Doxing-as-a-Service: Demystifying the Chinese Online Doxing Ecosystem

Yiran Gao*
Huazhong University of Science and Technology
Wuhan, China
gaoyiran@hust.edu.cn

Pengcheng Xia*
Huazhong University of Science and Technology
Wuhan, China
xpc357@hust.edu.cn

Liu Wang†
Huazhong University of Science and Technology
Wuhan, China
liuwang@hust.edu.cn

Tianming Liu†
Huazhong University of Science and Technology
Wuhan, China
tmliu@hust.edu.cn

Haoyu Wang
Huazhong University of Science and Technology
Wuhan, China
haoyuwang@hust.edu.cn

Abstract

Doxing refers to the disclosure of personal information without consent, has evolved from sporadic acts of online vigilantism into a structured and commodified practice. In Chinese cyberspace, this shift has produced Doxing-as-a-Service (DaaS), a commercial model in which personal data is retrieved, organized, and traded as on-demand products. This industrialization of privacy violation lowers the barriers to doxing and amplifies its social harms, posing new challenges for building a responsible and safe web. Yet little is known about how DaaS operates or sustains itself, motivating our systematic, data-driven examination of its ecosystem and practices. This paper provides the first systematic study of the Chinese DaaS ecosystem. Analyzing 25,972 messages and 13.22 million subscriber links from 100 major channels on Telegram, we demystify its organization, operations, and user engagement. We find that the DaaS ecosystem operates through a three-tier supply chain linking data providers, service operators, and end users. Operators sustain illicit businesses through six major service categories, persistent advertising, and crypto-based payments, while users interact via specialized group spaces that enable real-time matching, large-scale identity exposure, and community-driven fraud mitigation. Our findings reveal a mature, resilient underground data market operating within mainstream messaging platforms, highlighting new challenges for online privacy and exposing critical vulnerabilities in platform governance and content moderation. This study provides empirical evidence for developing effective regulatory frameworks and accountability mechanisms to mitigate commodified online harms on encrypted messaging services.

CCS Concepts

• **Social and professional topics** → **Identity theft**; **Social engineering attacks**; • **Security and privacy** → *Social network security and privacy*.

Keywords

Doxing; Identity Theft; Online Abuse; Crime-as-a-Service

ACM Reference Format:
Yiran Gao, Pengcheng Xia, Liu Wang, Tianming Liu, and Haoyu Wang. 2026. Doxing-as-a-Service: Demystifying the Chinese Online Doxing Ecosystem. In *Proceedings of the ACM Web Conference 2026 (WWW '26), April 13–17, 2026, Dubai, United Arab Emirates.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3774904.3792296

1 Introduction

Doxing, the exposure of personal information to shame, harass, or intimidate others, has become a visible and harmful online practice [12]. Enabled by the participatory and searchable nature of the Web, doxing exploits the vast amount of personal data circulating across online platforms. Perpetrators aggregate fragments of data from leaked databases, social media profiles, and public records, reconstructing detailed personal dossiers that can then be disseminated through reposts, screenshots, or mirrored archives [2, 5]. Recent high-profile incidents [1, 3, 4] show that individuals who express controversial opinions can rapidly become targets of retaliatory exposure, underscoring how easily doxing escalates and spreads within online communities. The Web's openness and persistence make such disclosures difficult to contain, allowing private data to become permanently indexed and redistributed across networks. Thus, doxing undermines trust in communities, erodes the privacy expectations that sustain digital participation, and poses a challenge to building a responsible and safe web environment.

Prior research has examined doxing from multiple angles, including its origins in hacker subcultures [20, 38], its role in digital vigilantism [17, 40], and the psychological drivers of participation [13, 34]. Automated tools have also been developed to detect

*Yiran Gao and Pengcheng Xia are the co-first authors.
†Liu Wang and Tianming Liu are the corresponding authors.

The full name of the authors' affiliation is Hubei Key Laboratory of Distributed System Security, Hubei Engineering Research Center on Big Data Security, School of Cyber Science and Engineering, Huazhong University of Science and Technology.

WWW '26, Dubai, United Arab Emirates

ACM ISBN 979-8-4007-2307-0/2026/04
https://doi.org/10.1145/3774904.3792296

doxing on social media [21, 31], demonstrating the prevalence and severity of this threat. While most existing work has focused on English-speaking contexts, where doxing generally involves publicizing personal data for harassment or social shaming [32, 41], its manifestations vary across linguistic and cultural environments. In Chinese-speaking cyberspace, the practice known as *kaihe* ("unboxing") or *renrou sousuo* ("human-flesh search") has undergone a profound transformation. What began as sporadic, crowd-sourced investigations driven by collective outrage [18] has evolved into a commodified and service-oriented system. This phenomenon is termed as **Doxing-as-a-Service (DaaS)**, which operates primarily on Telegram, where personal data is systematically collected, packaged, and sold on demand. Recent reports [29, 30] suggest that DaaS has become increasingly organized and profitable, offering subscription-based access to private data.

Despite growing evidence of this industrialization, the DaaS ecosystem remains largely unexamined. We know little about how such services are structured, how they monetize personal information, or how users participate in these illicit transactions. The absence of systematic analysis limits our understanding of how illicit data practices have adapted to mainstream platforms and global communication infrastructures. To fill this gap, this paper presents the first large-scale empirical study of DaaS in Chinese-language contexts. Drawing on a dataset of 25,972 messages collected from 100 largest Chinese-language Telegram channels, we demystify the Chinese online doxing ecosystem by mapping its organizational structure, examining its operational mechanisms, and analyzing user participation. Our investigation reveals several key findings:

- We clarify the industrial logic of commodified doxing, i.e., conceptualizing the DaaS ecosystem as a three-tier supply chain structure connecting upstream data providers, midstream service operators, and downstream end users, supported by Telegram-based channels and bots.
- We reveal how DaaS operators sustain commodified doxing through six major service categories, templated and pinned advertisements, and revenue streams from paid services and bots. Insider data supply, cryptocurrency payments, and risk-avoidance measures enable stable operations.
- We show that user engagement unfolds through specialized Telegram group spaces. Discussion and query spaces facilitate real-time matching and large-scale data exposure, while fraud and scam reports within user discussions highlight distrust and community-driven mitigation efforts.

