# The Imitation Game: Exploring Brand Impersonation Attacks on Social Media Platforms

Bhupendra Acharya
*CISPA*

Dario Lazzaro
*University of Genoa*

Efrén López-Morales
*Texas A&M University-Corpus Christi*

Adam Oest
*PayPal Inc.*

Muhammad Saad
*PayPal Inc.*

Antonio Emanuele Cinà
*University of Genoa*

Lea Schönherr
*CISPA*

Thorsten Holz
*CISPA*

## Abstract

The rise of social media users has led to an increase in customer support services offered by brands on various platforms. Unfortunately, attackers also use this as an opportunity to trick victims through fake profiles that imitate official brand accounts. In this work, we provide a comprehensive overview of such *brand impersonation attacks* on social media.

We analyze the fake profile creation and user engagement processes on X, Instagram, Telegram, and YouTube and quantify their impact. Between May and October 2023, we collected 1.3 million user profiles, 33 million posts, and publicly available profile metadata, wherein we found 349,411 *squatted* accounts targeting 2,625 of 2,847 major international brands. Analyzing profile engagement and user creation techniques, we show that squatting profiles persistently perform various novel attacks in addition to classic abuse such as social engineering, phishing, and copyright infringement. By sharing our findings with the top 100 brands and collaborating with one of them, we further validate the real-world implications of such abuse. Our research highlights a weakness in the ability of social media platforms to protect brands and users from attacks based on username squatting. Alongside strategies such as customer education and clear indicators of trust, our detection model can be used by platforms as a countermeasure to proactively detect abusive accounts.

## 1 Introduction

Fraudsters perform social engineering attacks by mimicking popular brands and tricking victims into giving away sensitive information for financial gain. Traditionally, online brand impersonation attacks have been launched using website cloning (e.g., social engineering via phishing [21, 39] and typosquatting [3, 48, 74, 75, 81]) or offline engagement (e.g., fake technical support calls [47, 56, 72]). In both types of scams, fraudsters trick their victims into disclosing sensitive personal information that can be monetized. Over time, these brand impersonation scams have evolved and fraudsters have adopted increasingly complex methods to deceive their victims [43, 77, 80, 83].

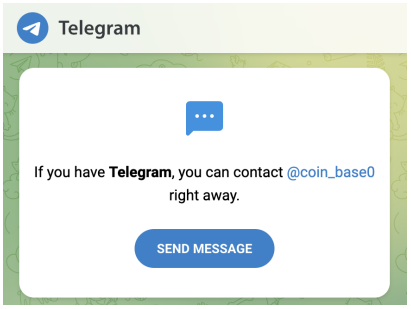
Brand impersonation, also commonly referred to as *brand spoofing*, is a well-known problem in which cybercriminals perform social engineering tricks to represent themselves as official employees or brand owners [13, 51]. These impersonations are not limited to traditional email-based phishing attacks, but are also widely found in other areas such as technical support, e-commerce, job offers, law and legal entities, and social media [19, 20, 71]. Brand imposters often engage with brand users, posing as reputable brands to obtain sensitive information or to sell counterfeit products. Once users trust these imposters, they become susceptible to attacks, potentially resulting in the theft of financial and personal information.

As the number of social media users continues to increase, it is projected that there will be approximately 5.04 billion social media users worldwide in 2024 [16, 22]. Consequently, brands are using their social media profiles to handle customer support and engagement [28, 36]. Many companies use social media platforms such as X [7, 14, 23, 52] and Instagram [4, 24, 31] to actively communicate and connect with their customers through both private messaging and public responses. The expectations of customers have changed with time in engagement with the brand. For example, Sproutsocial reported in 2021 [28] that 80% of the users expect companies to interact through social media profiles. On the other hand, this has created imposters a perfect ground for imposters to launch social engineering-based attacks that target the users of these brands. Several brand impersonation examples are shown in Figure 1.

According to the Federal Trade Commission (FTC), there was an 85% increase in brand impersonation between October 2020 and September 2021, resulting in losses of $2 billion [26]. In 2022, imposter-based scams caused losses of $2.6 billion. Taking into account scams based on only social networks, losses of $2.7 billion were reported from January 2021 to June 2023, excluding unreported losses [37]. Regrettably, the number of fake social media profiles, which often


(a) Typosquatting attack targeting PayPal.


(b) Combosquatting attack targeting Apple.




(c) Fuzzysquatting attack targeting Coinbase.

Figure 1: Representative examples of brand impersonation attacks where fraudsters applied typosquatting, combosquatting, and fuzzysquatting to target PayPal, Apple, and Coinbase. We can observe that username squatting is a key factor in creating a profile impersonation. Moreover, by adding official brand logos (Figure 1a and Figure 1b), fraudsters try to establish a notion of trust.

include brand impersonators, is on the rise across various platforms [57, 61, 64, 65]. Efforts to combat these fraudulent accounts via various techniques, such as the introduction of verified account badges, have proven not only ineffective but also counterproductive. Imposters have exploited this system by obtaining verification badges, further deceiving users by appearing as official brand representatives [5, 6, 18, 59, 65]. Although such brand impersonation attacks are known, there is no systematic study that comprehensively analyzes their *modus operandi*, scale of abuse, and monetary impact.

In this work, we close this gap by collecting a dataset of brand impersonation attacks, analyzing account setup and user engagement processes, quantifying their impact using loss metrics, and proposing robust countermeasures. More specifically, we perform the first large-scale study of brand impersonators that target the top 10K Tranco domains [44] on four social media platforms (X, Instagram, Telegram, and YouTube). Through the brand's account name similarity search, we collected 1.3 million users and 33 million posts from all four social media platforms. We identified almost 350,000 usernames performing various squatting techniques to impersonate 2,625 popular brands across the four platforms. To understand the scam life cycle, we further analyzed the profile metadata and user engagement of these username-squatted profiles revealing key characteristics of account creation techniques and several emergent and traditional threats. Furthermore, we analyzed the effectiveness of detecting scam profiles on social media platforms and identified that more than 98% of the scam accounts were alive, ranging from account creation dates between 2007 and 2023. Our findings show that one-third of username squatting profiles display official logos and are common among technology-related brands that target the users of the brand with technical support issues. Finally, we recommend defense mechanisms to combat such scams.

In summary, our key contributions are as follows:

- We perform the first systematic and large-scale analysis of username squatting-based brand impersonating attacks across four social networks. Our method and implementation are scalable to finding popular and non-popular brands impersonating social media profiles.
- In our empirical measurement study, we uncover brand-impersonating abusive social media accounts that target the top 10K brands from nine web categories. We uncover active attacks that these abusive accounts are carrying out and detect new and old threats via username and engagement creation.
- We demonstrate that the impact of brand impersonation extends beyond the fraud landscape and poses security and privacy concerns for victims. In particular, we provide evidence that fraudsters obtain sensitive information from their victims, which is then monetized for identity theft or monetary gains.
- Finally, we provide recommendations and defense mechanisms to combat these fraudulent accounts to mitigate existing and emerging threats.

To foster research, we publish our implementation [73]. For data protection reasons, we will only share data on fraudsters with interested academics or entities upon request.

**Ethical Considerations and Disclosure.** We conducted a coordinated disclosure of the 100 most impersonated brands and reported the identified squatting accounts. This includes brands such as *Google*, *PayPal*, *Facebook*, *Netflix*, *Amazon*, and others. We also shared 5K email addresses and 22K *payme* links from engagement posts with *PayPal*. PayPal's internal analysis confirmed that at least 43% of the 16K reported data had already been restricted at the time of data sharing, thus further observing fraudulent activities on other accounts. Additionally, we received confirmation from Google and we were acting on it. Both Google and X showed interest in further collaboration to identify similar fraud. Overall, our work received positive recognition, validating the scam life cycle of these impersonating accounts.

## 2 Technical Background

In impersonation attacks, including those on social media platforms, scam profiles often employ the target brand's official logos and description texts without authorization. At first glance, this official content naturally imparts a notion of credibility to scam profiles. In addition, fraudsters opt for usernames closely resembling those of official accounts to further add legitimacy. This behavior largely reflects web-based typosquatting attacks, where fraudsters register misspelled names of popular domains to capitalize on the customers' typing errors (e.g., *amazn.com* or *paypsl.com*). However, username squatting in social media-based brand impersonation is more intricate than conventional web-based typosquatting, and it can be broadly categorized as typosquatting, combosquatting, and fuzzysquatting. In the following, we discuss each type of username squatting along with representative examples.

**Typosquatting.** Typosquatting is a well-known technique in which attackers set up a fake profile with minor changes to the username, in order to lure victims who make a typing error while searching for the target brand [3, 81]. For instance, a typosquatter can register the X profile https://twitter.com/paypal7 and intercept users who mistakenly type "7" after the intended profile name https://twitter.com/paypal. In Figure 1a, we show a typosquatting attack targeting PayPal.

**Combosquatting.** Combosquatting is a variant of typosquatting in which attackers concatenate two or more strings to encapsulate the target brand's name and provide some additional context. For instance, if "apple" is the username of an official profile, a scammer can create "apple_support_US" or "apple_helpdesk2024" to appear as an official Apple support account, which can trap unsuspecting Apple users requiring technical support. Combining official names with context nouns is called combosquatting [40], and in Figure 1b, we show a combosquatting profile that targets Apple.

