 for location information on Chinese phone numbers.

Ethical considerations Reaves et al. (2018) argues that for the user, it is clear that the gateways are public, and the use of it is „opt-in". So privacy concerns arise for institutions or individuals that are sending SMS messages to public SMS inboxes and are not aware of it. As it is unlikely that the analysis of institutional messages might cause harm, they are analyzed in the study, but personal messages are not. They state that since personally identifiable information was found, the dataset is not published not to prolong this information's exposure. Additionally, they do not take advantage of this information, e.g. by attempting to access accounts. Regarding the combination with other data sources, they argue that the study does not make it possible to identify individual users of a public SMS inbox or the operator of such a website. Cheng et al. (2020) performs a case study on privacy leakage in online travel agencies. Therefore they access accounts registered with numbers and OTPs from public SMS inboxes. Considering ethical aspects, they state not to disclose obtained personal data or alter data on these platforms.

Algorithmic classification Two studies enact algorithmic classification on the data. Reaves et al. (2018) are automatically clustering messages and labelling them based on message intention. Berenjestanaki et al. (2019) sets up an algorithm to determine if a message contains a verification code independent of language.

Limitations Only Reaves et al. (2018) elaborate on the limitations of their study. Since users might be aware of the publicity of the SMS messages, the findings on sensitive data in SMS messages might be underestimated. Also, because numbers change on public SMS Inbox websites regularly, users may be unlikely to use them as a second factor in two-factor authentication schemes.

3.3.2 Analysis

Characteristics

Domain registration date and location Cheng et al. (2020) and Reaves et al. (2018) analyzed the registration date and location from the WHOIS entry of the respective domain. We must note that WHOIS information is not reliable. While cross-checking the WHOIS creation date of a domain with its first occurrences in the Wayback Machine, we found discrepancies. The reason is that the creation date only refers to the creation of the WHOIS entry, not the domain itself. So the WHOIS creation date might be more recent than the actual date the domain registration happened. Also, the location information in WHOIS entries is unreliable, as domain registrars do not validate geographic information provided by the user who registers the domain (Watters et al., 2013).

Monetization Reaves et al. (2018) speculate that the websites' monetization is mainly based on ad revenue and some selling private numbers and phone-verified accounts.

Shared numbers Reaves et al. (2018) found that some gateways share numbers, which indicates collaborating parties. Cheng et al. (2020) state that 40% of numbers in the studied public SMS inboxes are shared. They conjecture that this is due to collaborating parties, that some websites crawl other websites in real-time, or that numbers are assigned at an underlying platform upon which public SMS inbox websites are built.

Number similarity Reaves et al. (2018) found that in six of the eight studied websites, more than 40% of phone numbers were similar with a Hamming distance of two or less. More than 80% of these similar numbers were mobile numbers, not VoIP.

Networks Reaves et al. (2018) identified 65 networks, of which are 55 mobile, seven VoIP, and three landlines. The landlines could be mislabeled. Cheng et al. (2020) analyzed only Chinese phone numbers, of which 78% are issued by virtual phone networks, which resell services but do not operate telecommunications infrastructure.

Geography Reaves et al. (2018) identified 30 different country prefixes in the phone numbers, of which the US prefix represents the highest share. Via CNAM, they could get some location data on the US-based numbers, indicating that their location is spread around the country and not in large population centres. Cheng et al. (2020) identified 44 different country prefixes, where US and China prefixes accounted for 61% of the numbers.

Lifetime Reaves et al. (2018) analyzed that the median number lifetime is 20 days, and 74% of numbers do not last longer than a full billing cycle of 31 days. They speculate on two reasons for this short lifetime. First, numbers need to be replaced as they are used too often at services and are not useful anymore. Second, networks might shut off the line due to high message volume. Cheng et al. (2020) analyzed that the average lifetime of numbers is 33 days with a median of 24 days.

Activity Reaves et al. (2018) analyzed that the activity on a phone number peaks early in the lifetime. Cheng et al. (2020) concluded from their data that there is more activity during the day than at night.

Usage

Intent Reaves et al. (2018) clustered the messages according to the intent of the message. They found 67.6% contained a code for verification and authentication purposes, 7.6% contained a One-time password, 1.3% were password reset messages, and 0.8% were test messages. Berenjestanaki et al. (2019) found that 82% of messages include a code or one-time password.

Senders Cheng et al. (2020) analyzed, that 20 senders account for 33% of the total messages.

Language Reaves et al. (2018) used Google's „langdetect“ library and found the six top languages to be English, Russian, Portuguese, German, French, and Spanish, making up 80% of all messages. English messages account for 67%. In the study from Berenjestanaki et al. (2019), 62,59% of messages were written in English

Abuse in SMS By using cluster analysis and a spam classifier, spam campaigns from financial services, job search websites, and SMS gateways advertising their services were found by Reaves et al. (2018). In total, between 2-3% of messages were identified as spam. The maliciousness of URLs in SMS messages was probed. 5.4% of links in Russian language SMS were malicious, followed by 2.1% of URLs in English messages (Reaves et al., 2018).

Use Cases No in-depth analysis is given on the use cases of public SMS inboxes, and only some observations are made by the authors that provide hints but not a complete picture. They quote the exploitation of benefits for newly registered accounts, creating multiple accounts for fake reviews and ghost followers, and preserving privacy for pornsite users (Cheng et al., 2020). Reaves et al. (2018) observed the case of registrations for VoIP services to Cuba. Public SMS inboxes are generally used to circumvent the identity verification step. They are less likely used for login since numbers change and may not be usable after some time (Berenjestanaki et al., 2019). The evidence for phone-verified account evasion is backed by message activity skewed to the early phase of the lifetime of a phone number and that users and messages are in different locations (Reaves et al., 2018). Further, it was observed that public SMS inbox numbers are used to sign-up and obtain telephone numbers for free at VoIP providers, so-called „number chaining“ (Thomas et al., 2014), thereby further facilitating PVA evasion (Reaves et al., 2018).

Longitudinal study Reaves et al. (2018) performed a longitudinal analysis from 2016 to 2018. They found no significant change in the data over time.

Privacy leakage

PII and sensitive information Reaves et al. (2018) discovered several kinds of personally identifiable information or other sensitive information in the messages. Two virtual credit card providers, „Paytoo“, and „iCashCard“, send the credit card numbers and CCV2 codes over SMS. Also, the phone number and credentials sent via SMS could be used to login into these services. Several other instances of financial information (credit card numbers, IBAN, names, ...) were found. Some services distributed passwords via SMS, allowing users to authenticate with the corresponding phone number and password. By obtaining access to SMS messages, one could log into these services. Password reset links that allowed taking control of accounts were sent via SMS. Further, alerts and status reports, addresses, zip codes, and email addresses were present in the messages.

PII leakage via linked services Often it is made use of short links in SMS due to character limitations. Some of these URL shortening services leak the IP address of the user clicking on these links (Reaves et al., 2018). Several Chinese online travel agencies allow authentication with a phone number. Logging into accounts created with numbers from public SMS inboxes revealed that more than 83% of the accounts contained personal data, like name, gender, phone, birthday, ID, and booked journeys (Cheng et al., 2020).

One-time passwords and codes Most of the SMS messages in public SMS inboxes contain one-time passwords and codes. 35% of codes have four-digit length, 52% a 6-digit length (Berenjestanaki et al., 2019). Reaves et al. (2018) analyzed the entropy and found three services with significantly low entropy.

Prevention

Probing detection Cheng et al. (2020) is the only group that experimented with mapping out the detection rate of disposable phone numbers at internet services. They probed 76 of 100 services that appeared most often in their dataset of messages from public SMS inboxes. In total, they could log in with 47% of the phone numbers, with varying detection rates among these services. Furthermore, they found that 46% of the probed 77 services did not block any accounts using these phone numbers. 7% of the services generally block phone numbers from virtual networks. They observed from displayed messages that these accounts were detected by „abnormal activity“.

From our point of view, this experiment has several limitations. First, it is unclear if their probing of the „login“ is related to the account creation/registration or authentication phase. A probing of the account creation phase could yield results in the detection rate of the phone numbers, whereas a probing of the login phase implies a successful creation of an account with a phone number of a public SMS inbox. Second, suppose we assume that Cheng et al. (2020) probed the login phase as we interpret the description of the experiment. In that case, it is unclear to which extent the use of a phone number from a public SMS inbox contributed to the detection of this account, as they described the reasons for detection due to „abnormal account activity“. Related to this, it is also ambiguous if this abnormal activity resulted from the trial to log in with different devices / IP addresses by the researchers or from the activity of the account owners within these services. Third, as this study only probes services that most often sent SMS to public SMS inboxes, there is no information on services that may block these numbers so effectively that they do not send any messages to public SMS inboxes and therefore do not appear in the top list in this dataset. Fourth, there is no information on the lifetime of the probed numbers, which we assume correlates with the detection rate since the longer a number exists, the more likely it has been used before.

Possible countermeasures Several prevention mechanisms are discussed in the three papers. Performing a number lookup and blocking based on the phone network might be cost-intensive and lead to high false positive rates (Reaves et al., 2018; Cheng et al., 2020). Blocking numbers based on similarity might be feasible since 86,4% of mobile numbers are similar with Hamming distance of 2 or less. However, cloaking similar numbers still leads to a high false positivity rate (Reaves et al., 2018). Another possibility is to re-verify the phone number proposed by Thomas et al. (2014) in the context of PVA evasion. The problem is that half of the public SMS inbox numbers are available for up to 24 days (Reaves et al., 2018). Promising is the idea of a phone reputation service that shares abuse data between service providers (Reaves et al., 2018). Similar is the method to check the existence of a phone number in a public SMS inbox, supposedly by web-scraping (Berenjestanaki et al., 2019). Another approach is that the user sends a code to the service since the public SMS inboxes cannot send SMS (Cheng et al., 2020). Finally, a possible method is to limit the number of accounts associated with a phone number or detect anomalies in account behaviour (Cheng et al., 2020). This is probably already enacted by most services but can not prevent the creation of at least one account

per service per phone number from a public SMS inbox.

3.4 Identified research gaps

Concluding, the three, respectively four studies cover many different aspects of public SMS inboxes and give a detailed view of the topic. Nevertheless, still, more research is needed on this topic. Therefore, we identified several research gaps that might be valuable for the research community to investigate.

There needs to be an estimation of the scale and impact of public SMS inboxes. e.g. how many of these websites exist, how popular are they, how many messages do they receive, how many accounts are affected, what is the fraud potential and what damage do they incur on services?

What are the characteristics of public SMS inboxes concerning their monetization schemes, legitimacy, privacy schemes, and operations? Cheng et al. (2020) and Reaves et al. (2018) observed that public SMS inboxes share numbers, indicating collaborating parties. To what extent and how do they share the numbers, e.g. by scraping or shared gateways? Prior studies did not probe the public SMS inbox by sending SMS to it to see how reliable they work.

What is the source of these numbers, and what gateways are behind these websites? For example, are specific mobile networks more affected than others? Are these phone numbers also used in other means of communication, e.g. for outbound calls and sending messages?

What legitimate and illegitimate use cases for public SMS inboxes exist? Legit use cases are not considered in the existing literature. For illegitimate use cases, only case studies are given. Possibly, more illegitimate use cases can be found and systemized. The extent of legitimate and illegitimate use is not quantified yet. How does the public SMS inbox facilitate cybercrime as part of its value chain?

Regarding privacy leakage, only case studies on what kind of information is leaked by public SMS inboxes are given in the literature. It has yet to be researched if cybercriminals actively exploit this leaked information. An experiment can be set up to investigate this.

The one study on prevention mechanisms against public SMS inbox numbers at the service level is limited. Therefore, probing services with public SMS inbox numbers in the account creation phase might be a feasible experiment. However, since services send SMS to public SMS inboxes, we can conclude that no effective prevention mechanisms are in place at these services.

Prevention mechanisms are proposed in the literature, but each comes with limitations. The instrument of denylists is not investigated. Services exist that provide denylists or reputation scores for phone numbers. However, their effectiveness against public SMS inboxes needs to be examined.

Correlating the collected data on public SMS inboxes with third-party data sources could yield more insights into the ecosystem. For example, to what extent are public SMS inboxes related to SIM Box fraud and spam calls? Home Location Register (HLR) can be queried to get the SIM status of a phone number, roaming status, and which network

it is connected to. Telecoms can be contacted for further insights, e.g. if they know the phenomenon and which prevention mechanism they enact. Law enforcement could be contacted to investigate if these numbers are related to crimes, especially cybercrime schemes. Finally, internet services can be contacted to determine to which extent public SMS inbox facilitates abuse of their offerings (e.g. fake account creation, bots, and spam) and further to distinguish legitimate or illegitimate use of public SMS inbox.

4 Methodology

4.1 Research questions

This research aims to understand the ecosystem of public SMS inboxes better. As prior research exists, we want to broaden the scientific community's knowledge, building upon prior findings. Our analysis is based on our measurements of the research subject public SMS inboxes. Previous studies took a similar approach, and the possibility of reproducing results exists. Therefore we concentrate on the identified research gaps in 3.4 and deduct our research questions from it.

Our leading research questions are:

- What are the characteristics of websites providing public SMS inboxes?
- How are they used?
- What threats come by using these services?
- What approaches can be made to mitigate these threats?

4.2 Research design

We ground the evidence of our findings in data. For most aspects, this data are SMS messages collected by web-scraping from public SMS inboxes. To facilitate reproducibility and replicability for fellow researchers, we designed our web-crawling study according to the criteria of Demir et al. (2022).

		ID	Criterion	Fulfillment	Reference
Dataset		C1	State analyzed sites	✓	4.4
		C2	State analyzed pages	✓	4.4
		C3	State site or page selection	✓	4.4
		C4	Perform multiple measurements	✓	4.4
Experiment Design	Crawler	C5	Name crawling tech	✓	4.4
		C6	State adjustments to crawling tech	✓	4.4
		C7	Describe extensions to crawling tech	✓	4.4
		C8	State bot detection evasion approach	✓	4.4
		C9	Used crawler is publicly available	✓	4.6
		C10	Mimic user interaction	not applicable	
	Env.	C11	Describe crawling strategy	✓	4.4
		C12	Document a crawl's location	✓	4.4
		C13	State browser adjustments	not applicable	
		C14	Describe data processing pipeline	✓	4.4
Eval.		C15	Make results are openly available	✗ on request	4.6
		C16	Provide a result/success overview	✓	4.5
		C17	Limitations	✓	4.7
		C18	Ethical discussion	✓	4.9

Table 4.1: Criterias used to design the methodology of this study. According to Demir et al. (2022).

4.3 Identification of public SMS inboxes

To analyse the characteristics of public SMS inboxes, a dataset on URLs of public SMS inboxes is necessary. We use five data sources for URLs to identify possible websites: Google Search, Bing Search, Tranco List, Reaves et al. (2018) research paper, and Cheng et al. (2020) research paper.

Search engines For Google and Bing Search Engine, we composed five Search strings: „receive sms“, „receive sms free“, receive sms online, „temporary sms“, „temporary sms free“, and „public sms inbox“. The search query was executed in a python script using serpapi.com on 12.03.2022. The top organic results were scraped for each search query

and search engine. The API returned between 40 and 50 results per search query. For each result, the corresponding search string, search engine, target URL, and position in search results were saved in a CSV file. The result is a dataset of 234 URLs for Bing Search and 296 URLs for Google search.

Research papers We converted the list of domains given in research papers into a CSV file, with ten domains from Cheng et al. (2020) and eight domains from Reaves et al. (2018). Unfortunately, we did not consider the paper of Berenjestanaki et al. (2019) since we were unaware of it at the moment of this data collection.

Tranco List The Tranco list is a „Research-Oriented Top Sites Ranking Hardened Against Manipulation“ (Le Pochat et al., 2019). We use the Tranco list[1] generated on 9.03.2022 and filtered for entries containing the string „sms“, resulting in a list of 456 domains.

Labeling We combined all URLs from the five data sources in a list of candidate URLs. Then, we manually labelled this list with the help of a python script according to the following workflow:

1. Has the domain already been labelled?

   Yes: Skip this domain, and continue with the next domain at step 1.

   No: Continue with labeling

2. Does the domain offer an active public SMS inbox?

   (Public SMS inboxes that did not publish an SMS message in the last days were regarded as inactive.)

   Yes: Continue with labelling, add a sample URL of an inbox from this website.

   No: Label as isInbox = false, and continue with the next domain at step 1.

3. What type of monetisation has the website?

   (Multiple values possible)

   „Generic Ad“: Website displays generic ads from ad networks that are not related to SMS services.

   „Temporary Phone Number Ad“ Website displays ads for the lease of a temporary phone number (paid service).

   „Temporary Phone Number Ad“ Website displays ads for the lease of a temporary phone number (paid service).

   „Premium Service“ Website offers „premium“ SMS inboxes that are only accessible through one-time payment or subscription.

[1] Available at `https://tranco-list.eu/list/2L69/full`

„None“ No monetization recognizable.

4. Does the website have Terms of Service?

   Yes: Add URL of Terms of Service.

   No: Continue.

5. Does the website have a privacy policy?

   Yes: Add URL of privacy policy.

   No: Continue.

Through manual labelling of 1003 candidate URLs, we detected 72 unique domains that offer a public SMS inbox. For a complete list, reference A.1 in the Appendix.

4.4 Web scraping SMS messages

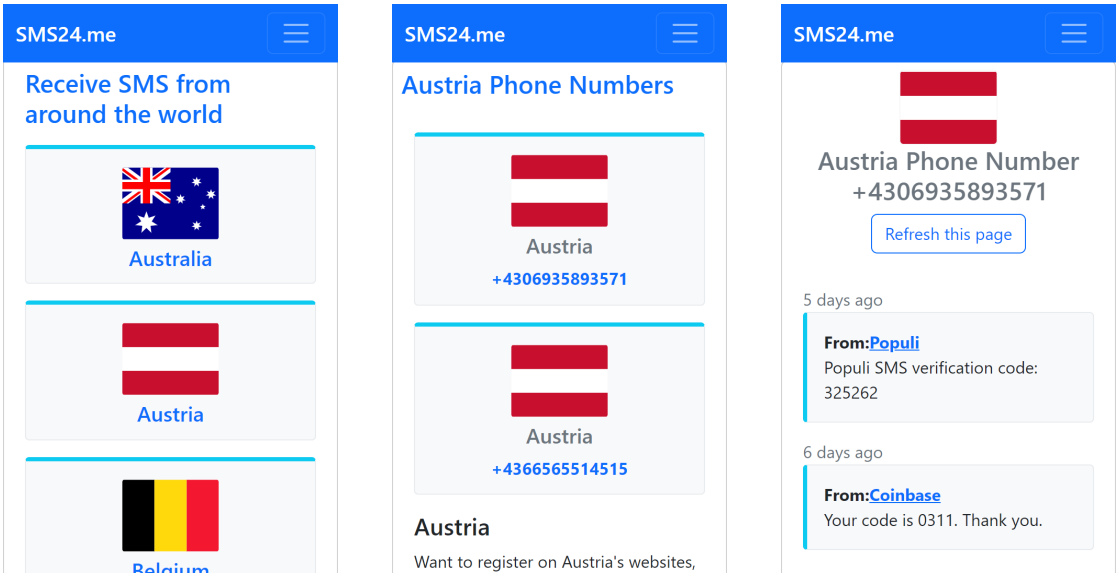
To analyse different aspects of public SMS inboxes, we collect a dataset of SMS messages with the help of web scraping from a set of websites. We focused on the highest ranked (according to Tranco ranking) domains of our list of SMS inboxes. Unfortunately, it was impossible to scrape the top-ten domains due to different bot-detection and prevention measures enacted by the websites. Nevertheless, we successfully coded scraping for ten websites.

Website
receive-smss.com
receive-sms-free.cc
sms24.me
receivesms.co
7sim.net
online-sms.org
receivesms.live
receivesmsfast.com
receivesms365.com
tesms.net

Table 4.2: List of scraped websites.

Implementation We built the program for web scraping in a Python environment with the help of the Scrapy[2] framework. We programmed a generic „Spider“ to crawl the given websites and extract the data. For each website, CSS selectors needed to be specified that get the corresponding HTML element content.

[2]See `https://scrapy.org/`



(a) Example of a country overview.

(b) Example of an inbox overview.

(c) Example of an inbox.

Figure 4.1: Example structure of a public SMS inbox website (`sms24.me`).

Websites of public SMS inboxes have a typical structure: A list of countries is given (see Fig. 4.1a). When clicked on a country, all inboxes (phone numbers) corresponding to this country are shown (see Fig. 4.1b). When clicked on a specific phone number, the inbox with the received SMS messages corresponding to the numbers is displayed (see Fig. 4.1c). The inbox lists SMS messages that display the messages' date, sender, and body. Some of the websites have pagination of the lists of countries, inboxes, or messages. However, some websites do not have an overview of countries, only a view of all available numbers.

Our scraping script starts at the list of countries and collects all links to the lists of inboxes. If no country overview exists, the script begins with the overview of all inboxes. Next, all links to the inboxes are collected from the overview of all inboxes. Next, the messages are scraped from every inbox, and if necessary, it navigates through the pagination.

Only newer messages than the last scrape timestamp are scraped and saved to reduce the amount of data. Otherwise, many duplicates would exist in the database. In addition, no concurrent requests are made per website to reduce the load on scraped websites.