2 Background and Related Work

2.1 Cyberbullying and Doxing

Cyberbullying, broadly defined as online harassment enabled by digital communication technologies, differs from traditional bullying in its persistent accessibility and enduring effects [15]. In this context, doxing refers to the non-consensual disclosure of personally identifiable information (PII) and has emerged as a harmful practice [9, 12]. Douglas [12] identified three forms of doxing: deanonymization, targeting, and delegitimization, motivated by exposing wrongdoing, humiliation, intimidation, or punishment. These disclosures exploit either dispersed public data [22] or deceptive social engineering [28].

Existing research explores doxing through its prevalence, motivations, and detection [8, 13, 21, 31, 33, 34]. For example, Snyder et al. [34] were among the first to systematically examine doxing at scale, analyzing posts on platforms such as *pastebin.com* and *4chan.org*. Using automated classifiers and longitudinal tracking of affected social-media accounts, they found that doxed users were significantly more likely to close or privatize their profiles after exposure. Furthermore, technical research has leveraged contextual embeddings and transformer models to achieve high-accuracy, explainable detection of doxing and its intersection with fake news [21, 31]. Overall, existing research has mainly focused on English-speaking contexts, where doxing is viewed as the public exposure of personal information to shame or retaliate against others. These studies highlight moral and emotional motives such as revenge or justice, portraying doxing as spontaneous online aggression. In contrast, we examine how doxing in Chinese cyberspaces has evolved into a commercialized, platform-based service model.

2.2 Doxing in Chinese Cyberspace

Parallel to early English doxing, Chinese cyberspace first saw "renrou sousuo" (literally "human-flesh search"), which referred to episodic, crowdsourced campaign efforts by large groups of internet users. Often reacting to perceived corruption or other wrongdoing, participants gathered personal information from public sources and shared it through ad-hoc coordination on forums and social media [11]. Scholars have modeled these campaigns via epidemic dynamics to explain their rapid spread through network coordination [11, 27]. Empirical studies further highlight their dual nature, acting as both civic vigilance and entertainment-driven privacy violations [18].

In recent years, the term kaihe (literally "unboxing") has moved from hacker slang into everyday online use in Chinese cyberspace [39]. It now appears in contexts ranging from celebrity gossip to personal disputes and political intimidation, showing how doxing has spread into mainstream culture [30]. Reports also note that many kaihe cases rely on so-called "social-engineering databases" [35], where personal data from past breaches is collected and sold. These sources make it easy and cheap to expose individuals. As these practices have become more widespread, the strong negative associations once tied to privacy violations have weakened, reflecting a broader normalization of doxing culture in online communities.

2.3 Doxing-as-a-Service

Underground data trading has evolved from fragmented dumps into specialized, platform-based infrastructures[10, 14, 25]. Recent studies highlight a systemic shift in market demand from financial data toward PII, accompanied by a migration of illicit activities to encrypted platforms like Telegram [25]. In the Chinese context, this ecosystem functions as a highly modular underground ecosystem, where data harvesters and brokers facilitate the resilient circulation of personal archives for downstream misuse [10, 14].

This professionalization mirrors the Crime-as-a-Service (CaaS) paradigm [19, 24], which lowers technical barriers by offering scalable, on-demand illicit services such as Ransomware-as-a-Service [26]. Doxing-as-a-Service (DaaS) represents the data-driven evolution of this model, where the exposure of personal information

is commercialized through automation and integrated payments on semi-public platforms. While existing research primarily examines data trading as a macro-level resource-gathering phase for subsequent financial fraud [10, 25], we investigate DaaS as a highly operationalized downstream service. In this model, static data assets are transformed into real-time, on-demand tools that significantly lower the technical threshold for targeted privacy violations.

3 Study Design

3.1 Research Questions

This study is guided by the following research questions (RQs):

- **RQ1: How is the DaaS ecosystem structured and operationalized?** This question takes a macro-level view, aiming to map the overall structure of the DaaS ecosystem. It reveals key components, actors, and infrastructures that sustain it, offering insights into how such illicit services scale, persist, and adapt.
- **RQ2: How do DaaS operators structure, promote, and sustain their services?** From the vendor perspective, this question examines the organizational and technical mechanisms that enable DaaS operations, including service provisioning, monetization, and marketing strategies that allow operators to scale and maintain illicit businesses despite enforcement pressures.
- **RQ3: How do users interact with and engage in DaaS transactions?** Focusing on the user perspective, this question analyzes user interactions and participations through Telegram-based group spaces, revealing how users engage with DaaS markets and navigate transactional processes.

3.2 Data Collection

To investigate the DaaS ecosystem, we collected a comprehensive dataset from Telegram, which has evolved into a central hub for illicit data-trading and doxing services. Our data collection followed a three-step workflow illustrated in Figure 1.

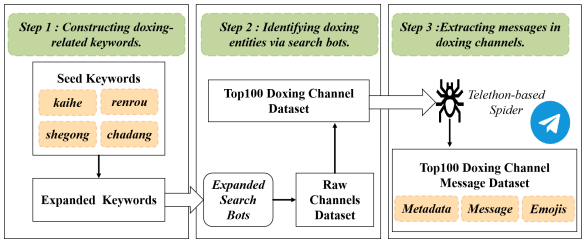

Figure 1: The data collection workflow.

