An Overview of 7726 User Reports: Uncovering SMS Scams and Scammer Strategies

Sharad Agarwal
University College London (UCL)
sharad.agarwal@ucl.ac.uk

Guillermo Suarez-Tangil
IMDEA Networks
guillermo.suarez-tangil@networks.imdea.org

Marie Vasek
University College London (UCL)
m.vasek@ucl.ac.uk

***Abstract*—Mobile network operators implement firewalls to stop illicit messages, but scammers find ways to evade detection. Previous work has looked into SMS texts that are blocked by these firewalls. However, there is little insight into SMS texts that bypass them and reach users. To this end, we collaborate with a major mobile network operator to receive 1.35m user reports submitted over four months. We find 89.16% of user reports comprise text messages, followed by reports of suspicious calls and URLs. Using our methodological framework, we identify 35.12% of the unique text messages reported by users as spam, while 40.27% are scam text messages. This is the first paper that investigates SMS reports submitted by users and differentiates between spam and scams. Our paper classifies the identified scam text messages into 12 scam types, of which the most popular is 'wrong number' scams. We explore the various infrastructure services that scammers abuse to conduct SMS scams, including mobile network operators and hosting infrastructure, and analyze the text of the scam messages to understand how scammers lure victims into providing them with their personal or financial details.**

I. Introduction

There has been a recent surge in SMS scams worldwide [1], [2], with over 300k fraudulent texts sent everyday [3]. Unlike phishing, where data is easily accessible through aggregators such as OpenPhish [4], Phishtank [5] and APWG eCX [6], studying SMS scams is a lot harder due to unavailability of updated public data. Blocking scam texts is not universally implemented by mobile network operators and the limited metadata available in SMS — sender ID and timestamp, makes detection challenging. In 2024, users in the US lost \$470m to text scams [7], including \$129k attributed to just toll-related smishing campaigns [8]. In the same period, the UK reported an 8% increase in SMS scams [9], contributing to more than £162m in losses from telecommunications-enabled authorized push payment (APP) fraud [10]. While Australia recorded over \$14m in losses due to smishing in 2024 [11], fraudsters stole over \$4.2m in just three months from users in New Zealand [12]. Notably, most scams are delivered via text messages and phone calls [9].

Network and Distributed System Security (NDSS) Symposium 2026
23-27 February 2026, San Diego, CA, USA
ISBN 979-8-9894372-8-3
https://dx.doi.org/10.14722/ndss.2026.240987
www.ndss-symposium.org

Despite the substantial financial losses caused by SMS scams, most government (or commercial) reports fail to analyze and publish detailed reports on SMS scams. For example, the FBI IC3 annual report groups smishing, vishing, and phishing into one, without differentiation [13]. The lack of insights into the SMS scam ecosystem can be attributed to the difficulties organizations face in data sharing.

Government telecom regulators in the US, Canada, UK, New Zealand, and Australia work with mobile network operators to combat the increasing amount of telecom-related scams. To this end, mobile network operators have implemented SMS firewall filters and detection systems [14], [15], [16], [17] to stop SMS scams, including smishing (SMS phishing). Similar to phishing, this has become a cat-and-mouse game where scammers create new URLs [18], and mobile network operators, along with threat intelligence organizations, detect and block known URLs and text messages containing them. Despite the mobile network operator's filters blocking malicious messages, scammers evade detection by changing sender IDs, message text, and URLs. Prior work has investigated the texts blocked by these filters [19] and we lack insights into scam texts that evade detection and reach users.

As mobile network operators continue to block scams, threat actors find new tactics to evade detection and deceive users. One such technique includes sending a text message and asking users to call or text back. Others found that prompting users to reply and click on a URL increases the odds of a response to a smishing text [20]. While 7 in 10 people receive a suspicious text [21], 1 in 10 fall victim [22]. Controlled studies show at least 17% of participants fall for a smishing attack [20], [23] and users focus on the SMS content rather than the sender ID to identify smishing [24]. To devise effective countermeasures, it is essential to understand the various successful types of scams and how scammers deceive their victims.

Mobile network operators in the US, Canada, UK, and New Zealand run a special reporting service called 7726 ('SPAM' on a mobile keypad), which allows users to report suspicious calls and SMS messages [25], [26], [27], [28]. Reported messages have evaded the mobile network operator's filters, and this service helps mobile network operators update their detection rules with new threats. Mobile network operators collaborate with Google and Apple to integrate one-click reporting into the 7726 data feed. As users confuse legitimate

and illegitimate text messages [29], the reports submitted to 7726 are undifferentiated and contain a mix of spam, scams, and legitimate texts.

Research Gap. Mobile network operators have implemented an SMS detection system to reduce the amount of smishing received by end users. While previous research has examined these blocked SMS texts to study an individual scam [30] or identify scam types [19], it is essential to understand the text messages that bypass the filters and get delivered to the users. The data collected by previous studies through online forums [31], public online SMS gateways [32] or crowdsourcing [33] does not differentiate between scams and spam. This distinction is vital to build better mechanisms to thwart monetarily damaging cybercrime, as scams pose a significantly greater threat — they manipulate trust to inflict harm, making their detection and prevention a priority.

Contribution. To fill this research gap, we collaborate with a major mobile network operator that provides four months of user reports weekly. Our collaborator integrated Google's one-click SMS reports during the data collection period. We investigate 1.35 million user reports, identify scam and spam text messages, and provide insights into the identified scams.

With this data, we ask the following research questions:

RQ1 What kind of reports do users submit to 7726?
RQ2 What scams exist in the reported SMS messages?
RQ3 Which mobile network operators are abused, and how long scammers use these mobile numbers?
RQ4 What infrastructure is abused by scammers to run smishing campaigns?
RQ5 How do scammers try to lure victims?

In answering these RQs, we contribute the following:

- We investigate 1,349,039 user reports over four months in the UK and present the distribution in §IV-A.
- We categorize SMS user reports into spam and scams, further breaking down scams into deeper subsets (§IV-B).
- We detail the infrastructure that scammers abuse to conduct SMS scams, including mobile network operators (§IV-C) and registrars (§IV-D).
- We investigate how scammers craft text messages, including lures that scammers use to deceive victims (§IV-E).

II. Background

With online messaging communication channels, there has been an overall decline in the total number of text messages sent/received. However, there has been an uptick in SMS texts for essential services such as order updates, bank transactions, and medical services. As of 2024, 97% of adults in the UK own or have access to mobile telephone [34], with $89.6m$ active mobile subscriptions at the end of Q2 2024 [35].

As the legitimate use of SMS increases, businesses have started using this communication channel by sending unsolicited marketing SMS texts, aka spam. At the same time, scammers abuse SMS to deceive victims into clicking on a URL or interacting with them by impersonating brands/organizations/individuals to steal users' personal or financial details, aka scams. While *spam* texts hinder the availability of SMS texts, *scam* texts intentionally cause victims financial and/or psychological harm. Hence, it is essential to differentiate between spam and scam texts so stakeholders, such as mobile network operators and governmental regulators, can take appropriate actions. Similarly, research communities wielding this distinction will be able to develop more efficient detection models and propose more effective countermeasures to stop both spam and scams.

Recent studies on suspicious SMS messages fail to differentiate between these two types of text messages [32], [31], [33], [36]. The messages collected in previous works consist of old URLs that cannot be resolved now, and the available datasets do not include the redirected URL. Using previously collected data to differentiate spam from scams makes it challenging; §VI further discusses related academic work.

In response to the uptake of fraud in text messages, mobile network operators have implemented advanced fraud prevention systems powered by real-time AI and machine learning technologies. These systems leverage message fingerprinting algorithms to detect and mitigate messaging fraud scenarios automatically. Throughout this paper, we use the term Extended Detection and Response (XDR) system to refer to these fraud prevention systems due to its widespread use in industry. The ongoing arms race between fraudsters and mobile network operators significantly complicates the detection of scam messages at scale, posing a challenge even to advanced automated systems like XDR.

In addition to their efforts to evade detection by XDR systems, scammers also carefully craft messages designed to deceive human recipients. They employ various lures to manipulate victims into falling for different types of scams, distinct from businesses' tactics to attract users to spam. Understanding these lures requires access to the text being sent. This allows us to inform educational resources to inform potential victims as well as interveners like law enforcement.

There exist six known SMS scam types – (1) Wrong Number, (2) Hi Mum/Dad, (3) Delivery, (4) Banking, (5) Telecom, and (6) Government [19]. Based on these scam types, Mobile network operators block text messages which their XDRs can identify. However, many messages (which contain unknown scams) successfully evade the XDRs' filters. Our research focuses on the SMS text messages delivered to final users, evading the mobile network operator's detection capabilities. We provide a methodological framework that systematically groups user reports, differentiates between spam and scam, and understands the lures scammers use to deceive victims into different scams.

III. Methodology

Mobile network operators in various countries run a special user reporting service — 7726, where users can report suspicious SMS texts either by forwarding the text message or through the one-click reporting system enabled by Apple and

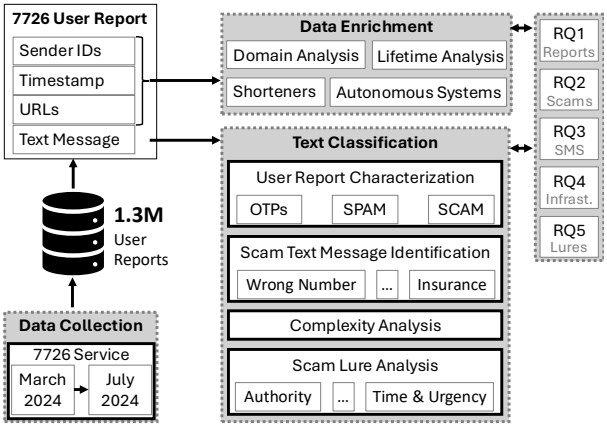


Fig. 1: Overview of our processing pipeline to characterize 7726 user reports and identify SMS scams.

Google in their messaging platforms.[1] If forwarding via 7726, users are asked to share the sender ID while reporting the text message. The one-click solution automatically shares the sender ID with the text message. In this section, we describe how we obtain, enrich, and examine this data to answer our research questions (**RQ1**-**RQ5**). Figure 1 shows an overview of our processing pipeline in a nutshell.

A. Data Collection

We collect weekly 7726 reports between March 7, 2024, and July 11, 2024, from a major mobile network operator. We received 1,349,039 user reports consisting of SMS, URLs, and calls. Since our paper focuses on understanding the SMS ecosystem, we exclusively retain SMSs. In particular, we filter out 129,467 reports flagging suspicious calls and 16,713 reports flagging 9,744 suspicious URLs (reports which contain URLs without the associated SMS text message or any relevant metadata). As a result, we investigate 1,202,859 reports flagging 530,555 unique text messages. These reports originate from users located in the United Kingdom. The remainder of this section describes the methods we use to examine these messages.

B. Data Enrichment

The 7726 user reports contain a text message, sender ID, timestamp, and URL (where available). We consider text separately in §III-C and describe next the steps we take to obtain contextual information about the remaining elements of user reports.

Sender IDs. SMS texts are generally sent using a mobile number, short codes via bulk messaging services, or email addresses (incl. Apple's iMessage). Unlike email, the SMS protocol only includes sender ID and a timestamp as metadata.

We investigate the mobile numbers users report via Home Location Register (HLR) lookup [37]. HLR lookup validates the queried phone number and provides the country to which the phone number belongs. It further provides details about a mobile number – current mobile network operator, original mobile network operator, and its current status (dead or alive). Current mobile network operator is the name of a network where the phone number is currently assigned, and the original mobile network operator is the network assigned to this telephone number range. These can change over time due to company name changes, mergers, acquisitions, and re-allocation of a number range [38].