By navigating through the link tree of the website for every crawl, we can detect the appearance of new numbers and the abandonment of numbers.

Bot detection We used one static User-Agent string to mock a Chrome browser in the requests toward the servers.

```
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36
```

```
(KHTML, like Gecko) Chrome/58.0.3029.110 Safari/537.36
```

While setting up the web scraping program, we encountered numerous bot detection and prevention measures. Sites that enact bot detection and prevention, which we recognised in the phase of developing the web scraping, were not considered sources for our data collection. Setting up a mechanism to circumvent bot detection would involve using captcha solvers, rotating proxies, or emulating browsers that would require additional financial and computing resources and more time for coding.

At `smsreceivefree.com` the page is protected by Cloudflare, and a Captcha needs to be solved to access the site, independent of whether it is accessed via a browser or a simple HTTP request. `temporary-phone-number.com` can be accessed without restrictions, but when more than 100 requests with the same IP are made, random SMS messages are returned instead of the originals. `freereceivesms.com`, `receive-sms-online.info`, `receive-sms.cc`, `quackr.io` and `pingme.tel` has some bot prevention from Cloudflare that depends on the characteristics of the browser. Plain HTTP requests with User-Agent Strings are blocked. That may be circumvented by emulating a browser for scraping. At `mytempsms.com`, the telephone numbers are not displayed as plain text but rendered as images. That would require OCR to scrape the numbers. `freephonenum.com` has no observed bot-detection but a retention time of only 10-20 minutes for SMS messages. That would require scraping the website in a much shorter interval.

Data Processing pipeline After scraping the data from the websites, some data transformation is necessary. The timestamps of messages are often relative. Absolute timestamps of messages are derived from relative dates (5 min ago). Precision loss (5”00’ or 5”59’?) is accounted for by calculating a message’s earliest or latest possible timestamp. Absolute URLs are derived from relative URLs (e.g. /inbox → example.com/inbox). Phone numbers are often not depicted in plain international format (e.g. „+1 234 / 00“). Instead, they are reformatted to standard international format (+12345) without any characters except numbers and the leading +.

Data fields For each scraped message, a row is saved in a table in a MySQL database table. The following fields are saved:

Field	Description
scrape_start	Unix timestamp when the current scrape run for this website started
website	Domain of the public SMS inbox website that is scraped
country_page	URL of the page where all available countries are listed, parent of the scraped message
country_name	Name of the country as depicted on the country_page
inbox_page	URL of the page where available inboxes/numbers are listed. Parent of the scraped message
inbox_number	Phone number of the inbox as depicted on the page
message_page	URL of the page from which the message is scraped
message_scrape_timestamp	Timestamp when the message is scraped
message_number	Phone number of the message as depicted on the message_page
sender	Sender of the message as depicted on the message_page
body	Message content as depicted on the message_page
date	Date of the message as depicted, often relative, e.g. „5 minutes ago“
timeframe_earliest	Timestamp calculated from relative date, earliest possible value
timeframe_latest	Timestamp calculated from relative date, latest possible value
max_message_age	Timestamp of the last scrape run, the parameter used to scrape only messages newer than the last scrape

Table 4.3: Description of saved data fields.

Infrastructure The scraping program is deployed on a virtual machine at DigitalOcean[3] in a german datacenter. The results are saved into a managed MySQL database instance running in the same data center. The script schedules itself to run every 55 minutes after the last run. Instead of a smaller interval, this interval was chosen to reduce the load on the scraped websites compared to a smaller interval while also ensuring a precise timestamp derivation. In comparison, an interval larger than 1 hour would lead to an uncertainty of +- one hour.

Duration The scraping process started end-March 2022 and stopped in mid-June due to an unknown error. A succeeding thesis continues the scraping process, so longitudinal

[3] `digitalocean.com`

data beyond this point will be available. This study analyses a subset of the data from 1st April to 31st May.

4.5 Data validation

For further analysis, we processed the raw database of scraped SMS messages.

We omitted all messages older than 1st April 2022 and those newer than 1st June. Duplicates of the same message exist, as it is technically possible that a message is scraped twice by following scrape runs. Thus, we deduplicated our dataset and removed 6,985 rows. This process only removed duplicates of messages that were scraped twice. The dataset still includes duplicates because some messages were sent to a public SMS inbox multiple times.

The data is validated manually for empty fields. We encountered two runs where the field „message_number“ was empty. We recovered the value via the „inbox_number“ field. Six rows are missing the „sender“ field. 2946 rows were missing values in the „body“ field. In 4,2917 rows, the „body“ field also contained the strings from the „date“ and „sender“ fields. We corrected these rows by removing the unsolicited parts. 89,458 rows wrongly contained Javascript related to Google ads in the „body“ field, which is placed between messages on the website and was wrongly scraped. We deleted these rows since they did not contain any information.

We checked if our scraping infrastructure consistently scraped every public SMS inbox every hour. We found that for four websites, we have data for every interval. For four websites, we miss data for precisely one interval, and for one website, we miss data for two intervals. Missing intervals may be caused by the inability to reach the server. For one website, we miss data for 248 intervals. Further investigation revealed that this website does receive only a few messages per hour and that, for some intervals, no messages were received, leading to no entry in the table. We can conclude that our scraping infrastructure reliably collected the messages in the public SMS inbox for over two months without outages. The average time between two scraping runs was 63 mins and 53 seconds.

Website	Missed intervals
receive-smss.com	0
receive-sms-free.cc	1
sms24.me	0
receivesms.co	0
7sim.net	1
online-sms.org	0
receivesms.live	1
receivesmsfast.com	1
receivesms365.com	248
tesms.net	2

Table 4.4: Number of intervals for which websites miss data.

Due to cloaking Invernizzi et al. (2016) and bot detection, we need to find out if the contents scraped by us are the same as the user experience. So we cross-checked a sample of messages from our database whether they are displayed on the public SMS inbox websites in a browser with a different IP the same way. We did not find any discrepancies.

The final dataset used in the analysis contains in total of 13,812,753 messages. 2,047,546 of these messages occur more than once in a specific inbox. Compared to previous studies, the dataset is significantly larger than Reaves et al. (2018), but smaller than Cheng et al. (2020).

4.6 Tools and publication

Toolchain

For the analysis, we copied the MySQL database to a local instance to prepare the database independent of the scraping infrastructure and run performant queries. The SQL queries extract and aggregate the data needed for the corresponding aspect of the analysis, which is more performant than loading the whole dataset into a Python Pandas Dataframe. Third-Party Datasources, like APIs, are queried by a Python Script. The query results are saved into CSV files for reproducibility and then further analysed with Python Pandas[4].

[4]See `https://pandas.pydata.org/`

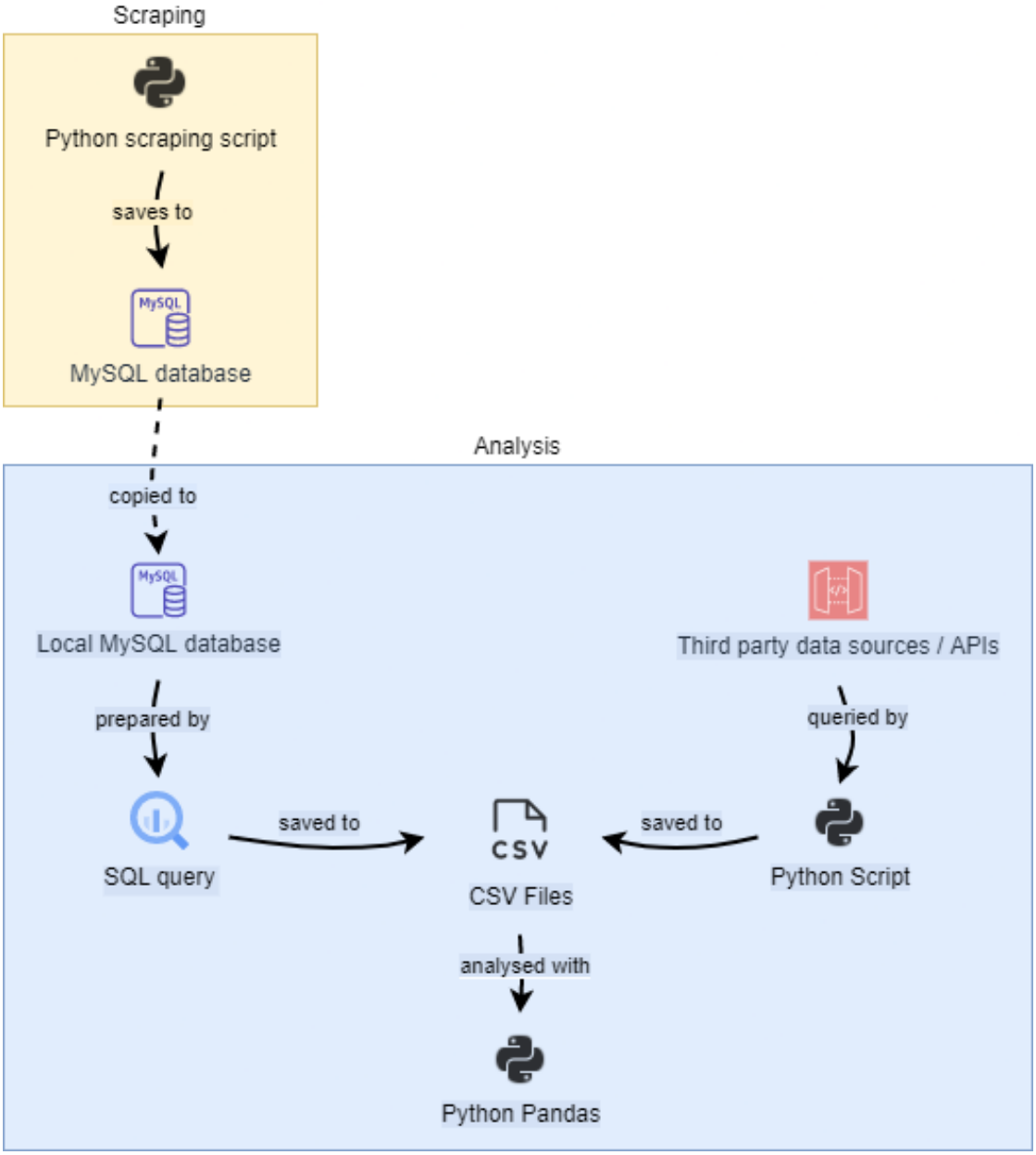


Figure 4.2: Toolchain used to analyse the data.

Publication

We publish the identification, web scraping, and data analysis scripts under an open-source license for fellow researchers as a public repository on GitHub accessible via `https://github.com/z8leo/master-thesis-public-sms-inboxes`. We publish the list of identified public SMS inboxes in this thesis and the corresponding public source

code repository. However, the dataset of scraped messages is not published but available for other researchers on request, as discussed in 4.9.

4.7 Limitations

The scraping interval of approximately one hour may lead to some SMS messages and inboxes not being captured. When an inbox is listed at times between our scraping runs, the messages and corresponding inbox phone number will not be scraped. Also, the messages after the last scraping run are not collected when an inbox gets delisted.

These current shortcomings could be mitigated by separating the crawler into three parallel threads. The first thread scrapes the page which lists the countries (country_page) in a shorter interval (e.g. 1 minute) and saves the URLs (inbox_page) to a separate table. The second thread scrapes all these URLs (inox_page) in a short interval and saves the URLs to the inboxes (message_page) in a separate table. The third thread scrapes the messages from all message_page URLs in a large interval (e.g. 1 hour). Message pages could still be reached and scraped, even when the inbox is not listed anymore.

4.8 Third-party datasources

Our research is facilitated by third-party data correlated with our web measurement data. We use data via the API products from IPQualityScore[5] for phone number reputation, HLRLOOKUPS[6] for intelligence on the phone number and „have I Been Pwned“[7] for information on data breaches.

We collaborated with the company Araxxe[8], which specialises in fraud prevention in the telecommunication industry, and provided us data on phone numbers related to „SIM-Boxers“.

Insights into the characteristics of the websites are provided by comparing the HTML source code of these and by qualitative content analysis of terms of use and privacy policies.

4.9 Ethical considerations

From a legal perspective, we intend to scrape publicly available data for research purposes. Users should be aware that any message sent to a public SMS inbox is eventually publicly available for some time. Regarding privacy and ethics, the collected dataset may contain personally identifiable information (PII), such as names, user names, addresses, or passwords. Furthermore, there is a risk that some phone numbers may originate from stolen SIM cards. The involved researchers will anonymise any detected PII whenever

[5] See `https://www.ipqualityscore.com/`
[6] See `https://www.hlr-lookups.com/`
[7] See `https://haveibeenpwned.com/`
[8] See `https://www.araxxe.com/`

possible. We will not use potential login credentials in text messages to access private accounts. Alternatively, to demonstrate attack vectors, accounts could be set up for staging by the researchers. The collected dataset will not be published publicly. However, for the reproducibility of results, they will be made available to other researchers upon a written request and a signed non-disclosure agreement. The corresponding ethical board review at the University of Innsbruck is in Appendix A.2.

5 Analysis

5.1 Characteristics of public SMS inboxes

5.1.1 Website characteristics

Similarity clusters

While manually labelling the public SMS inboxes, it was apparent that many sites are similar in their look and navigation. To quantify the similarity of the websites, we methodologically compare them by calculating a similarity score. The results help us uncover networks of websites or common technology among these websites.

We calculated the similarity of websites based on structure and style similarity according to Gowda and Mattmann (2016). As input for the calculation, a sample inbox site of the public SMS inboxes, gathered in the process of manual labelling, is used. The page source HTML code is downloaded, and a joint similarity score is calculated between each inbox with the python implementation[1]. The joint similarity is calculated with weight k = 0.3 for structural similarity and (1-k) for style similarity, as these values are recommended in the paper. Finally, the joint similarity between each website is structured into a similarity matrix for further analysis.

To uncover networks, we clustered the public SMS inboxes. We decided on the DBSCAN[2] algorithm, as it includes outlier removal and no minimal cluster size. Compared to other clustering algorithms, it fits our case since we assume some public SMS inboxes may be individual sites not belonging to a cluster, so there are outliers. Second, we expect different cluster sizes with a minimum member size of 2. Third, we can define the „eps“ value, that in our case, defines the threshold of the similarity score to decide when two sites are neighbours or not. We tried different eps values for our clustering and achieved good results with eps = 0.8 (See Fig. 5.1).

We identified a total of five clusters and attributed the similarity within the clusters to several different factors (Table 5.1).

[1] See `https://pypi.org/project/html-similarity/`
[2] See `https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html`

Cluster	# members	Description
-1	26	Outliers which do not belong to a cluster
0	25	These website use the Bootstrap Framework[3]. We did not identify a common template, but many sites look similar. Some sites may use others as inspiration or copy HTML code.
1	7	In cluster 1 are websites that use the „tSMS - Temporary SMS Receiving System - SaaS - Rent out Numbers“ [4] Template for WordPress from Envato market.
2	2	No source for mutuality identifiable
3	2	Consist of identical websites but serve different languages.
4	2	No source for mutuality identifiable

Table 5.1: List of identified clusters.

[3]See `https://getbootstrap.com/`

[4]See `https://codecanyon.net/item/tsms-temporary-sms-receiving-system/23244962`

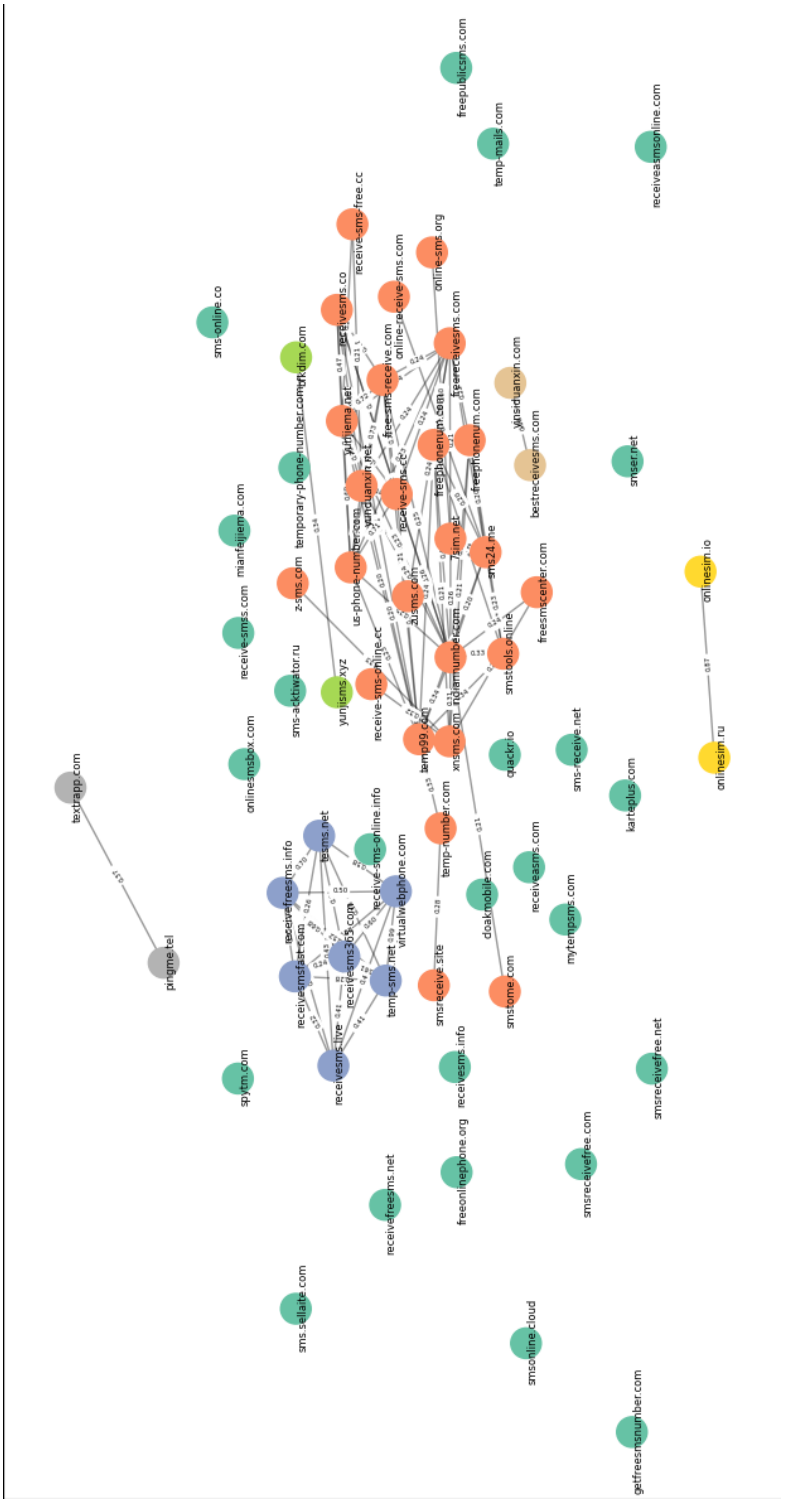


Figure 5.1: Cluster of websites based on similarity.

Each website is a node. Only edges where the similarity score is greater than 0.2 are displayed. Value on edges denotes similarity score. Each cluster has a different colour.

Monetisation

During labelling, we also evaluated which kinds of monetisation the websites used. With monetisations, we refer to how a website generates revenue from its users. We identified four schemes: „Generic Ads“, „Temporary Number Ads“, „Premium Services“, and „Gateway Services“. A website or public SMS inbox can have multiple types of monetisation that are not exclusive.

The most common type of monetisation is „Generic Ads“ which have 59 out of 71 public SMS inboxes. „Generic Ads“ are ads from Ad Networks like Google AdSense with no reference or relation to SMS services. This differentiates it from „Temporary Number Ads“, the second most common monetisation type, with 11 websites out of 71. „Temporary Number Ads“ are ads for services which enable the user to temporarily lease a virtual phone number to receive SMS. The difference to a public SMS inbox is that the number is dedicated to the user for a certain amount of time and that this service is not free.

Two public SMS inboxes offered a premium service with access to an SMS inbox for an extra charge. The difference is that this inbox is not free to use but is kind of public, as everyone can buy this premium service.

One website offered a paid SMS gateway service via an API to send and receive SMS.

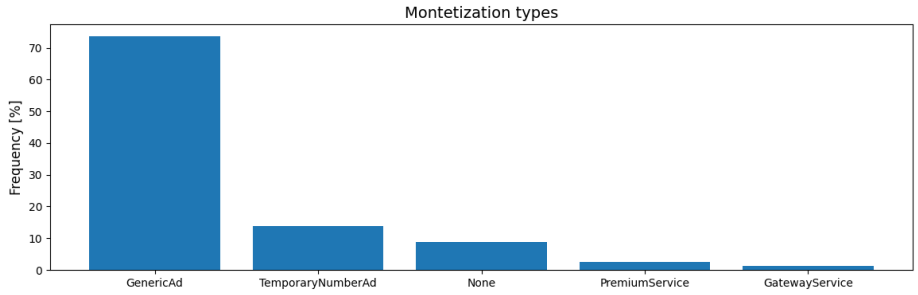


Figure 5.2: Frequency of monetisation schemes on a public SMS inbox website.

Privacy policies

The manual labelling of the public SMS inboxes contained the check for the existence of a privacy policy on the website. Of a total of 71 websites, we found 44 websites that have one and 27 websites that do not have one.