3.2.1 Constructing Doxing-related Keywords. We adopted a two-step workflow to construct a robust keyword set to identify doxing-related entities. We listed four vernacular seed terms transliterated *Pinyin*: *kaihe* ("unboxing"), *renrou* ("human-flesh"), *chadang* ("check-records") and *shegong* ("social engineering"). These terms were chosen because they represent both historical practices (such as collective crowd-sourced identification) and more recent, commodified forms of doxing, providing coverage across the main stages of its evolution in Chinese online discourse. We then expanded them into a robust lexicon using derived variants found in early retrievals and bot menus such as *chadangjiqiren* ("social-engineering bot") and *shegongku* ("social-engineering database").

3.2.2 Identifying Doxing Entities via Search Bots. Telegram's search bots act as automated directories that index channels, groups, and files, providing entry points into illicit markets. We first compiled 13 search bots from previous literature [16], of which 7 were functional. By cross-referencing metadata from *TGStat* [37] and *Telemetr.io* [36], we identified 3 additional bots, yielding 10 active bots in total. We queried our expanded keyword list and retrieved a raw dataset of 312 candidate channels. All candidates underwent rigorous manual screening to verify their explicit doxing relevance. From these verified candidates, we selected the top 100 channels by subscriber count as our representative corpus, as these high-visibility hubs concentrate the majority of illicit activity and information exchange.

3.2.3 Extracting Messages in Doxing Channels. For each sampled channel, we subscribed to its feed and retrieved full historical archives up to July 22, 2025, including text messages, price lists, advertisements, transaction records, and media attachments. Data collection used the official Telegram API [7] with daily incremental updates. The final dataset comprises 25,972 messages with associated metadata, including 325M cumulative views and 2.27M forwards. Channel activity is uneven, with message counts ranging from 1 to 2,981 per channel and subscriber numbers spanning from 76,069 (median) to more than 550,000 (maximum). All statistics are derived from channel metadata and messages; no personally identifiable information (PII) was downloaded or stored.

4 RQ1: The Ecosystem of DaaS

4.1 The Structure of the Ecosystem

Our analysis conceptualizes the DaaS ecosystem as a three-tier supply chain that connects data generation, service operation, and end-user consumption, as shown in Figure 2. Within this structure, data flows downstream from providers to users, while money flows upstream, sustaining the illicit market.

4.1.1 Upstream Data Providers. This layer consists of data suppliers including Hackers and insiders who leak raw data via database breaches, traffic interception, credential cracking, or SDK-based and DPI-based harvesting. This layer provides the fragmented "raw material" for the market-an opportunistic collection of datasets varying in type, structure, and reliability that together form the essential input feeding the DaaS market.

4.1.2 Midstream DaaS Operators. At the core of the ecosystem are the service operators, who aggregate, sanitize, and repackage leaked, fragmented raw material into marketable products. Hosted primarily on Telegram, operators deploy multi-component infrastructures combining manual doxing channels and social-engineering bots. They fulfill two primary functions: (*i*) Data integration and repackaging: converting heterogeneous leaks into standardized, queryable "washed data"; and (*ii*) Market mediation: providing advertising, payment processing, and post-sale support under recognizable brand identities. Financial flows from end users and affiliates are funneled back to operators, enabling continuous data acquisition and infrastructure maintenance. Through this layer, fragmented raw data are

professionalized and transformed into scalable, service-oriented offerings.

4.1.3 Downstream End Users. Consumers utilize DaaS outputs for various purposes, including illegal debt collection, precision fraud, underground market advertising, lending risk assessment, and private retaliation. DaaS users range from professional offenders to ordinary individuals seeking personal information, as the accessibility of Telegram channels lowers participation barriers and obscures the boundary between illicit and routine misuse.

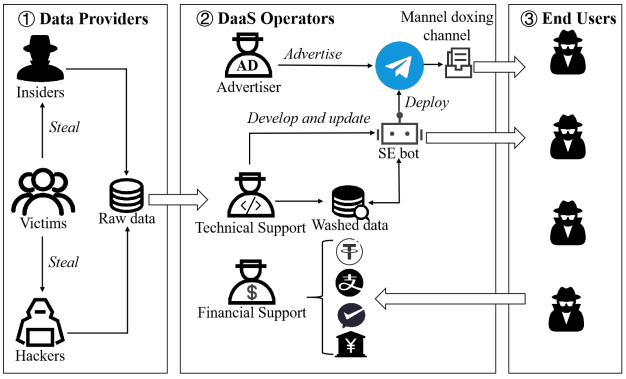

Figure 2: Three-tier supply chain of DaaS.

Finding 1: *The DaaS ecosystem functions as a three-tier structure linking upstream data providers, midstream service operators, and downstream users. This layered organization transforms fragmented leaks into standardized, on-demand services, illustrating how doxing has evolved from scattered acts into a coordinated service model.*

4.2 Operational Infrastructure of the Ecosystem

We next characterize the core operational infrastructure of the DaaS ecosystem, including channels that act as storefronts and automated bots that support data retrieval.

4.2.1 DaaS Channels. The channels serve as the public-facing storefronts of the DaaS ecosystem where services are advertised, transactions are initiated, and customers are engaged.

Channel Categorization. Based on content and operational patterns, one author conducted the initial coding of the 100 sampled channels, and all ambiguous cases were jointly reviewed and resolved with a second author, resulting in four functional categories: (*i*) *Manual doxing channels* (65 channels) constitute the majority. These operate as interactive service desks that post price lists, handle custom requests via direct messaging, and perform near-real-time data lookups for customers. (*ii*) *Social-engineering bot notification feeds* (31 channels) represent the second largest group. They serve as broadcast channels dedicated to disseminating bot-related updates, commands, and usage instructions. Notably, three seemingly independent manual doxing channels were found to promote the same SE bot, share identical messages, and redirect users to the same human service representative, revealing a hybrid operational model where multiple front-end storefronts conceal a unified automated backend. (*iii*) *Aggregators* (2 channels) curate and share links to various social-engineering bots, functioning as directories within the broader DaaS ecosystem. (*iv*) *Backup channels* (2 channels) are created to replicate content and maintain service availability during suspension or takedown events affecting primary channels. This distribution indicates a marketplace dominated by direct, human-mediated service provision, but increasingly integrated with automated components that streamline operations and expand scalability.