**Fuzzysquatting.** In our preliminary evaluation of social media scam accounts, we identified usernames that could not be solely attributed to typosquatting or combosquatting. For instance, we found scammers targeting Amazon with usernames including amaz0n_h3lp_ds3k and a3mz0n@_supp0rt__dkz. Note that amaz0n_h3lp_d3sk contains a typosquatted brand name (amazon -> amaz0n) and a combosquatted context noun (help_desk -> h3lp_ds3k), with key letters (o and e) substituted with digits (0 and 3). We label these accounts as *fuzzysquatting*. To identify fuzzysquatting, we create two rules: (i) Initially, we apply similar techniques to identifying combosquatting to find a series of words (or segments), and (ii) For each of the words (or segments), we compare whether one of the words contains an official social media handle with a non-zero Damerau-Levenshtein distance of two or less. We choose fuzzysquatting techniques purely based on a manual analysis of impersonating accounts on exclusion to the typosquatted and combosquatted lists. We suspect fraudsters adopting such username-squatting approaches for two main reasons: (i) allowing a larger number of non-conflicting registers of a user handle in social media platforms and (ii) obfuscating and potentially thwarting mechanisms against social media in detecting the squatting-based logic. Previous works on squatting have explored typosquatting and combosquatting at great length [40, 63, 81]. However, this novel form of username squatting, particularly in the context of brand impersonation on social media platforms, has not been extensively investigated. In Figure 1c, we provide an example of a fuzzysquatting attack targeting Coinbase on Telegram.

## 3 Measurement Setup

To study brand impersonation attacks on social networks, we have developed an analysis pipeline, which we present in the following. Our data collection process involves (i) selecting popular brands that could be targeted with impersonation attacks, (ii) collecting accounts linked to those brands from four social media platforms, and (iii) obtaining a candidate set of scam accounts by removing potential false positives from our account population. We now elaborate on each step of the data collection process.

### 3.1 Brand Domain Identification

To identify brands that are frequently targeted by fraudsters, we examined the top 10K domains of the highest ranked sites listed on *Tranco* [44]. From these top 10K domains, we first automated the filtering process by excluding unreachable domains and querying the web category of each domain using an external API [41] that categorizes domains based on the web content [17]. Subsequently, we filtered out domains that were not associated with e-commerce or consumer-oriented categories. Our analysis revealed nine main web categories that could be lucrative targets for fraudsters. In Table 1, we present a breakdown of the nine web categories for the 2,847 brand domains collected after applying the filtering process.

Our motivation to focus on the top 10K brands was inspired by the 2023 Phishing Threats Report from APWG and Cloudflare [12, 25]. APWG's report provides an overview of phishing trends across various industry sectors (financial, gaming, crypto, e-commerce, payment, social media, etc.), while Cloudflare's data highlights the top 1,000 threats associated with brand impersonation. These sources indicated a preference among scammers for targeting popular e-commerce websites and consumer-oriented businesses due to their greater potential for financial gain. The curation of candidate domains for our study consists of the following two steps:

- **Automated Filtering.** For each domain, we initially performed a check whether a given domain is alive. We used the Python GET request call to ensure that the response received was alive. For each of the live domains,

Table 1: Web categorization results for the popular brands considered in our work. In this table, we present the total number of unique brand names that we study based on each web categorization. Note that most brands were categorized as Business & Industrial.

| Web Categories | Unique Brand |
|---|---|
| Online Communities | 65 |
| Auto & Vehicles | 15 |
| Travel | 75 |
| Finance | 163 |
| Shopping | 118 |
| Arts & Entertainment | 601 |
| Internet & Telecom | 340 |
| Computers & Electronics | 656 |
| Business & Industrial | 814 |
| Total | 2,847 |

we further queried to obtain web content categories using a third-party API service [41] which categorizes domains, based on the Interactive Advertising Bureau (IAB) classification [17]. From this, we obtained 27 different web categories of these domains such as adult, entertainment, auto and vehicles, finance, games, food and drinks, hobbies, news, online communities, etc.

- **Manual Filtering.** For each of the 27 categories collected from the automated filtering, we manually compare them against the phishing trend report from APWG and Cloudflare [12, 25] to identify whether those categories represented the industry sector belonging to e-commerce websites and consumer-oriented businesses that are targeted by the fraudsters. This resulted in 9 such web categories as listed in Table 1.

## 3.2 Social Media Accounts Collection

To collect social media profiles associated with each brand, we selected four social media platforms: X, Instagram, Telegram, and YouTube. Using API services [10, 11, 54, 55, 78], we collected data related to social media profiles (e.g., name, description, profile picture, date of creation, account status, etc.) between May 2023 and October 2023. We then combined the brand domain's second-level domain (2LD) name with eight popular keywords, namely *rewards*, *recover*, *hack*, *support*,*help*, *assist*, *contact*, and *team* to create a search query for account collection. For instance, given a domain name *paypal.com*, our search query contained the following keywords: *paypal*, *paypal hack*, *paypal support*, *paypal help*, *paypal assist*, *paypal contact*, *paypal team*, *paypal recover*, and *paypal contact*. We ran queries on each platform to collect accounts as well as their profile metadata, including profile pictures, posts, interactions, etc. In Table 2, we provide an overview of all accounts collected from each platform, along with the number of associated brands. In total, we collected 1,310,962 accounts from the four platforms.

Table 2: Overview of the raw dataset obtained by performing search queries across four social media platforms. We can observe that X hosts the largest number of accounts.

| Platform | Unique Brand | Number of Accounts |
|---|---|---|
| X | 2,628 | 1,206,250 |
| Instagram | 1,717 | 13,545 |
| Telegram | 2,847 | 63,187 |
| YouTube | 1,784 | 27,980 |
| Total (Distinct) | 2,847 | 1,310,962 |

In a cursory view, our data invariably reveal a large number of potential fraudulent accounts. For example, a simple estimate suggests that on average, there are about 460 accounts associated with each brand across all platforms (1,310,962 accounts / 2,847 brands). It is logical to assume that a brand will not own such a large number of social media accounts, as it would create confusion among customers. On the other hand, the numbers suggest that there must be scam accounts that impersonate legitimate brands to deliberately create confusion and lure customers. In Appendix A, we provide further details on the brand social media account collection process and the rationale of keywords search.

## 3.3 Data Filtration

Having gathered roughly 1.3 million accounts, we conducted a series of filtering experiments to clean up our dataset and eliminate potential false positives. In the following, we present our data filtration methodology.

**Official Accounts.** We conducted two experiments to collect official accounts of the 2,847 brands. In the first experiment, we crawled the web pages of these brands and parsed their content to extract accounts linked to available social media platforms. We collected a total of 8,527 social media handles from the three platforms that are related to the official brands X(3,537), Instagram (2,429), and YouTube (2,561). As for Telegram, it is also worth mentioning that prominent brands do not use Telegram as their preferred communication channel. However, Telegram is a popular communication medium among fraudsters, and we expected to find scam accounts shared on Telegram, as previously shown in Figure 1c. In the second experiment, we used an external third-party API service [41] that monitors the official social media pages of popular brands. During our account filtering process, we observed that not all brands maintain official accounts on all four social media platforms. As a result, if an official account, e g., from X, was observed in our dataset, we removed it during the filtering process while keeping all other accounts.

**Verified Accounts.** In the next filtering step, we remove all verified accounts from our dataset. Studies suggest that subscription-based account verification models have enabled fraudsters to easily obtain verified accounts [6, 18, 59]. However, in the absence of any empirical evidence that quantifies the volume of verified scam accounts, we adopted a conservative approach by implicitly assuming verified accounts as legitimate accounts and removing them from our dataset. In total, we excluded 8,415 verified accounts in this step.

**Brand Subsidiary Accounts.** It is a common observation that while searching for a particular brand (e.g., *PayPal*), social media platforms also return associated brands that could be the subsidiaries of the main brand (e.g., *Venmo* in this example). Such results are typically returned due to account associations inferred from user interests. Therefore, we anticipated that our search results could include such accounts, which may introduce false positives in our dataset.

To remove such accounts, we collected the top 1 million Tranco domains and extracted second-level domains as brand names. Our intuition was that for a popular brand in Tranco's top 10K domains, its subsidiary brand would at least be among the Tranco 1M domains. When observing an overlap between our dataset and the second-level domains of the top 1M domains, we removed 16,492 entries from our dataset.

**Low-impact Accounts.** In the final step of account filtering, we filtered the remaining dataset to identify accounts whose username handle creation does not align with the criteria of any of the squatting techniques (i.e., typosquatting, combosquatting, and fuzzysquatting) by cross-referencing them with the username of the official brand handle. These identified accounts are labeled as low-impact accounts. While conducting our manual inspection, we noticed accounts whose names are similar to our search query but whose usernames are not similar to the target brand. For example, if an account displayed the name *amazon tech support* but had an innocuous username (e.g., lead_vocalist701), this indicated that the account owner was not engaged in username squatting, which we consider a fundamental aspect of identifying brand impersonation attacks. Such accounts naturally have limited deception potential and are less likely to trap victims, so they are categorized as low-impact accounts. Using this heuristic, we removed 927,767 low-impact accounts.

Starting with 1,310,962 accounts, we removed 8,527 (official accounts), 8,415 (verified accounts), 16,492 (brand subsidiary), 927,767 (low-impact accounts), and 350 (API response as invalid accounts upon metadata query) entries in the different filtering steps. This narrowed down our analysis to a candidate set of 349,411 accounts in total. We acknowledge that our data filtration is conservative and we might have overlooked accounts that could be involved in brand impersonation or other types of scams. However, we wanted to limit our analysis to scammers that are very similar to legitimate brands and have a higher likelihood of entrapping users.

Table 3: Summary of username squatting accounts targeting the brands. Note that most usernames were involved in combosquatting, targeting 2625 unique brands.

| Squatting | Squatted Accounts | Targeted Brands |
|---|---|---|
| TypoSquatted | 8,473 | 1,572 |
| ComboSquatted | 342,892 | 2,625 |
| FuzzySquatted | 4,523 | 1,035 |
| Total (Distinct) | 349,411 | 2,847 |

## 4 Account Setup Analysis

Based on our dataset, we now analyze the key characteristics of our candidate accounts, including username squatting trends and unauthorized usage of official brand content. As mentioned in Section 2, username squatting and brand content analysis provide key insights into the onboarding strategies applied by fraudsters. In addition, we study the distribution of targeted brands and their platform-wise distribution across social media.