We considered further qualitative analysis (Mayring, 2015) of the privacy policies but realised that most of the privacy policies were generic. We identified only two privacy policies that had custom statements. We manually coded these statements under the research question „What statements regarding privacy do they make?“.

The two websites `temp-mails.com` and `receive-smss.com` warned the user not to receive sensitive information and that the content of the messages is available to all users. temp-mails.com further informed that only the last 100 messages are available and that the number of the sender will be partially masked (Appendix A.3).

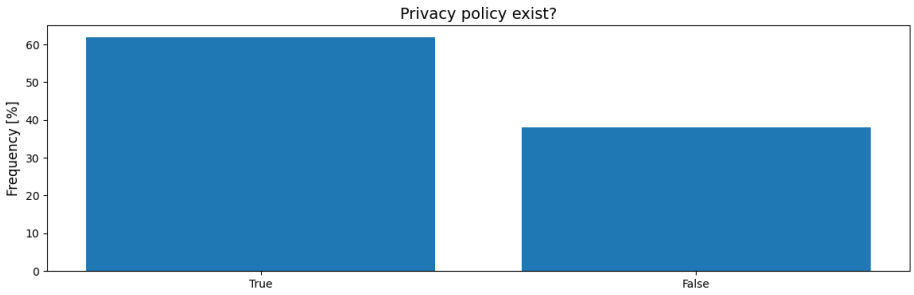


Figure 5.3: Prevelance of privacy policies.

Terms of Service

We conducted the same analysis with the terms of service and collected the information if a website has terms of service during the manual labelling. We found that 37 of the 71 websites provided terms of service.

We checked the terms of service for information regarding usage specific to public SMS inboxes. Like the privacy policies, most of these statements were generic, and only eight terms of service contained relevant information. We used qualitative content analysis (Mayring, 2015) to manually code the terms of service to determine the desired and undesired usage of public SMS inboxes according to the statements of these websites.

Regarding intended usage, most commonly, seven of the eight terms of service we coded mention using these inboxes for anonymisation of the user. Further, `smstools.online` mentions the desired use for exploiting promotions, circumventing advertisements, bypassing geo-blocking and creating multiple accounts. Also, use cases undesired by the public SMS inbox websites are mentioned. Five terms of service warn the user not to receive sensitive information. One website explicitly mentions the case of fraudsters possibly exploiting the information and informs that „[...]messages coming from the list of the forbidden addresses (such as bank/ credit organisations, electronic payment systems and similar projects[...] will not be shown on the page of the incoming SMS“ (Appendix A.4). Five websites forbid any „illegal“ use, and two websites the use of „automation“ without further clarification (Appendix A.4).

Desired Usage	**Undesired Usage**
Anonymisation of user	Receiving sensistive inforamtion
Bypassing geo-blocking	Illegal activity
Exploiting promotions	Automation
Multiple accounts	
Prevent advertisements	

Table 5.2: Desired and undesired usage according to terms of service.

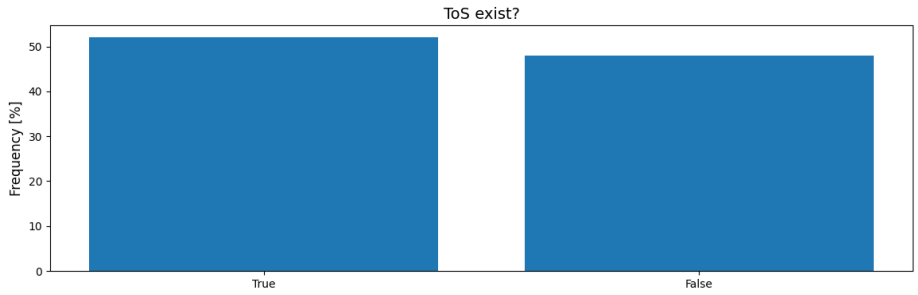


Figure 5.4: Prevelance of terms of service.

Popularity

The Tranco List is a „Research-Oriented Top Sites Ranking“ (Le Pochat et al., 2019) that gives us a rank which reflects the general popularity of domains. We studied whether the popularity of a public SMS inbox website depends on the number of inboxes they provide and the number of messages they receive. For this, we calculate Pearson's correlation coefficient and the corresponding p-value.

The values for the Tranco Rank are taken from the Tranco List dated 9.03.2022[5]. The value for the number of inboxes is calculated by counting the inboxes that a website provided in the timeframe of our dataset (1st April to 31. May 2022). The number of messages is calculated by counting the messages per website over the same period.

For this small statistical analysis, we state the following hypothesis that we will accept or reject with a significance level of $\alpha = 0.05$:

- H_0^0 Tranco Rank is not correlated with the number of inboxes
- H_1^0 Tranco Rank is correlated with the number of inboxes
- H_0^1 Tranco Rank is not correlated with the number of messages
- H_1^1 Tranco Rank is correlated with the number of messages

The R-values of -0.47 respectively -0.54 indicate a weak negative linear correlation. Unfortunately, the p-values are above our significance level, so we cannot reject our null hypothesis H_0^0 and H_0^1. Thus, we can not make a reliable statement if the number of inboxes and messages correlates to the Tranco Rank. (Table 5.3)

We must note that this experiment has several limitations. The Tranco ranks used for this calculation are tied to one day but change daily. Nevertheless, we assume that Tranco ranks do not fluctuate much and are pretty stable over time. Also, the number of inboxes collected over the distinct period may be a weak predictor since users might value other factors like the available countries or the freshness of phone numbers.

[5] Available at `https://tranco-list.eu/list/2L69/full`

Anyhow, the fact that the number of messages does not positively correlate to popularity, though we can not make a reliable claim here, sparks interest. We argue that a user visiting the website initiates one or many SMS messages to these inboxes. Consequently, more popular websites should get more users that initiate messages. We suspect that not all messages in public SMS inboxes are user-initiated. Possible reasons for that can be shared gateways (see chapter 5.1.2), spam and phishing campaigns (see chapter 5.3.4) or uptime (see chapter 5.1.2).

	r-value	**p-value**
Tranco Rank ∼ No. of inboxes (H^0)	-0.47	0.17
Tranco Rank ∼ No. of messages (H^1)	-0.54	0.10

Table 5.3: Pearson's correlation coefficients for Tranco Rank, total inboxes and total messages.

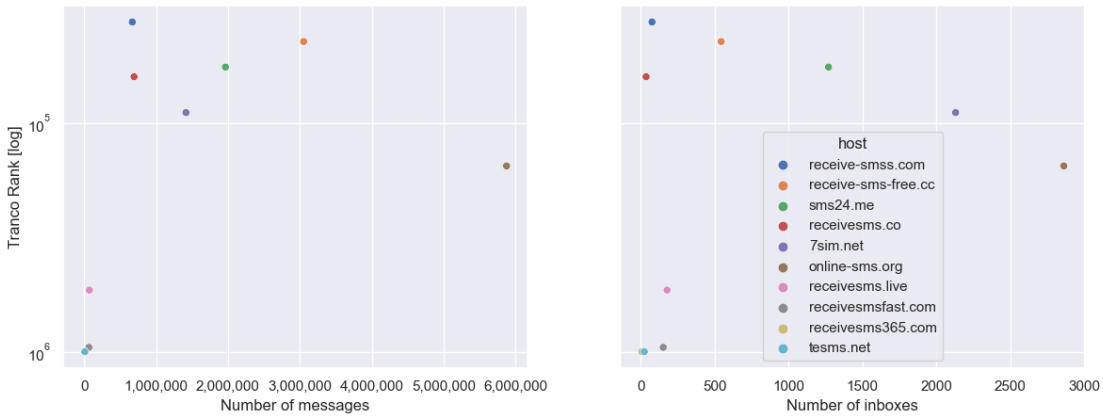


Figure 5.5: Scatterplot Tranco rank, number of inboxes and number of messages.

5.1.2 Inbox characteristics

Number lifetime and uptime

The phone numbers published on the public SMS inbox websites change over time. Inboxes, and respective numbers, appear and disappear on these websites. It is known that numbers are not offered anymore after a certain amount of time (Reaves et al., 2018). We refer to this as the lifetime of the numbers to describe how long they are used and when they are abandoned. Further, we analyse how often these numbers go on- and offline within their lifetime and refer to this as uptime. Our data shows an average lifetime of 23 days for the numbers, but with huge differences between and in the websites indicated by a standard deviation of 20 days (Table 5.4). The uptime in the lifetime is, on average, only six days, corresponding to an average uptime of 40%.

Due to our scraping methodology, a number is considered online if an SMS was received in the last $\tilde{5}5$mins and the inbox number is displayed on the website. Further, our dataset is limited to a timeframe of two months. Consequently, numbers could have appeared before this period or still exist after this period. Further, they could have been online between two scraping intervals, so our system did not detect them. For more precise estimations, observations over a longer period must be made. The status that an inbox is listed on the website and that an inbox receives messages must be considered.

Website	μ Avg. Lifetime [days]	σ Stddev. Lifetime [days]	μ Avg. Uptime [days]	σ Stddev. Uptime [days]	uptime in lifetime %
7sim.net	32.46	24.22	2.49	2.97	32 %
online-sms.org	17.27	15.67	2.70	8.03	30 %
receive-sms-free.cc	38.13	22.48	35.52	21.07	93 %
receive-smss.com	21.65	8.70	21.07	8.24	98 %
receivesms.co	49.70	16.79	49.15	16.39	99 %
receivesms.live	18.93	10.39	8.33	7.08	39 %
receivesms365.com	46.42	14.95	11.90	6.16	26 %
receivesmsfast.com	20.94	9.50	9.40	5.34	44 %
sms24.me	11.92	9.26	4.05	4.33	50 %
tesms.net	3.34	4.80	2.79	2.23	85 %
Overall	22.65	20.25	6.05	12.40	40 %

Table 5.4: Average lifetime and uptime of inboxes per website.

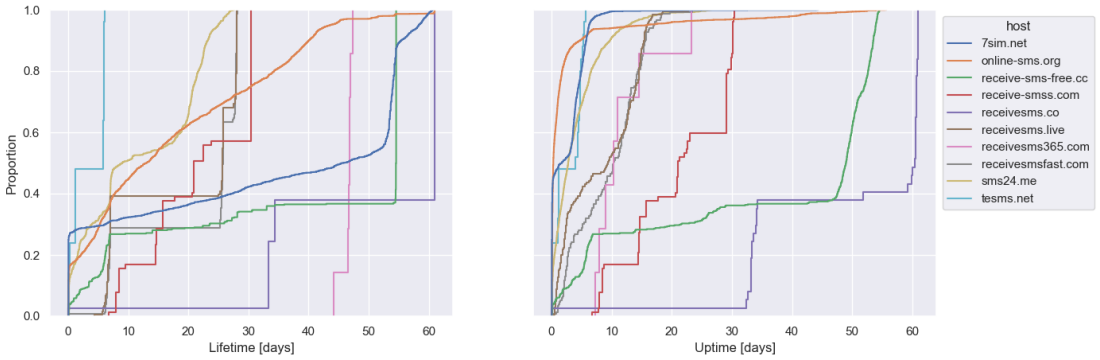


Figure 5.6: Cumulative distribution for lifetime and uptime of inbox numbers.

Shared numbers

Previous research recognised that public SMS inboxes share numbers to some extent (See 3.3.2). We analysed our dataset for this aspect and found that 69.29 % of all numbers occur on more than one website (Table 5.5). Not more than five websites shared a number. Concerning which websites share numbers, it is observable that websites that offer many numbers also share many of these numbers with others. Sharing numbers also happens between websites that do not belong to the same cluster (Fig. 5.7). Most numbers (1,712) are shared between `7sim.net` and `online-sms.org`.

According to Cheng et al. (2020), the source of shared phone numbers could be mutual scraping or a common underlying gateway. Further, we argue that they could lease these numbers for a certain amount of time from a gateway, after which another party leases the number. Theoretically, it is possible to distinguish the three cases. When scraping messages from another party, the timestamp and appearance of the messages in one inbox would always be later than the one scraped from. In the case of shared gateways, we expect the messages to have roughly the same timestamp with only minor variations. With leased numbers, we expect the number to not be active on two websites concurrently, and a message to not appear in more than one inbox (Table 5.6).

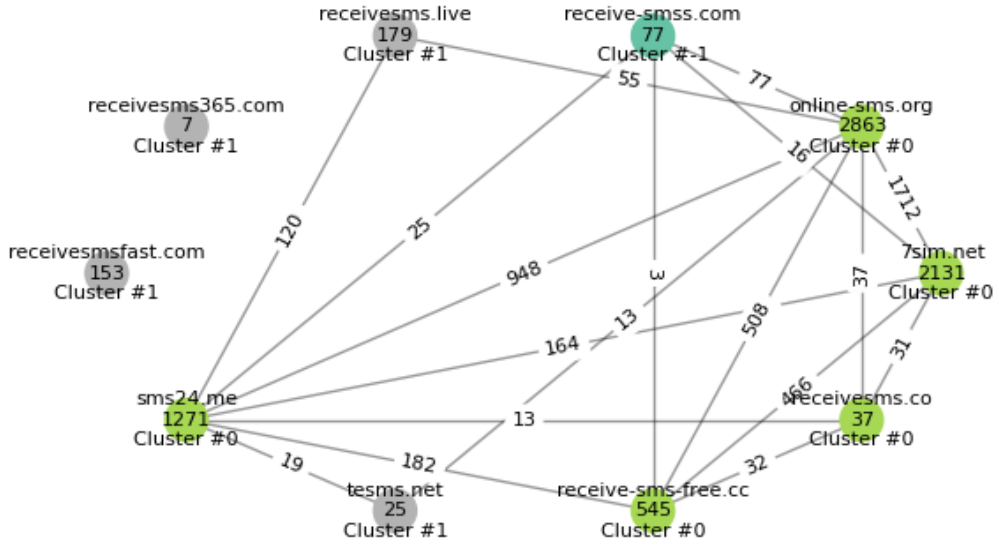


Figure 5.7: Websites and the amount of numbers they share between each other.

Nodes: Public SMS inbox websites, the count of unique phone numbers they provide, and the cluster they belong to.
Edges: Count of numbers they share with others.

# Websites	# numbers	% of numbers
1	1180	30.71 %
2	2067	53.80 %
3	418	10.88 %
4	170	4.42 %
5	8	0.20 %

Table 5.5: Amount of same numbers occurring on multiple websites.

	Time delay	Shared messages	Concurrently active
Scraping	✓	✓	✓
Shared gateways	✗	✓	✓
Leased numbers	✗	✗	✗

Table 5.6: Possible sources for shared numbers.

Leased numbers To distinguish between the case of leased numbers, scraping or shared gateways, we checked whether there are numbers that appear on multiple websites concurrently. For most websites, the numbers they share appear concurrently in other websites (Table 5.7). However, except for three websites (`7sim.net`, `online-sms.org`, `tesms.net`), only about half of the numbers are active concurrently. For these three websites, the reason can be the shorter uptime (32% resp. 30%) or the short lifetime (3.3 days for `tesms.net`), so phone numbers might appear simultaneously but not get captured by our scraping infrastructure. So we conclude that number leasing is unlikely and the reason for shared numbers is probably due to shared gateways or mutual scraping.

Scraping and shared gateways We can potentially differentiate scraping and the use of a shared gateway by comparing the average time delay of the same message between inboxes. We assume that a website scraping messages derives the timestamp it displays from the time it scrapes the message from another party and not the displayed timestamp from the source. In case they share gateways, we assume messages to arrive roughly at the same time at the different websites, so there would be no significant delay. We calculate the delay parameter by comparing a subset of unique messages for a phone number by subtracting their timestamps from the two websites where they occur. Comparing non-unique messages (duplicates) would elude the results. Further, we calculate the percentage of shared messages. We only compare messages with timestamps of precision of +- one minute.

From our data (Fig. 5.8), we see no apparent positive or negative delay for each website. We would expect that there is a significant positive time delay for a website that is scraping others. We see that the distribution of delays is shifted for some websites to positive or negative, but still, some delays are outliers to this trend. A delay in shared messages between websites can also happen in the case of shared gateways since we do not know the delay between the shared gateway and the displayed inbox. Further, counter to our assumptions, the timestamps given by the website might not be correct.

While analysing our dataset, we found messages with the body „Copied from receive-smss“ in three websites (Table 5.8). From this observation, we can conclude that these websites scrape, at least to some extent, messages from one or many other websites. These three websites also have a noticeable time delay in their messages (Fig. 5.8). Nevertheless, we cannot reject the case that also websites share gateways. Possibly both cases exist, with evidence of mutual scraping by some websites.

Website	# numbers	# shared	# concurrently active	% unique messages shared
7sim.net	2,131	1,736	911	12.74 %
online-sms.org	2,863	2,568	1,701	19.12 %
receive-sms-free.cc	545	529	505	27.36 %
receive-smss.com	77	77	75	25.00 %
receivesms.co	37	37	37	44.94 %
receivesms.live	179	120	109	9.77 %
receivesms365.com	7	0	0	N/A
receivesmsfast.com	153	0	0	N/A
sms24.me	1,271	1,022	956	14.68 %
tesms.net	25	19	9	43.35 %

Table 5.7: Amount of shared and concurrently active numbers and percentage of unique messages per website.

Website	# occurences
online-sms.org	62,006
sms24.me	1,507
7sim.net	22

Table 5.8: Occurence of message „Copied from receive-smss“ in websites.

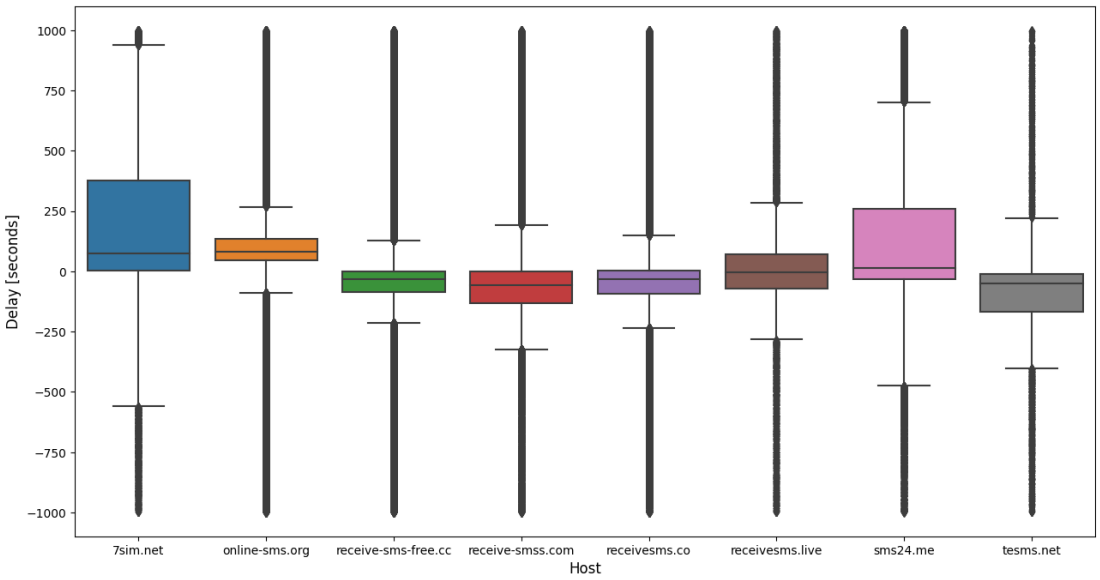


Figure 5.8: Boxplot delay of same messages sent to shared numbers.

Message delay in second. Negative values indicate that the message appeared first on the website. Positive values indicate that the message appeared on another website first, so it is delayed on the corresponding website.

Similarity

Previous studies showed that 40% of the numbers used as public SMS inboxes are similar with a Hamming distance of two or fewer (Reaves et al., 2018). We measured the hamming distance, which gives the number of positions that are different between two equal-length strings, for the inbox numbers in our dataset. Our results show that only 13,48% of the inbox phone numbers have a hamming distance of two or less (Table 5.9).

With the hamming distance, a character can be changed at any position in the string. Therefore, we further investigated how many numbers share the same number block. The hypothesis is that when numbers are bought in bulk, they share the same number blocks, and only the last digits change. E.g. +4915200001 and +4915200002 share the same number block of +491520000X. Our results show that 20,19 % of all numbers share number blocks where only the last two digits vary (Table 5.10).

Hamming distance	# numbers	Percentage
1	423	11.01%
2	518	13.48%
3	637	16.58%
4	857	22.30%
5	1,213	31.56 %
6	1,767	45.98 %
7	2,633	68.51 %
8	3,401	88.50 %
9	3,672	95.55 %
10	3,781	98.39 %
11	3,831	99.69 %
12	3,839	99.90 %
13	3,843	100.00 %

Table 5.9: Hamming distance.

# Masked chars	Example	# numbers	% of numbers
1	+1201277146x	473	12.1%
2	+120127714xx	776	20.19%
3	+12012771xxx	1,169	30.42%
4	+1201277xxxx	1,666	43.35%
5	+120127xxxxx	2,320	60.37%
6	+12012xxxxxx	3,045	79.23%
7	+1201xxxxxxx	3,692	96.07%
8	+120xxxxxxxx	3,815	99.27%
9	+12xxxxxxxxx	3,832	99.71%
10	+1xxxxxxxxxx	3,835	99.79%

Table 5.10: Number blocks.