To identify the country, the current and original mobile network operator, and its current status, we partner with Stour Marine, who offered HLR lookups [39]. We make these lookups against all newly used phone numbers as soon as we receive the user reports. We additionally query all phone numbers that were live the previous week. We combine these results to calculate each mobile number's lifetime and find the various mobile network operators scammers abuse.

Survival Analysis. We calculate the number of weeks 78,906 unique mobile numbers were active using the availability status we retrieve weekly from the HLR lookups.

We conduct survival analysis, which helps reveal patterns of activeness of mobile numbers scammer abuse. In particular, this technique considers intermittently unavailable data points to be "right-censored"; we only monitor these mobile numbers until July 18, 2024, and censor those still active on that date. We use a Kaplan-Meier estimator [40] to estimate the survival function $S(t)$ from the lifetime we observe in the data. Intuitively, this measure illustrates the fraction of mobile numbers that become inactive after a given date. This helps us estimate the lifetime of these numbers by using the probability of a mobile number abused by a scammer being active after x weeks.

URLs. The majority of user-reported SMS texts contain URLs. We query all collected URLs on VirusTotal [41], an antivirus aggregator service with over 70 antivirus scanners. We extract threat intelligence from VirusTotal to determine if URLs are malicious. We consider a URL malicious if it has been flagged as such by at least one scanner.

We next describe how we further enrich our dataset of URLs with DNS telemetry, dynamic URL redirections, domain registrars, and Autonomous Systems.

Passive DNS: Spamhaus [42] provides us with passive DNS (pDNS) API that returns the first time Spamhaus saw a domain and the IP addresses it resolved to in the last year [43].

URL Shorteners: Scammers use URL shorteners to hide the redirected malicious URL and send that in the message's text. To this end, we create a list of 27 commonly used URL shorteners and query them against the collected URLs from the text message reports. Some of these are bespoke for a single service, like wa.me URLs that redirect to WhatsApp, while others are more generic, like bit.ly.

Lifetime Analysis: The lifetime of domains helps us understand the impact of scams, as longer-running websites defraud more victims. We consider the beginning of the website to be when we first see it and the last date as per the passive DNS

[1]During our data collection period, Apple one-click was not enabled by our data partner.

results. Note that we exclude shortened URLs and other third-party domains.

Domain Registrars: To investigate the domains scammers abuse, we first extract *Top-level Domains (TLDs)*. We remove shortened URLs and other third-party domains (such as those using the '.me' and '.sbs' TLDs or popular domains listed among Alexa's top $10k$ sites, since these domains being abused is not reflective of the registrar.

We identify registration data for each domain using a WHOIS API service. We query WhoisXMLAPI [44] at the time users report SMSs, which allows us to obtain current registration data for each domain name.

Autonomous Systems (ASes): We query the IP addresses returned by Spamhaus against IPinfo's database [45] to identify the corresponding Autonomous System (AS) to the IP and the IP's geographical location [46]. We note that IPinfo is unable to identify the AS for 715 IP addresses.

Timestamps. Our collected timestamps are in UTC-8. We convert the timestamp into British Summer Time (BST) which reflects local time during our collection period.

C. Text Classification

We perform text analysis to characterize user reports, classify scam messages, and identify the lures that scammers use.

User Report Characterization. We categorize the different types of user reports to whitelist messages that are not scams. In particular, we look into the text messages and search for specific keywords in the text as described in Table I to identify straightforward cases: (1) text messages containing one-time passwords (OTPs) and (2) spam text messages. OTP messages are occasionally reported to 7726 when they are unintentionally delivered to the wrong recipient,[2] but they are easy to flag. Likewise, spam messages are often reported to 7726. These messages are easier to characterize than scam messages because they follow more predictable patterns and adhere to explicit opt-out mechanisms. After filtering out the more straightforward cases, we next outline the mechanism employed to further identify and analyze scam messages.

Scam Text Message Identification. We investigate the collected suspicious text messages to identify spam and scam text messages. To this end, we create our prompt (cf. Appendix A) and query the unique suspicious text messages using OpenAI's API GPT-4o model. We select GPT-4o as this was the best available model during our research period.

First, we run all unique suspicious text messages with our prompt via OpenAI's model to categorize the text messages into spam. Next, we query the URLs from the suspicious text messages for antivirus detection on VirusTotal and query all the identified malicious domains against the domains in the text of all unique suspicious text messages with URLs. Lastly, we classify suspicious text messages with and without URLs using OpenAI's GPT-4o model into various scam types. In particular, we identify the six known SMS scam types in the UK introduced in Section II, i.e.: (1) Wrong Number, (2) Hi Mum/Dad, (3) Delivery, (4) Banking, (5) Telecom, and (6) Government [19]. The remaining suspicious text messages with URLs where VirusTotal does not flag the URL are 'Unknown.'

To our surprise, the initial results for scam type classification had almost 50% of the scam text messages marked as 'Others'. To this end, we update our prompt (cf. Appendix A) and ask OpenAI to suggest a category if the initial classification is 'Others.'

Complexity Analysis. Gunning Fog Index helps explain the educational level required to understand the scam text message the Gunning fog index [47]. The scam text messages collected in our dataset are in English, as they target users in the UK. We tokenize the text of the SMS messages reported by users and remove the stop words. Stop words are common words that do not provide meaningful information about the topic. To calculate the complex words, we use the conventional criteria of the Gunning Fog Index and filter out the words with less than three syllables. We additionally remove proper nouns, such as brand names being impersonated. The formula to calculate the Gunning Fog Index is as follows:

$$\text{GFI} = 0.4^{*}\left(\left(\frac{words}{sentence}\right) + 100\left(\frac{complexwords}{words}\right)\right)$$

where words and sentences are counts as normally defined and complex words are words with three or more syllables, excluding proper nouns, familiar jargon, or compound words.

Scam Lure Analysis. We want to understand the various lures that scammers use to deceive victims into taking actions mentioned in a scam message. Towards this end, we adopt the ontology of lures from Stajano and Wilson [48]. The n-gram analysis for multiple scam types with thousands of messages is not feasible to do manually. Thus, we use OpenAI's GPT-4o to categorize our messages into the different lures using the prompt listed in Appendix B.

D. Evaluation

To evaluate the performance of our methodology, we extract 384 random texts from our dataset, ensuring (95% confidence) that the random sample is representative of the complete dataset. We manually label them as an OTP, spam, or scam. Next, we extract another random sample of 384 scam texts and manually classify them into seven scam types (including other) along with the scammer lures used in each scam text. We use human-labeled texts as ground truth and calculate the inter-rater reliability (IRR) between ground truth and OpenAI's annotation. We use Cohen's κ [49], a standard metric for IRR. There is near-perfect agreement for OTP and scam types and substantial agreement for spam and lures (see Table II for Cohen's κ coefficient and F1-score).

In addition to prompting AI to label scams based on existing scam types, we also asked AI to provide a new

[2]For example, a sender may mistype the intended recipient's phone number, resulting in the OTP being sent to an unrelated individual. This can lead the recipient to mistakenly perceive the message as spam or a phishing attempt, prompting them to report it to 7726.

Classification	Keywords Used
Spam Text Messages	STOP, optout, opt out, opt-out, won the draw, claim your prize and offer ends.
Text Messages with One Time Password (OTP)	otp and verification code

TABLE I: Keywords used to differentiate suspicious text messages from spam and messages with OTPs.

Category	κ coeff.	Agreement level	F1-score
One Time Password (OTP)	0.81	Near-perfect/ Strong	91%
Spam	0.76	Substantial/Moderate	88%
Known Scam Types	0.84	Near-perfect/ Strong	89%
New Scam Types	0.71	Substantial/Moderate	80%
Lures	0.73	Substantial/Moderate	78%

TABLE II: Cohen's κ Coefficient for evaluating the inter-rater reliability between ground truth and our methodological framework using OpenAI ($n = 384$).

scam category rather than "other." However, this new scam category suggestion has only a moderate agreement level (see Table II). We also find that its category recommendations are inconsistent for the same scam type. For example, in the case of 'job scam' texts, OpenAI provides a variety of new scam categories — 'job/recruitment', 'employment', 'joboffer', 'jobopportunity', and in 12.3% cases, it even returns no new category or 'Unclear.' The inconsistency here justifies our two-stage approach – one with a fine-grained classification under a closed-world of scam types and another with an open-world assumption refining those in the "others" category.

IV. Results

We observe 1,202,859 SMS reports over four months. This section presents the distribution of these user reports over time and the scam text messages we identify. We study existing scam types and the infrastructure scammers abuse to conduct them, including the mobile network operators that they abuse. Lastly, we highlight the lures scammers use and examine the readability of the scam text messages.

A. User Reports Over Time

Table III presents the weekly distribution of user reports along with associated sender IDs.[3] The average number of reports submitted by users weekly is 66.8k, with a median of 64.9k. From these reports, we see a weekly average and median of 11.8k unique sender IDs. Fig. 2 shows the time of the day per week when users report suspicious messages in BST. Most users report between 10:00 - 21:00 BST (with medians: Mon - 16:10:28, Tues - 16:15:00, Wed - 16:22:25, Thurs - 16:05:34, Fri - 16:26:41, Sat - 15:16:26, Sun - 16:11:25). This pattern aligns with the opportunistic nature of scam campaigns, which are often driven by specific events or timing rather than a steady stream of messages. In contrast, spam campaigns may follow consistent schedules or promotional cycles.

Takeaway. We see that users mostly submit reports between 10:00 and 21:00 BST daily. Prior work found that scammers interact with Hi mum/dad scam victims during 10:00-15:00 UK time [30], indicating that users report SMS messages or calls without significant delays.

[3]The identity of the user submitting a report remains anonymous and should not be confused with sender IDs.

Dates (2024)	Reports (k) Total	Reports (k) Unique	Sender IDs (k) Total	Sender IDs (k) Unique
Mar 8 - 14	73.3	36.2	45.9	13.2
Mar 14 - 21	71.9	37.7	43.9	11.9
Mar 21 - 28	70.8	39.2	43.1	12.9
Mar 28 - Apr 4	59	33.8	36.5	11.7
Apr 4 - 11	62.1	34.8	38	11.9
Apr 11 - 18	61.4	34.3	37.9	12
Apr 18 - 25	62.8	34.5	39.1	12.9
Apr 25 - May 2	72.9	34.1	47	11.9
May 2 - 9	66.9	32.5	43.9	11
May 9 - 16	57.8	33.1	36.1	11.6
May 16 - 23	63.6	34.9	40.9	11.6
May 23 - 30	69.9	41.4	45.8	12
May 30 - Jun 6	62.9	34	40.3	11.4
Jun 6 - 13	65.3	33.6	42.6	11.5
Jun 13 - 20	63	28.9	40.8	10.6
Jun 20 - 27	64.5	28.6	41.6	10.3
Jun 27 - Jul 4	74.3	31.7	48.9	11.5
Jul 4 - 11	80	33.7	51.4	12.4

TABLE III: Distribution of all text message reports ($n = 1,202,859$) received weekly.

Fig. 2: Time of the day per week when users report suspicious text messages ($n = 1,202,859$). The pair-wise two-sample KS test is significant with $p < 0.05$.

B. Characterization of User Reports

To answer **RQ1**, we study the type of suspicious text messages in §IV-B1. We characterize the type of scams in §IV-B2 to answer **RQ2**.