### 4.1 Username Squatting

In the first step, we analyze the number of accounts that performed typosquatting, combosquatting, and fuzzysquatting in their usernames.

#### 4.1.1 Detection Methodology

For typosquatting detection, we measure the *edit distance* using the *Damerau-Levenshtein distance* of one, which measures the distance between two strings by counting the minimum number of deletions, insertions, and substitutions of characters required to transform one string into the other [3, 79]. We applied the Damerau-Levenshtein distance of one to check if a username had a missing character, character permutation, substitution, or duplication typo. For combosquatting detection, we compare the account usernames of the official brand and the impersonating brand using the *Word Ninja* [8] library, which uncouples strings containing multiple words. We applied *Word Ninja* on the account usernames to analyze the appearance of a target brand in the resulting text (i.e., amazon in amazonhelpdesk90). To identify fuzzysquatting, we first replicated the combosquatting detection method to split the strings, followed by the typosquatting detection method on the brand name. If the output contained a typosquatting brand name and a context noun, we labeled it as a fuzzysquatting username.

#### 4.1.2 Results and Key Findings

Our analysis revealed that fraudsters apply all three forms of username squatting techniques, with combosquat-

ting being the most prevalent method accounting for 98.13% (342,892/349,411) of the accounts targeting 92.20% (2,625/2,847) brands. Typosquatting was the second most prevalent technique, and it was observed across 2.42% (8,473/349,411) accounts that targeted 55.21% (1,572/2,847) brands. Finally, fuzzysquatting was detected across 1.29% (4523/349411) accounts that targeted 36.35% (1,035/2,847) brands. In Table 3, we provide a breakdown of username squatting types along with the number of targeted brands. In the following, we report our key findings about each username squatting technique.

**Typosquatting Techniques.** In typosquatting, our key observation was exploiting unique username creation rules of social media platforms to amplify scam deception. The creation of typosquatting in the handle depends on the length and allowed characters as permitted by each social media platform [33, 34, 76, 85]. Our study revealed that among impersonating accounts using typosquatting, 87.65% (7427/8473) added a single character as a prefix or suffix, while the remaining 12.34% (1046/8473) used omission or replacement.

**Combosquatting Techniques.** In combosquatting, we observed four popular trends in scam profiles namely: (1) random digits suffixed to the brand name (e.g., paypal12233243), (2) multiple words concatenated with the brand name (e.g., paypalhelpdesknow), (3) alphanumeric characters added to the brand name (e.g., paypaldsk3434), and (4) combinations of the aforementioned techniques along with characters and identifiers allowed by the respective platform. In Figure 2, we present the distribution of each pattern in our dataset of combosquatting accounts. We also noticed affinities in the usage of English language words. For instance, across all prominent brands, the terms *help*, *desk*, *support*, and *deal* were most commonly used with the brand name. In Figure 3, we provide a breakdown of the top 25 English keywords used by the combosquatting accounts.

**Fuzzysquatting Techniques.** Although fuzzysquatting accounted for a small percentage of our dataset, it still revealed some interesting insights. Notably, 973 fuzzysquatting usernames were derived from combosquatting usernames, which we detected by reverse substituting digits with letters.

## 4.2 Profile Image Analysis

As discussed in Section 2, we consider username squatting as the baseline attribute of impersonating profiles. However, a scam profile may not be convincingly deceptive if—in addition to username squatting—it does not use official logos of the target brand. In this section, we analyze how scam profiles attempt to project an illusion of credibility by using the official logos of the target brands in their profile pictures. We believe that analyzing such profiles is pertinent for two main reasons: (i) Highly deceptive profiles will invariably trap a large number of unsuspecting customers, thereby posing a threat to legitimate brands and their customers, and (ii) By analyzing affinities among such profiles, we can propose account setup controls for social media platforms, which could prevent the scam account creation at the sign-up level.

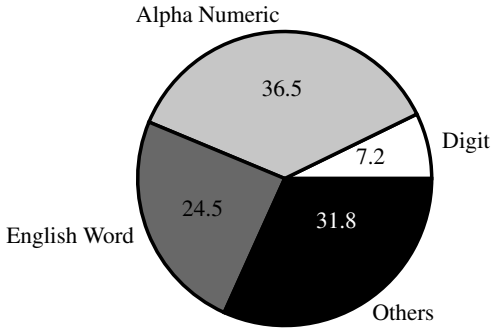



Figure 2: Distribution of combosquatting segments in usernames. Among five combosquatting segments, Alpha Numeric segments accounted for the most combosquatted usernames.

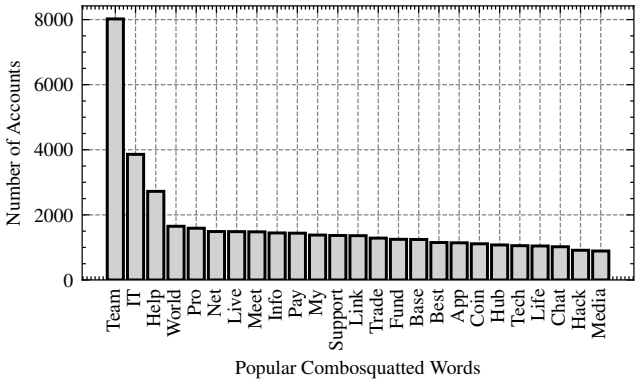



Figure 3: Overview of the top 25 keywords that combosquatters used to create user handles. Among these, the *Team* keyword was found most frequently.

### 4.2.1 Detection Methodology

In our analysis of profile images, we excluded pictures of squatted handles that use the default logo of the respective social media platform. This resulted in 219,465 squatted handle profile pictures belonging to the username creation of 2,484 brands. We employ the pre-trained visual model CLIP [66] [1] for feature extraction. We resized each profile picture to a resolution of $224 \times 224$ pixels, extracted the corresponding CLIP token embedding, and used cosine similarity to measure

[1] Specifically, we used a CLIP-ViT-B-16 model: https://huggingface.co/laion/CLIP-ViT-B-16-laion2B-s34B-b88K

their similarity to all CLIP token embeddings of the original logos [2, 70]. Subsequently, we sorted the obtained similarity scores for each picture and retained the most similar original logos. To enhance our pipeline's quality and reduce false positives, we filtered profile pictures where the affinity with the most similar logos scored below 80%.

#### 4.2.2 Results and Key Findings

After conducting the profile picture analysis, we identified 9,473 accounts that impersonated 1,701 brands using their official brand logos as display pictures. In Table 4 we provide a breakdown of image impersonation attacks performed by the scam account. Overall across all four social media platforms, we found 59.74% (1701/2847) of the total brands impersonated via profile image creation by 9,473 username squatting handles. On Instagram, among the 1717 brands we studied, 483 brand profiles were found to be impersonated by 1065 squatted handles, resulting in the highest brand impersonation (28.13%). Among these 9,473 impersonated profile images from all four social media platforms, X was found to be the highest at 45.02% (4265/9473), and Telegram was second most impersonated at 43.33% (4105/9473). The image cluster size of impersonating accounts was found to be an overall median size of 5 and as high as 308 squatted handles in a cluster. Among these image clusters, X (308) and Telegram (228) were found to contain a higher number of image cluster sizes with a median of 9 and 6, respectively. In the following, we provide a platform-specific analysis to show the popular brands being targeted on each platform.

**X.** On X, we discovered 488 brands that were impersonated by 4,265 scam accounts. Among them, the top 10 brands were *Apple, Microsoft, Samsung, Binance, Netflix, AT&T, Slides, Salesforce, Accenture, and Walmart*, and they were targeted by 38.9% of the scam accounts. These results are unsurprising given that the top 10 brands are indeed high-profile brands with a larger customer base.

**Instagram.** On Instagram, we found 483 brands in total that were targeted by 1,065 scam accounts. The top 10 most frequently targeted brands were *Sony, Microsoft, DHL, Samsung, Unilever, TechnoMobile, Orange, Vivo, BSIGroup, and Garmin.*, and they were targeted by 11.9% of the scam accounts. A surprising finding in our results is the small overlap (2/10) between the top 10 brands on X and Instagram. The account following of these brands may vary across each platform, which eventually determines the targeting patterns of fraudsters.

**Telegram.** On Telegram, we observed 698 brands being targeted by 4,105 unique scam accounts. The top 10 most targeted brands were *BSIGroup, Netflix, Telegram, Instagram, Aliexpress, Facebook, Spotify, Paytm, Android, and Binance*, and they were targeted by 26.2% of the scam accounts. Similar to Instagram, we observed a small overlap in the top 10 targeted brands on Telegram compared to X and Instagram. In fact, 3 out of the top 10 brands on Telegram were common with other platforms namely Instagram, Facebook, and Telegram. This pattern complements our earlier statement that Telegram is not a preferred communication channel for the top brands. Telegram users (including scammers) might be aware of it and instead use Telegram to target popular non-ecommerce brands such as social media platforms.

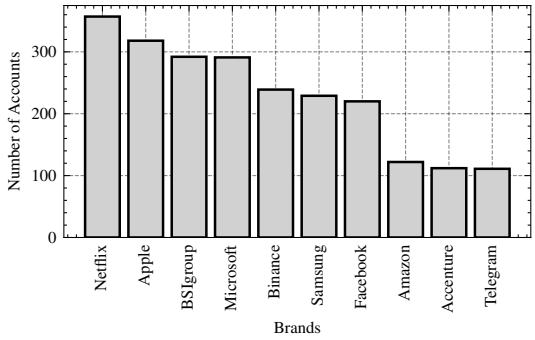



Figure 4: Top 10 most commonly targeted brands across all social media platforms. Technology-related brands such as Netflix, Apple, and Microsoft appear to have a high-risk exposure with over 200 scam profiles targeting each brand.