5.1.3 Phone number characteristics

Networks

For each telephone number, there is a network that provides the telecommunication service for the subscriber. We can determine the line type and the name of the original network by performing a number type lookup. Further, we can query „Home Location Register“ (HLR) databases that „[...]are databases maintained by mobile operators containing information about the current status of a phone number, e.g. the International Mobile Subscriber Identity (IMSI), roaming status, and roaming operator“ (Costin et al., 2013). To perform HLR lookups at mobile networks, we need access to the SS7 network (Engel).

We make use of the service „`hlr-lookups.com`“[6] that provides querying HLR registries via an API. As HLR lookups are not available on every phone line, we perform a number type lookup beforehand that gives us the information if an HLR lookup is possible. We queried the number type for all 3843 distinct phone numbers from our dataset advertised as public SMS inboxes. As a result, we were able to retrieve HLR records for 3721 (97 %) of these numbers.

Line type Most numbers were identified as a „Mobile“ or as „Mobile or Landline“. We suspect „Mobile or Landline“ labels are numbers served from VoIP networks. A minority of the numbers could not be identified or might be wrongly labelled as Landline (Fig. 5.9). Noteworthy is that no numbers of the type „Toll free“, „Shared cost“, or „Premium rate“, which charge the caller, were detected in our set. Our findings are similar to previous studies, where most were mobile or VoIP lines, with a few landlines suspected to be wrongly labelled (Reaves et al., 2018).

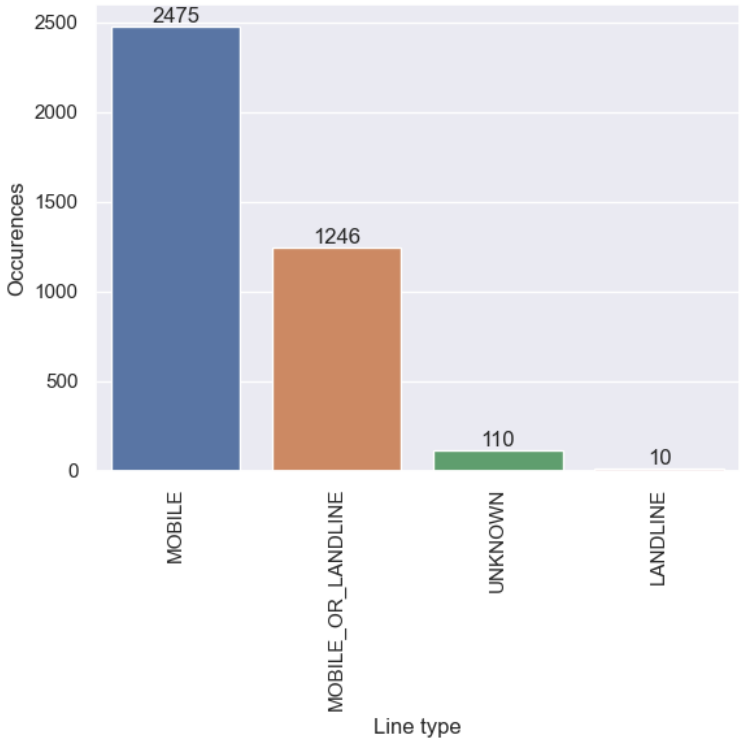


Figure 5.9: Line type of phone numbers.

Geography The Number Type Lookup dataset includes information on the country to which a phone number belongs according to its prefix. Roughly one-third of the

[6] See `https://www.hlr-lookups.com/`

numbers are based in the United States, followed by the Russian Federation and the United Kingdom (Fig. 5.10). Compared to previous studies, our distribution of countries differs but is easily explainable. What kind of countries are present highly depends on which numbers the public SMS inbox website offers.

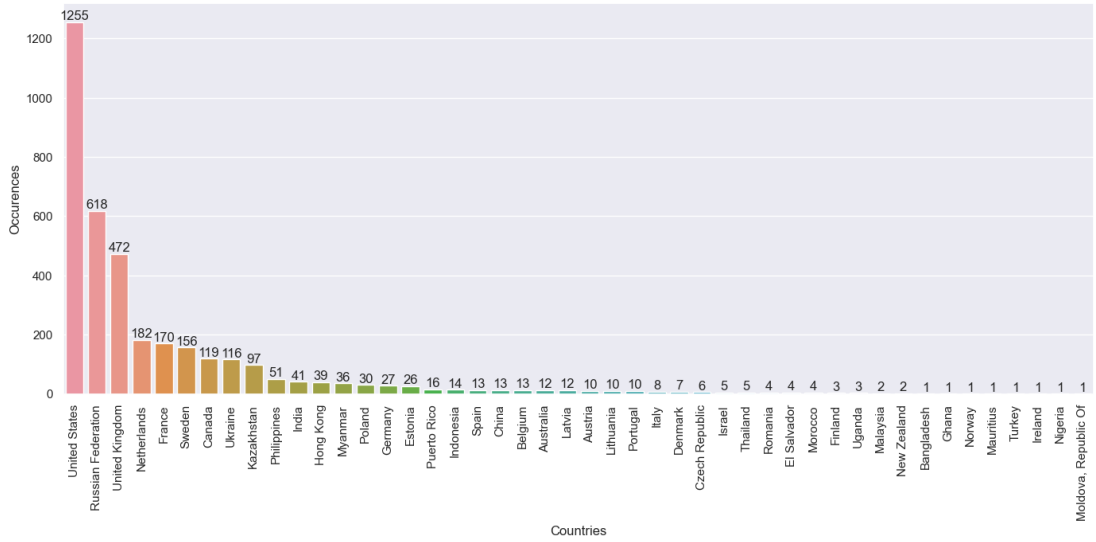


Figure 5.10: Top 30 countries in which the phone numbers are located.

Original, Ported and Roaming Networks The original network of a phone number is determined by the area code prefix allocated to a mobile phone network (wikipedia.org, b). However, the original network must not necessarily be the network currently serving the phone number, as numbers can be ported to other networks. For some networks, we see a large share of numbers to be ported (Fig. 5.11). An HLR lookup at the original network gives us the information if a number is ported and to which network. With this information, we can determine the network currently serving the number. By default, this is the original network, but if the number is ported, the ported network is considered the serving network. The network which serves the most numbers is „Bandwith.com“, a „[...]Communications Platform as a Service Provider [...] [that sells] APIs for voice and messaging, using their own IP voice network.“ (wikipedia.org, a). The second and third most prevalent networks are „Verizon Wireless“, and „VimpelCom PJSC“, both major mobile network operators. In total, 1011 numbers are marked as ported (Fig. 5.13).

A mobile subscriber can also be roaming in another network in another country. This state is also represented in the HLR lookup data. We identified 42 phone numbers currently roaming and determined in which network they are roaming, a majority within „Orange France“ network (Figure 5.14). Which networks are used for public SMS inboxes has not been published before. We see that the network „Bandwith.com“, a network specialised in providing SMS messaging APIs, is powering 7.11% of the numbers. However, they are not the only network. Various networks serve the numbers, and classical mobile networks, like „Lycamobile“, that does not offer any SMS messaging APIs. This implies

that different APIs and gateways, which bridge SMS/mobile networks to the Web, exist. We investigate this further in 5.1.3.

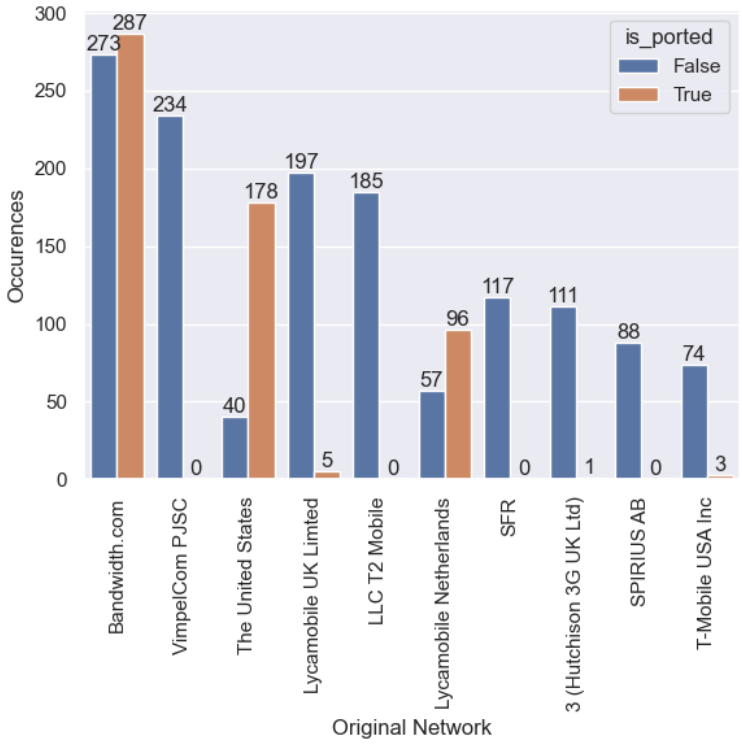


Figure 5.11: Top 10 original networks and how many of their numbers are ported

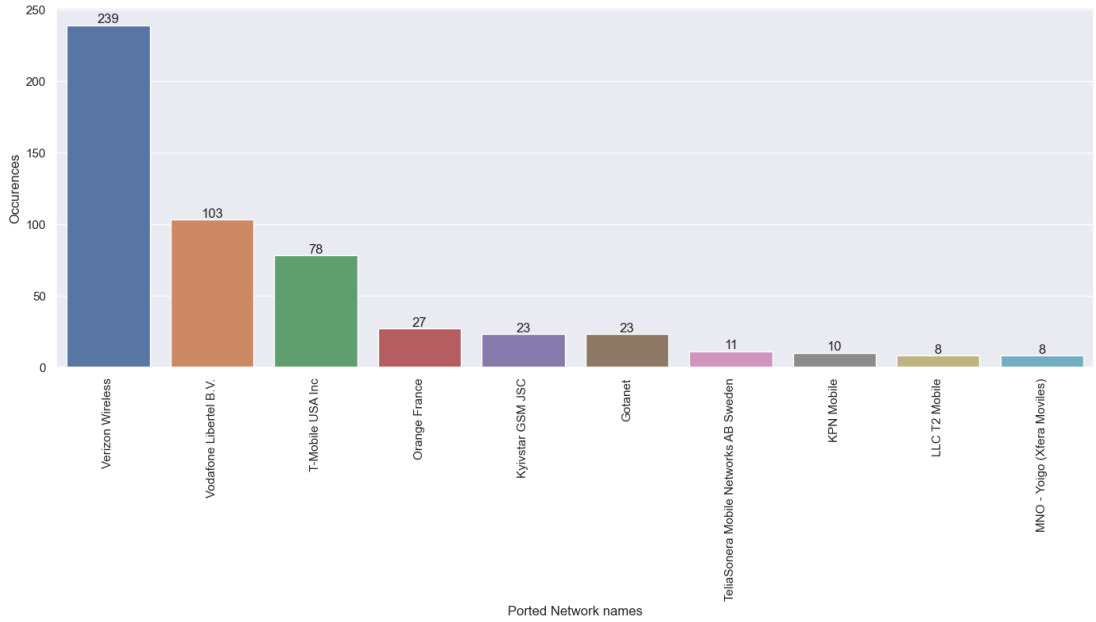


Figure 5.12: Networks to which the numbers were ported.

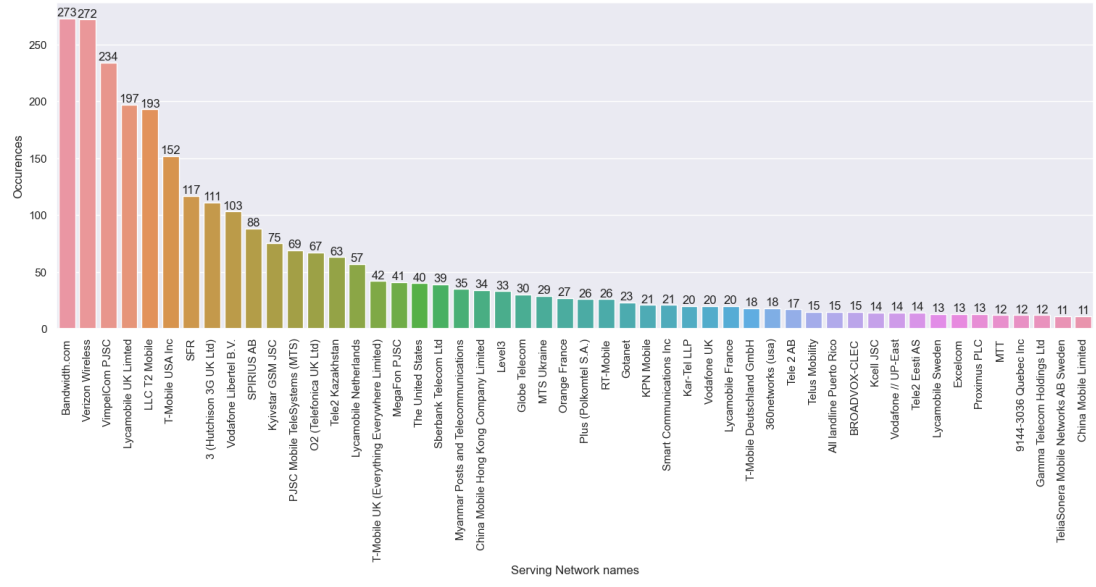


Figure 5.13: Serving networks.

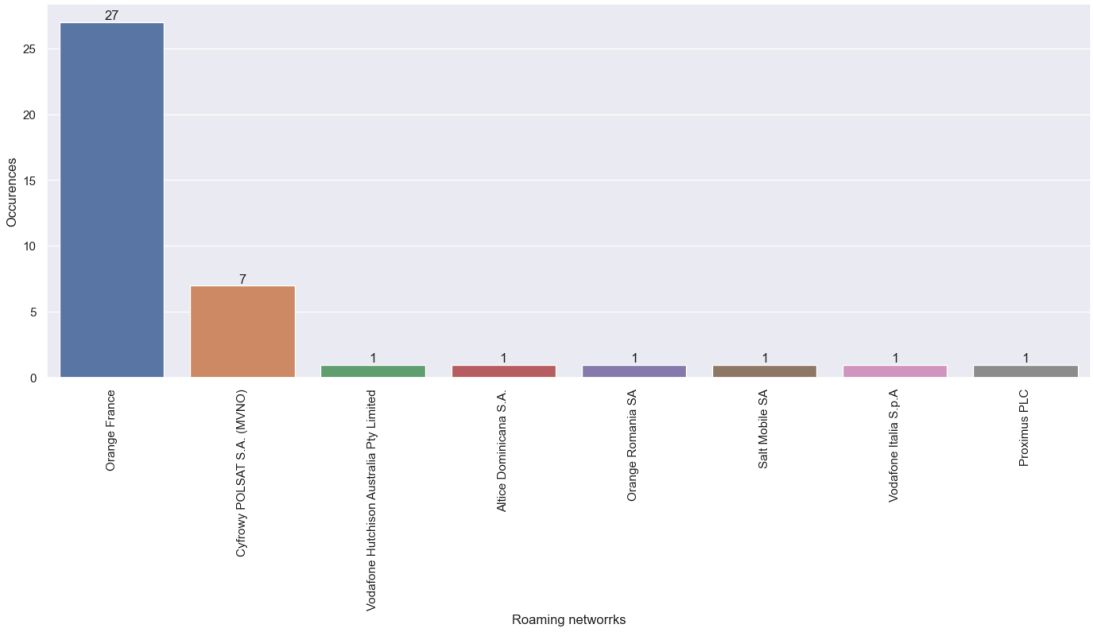


Figure 5.14: Networks in which subscribers are roaming.

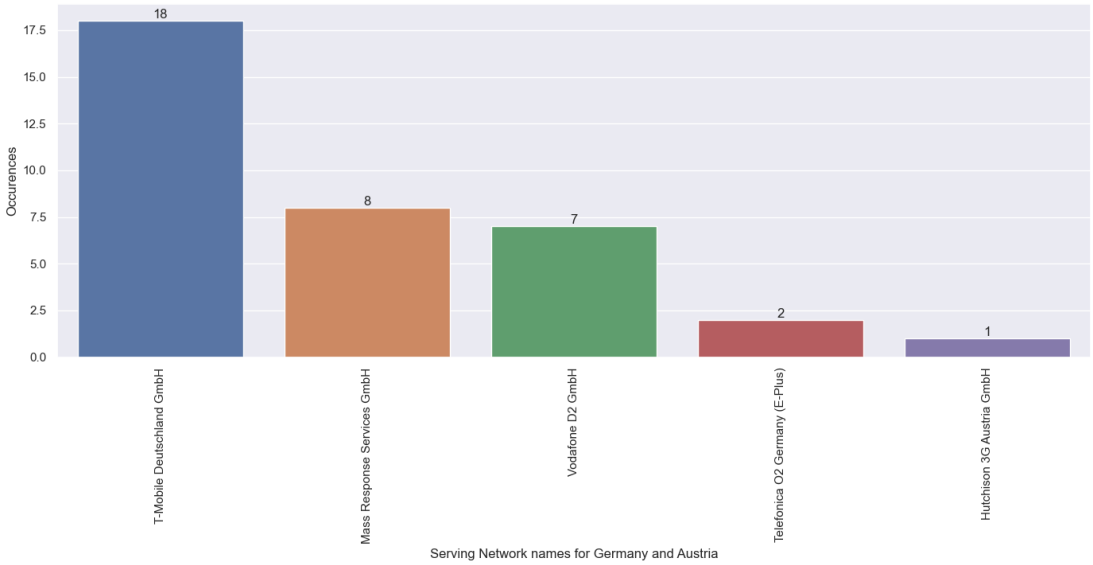


Figure 5.15: Serving network for Germany and Austria.

Connectivity status With the HLR lookup, we can determine the connectivity status of the subscriber, which belongs to the phone number.

> CONNECTED Indicates that the number is valid and the target handset is currently connected to the mobile network (and reachable).
> ABSENT Indicates that the number is valid, but the target handset is currently switched off or out of network reach.
> INVALID_MSISDN indicates that the number is not currently assigned to any subscriber on the mobile network or is otherwise invalid.
> UNDETERMINED indicates that the connectivity status could not be determined and the connectivity status is unknown. This can be caused by invalid numbers, due to lack of connectivity to the target network operator, or other exceptions and errors.

(hlr lookups.com)

Since the phone numbers were collected in April and May 2022, and the HLR lookup was performed in September 2022, we expect many phone numbers to be shut down by subscribers or shut down by the networks. Surprisingly, about half of the phone numbers are still in use („ABSENT“, or „CONNECTED“) (Fig. 5.16). We compare the different mobile network operators on how many numbers they are serving are now online („CONNECTED“, or „ABSENT“) or offline („INVALID“, or „UNDETERMINED“). By this, we try to grasp which mobile network operators are deactivating phone numbers under the assumption that public SMS inbox websites try to use a phone number as long as possible (Reaves et al., 2018). In figure 5.17 we can see a huge difference between the serving networks. „VimpelCom PJSC“, „Lycamobile UK Limited“, and „T-mobile USA“ barely took numbers offline, whereas „Bandwith.com“, „SFR“, and „SPIRIUS AB“ took offline all numbers they served. Within other major networks, the rate between on- and offline numbers is balanced, e.g. „Verizon Wireless“, and „LLC T2 Mobile“. The reason for numbers becoming online or offline can vary, e.g., abnormally high usage (amount of message), detection of number in SIM box fraud or expiration of the contract.

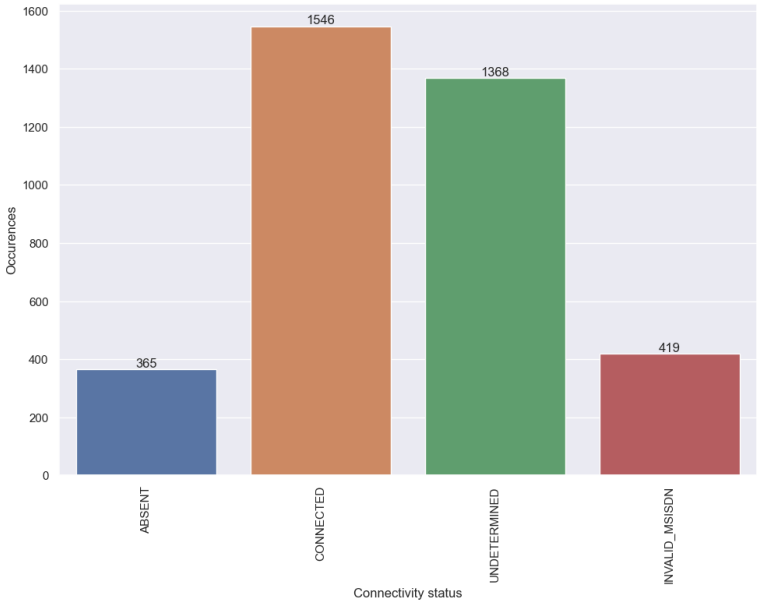


Figure 5.16: Connectivity status of phone numbers.