1) Suspicious Text Message Classification: The first step of our text classification method focuses on identifying and whitelisting clear cases of fraudulent and non-fraudulent messages. Table IV presents the results of our initial report characterization, where we find 7,143 (1.35%) text messages delivering OTPs, 186,325 (35.12%) spam messages, and 213,659 (40.27%) scam messages. This result indicates that 7726 user-reported messages contain a significant amount of non-fraudulent messages, prompting subsequent steps in our

methodology to filter out irrelevant content and further classify types of fraud. The results of the subsequent steps in our methodology are presented in §IV-B2. We now delve deeper into these results, providing detailed discussions and examples of the different types of messages observed and reasons that may drive users to report non-fraudulent messages.

Type	Unique Text Messages		Total	
	w/ URLs	w/o URLs	(#)	(%)
OTPs	-	7,143	7,143	1.35
Spam	164,826	21,499	186,325	35.12
Scam	46,635	167,024	213,659	40.27
Unknown	123,428	-	123,428	23.26
Total	334,889	195,666	530,555	100

TABLE IV: Categorization of all unique text messages ($n = 530,555$) into spam and scam, both with and without URLs.

Non-fraudulent messages. We identify 36.47% of the messages as non-fraudulent (7,143 OTPs, 186,325 Spam). While most of these are spam (35.12%), we see a small fraction of unique reports containing OTPs (1.35%). For example,

> <brand name>: 1041 is your verification code. It expires in 15 minutes. Don't share this with anyone.

Examples of spam messages include:

> Join now to receive a <brand> 100 FS plus up to 2000 GBP and enjoy weekly bonuses [URL] OptOut: [URL].

These messages are generally unsolicited marketing messages sent by companies or unknown senders. Spam is often annoying and reduces the availability of messages on a user's mobile phone, which can prompt users to report them to 7726. Some of these reports might be erroneous. As reported by Ofcom, the built-in reporting function on users' mobile phones was the most used channel to report suspicious messages [50]. The user interface has 'delete' and 'delete and report' buttons next to each other, which could confuse users and send their reports to 7726 instead of simply deleting the text message.

Fraudulent messages. We identify 40.27% (213,659) of the unique messages as fraudulent. Out of these, we see 46,635 with URLs and 167,024 without URLs. Overall, users report 334,889 messages with URLs, which means that most of the messages reported with URLs are either spam or the URL is not flagged by an AV vendor on VirusTotal. VirusTotal only flags 24,546 (44.77%) of our unique URLs as suspicious (9,285) or malicious (20,482). This indicates that mobile network operators are more effective in blocking scam texts with URLs. Alternatively, it could also mean that the scammers are shifting towards scams without initial URLs.

While URLs are more common in spam than in scams, they can play a crucial role in identifying fraudulent messages. The following illustrates a smishing attack, featuring a seemingly benign SMS, where the only distinguishing factor between a legitimate and malicious message lies in the link's behavior:

> <Brand name>: Hi, unfortunately you have missed your delivery. Please visit [URL] to schedule a redelivery.

Unknowns. We label the remaining 123,428 (23.3%) unique suspicious text messages with URLs as 'Unknown'. Considering that these messages have been reported by users and flagged as suspicious, we see them as potentially malicious rather than benign due to the various evasion techniques deployed by phishing websites [51] and the low recall of AV vendors in detecting phishing websites [52], [36].

2) Scam Type Classification: We classify scam text messages into six known SMS scam types [19] and identify six new scam types using our methodological process described in §III-C. Table V outlines messages by category. Out of the 119,398 scam texts overall, we discover that the most popular scam type is the Wrong Number scam (16.36%), followed by Banking (9.14%) and Delivery/Parcel (6.81%) scams. We plot the distribution of the these scams over time in Fig. 3a.

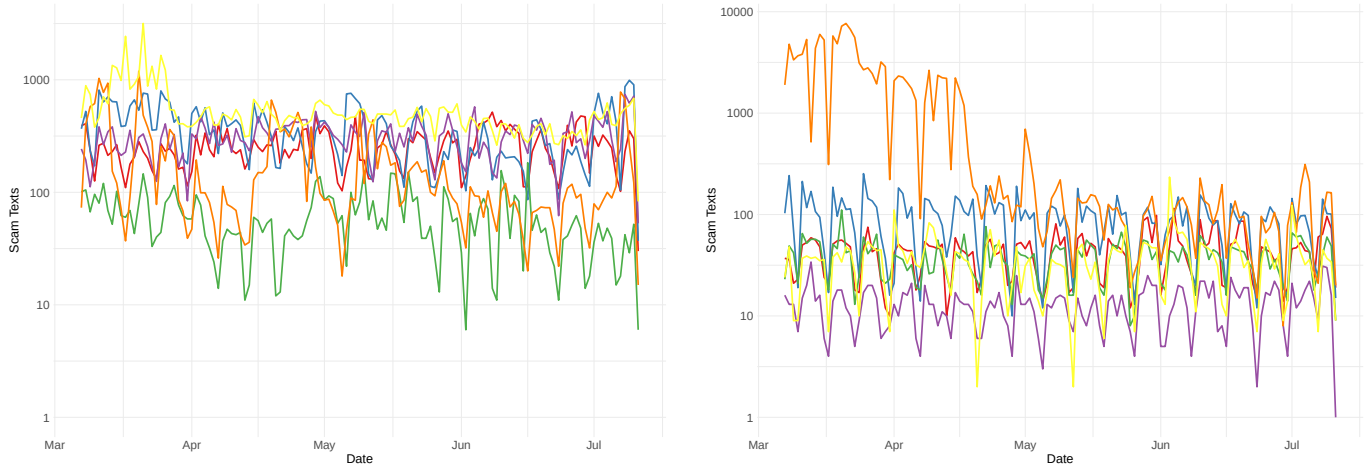
	Category	Unique Scam Texts		Total	
		w/o URL	w/ URL	(#)	(%)
Known [19]	Wrong Number	34,863	86	34,949	16.36
	Banking	14,744	4,784	19,528	9.14
	Delivery/Parcel	1,834	12,725	14,559	6.81
	Hi Mum/Dad	6,604	190	6,794	3.18
	Telecom	3,743	2,083	5,826	2.73
	Government	1,880	2,229	4,109	1.92
New	Job	10,407	528	10,935	5.12
	Debt	7,188	3,185	10,373	4.85
	Appointment	4,087	458	4,545	2.13
	Finance	2,800	1,018	3,818	1.79
	Utility	1,253	1,355	2,608	1.22
	Insurance	1,201	153	1,354	0.6
	Sub-Total	90,604	28,794	119,398	55.88
	Others	76,420	17,841	94,261	44.12
	Total	167,024	46,635	213,659	100

TABLE V: Identified categories of scam text messages ($n = 213,659$), both with and without URLs.

From the new scam categories, we find that job-related scams (5.12%) are the most reported category, followed by debt (4.85%) and appointment-related scams (2.13%). Job-related scams lure victims by providing fake employment opportunities [53]. Debt scams are text messages that ask users to pay an outstanding debt balance. Appointment scams are similar to appointment reminders, asking users to call or text back on the phone number but then luring victims into transferring funds or stealing financial or personal details [54], [55]. Example of job-related scams is (cf. Appendix C for other types):

> Hi, I'm <First Name> from <Fake HR>. I'd like to introduce you to a employment opportunity here, can I share some detail with you?

Some scam texts contain a malicious URL and are known as smishing or SMS phishing. These messages contain a malicious URL enticing a user to click; this redirects to a phishing page impersonating a brand which deceives victims into providing personal or financial details. Unsurprisingly, Delivery/Parcel (12,725) and Banking (4,784) impersonation scams are the top two categories for URL-based scams (Table V). From the new categories, for URL-based scams, we find Debt (3,185) and Utility (1,355) scams to be the top two. Utility scams impersonate a utility company, such as gas or water, and aim to steal users' details and banking credentials [56]. Here is an example Banking scam text with a URL where scammers try to lure victims into providing credentials:



(a) Distribution of known scams' texts over time ($n = 222,143$). Yellow: Wrong number, Red: Banking, Blue: Delivery/Parcel, Green: Government, Purple: Hey Mum/Dad, and Orange: Telecom scams.

(b) Distribution of new scams' texts over time ($n = 163,808$). Orange: Job, Blue: Debt, Red: Appointment, Green: Finance, Yellow: Utility, and Purple: Insurance scams.

Fig. 3: Distribution of known and new scam-type messages over time. Y-axis is on log scale. Colors represent scam types.

> <Bank Name>: A scheduled payment to <Brand Name> has been made, please verify your credentials via: [URL]

Conversational scams, on the other hand, are text messages asking users to interact with the scammer directly via text. For example, in Hi Mum/Dad scams or Wrong Number scams, scammers send an initial text message and deceive the potential victim into replying/initiating the conversation on the same or a new mobile phone number. For example,

> Now then mate you well - just a txt to say hello buddy . X Sending love - and sorry I've not been in contact x

We do not expect conversational scams to contain URLs. However, we see a handful of these (86 Wrong Number, 190 Hi Mum/Dad) with URLs, overwhelmingly from scammers attempting to move the conversation to online messaging platforms, e.g., WhatsApp. Contrary to URL-based scams, scams without URLs feature Wrong Number scams (34,863), followed by Banking scams. These are more difficult to identify than a malicious URL in a smishing text. While Hi Mum/Dad scams are a well-known authorized push payment (APP) fraud which cause significant financial loss to victims [30], these are actively blocked by mobile network operators in the UK [19] and thus lower in ranking by volume (6,604). Investigating banking scam messages without a URL, we find that these messages mention an OTP or a transaction and ask the victim to call the provided phone number. For example,

> <Bank name> Bank: Transaction <brand name> £442.26 on 19/05/2024 15:30pm .Your code is 446228. If this was NOT you call us on <phone number>

This indicates a rise in text and call-back scams. Here, scammers lure victims through text messages by asking them to text or call the phone number provided in the SMS text. These messages often include a OTP or mention a fake transaction, instructing recipients to call or text a provided phone number if they did not request the OTP or initiate a transaction. We do not directly engage with phone numbers provided, either by texting or calling; telephony honeypots present a potential avenue for deeper investigation [57] discussed in §V.

We note new scam categories have higher volumes for non-URL-based scams, particularly Job scams (10,407) and Debt-based scams (7,188) [58], [59], [53]. Job scams exploit users' vulnerability by contacting them about unrealistic job opportunities, offering high salaries to lure them into a scam [60], [61]. We plot the distribution of the new scam types over time in Fig. 3b.

Takeaway from RQ1 and RQ2. We identify fraudulent, non-fraudulent, and potentially malicious texts from user reports. Over 40% of reported texts are scams, with Wrong Number scams as the most reported type. This indicates that scammers are able to evade mobile network operator's XDR system using conversational scams instead of URL-based scams. Alternatively, mobile network operators are better at blocking scam texts with URLs than ones without URLs. Our findings could help mobile network operators update their XDR system to block new scams and save users from falling prey.

C. Originating Sender ID Distribution

Unlike phishing over email, SMS messages or calls only have sender ID and timestamp as metadata. As discussed in §III-A, we collect the sender IDs reported by users, either forwarded by users to 7726 or automatically via Google's one-click reporting system, and analyze them to answer **RQ3**.

1) Sender IDs: This subsection investigates the sender IDs reported by users and abused by scammers to conduct scams.

Distribution. There are four types of sender IDs used to send SMS — phone numbers, alphanumeric shortcodes, number-only shortcodes, and email addresses. Table VI shows the sender ID distribution of text message reports. Unsurprisingly, the majority of these are phone numbers. Curiously, while 83% of overall reports are from phone numbers, about 92%

of scams originate from phone numbers. This material discrepancy between reports and scam texts highlights the utility of dividing out scams and considering them separately.