Channel Lifespan. Analysis of channel lifespans reveals that most DaaS channels remain active for one to three years, with a smaller subset persisting even longer. This sustained activity indicates a relatively stable and resilient ecosystem, supported by recurring demand and adaptive infrastructure. Such persistence may also reflect limited or uneven enforcement, highlighting the importance of examining DaaS as an enduring socio-technical ecosystem and the need for more coordinated regulatory interventions.

Finding 2: *DaaS channels cluster into four functional types—manual service desks, social-engineering bot feeds, aggregators, and backups–revealing a hybrid ecosystem where human-operated and automated components coexist. Their lifespans suggest a stable, enduring market sustained by recurring demand and limited enforcement.*

4.2.2 Social Engineering Database Bots. Social-engineering database bots represent the core technical infrastructure of the DaaS ecosystem. We identified 31 distinct bots, which together have 882,910 monthly active users. Channels and bots exhibit a many-to-many relationship, with channels embedding multiple bots and bots promoted across sub-channels.

User Interaction. To understand the interaction modes of these bots, we conducted a manual engagement experiment and found 22 bots were functional, the remainder were inactive. We found that all active bots share a standardized interaction template. Upon issuing the canonical `/start` command, users receive a welcome payload containing an inline keyboard, subscription prompts, recharge buttons, and referral links. Figure 4 of the Appendix illustrates an example of user-bot interaction. The bots consistently support three canonical query modes: (*i*) *direct lookup*, retrieving results for exact identifiers (e.g., phone number, ID); (*ii*) *completion queries*, auto-filling partial inputs; (*iii*) *fuzzy or hunter*, returning candidate records based on vague descriptors such as surname or city.

Incentive Design. Most DaaS bots embed incentive and access systems that simultaneously drive engagement and regulate participation. The dominant model (22 active bots) operates under a *points-based regime*, in which each query consumes one point that users can replenish through daily check-ins, successful referrals, or direct purchase. Multi-level referral systems are a universal component of this design. Referrers receive bonuses or recurring commissions from invitee expenditures, forming hierarchical promotion chains. Meanwhile, access to bot functions is gated by mandatory channel subscriptions. Compliance is automatically verified via Telegram's `getChatMember` API [6], and non-subscribers are denied functionality. Consequently, these bots act not only as lookup utilities but also as integrated engines for traffic funneling, monetization, and user retention, reinforcing the cohesion of the broader DaaS ecosystem.

Finding 3: *Social-engineering bots operate as both technical infrastructure and commercial engines within DaaS. By combining automated data retrieval and referral-based growth loops, they transform*

Table 1: DaaS services and their price range.

Primary Category	Representative Content	Price (USDT)
Identity & Personal	National ID numbers, household registration, marriage history, family relations	15 – 150
Financial & Assets	Bank balance, bank statement, property ownership	300 – 1500
Communication & Online	WeChat ID reverse lookup,douyin reverse lookup, WeChat friend list dump	300-2000
Judicial & Security	Criminal record check, court testimony excerpt, bank-card freeze reason	100 – 1700
Location & Travel	Mobile geolocation,hotel stay-with-companion log, air & rail travel history, express delivery address history	570 - 1200
Social Benefits & Resume	Education verification, social security log	100 - 200

individual lookup tools into scalable, self-sustaining service platforms that bind together the broader ecosystem.

5 RQ2: Service Operators in DaaS

5.1 Taxonomy of DaaS Services

We next examined the range of services offered by DaaS operators. Manual analysis of representative channels shows that most offer a similar comprehensive portfolio, often advertised as *"customizable"* or capable of finding *"anything you want."*

We derived a six-category taxonomy from explicit operational artifacts such as itemized price sheets, service menus, and pinned posts, which two authors iteratively merged into higher-level categories. (see Table 1): (*i*) The **Identity & Personal** category forms the foundation of most doxing operations, covering core identifiers such as national ID numbers, household registration records, marriage history, and family relationships. (*ii*) **Financial & Assets** services extend this to economic profiling, offering access to bank balances, account statements, and property ownership data—often sourced from state or commercial financial institutions. (*iii*) **Communication & Online** services target digital presence and interpersonal connections, including mobile geolocation, WeChat ID reverse lookups, and full WeChat friend list dumps. (*iv*) **Judicial & Security** category explicitly claims to provide access to criminal records, court testimony excerpts, border control status, and fugitive listings. Some also advertise these services to clients abroad seeking to assess their risk of arrest, highlighting the downstream customer base targeted by DaaS operators and their embeddedness within broader cybercrime networks. (*v*) **Location & Travel** data includes hotel stay logs, transportation histories, and express delivery addresses, enabling reconstruction of an individual's mobility patterns. (*vi*) **Social Benefits & Resume** services focus on education verification and social security data, allowing reconstruction of a person's professional and institutional background.

Finding 4: *DaaS operators offer services spanning six categories, including identity, financial, communication, travel, and judicial records, enabling on-demand access to a wide range of personal data.*

5.2 Promotion of DaaS Services

DaaS operators usually attract customers and maintain visibility via advertisements (ads). We analyze their advertising and promotional practices across Telegram channels. This includes examining common ad message formats and pinned promotional content.