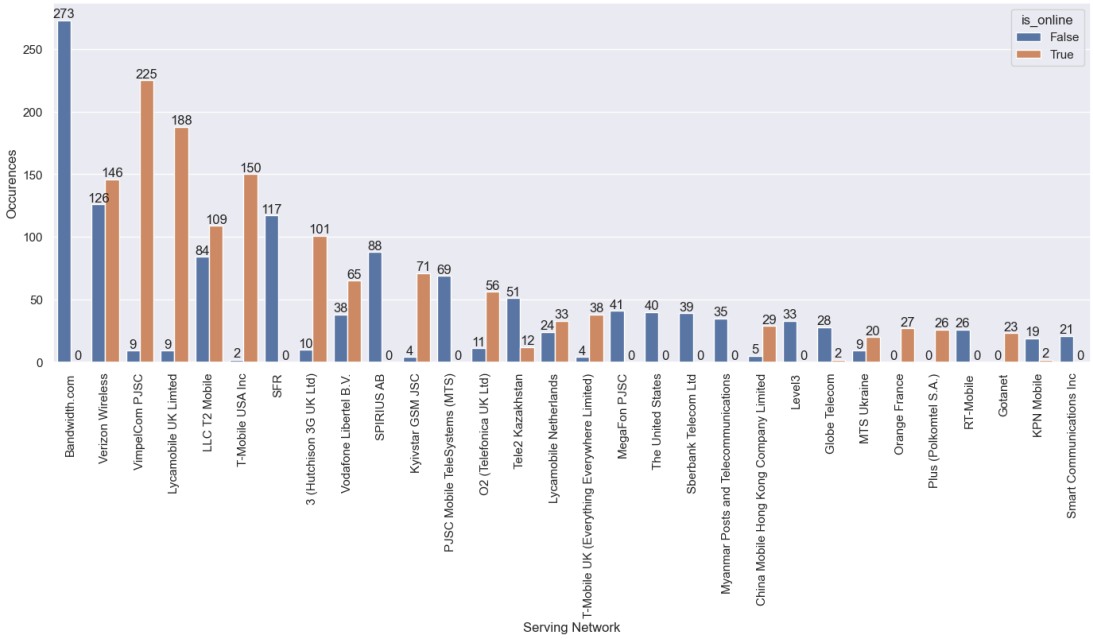


Figure 5.17: Online / Offline numbers per serving network (Top 30).

Gateways

Public SMS inboxes must rely on a gateway that bridges the SMS network and the Web, enabling the websites to display SMS messages (Fig. 2.2). Even if some websites are sourcing the numbers and messages by scraping others, there must be other websites (from which they scrape) that are somehow connected to the SMS network. There is no description in previous literature on what kind of gateways are used. However, from our analysis of phone number networks, observations in online forums and related literature work, we can identify three different kinds of possible gateways: Messaging APIs, SIM-Boxes and Malware.

Messaging APIs The obvious choice for public SMS inbox websites would be to interface the SMS network via messaging APIs. However, our analysis in 5.1.3 revealed that a share of these numbers is served in the network of „Bandwith.com“, which solely offers Messaging and Voice APIs.

Another hint for messaging APIs is served by the discussion forum of the Codecanyon template[7], where users discuss the use of messaging APIs. We extracted the discussed APIs and looked up their pricing (Table 5.11). These message APIs charge only for the number and not per inbound message. This is an important factor towards the economics of public SMS inboxes. Nevertheless, messaging APIs are not the sole source of numbers since also pure mobile networks (e.g. Lycamobile) serve numbers that can only be connected to via a SIM card and mobile gateway (e.g. mobile phone).

API Provider	Cost per US number	Cost per Austrian number	Cost per inbound SMS
`messagebird.com`	from 1€	from 8€	free
`sms.to`	N/A	N/A	free
`txtsync.com`	5.85€ (dedicated)	free (shared number)	free
`proovl.com`	from 0.99€ (79.00€ for SIM)	189€ (german SIM)	N/A

Table 5.11: Possible messaging APIs extracted from an online forum.

SIM-Boxes SIM Boxes are devices that can handle multiple SIM cards to connect to mobile networks (Murynets et al., 2014). They are used in different telecommunication fraud schemes, most commonly to terminate traffic (SMS and voice) in a destination network that bypassed legal interconnects (and their fees) over IP networks (See Fig. 5.18) or to deliver Robocalls (Sherman et al., 2020). Sim boxes can be centralised, where one machine is operating a large quantity of SIM cards, or decentralised, where consumers

[7] See `https://codecanyon.net/item/tsms-temporary-sms-receiving-system/23244962/comments`

download apps that allow SIM farm operators to remotely use the device for sending and receiving SMS (Forum, 2020).

Since bypass fraud incurs severe losses to telecommunication networks, an industry that tries to identify these SIM boxes has evolved. Therefore, we contacted the two companies Araxxe[8] and A1 Telekom Austria[9] that are providing such services with the intent to identify which phone numbers (inboxes) from our dataset have an affiliation with SIM Boxes.

We were able to cooperate with the company Araxxe. They compared our list of phone numbers with their list of identified SIM boxes, which resulted in 20 positive hits (Perron and Araxxe, 2022). In addition, we looked up which websites make use of SIM boxes and identified five of the ten scraped websites (Table 5.12). This result indicates that some public SMS inboxes collaborate with SIM Boxers and therefore use illegal gateways to some extent. More than 20 numbers are possibly sourced from SIM box operators but might not be detected yet or are not in the scope of the company's surveyed networks.

Without insight into the operations of SIM Boxers, we can only speculate on the incentive to act as a gateway for public SMS inboxes. One reason could be that they power illegal Messaging API services on which public SMS inboxes rely. Another reason could be that they use inbound SMS traffic to cover up their outbound voice and messaging activity so that they appear more like a regular subscriber, which would be more challenging for telecommunication companies to detect by anomalies.

Website	# detected SIM Boxes	# total numbers
7sim.net	19	2,131
online-sms.org	17	2,863
receive-smss.com	1	77
receivesms.live	0	179
receivesms365.com	0	7
receivesmsfast.com	0	153
sms24.me	0	1,271
tesms.net	0	25
receive-sms-free.cc	8	545
receivesms.co	2	37

Table 5.12: Websites and how many numbers they source from SIM Boxers.

[8]See `https://www.araxxe.com/`
[9]See `https://www.a1.group/en/wholesale/sim-box-detection-service`

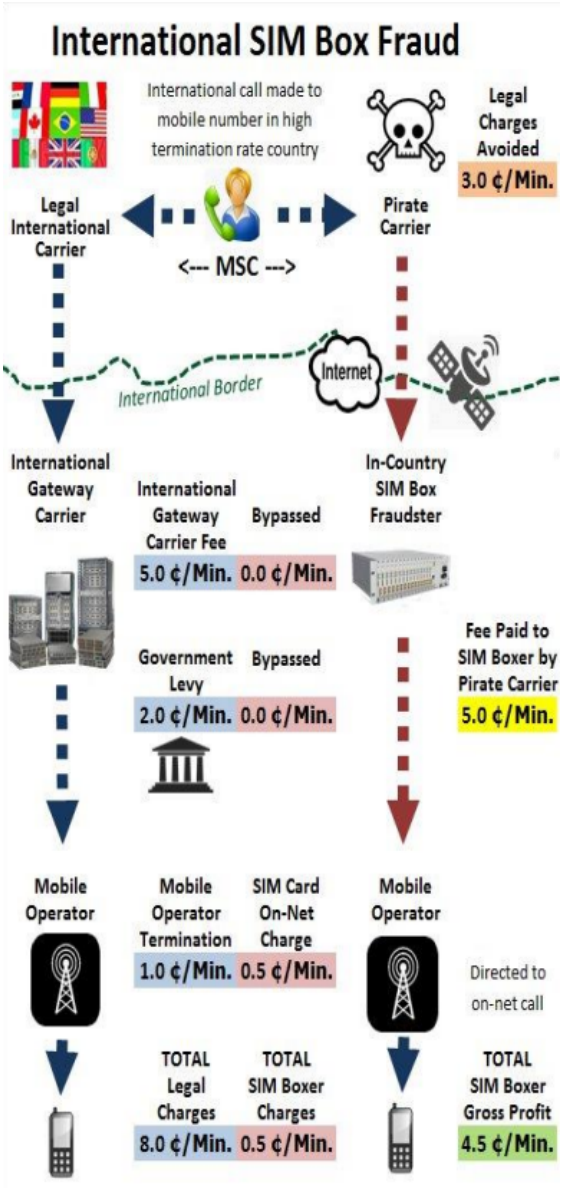


Figure 5.18: SIM box fraud (Murynets et al., 2014).

Figure 5.19: Example of SIM box devices (addpac.su, 2010).

Malware The third kind of gateway could be compromised phones. Dong et al. (2022) describes detailed how malware on Android intercepts specific SMS messages by a Regex expression it gets from Command and Control server. So far, there is only evidence that these schemes are only used for phone-verified account evasion by intercepting SMS messages for one specific sender for a short time on the compromised phone. Then, the phone number of the compromised phones is sold as a temporary phone number to lease for receiving SMS for this specific sender.

The malware described could be easily adapted to intercept all SMS messages to power public SMS inboxes. Though there is no evidence of such malware yet, there is no reason that compromised phones do not become a commodity in black markets (Thomas et al., 2014).

5.2 Usage of public SMS inboxes

5.2.1 Message characteristics

Common sender and message content

We look at the most common sender and message content in our dataset. We observed that the results are different for each website, so we compare on a per-website basis. The complete list of the most common sender and message content can be found in Appendix A.5. The most common messages in „`7sim.net`“, „`receive-sms-free.cc`“, and „`tesms.net`“ are phishing messages, that we further analyze in 5.3.4. „Copied from receive-smss“ is most common in „`online-sms.org`“, thereby indicating that this website scrapes „`receive-smss.com`“, but obviously fails due to this message. With „`receive-smss.com`“, „`receivesms.co`“, „`receivesms.live`“, „`receivesms365.com`“, „`receivesmsfast.com`“ and „`sms24.me`“ most common messages are welcome messages to services (e.g. Uber) or one-time passwords for phone verification.

Regarding the senders, we observed that in six websites, senders are represented with names, and popular internet services are most common (e.g. Amazon, Skype). Against this, four websites represent the senders with partially masked numbers (e.g. +122678XXXXX) (Table A.6). It is unclear if the websites derive the sender's name from a lookup list via the sender's phone number or if they display the alphanumeric sender ID (thesmswork.co.uk).

Multipart SMS

In 1,142,225 rows, the string in the „body“ field was longer than 140 bytes, which is the maximum length of a single SMS (Amri). If an SMS message exceeds 140 bytes, it is sent as multipart SMS. Therefore, the presence of long SMS indicates that the receiving gateway can assemble multipart SMS messages.

Usage patterns and correlation

Usage, respective to the amount of messages, differs over time. A fluctuation in the amount of messages in public SMS inboxes is observable over two months within our dataset (Fig. 5.20). Similar to findings of previous studies, it is observable in our data that usage slightly increases during daytime (Fig. 5.21).

To estimate the influence of lifetime and uptime on the amount of messages, we calculated the correlation coefficient overall and for every website. For most websites, between the amount of messages and lifetime, we observe a moderate positive correlation and a fairly strong positive correlation between uptime and amount of messages (Table 5.13). However, the amount of messages websites receive, lifetime and uptime differ significantly (Fig. 5.22), so the overall correlation for the lifetime is fairly weak and moderate for uptime.

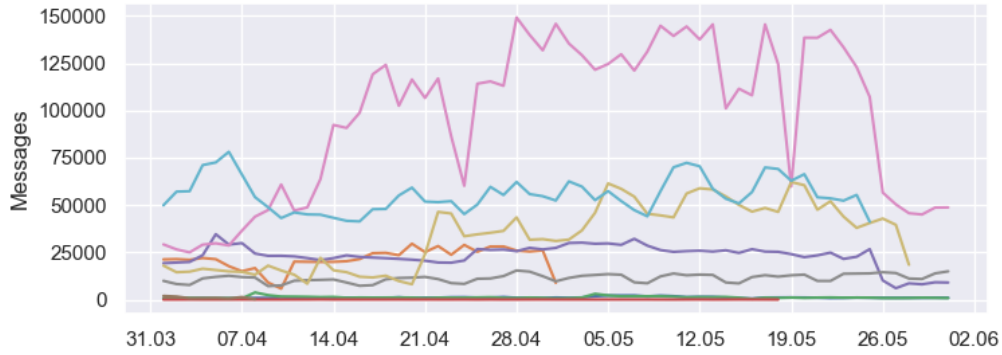


Figure 5.20: Daily Usage pattern over the observation period.

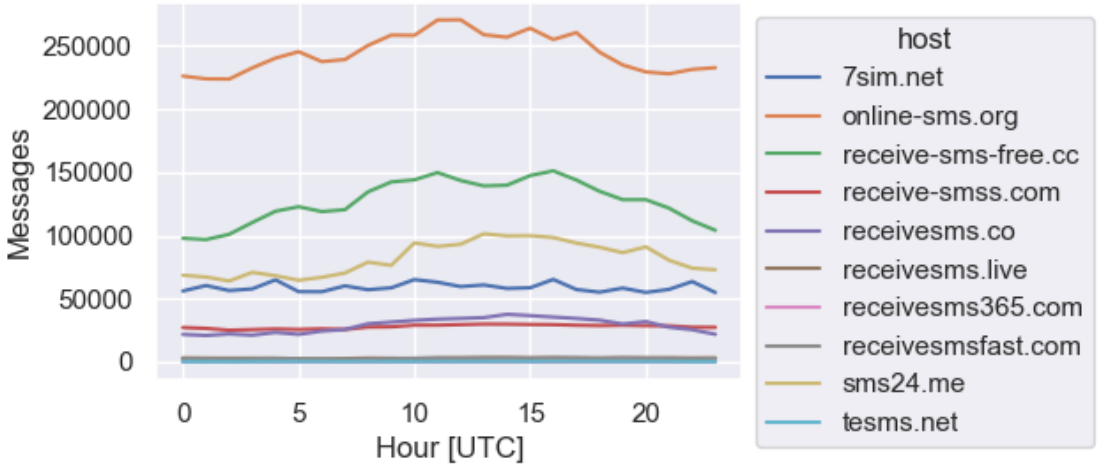


Figure 5.21: Hourly usage pattern over the observation period.

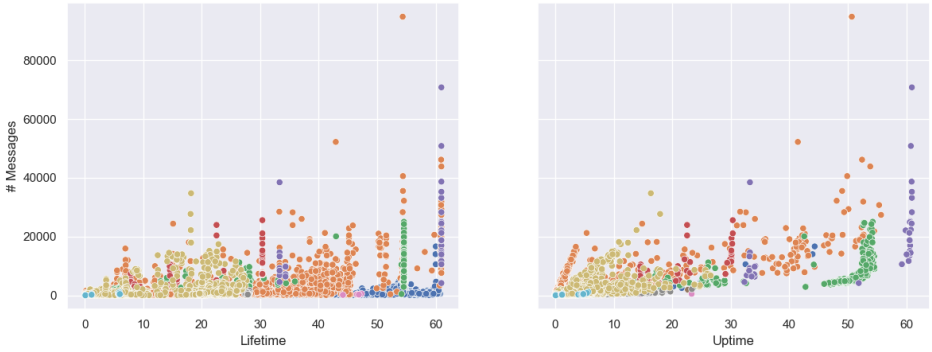


Figure 5.22: Scatterplot between the amount of messages, lifetime and uptime.

Website	Lifetime	Uptime
7sim.net	0.50	0.70
online-sms.org	0.55	0.80
receive-sms-free.cc	0.59	0.64
receive-smss.com	0.57	0.60
receivesms.co	0.48	0.51
receivesms.live	0.79	0.95
receivesms365.com	0.50	1.00
receivesmsfast.com	0.66	0.92
sms24.me	0.29	0.56
tesms.net	0.70	0.79
Overall	0.36	0.66

Table 5.13: Correlation coefficient of the amount of messages with uptime and lifetime per website.

5.2.2 Use cases

We have several sources like previous literature, terms of service, privacy policies, observations and message characteristics that hint at what public SMS inboxes might be used for. To methodically structure the usage, we created a list of use cases. We differ in benevolent usage, where users do not intend to harm a service and malicious usage of public SMS inboxes, intended to harm and exploit services. Further, we categorise the different use cases.

Intent	Category	Use case
Benevolent	Living abroad	Registration at service requires a local phone number.
	Testing	Developers use numbers for testing A2P SMS functionality.
	Multiple accounts	User wants to create multiple accounts for different purposes.
	Privacy	User wants to prevent: - receiving spam - leakage of a phone number to the public - association with his identity via phone number - restriction circumvention - handing out a phone number to other individuals for communication via SMS
Malicious	PVA evasion	User circumvent phone verification to create accounts for malicious activity (to exploit sign-up incentives, free-tiers, trial periods, ...)
	Number chaining	User verifies accounts at VoIP providers with public SMS inbox numbers to obtain fresh numbers (e.g. Google Voice)
	Identity obfuscation	Cybercriminals use phone numbers to hide their identity, e.g. in e-commerce fraud

Table 5.14: List of use cases.

Living aborad

The use case of living abroad is given when a user needs a phone number to register at a service for which a specific country-prefixed number is needed. The user might be unable to register without such a number and therefore relies on numbers obtained from public SMS inboxes. We cannot quantify how many messages in public SMS inboxes relate to this use case, but Reaves et al. (2018) observed by aggregating click statistics from shortened URLs that „the locations of the gateways' users significantly differs from the services sending message“. Though Reaves et al. (2018) conclude that the primary purpose is PVA fraud, it may also be an indicator for the „Living abroad“ use case.

Testing

Reaves et al. (2018) labeled 0.9%´of all messages as test messages. In our dataset, we also found messages that appear to be sent in a testing use case (Table 5.15). We see proof that developers use public SMS inboxes to test Application-to-Person SMS functionality.

Sender	Message	Website
Ello	Hello bruv. testing code	receive-sms-free.cc
18448575XXX	Campaign Testing 127969elementum	online-sms.org
18448575XXX	Campaign Testing 3109263nostrum	online-sms.org
+40770924XXX	Test	online-sms.org

Table 5.15: Example of test messages.

Multiple accounts

Some services use phone number verification to prevent the creation of multiple accounts. Public SMS inboxes might serve the use case in which a user wants to create multiple accounts but only has one phone number. This is phone-verified account evasion (PVA), but with benevolent intent. Nevertheless, there are some restrictions to creating a long-term account with a public SMS inbox number. At some services, continuous access to the number for receiving SMS must be available, e.g. when the phone number is used as a second factor in authentication. Further, if phone numbers are used as the first factor in authentication, other users might be able to authenticate and intentionally or unintentionally hijack the account. Therefore, we conclude that the creation of long-term accounts is rather unlikely.

Privacy

Prevent spam We came up with different reasons why a user might use a public SMS inbox for privacy reasons. Users might want to redact their phone numbers not to get marketing/spam messages from that service. However, Reaves et al. (2018) identified that only 1.0% of all SMS messages in public SMS inboxes were actual spam. Therefore, the fear of spam abuse might be exaggerated or other motives prevalent.

Information leakage Handing someone's phone number to a service poses the threat of this information getting leaked in a data breach. So users might be hesitant to enter their phone number to mitigate that risk. We checked „Have I Been Pwned“[10] for how many of the phone numbers in our dataset the information was breached. From the 3843 phone numbers, 78 were leaked in the Facebook data breach, but nowhere else (haveibeenpwned.com). It is unclear if the phone numbers were already active as public SMS inboxes when the data breach occurred in 2019 or if they were coincidentally reassigned to SIM cards that public SMS inboxes are using.

Identifiability Users might prefer to use public SMS inboxes to prevent the potential revelation of their identity. The subscriber of a phone number can usually be identified,

[10] See `https://haveibeenpwned.com/`

as it is a requirement by law (privacyinternational.com) or via billing information.

Restriciton circumvention Similar to the „living abroad“ use case, an individual might need a phone number from abroad to circumvent certain restrictions in a country. For example, China is banning all virtual currency trading. To enforce this policy, cryptocurrency exchange platforms like Binance do not allow registration with Chinese phone numbers (Dong et al., 2022). Consequently, public SMS inboxes become a tool to circumvent local restrictions.

Personal communication We investigated if users possibly do handle person-to-person communication over public SMS inboxes. However, neither Reaves et al. (2018) encountered a significant amount of personal messages during their clustering, nor did we encounter personal communication in our dataset. Therefore, we conclude that this use case is negligible.

Phone verified account evasion

Many account systems are based on a username or e-mail as the primary user ID. Creating a new e-mail address or username comes at no cost, and services exist that provide temporary e-mail addresses for free (Hu et al., 2019). Many services use phone verification to combat abuse (e.g. social bots), for security (two-factor authentication) and account recovery purposes. It is assumed that phone numbers are tied to individuals and to obtain additional phone numbers involves cost and substantial effort (Thomas et al., 2014). So platforms demand a phone number when signing up and verify them by sending a one-time password (OTP). Usually, there is a hard limit on how many accounts can be associated with a phone number. E.g. for Google accounts, a maximum of 10 accounts can be linked to a particular number (Xie et al., 2019). Consequently, miscreants require a constant stream of new phone numbers to create accounts (Thomas et al., 2014).

Using the resources of arbitrary phone numbers to create accounts at services instead of a phone number belonging to the user is referred to as phone-verified account evasion (PVA). Previous studies of Berenjestanaki et al. (2019), Cheng et al. (2020), and Reaves et al. (2018) conclude that public SMS inboxes are mainly used for PVA evasion.