Type	Unique Sender IDs	
	User Reports	Scam Texts
Phone numbers	103,301	79,894
Alphanumeric shortcodes	19,318	6,047
Email addresses	1,191	943
Number shortcodes	429	207

TABLE VI: Distribution of all unique sender IDs (124,239 total) for all text message user reports and identified scam texts.

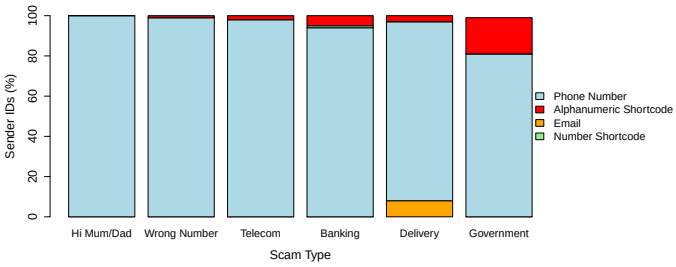


Fig. 4: Sender IDs which scammers abuse to send six types of scam texts. Y-axis is normalized by the total number of scams in each category.

We analyze the distribution of sender IDs in relation to the types of scams to identify patterns that enhance our understanding of existing threats. We find that phone numbers dominate the breakdown by scam type. We notice that 18,179 phone numbers are abused to send Wrong Number scams, followed by 10,429 phone numbers for Delivery scams and 8,741 for Hi Mum/Dad scams. Due to the absence of Know-your-customer (KYC) checks in the UK, scammers can procure multiple SIM cards.

We also see cases where other sender IDs are abused. Fig. 4 shows that scammers abuse 875 email addresses to send delivery/parcel scams. The use of email addresses indicates that these scam text messages are sent to users on iMessage. The email address in the sender ID is available on Android devices but is more commonly used to send iMessage on Apple devices. Future work integrating one-click reporting on iMessage will show us if this trend holds with additional data.

While Delivery scams contribute toward the abuse of 363 alphanumeric shortcodes, 341 alphanumeric shortcodes are abused for banking scams. Scammers use similar-looking alphanumeric shortcodes to impersonate various delivery and banking entities to lure victims into the fraud: 'evroi' instead of 'evri' and 'santandar' instead of 'santander.'

We investigate overlapping sender IDs scammers abuse to send different scam types (Fig 5). We find 490 phone numbers abused to send both Hi Mum/Dad and Wrong Number scams. This indicates that some scammers conduct both types of conversational scams. An additional 183 sender IDs were used to send both Delivery and Banking scams (cf. Table XIV in Appendix for all numbers). One explanation could be scammers using the same third-party service to broadcast their scams.

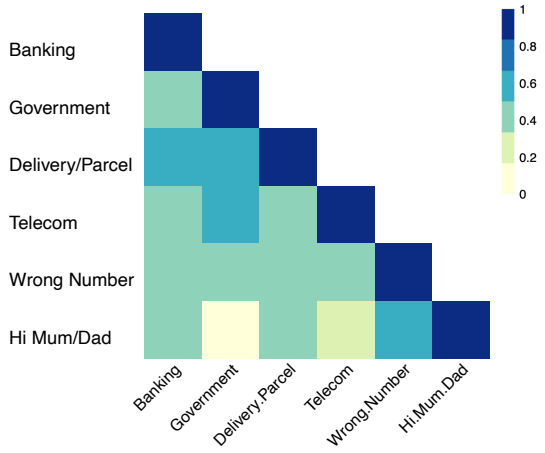


Fig. 5: Heatmap of common Sender IDs used in different scams. Values normalized by the total common sender IDs.

Countries. We analyze the origin of each phone number using our collected HLR lookup data. We find that 73,815 (92.4%) phone numbers abused to send scam texts originate in the UK. This makes sense as it is easy and cheap to procure UK pay-as-you-go SIM cards without KYC. Additionally, it is more believable to target victims using the same country's number as theirs.

Table VII shows that more than 99% phone number scammers abuse for the Wrong Number, Hi Mum/Dad, and Government scams originate in the UK While more than 91% of phone numbers originate from the UK for Delivery and Banking scams, we notice phone numbers originating from the USA and the Philippines abused for Delivery scams and numbers from Japan abused for Banking scams. This is likely due to preferences of a few attackers. Scam campaigns impersonating only two particular UK banks account for almost all of the Japanese number (8.49% of the total banking scams). Similarly, we identify a campaign impersonating one Delivery brand with nearly identical text explaining most of the US numbers (and 3.09% of delivery scams) and two campaigns originating from the Philippines, one of which belongs to the 'darcula' iMessage and RCS smishing attacks [62].

Mobile Number Lifetime. We study the lifetime of mobile numbers to understand how long scammers have used the same mobile number to scam their victims. To this end, we plot a survival curve to visualize the lifetime of the mobile numbers with overall survival probability in Fig. 6.

We find that the median lifetime of all phone numbers scammers abuse is 48 days (6.86 weeks). While 75.3% are active after 2 weeks, only 35.8% are active after 10 weeks. For Hi Mum/Dad scams, we find the median to be 4.14 weeks, i.e., 29 days, whereas previous research found the median lifetime to be 14 days [30]. We attribute this difference to the fact that previous work only studies scam messages detected by existing XDR filters, missing those that permeate through

Scam Type	Country	%	Country	%	Country	%	Country	%	Country	%
Wrong Number	UK	99.14	USA	0.32	Nigeria	0.09	Canada	0.06	Ireland	0.06
Hi Mum/Dad	UK	99.88	Canada	0.05	Japan	0.05	Nigeria	0.02		
Delivery/Parcel	UK	91.72	USA	3.09	Philippines	2.85	Thailand	0.73	Japan	0.33
Telecom	UK	95.12	Philippines	1.53	Tajikistan	0.98	India	0.68	Vietnam	0.3
Banking	UK	91.35	Japan	8.49	USA	0.05	Channel Islands	0.04	Indonesia	0.02
Government	UK	99.37	Japan	0.21	France	0.1	Jersey	0.1	Poland	0.1

TABLE VII: Distribution of top 5 origin countries of the sender ID phone numbers for six known scam types.

their detection systems. Instead, our work investigates phone numbers reported by users. While we discover that the median lifetime of phone numbers abused to conduct delivery and banking scams is the same as the Hi Mum/Dad scams, the median lifetime for telecom is 9 weeks and over 9 weeks for wrong number scams. This indicates scammers being able to evade mobile network operator's XDR systems.

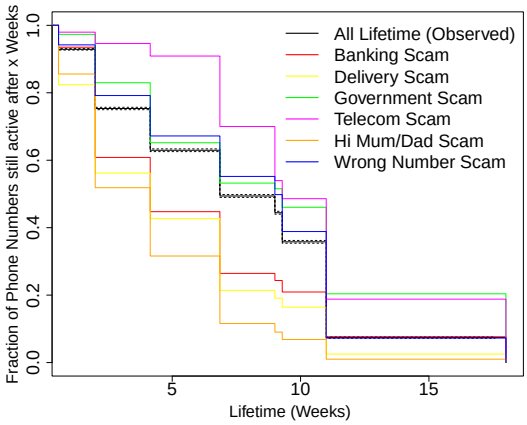


Fig. 6: The number of weeks a mobile number (originating sender ID) is alive after being used to send a scam text message ($n = 78,906$). Black dashed line is the 95% CI for all phone numbers (observed) lifetime.

2) Mobile Network Operators: Mobile network operators (MNOs) are one of the main stakeholders that scammers abuse to send scam texts and we use our HLR lookup data to uncover the mobile network operator of a mobile number. Table VIII presents the top 10 different mobile network operators.[4] Mobile network operators can be (1) Physical — issues a physical SIM card, (2) Virtual — does not issue a physical SIM card, and (3) Mobile Virtual Network Operator (MVNO) — issues a physical SIM card but rents space from other mobile network operators instead of running their own.

UK mobile network operators that provide physical SIM cards are more abused than Virtual mobile network operators. This could be because scammers broadcast thousands of scam texts using a SIM box/farm [63]. A SIM box is an SMS gateway device connected to a computer with multiple SIM slots. MNO 1 is the most preferred mobile network operator, followed by MNO 2, 3, and 4. While MNO 1, 2, and 3 support GSM technology that works with SIM boxes, MNO 4 does not support GSM. The desirability of mobile network operators by scammers could also depend on the ease of availability to procure the SIM cards, network availability where the scammers are based, or the XDRs that scammers evade to detect these scam texts.

[4]We refrain from naming mobile network operators in line with the confidential agreement with our partner.

MNO	Country	Type	Current MNO	Original MNO
MNO 1	UK	Physical	19,576	25,629
MNO 2	UK	Physical	13,572	20,938
MNO 3	UK	Physical	9,885	12,051
MNO 4	UK	Physical	8,718	8,614
MNO 5	UK	Virtual	1,318	1,318
MNO 6	UK	Physical (MVNO)	1,233	769
MNO 7	UK	Physical (MVNO)	999	2,136
MNO 8	UK	Virtual	619	654
MNO 9	UK	Virtual	296	289
MNO 10	USA	Physical	292	134

TABLE VIII: Top 10 mobile network operators abused to send scam text messages. *Current MNO* is the network where the phone number is currently assigned. *Original MNO* is the network assigned to this telephone number range.

Next, we explore the popularity of mobile network operators based on scam types. While MNO 1 is least abused when sending Telecom scams, MNO 2, 3, and 4 are least abused when it comes to Hi Mum/Dad and Government scams. This could mean that MNO 1 is better at detecting Telecom scams, whereas MNO 2, 3, and 4 are better at blocking Hi Mum/Dad and Government scams. While most Virtual numbers are primarily abused to send Wrong Number scams, MVNOs are abused for Telecom, Hi Mum/Dad, and Wrong Number scams. Virtual numbers provide scammers the advantage of sending scam texts from outside the UK. However, with MVNOs, even though they use the network of a physical MNO, the text messages go through their own XDR systems (if implemented). Looking into MNO 10, the only non-UK mobile network operator in Table VIII, we find similar text messages targeting one Delivery company. This indicates that a single threat actor abused 3.09% of all sender IDs that belong to the USA. For Hi Mum/Dad scams, MNO 1 is significantly more abused (86.4%) than others [30] showcasing that scammers abuse different types of mobile network operators for various scams and some more than others.

Takeaway from RQ3. We find that scammers prefer certain MNOs over others depending on scam types. This indicates that MNO's XDR systems use independent rules and do not share intelligence. We suggest MNOs work collaboratively and share their best practices and XDR rules to make the blocking a collective effort. Investigating user reports can

help MNOs identify and block new originating sender IDs so scammers cannot abuse them for long. As scammers also abuse shortcodes, we suggest implementing a central sender ID registry operated by the regulator that could stop the abuse of impersonation scams.

D. Domain Analysis

Scammers trick victims into clicking on malicious URLs sent in text messages and lure them into providing their personal or financial information. We examine the various infrastructure services scammers abuse to host phishing pages shared via scam texts, answering **RQ4**.

URL Shorteners. We identify a significant amount of URLs in the text message reports that are shortened URLs, inline with prior research [32]. Table IX presents the number of unique URLs belonging to the 10 most popular URL shorteners abused to conduct scams, with the two most common services being bit.ly and t.ly. Scammers abuse URL shortening services to evade detection from XDRs and antivirus vendors. Shortened URLs also can be created for free, fit in an SMS character limit, and redirect to the malicious URL only when a user clicks.