**YouTube.** Finally, on YouTube, we observed a small population of 32 brands being targeted by only 38 scam accounts. The top 10 targeted brands were *BSIGroup, Freshhooks, Sonicwall, PeacockTv, UPS, Tucows, USAA, Opensooq, Spotify, and Ushmm*. Compared to all other social media platforms, YouTube appears to be the least popular platform among fraudsters, and a plausible reason could be the limited user-to-user engagement on YouTube compared to other platforms. For instance, on X, users can easily search for other users, track their activities, and interact with them. However, YouTube users do not have a direct method to identify other platform users and most user-to-user interactions occur in video comments. Moreover, X and Instagram also provide direct messaging or private messaging options, which can be used by scammers to discretely communicate with their victims after luring them. YouTube does not offer such functionality to its users. Therefore, due to the limited target audience and lack of private communication options, scammers might not prefer YouTube as a medium for brand impersonation attacks. In the supplementary material [73], we provide further details on image clustering limitations and metrics.

#### 4.2.3 Holistic Brand Risk Exposure

Taking into account the most commonly targeted brands across all platforms, we now consolidate our findings to present the overall risk exposure of the popular brands. For that purpose, we collect the top 10 brands based on the number of scam accounts and present our results in Figure 4. We

Table 4: Summary of image impersonation by squatted handles of social media profiles mimicking various official brands. In this table, we provide several search accounts that impersonated the official brands from each social media platform we studied.

| Platform | Impersonated Brands | Scam Accounts | Scam Clstr. Median | Scam Clst. Max | Overall Brand Target % |
|---|---|---|---|---|---|
| X | 488 | 4,265 | 9 | 308 | 18.56 |
| Instagram | 483 | 1,065 | 3 | 17 | 28.13 |
| Telegram | 698 | 4,105 | 6 | 228 | 24.51 |
| YouTube | 32 | 38 | 1 | 4 | 1.79 |
| Total | 1,701 | 9,473 | 5 | 308 | 59.74 |

can observe that Netflix is the most commonly targeted brand across all social media platforms, followed by Apple and BSIGroup. Our results strongly complement a recent report published by a software solution company called Egress [82], and they have observed a 78% increase in brand impersonation attacks on Netflix. They also observed that scammers targeting Netflix often use typosquatted user names along with context nouns such as "Help Desk." On one hand, similarities in our results with those reported in [82] acknowledge the correctness of our methodical approach in discovering scam accounts. On the other hand, it also raises concerns about the growing risks of brand impersonation attacks which mandate remedial actions to be taken by social media platforms.

**Key Takeaways.** Our analysis of account setups leads us to several important findings. X is the primary platform for brand impersonation attacks, with fraudsters frequently using combosquatting in their usernames. Roughly a third of these deceptive profiles also use official logos to appear more legitimate. While a variety of brands are impersonated on social media, technology brands are particularly vulnerable, often targeted because their users seek technical support.

## 5 User Engagement Analysis

To effectively understand these scams, it is crucial to thoroughly examine scam profiles, understand how they engage with users, learn about their victim demographics, and analyze the fake services they offer. In this section, we delve into the details of scam profiles to provide a comprehensive view of the entire scam lifecycle.

### 5.1 Profile Metadata Analysis

We start by analyzing scam profile metadata, including name, geolocation, account type, languages, external links, account creation timelines, and public metrics.

**Display Name.** We applied text similarity analysis to study patterns and affinities in the display names of scam profiles. We found notable patterns in display names, including the usage of (i) the brand name, (ii) the brand username as display name, (iii) the brand name followed by context nouns such as *help*, *support*, *deal*, etc., (iv) multiple brand names such as Netflix, Facebook Help, and (v) unrelated keywords including human names and job titles such as *Shopify Expert & Ecomm Strategist* and *Amazon Products Guru.*

**Geolocation.** Geolocations are voluntarily disclosed by social media profiles. Typically, official brands tend to put head office locations on their profiles. Since top brands in our dataset have their head offices in the US, unsurprisingly, we found the highest number of valid geolocations in scam profiles set to the US (56,348). Other prominent geolocations were the UK (8,925), Japan (5,456,), India (5,134), Nigeria (4,231), Indonesia (2,321), Canada (2,134), Australia (1,993), and France (1,243). Overall, we found 51,919 distinct geolocations mentioned on the scam profiles, with 50,549 invalid locations such as the north pole, gaming, blackhole, etc.

**Languages.** Account language usage is a significant component of our analysis, as it reveals specifics of the user segment being targeted. For instance, if an account language is set to "Germany, English", it would likely mean that the scammer is targeting users from the region of Germany who can communicate in English. Diversity in language usage can shed light on a wide range of user segments being targeted. Our dataset showed that 38.9% (135,820/349,411) of the accounts do not contain language settings. Among 61.11% (213,591/349,411) of the accounts where languages were found, we identified 273 distinct language settings targeting 56 language-speaking users from 66 different regions. Among the scammers who set languages,12.38% (26,447/213,591) specified targeted region but missing speaking language users whereas the remaining 87.61% (187,144/213,591) specified region and language. Overall, the top 5 regions are India (7,623), Indonesia (6,947), Iran (6,400), Russia (5,473), and Uzbekistan (3,127) and the top 5 languages are Russian (4,367), Farsi (2,968), English (2,076), Arabic (1,620), and Spanish (1,066).

**External Link.** The description section of the profile metadata allows profiles to embed URLs which often point to a website outside of the platform. Among the four platforms in our study, X creates tiny links (i.e., *https://t.co/****) upon adding a URL. Tiny links are always unique, even if they re-

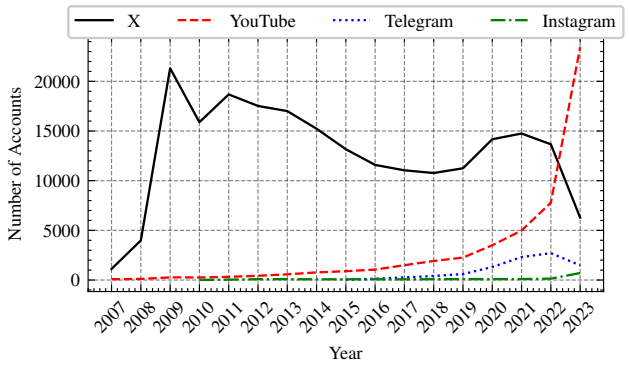



Figure 5: Yearly scam account creation volume for all four platforms. Note that X has the highest account volume with the maximum number of accounts reported in 2009. All the other platforms have been recently targeted by fraudsters, with YouTube showing exponential account growth.

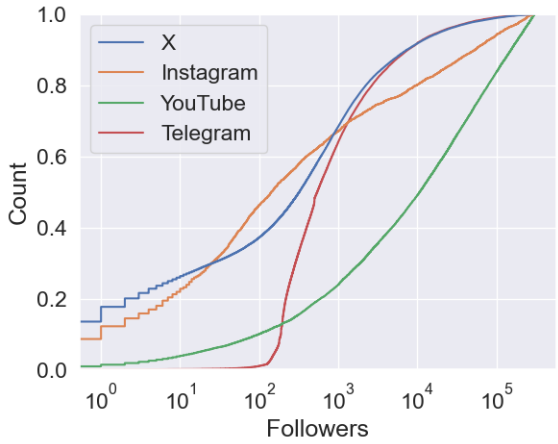



Figure 6: Followers count: This graph displays the number of followers on each social media platform that we studied.

solve to the same website. We collected 101,633 URLs from scam profiles and analyzed their security posture. We found 23,606 URLs that returned the 404 response code, indicating that the domains were taken down. For the remaining URLs, we performed extensive crawling to collect website metadata and other URLs mentioned on the websites. In total, we obtained 1,519 unique URLs through crawling, confirming that tiny URLs were indeed mapping onto the same set of websites. We then evaluated the security posture of 1,519 URLs using the VirusTotal API and found that 84 URLs were labeled as malicious, while others were labeled benign. A manual inspection of benign URLs revealed that they were linked to famous brands, indicating that scammers put legitimate URLs in their descriptions to add legitimacy to their profiles.

**Account Creation Age.** Account creation date is an important feature that sheds light on some of the key attributes of scam accounts. For example, if scam accounts have been operating for several years, this shows that their hosting platform needs to apply better controls to detect and restrict them. Moreover, if there are anomalous trends in the account sign-up timeline, this might indicate coordinated scam campaigns being launched to target brands. Our dataset showed that scam accounts were created between 2007 and 2023 with a median value of 16,158 per year. In Figure 5, we show the scam account creation timeline across all four platforms. In terms of year-wise distribution, the highest number of accounts created on the four different platforms represented by (year, total count) are X (2009, 21,316), Instagram (2023, 705), Telegram (2021, 23,113), and YouTube (2019, 23,443). The account creation trend on X was distinctly different from other platforms since the account volume varied over time. In contrast, other platforms showed a continuous growth of scam accounts over the years, with YouTube's count growing exponentially since 2022. From our results, we conclude that X has been historically a popular choice for scam account creation. However, the trend has recently been diminishing on X and growing across the other three platforms.

**Followers and Posts.** In the next step of our study, we analyzed the popularity and engagement tendency of scam profiles. A strawman method to measure an account's popularity is by observing its followers and following counts. Our analysis revealed that most scam accounts had a mundane outlook with few followers. In Figure 6, we present the number of followers across different social media platforms. We found that 70% of profiles on X, Instagram, and Telegram have fewer than 1,000 followers, while on YouTube, over 75% of accounts have more than 1,000 followers. The median follower counts for X, Instagram, YouTube, and Telegram were 321, 142, 10,700, and 547, respectively. Apart from YouTube, the follower count for all other platforms is considerably lower than the expected follower count of a popular brand account.

Similarly, account engagement tendency can be broadly analyzed by counting the number of posts generated by an account. We observed that more than 95% of the accounts had generated a post with a median post count of 472, 30, 33, and 394 across X, Instagram, YouTube, and Telegram, respectively. In Figure 7, we provide a CDF for all posts generated by scam accounts. Our CDF shows that many accounts ($>$80%) generated fewer than 20,000 posts. Hence, it can be concluded that $\approx$20% of the scam accounts are actively engaged on the platforms through posts.