Dei (2019) describes the use of a public SMS inbox in the flow of creating social bots on Twitter. Xie et al. (2019) mention the free access to phone numbers in public SMS inboxes while exploiting the account sign-up and verification process at Google. Dong et al. (2022) identified the use of paid temporary phone numbers in PVA evasion for buy-now-pay-later microfinancing, illicit purchasing, money laundering, fake accounts for misinformation, exploiting sign-up and game bonuses and other scam and fraud schemes. We investigate the threats and impact of malicious activity enabled by public SMS inboxes in 5.3.

Number chaining

Public SMS inboxes can also be used as an intermediate step to obtain a phone number at a VoIP provider. Thomas et al. (2014) describe this practice as number chaining. Miscreants may be able to register multiple virtual phone numbers with a so-called seed number. Repeating this process, they can acquire a tree of virtual phone numbers in a many-to-one relationship to seed numbers used in phone verification (Thomas et al., 2014). This practice might not only be used by individuals but also possibly by public SMS inboxes to acquire new numbers. One-third of the phone numbers in our dataset are VoIP numbers (Fig. 5.1.3). Reaves et al. (2018) observed the registration for VoIP services to Cuba (Reaves et al., 2018).

In our dataset, we observed 10,045 messages containing verification codes to set up an account at Twilio, a popular online VoIP provider, 143 account verification messages from the VoIP app „Ringberry“, and 6,191 messages containing codes for Google Voice. The prevalence of messages from VoIP providers clearly shows that number chaining is a relevant issue and that such providers do not take effective countermeasures against it.

Identity obfuscation

Because using public SMS inboxes provides privacy, Cybercriminals might use such numbers to obfuscate their identity and not be tracked down. However, this use case can not be differentiated from PVA evasion since it can be assumed that miscreants want to achieve both PVA evasion and anonymity. Nevertheless, PVA evasion tends more towards creating multiple accounts and exploiting the service, whereas identity obfuscation can be only a commodity in other (cyber) crimes, e.g. for credit card fraud.

5.3 Threats

The usage of public SMS inboxes can pose a variety of threats. We distinguish between threats for the users themselves, independent of malicious or benevolent intent, threats for services and threats for society (third parties) when public SMS inboxes enable cybercrime. Nevertheless, it might not always be distinguishable which parties are affected. For example, the leakage of private information is neither favoured by users nor services.

Category	Threat	Affected parties	Risk
Privacy leakage	PII in messages	user, service	low
	PII in linked services	sser, service	low
	Metadata	user, third-party	low
SMS Authentication	as first factor	user	high
	as second factor	user	low
PVA evasion	abuse of sign-up incentives	service	high
	resource exhaustion	service	high
	as commodity in cybercrime	service, third-party	high
	implications for identity providers	service, third-party	high
	phone number chaining	VoIP service	high
Scam messages	SMS Phishing	user	medium
	lottery scam	user	medium

Table 5.16: List of threats.

5.3.1 Privacy leakage

Personally identifiable information in messages

SMS messaging is assumed to be a private channel, though vulnerabilities have been known for a long time (Androulidakis, 2016). Public SMS inboxes make the contents of SMS messages public, thereby leaking private identifiable and sensitive information. Previous studies found credit card numbers along with CCV2 codes, login credentials, IBAN, names, passwords, TANs, password reset links, addresses, and status reports in the messages (Reaves et al., 2018).

We searched our dataset of messages with Regex expression to identify if they contain e-mail addresses, phone numbers, credit card numbers, social security numbers or International Bank Account Numbers (IBAN). Other than the previous studies, we did not find any evidence of this kind of personally identifiable data in our dataset, except for one case of messages from a job market that leaks phone numbers that are not related to the sender party. We cannot identify more PII because there is either awareness at the sending parties, public SMS inboxes redacting PII, or our regular expressions do

not catch information that may not be formatted appropriately. We see evidence that some public SMS inboxes redact e-mail addresses in messages by replacing them with the string „[e-mail protected]“.

Message
7Recruiting part time staff: earn ZAR 700-1000 a day,do it at home ,use your mobile phone,is simple and easy,Kindly pm https://wapp.my/+27672730***/ if interested

Table 5.17: Example of a message leaking private telephone number.

We partly redacted the phone number for privacy reasons.

Privacy leakage via linked services

Cheng et al. (2020) studied privacy leakage in Chinese online travel agencies that allowed sign-up and authentication with phone numbers. In the study, they were able to access private information like name, birthday, gender, ID and journey.

In our dataset, we found a similar privacy leak for the online travel agency „booking.com“. We encountered messages in which they notified customers that the property could not process the credit card. These messages send a direct link to the booking confirmation. With this link, information on the booking can be accessed without authentication, and the information within can be changed.

Message
The property couldn't process your credit card. To keep your reservation, please update your details within 2 hours: booking.com/55f4d654****

Table 5.18: Example of a message from booking.com leaking the secret link to view and modify a booking.

We redacted the last four symbols of the link.

Metadata leakage

Sender phone number Besides the message content, some public SMS inboxes also display the phone number of the sending gateway. We analysed our dataset for phone numbers in the sender field and identified 90,922 distinct sending phone numbers. Since we expect all messages to be application-to-person messages, these phone numbers belong to sending gateways. There is little risk that they belong to individuals, but we can not be sure that these phone numbers may belong to infected private devices that are part of botnets.

IP address Reaves et al. (2018) explains the leakage of the user's IP address by clicking on shortened URLs. URL shortening services record the IP addresses of visitors. This information was publicly available at the time of study from Reaves et al. (2018). We were unable to reproduce this leakage since, in the study, it is not explained which URL shortening services leak the IP address and the services we identified in our dataset make click statistics not publicly available.

5.3.2 SMS authentication

As single-factor

The use of public SMS inbox numbers for services where the phone number is the only authentication factor poses a severe security threat. Since the numbers and messages are public, there is the risk that a third party can easily log in to the account. Popular messengers like Whatsapp, Telegram and Signal rely on the phone number as the only factor for authentication by default. Especially accounts used for financial services like „bankoff.co“, which offers banking services via Telegram messenger, might be a lucrative target for account takeover. When a user uses a public SMS inbox phone number as the first and only factor for authentication, the risk is very high that this account gets compromised. We can construct the potential abuse scheme that public SMS inbox websites are potentially involved in the takeover of such accounts after users have set them up to exploit them. The threat can be mitigated by enabling multi-factor authentication.

As second-factor

The threat of using public SMS inbox numbers in multi-factor authentication schemes is relatively low. First, the compromise of one factor (the phone number) does not pose a direct threat as long as the second factor is not compromised. Second, as mentioned before, we do not expect accounts to be used in the long term since phone numbers must be reauthenticated from time to time. Public SMS inboxes are only available for a short period, on average 23 days, so users cannot access the codes sent to inboxes in an authentication process after some time. Nevertheless, if attackers can guess the phone number used as the second factor and are able to access the corresponding inbox, the security of two-factor authentication is severely weakened.

5.3.3 Phone-verified account evasion

Abuse of sign-up incentives

A major threat to services that offer some incentives for users to sign-up, like a free tier or coupon, is that miscreants exploit this benefit. Services enact measures, most commonly phone verification, to prevent the sign-up of arbitrary accounts. However, when miscreants are now able to access an arbitrary amount of phone numbers from public SMS inboxes, they can create multiple accounts to take advantage of the incentives.

Dong et al. (2022) describes the case of Starbucks in China, which offered coupons for a coffee on sign-up. Criminals created multiple accounts, reaping the coupons and selling them later for a discount.

Food delivery services In our dataset, we found evidence for the creation of accounts at different services that offer an incentive for sign-up. For example, we found 1,777 distinct messages for sign-up at the Austrian food delivery service „Mjam“. „Mjam“ gives a discount of 10€ for every new customer. The potential loss for the company is 17,770 € over two months. We can assume the damage is five to ten times bigger since our dataset only contains messages from a small set of public SMS inboxes. We found a similar amount of messages for other food delivery services that offer a similar amount of discount for new customers: Foodora (2,962 messages), DoorDash (1,1220 messages) and Deliveroo (9,922 messages).

Ride sharing Similar to food delivery services, we found messages from ride-sharing companies offering a discount on sign-up: Uber (44,443 messages), Bolt (35,690 messages) and Lyft (2,864 messages).

Cloud computing Cloud computing providers offer generous free tiers to try their services and onboard new customers. We found messages for Amazon Web Services (9,820 messages), Google Cloud (1,698 messages), and Microsoft Azure (1,366 messages). In terms of exploiting cloud computing tiers, it may also be a motivation for cybercriminals to redact their identity while running their services on cloud providers.

Ressource exhaustion

Thomas et al. (2014) theorises on the possibility of launching a resource exhaustion attack against services. Miscreants could use thousands of phone numbers to target phone verification processes at services. Since every SMS message is costly to send for services, attackers can incur damage. Further, the magnitude of requests a service causes for services might trigger latency in sending SMS messages or prevention mechanisms due to abnormal traffic. We could not find evidence for such an attack in our dataset.

As a commodity in (cyber)crime

Cybercrime nowadays heavenly depends on multiple resources provided by underground markets. The circumvention of phone verification is a commodity in the ecosystem of cybercrime (Thomas et al., 2015). With the availability of phone numbers, not only from public SMS inboxes, miscreants can create multiple accounts and hide their identities. We assume public SMS inboxes to be also part of this black market, though they provide a free service. Since there is benevolent and malicious usage of public SMS inboxes (see 5.2.2), it is impossible to tell to what extent they play a role in cybercrime.

Nevertheless, we see examples of messages in our dataset that may be linked to cybercrime. For example, we see a vast amount of messages for financial services, like

„CashApp“ and „PayPal“, an app for transferring money between peers, „Bankoff“ providing digital banking services, and „Coinbase“, and „Binance“, both cryptocurrency exchanges. These financial services enable criminals to launder money when bypassing security measures like phone verification. For example, there are news reports of drug traffickers using „CashApp“ to transfer money (Davis, 2021).

Phone numbers from public SMS inboxes might also be used to purchase goods with stolen credit cards. We see a considerable amount of confirmation messages from online shops. Also, the message stating that the credit card failed at the booking platform booking.com (Table 5.18) might be linked to credit card fraud. Additionally, we found 7,708 messages from Klarna, a pay-later service, that miscreants might abuse.

Implications for identity providers

Single sign-on schemes where users can log in to third-party apps with their, e.g. Facebook or Google accounts get more and more popular due to convenience and security reasons (Bauer et al., 2013). Third-party services that allow sign-in with the accounts of identity providers rely on their phone verification process. Consequently, if the phone verification is bypassed, this has implications not only for the identity provider but also for every service where this account is used for authentication. We found 60,345 messages containing one-time passwords for Google, respectively 11,788 messages for Facebook and 7,372 for Apple.

Phone number chaining

We described the use case of Phone number chaining in 5.2.2. The possibility to chain phone numbers not only poses a threat towards phone verification schemes. Also, telecommunication and VoIP providers are affected when their resources get abused by phone number chaining.

5.3.4 Scam messages

Our huge dataset of messages from public SMS inboxes provides a great opportunity for insight into the matter of scams in SMS messaging. A lot of different scam schemes are present in SMS messages. We provide insight and give examples of two kinds of them, phishing and lottery scams. SMS spearphishing is a topic relatively new to the scientific community, compared to phishing attacks in e-mail and voice communication (Liu et al., 2021).

Phishing

We encountered different kinds of SMS phishing messages in our dataset. One kind of message contains the alert that the app for the german bank „Volksbank“ must be reactivated and sends a suspicious link alongside. Messages that inform about a withdrawal on „Binance“ with a suspicious link to login in and stop the transaction are

also present. Another kind of message pretends to leak login credentials for a bitcoin wallet with a link that leads to a login form.

The latter kind of messages might be different from typical phishing messages like the „Volksbank“ case since they are probably targeted at public SMS inbox users and not individuals on their phones. These „leakages“ might seem tempting for users and visitors since they look like legitimate SMS messages. The aim of this scam is probably to steal private keys from individuals. However, we cannot tell if these messages are sent as SMS towards the inbox or if the gateway or website itself ingests them. In 5.2.1, we observed that phishing messages are the most common type of message on some websites. Therefore, we cannot dismiss the suspicion that public SMS inbox websites may ingest these phishing messages.

Message
Ihre VoIksbank 100 App ist abgelaufen. Klicken Sie hier, um es zu reaktivieren: https://***-volks.com/volksbank30dd361
Withdrawal Code: 384738. If you did not request this code, login securely cancel626985-binance****.w**.app and cancel immediately
btc-bank***.net Hello mturla78 password:mon1978 balance:34.87BTC

Table 5.19: Examples of phishing messages.

URLs redacted for safety reasons.

Lottery scams

Classic fortune-telling scams, known from e-mail spam campaigns, are also present in SMS communication. We present one example encountered in our dataset below:

Message
„+19179605819 - My name is Franco Robb the chief operating officer of FACEBOOK nice meeting you!!!. I was assigned to contact you from the CEO of FACEBOOK MARK ZUCKERBERG, There's an online draws that was conducted by a random selection of E-mails and numbers you were picked by an advanced automated random computer search from your phone carrier in other to claim your five hundred thousand US dollars ($500,000.USD)..text the Agent in charge of your winnings (863)225-8437 Congratulations in advance!“

Table 5.20: Examples of fortune-telling scam.

URLs redacted for safety reasons.

5.4 Mitigation

In section 5.2.2 Use cases and 5.3 Threats, we showed the potential use of public SMS inboxes and the resulting threats. From the perspective of a legitimate internet service, it might be desirable to mitigate these threats to ensure the integrity of the userbase, protect data and prevent losses. On the other hand, from the user's perspective, it is desirable to protect one's information and privacy while also being able to hide one's private identifiable information (phone number) against internet services. We propose several approaches to mitigate the stated threats in the following.

5.4.1 Blocking phone numbers

From the perspective of internet services, a possibility to mitigate the use of phone numbers from public SMS inboxes is to block their use. This mitigation feature could be integrated at the level of the service (e.g. during sign-up) or at the level of the sending gateway. Blocking phone numbers would require some denylist. We test if today's available anti-fraud and phone-reputation products, which can be queried with a phone number, effectively detect phone numbers from public SMS inboxes. Further, we propose a mechanism to build a denylist of public SMS inboxes and their phone numbers that could be integrated into these products or at the service level.

Probing APIs

We identified two suppliers that claim to be able to check phone numbers. Telesign[11] and IPQualityScore[12] provide such services that are accessible via an API. We contacted the vendors and asked for a free tier in cooperation. Unfortunately, we could not get free access to these APIs and therefore are limited to the free tier of 39 requests towards the IPQualityScore API. We queried this API with a randomly selected set of phone numbers from our dataset on 27.10.2022. IPQualityScore returns a risk score ranging from 0 to 100 associated with the number. Our probing of 39 numbers resulted in 35 numbers labelled with a fraud score higher than 85, implying a high risk and recent abusive behaviour (IPQualityScore, 2022). Four numbers were labelled with a risk score of 0, indicating no abusive behaviour associated with this number. Even though the sample size of our experiment is too small to make a statistically viable statement, the results hint that phone reputation APIs might be a performant deterrent against public SMS inboxes.

Building denylists

It is not clear how phone reputation APIs, in particular, „IPQualityScore“, gather intelligence of information. Based on marketing information from their website, it seems to be based on feedback from services using that API and the linkage with other fraud

[11] See `https://www.telesign.com/products/intelligence`
[12] See `https://www.ipqualityscore.com/solutions/phone-validation`

signals. Due to the limited amount of requests, we were unable to study the factor of time, particularly how fast public SMS inbox phone numbers get detected. As a possible alternative deterrent, we propose building denylists by scraping the phone numbers displayed on public SMS inbox websites.

5.4.2 Using private apps

For users that wish to hide their phone number for privacy reasons, apps like „Google Voice“[13] or „burnerapp.com“[14] might provide a viable alternative. Although we cannot say that using phone numbers from these services is untraceable, they might provide the amount of privacy most users are longing for.

5.4.3 Not sending PII

Due to the insecurity of SMS in general, services should never send private identifiable information in SMS messages or information that makes them accessible (e.g. hyperlinks). When users decide to use public SMS inbox phone numbers, they should be very cautious about for which services they are using them and that their private information might get leaked. Especially the common procedures in online travel agencies to share and make information accessible to customers, that Cheng et al. (2020), and we studied in 5.3.1, pose a high risk for customers. Nevertheless, from our point of view, there is no risk if the SMS and public inboxes are used as a channel for signalling messages, e.g. notifications.

5.4.4 Authentication

Due to the high risk of account takeover when a phone number is used as the only factor for authentication, users should not use numbers from public SMS inboxes at such services (e.g. messengers like WhatsApp, Telegram and Signal). In case someone uses it for such services, it is advisable to enable two-factor authentication. As stated, we regard the threat when using these phone numbers in two-factor authentication schemes as low but still present. A general alternative to SMS as a second factor in online services can be biometric or U2F/Fido keys (Jin et al., 2021).

5.4.5 Preventing phone number chaining

The issue of phone number chaining, to which extent is not studied yet, can be impactful due to the exponential propagation of threats by enabling miscreants access to phone numbers. Thereby providers of virtual phone services can get a commodity in cybercrime themselves while also potentially getting abused of their resources. However, the threat is limited to those companies offering these communication services, and it is advisable for them to mitigate the threat of PVA evasion for their own sake and to reduce the risk of abuse for other services. Potential measures include the ones mentioned in this chapter.

[13] See `https://voice.google.com/u/0/about`
[14] See `https://www.burnerapp.com/`

5.4.6 Limiting reuse of phone number

Some services allow the reuse of phone numbers to create accounts, e.g. it is reported that 10 Google accounts can be associated with one phone number (Xie et al., 2019). Potential mitigation to reduce the risk of abuse overall can be to limit and reduce the number of accounts that are allowed to be associated with a phone number (Thomas et al., 2014).

5.4.7 Re-verifying phones

With public SMS inboxes, users cannot send messages or make calls from the numbers they use on these websites. This limitation can be taken advantage of to mitigate the risk of using public SMS inboxes by forcing users to respond to an SMS message or accept a call during the phone verification step. However, we see a problem with this approach, as users may be unwilling to send SMS or receive calls due to their mobile tariffs. Such additional steps may harm user experience, but this mitigation could only be enacted on suspicious phone numbers to reduce false-positive and false-negative rates while blocking these numbers.

A second deterrent can be re-verifying the phone number after a certain time. We have shown in 5.2.1 that the lifetime of these numbers is limited and that they have less than 100% uptime. When a service enforces to re-verify phone numbers after a certain amount of time, the fact that users cannot, hints at the usage of public SMS inboxes. Consequently, services can disable these accounts.

6 Conclusion

The thesis aimed to study the ecosystem of public SMS inboxes exploratively. The central research questions were:

- What are the characteristics of websites providing public SMS inboxes?
- How are they used?
- What threats come by using these services?
- What approaches can be made to mitigate these threats?

Some prior research has already studied various aspects of public SMS inboxes. We synthesised this literature to state the current academic research on this topic. By doing this, we discovered limitations in these studies and identified research gaps. In our study, we focused on exploring and answering these research gaps to expand the scientific community's knowledge. Whereas this was possible in many aspects by combining findings from previous research, our own collected dataset and third-party data, we hit barriers where aspects either turned out not to be that relevant, where we lacked data or where it might not be possible to collect viable data at all.

Many public SMS inbox websites share similarities in HTML and Style. We identified clusters and some reasons for it, e.g. using a commercially available theme. The main monetisation scheme for these websites is ads, while also the case of involvement in account takeover and exploitation can be made up but not be proven. We could not make significant findings by analysing terms of service and privacy policies but found insights into the desired and undesired usage of these services. The amount of messages in a public SMS inbox is not tied to their popularity. The lifetime of the public SMS inboxes we collected data on is, overall, on average, 23 days, with an average uptime of 40% in this lifetime. We uncovered that different public SMS inbox domains have ties with each other as they share, to some extent, the same phone numbers. We can not clearly state if this happens due to shared gateways or collaboration, and we also found hints for mutual scraping. Due to the similarity of phone numbers, a share of inbox numbers might be tied to contracts or SIM cards bought in bulk. The provided phone numbers are of mobile or VoIP line type. Roughly one-third of phone numbers are ported to another network. We see huge differences between networks if they take numbers offline. We discovered that the gateways to the mobile networks are not only SMS messaging APIs but also illegal „SIM Boxes“. We were able to prove the connection to SIM boxes by correlating our data with a company specialised in detecting these.

With public SMS inboxes, the most common messages are one-time passwords and codes from internet services or phishing messages. Multipart SMS are present and

assembled. There is a weak correlation between messages and lifetime but a moderate between uptime. The usage varies over the day, with its peak around noon. Several use cases exist, and we can distinguish them into benevolent and malicious usage. For benevolent users, this includes the case of living abroad, testing and privacy reasons. However, also malicious usage is possible, which is not distinguishable from the data. For example, criminals use public SMS inboxes for PVA evasion, number chaining and identity obfuscation.