URL Shortener	Unique URLs	Scam Types					
		W	H	T	B	G	D
bit.ly	2,280	0	0	12	50	48	73
tinyurl.com	1,094	0	0	48	8	16	241
is.gd	814	0	0	5	1	2	393
wa.me	712	8	190	1	4	3	0
rb.gy	611	0	0	185	2	6	250
cutt.ly	462	0	0	95	4	13	61
qrco.de	447	0	0	82	1	3	354
rebrand.ly	434	0	0	314	2	1	21
t.ly	302	0	0	5	16	0	49
tiny.cc	18	0	0	0	0	2	0

TABLE IX: Distribution of Top 10 URL shorteners abused by scammers to send scam texts. (W: Wrong Number, H: Hi Mum, T: Telecom, B: Banking, G: Government, D: Delivery)

In addition to the URL shorteners in Table IX, we identify 173 unique '.sbs,' and 62 unique '.me' URLs abused in scam text messages. For example, we find six second-level domains that try to impersonate EVRI that contain 'evri' plus one more character as the second-level domain name for '.sbs' URLs. The registrar cannot take down malicious shortened URLs; deleting the entire domain would cause harm to other, non-malicious shortened URLs. Instead, a URL shortener requires investigating and take down on URL shortener service's side.

Registrars. We find that NameSilo (26.3%) is the most abused registrar for SMS scam URLs, followed by Hosting Concepts B.V. (13.2%). Previous research on newly registered phishing domains also found NameSilo as the most abused registrar [18]. On the other hand, a smishing research using US smish reports identified NameCheap [33] as the most abused registrar. This highlights differences in our collected data as well as the impact of our spam/scam distinction. While domains in delivery and telecom scams primarily abuse NameSilo, MarkMonitor is the most abused registrar for banking and government scams. This highlights that different scams abuse different registrars, likely reflecting preferences of different groups of threat actors.

Top-level Domains (TLDs). We present the ten most abused top-level domains (TLDs) by scammers to register domains abused in scam text messages in Table X. We differentiate TLDs by domains abused by scammers and domains with URL shorteners as URL shorteners are third-party services, so their TLDs should not misunderstood for abuse.

We find that '.com' remains the most abused TLD, followed by '.top' and '.co.uk.' This is in line with recent research [18], [64], [65], [66]. Investigating individually by scam types – '.com' remains the most abused for delivery, banking, telecom, government, and wrong number scams. After '.com,' we uncover that the delivery scams abuse .top followed by .xyz. On the other hand, we find that government scams abuse 'co.uk' and '.uk' TLDs making them more convincing for potential victims. For example, arrange-test-kit[.]co[.]uk is used to impersonate a health service text. While '.buzz' is the second most abused TLD by scammers for telecom scams, they abuse 'web.app' for banking scams. For example, scammers set up attempted-logon[.]web[.]app to lure victims into clicking on the malicious link.

URLs		URL Shorteners	
TLD	Unique URLs	TLD	Unique URLs
com	3,593	ly	3,480
top	897	com	1,097
co.uk	390	gy	611
xyz	370	gd	814
buzz	166	me	712
info	156	de	447
web.app	143	cc	18
sbs	140	co	15
cyou	101	ws	7

TABLE X: Top 10 Top-level Domains (TLDs) abused in all unique scam URLs.

Autonomous Systems (ASes). We identify 56,092 unique IP addresses that the identified scam domains abused. We find that 7,110 (12.7%) unique IP addresses belong to Cloudflare, a proxy service which hides the server IP address. Previous research also found most domains in their dataset abusing Cloudflare [32], [18]. Table XI presents the five most abused Autonomous Systems (ASes) to host the identified scam URLs, excluding Cloudflare. It also shows the geographical location of the IP addresses abused to host scam domains.

While scam messages target individuals in the UK and significantly dominate UK phone numbers used to send scam texts, surprisingly, we only see 1,267 IP addresses abused to host domains based in the UK. It might be that scammers abuse the infrastructure outside their target country to make it challenging for takedown companies and law enforcement, giving themselves more time before the domain is taken down.

Domain Randomness. We evaluate the randomness of second-level domains (SLDs) abused in smishing attacks. We find that 60.3% SLDs contain a (nontrivial) dictionary word, indicating use of meaningful words. This may imply the use

AS	Unique IPs	Country	Unique IPs
Amazon	42,266	United States	20,452
Akamai	1,824	China	9,741
Hostinger	1,463	India	8,080
Cogent	502	Brazil	4,069
Tencent	166	Ireland	2,801

TABLE XI: Top 5 Autonomous Systems (ASes) and countries where scam URLs are hosted, excluding Cloudflare.

of words such as 'cancel,' 'track,' and 'ship' along with brand names (including typosquatted brand names), which would be consistent with common impersonation tactics inline with previous research [67], [18], [68], [69], [70]. While the mean Shannon entropy is 2.88, suggesting moderate randomness, the average vowel-to-consonant ratio is 0.64, higher than expected for randomly generated strings. This indicates a bias toward pronounceable patterns. We select a random sample of 100 domain names and find 67% to be brand typos (e.g., evrlgb-couriers, verify-myapplepay, hsbc-cancel-payment, icloud-uk), also known as combosquatting [70]. The rest contain arbitrary character sequences (e.g., dcfcy, onapuw, azirew, acozir).

Domain Lifetime. We find the lifetime of 5,854 unique domains (excluding URL shorteners), with a median of 118 days. We use survival analysis to investigate lifetime of all identified scam domains with overall survival probability (Fig. 7). The dotted lines in the plot are the 95% confidence interval. While 83.3% of the domains are active after 10 days, 65.8% are active after 100 days. This demonstrates that longer lived domains live longer. Previous work found that more than 40% of the domains from the US Federal Trade Commission (FTC) user reports were active for over 100 days [36]. On the contrary, Nahapetyan et al. observe an average lifetime of 12.241 days with a median of 0.53 hours [32]. Others identify the average lifetime of scam domains as 59.4 days [71].

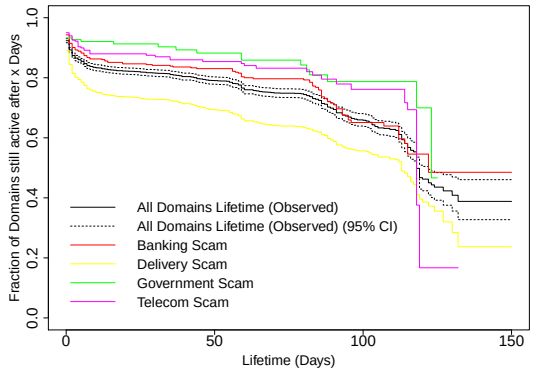


Fig. 7: Survival analysis of the lifetime of domains scammers abuse to conduct SMS scams.

The aggregate lifetime of all domains represents a broad view; we look into the domains scammers abuse to conduct different scams. Delivery scam domains have a median lifetime of 113 days, while government scam domains last longer, with a median of 123 days. For delivery scams, 55.7% of domains remain active after 100 days, but 78.7% of domains abused to conduct government scams are active after 100 days. For example, tax-rebate[.]top was active only for 9 days, while online-uk-claims[.]com remained active for 123 days. This shows the different takedown practices or domain registration preferences for different types of scams.

Particularly, the identified domains last so long as scammers register and re-register domain names, continuing to abuse them over time [72]. For example, we see uk-delivery[.]com was first seen on April 20, 2014, and has been likely re-registered multiple times since then.

Takeaway from RQ4. Scammers abuse third party services like URL shorteners and proxy services to evade detection. Infrastructure services should perform Know-Your-Customer (KYC) checks before allowing users to use their services and collaborate with threat intelligence organizations and mobile network operators to perform more effective takedowns. Domains abused to conduct scams have a median lifetime of 118 days, indicating that scammers are able to keep the domains active for a long time. Registrars need to proactively identify scam domains, particularly potential re-registeried scam domains, and perform takedowns on abused domains.

E. SMS Text Analysis

The text of the scam message provides insights into how scammers craft these messages to deceive victims into taking action. This subsection answers **RQ5** by working to understand the various lures and readability of the scam messages.

Scam Lures. Scammers design text messages to deceive victims, asking them to take action. We work to understand this deception by classifying the scams by lure using the typology identified by Stajano and Wilson [48]. Table XII presents the scams we classify by lures. Scammers use the time/urgency lure in all scam types, forcing users to make an impulsive decision. For example, using words like 'now,' 'urgent,' or 'immediately.' We find that scammers lure victims into Hi Mum/Dad scams by also employing distraction and kindness, similar to prior findings [30]. For example, scammers pretend to be a victim's child, provide a random reason for reaching out from a different number and request immediate help. Similarly, for Wrong Number scams, scammers provide unrelated details and asking a random question or sounding like a friend to deceive potential victims into replying, which can turn into a pig butchering scam [73].

Scammers use need and greed to lure victims into Government (12.6%), Telecom (13.7%) and Banking (11.7%) scams by offering tax refunds, points, cashback, or mentioning a suspicious payment, tempting the user to take the requested action in the text. Government (67%), Telecom (47.1%), Banking (44.2%), and Delivery (37.8%) scams also invoke authority, making it more convincing by impersonating government and private entities such as HMRC, EE, Barclays, or EVRI. This indicates that scammers do not just randomly draft scam texts but take advantage of various lures to deceive potential victims into taking an action. Depending on the type

Scam Lure	Definition	Scam Types
Authority	Scammers refer to trusted third parties and convince users to do things they would not do otherwise	Banking, Government, Delivery/Parcel & Telecom
Dishonesty	Scammers invite users willingly and knowingly into taking fraudulent action	-
Distraction	Scammers provide unrelated details to distract the user	Hi Mum/Dad & Wrong Number
Need & Greed	Scammers leverage users' greed and offer attractive benefits	Banking, Government & Telecom
Herd	Scammers convince that others have won taking the same risk	-
Kindness	Scammers leverage the willingness of people to help others	Wrong Number & Hi Mum/Dad
Time & Urgency	Scammers put time pressure on users so they make an irrational decision	All

TABLE XII: Scam types categorized by lures (adapted from Stajano and Wilson [48]).

of scam, scammers use one or more lures in the text of the message. Unsurprisingly, we do not find SMS scammers using herd and dishonesty lures despite these commonly used in cryptocurrency scam ads [74], [75].

Gunning Fog Index. Scammers design the text of the scam messages depending on their target victims. The text message's readability is essential as not all potential victims can understand complex sentences in scam text messages. To target a broader range of individuals, one would expect the scam text messages to be simple. Towards this end, we use the Gunning Fog Index (GFI), designed to compute the years of education required to understand a given passage/text [47]. We find that the mean of the Gunning Fog Index for conversational scam text messages is lower than others. While Hi Mum/Dad and Wrong number scams have a Gunning Fog Index of 4.6 and 3.9, respectively, Delivery/Parcel, Banking, Telecom, and Government impersonation scams have the Gunning Fog Index above 5.6, the maximum being 8.3 for Delivery/Parcel scams (Fig 8 in Appendix shows the GFI for all scam types). The Gunning Fog Index does not significantly differ between scam types. This indicates that scammers design straightforward scam texts that anyone with a maximum of 8th-grade education can read and understand.

Takeaway from RQ5. Dividing scams by lures used can help educate victims on common ruses. Scammers use authoritative lures for Banking, Government, Delivery, and Telecom scams. Users should ensure messages are from legitimate senders with trustworthy URLs. We find that most scam texts can be read and understood by those with a middle school education level, with no real differences between types.

V. Discussion

We next discuss findings we derive from analyzing user text reports in the context of our limitations.

Differentiating fraudulent from non-fraudulent messages is crucial in an increasingly challenging landscape. Our analysis reveals that over 40% of text messages are scams, observing on the other end a significant number of spam reports. However, we see that messages with fraudulent intent show different patterns than those without. Quite a number of our findings vary with intent like the likelihood of using a phone number vs. other sender IDs or the likelihood of using a URL. Previous research fails to differentiate between the scam and spam text messages, limiting its effectiveness in combating smishing [36], [31], [33], [32]. This distinction is of heightened importance when factors like the infrastructure or text patterns used are key ones that feed into XDRs at mobile network operators. When these factors differ, stakeholders require to adopt distinct strategies to address fraudulent messages effectively using user reports [15]. We call upon other researchers to also make this vital distinction in order to positively protect users from harm.