5.2.1 Ad Message Format. We examined 100 randomly sampled ad messages and found that DaaS operators follow a highly standardized and reusable format optimized for Telegram's fast-scrolling interface (see Figure 5 in the Appendix). Each ad typically begins with a *welcome message* ("Welcome to our channel!"). This immediately leads into a clearly structured service section, consisting of the *service name* and a brief *description* outlining its function (e.g., WeChat-ID reverse lookup), occasionally highlighting success rates or add-on features such as partial bank card retrieval. Ads explicitly specify *pricing*, quoted in USDT or RMB, and the *delivery cycle*, ranging from "instant" to several working days. A *contact* section then provides communication handles for both human sellers and bot interfaces. Many ads conclude with a mention of a linked *social-engineering bot*, reinforcing integration between manual and automated services. To enhance credibility, a *sample screenshot* (or document) is often attached, typically showing examples of data outputs relevant to the advertised service (see the upper gray box of Figure 5 containing records with fields of WeChat ID, name, national ID, and phone number). These samples serve as live demonstrations of claimed capabilities and data authenticity.

5.2.2 Ads in Pinned Messages. Pinned posts on Telegram channels serve as persistent, high-visibility anchors that operators often use to showcase core advertisements and maintain constant promotional exposure. Across the 100 sampled channels, 82 had pinned posts, yielding 237 pinned messages in total. Promotional content dominates this corpus: 74 pinned posts (31%) are explicitly commercial advertisements, while another 47 (20%) present consolidated service menus in template form. Other pinned content, such as insider recruitment, data bids, anti-scam notices, and tutorials, appears less frequently and serves primarily logistical or credibility-related purposes.

Pinned placement substantially increases visibility. On average, pinned messages received 81.8k views, compared to 53.2k for non-pinned messages, corresponding to a 1.5× overall uplift. At the channel level, the average visibility uplift reached 2.73×. These results indicate that pinning significantly amplifies exposure, making it a central promotional mechanism in DaaS channels.

Finding 5: *DaaS ads follow standardized templates featuring clear service descriptions, pricing, delivery cycles, contacts, and sample proofs. Pinned posts in DaaS channels are dominated by such promotional ads whose fixed, high-visibility placement significantly amplifies exposure, making them a core tool for sustained marketing.*

5.3 Monetization

To understand how DaaS operators generate revenue, we examine the economic mechanisms underpinning this ecosystem. Monetization primarily occurs through two channels: priced doxing services and pay-per-query bots, while cryptocurrency-based payment systems facilitate seamless and scalable transactions.

5.3.1 Pricing of DaaS Services. To understand the market benchmark for DaaS service pricing, we collected 152 price observations from 23 channels covering a variety of service categories in our dataset. Pricing patterns vary notably both across and within categories (see Table 1), reflecting differences in data sensitivity, acquisition difficulty, and perceived commercial value. *Identity & Personal*

(15–150 U[1]) and *Social Benefits & Resume* (100–200 U) services are generally the cheapest. *Financial & Assets* data (300–1500 U) and *Communication & Online* records (300–2000 U) command higher prices, consistent with their restricted access and greater misuse potential. *Judicial & Security* records (100–1700 U) and *Location & Travel* data (570–1200 U) occupy the mid-to-high range, balancing limited supply with strong buyer interest in profiling or tracking. Generally, the market remains highly decentralized. Each channel independently sets its tariffs, applies promotions, and negotiates custom prices, indicating that pricing power is distributed among individual operators rather than centralized authorities.

Finding 6: *The pricing of DaaS services is supported by a decentralized pricing structure where routine identity and social-record lookups remain inexpensive while financial, communication, and high-risk data command premium rates.*

5.3.2 Profitability of Bots. Although often advertised as "free tools," the bots are tightly coupled with gray-market monetization systems. As aforementioned, the dominant points-based model charges credits per query, which users can replenish through check-ins, referrals, or direct purchases, effectively turning interaction into pay-per-use access. Besides, a smaller subset of bots adopts a tiered access model: sensitive data fields are partially redacted in free queries, and users must complete micropayments or membership upgrades to reveal records. These models convert ostensibly open utilities into consistent revenue sources within the DaaS ecosystem.

5.3.3 Payment Channels. Payment practices within the DaaS ecosystem reveal a structured process of monetization. Across all examined channels, TRC20–USDT functions as the de facto transaction medium. Its fast confirmation speed, low transfer fees, and 1:1 U.S. dollar peg make it ideal for cross-border microtransactions, while its pseudo-anonymous nature minimizes traceability compared to fiat-based channels. Users initiate payments directly through wallet addresses embedded in ads or bot interfaces. Once the transfer is verified, operators issue digital "credits" or query quotas within minutes, enabling an automated and frictionless exchange between cryptocurrency and data access. More than half of the observed channels additionally support limited fiat alternatives, including Alipay, "password red packets" (a claim-based Alipay transfer requiring a specific phrase), WeChat Pay, or direct bank transfers. These methods are primarily offered for domestic buyers who lack crypto wallets, reflecting operators' attempts to maintain accessibility while prioritizing crypto as the default settlement mode.

Finding 7: *DaaS operators profit mainly through paid doxing services and pay-per-query bots, with transactions predominantly settled in TRC20-USDT. This dual revenue structure supports scalable income generation and commercial sustainability of DaaS ecosystem.*

5.4 Data Acquisition

We next examine how DaaS operators obtain and replenish data. Thus, we extracted data-related posts involving 1, 284 insider recruitment posts and 60 bulk dataset purchase bids, alongside 207 ads promoting technically exfiltrated datasets. Our analysis reveals two broad yet complementary modes of data acquisition.

5.4.1 Insider-Enabled and Market-Driven Acquisition. The first category revolves around the human supply chain that connects privileged insiders with data brokers. A total of 68 channels openly published recruitment messages targeting employees in telecom carriers, express delivery companies, financial institutions, and major Internet platforms. These posts often mimic legitimate corporate hiring templates, complete with salary ranges, onboarding instructions, and "safety training" for cryptocurrency laundering, blurring the boundary between illicit labor markets and formal employment. Such insiders sell or leak access to backend systems containing national ID records, call logs, financial data, and logistics information, forming the foundation of DaaS's high-credibility datasets.