## 5.2 Qualitative Evaluation: Account Age, Account Creation, and Long Tail Followers

Based on the profile data analysis of scammers on the date of creation, we perform further analysis to understand the scammers' account activity and ensure that the data does not contain false positives. Below we discuss three qualitative

studies we performed.

**Random Account Analysis from Last Decade.** We randomly picked 50 X accounts from 2007 to 2017 and inspected these accounts in the web UI. Overall, 13/50 of these accounts were found to be blocked on the X platform for some form of policy violation. For the remaining 37 accounts, we found that 6 accounts were either deactivated or had their user handle changed. For the 31 accounts, we observed that 9 did not contain any posts and had a similar logo to the squatted brands, while 22 accounts contained posts and interactions with users related to the squatted brands. All of these interacting accounts showed signs of maliciousness corresponding to the attack categories we will discuss in Section 5.3.2. However, the tweets interaction did not contain any interaction before 2023.

**Buy and Sale of Social Media Profile.** We observed that social media profiles on platforms such as X, Instagram, YouTube, and others can be bought and sold on multiple open marketplaces in public trading [1, 53, 86]. Scammers are also likely to create mass social media profiles and sell these accounts, which are later used for various types of scams.

**Long Tail Followers.** Additionally, we conducted a qualitative analysis of followers of squatted profiles from three popular brands: Netflix, Amazon, and Binance. From each of these brands, we selected 10 profiles from the X platform with over 300 followers. Below we provide details on our analysis of these profiles.

- **Followers with no posts.** 14/30 profiles had no posts on their timelines and did not interact with public posts, despite having over 300 followers. These accounts appeared to be suspicious, often protected, missing posts, or shared followers.
- **Followers with posts.** 16/30 profiles had between 3 and 117 posts. The posts from Amazon-related profiles often targeted job seekers with offers for work-from-home or remote jobs. Netflix-related posts promoted creating unlimited Netflix accounts, direct messaging for further information, or watching free movies without paying. Binance-related posts focused on growth, pump-and-dump schemes, and investment opportunities.

## 5.3 Scam Post Analysis

The last component of the scam lifecycle analysis is the study of scam posts to understand how fraudsters trap victims through fake incentives or offers. For this purpose, we collected all available posts from scam profiles and applied topic modeling techniques to group them in clusters. We collected 33,768,759 posts in total [X (33,588,977), Instagram (19237), Telegram (133,399), and YouTube (27,146)] and applied the following technique to extract prominent fake incentives offered in them.

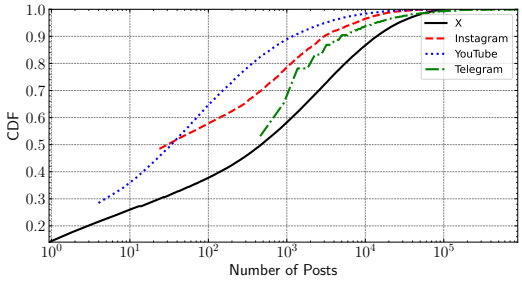



Figure 7: Graph of posts interaction on the social media platform. The graph provides a breakdown of the squatting profile posts count from four different social media platforms.

### 5.3.1 Clustering Methodology

First, we picked a maximum of 15 most recent posts from each account, acknowledging that not all accounts necessarily had 15 or more posts available. Consequently, the total number of posts analyzed in the clustering process amounted to 322,327 after excluding posts that were not in English. For language identification and filtering, we use the CLD2 library [15]. We then vectorized the posts using the *all-mpnet-base-v2* sentence transformer model [67]. Subsequently, we processed the posts using the BERTopic library [30] to remove redundant information, such as stop words. Finally, we combined UMAP [50] and HDBSCAN [49] for clustering, followed by the KeyBERT [29] model to refine topic representations within each cluster. We selected the top 100 clusters in our results, from which we extracted ten prominent scam categories that were prevalent in our results. Further details on the text processing and clustering are reported in the supplementary material [73].

### 5.3.2 Results and Key Findings

We now discuss the top ten most prominent scam offers observed in our clustering results. We performed manual analysis on each of these top ten clusters arranged by size and provided details on brand-impersonated attacks. The clusters that do not contain brand impersonation are excluded from our study.

**Free Giveaways.** The most common scam offer in our results was fake free giveaways, such as cryptocurrencies, free streaming service sign-ups, purchase deals, etc. A total of 3,735 scam accounts were involved in giveaways, frequently targeting brands including Amazon, Netflix, Binance, and Coinbase.

**Fund Raiser.** Fund raise scams include a request to support emerging businesses such as future health care, wealth summit, startups, etc. We found 3,053 accounts involved in such fundraising scams, and they mostly targeted Tesla, Salvation Army, Ford, GoFund, and John Hopkins.

**Fake Technical Support.** Fake technical support includes scammers offering fake account recovery or device diagnostics in return for a high fee. We found 2,115 accounts offering fake technical support and targeting brands including Moodle, Cloudflare, WhatsApp, and Amazon.

**Third-Party Service Representative.** In third-party service scams, fraudsters offer technical support for app or software installations of popular brands. We found 1,979 scam accounts targeting mostly finance-driven brands such as HDFS bank, PayPal, and Paytm.

**Advertisement Booster.** In this type of scam typology, fraudsters pretend to be an official representative of brand advertisement segments such as Google Ads or Facebook Ads. They use such fake affiliations to offer fake boosting services for other users. We found a total of 1,597 accounts that were involved in this scam.

**Fake Recruitment.** Recruitment scams involve fake job offers at the target brand. For instance, we observed fraudsters offering fake jobs such as Amazon drivers, sales and marketing leadership roles, and remote work. We found a total of 1,442 scammers in this category, and they mainly targeted Amazon, Sears, LinkedIn, and SalesForce.

**Fake Discounts and Giftcards.** In this scam type, fraudsters offer brand discounts on popular events or gift cards for VIP membership, store anniversaries etc. We found 1,367 scam profiles involved in fake discounts and gift card scams, primarily targeting Walmart, Amazon, Shopify, and GoDaddy.

**Brand Union.** In the Brand Union scam, fraudsters pose as the representative of a brand workforce union that supports various causes such as allowance, compensations, retirement perks, etc. They typically use noble causes to raise funds for their fake union. We identified 1,186 accounts linked to fake brand union scams.

**Fan Page.** In this type of scam, fraudsters claim to support fake events linked to a given brand. In general, their motivation is to either increase the profile followers or request payments for fake event organizations. We found 946 accounts in this scam type.

**Sales Training and Leadership.** In sales training and leadership scams, fraudsters pretend to be official brand representatives and offer paid training programs for leadership and growth opportunities. Unsuspecting users implicitly assume that receiving such training from popular brand officials will help them in their professional careers. However, as common with other scams, fraudsters receive training fees without providing a service in return. We found 931 accounts with posts indicating affiliations with such scams.

**Key Takeaways.** By combining account setup and user engagement analysis, we map out the end-to-end scam lifecycle and derive the following key insights: (i) Brand impersonation is a multi-step process in which fraudsters set up fake profiles using tactics such as username squatting and displaying official brand logos, (ii) Fraudsters engage with customers through posts and offer them phony incentives as bait, (iii) Scammers tend to target a wide range of victims, as indicated by the diverse geographical locations and language usage, and (iv) Brand impersonation scams have been ongoing for over a decade and are quickly proliferating from X to other social media platforms.

Table 5: Overview of blocked squatted profiles - The table provides the summary of the effectiveness of blocking username squatting from social media based on squatting techniques.

| Squatting | Blocked Accounts | Blocked % |
|---|---|---|
| Typosquatting | 65 | 0.76 |
| Combosquatting | 2,367 | 0.69 |
| FuzzySquatting | 57 | 1.26 |
| Total | 2,489 | 0.71 |

## 6 Scam Profiles Detection Efficacy

In this experiment, we analyze the efficacy of social media platforms in detecting and blocking scam profiles. We perform two experiments to analyze the control gaps that exist in detecting scam profiles and later apply our insights to propose effective countermeasures.

### 6.1 Control Gap Detection through Dataset

In the first experiment, we collected all scam profiles in our dataset (Section 4) and revisited their profiles after completing our analysis. We hypothesized that towards the end of our study (spanning seven months), it is possible that social media platforms have detected scam profiles and blocked them for violating the platform's terms and services. Surprisingly, only 2,489 (0.71%) of the scam profiles were blocked when we revisited them. In Table 5, we present the number of blocked accounts along with the type of username squatting they performed. Our results clearly show a control gap in social media platforms, as indicated by the low blocking rate despite the time lapse in our data collection and experiment completion.

We also cross-referenced the blocked accounts with our brand labels to check the famous brands being protected by account blocking. On X, the top 5 brands were Binance, Netflix, AirBnB, PayPal, and Toyota, with the total number of blocked accounts being 80, 73, 69, 65, and 57, respectively. The top 5 brands on Telegram were Netflix, Box, Sony, Download, and Marvel, with the total number of blocked accounts being 71, 36, 35, 33, and 32, respectively. On Instagram, the top 5 brands were Online (28), Business (21), Mail (18), Site (17), and WordPress (16), and for YouTube, the top brands were Ask (9), About (6), AT&T (6), Money (6), and NTV (5). Note

Table 6: Experiment evaluation of proactively finding brand impersonation accounts targeting the brands.

| Web Cat. | Brand | Generated | NotFound | Suspended | Active |
|---|---|---|---|---|---|
| Arts & Ent. | Netflix | 100 | 63 | 9 | 28 |
| Auto & Veh. | Toyota | 100 | 51 | 7 | 42 |
| Bus. & Indus. | PayPal | 100 | 62 | 4 | 34 |
| Compt. & Electr. | Microsoft | 100 | 52 | 5 | 43 |
| Finance | Binance | 100 | 67 | 10 | 23 |
| Intr. & Telecom | Google | 100 | 47 | 2 | 51 |
| Online Comm. | Facebook | 100 | 43 | 6 | 51 |
| Shopping | Amazon | 100 | 44 | 2 | 54 |
| Travel | AirBnB | 100 | 67 | 3 | 30 |
| Total | - | 900 | 496 | 48 | 356 |

that there is a small overlap between the top brands being targeted (Section 4) and the top brands being protected through blocking. Ideally, brands in both categories should have been the same due to scam account volume and consistently similar account setup techniques. However, we only observed a slight overlap, likely indicating varying rules applied to evaluate scam profiles.