Scammers shift toward non-URL-based lures, which bypass mobile network operators's XDR systems. With the uptake in SMS scams, we see "Wrong Number" and other conversational scams dominating the submissions — here we identify six new scam types above narrower prior work [19], [30]. These new types of conversational scams bypass mobile network operators' XDR systems, which are more effective at detecting URL-based threats. This indicates a critical gap in current detection mechanisms, as most reported scams are texts without URLs, and shows the need to enhance algorithms for identifying such messages. This feature also highlights a key distinction from email phishing, posing challenges to the adoption of effective defense mechanisms [76], [77]. The shift in SMS scams echoes an evasion strategy similar to technical support scams, which have advanced from URL-based scareware URLs [78] to call-based attacks via malvertising [79], bypassing conventional URL-focused defenses.

We note that Telecom regulators maintain a list of phone numbers (aka do not originate or DNO) from organizations like banks to combat call spoofing [80]. We suggest that mobile network operators query phone numbers in the text of an SMS that posits to be an organization against the DNO list as an effective legitimacy checking mechanism, in the direction of Domain-based Message Authentication, Reporting & Conformance (DMARC) used in mail servers [81] and telephony blocklists [82].

Exploiting online encrypted messaging platforms to send scam texts. We also notice scammers use email addresses for delivery scams indicating RCS/iMessage scams. Rich communication service (RCS) and iMessage allow users to send encrypted SMS over the internet instead of SMS over mobile network operator. In addition to emails, we identify a scam campaign with over 44 deliver scam texts known to send RCS/iMessage scam texts called 'darcula' — a phishing-as-a-service platform [62], [83]. As mobile network operators improve their XDRs, scammers are turning to similar and cheaper

alternatives like iMessage and RCS to bypass detection. The shift in smishing from traditional SMS to encrypted messaging platforms mirrors how traditional email-based phishing [84] has evolved towards cryptocurrency platforms [85], [86]. As RCS/iMessage texts directly bypass the mobile network operator's XDRs, we suggest services like Apple and Google collaborate with mobile network operators to identify and proactively block scams, enabling a more coordinated and comprehensive mitigation strategy.

Limitations. We receive the user reports from only one mobile network operator. The unavailability of SMS scam data is a general limitation here. We make significant progress on this limitation for one mobile network operator by devising a processing pipeline that can characterize popular, known scam types and identify novel ones. As a key contribution, we see that the lures scammers use to deceive victims into taking an action differs based on the scam type. Understanding the lures used could assist interveners to create more effective warnings and better user education about ongoing scams. Interveners include private companies like the Students Loans Company [87] and governments [88].

While we identify spam and scams, we also flag 23.26% text reports as 'Unknown.' Even though the URLs in the reports are not flagged by an AV vendor, previous research has shown that more than 94% of the blacklisted domain names do not appear in public blacklists for several weeks or even months after they are first reported in abuse complaints [36]. There is room for improving our text characterization and scam type identification. Some of the misclassifications we observe reveal fraudulent messages mimicking legitimate messages, with some being as simple as 'Hi' — requiring more sophisticated detection mechanisms.

However, the scope of our paper is not to devise such detection mechanisms, but to offer a first look into SMS scams and scammer strategies through the lens of a large base of recent user reports. Our evaluation metrics (88% for spam messages, 89% for known scam types, 80% new types) demonstrate high standards to answer our research questions, especially considering the scale and real-world conditions of our evaluation. While a deeper understanding of *unknowns* and *other* new types of scams is the scope of our future work, we reveal six emerging scam types (that group together) (§IV-B); finer breakdowns will be a valuable future direction. A challenge to address in this direction is to adopt novel detection mechanisms [89] to identify and classify new trends.

Despite the limitations of our research, we present the first measurement of the SMS user report ecosystem, shedding light on the services exploited by scammers to perpetrate this fraudulent activity and de-conflating noisy data such as spam.

VI. Related Work

Our work fits into the broader literature on suspicious SMS messages. We divide them by data source. Previous researchers have investigated public *SMS gateways* to understand the SMS ecosystem and identify SMS phishing text messages. While Reaves et al. [90], [91] and Moreno et al. [92] mention a minority of SMS messages containing malicious URLs, but they report a wide use of URL shorteners that cannot resolve. Instead, by enriching our dataset with telemetry captured as SMSs are reported, we gain the ability to, for instance, resolve these shorteners and reliably query relevant contextual information. Nahapetyan et al. identify over $67.9k$ SMS phishing messages using identifiers such as phone numbers, email addresses, and One-Time Codes without removing spam [32].

Some have collated victim reports [93], [31] or crowdsourced them [94], [95]. Others used intensive techniques like in-depth interviews [20], [96] and analyzing news articles [97]. More relatedly, some researchers investigated SMS messages caught by XDRs and interacted with the scammers behind the messages, albeit for only one type of scam (Hi Mum/Dad) [30]. Similarly, Agarwal et al. investigated two months of blocked SMS text messages from one UK mobile network operator and manually categorized them into six scam types [19]. Our methods build on their insights into attacker strategies, but we see that just focusing on these six scam types offers an incomplete picture of the landscape.

It is common to mix together spam and scam messages. A few groups collect spam and smishing (scam) text messages reported by users on Twitter, but treat them all as spam [31], [98]. Another recent work crowdsources suspicious text messages from users in the US, categorizing them all as scams [33]. Srinivasan et al. consider the SMS reports users submit to the Federal Trade Commission (FTC) and third parties as spam [36]. Contrastingly, we break down this difference which we show to materially change our results.

Previous work has devised machine learning models to detect smishing messages [99], [100], [101], [102], [103], [104], [105], [106], [107], using old spam datasets [108], [95], [109]. We provide a methodological framework to differentiate spam and identify new scam types from user reports that could help researchers enhance these models.

VII. Conclusion

It is mandatory in the UK to reimburse fraud victims after they fall victim to a scam and directly send the fraudster money [110]. Because of this, UK infrastructure operators are increasingly being pressured by banks to reduce fraud. These incentives have played out in our work investigating hundreds of SMS scams reported during a 4 month period in 2024. We winnow down over $1m$ user reports into $213k$ scam texts using a careful, multi-layered methodology. We investigate these scam texts and find that most that get through the mobile network operators's XDR have no URL. We hypothesize this is from direct pressure and vast industry experience blocking URL-based scams. We highlight the emerging trend of call back scams, where scammers entice victims to call them on a phone number. We encourage those designing filtering products like XDRs for mobile network operators to more carefully consider conversation scams, since they often lead to fraud, even if not from the start.

ACKNOWLEDGMENTS

We would like to thank *hidden for anonymous submission*.

ETHICS CONSIDERATIONS

Our work has some ethical considerations, we next explain how we mitigate risks. We get access to all 7726 reports submitted by users to a major mobile network operator. The first set of safeguard measures is taken by the mobile network operator we collaborate with, who redact users report to anonymize the identity of the users who reports the messages and remove personally identifiable information (PII) from text messages. For example, we receive the domain name instead of the complete URL, where a URL might contain personally identifiable information (PII) such as phone number. Our collaborators also ensure that our sharing agreement abides by the terms of the 7726 service.

The second set of safeguards is taken by the authors of this paper, who have designed a protocol and an impact assessment to ensure user reports are stored and processed safely. Most critically, we receive and process sender IDs without having explicit consent of the actual sender who has been reported. This includes a phone number or an alphanumeric shortcode used to send the message or call a user. To protect these users, we avoid any attempts to use this information to identify individuals and we do not interact with any of the phone numbers. Our experiments are designed to work over aggregates, and our RQs aim at addressing gaps that benefit the research community. Our goal is to make strides in enhancing the safety of mobile users — broader societal benefits — by providing a deeper understanding of the types of scams being reported and the lure strategies employed by scammers; and by identifying the mobile network operators that they abuse and their current network status. Our department's research ethics committee evaluated and approved our measures to minimize risks and this study.

REFERENCES

[1] Trend Micro, "Scam texts on the rise: February sees a 73% increase in fraudulent sms," https://news.trendmicro.com/en-au/2025/03/16/sms-scams-skyrocket-in-feb/, 2025.

[2] S. Fadilpašić, "Billions of fake sms messages were sent last year — here's why it's a growing problem for all of us," https://www.techradar.com/pro/security/billions-of-fake-sms-messages-were-sent-last-year-heres-why-its-a-growing-problem-for-all-of-us, 2024.

[3] RTE, "Mwc hears sms fraud a headache for telecom operators," https://www.rte.ie/news/business/2024/0228/1434908-mwc-hears-sms-fraud-a-headache-for-telecom-operators/, 2024.

[4] OpenPhish, "OpenPhish - Phishing Intelligence," https://openphish.com/index.html, 2024.

[5] Cisco Talos Intelligence Group, "PhishTank," https://www.phishtank.com, 2024.

[6] Anti-Phishing Working Group, "The APWG eCrime Exchange (eCX)," https://apwg.org/ecx/, 2004.

[7] Federal Trade Commission (FTC), "Top text scams of 2024," https://www.ftc.gov/news-events/data-visualizations/data-spotlight/2025/04/top-text-scams-2024, 2025.

[8] Federal Bureau of Investigation, "2024 internet crime report," https://www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf, 2025.

[9] Global Anti-Scam Alliance (GASA), "The state of scams in the united kingdom 2024," https://www.gasa.org/_files/ugd/7bdaac_bc34e713c6434551a9c8f25207e1be9d.pdf, 2024.

[10] UK Finance, "Annual fraud report 2025," https://www.ukfinance.org.uk/policy-and-guidance/reports-and-publications/annual-fraud-report-2025, 2025.

[11] ACCC Scamwatch, "Scam statistics," https://www.scamwatch.gov.au/research-and-resources/scam-statistics, 2025.

[12] Radio New Zealand, "Scams cost new zealanders $4.2m, with 'smishing' on the rise," https://www.rnz.co.nz/news/national/497904/scams-cost-new-zealanders-4-point-2m-with-smishing-on-the-rise, 2023.

[13] Federal Bureau of Investigation, "2023 internet crime report," https://www.ic3.gov/AnnualReport/Reports/2023_IC3Report.pdf, 2024.

[14] Canadian Telecommunications Association, "Telecom providers' anti-fraud initiatives," https://canadatelecoms.ca/consumer_resource/telecom-providers-anti-fraud-initiatives/, 2025.

[15] Mavenir, "Spamshield messaging fraud," https://www.mavenir.com/portfolio/mavapps/fraud-security/spamshield-messaging-fraud/, 2024.

[16] TCF New Zealand, "Scam prevention measures," https://www.tcf.org.nz/digital-living/online-safety/scam-prevention-measures, 2025.

[17] Telstra, "Helping to reduce cyber threats by blocking malicious sms," https://www.telstra.com.au/cyber-security-and-safety/active-scams/block-malicious-sms-with-sms-scam-filter, 2025.

[18] S. Agarwal and M. Vasek, "Examining newly registered phishing domains at scale," in *Workshop on the Economics of Information Security (WEIS)*, 2025.

[19] S. Agarwal, E. Harvey, and M. Vasek, "Poster: A comprehensive categorization of sms scams," in *Proceedings of the 2024 ACM on Internet Measurement Conference*, ser. IMC '24. Association for Computing Machinery, 2024, p. 755–756.