Complementing insider recruitment, at least 15 channels issued public dataset purchase requests, each requiring 10, 000 or more records per batch. The coexistence of lucrative insider salaries and stringent procurement requirements demonstrates the scarcity and market value of first-hand data access.

5.4.2 Technical Exploitation and Automation. Technical exploitation takes several complementary forms. *Carrier DPI* misuse refers to illicit access to deep-packet inspection or flow feeds, yielding traffic-level identifiers and session/location metadata useful for re-identification. *Malicious or compromised SDKs and API* abuse denote trojanized third-party libraries or the misuse of poorly protected service APIs to harvest PII, device fingerprints, and bulk query results at scale. *Order decryption* covers efforts to recover plaintext shipping or order details from obfuscated e-commerce/logistics records, a process frequently enabled by backend access. These vectors automate data collection and greatly increase throughput.

By contrast, *qiangdeng* ("forced-login") denotes an automated account-takeover pipeline that programmatically harvests live account artifacts (order histories, saved addresses, payment traces). In practice, forced-login deployments typically leverage the techniques above—credential or token leaks, SDK/API-derived credentials, DPI-informed targeting, or insider-provided access—to scale and monetize account extraction. The presence of multiple advertisements for forced-login sourced packages in our corpus indicates that this hybrid, automation-enabled pipeline has already been commercialized within the DaaS ecosystem.

Finding 8: *DaaS operators source data through human-enabled insider access and large-scale technical exploitation. This combination blends the reliability of privileged data with the scalability of automation, ensuring a continuous and diversified data supply.*

5.5 Risk Avoidance

We identify three primary risk mitigation strategies that DaaS operators use to sustain operations under enforcement pressure.

5.5.1 Backup Channels. A key operational safeguard we observed is the systematic creation of backup channels to ensure service continuity amid takedowns or bans. Within our dataset, two dedicated backup channels were explicitly identified; however, this number likely underrepresents the true scale of the practice, as many operators maintain secondary channels beyond our sampling frame or conceal them through restricted-access links. Backup channels

[1]TRC-20 USDT is the Tether stablecoin issued on the Tron blockchain, offering low transaction fees and fast transfers commonly used in gray-market payments.Henceforth, we use "U" to denote USDT.

typically mirror the content and structure of primary ones, enabling rapid recovery with minimal audience loss after enforcement.

5.5.2 Transaction Anonymity. Payment systems also serve as a mechanism of risk mitigation. Although many operators accept RMB-based options for convenience, a growing number have explicitly restricted or discontinued domestic payment channels to avoid financial traceability and platform scrutiny. Several channels display disclaimers such as *"no CNY accepted"* or *"crypto only"*, underscoring heightened sensitivity to anti-money-laundering enforcement and account freezes on Chinese platforms. Furthermore, for those channels that still accept RMB, an additional 10% fee is typically imposed—a practice that effectively encourages users to transition toward more discreet cryptocurrency channels like USDT. We also find operators frequently route crypto receipts through multiple intermediary addresses (multi-hop transfers) before attempting to on-ramp the proceeds; in some cases, these chains ultimately interface with major centralized exchanges (e.g., Binance, OKX) for fiat withdrawal. This reflects a payment strategy that reduces traceability while maintaining convenient access to liquidity.

5.5.3 Refusal Policies. There are 57 channels publicly posting refusal policies disclaiming services for politically exposed or sensitive individuals. The excluded categories, often listed verbatim across channels, include law enforcement and military personnel, civil servants, celebrities, journalists, executives of state-owned enterprises, deceased officials, public-sector scholars, and minors. Buyers are required to attest that their target is not a sensitive figure; if subsequent verification proves otherwise, the operator cancels the request and withholds the payment as a penalty.

Finding 9: *DaaS operators actively manage platform and legal risks through redundant infrastructure, anonymous payment mechanisms, and symbolic refusal policies, enabling a fragile yet resilient operational equilibrium under continuous scrutiny.*

6 RQ3: User Engagement in DaaS

6.1 Overview of Doxing Groups

Interactive doxing groups serve as a dynamic space where buyers, sellers, and intermediaries coordinate transactions, verify data authenticity, and negotiate disputes in real time. To understand how users interact with and engage in DaaS transactions, we identified linked groups by inspecting each channel's bio, pinned posts, and message history for hyperlinks and plain-text aliases, then verified candidates through Telegram's native search and third-party indexes. We retained only groups that were reachable and visibly active at the time of inspection, yielding 15 public groups. Their membership ranges from approximately 400 to 5,500 users, and all groups were joinable during data collection. Because private or invite-only groups are excluded, this sample provides a conservative, lower-bound view of the broader interaction landscape.

Based on observed communication patterns and functions, we categorize them into four main types: (*i*) **Notification groups** (n=2), which relay service updates, bot announcements, and promotional messages, capturing user attention and feedback signals. (*ii*) **Discussion groups** (n=6), where participants often debate data reliability, report scams, or share experiences of trust and fraud. (*iii*) **Query groups** (n=5), where users post or crowdsource doxing requests, revealing market demand patterns and evolving interests in personal data exposure. (*iv*) **Escrow groups** (n=2), which mediate payments and dispute resolutions through informal mechanisms that reduce fraud risk and sustain user confidence.

Finding 10: *DaaS groups provide the interactive backbone of the ecosystem, facilitating communication, and coordination among participants across notification, discussion, query, and escrow functions.*

6.2 Community Discussion

Discussion groups serve as a primary interactive space for user engagement within the DaaS ecosystem. In moderated groups, discussions focus on concrete data requests, effectively functioning as gray-market matching spaces that reduce search frictions between buyers and sellers. Sellers are predominantly independent operators without their own channels or bots, who monitor group activity and proactively contact buyers in response to posted queries.