## 6.2 Control Gap Detection through Shadow Profile Setup

**Overview** In our second experiment, we took a methodical approach to create *shadow* scam profiles and use them to measure the detection accuracy of social media platforms. We applied our knowledge from username squatting to mimic the account setup process flows and validated the corresponding username's status on the social media platform. Our second experiment served two main purposes, namely (i) Confirmation of control gaps identified in the first experiment, and (ii) Exposure to the methods used by fraudsters to evade detection, which can be applied for countermeasures.

**Username Generation** For account generation rules, we selected X as our evaluation platform since X hosted the largest number of scam accounts in our dataset. We selected nine random popular brands from each web category as listed in Table 1. For each brand, we selected all its squatted usernames and identified common patterns in those usernames. For instance, if the target brand is *paypal* and its squatted usernames are (paypal_39, pay_pal3, h3lp_paypal), the squatting patterns include (1) usage of underscores followed by suffixed digits, and (2) prefixed context noun before the brand name. We collected squatting patterns in a given brand name (e.g., PayPal) and applied those patterns to generate new usernames for another brand (e.g., Netflix). For each target brand, we generated 100 usernames and a total of 900 usernames across all brands. We ensured that the new usernames were not duplicated in our previous dataset. As a result, we generated fresh usernames that could likely be usernames of scam profiles.

**Username Monitoring** After selecting 900 usernames, we then queried X to check if (1) the username was present on the platform, and (2) X had taken any actions on those usernames. Please note that we did not register those usernames on X. Our exercise was intended to create usernames of scam profiles and check if they were present on X at the time and if X had taken any actions to block them. Moreover, since the experiment was conducted long after our previous dataset collection, it would mean the accounts would be fresh and, therefore, not captured during our dataset collection period. After querying X, we observed that out of 900 total usernames, 404 (44%) were present out of which only 48 (5.3%) were suspended. The remaining 356 were active and operational. Moreover, 496 accounts were not found on X, and we consider them as potential scam profiles that could be registered in the future. In Table 6, we present the results of our second experiment.

**Key Takeaways** Our analysis confirms that social media platforms do not efficiently detect scam profiles as indicated by a small percentage (0.71%) of suspended accounts in seven months. Moreover, there are control gaps in the account onboarding process which are exploited by fraudsters to set up scam profiles. An interesting observation in our experiments is the brand-based prioritization of scam profile detection on social media platforms. As shown in Table 6, X scam detection rules have better coverage of the finance category than the shopping category. Although the root cause for this gap in coverage remains opaque to us, the observation alone provides meaningful insights for scam countermeasures.

# 7 Scam Validation and Countermeasures

After examining the end-to-end scam lifecycle and identifying control gaps on social media networks, we performed scam validation to confirm that brand impersonation attacks indeed lead to monetary losses. For this purpose, we decided to partner with an organization that is a brand impersonation victim in our dataset and has insights into the payment ecosystem. We partnered with PayPal and shared our dataset containing 5,980 email addresses and 11,606 paypal.me profile links.

## 7.1 Scam Validation Results

As per our agreement with PayPal, they only shared aggregate data and insights with us, without providing account-level details or payment details. Using our data as seed intelligence, PayPal identified more than 22,000 accounts that either matched emails or paypal.me profiles in our dataset or were closely linked to them through other attributes, including phone numbers. More than 43% of the accounts were already restricted at the time of data sharing, indicating that PayPal was already observing fraudulent activities on those accounts. At the time of writing this paper, they were investigating the remaining account population and determining their risk posture.

Given that PayPal confirmed fraudulent activities performed by the accounts we escalated, their feedback also led to some interesting conclusions. Among the 43% of accounts they had restricted, prominent fraud and risk indicators included (i) Failure of mandatory Know Your Customer (KYC), (ii) Use of stolen identities, (iii) Credit card fraud, (iv) Scripted activities, (v) Violation of payment limits. In sharing the data, we assumed that fraudulent accounts would only be associated with payment fraud reported by victims and that this would be a sufficient outcome to show the impact of our work. However, the feedback we have received indicates that fraudsters are not just pickpocketing benign users. In fact, they are collecting sensitive information, including victims' identities and credit card details, to engage in other fraudulent activities, clearly demonstrating the security and privacy risks associated with brand impersonation attacks. In other words: if the fraudsters were simply receiving payments from their victims, we might have assumed a notable but small-scale impact of brand impersonation attacks. However, with the evidence provided by PayPal suggesting that in addition to receiving payments, fraudsters are also harvesting credentials and monetizing them for other nefarious activities, we argue that brand impersonation attacks pose a significant security and privacy threat to the ecosystem. It is therefore pertinent to address the problem through collaborative efforts, and in the following section, we propose a few recommendations as countermeasures.

## 7.2 Recommendations for Countermeasures

Consolidating all our insights, we now propose countermeasures against brand impersonation attacks. Our recommendations involve methods and techniques that can be adopted by social media platforms and brands to collaboratively mitigate brand impersonation attacks. In the following, we elaborate on those methods.

**Brand-based Account Sign-up Rules.** We recommend social media platforms perform brand validation at the account registration step. When a new account signs up, the social media platform can enforce official email address verification along with the 2LD domain name mapping. For instance, an account signing up with *Amazon* username must be verified through *@amazon.com* email, which can also be checked in the 2LD domain. Moreover, additional safety checks (e.g., domain age) can be applied when the account applies for official verification. As a result, only official brands can set up accounts on social media and acquire verified profiles. Once official accounts are set up, username squatting models (Section 4.1) can be applied to block any subsequent account that attempts to impersonate a brand. Finally, for account blocking, we suggest unbiased detection models to be enforced across all top brands. As shown in Table 6, scam account detection rules vary across different web categories, which can be invariably learned by fraudsters to select target brands that are not well-protected by the social media platform. A uniform account-blocking approach can prevent such situations and reduce the number of scam accounts on social networks.

**Customer Education.** We suggest brands communicate with their customers to be aware of standard engagement protocols and inform users not to share personal information on social media. Additionally, they can share key characteristics of scam profiles (e.g., username squatting techniques) with their customers to help them distinguish between scam accounts and legitimate accounts.

**Auto-respond Adoption.** Recently, popular brands including *Coinbase*, *MetaMask*, and *Google* have started using the auto-respond feature to reply to customers that post any technical issue. If more brands start using the auto-respond feature and social media platforms also prioritize their responses, official brands can engage with customers before scammers. It will give them a head-start advantage in establishing reliable communication channels with their customers before they are intercepted by scammers.

**Active Monitoring and Reporting.** We recommend that brands actively monitor the use of their brand name, logos, and products, as well as the creation of usernames that may be squatting on their brand identity. This will prevent fraudsters from impersonating individuals or carrying out social engineering attacks against the brand's users. By proactively monitoring these activities, fraudulent profiles on social media can be identified and reported to the relevant platforms so that they can take appropriate action, such as deletion.

**Key Takeaways.** Our collaboration with PayPal confirms the impact of brand impersonation attacks on victims and payment platforms. It is logical to assume that as more brands appear in the market and build social media profiles, the scale of abuse will likely increase. Taking into account the emerging threat and its ramifications, we propose effective techniques that require collaborative efforts between social media platforms and popular brands. Our propositions involve leveraging the techniques used in our study to strengthen official profile verification methods, indiscriminately block scam profiles, and enhance customer education.

# 8 Related Work

To the best of our knowledge, our work is the first attempt to perform a large-scale systematic study of social media-based end-to-end analysis on brand impersonation attacks. Below we reference some of the prior studies and provide the novelty of our work toward building scam validation across multiple social media profiles performed by these brand impersonators.

**Early studies on squatting.** Typosquatting has been a topic of interest for over 15 years [3, 38, 40, 58, 68, 74, 75, 81]. The early study on typosquatting from Wang et. al [81] dates back to 2006 which showed that attackers exploit typing errors to lure victims to a counterfeit domain through which

they harvest user credentials. In 2013, Nikiforakis et. al [63] studied *bitsquatting*, in which attackers abuse random bit-related errors in the memory and redirect traffic to the counterfeit domains. The study analyzed bit-squatted domains among the top 500+ Alexa domains and further provided abuse categorization. The authors continued following trails of domain-based squatters to study several alternative versions of squatting-based attacks through elaborated independent studies. In 2014, their work on *soundsquatting* [62] explored squatting based on words that sound alike. In 2015, a longitudinal study of typosquatting domains studied HTML page content usage for monetization strategies and showed that there is little protection against the registration of such domains that target trademark owners [3]. In 2017, an empirical study of combosquatting domains showed that attackers register domains by combining trademarks with words [40]. The study also highlighted that the domains were used for phishing and trademark abuse.

**Email squatting.** In the domain of email squatting, Janos et. al [74] studied email typosquatting by registering 76 typosquatting domain names. Throughout a seven-month study, the authors observed receiving millions of emails containing sensitive personal information. They also studied 1,200+ typosquatting domains in the wild by sending honey emails and found that most email responses received were used for spam.

**Social media abuse and impersonation study.** Towards username manipulation in social media, Jain et. al [35] studied the behavior of 8.7 million X users and monitored their username-changing behavior. They monitored 10% of users who changed their usernames to study the root cause for their behavior. They concluded X users often change handles for space gain, followers gain, and username promotion. Similarly, Lepais et. al studied username squatting profiles on X impersonating celebrity profiles [45]. The work from [27] studied the impersonation accounts on X where malicious attackers copy the profiles of legitimate users to create fake accounts that are later used for the illegal promotion of content on X. All these studies do not provide holistic measurements that target brands across multiple social media platforms.