[20] M. L. Rahman, D. Timko, H. Wali, and A. Neupane, "Users really do respond to smishing," in *Proceedings of the Thirteenth ACM Conference on Data and Application Security and Privacy*, ser. CODASPY '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 49–60. [Online]. Available: https://doi.org/10.1145/3577923.3583640

[21] Ofcom, "45 million people targeted by scam calls and texts this summer," https://www.ofcom.org.uk/phones-and-broadband/scam-calls-and-messages/45-million-people-targeted-by-scams/, 2021.

[22] CIFAS, "£7.5 billion stolen as 1 in 10 britons fall victim to scams in 12 months," https://www.cifas.org.uk/newsroom/stateofscams, 2023.

[23] D. Timko, D. H. Castillo, and M. L. Rahman, "Understanding influences on sms phishing detection: User behavior, demographics, and message attributes," in *Symposium on Usable Security and Privacy (USEC) 2025*, 2025.

[24] S. Tabassum, C. Faklaris, and H. R. Lipford, "What drives SMiShing susceptibility? a U.S. interview study of how and why mobile phone users judge text messages to be real or fake," in *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*. Philadelphia, PA: USENIX Association, Aug. 2024, pp. 393–411. [Online]. Available: https://www.usenix.org/conference/soups2024/presentation/tabassum-sarah

[25] Ofcom, "How to report scam texts and mobile calls to 7726," https://www.ofcom.org.uk/phones-and-broadband/scam-calls-and-messages/7726-reporting-scam-texts-and-calls/, 2024.

[26] AT&T, "Report unwanted text messages," https://about.att.com/pages/cyberaware/ni/blog/forward_7726, 2024.

[27] D. of Internal Affairs, "Report txt spam," https://www.dia.govt.nz/Spam-Report-TXT-Spam, 2024.

[28] Government of Canada, "Reporting spam text messages to 7726," https://www.getcybersafe.gc.ca/en/blogs/reporting-spam-text-messages-7726, 2023.

[29] D. Timko, D. H. Castillo, and M. L. Rahman, "A quantitative study of SMS phishing detection," 2024, https://arxiv.org/abs/2311.06911.

[30] S. Agarwal, E. Harvey, E. Mariconti, G. Suarez-Tangil, and M. Vasek, "'Hey mum, i dropped my phone down the toilet': Investigating hi mum and dad sms scams in the united kingdom," in *Usenix Security Symposium*. USENIX Association, 2025.

[31] S. Tang, X. Mi, Y. Li, X. Wang, and K. Chen, "Clues in tweets: Twitter-guided discovery and analysis of sms spam," in *ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '22, 2022, p. 2751–2764.

[32] A. Nahapetyan, S. Prasad, K. Childs, A. Oest, Y. Ladwig, A. Kapravelos, and B. Reaves, "On SMS phishing tactics and infrastructure," in *IEEE Symposium on Security and Privacy*, 2024, pp. 169–169.

[33] D. Timko and M. L. Rahman, "Smishing dataset I: Phishing SMS dataset from Smishtank.com," in *ACM Conference on Data and Application Security and Privacy*, ser. CODASPY '24, 2024, p. 289–294.

[34] Ofcom, "Mobile telephony adoption in the United Kingdom (UK) from 2007 to 2024," https://www.statista.com/statistics/272221/mobile-telephony-adoption-in-the-united-kingdom-uk/, 2024.

[35] ——, "Telecommunications market data update q2 2024," https://www.ofcom.org.uk/phones-and-broadband/telecoms-infrastructure/telecommunications-market-data-update-q2-2024/, 2024.

[36] B. Srinivasan, P. Gupta, M. Antonakakis, and M. Ahamad, "Understanding cross-channel abuse with SMS-spam support infrastructure attribution," in *21st European Symposium on Research in Computer Security (ESORICS)*. Springer, 2016, pp. 3–26.

[37] D. Morris, "What is a HLR Lookup?" https://www.hlrlookup.com/what-is-a-hlr-lookup/, 2021.

[38] ——, "Full API Result - Response Body Format and Explanations," https://www.hlrlookup.com/knowledge/full-api-result, 2021.

[39] S. Marine, "Mobile network operator," https://www.stourmarine.com.

[40] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.

[41] VirusTotal, "VirusTotal," https://docs.virustotal.com/docs/how-it-works.

[42] Spamhaus, "Strengthening trust and safety across the internet," https://www.spamhaus.org/, 2024.

[43] ——, "Passive DNS," https://docs.spamhaus.com/pdns/docs/source/index.html, 2024.

[44] WhoisXMLAPI, "Whois api offers unified & consistent data," https://whois.whoisxmlapi.com, 2024.

[45] IPinfo, "Trusted ip data provider, from ipv6 to ipv4 - ipinfo.io," https://ipinfo.io/, 2024.

[46] ——, "Data downloads," https://ipinfo.io/developers/database-download, 2024.

[47] R. Gunning, "The fog index after twenty years," *Journal of Business Communication*, vol. 6, no. 2, pp. 3–13, 1969.

[48] F. Stajano and P. Wilson, "Understanding scam victims: seven principles for systems security," *Commun. ACM*, vol. 54, no. 3, p. 70–75, Mar. 2011.

[49] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[50] Ofcom, "Experiences of suspicious calls, texts and app messages," https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/telecoms-research/scams-research/experiences-of-suspicious-calls-texts-and-app-messages-research-2024.pdf, 2024.

[51] A. Oest, Y. Safaei, P. Zhang, B. Wardman, K. Tyers, Y. Shoshitaishvili, and A. Doupé, "PhishTime: Continuous longitudinal measurement of the effectiveness of anti-phishing blacklists," in *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, Aug. 2020, pp. 379–396.

[52] P. Peng, L. Yang, L. Song, and G. Wang, "Opening the blackbox of virustotal: Analyzing online phishing scan engines," in *Proceedings of the Internet Measurement Conference*, ser. IMC '19. Association for Computing Machinery, 2019, p. 478–485.

[53] T. Ramsey, "5 text scams to watch out for," https://www.which.co.uk/news/article/5-text-scams-to-watch-out-for-asQ611R5pfQM, 2024.

[54] AppointmentReminders.com, "Fake appointment text messages," https://www.appointmentreminders.com/fake-appointment-text-messages/, 2024.

[55] PowerTextor, "Avoiding fake appointment text messages: Tips for patients," https://powertextor.com/blogs/avoid-fake-doctor-appointment-text-patients/, 2024.

[56] Electricity North West, "How to avoid an energy scam," https://www.enwl.co.uk/power-cuts/extra-care/how-to-avoid-energy-scams/, 2025.

[57] N. Miramirkhani, O. Starov, and N. Nikiforakis, "Dial one for scam: A large-scale analysis of technical support scams," in *Proceedings of Network and Distributed System Security (NDSS) Symposium*, 2017.

[58] UK Debt Collection News, "Debt collection sms text scam warning," https://www.ukdcnews.co.uk/debt-collection-2/debt-collection-scam-warning/, 2023.

[59] Federal Trade Commission (FTC), "That random text offering you a job? it's probably a scam," https://consumer.ftc.gov/consumer-alerts/2024/11/random-text-offering-you-job-its-probably-scam, 2024.

[60] Which?, "Job scams and employment fraud," https://www.which.co.uk/consumer-rights/advice/job-scams-aFQaP4P9btJv, 2024.

[61] Federal Trade Commission (FTC), "New ftc data show skyrocketing consumer reports about game-like online job scams," https://www.ftc.gov/news-events/news/press-releases/2024/12/new-ftc-data-show-skyrocketing-consumer-reports-about-game-online-job-scams, 2024.

[62] H. Everett, "Out of the shadows – 'darcula' iMessage and RCS smishing attacks target USPS and global postal services," https://www.netcraft.com/blog/darcula-smishing-attacks-target-usps-and-global-postal-services/, 2024.

[63] U. Finance, "Men sentenced for using SIM farm to steal over £220,000 from banking customers," https://www.ukfinance.org.uk/news-and-insight/press-release/men-sentenced-using-sim-farm-steal-over-ps220000-banking-customers, 2024.

[64] M. Korczynski, S. Tajalizadehkhoob, A. Noroozian, M. Wullink, C. Hesselman, and M. Van Eeten, "Reputation metrics design to improve intermediary incentives for security of tlds," in *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*, 2017, pp. 579–594.

[65] C. Lever, R. Walls, Y. Nadji, D. Dagon, P. McDaniel, and M. Antonakakis, "Domain-z: 28 registrations later measuring the exploitation of residual trust in domains," in *2016 IEEE Symposium on Security and Privacy*, 2016, pp. 691–706.

[66] N. Kumar, S. Ghewari, H. Tupsamudre, M. Shukla, and S. Lodha, "When diversity meets hostility: A study of domain squatting abuse in online banking," in *2021 APWG Symposium on Electronic Crime Research (eCrime)*, 2021, pp. 1–15.

[67] P. Agten, W. Joosen, F. Piessens, and N. Nikiforakis, "Seven months' worth of mistakes: A longitudinal study of typosquatting abuse," in *Proceedings of Network and Distributed System Security (NDSS) Symposium*, 2015.

[68] T. Moore and B. Edelman, "Measuring the perpetrators and funders of typosquatting," in *International Conference on Financial Cryptography and Data Security*. Springer, 2010, pp. 175–191.

[69] J. Szurdi, B. Kocso, G. Cseh, J. Spring, M. Felegyhazi, and C. Kanich, "The long "Taile" of typosquatting domain names," in *23rd USENIX Security Symposium (USENIX Security 14)*. San Diego, CA: USENIX Association, Aug. 2014, pp. 191–206.

[70] P. Kintis, N. Miramirkhani, C. Lever, Y. Chen, R. Romero-Gómez, N. Pitropakis, N. Nikiforakis, and M. Antonakakis, "Hiding in plain sight: A longitudinal study of combosquatting abuse," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 569–586.

[71] P. Kotzias, M. Pachilakis, J. Aldana-Iuit, J. Caballero, I. Sánchez-Rola, and L. Bilge, "Ctrl+ alt+ deceive: Quantifying user exposure to online scams." in *Proceedings of Network and Distributed System Security (NDSS) Symposium*, 2025.

[72] S. Hao, M. Thomas, V. Paxson, N. Feamster, C. Kreibich, C. Grier, and S. Hollenbeck, "Understanding the domain registration behavior of spammers," in *Proceedings of the 2013 Conference on Internet Measurement Conference*, 2013, p. 63–76.

[73] M. Ordekian, A. Papasavva, E. Mariconti, and M. Vasek, "A sinister fattening: Dissecting the tales of pig butchering and other cryptocurrency scams," in *2024 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2024.

[74] G. A. Siu, A. Hutchings, M. Vasek, and T. Moore, ""Invest in crypto!": An analysis of investment scam advertisements found in bitcointalk," in *2022 APWG Symposium on Electronic Crime Research (eCrime)*, 2022, pp. 1–12.

[75] S. Agarwal, G. Atondo-Siu, M. Ordekian, A. Hutchings, E. Mariconti, and M. Vasek, "Short paper: DeFi deception—uncovering the prevalence of rugpulls in cryptocurrency projects," in *Financial Cryptography and Data Security*. Springer, 2024, pp. 363–372.

[76] A. Oest, Y. Safei, A. Doupé, G.-J. Ahn, B. Wardman, and G. Warner, "Inside a phisher's mind: Understanding the anti-phishing ecosystem through phishing kit analysis," in *2018 APWG Symposium on Electronic Crime Research (eCrime)*, 2018, pp. 1–12.

[77] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from urls," *Expert Systems with Applications*, vol. 117, pp. 345–357, 2019.