Across active discussion groups, conversations predominantly cluster around three themes: supply–demand interactions, fraud and scam reports, and technical discussions. Supply–demand content dominates these spaces: in one large group with over 5,500 daily messages, more than 90% of posts involve data requests or seller responses, primarily concerning identity data and social-media-based reverse lookups (e.g., WeChat, Douyin, Kuaishou). Discussions of fraud and technical limitations reveal that DaaS services are far from uniformly reliable. Sellers acknowledge low success rates for certain queries, such as Kuaishou reverse lookups, and some are reported to fabricate results when legitimate data cannot be obtained. This mismatch between strong demand and inconsistent supply highlights the fragility of DaaS trust networks, where information asymmetry enables persistent deception despite ongoing user scrutiny.

Finding 11: *Discussion groups facilitate real-time matching between buyers and sellers while exposing the ecosystem's instability, where exchange, opportunistic fraud, and peer regulation coexist.*

6.3 Data Query Patterns

Five query groups embed a social-engineering database bot with daily message volumes of 100-4500. Users issue text commands to query the bot, while non-query posts such as ads or off-topic messages are removed, and repeated violations lead to loss of access.

To quantify the extent of data exposure, we analyzed 411, 707 messages collected from five query groups from May 30, 2025 to August 25, 2025. After removing failed requests and routine sign-in logs, we found that personal identity information of over 300,000 unique individuals had been directly exposed within these publicly accessible groups during 3 months. The leakage remains ongoing, with an average of about 5, 100 new successful queries per day from these five groups alone. Because most users delete their query commands shortly after submission to conceal their identity, we examined 100 undeleted query records that remained visible. Among these, more than half used phone numbers as search keys, nearly one-third relied on national ID numbers, and a smaller share used names as identifiers. A small but non-trivial fraction referenced QQ numbers (a kind of instant messaging identifiers), indicating the continued use of legacy social-network identifiers in doxing

practices. These patterns collectively demonstrate a dynamic and self-sustaining data-leakage process within DaaS query services. Given that these figures originate solely from public groups, the overall scale of DaaS data exposure, including transactions through private bots and manual services, is likely to be far greater.

Finding 12: *DaaS query services exhibit large-scale, continuous, and self-concealing identity exposure, suggesting that the visible leakage represents only a fraction of the overall ecosystem.*

6.4 Fraud Risks and Mitigation Practices

Beyond doxing requests, DaaS group discussions often include scam complaints and disputes, reflecting the structural nature of fraud driven by anonymity, unenforceable transactions, and information asymmetry. We examined how participants address these risks and manage transactional uncertainty.

6.4.1 Transactional Fraud in DaaS Markets. A keyword search for *"pian"* (scam) across the group corpus yielded approximately 22,000 mentions (about 10% of total messages), including 3,280 self-reported victimization cases and 17,000 accusations directed at service providers. This indicates that fraud is systemic rather than exceptional, reflecting persistent distrust and repeated failures of transactional guarantees within DaaS markets. Further inspection reveals three recurring fraud patterns: delivery failure (collecting payment but failing to deliver or ceasing contact), bait-and-switch schemes (providing falsified or low-quality data), and advanced-fee scams that involve deceptive top-up offers, where users are persuaded to pay large upfront deposits for supposed discounts.

6.4.2 Clone Channels. Prior work has shown that Telegram hosts numerous clone groups [23], and the DaaS ecosystem is no exception. Our analysis revealed a case in which three seemingly independent service channels were in fact operated by the same team, sharing identical bots, promotional content, and customer contacts. Building on this observation, we searched channel names to identify additional clones and found 124 channels that exhibited similar naming and content patterns but were not listed as official backups of the originals. Among them, the most impersonated original channel has nine copies. Given their lack of verified affiliation, we consider these channels high-risk entities that likely aim to divert payments or impersonate established DaaS providers.

6.4.3 Fraud Mitigation Mechanisms. Given the pervasive fraud documented in DaaS transactions, participants have developed both community-driven and escrow-based mechanisms to mitigate risk and restore transactional trust.

Community-driven Mitigation. We first observed a few community-driven anti-fraud channels that originated within discussion groups. These channels were typically created by users who had been redirected or scammed and sought to expose fraudulent vendors. Acting as informal bulletin boards, they collect scam reports, screenshots, and vendor blacklists contributed by other members. Although their number is limited and moderation is inconsistent, these channels represent grassroots efforts to improve transparency and mutual awareness within an unregulated market.

Escrow-based Mitigation. At the same time, large DaaS operators have introduced more structured systems. Prominent among these are escrow services with deposit models, which replace interpersonal trust with procedural assurance. In this workflow, vendors must place a deposit with the escrow operator, and buyers can verify the vendor's bond status through an escrow bot before proceeding. Transactions are completed only after both parties confirm payment and data delivery, at which point the bot releases the funds. We identified two escrow groups implementing this process, along with smaller invite-only variants that restrict access to vetted participants. While these systems reduce risks, they centralize power in escrow moderators and impose non-trivial entry costs.

Finding 13: *Despite pervasive fraud and impersonation, DaaS participants maintain transactions through community reporting and escrow-based mechanisms that partially restore trust.*

7 Discussion

Data Leakage Risks beyond Doxing. Our observations suggest that DaaS amplifies the impact of data breaches beyond individual doxing incidents. By repackaging leaked data for downstream uses such as fraud and social engineering, DaaS acts as an intermediary infrastructure that extends data lifecycles and turns isolated breaches into persistent, large-scale privacy and security risks.

Mitigations. Mitigation efforts should address the entire data pipeline. Organizations handling sensitive data should adopt audit trails and anomaly-based logging to detect unauthorized access, while developers should monitor abnormal SDK or API usage indicative of data exfiltration. Platforms such as Telegram also bear responsibility for moderating data-trading channels and removing automated bots that facilitate DaaS transactions. In addition, the widespread use of cryptocurrency payments suggests that blockchain tracing and exchange-level cooperation may help disrupt DaaS monetization. Overall, reducing DaaS resilience requires coordinated technical, organizational, and financial interventions.