**Impersonation study on network.** Impersonation attacks are well-studied in areas of communication and network protocol. For example, Antonio et. al studied a bluetooth impersonation attack exploited via missing permission and authentication [9]. In [69], Rupprecht et. al studied impersonation attacks in 4G networks that are exposed via cross-layer communication. Similarly, Yilmaz et. al studied impersonation attacks in the wireless network via spoofing signaling [84].

**Squatting beyond domains.** Squatting-based attacks have evolved over time, and they can be observed across multiple platforms. Nowadays, attackers are also targeting (i) mobile apps by registering similar app names and identifiers [32], (ii) software packages by mimicking package name imports [60], (iii) container technology by registering similar container images [46], and (iv) commercial Internet-of-Things (IoT) devices such as speech recognition-based commands [42]. Although all these notable works focus on some form of social engineering attack, our work is unique as it explores a pertinent and understudied attack type involving brand impersonation on social media using various deception techniques.

The above prior studies show a gap in large-scale studies of brand impersonation attacks across multiple social platforms. Our work illustrates the prevalence of ongoing impersonation attacks occurring in the real world and confirms scam validation through fraudulent payment-related association with these fraudsters' accounts via username and engagement creation. With this study, we lay the foundation for future research to further investigate brand impersonation-based attacks on social media platforms.

## 9 Conclusion and Future Work

In our research, we delve into the underexplored realm of brand impersonation on social media, focusing on well-known brands and abuse in this context. We found that scammers create fake profiles targeting popular brands, using tactics like username squatting and unauthorized use of trademarks to appear legitimate. We performed the first large-scale measurement study to examine the scam lifecycle on four social media platforms, analyzing the activities of 1.3 million users and their public interactions. We uncovered about 350,000 profiles engaged in username squatting, employing three distinct methods. The impersonators launch a variety of attacks against both high-profile and less well-known brands worldwide. Additionally, our scrutiny of major payment services highlights the financial damage within the scam lifecycle. Our findings indicate that social media platforms are currently ineffective in safeguarding brands and users from such threats, with over 98% of these deceptive accounts remaining active and their numbers growing annually. In a practical experiment, we applied our own rules for detecting unexposed squatting accounts and found that half of these are still active, continuing to conduct impersonation attacks. Drawing from our research, we offer recommendations for social media platforms to combat these fraudulent activities.

**Acknowledgements** This work was funded by the German Federal Ministry of Education and Research (grant 16KIS1900 "UbiTrans"). Additionally, this work was partially supported by the European Union—NextGenerationEU (National Sustainable Mobility Center CN00000023, Italian Ministry of University and Research Decree n. 1033—17/06/2022, Spoke 10). Lastly, this work has been carried out while Dario Lazzaro was enrolled in the Italian National Doctorate on Artificial Intelligence run by Sapienza University of Rome in collaboration with University of Genoa.

## References


[1] ACCS. Quick and secure social media marketplaces. https://accs-market.com/twitter.

[2] Bhupendra Acharya, Muhammad Saad, Antonio Emanuele Cinà, Lea Schönherr, Hoang Dai Nguyen, Adam Oest, Phani Vadrevu, and Thorsten Holz. Conning the crypto conman: End-to-end analysis of cryptocurrency-based technical support scams. In *IEEE Symposium on Security and Privacy (S&P)*, 2024.

[3] Pieter Agten, Wouter Joosen, Frank Piessens, and Nick Nikiforakis. Seven months' worth of mistakes: A longitudinal study of typosquatting abuse. In *Symposium on Network and Distributed System Security (NDSS)*, 2015.

[4] Airfrance instagram customer support. https://www.instagram.com/airfrance.

[5] Zeeshan Aleem. Twitter, facebook and instagram once had a decent tool to fight misinformation. it's gone. https://www.msnbc.com/opinion/msnbc-opinion/meta-facebook-instagram-verified-badges-twitter-rcna72058, Feb 24, 2023.

[6] Addressing executive & social media impersonation: Protecting leaders that lack an online presence. https://alluresecurity.com/addressing-executive-social-media-impersonation-protecting-leaders-that-lack-an-online-presence/, 2023.

[7] Amazon twitter customer support. https://twitter.com/AmazonHelp.

[8] Derek Anderson. Word ninja. https://github.com/keredson/wordninja, 2023.

[9] Daniele Antonioli, Nils Ole Tippenhauer, and Kasper Rasmussen. Bias: bluetooth impersonation attacks. In *IEEE symposium on security and privacy (S&P)*, 2020.

[10] Apify. Apify instagram scraper api. https://apify.com/apify/instagram-scraper, 2023.

[11] Apify. Youtube scraper. https://apify.com/streamers/youtube-scraper, 2023.

[12] APWG. Phishing activity trends, first quater 2023. https://docs.apwg.org/reports/apwg_trends_report_q1_2023.pdf, 2023.

[13] Brand impersonation. https://www.barracuda.com/support/glossary/brand-impersonation.

[14] Bank of america twitter customer support. https://twitter.com/BofA_Help.

[15] Greg Bowyer. CLD2-CFFI – Python (CFFI) Bindings for Compact Language Detector 2, 2016. https://github.com/GregBowyer/cld2-cffi.

[16] Dave Chaffey. Global social media statistics research summary 2024. https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/, Feb 01, 2024.

[17] Interactive Advertisement Bureau (IAB) Domain Classification. Github repo on IAB Domain Classification. https://github.com/InteractiveAdvertisingBureau/Taxonomies/blob/main/Content%20Taxonomies/Content%20Taxonomy%203.0.tsv, 2023.

[18] Corsearch. Why brand impersonation is increasing on twitter & how to combat it. https://corsearch.com/content-library/blog/why-brand-impersonation-is-increasing-on-twitter-how-to-combat-it/, Dec 11, 2022.

[19] Casey Crane. What is brand impersonation? a look at mass brand impersonation attacks. https://www.thesslstore.com/blog/what-is-brand-impersonation-a-look-at-mass-brand-impersonation-attacks/, Oct 20, 2022.

[20] DarkTrace. What is brand impersonation? https://darktrace.com/cyber-ai-glossary/brand-impersonation.

[21] Rachna Dhamija, J Doug Tygar, and Marti Hearst. Why phishing works. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 2006.

[22] Stacy Jo Dixon. Number of social media users worldwide from 2017 to 2027. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/, Aug 29, 2023.

[23] Ebay twitter customer support. https://twitter.com/AskeBay.

[24] Wellsfargo instagram customer support. https://www.instagram.com/wellsfargo.

[25] Cloud Flare. 2023 phishing threats report. https://www.cloudflare.com/lp/2023-phishing-report/.

[26] Ftc launches rulemaking to combat sharp spike in impersonation fraud. https://www.ftc.gov/news-events/news/press-releases/2021/12/ftc-launches-rulemaking-combat-sharp-spike-impersonation-fraud, 2021.

[27] Oana Goga, Giridhari Venkatadri, and Krishna P Gummadi. The doppelgänger bot attack: Exploring identity

impersonation in online social networks. In *Internet Measurement Conference (IMC)*, 2015.

[28] Ronnie Gomez. 8 ways customers interact and engage with your brand on social. https://sproutsocial.com/insights/social-media-interaction/, 2024.

[29] Maarten Grootendorst. Keybert: Minimal keyword extraction with bert. https://doi.org/10.5281/zenodo.4461265, 2020.

[30] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.

[31] H&M Instagram Customer Support. https://www.instagram.com/hm.

[32] Yangyu Hu, Haoyu Wang, Ren He, Li Li, Gareth Tyson, Ignacio Castro, Yao Guo, Lei Wu, and Guoai Xu. Mobile app squatting. In *Web Conference*, 2020.

[33] Twitter Inc. How long can usernames and names be? https://help.twitter.com/en/managing-your-account/.

[34] Instagram. Creating an account & username. https://help.instagram.com/182492381886913.

[35] Paridhi Jain and Ponnurangam Kumaraguru. @ i to@ me: An anatomy of username changing behavior on twitter. In *Data Science (CODS)*, 2014.

[36] Alita Joyce. Companies on social media: 6 types of user interactions with business. https://www.nngroup.com/articles/companies-social-media/, 2021.

[37] Tom Huddleston Jr. Americans are being scammed out of billions on social media—look for these 7 red flags. https://www.cnbc.com/2023/10/12/americans-lose-billions-to-social-media-scams-red-flags-to-spot.html, 2023.

[38] Mohammad Taha Khan, Xiang Huo, Zhou Li, and Chris Kanich. Every second counts: Quantifying the negative externalities of cybercrime via typosquatting. In *IEEE Symposium on Security and Privacy (S&P)*, 2015.

[39] Mahmoud Khonji, Youssef Iraqi, and Andrew Jones. Phishing detection: a literature survey. *IEEE Communications Surveys & Tutorials*, 15(4), 2013.

[40] Panagiotis Kintis, Najmeh Miramirkhani, Charles Lever, Yizheng Chen, Rosa Romero-Gómez, Nikolaos Pitropakis, Nick Nikiforakis, and Manos Antonakakis. Hiding in plain sight: A longitudinal study of combosquatting abuse. In *ACM Conference on Computer and Communications Security (CCS)*, 2017.

[41] Klazify. Klazify web category api. https://www.klazify.com/, 2023.

[42] Deepak Kumar, Riccardo Paccagnella, Paul Murley, Eric Hennenfent, Joshua Mason, Adam Bates, and Michael Bailey. Skill squatting attacks on amazon alexa. In *USENIX Security*, 2018.

[43] Rafael Laurenco. From phishing to friendly fraud: Anticipating 2024's fraud dynamics. https://securityboulevard.com/2024/01/from-phishing-to-friendly-fraud-anticipating-2024s-fraud-dynamics/, 2024.