The usage of public SMS inboxes introduces some threats towards the users of its internet services or third parties. This includes the leakage of private identifiable information either in messages or via links, in the metadata or the breach of authentication, primarily where phone numbers are used as the only factor. However, the biggest threat, especially towards internet services in terms of scale and financial losses, is the enablement of phone-verified account evasion. Miscreants can exploit sign-up incentives, exhaust resources, chain phone numbers, undermine single-sign-on and use the phone numbers as a resource in other cybercrime schemes. In addition, scams and phishing messages are prevalent and pose a threat towards users and visitors. Also, we recognised a phishing message that is probably targeted towards public SMS inbox users.

In response to the identified threats, we propose and probe countermeasures to mitigate the risks. First, we showed that blocking phone numbers based on scoring from phone reputation APIs is a promising approach. We propose the method of building denylists that can enhance phone reputation algorithms by scraping public SMS inboxes. Second, to mitigate privacy leakage, services should avoid sending PII or links to PII in SMS, and users should avoid this kind of service when using public SMS inboxes. Third, we discourage using phone numbers as the only factor in authentication. Services providing VoIP services must consider the abuse case of phone number chaining not to get a commodity themselves and thereby facilitate cybercrime. To limit the impact and detect PVA evasion, services can limit the number of accounts associated with one particular phone number, requesting a message or call response during the verification step or re-verifying phone numbers after a certain time. A risk-based approach might be viable to reduce the impact on user experience.

From our observations, we conclude that public SMS inboxes do not appear as legitimate businesses and rather operate in the grey area. This is due to their vague policies, their inscrutable source of phone numbers (mutual scraping, SIM boxes, malware, bulk SIM cards, shut-down of numbers by networks) and phishing attempts on their website, though illegal behaviour of the websites can not be proven. By providing their services for free, they lower the barrier for PVA evasion since everyone can get an unlimited supply of phone numbers. This must not always result in cybercrime since legitimate use cases exist. However, as with most privacy-enabling technology, these services are also a facilitator of and a commodity in cybercrime. We gave examples of malicious behaviour and can even quantify some damage associated with exploiting internet services. Though we studied only a subset of websites, we can assume that the impact is far more significant than what we observed. By this, it is demonstrated that public SMS inboxes substantially impact real-world internet services that can not be neglected. Consequently, internet

services should evaluate possible mitigation measures.

The most prevalent issue with public SMS inboxes is PVA evasion, confirmed by prior studies of Berenjestanaki et al. (2019) and Reaves et al. (2018). Surprisingly, to our knowledge, the circumvention of phone verification and its consequences are barely researched in the scientific community despite its prevalence as a security feature in many internet services nowadays. PVA evasion is enabled by public SMS inboxes and VoIP and Messaging API providers through number chaining and paid temporary phone numbers. We propose further research on temporary phone numbers in general and PVA evasion in internet services. Temporary phone numbers can be framed as a commodity enabling various kinds of cybercrime, e.g. e-commerce fraud, messenger scams, fake accounts or undermining platform integrity. We speculate that temporary phone numbers substantially impact today's internet economy. Public SMS inboxes and the messages gathered from them can provide insight into the cause.

Still, many aspects of public SMS inboxes exist that are not yet studied and are not part of this thesis. We propose further research that might provide more insight and intersect with other research topics. So it is unclear to what extent account takeover is a concern when phone numbers are used as the only authentication factor. An experiment can be set up by registering accounts with these public phone numbers and monitoring possible takeover attempts. More data sources can be incorporated to understand the usage better and define more granular use cases. E.g. it can be analysed for what the registered accounts are used, e.g. by importing public SMS inbox phone numbers as contacts, identifying social media and messenger profiles via „find your friends“ functionality and then surveying these accounts.

The resource of millions of SMS messages can be used to study the topics of SMS spam and phishing, e.g. in the context of machine-learning classifiers. Nevertheless, it must be considered that not all messages might be authentic SMS traffic due to the possible injection of messages by the websites and receiving gateways, e.g. for phishing attempts. Therefore, the reliability and authenticity of the message flow from messaging APIs, sending gateways, and receiving gateways towards the public SMS websites can be studied by initiating or sending SMS messages towards these numbers. Further, comparing public SMS inbox websites on what messages they might receive can uncover possible traffic manipulation.

Bibliography

addpac.su. Derive from "addpac gsm voip gateways" [jpg]". Available online https://commons.wikimedia.org/wiki/File:Addpac_gsm_voip_gateways.jpg, 2010. Accessed: 2022-09-25.

Kuross Amri. Available online https://web.archive.org/web/20160511143408/http://services.eng.uts.edu.au/userpages/kumbes/public_html/ra/sms/. Accessed: 2022-09-15.

Iosif I. Androulidakis. *SMS Security Issues*, pages 71–86. Springer International Publishing, Cham, 2016. ISBN 978-3-319-29742-2. doi: 10.1007/978-3-319-29742-2_5. URL https://doi.org/10.1007/978-3-319-29742-2_5.

Lujo Bauer, Cristian Bravo-Lillo, Elli Fragkaki, and William Melicher. A comparison of users' perceptions of and willingness to use google, facebook, and google+ single-sign-on functionality. In *Proceedings of the 2013 ACM Workshop on Digital Identity Management*, DIM '13, page 25–36, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450324939. doi: 10.1145/2517881.2517886. URL https://doi.org/10.1145/2517881.2517886.

BBC. Hppy bthdy txt! Available online https://web.archive.org/web/20220913093119/http://news.bbc.co.uk/2/hi/uk_news/2538083.stm, 2002. Accessed: 2022-09-13.

Md. Hajian Berenjestanaki, Mauro Conti, and Ankit Gangwal. On the exploitation of online sms receiving services to forge id verification. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*, ARES '19, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450371643. doi: 10.1145/3339252.3339276. URL https://doi.org/10.1145/3339252.3339276.

Yanan Cheng, Han Wang, Zhaoxin Zhang, and Ning Li. Characterizing the security threats of disposable phone numbers. In Guangquan Xu, Kaitai Liang, and Chunhua Su, editors, *Frontiers in Cyber Security*, pages 491–507, Singapore, 2020. Springer Singapore. ISBN 978-981-15-9739-8.

Andrei Costin, Jelena Isacenkova, Marco Balduzzi, Aurélien Francillon, and Davide Balzarotti. The role of phone numbers in understanding cyber-crime schemes. In *2013 Eleventh Annual Conference on Privacy, Security and Trust*, pages 213–220, 2013. doi: 10.1109/PST.2013.6596056.

Phil Davis. Seven people charged with running 'cashapp' drug trafficking ring in west baltimore, police say. Available online https://web.archive.org/web/20211104082425/https://www.baltimoresun.com/news/crime/bs-md-ci-cr-cashapp-gang-baltimore-arrests-20210624-fkycnhoqtbb23h24apry47n754-story.html, 2021. Accessed: 2022-11-03.

Virginia Dei. Design und implementierung von social bots: Methoden der künstlichen intelligenz zur generierung von automatisierten nachrichten. Master's thesis, Hochschule Wismar, 2019.

Nurullah Demir, Matteo Große-Kampmann, Tobias Urban, Christian Wressnegger, Thorsten Holz, and Norbert Pohlmann. Reproducibility and replicability of web & measurement studies. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 533–544, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450390965. doi: 10.1145/3485447.3512214. URL https://doi.org/10.1145/3485447.3512214.

Zhengyu Dong, Ryan Flores, Vladimir Kropotov, Paul Pajares, and Fyodor Yarochkin. Sms pva: An underground service enabling threat actors to register bulk fake accounts. Technical report, Trend Micro Research, 2022. Available online https://documents.trendmicro.com/assets/white_papers/wp-sms-pva-underground-service-enabling-threat-actors-to-register-bulk-fake-accounts.pdf.

Tobias Engel. Available online https://berlin.ccc.de/~tobias/31c3-ss7-locate-track-manipulate.pdf. Accessed: 2022-09-19.

Mobile Ecosystem Forum. Available online https://mobileecosystemforum.com/wp-content/uploads/2020/11/MEF-Business-SMS-SIM-Farms-The-Data-Protection-Risk.pdf, 2020. Accessed: 2022-09-24.

Thamme Gowda and Chris A. Mattmann. Clustering web pages based on structure and style similarity (application paper). In *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, page 175–180. IEEE Press, 2016. doi: 10.1109/IRI.2016.30. URL https://doi.org/10.1109/IRI.2016.30.

haveibeenpwned.com. Have i been pwned. Available online https://web.archive.org/web/20221025181257/https://haveibeenpwned.com/PwnedWebsites. Accessed: 2022-10-25.

hlr lookups.com. Available online https://www.hlr-lookups.com/en/api-docs. Accessed: 2022-09-21.

Hang Hu, Peng Peng, and Gang Wang. Characterizing pixel tracking through the lens of disposable email services. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 365–379, 2019. doi: 10.1109/SP.2019.00033.

Luca Invernizzi, Kurt Thomas, Alexandros Kapravelos, Oxana Comanescu, Jean-Michel Picod, and Elie Bursztein. Cloak of visibility: Detecting when machines browse a different web. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 743–758, 2016. doi: 10.1109/SP.2016.50.

IPQualityScore. Phone number validation api documentation. Available online `https://web.archive.org/web/20221107093842/https://www.ipqualityscore.com/documentation/phone-number-validation-api/overview`, 2022. Accessed: 2022-11-07.

Weizhao Jin, Xiaoyu Ji, Ruiwen He, Zhou Zhuang, Wenyuan Xu, and Yuan Tian. Sms goes nuclear: Fortifying sms-based mfa in online account ecosystem. In *2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*, pages 7–14, 2021. doi: 10.1109/DSN-W52860.2021.00013.

Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. Tranco: A research-oriented top sites ranking hardened against manipulation. In *Proceedings of the 26th Annual Network and Distributed System Security Symposium*, NDSS 2019, February 2019. doi: 10.14722/ndss.2019.23386.

Mingxuan Liu, Yiming Zhang, Baojun Liu, Zhou Li, Haixin Duan, and Donghong Sun. Detecting and characterizing sms spearphishing attacks. In *Annual Computer Security Applications Conference*, ACSAC '21, page 930–943, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450385794. doi: 10.1145/3485832.3488012. URL `https://doi.org/10.1145/3485832.3488012`.

Philipp Mayring. *Qualitative Inhaltsanalyse: Grundlagen und Techniken. 12., überarb. Aufl.* Beltz, Weinheim u.a, 2015. ISBN 3407293933.

Mobilesquared. Nubmer of businesses using a2p sms leaps 20% in response to pandemic. Available online `https://web.archive.org/web/20220913092333/https://mobilesquared.co.uk/2021/07/05/number-of-businesses-using-a2p-sms-leaps-20-in-response-to-pandemic/`, 2021. Accessed: 2022-09-13.

Ilona Murynets, Michael Zabarankin, Roger Piqueras Jover, and Adam Panagia. Analysis and detection of simbox fraud in mobility networks. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pages 1519–1526. IEEE, 2014.

Okta. What is sms authentication and is it secure? Available online `https://web.archive.org/web/20220913093339/https://www.okta.com/blog/2020/10/sms-authentication/`, 2020. Accessed: 2022-09-13.

Lilian Perron and Company Araxxe. Personal communications, 2022. `https://www.araxxe.com/`.

privacyinternational.com. Sim card registration. Available online https://web.archive.org/web/20221025181234/https://www.privacyinternational.org/learn/sim-card-registration. Accessed: 2022-10-25.

Bradley Reaves, Nolen Scaife, Dave Tian, Logan Blue, Patrick Traynor, and Kevin R. B. Butler. Sending out an sms: Characterizing the security of the sms ecosystem with public gateways. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 339–356, 2016. doi: 10.1109/SP.2016.28.

Bradley Reaves, Luis Vargas, Nolen Scaife, Dave Tian, Logan Blue, Patrick Traynor, and Kevin R. B. Butler. Characterizing the security of the sms ecosystem with public gateways. *ACM Trans. Priv. Secur.*, 22(1), dec 2018. ISSN 2471-2566. doi: 10.1145/3268932. URL https://doi.org/10.1145/3268932.

serpapi.com. Available online https://serpapi.com. Accessed: 2022-09-15.

Imani N Sherman, Jasmine D Bowers, Keith McNamara Jr, Juan E Gilbert, Jaime Ruiz, and Patrick Traynor. Are you going to answer that? measuring user responses to anti-robocall application indicators. In *NDSS*, 2020.

thesmswork.co.uk. A guide to sms sender id. Available online https://web.archive.org/web/20221008093500/https://thesmsworks.co.uk/blog/sms-sender-id/. Accessed: 2022-10-08.

Kurt Thomas, Dmytro Iatskiv, Elie Bursztein, Tadek Pietraszek, Chris Grier, and Damon McCoy. Dialing back abuse on phone verified accounts. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, page 465–476, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450329576. doi: 10.1145/2660267.2660321. URL https://doi.org/10.1145/2660267.2660321.

Kurt Thomas, Danny Huang, David Wang, Elie Bursztein, Chris Grier, Thomas J. Holt, Christopher Kruegel, Damon McCoy, Stefan Savage, and Giovanni Vigna. Framing dependencies introduced by underground commoditization. In *Workshop on the Economics of Information Security*, 2015.

Paul A. Watters, Aaron Herps, Robert Layton, and Stephen McCombie. Icann or icant: Is whois an enabler of cybercrime? In *2013 Fourth Cybercrime and Trustworthy Computing Workshop*, pages 44–49, 2013. doi: 10.1109/CTC.2013.13.

Jane Webster and Richard T. Watson. Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2):xiii–xxiii, 2002. ISSN 02767783. URL http://www.jstor.org/stable/4132319.

wikipedia.org. Available online https://en.wikipedia.org/wiki/Bandwidth_(company), a. Accessed: 2022-09-22.

wikipedia.org. Available online https://en.wikipedia.org/wiki/E.164, b. Accessed: 2022-09-22.

Tian Xie, Sihan Wang, Guan-Hua Tu, Chi-Yu Li, and Xinyu Lei. Exploring the insecurity of google account registration protocol via model checking. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 3087–3096. IEEE, 2019.

A Appendix

A.1 List of identified public SMS inboxes

source	searchString	rank	host
tranco	sms	35952	receive-smss.com
tranco	sms	43785	receive-sms-free.cc
tranco	sms	51937	smsreceivefree.com
tranco	sms	56657	sms24.me
tranco	sms	62437	receivesms.co
tranco	sms	63741	freereceivesms.com
tranco	sms	67428	receive-sms.cc
tranco	sms	79431	receive-sms-online.info
tranco	sms	81847	mytempsms.com
tranco	sms	136635	zusms.com
tranco	sms	148567	z-sms.com
tranco	sms	153654	online-sms.org
tranco	sms	155547	getfreesmsnumber.com
tranco	sms	228623	sms-online.co
tranco	sms	251107	receive-sms.com
tranco	sms	255851	sms-acktiwator.ru
tranco	sms	266118	sms-receive.net
tranco	sms	286212	receivesmsonline.net
tranco	sms	412300	smsreceive.site
tranco	sms	458794	receiveasms.com
tranco	sms	537287	receivesms.live
tranco	sms	562938	receivefreesms.net
tranco	sms	622696	yunjisms.xyz
tranco	sms	625243	temp-sms.org
tranco	sms	646473	smstools.online
tranco	sms	656010	receive-sms-online.cc
tranco	sms	716974	freesmscenter.com
tranco	sms	750099	xnsms.com
tranco	sms	864914	smstome.com
tranco	sms	910094	smsonline.cloud
tranco	sms	943229	smser.net
tranco	sms	951569	free-sms-receive.com
tranco	sms	957570	receivesmsfast.com
tranco	sms	987309	bestreceivesms.com
tranco	sms	991740	smsbo.com
google	receive sms	9	freephonenum.com
google	receive sms	14	pingme.tel
google	receive sms	28	temporary-phone-number.com
google	receive sms	42	textrapp.com
google	receive sms	44	us-phone-number.com
google	receive sms free	8	freephonenum.com
google	receive sms free	14	7sim.net
google	receive sms free	15	quackr.io
google	receive sms free	28	onlinesim.ru
google	receive sms free	29	tesms.net
google	receive sms free	34	onlinesim.io
google	receive sms free	38	bfkdim.com
google	receive sms free	42	sms.sellaite.com
google	receive sms free	47	spytm.com
google	receive sms online	39	freeonlinephone.org
google	temporary sms	9	temp-number.com
google	temporary sms	27	temp99.com
google	temporary sms	36	number4sms.com
google	temporary sms free	49	karteplus.com
google	public sms inbox	38	temp-mails.com
bing	receive sms	4	receivesms.info
bing	receive sms	10	smsreceivefree.net
bing	receive sms	17	cloakmobile.com
bing	receive sms free	17	online-receive-sms.com
bing	receive sms free	19	onlinesmsbox.com
bing	receive sms free	22	receivefreesms.info
bing	receive sms free	24	receivesms365.com
bing	receive sms free	30	virtualwebphone.com
bing	receive sms free	40	freepublicsms.com
bing	receive sms online	9	receiveasmsonline.com
bing	receive sms online	32	indiannumber.com
bing	temporary sms	39	temp-sms.net
cheng			yinsiduanxin.com
cheng			mianfeijiema.com
cheng			yunduanxin.net
cheng			yunjiema.net

A.2 Ethical board review

Beirat für ethische Fragen in der wissenschaftlichen Forschung, Universität Innsbruck
(Geschäftsstelle: Robert Rebitsch / projekt.service.büro / Vizerektorat für Forschung)

An das
Büro der Vizerektorin für Forschung der Universität Innsbruck
zu Handen des betreffenden Review Boards der Fakultät

Beirat für ethische Fragen in der wissenschaftlichen Forschung Geschäftsstelle: Robert Rebitsch, projekt.service.büro **@: Robert.Rebitsch@uibk.ac.at**			
Review Board „Sportwissenschaft“ Institut für Sportwissenschaften z.Hd. Fr. Laura Rietzler @: Laura.Rietzler@uibk.ac.at	Review Board „Psychologie“ Institut für Psychologie z.Hd. Hr. Univ.-Prof. Dr. Pierre Sachse @: Pierre.Sachse@uibk.ac.at	Review Board „Sozialwissenschaften“ Fakultät für Betriebswirtschaft und Fakultät für Volkswirtschaft und Statistik z.Hd. Hr. Univ.-Prof. Dr. Heiner Schumacher @: Heiner.Schumacher@uibk.ac.at	Review Board „LehrerInnenbildung“ Fakultät für LehrerInnenbildung z.Hd. Fr. Univ.-Prof. Dr. Suzanne Kapelari @: Suzanne.Kapelari@uibk.ac.at

Eingangsvermerk: ……

Erledigungsvermerk: ……

Ansuchen um Prüfung der Unbedenklichkeit eines Forschungsprojektes

Name des/der Projektleiters/Projektleiterin: Dr. Svetlana Abramova & Dr. Daniel Woods

Projekttitel: End-to-end study of the ecosystem of public SMS inboxes

Institut: Institut für Informatik

Datum: 19.07.2022

Projektbeschreibung – max. zwei Seiten
Hat zu enthalten: Beschreibung der Zielsetzung und des wissenschaftlichen Hintergrundes der Studie, ihres Neuwerts, verwendete Methodik, das mögliche Risiko für die Proband/inn/en, ethisch relevante Punkte, Globalbudget oder Drittmittel (Forschungsförderung/Auftragsforschung), eine Abwägung des Nutzen-Risiko-Verhältnisses, Fallzahlschätzung sowie die wichtigste wissenschaftliche Literatur; Begründung, weshalb eine Probanden/Probandinnen-Versicherung nötig bzw. nicht nötig ist; Nachweis der für die Studie erforderlichen wissenschaftlichen oder sonstigen Qualifikationen der/des Studienleiter/in/s und sonst maßgeblich an der Studie beteiligten Personen.

Projektbeschreibung bitte hier einschreiben

In this project, we plan to scrape and analyse publicly available data from SMS inbox websites. To be completely clear, we will study public SMS inboxes that already exist. The service providers advertise a phone number. Whenever someone sends an SMS message to that phone number, it is subsequently posted on a public website. You can see an example here: https://receive-smss.com/sms/447846037301/. Such services are completely open and do not require a registration / login. They can be used, for example, for getting a one-time password when someone wants to register an online account wihout sharing her personal number with a third-party serice provider. A required authentication code will be then sent to the public SMS inbox and published on the public website, which can be looked up by the registering user.
Public SMS inboxes are studied only exploratively in [1, 2, 3]. We recognise the research value of analysing public SMS messages for the potential to identify non-apparent security and privacy risks for users using such services. We propose interacting with the SMS inbox in two ways:

Beirat für ethische Fragen in der wissenschaftlichen Forschung, Universität Innsbruck
(Geschäftsstelle: Robert Rebitsch / projekt.service.büro / Vizerektorat für Forschung)

(1) Collecting the messages in the public SMS inbox, which are already in the public domain. It is possible these messages contain personal data under the GDPR. We justify the research based on our legitimate interest as academics conducting research in the public interest and the fact that we can not obtain consent because we have no way of interacting with users of the service. Note, we never send messages to users of the service.

(2) We may use the public SMS phone number to register for services. This would involve giving the phone number to other websites as part of the registration process. Notably, the phone number is owned by a business (i.e., a provider of a public SMS inbox) and is not personal data. We may also send our own messages to the public SMS inbox, but again this is owned by a business.