[78] B. Stone-Gross, R. Abman, R. A. Kemmerer, C. Kruegel, D. G. Steigerwald, and G. Vigna, "The underground economy of fake antivirus software," in *Economics of information security and privacy III*. Springer, 2012, pp. 55–78.

[79] N. Miramirkhani, O. Starov, and N. Nikiforakis, "Dial one for scam: Analyzing and detecting technical support scams," in *Proceedings of Network and Distributed System Security (NDSS) Symposium*, 2016.

[80] Ofcom, "The Do Not Originate (DNO) list," https://www.ofcom.org.uk/phones-and-broadband/scam-calls-and-messages/do-not-originate/, 2023.

[81] M. Kucherawy and E. Zwicky, "Domain-based message authentication, reporting, and conformance (dmarc)," 2015, https://dmarc.org/overview/.

[82] S. Pandit, R. Perdisci, M. Ahamad, and P. Gupta, "Towards measuring the effectiveness of telephony blacklists." in *Proceedings of Network and Distributed System Security (NDSS) Symposium*, 2018.

[83] U. Akyazi, M. van Eeten, and C. H. Gañán, "Measuring cybercrime as a service (caas) offerings in a cybercrime forum," in *Workshop on the Economics of Information Security (WEIS)*, 2021.

[84] T. Moore and R. Clayton, "An empirical analysis of the current state of phishing attack and defence." in *Workshop on the Economics of Information Security (WEIS)*, 2007.

[85] Z. Chen, D. Luo, Y. Hu, L. Wu, B. He, and Y. Zhou, "Dissecting payload-based transaction phishing on ethereum," in *Proceedings of Network and Distributed System Security (NDSS) Symposium*, 2025.

[86] L. Chen, J. Peng, Y. Liu, J. Li, F. Xie, and Z. Zheng, "Phishing scams detection in ethereum transaction network," *ACM Trans. Internet Technol.*, vol. 21, no. 1, Dec. 2020.

[87] Student Loans Company, "Student smishing scams on the rise," https://www.gov.uk/government/news/student-smishing-scams-on-the-rise, 2024.

[88] Federal Trade Commission (FTC), "How to recognize and report spam text messages," https://consumer.ftc.gov/articles/how-recognize-and-report-spam-text-messages, 2022.

[89] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal processing*, vol. 99, pp. 215–249, 2014.

[90] B. Reaves, N. Scaife, D. Tian, L. Blue, P. Traynor, and K. R. Butler, "Sending out an SMS: Characterizing the security of the SMS ecosystem with public gateways," in *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2016, pp. 339–356.

[91] B. Reaves, L. Vargas, N. Scaife, D. Tian, L. Blue, P. Traynor, and K. R. Butler, "Characterizing the security of the SMS ecosystem with public gateways," *ACM Transactions on Privacy and Security (TOPS)*, vol. 22, no. 1, pp. 1–31, 2018.

[92] J. M. Moreno, S. Matic, N. Vallina-Rodriguez, and J. Tapiador, "Your code is 0000: An analysis of the disposable phone numbers ecosystem," in *2023 7th Network Traffic Measurement and Analysis Conference (TMA)*. IEEE, 2023, pp. 1–10.

[93] S. S. Roy, U. Karanjit, and S. Nilizadeh, "Evaluating the effectiveness of phishing reports on Twitter," in *2021 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2021, pp. 1–13.

[94] D. Timko and M. L. Rahman, "Commercial anti-smishing tools and their comparative effectiveness against modern threats," in *ACM Conference on Security and Privacy in Wireless and Mobile Networks*, ser. WiSec '23, 2023, p. 1–12.

[95] T. Chen and M.-Y. Kan, "Creating a live, public short message service corpus: the NUS SMS corpus," *Language Resources and Evaluation*, vol. 47, pp. 299–335, 2013.

[96] J. S. Downs, M. B. Holbrook, and L. F. Cranor, "Decision strategies and susceptibility to phishing," in *Proceedings of the Second Symposium on Usable Privacy and Security*, ser. SOUPS '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 79–90. [Online]. Available: https://doi.org/10.1145/1143120.1143131

[97] M. Bitaab, H. Cho, A. Oest, P. Zhang, Z. Sun, R. Pourmohamad, D. Kim, T. Bao, R. Wang, Y. Shoshitaishvili, A. Doupé, and G.-J. Ahn, "Scam pandemic: How attackers exploit public fear through phishing," in *2020 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2020, pp. 1–10.

[98] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam: the underground on 140 characters or less," in *Proceedings of the 17th ACM Conference on Computer and Communications Security*, ser. CCS '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 27–37.

[99] S. Hosseinpour and S. Das, "Poster: A multi-signal model for detecting evasive smishing," in *18th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, ser. WiSec 2025. New York, NY, USA: Association for Computing Machinery, 2025, p. 292–293.

[100] A. K. Jain, S. K. Yadav, and N. Choudhary, "A novel approach to detect spam and smishing sms using machine learning techniques," *International Journal of E-Services and Mobile Applications (IJESMA)*, vol. 12, no. 1, pp. 21–38, 2020.

[101] A. K. Jain and B. B. Gupta, "Feature based approach for detection of smishing messages in the mobile environment," *Journal of Information Technology Research (JITR)*, vol. 12, no. 2, pp. 17–35, apr 2019.

[102] D. Goel and A. K. Jain, "Smishing-classifier: a novel framework for detection of smishing attack in mobile environment," in *Smart and Innovative Trends in Next Generation Computing Technologies*. Springer, 2018, pp. 502–512.

[103] A. K. Jain and B. B. Gupta, "Rule-based framework for detection of smishing messages in mobile environment," *Procedia Computer Science*, vol. 125, pp. 617–623, 2018.

[104] J. W. Joo, S. Y. Moon, S. Singh, and J. H. Park, "S-detector: an enhanced security model for detecting smishing attack for mobile computing," *Telecommunication Systems*, vol. 66, pp. 29–38, 2017.

[105] S. Mishra and D. Soni, "Smishing detector: A security model to detect smishing through sms content analysis and url behavior analysis," *Future Generation Computer Systems*, vol. 108, pp. 803–815, 2020.

[106] ——, "A content-based approach for detecting smishing in mobile environment," in *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM)*, 2019.

[107] G. Sonowal and K. Kuppusamy, "SmiDCA: an anti-smishing model with machine learning approach," *The Computer Journal*, vol. 61, no. 8, pp. 1143–1157, 2018.

[108] T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami, "Contributions to the study of SMS spam filtering: new collection and results," in *ACM Symposium on Document Engineering*, ser. DocEng '11, 2011, p. 259–262.

[109] S. J. Delany, M. Buckley, and D. Greene, "Sms spam filtering: Methods and data," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9899–9908, 2012.

[110] UK Finance, "Authorised push payment fraud reimbursement," 2025, https://www.ukfinance.org.uk/authorised-push-payment-fraud-reimbursement.

APPENDIX

A. Open AI Prompt: Message Classification

You will receive a json object with an ‘id’ and a ‘message’. The ‘id’ is the id of the message and the ‘message’ is text of a SMS. Based on the instructions below, process the message and return a json object. Instructions: 1. Identify the brand or organization that the message is trying to impersonate in the text. Return empty if none. (”named_entity” key in the json object. This key should always be returned in the json.) 2. Classify the type of smishing message (”scam_type” key in the json. This key should always be returned in the json.) The scam_types can be: ”Hey mum/dad” - text addressed to mum/mom or a dad and asking to text/call back potentially giving a reason about using a different mobile number. ”Delivery/Parcel” - text impersonating a parcel/delivery company asking to click on a link, text back or call on a number ”Banking” - text impersonating a bank or a financial institution asking to click on a link, text back or call on a number ”Government” - text impersonating a government organization asking to click on a link, text back or call on a number ”Telecom” - text impersonating a mobile network operator asking to click on a link, text back or call on a number ”Wrong number” - text with a normal greeting or asking about someone or to reply back ”Spam” - illicit marketing message including casino, betting, random draws, etc ”Others” - If it is does not fit as one of the above category 3. If the ”scam_type” classified is ”Others”, then identify a category for the text. Return empty otherwise. (”other_category” key in the json object. This key should always be returned in the json.) 4. Every json object should include the ”id” of the message being classified.

B. Open AI Prompt: Detecting Scam Lures

You will receive a json object with an ‘id’ and a ‘message’. The ‘id’ is the id of the message and the ‘message’ is text of a scam SMS. Based on the instructions below, process the message and return a json object. Instructions: 1. Provide which lure principles apply for each text message (”lure_principles” key should be a list and always be provided in the json object. If you cannot detect any lure principles, leave the list empty.) Lure principles are: a) Distraction Principle - providing various reasons to distract the user. b) Authority Principle - providing trust to the user to not question authority. could be done by making references to legitimate entities. c) Herd Principle - encouraging a user to not miss out on opportunities by relating to the popularity of a scheme. convincing how others have won things or take the same risk. d) Dishonesty Principle - inviting users willingly and knowingly to participate in a fraudulent scheme. e) Kindness Principle - Fraudsters leverage the willingness of people to help others. f) Need and Greed Principle - leveraging user’s greed and offering attractive (monetary) benefits so user would take an action asked in the text. g) Time/Urgency Principle - putting time pressure on user so they make a rushed decision. 2. Every json object should include the ”id” of the message being classified.

C. Other Scam Types Examples

An Enforcement Agent has been scheduled to attend your property, call <name> NOW ON <phone number> to prevent this action. Quote ref <reference number>

Can you do this EPC? <postcode> Requested day/time: Call to confirm Your fee: £40 Accept job here: [URL]

<First Name> is due a Kennel Cough Vaccine on 20/4/24. Call us on <phone number> to book. Thank you.

Ensure we hold the correct student finance information for your April payment by visiting:[URL]

<customer name>, Customer ID: <id>. You should receive a letter from us soon regarding <brand name>. It is important that you resolve this matter online as soon as possible by visiting [URL] Thanks, <brand name>.

[URL] You have not yet paid £39.39 due to <brand name> INSURANCE SERVICES GR. Got a question about your account? Text us on <phone number>. Ref JCF000052164

D. The Gunning Fog Index

Fog Index	Reading level by grade
17	College graduate
16	College senior
15	College junior
14	College sophomore
13	College freshman
12	High school senior
11	High school junior
10	High school sophomore
9	High school freshman
8	Eighth grade
7	Seventh grade
6	Sixth grade
5	Fifth grade

TABLE XIII: Education level required as per the Gunning fog index.

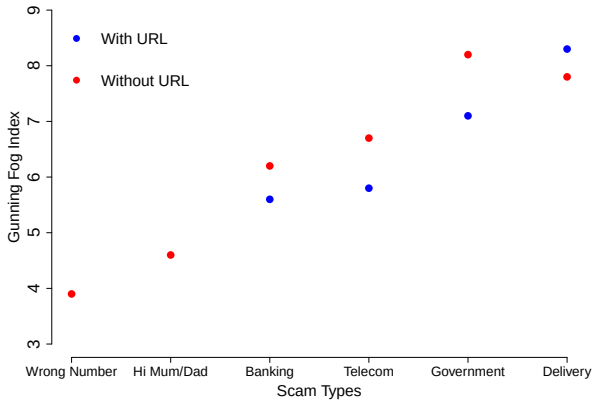


Fig. 8: Gunning Fog Index for six scam types with and without URLs.

Scam Type	Banking	Government	Delivery/Parcel	Telecom	Wrong Number	Hi Mum/Dad
Banking	6,327					
Government	32	1,222				
Delivery/Parcel	183	52	11,648			
Telecom	47	88	57	5,742		
Wrong Number	47	13	29	35	18,404	
Hi Mum/Dad	51	0	52	10	490	8,748

TABLE XIV: Common Sender IDs scammers abuse to send multiple scams.