[44] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. A research-oriented top sites ranking hardened against manipulation-tranco. In *Network and Distributed Systems Security (NDSS)*, 2019.

[45] Anastasios Lepipas, Anastasia Borovykh, and Soteris Demetriou. Username squatting on online social networks: A study on x. In *ACM ASIACCS*, 2024.

[46] Guannan Liu, Xing Gao, Haining Wang, and Kun Sun. Exploring the unchartered space of container registry typosquatting. In *USENIX Security*, 2022.

[47] Jienan Liu, Pooja Pun, Phani Vadrevu, and Roberto Perdisci. Understanding, measuring, and detecting modern technical support scams. In *IEEE European Symposium on Security and Privacy (EuroS&P)*, 2023.

[48] Tingwen Liu, Yang Zhang, Jinqiao Shi, Ya Jing, Quangang Li, and Li Guo. Towards quantifying visual similarity of domain names for combating typosquatting abuse. In *IEEE Military Communications*, 2016.

[49] Leland McInnes, John Healy, and Steve Astels. hdbscan: Hierarchical density based clustering. In *Journal of Open Source Software*, 2017.

[50] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[51] Brand impersonation: Phishing emails that impersonate well-known brands. https://www.meshsecurity.io/brand-impersonation.

[52] Metamask twitter customer support. https://twitter.com/MetaMaskSupport.

[53] MidMan. Buy social media account. https://mid-man.com/.

[54] Daniel Milevski. Apify telegram scraper api. https://apify.com/danielmilevski9/telegram-channel-scraper, 2023.

[55] Daniel Milevski. Telemetrio telegram scraper api. https://telemetr.io/, 2023.

[56] Najmeh Miramirkhani, Oleksii Starov, and Nick Nikiforakis. Dial one for scam: A large-scale analysis of technical support scams. In *Network and Distributed System Security Symposium (NDSS)*, 2017.

[57] Martin Moore. Fake accounts on social media, epistemic uncertainty and the need for an independent auditing of accounts. https://policyreview.info/articles/analysis/fake-accounts-social-media-epistemic-uncertainty-and-need-independent-auditing, Feb 07, 2023.

[58] Tyler Moore and Benjamin Edelman. Measuring the perpetrators and funders of typosquatting. In *International Conference on Financial Cryptography and Data Security (FC)*, 2010.

[59] Satnam Narang. Mrbeast scams: Verified accounts, deepfakes used in impersonations to promote fake giveaways on youtube and tiktok. https://www.tenable.com/blog/mrbeast-scams-verified-accounts-deepfakes-used-in-impersonations-to-promote-fake-giveaways-on, Oct 4, 2023.

[60] Shradha Neupane, Grant Holmes, Elizabeth Wyss, Drew Davidson, and Lorenzo De Carli. Beyond typosquatting: An in-depth look at package confusion. In *USENIX Security Symposium*, 2023.

[61] Christina Newberry. 6 tips to protect your brand from fake social media accounts. https://blog.hootsuite.com/fake-social-media-accounts/, 2023.

[62] Nick Nikiforakis, Marco Balduzzi, Lieven Desmet, Frank Piessens, and Wouter Joosen. Soundsquatting: Uncovering the use of homophones in domain squatting. In *Information Security International Conference (ISC)*, 2014.

[63] Nick Nikiforakis, Steven Van Acker, Wannes Meert, Lieven Desmet, Frank Piessens, and Wouter Joosen. Bitsquatting: Exploiting bit-flips for fun, or profit? In *World Wide Web (WWW)*, 2013.

[64] Exclusive report: The state of online consumer brand impersonations in 2023. https://alluresecurity.com/exclusive-report-the-state-of-online-consumer-brand-impersonations-in-2023/, Nov 16, 2023.

[65] Dramatic increase detected in impersonation attacks on social media. https://www.phishlabs.com/blog/dramatic-increase-detected-in-impersonation-attacks-on-social-media, June 02, 2022.

[66] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning (ICML)*, 2021.

[67] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Empirical Methods in Natural Language Processing (EMNLP)*, 2019.

[68] Richard Roberts, Yaelle Goldschlag, Rachel Walter, Taejoong Chung, Alan Mislove, and Dave Levin. You are who you appear to be: A longitudinal study of domain impersonation in tls certificates. In *ACM Conference on Computer and Communications Security (CCS)*, 2019.

[69] David Rupprecht, Katharina Kohls, Thorsten Holz, and Christina Pöpper. Imp4gt: Impersonation attacks in 4g networks. In *Network and Distributed System Security Symposium (NDSS)*, 2020.

[70] Christoph Schuhmann, Richard Vencu, Romain Beaumont, Robert Kaczmarczyk, Clayton Mullis, Aarush Katta, Theo Coombes, Jenia Jitsev, and Aran Komatsuzaki. Laion-400m: Open dataset of clip-filtered 400 million image-text pairs. *ArXiv*, abs/2111.02114, 2021.

[71] Stu Sjouwerman. Four impersonation attacks organizations should be wary of. https://www.forbes.com/sites/forbestechcouncil/\2022/11/23/four-impersonation-attacks-organizations-should-be-wary-of, Nov 23, 2022.

[72] Bharat Srinivasan, Athanasios Kountouras, Najmeh Miramirkhani, Monjur Alam, Nick Nikiforakis, Manos Antonakakis, and Mustaque Ahamad. Exposing search and advertisement abuse tactics and infrastructure of technical support scammers. In *World Wide Web (WWW)*, 2018.

[73] SysSec GitHub. Github repo on Brand Impersonation. https://github.com/CISPA-SysSec/brand_impersonation, 2024.

[74] Janos Szurdi and Nicolas Christin. Email typosquatting. In *Internet Measurement Conference (IMC)*, 2017.

[75] Janos Szurdi, Balazs Kocso, Gabor Cseh, Jonathan Spring, Mark Felegyhazi, and Chris Kanich. The long taile of typosquatting domain names. In *USENIX Security*, 2014.

[76] Telegram. Telegram limits. https://limits.tginfo.me/en.

[77] Eran Tsur. 2024 cyber threat projections – what lies ahead. https://www.memcyco.com/home/2024-cyber-threat-projections/, 2024.

[78] Twitter. User detail twitter api. https://developer.twitter.com/en/docs/twitter-api/v1/accounts-and-users/follow-search-get-users/api-reference/get-users-lookup, 2023.

[79] Nikita Voronov. textdistance. https://github.com/life4/textdistance, 2023.

[80] Danielle Walter. As phishing websites flourish, brands seek protection from impersonation. https://www.akamai.com/blog/security/brands-seek-protection-from-impersonation, 2023.

[81] Yi-Min Wang, Doug Beck, Jeffrey Wang, Chad Verbowski, and Brad Daniels. Strider typo-patrol: Discovery and analysis of systematic typo-squatting. In *USENIX Security*, 2006.

[82] Marcus White. Netflix email impersonation attacks up by 78. https://www.egress.com/blog/phishing/netflix-impersonation-phishing-emails, 2023.

[83] Andrew Williams. Impersonation attacks rise 12 percent in q3 2023. https://www.mimecast.com/blog/impersonation-attacks-rise-12-percent-q3-2023/, 2023.

[84] Mustafa Harun Yılmaz and Hüseyin Arslan. Impersonation attack identification for secure communication. In *IEEE Globecom Workshops*, 2013.

[85] YouTube. Policies & guidelines. https://www.youtube.com/creators/how-things-work/policies-guidelines/.

[86] Z2U. Buy and sell items and services without any intermediaries. safe and hassle-free. https://www.z2u.com/.

# A Brand Social Media Accounts Collection and Search Keywords Rationale

In this section, we touch upon the technical aspect of the process for collecting social media accounts and discuss the rationale behind using specific keywords to identify accounts targeting brand impersonation on social media platforms.

**Social Media Account Collection Process.** To gather social media profiles associated with candidate domains, we utilize two methods. First, an in-house automated script scans the domain's webpage to collect profiles from platforms such as X, Instagram, and YouTube. Second, we utilize an external third-party API service [41] to retrieve additional social media profiles. We cross-reference the profiles obtained from both sources and manually review them for any discrepancies. We noted that not all domains host multiple social media profiles on their web pages or utilize multiple social media accounts. Domains lacking a particular social media profile are excluded from the corresponding platform study. For example, if a domain "example.com" maintains X and YouTube accounts but lacks an Instagram presence, it will be included in the X and YouTube analyses but not included in the Instagram lists. Additionally, we incorporate Telegram into our study, despite lacking Telegram presence on candidate domains' home page and external API [41] social media accounts fetch. Due to its widespread usage among generic users, we added Telegram as one of the social media for our study. This inclusion allows us to examine abusive groups that may conduct brand-based attacks through public Telegram channels.

**Keywords Selection Rationale.** The rationale for selecting 8 popular keywords was based on our incubatory investigation during the design of the experiment and prior research works. We visit each point below.

1. **Incubatory Investigation.** During the initial phase of the experiment, we searched similar accounts associated with over 50 popular brands such as PayPal, Netflix, and Amazon using the UI of X and Instagram. We conduct searches using the second-level domain of each brand as a keyword and observe that fraudsters commonly append keywords such as "recover," "hack," "support," and "help" to brand names as part of an impersonation attack.

2. **Previous Research.** Additionally, our methodology is influenced by two prior studies [2, 72]. Acharya et al. [2] investigated technical support scams prevalent on popular social media platforms such as X, Instagram, and Telegram. The authors found that scammers present themselves as experts, support groups, help desks, traders, and engineers as part of fake technical support scams. Similarly, Srinivasan et al. examined abuse related to technical support scams in search and advertising. They employed a similar technique to identify abusive advertisements as part of defining campaign levels.

Thus, from the above two points, we devised eight popular keywords, namely *rewards*, *recover*, *hack*, *support*, *help*, *assist*, *contact*, and *team* to create a search query for account collection.