For (1), we will scrape public SMS messages and conduct a semi-automated analysis of the collected data using rule-based classification and Natural Language Processing (NLP) algorithms. The goal is to characterize messages into subcategories related to a specific use case or application scenario. Based on this categorization, we plan to identify security threats & attack vectors of public SMS messages and suggest potential countermeasures for users as well as organisations. Furthermore, we plan to scrape Terms of Service and Privacy Policies published by SMS inbox providers to do a qualitative analysis of their legal rules and bases of operation.

From a legal perspective, we intend to scrape publicly available data for research purposes. Users should be aware that any message sent to a public SMS inbox is eventually publicly available for some period of time. In terms of privacy and ethics, the collected dataset may contain personally identifiable information (PII), such as names, user names, addresses, or passwords. Furthermore, there is a risk that some phone numbers may originate from stolen SIM cards. The involved researchers will anonymize any detected PII whenever possible. We won't make use of potential login credentials found in text messages to access private accounts. The collected dataset will not be published publicly. However, for the purpose of reproducibility of results, it will be made available to other researchers upon a written request and a signed non-disclosure agreement. The collected dataset will be treated securely and confidentially (using the encryption + password protection).

For (2), the researchers plan to use public phone numbers and send test messages to them, either directly using a new SIM card, or when registering an account at other providers (e.g., Mjam, Whatsapp etc.) or via Voice over IP (VoIP). The involved researchers are advised against using own mobile SIM cards for this project. In order to reduce costs on the SMS inbox provider's side and not to spam third-party providers, no more than 3 messages per a public number will be sent on average.

The project will be carried out by Dr. Svetlana Abramova and Dr. Daniel Woods, with a technical support of two Master students, Benjamin Vettori and Leonhard Zacharias, from the Departments of Computer Science and Informaton Systems, respectively. The project leaders plan to submit and publish the final results in an abstract and aggregate form in a conference's proceedings. The scraping code (without data) will be released to the public.

Beirat für ethische Fragen in der wissenschaftlichen Forschung, Universität Innsbruck
(Geschäftsstelle: Robert Rebitsch / projekt.service.büro / Vizerektorat für Forschung)

[1] B. Reaves, N. Scaife, D. Tian, L. Blue, P. Traynor and K.R.B. Butler. (2016). "Sending Out an SMS: Characterizing the Security of the SMS Ecosystem with Public Gateways," IEEE Symposium on Security and Privacy, 2016, pp. 339-356, doi: 10.1109/SP-2016.28.

[2] B. Reaves, L. Vaargas, N. Scaife, D. Tian, L. Blue, P. Traynor and K.R.B. Butler. (2019). "Characterizing the Security of the SMS Ecosystem with Public Gateways," ACM Transactions on Privacy and Security 22, 1, doi: 10.1145/3268932.

[3] Y. Cheng, H. Wang, Z. Zhang and N. Li. (2020). „Characterizing the Security Threats of Disposable Phone Numbers," In: G. Xu, K. Liang, C. Su (eds) Frontiers in Cyber Security. FCS 2020. Communications in Computer and Information Science, vol. 1286. Springer, Singapore, doi: 10.1007/978-981-15-9739-8_37.

Unterschrift:

.....
(Projektleiterin/Projektleiter)

Beirat für ethische Fragen in der wissenschaftlichen Forschung, Universität Innsbruck
(Geschäftsstelle: Robert Rebitsch / projekt.service.büro / Vizerektorat für Forschung)

Beizulegen sind:

A) Proband/inn/eninformation und Einwilligungserklärung für ProbandInnen

Die Proband/inn/en sind über folgende Punkte zu informieren: Projektziele, Projektverantwortliche, vorgesehene Laufzeit, mögliches Risiko bzw. Belastungen, Aufwandsentschädigungen, Grundlagenforschung oder Auftragsforschung (bei Auftragsforschung ist Angabe des Unternehmens nicht verpflichtend). Die Information ist zielgruppenorientiert und verständlich zu gestalten (z.B. bei Projekten mit Kindern).

Einwilligungserklärungen müssen zumindest enthalten:
- Genaue Angabe des Studientitels
- Hinweise zum Datenschutz (anonymisiert/pseudonymisiert; vertrauliche Behandlung aller personenbezogenen Daten bzw. indirekt personenbezogenen Daten, Recht auf Einsichtnahme)
- Ausdrückliche Formulierung der Einwilligung der/des Studienteilnehmer/in/s plus Unterschrift (gegebenenfalls auch der gesetzlichen Vertreter – siehe unten)
- Möglichkeit des Widerrufs durch Studienteilnehmer/innen
- Kontaktdaten der/des Projektleiter/in/s bzw. der Projektverantwortlichen plus Unterschrift

Im Falle, dass die Proband/inn/en eine Erklärung unterfertigen, bitte diese im Originaltext beilegen.

Unterschriftsbefugnisse: Bei Kindern und unmündigen Minderjährigen (bis zum vollendeten 14. Lebensjahr) muss die/der gesetzliche Vertreter/in unterfertigen, bei mündigen Minderjährigen (ab vollendetem 14. bis zum vollendeten 18. Lebensjahr) die Person selbst und deren gesetzliche/r Vertreter/in.

Für Formulierungen der Einwilligungserklärung ist die Geschäftsstelle behilflich.

B) Case Report Form (Dokumentationsbogen für die zu erhebenden Daten)

C) Versicherungspolizze

Im Fall einer notwendigen Versicherung ist die Versicherungspolizze beizulegen. Information zu den bestehenden Versicherungen der Universität finden Sie auf der Homepage der Rechtsabteilung unter: http://www.uibk.ac.at/rechtsabteilung/versicherungenschaeden.html#Versicherungen

WEITERES VERFAHREN:

Nach Einlangen der geforderten Unterlagen über das betreffende Review Board der Fakultät befasst die Vizerektorin für Forschung gegebenenfalls den Beirat für ethische Fragen in der wissenschaftlichen Forschung. Im Fall, dass an der einreichenden Fakultät noch kein Review Board institutionalisiert ist, erfolgt die Einreichung an das Büro der Vizerektorin für Forschung. Diese befasst dann den Beirat. Der Beirat berät über die Unbedenklichkeit des Projektes. Die Unbedenklichkeit kann unter aufschiebenden Bedingungen (Änderungswünsche, Auflagen) gewährt werden oder die vorläufige Projektablehnung vorangehen. Etwaige Änderungswünsche oder Auflagen werden den AntragstellerInnen schriftlich mitgeteilt. Alle Projektunterlagen, Sitzungsprotokolle und Schriftverkehr werden durch die Geschäftsstelle zentral archiviert. Die Unterlagen unterliegen der Verschwiegenheitspflicht und dem Datenschutz; eine Einsicht ist nach Abschluss des Verfahrens nur bei Vorliegen eines rechtlichen Interesses möglich und bei der Vizerektorin für Forschung zu beantragen.

Bei positiver Erledigung wird eine Unbedenklichkeitsbescheinigung durch das Büro der Vizerektorin für Forschung ausgestellt!

Für Nachfragen:
Priv.-Doz. Mag. Dr. Robert Rebitsch
projekt.service.büro der Universität Innsbruck
M: Robert.Rebitsch@uibk.ac.at
T: 0512 507 34407

A.3 Qualitative content analysis - privacy policies

document	quotation	codes
https://www.temp-mails.com/number	1)- This website is public, so DO NOT USE these numbers to receive important messages.	Warning sensitive infroamtion
https://www.temp-mails.com/number	2)- All people can see only the last 100 messages or messages that were received in the last 7-8 days.	Publicly availability
https://www.temp-mails.com/number	3)- You can not see the full number of the sender and it will look like this: +1333333####.	Sender annonymization
receive-smss.com	Please do not use these phone numbers to receive important messages or PINs.	Warning sensitive infroamtion
receive-smss.com	The content of the message is accessible by every user.	Publicly availability

A.4 Qualitative content analysis - terms of service

document	quotation	codes
https://7sim.net/	Any attempts to use automation programs will be blocked.	Automation, Undesired Usage
https://7sim.net/	Any illegal activity related to the use of our numbers is strictly prohibited, and your data can be transferred to the appropriate authorities.	Illegal acitivity, Undesired Usage
https://smstools.online/receive-free-sms/#terms-of-use	ensure yourselves against intrusive advertising.	Prevent Advertising, Intended Usage
https://smstools.online/receive-free-sms/#terms-of-use	Event and distribution participation. A large number of websites run campaigns and free distribution of different worth things (for example, digital keys for some software), and they add an association to a phone number in order to not allow to people to overdo their actions, taking a large number of same presents per customer. Free SMS numbers allow to pass these issues and to collect perks by the hundreds literally, either you want to keep or to resale it.	Exploit Discounts, Intended Usage
https://temporary-phone-number.com/privacy-terms/	hese phone numbers provided on this APP is only used to register some APPs to prevent being harassed.	Annonymisation of user, Intended Usage
https://smstools.online/receive-free-sms/#terms-of-use	If the messages sent contain such private information as the login, password, info with the help of which the frauds can log in your account, as well as the service (sending you the message), everyone might use it. So, be aware of this before sending such private information.	Sensitive Information, Undesired Usage
https://smstools.online/receive-free-sms/#terms-of-use	Passing geography-specific issues. It is known that there are cases when one or another project on the net do not allow to register for people from some countries. It may happen because of some bureaucratic stuff, for example, if one of the business partners of some service buys out exclusive rights on working with users of your region, but often the reason is so common at all. SMS of local operator just do not arrive in the foreign addressee. Our service deals with this issue too, providing free SMS numbers from different countries.	Bypass Geo Blocking, Intended Usage

https://receive-sms-online.cc/Terms/	Please do not use this phone number to receive important content.	Sensitive Information, Undesired Usage
https://receivesms365.com/terms	Please under no circumstances use the phone numbers detailed on this site to receive important or sensitive content.	Sensitive Information, Undesired Usage
https://smstools.online/receive-free-sms/#terms-of-use	Protection against intruders. Safety for Internet users is very important, especially for who makes financial business in an online environment and keeps meaningful data on the computer. If the website, on which you want to register, carry little credibility, it is a headlong decision to enter personal information, whether that be your full name, account number or phone number. Making a “fake” save you from risk because free SMS numbers are not assigned to neither your personality nor your IP-address you come on the website of our service.	Annonymisation of user, Intended Usage
https://smstools.online/receive-free-sms/#terms-of-use	Realize a plural of registries on websites. It is often that creates an account needs not only to note your telephone number but limits users on the principle of “one account – one number”. But you can make a large number of profiles, for example, on Facebook, Google or eBay, as many as you want. It may be useful specifically to publishers and SMM proficients using accounts on social networks for promoting their projects and making money on this.	Multiple accounts, Intended Usage
https://smstools.online/receive-free-sms/#terms-of-use	Save the anonymity. It is notorious knowing the number allows finding out a lot of facts about its owner, up to a full file and location address. Far from everybody can make mind to it. Fortunately, free SMS numbers save you from the necessity to show your real number on the Internet.	Annonymisation of user, Intended Usage
https://smstools.online/receive-free-sms/#terms-of-use	The messages coming from the list of the forbidden addresses (such as bank\ credit organizations, electronic payment systems and the similar projects – you can see this list on our website) will not be shown on the page of the incoming SMS.	Sensitive Information, Undesired Usage
https://receivesms365.com/terms	The numbers listed on this site are marketed for use where privacy is a concern and should be treated as such.	Annonymisation of user, Intended Usage
https://receive-sms.cc/Privacy/	The phone number provided on this website is only used to register some websites to prevent being harassed.	Annonymisation of user, Intended Usage

https://receive-sms-online.cc/Terms/	The phone number provided on this website is only used to register some websites to prevent being harassed.	Annonymisation of user, Intended Usage
https://receivesms365.com/terms	The phone numbers detailed on this site are to be used to purely for lawful or educational reasons, it is your responsibility to ensure by using the telephone numbers listed here that you are not violating a websites terms of service	Illegal acitivity, Undesired Usage
https://www.free-sms-receive.com/sites/policies.html	The phone numbers provided on this website are only used to register some websites to prevent harassment	Annonymisation of user, Intended Usage
https://smstools.online/receive-free-sms/#terms-of-use	The SMS receiving for free should be available to every user, and those who use the automated methods can just interfere with the normal operation of the resource. If you try something like that, you will be blocked.	Automation, Undesired Usage
https://smstools.online/receive-free-sms/#terms-of-use	The virtual numbers for SMS receiving are not created for the criminal or illegal activity usage (such as fraudulent operations, the distribution of drugs and something like that) and if the law enforcement agencies ask for the information connected to the cases like those, the service will give them the necessary data.	Illegal acitivity, Undesired Usage
https://www.freereceivesms.com/privacy/	These phone number should not be used for any sensitive transactions And any illegal use and any terrorist uses.	Sensitive Information, Illegal acitivity, Undesired Usage
https://temporary-phone-number.com/privacy-terms/	These phone number should not be used for any sensitive transactions And any illegal use and any terrorist uses.	Illegal acitivity, Sensitive Information, Undesired Usage
https://www.freereceivesms.com/privacy/	These phone numbers provided on this APP is only used to register some APPs to prevent being harassed.	Annonymisation of user, Intended Usage

A.5 Most common messages

Host	Message content	#
7sim.net	"[www.usdtmarket.in] Your transfer has been successful, current balance: $1,289,287.52 USDT, account number: david p***word: 525252, please do not disclose"	154
7sim.net	"[www.paxxusdt.com] Your transfer has been successful, current balance: $1,189,287.52 USDT, account number: david p***word: 525252, please do not disclose"	142
7sim.net	"[www.paxxusdt.com] Hi David! New account: David, p***word: 525252, current balance: $1,189,287.52 USDT, please do not share this information with anyone"	128
7sim.net	.copied content from https://receivesms.cc/sms/447533431353	100
7sim.net	"[www.usdtmarket.in] Hi David! New account: david, p***word: 525252, current balance: $1,289,287.52 USDT, please do not share this information with anyone"	92
online-sms.org	Copied from receive-smss	4029
online-sms.org	Copied from receive-smss 34681999929	2792
online-sms.org	Copied from receive-smss 34681993330	2116
online-sms.org	Copied from receive-smss 447548032886	2008
online-sms.org	Copied from receive-smss 447548032890	1428
receive-sms-free.cc	"[trcausdvip.com]Account: lee1960 Password: 19601206 Balance: 1,840,768.67USDT"	47134
receive-sms-free.cc	"[usdtcoinbb.vip] Hi Dannis! New acc0unt: PeX689 Passw0rd: 525252 Bal: 1,189,287.52 USDT, please do not share this information with anyone"	35033
receive-sms-free.cc	"[usdtcoinbb.vip] Your account key has been reset, please keep it safe. Username: Pex689,Password: 525252,Bal: 1,189,287.52USDT. Only website : usdtcoinbb.vip"	31002
receive-sms-free.cc	"New Login [usdtkkc.com]Account:PeX689 Password:525252 Balance:1,091,768USDT"	30853
receive-sms-free.cc	"[usdtkkc.com]Hello[PeX689] password:525252 balance:1,091,768USDT"	30678
receive-smss.com	Welcome to Uber! Your app is designed to make your booking experience smooth. Learn more: t.uber.com/hey	15441
receive-smss.com	"Dear Rider, welcome to Uber! Your Uber app is designed to make your booking experience smooth. Learn how to request an AC car in minutes here: t.uber."	15339
receive-smss.com	WhatsApp code 401-572	13550
receive-smss.com	VK: 32796 - use this code to activate your VK profile.	13412
receive-smss.com	Your DENT code is: 817302	11713

receivesms.co	An existing Discord account is already using this number. Please remove it before it can be used with a new account.	3584
receivesms.co	”Doh, you already have a profile attached to this number and we currently only allow one profile per mobile number.”	2364
receivesms.co	36 Driver is delayedReply STOP to opt out	2278
receivesms.co	A new seller near you just requested a cash offer! Check marketplace for details. To unsubscribe txt STOP	2094
receivesms.co	”[usdtcoinbb.vip] Hi Dannis! New acc0unt: PeX689 Passw0rd: 525252 Bal: 1,189,287.52 USDT, please do not share this information with anyone”	1853
receivesms.live	”Dear Customer, your PayBy Password has been successfully changed. If you did not make this change, please contact us immediately on ¡¿.”	101
receivesms.live	Bienvenue sur Amazon Prime. Profitez de vos avantages dès maintenant. RDV sur amazon.fr dans Votre Compte ¿ Centre de messages pour plus ’informations.	67
receivesms.live	Cash App: Choprgetchop requested $1.05. Open Cash App to approve or decline	65
receivesms.live	Cash App: Choprgetchop requested $2.01. Open Cash App to approve or decline	55
receivesms.live	”Welcome to SMS messages from Square - Reply w/ ””HELP”” for more or ””END”” to unsubscribe from receiving messages, std rates apply”	43
receivesms365.com	Hi	37
receivesms365.com	None	26
receivesms365.com	Hey	7
receivesms365.com	¡#¿ Your WhatsApp code: 494-120You can also tap on this link to verify your phone: v.whatsapp.com/494120Don’t share this code with others4sgLq1p5sV6	7
receivesms365.com	Account notification: The password for your Google Account was recently changed. google.com/password	6
receivesmsfast.com	”Your Amazon account is temporarily on hold. To resolve, sign in now.”	107
receivesmsfast.com	An existing Discord account is already using this number. Please remove it before it can be used with a new account.	104
receivesmsfast.com	Welcome to Turo trip alerts! Message and data rates may apply. Update your notification settings at https://turo.com/account.	96
receivesmsfast.com	”Dear Customer, your PayBy Password has been successfully changed. If you did not make this change, please contact us immediately on ¡¿.”	87

receivesmsfast.com	"Doh, you already have a profile attached to this number and we currently only allow one profile per mobile number."	75
sms24.me	WhatsApp code 401-572	1987
sms24.me	28843 e o codigo de confirmacao do Facebook de Pedro Davi #fb	1814
sms24.me	Use 248173 as your login code for Tinder. (Account Kit by Facebook)	1776
sms24.me	[#] Your Uber code is 3868 qlRnn4A1sbt	1699
sms24.me	Welcome to Uber! Your app is designed to make your booking experience smooth. Learn more: t.uber.com/hey	1662
tesms.net	"[www.dvxios.com] Your withdrawal has been received, the balance is $868,562.84USDT, account number: shun5678, password: 561783"	244
tesms.net	"[www.usdtmarket.in] Your transfer has been successful, current balance: $1,289,287.52 USDT, account number: david password: 525252, please do not disclose"	86
tesms.net	?i?t?h? ?a?n?y?o?n?e? ?e?l?s?e?.? ?h?t?t?p?s?:?/?/?g?o?o?.?g?l?/?U?E?R?g?F?7? ??o?w?B?E?k?0?t?b?e?f?D	54
tesms.net	"[www.usdtmarket.in] Hi David! New account: david, password: 525252, current balance: $1,289,287.52 USDT, please do not share this information with anyone"	38
tesms.net	"[www.usdtmarket8.com]Your $889,990.36USDT has been credited. Account: Win188 Password: W898989 Quick withdrawal"	12

A.6 Most common senders

Host	Sender	#
7sim.net	Skype	20016
7sim.net	Discord	7382
7sim.net	Instagram	6428
7sim.net	Google	6151
7sim.net	Facebook	6089
online-sms.org	Amazon	77909
online-sms.org	22XXX	67474
online-sms.org	Discord	67133
online-sms.org	Facebook	64278
online-sms.org	Google	37784
receive-sms-free.cc	Google	257098

receive-sms-free.cc	TikTok	242627
receive-sms-free.cc	Ello	137605
receive-sms-free.cc	Amazon	83707
receive-sms-free.cc	Netease	64785
receive-smss.com	Amazon	12968
receive-smss.com	FACEBOOK	10308
receive-smss.com	Skype	9309
receive-smss.com	FolkeSMS	7854
receive-smss.com	16148776293	7734
receivesms.co	18882375129	12374
receivesms.co	39041	8805
receivesms.co	Info	8110
receivesms.co	FACEBOOK	7315
receivesms.co	DISCORD	7267
receivesms.live	+XXXXX	5567
receivesms.live	+122678XXXXX	4224
receivesms.live	+120190XXXXX	3772
receivesms.live	+141299XXXXX	2827
receivesms.live	+183398XXXXX	2289
receivesms365.com	+XXXXX	420
receivesms365.com	+140297XXXXX	118
receivesms365.com	+183390XXXXX	84
receivesms365.com	+7XXXXX	83
receivesms365.com	+183358XXXXX	57
receivesmsfast.com	+120190XXXXX	5021
receivesmsfast.com	+XXXXX	4921
receivesmsfast.com	+122678XXXXX	3025
receivesmsfast.com	+183390XXXXX	1795
receivesmsfast.com	+183380XXXXX	1792
sms24.me	Google	62721
sms24.me	Amazon	57282
sms24.me	Instagram	53980
sms24.me	Discord	44174
sms24.me	Uber	43144
tesms.net	+XXXXX	1627
tesms.net	+011XXXXX	850
tesms.net	+163188XXXXX	365
tesms.net	+2XXXXX	198
tesms.net	+7XXXXX	161

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die wörtlich oder inhaltlich den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.
Ich erkläre mich mit der Archivierung der vorliegenden Masterarbeit einverstanden. Die vorliegende Arbeit wurde bisher in gleicher oder ähnlicher Form noch nicht als Magister-/Master-/Diplomarbeit/Dissertation eingereicht.

Datum	Unterschrift