Ethical Considerations. All data were passively collected from public Telegram channels, without interacting with users or operators. Sensitive information was masked before storage and analysis, and no private or restricted groups were accessed. The study follows institutional ethical standards for minimal-risk research.

8 Conclusion

This paper presents the first large-scale empirical study of the Chinese DaaS ecosystem based on Telegram data. We reveal a three-tier structure linking data suppliers, service operators, and users, and show how platform features such as bots, pinned posts, and groups enable the commodification of personal data into on-demand services. By examining user interactions, we further illustrate how illicit data transactions are coordinated and regulated in everyday practice. Our findings highlight the growing risks posed by such gray-market ecosystems and underscore the need to account for them in platform governance and data protection efforts.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (grant No.62572209) and the Hubei Provincial Key Research and Development Program (grant No. 2025BAB057).

References

[1] 2025. Doxing on Rise After Info Leaks About Judge in DOGE Case: Privacy Expert. https://www.newsweek.com/doxing-rise-after-info-leaks-about-judge-doge-case-privacy-expert-2032954,.

[2] 2025. The Escalating Threats of Doxxing and Swatting: An Analysis of Recent Developments and Legal Responses. https://www.naag.org/attorney-general-journal/the-escalating-threats-of-doxxing-and-swatting-an-analysis-of-recent-developments-and-legal-responses/,.

[3] 2025. Federal Grand Jury Charges Three Women with Following ICE Agent Home from Work and Livestreaming His Home Address on Instagram. https://www.justice.gov/usao-cdca/pr/federal-grand-jury-charges-three-women-following-ice-agent-home-work-and-livestreaming.

[4] 2025. People are getting fired for allegedly celebrating Charlie Kirk's murder. It looks like a coordinated effort. https://www.cnn.com/2025/09/13/business/charlie-kirk-death-fired-comments.

[5] 2025. Resources for Individuals on the Threat of Doxing. https://www.dhs.gov/sites/default/files/2024-01/24_0117_ope_resources-for-individuals-on-the-threat-of-doxing-508.pdf,.

[6] 2025. TDLib: getChatMember Class Reference. https://core.telegram.org/tdlib/docs/classtd_1_1td__api_1_1get_chat_member.html,.

[7] 2025. Telegram APIs. https://core.telegram.org/api.

[8] Briony Anderson. 2025. Doxxing to destroy: The convergence of transphobic hate speech and non-consensual disclosure on X. *Crime, Media, Culture* (2025), 17416590251345736.

[9] Briony Anderson and Mark A Wood. 2021. Doxxing: A scoping review and typology. *The Emerald international handbook of technology-facilitated violence and abuse* (2021), 205–226.

[10] Guangxuan Chen, Qiang Liu, Guangxiao Chen, and Anan Huang. 2025. Exploring illicit personal information trading behind telecom fraud in China. *Humanities and Social Sciences Communications* 12, 1 (2025), 1–11.

[11] Long Cheng, Lei Zhang, and Jinchuan Wang. 2012. A study of human flesh search with epidemic models. In *Proceedings of the 4th Annual ACM Web Science Conference*. 67–73.

[12] David M Douglas. 2016. Doxing: A conceptual analysis. *Ethics and information technology* 18, 3 (2016), 199–210.

[13] Stephen Foster and Jasmine Cross. 2024. Dark doxxing: How Dark Triad traits impact support for doxxing behaviors. *Personality and individual differences* 217 (2024), 112432.

[14] Taisiia Garkava, Asier Moneva, and E Rutger Leukfeldt. 2024. Stolen data markets on Telegram: a crime script analysis and situational crime prevention measures. *Trends in Organized Crime* (2024), 1–25.

[15] Gary W Giumetti and Robin M Kowalski. 2022. Cyberbullying via social media and well-being. *Current opinion in psychology* 45 (2022), 101314.

[16] Yanhui Guo, Dong Wang, Liu Wang, Yongsheng Fang, Chao Wang, Minghui Yang, Tianming Liu, and Haoyu Wang. 2024. Beyond App Markets: Demystifying Underground Mobile App Distribution Via Telegram. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 8, 3 (2024), 1–25.

[17] Isabella Hansen and Darren J Lim. 2019. Doxing democracy: influencing elections via cyber voter interference. *Contemporary Politics* 25, 2 (2019), 150–171.

[18] Cheng Suang Heng, Zhijie Lin, Xiaoying Xu, Ying Zhang, and Yixuan Zhao. 2019. Human flesh search: what did we find? *Information & Management* 56, 4 (2019), 476–492.

[19] Thomas J Holt, Jin R Lee, and Olga Smirnova. 2023. Exploring risk avoidance practices among on-demand cybercrime-as-service operations. *Crime & Delinquency* 69, 2 (2023), 415–438.

[20] Les Hutchinson. 2018. Wielding power and doxing data: How personal information regulates and controls our online selves. In *The Routledge handbook of digital writing and rhetoric*. Routledge, 303–316.

[21] Younes Karimi, Anna Squicciarini, and Shomir Wilson. 2022. Automated detection of doxing on twitter. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.

[22] Batuhan Kukul. 2023. Personal data and personal safety: re-examining the limits of public data in the context of doxing. *International Data Privacy Law* 13, 3 (2023), 182–193.

[23] Massimo La Morgia, Alessandro Mei, Alberto Maria Mongardini, and Jie Wu. 2021. Uncovering the dark side of Telegram: Fakes, clones, scams, and conspiracy movements. *arXiv preprint arXiv:2111.13530* (2021).

[24] Derek Manky. 2013. Cybercrime as a service: a very modern business. *Computer Fraud & Security* 2013, 6 (2013), 9–13.

[25] Tina Marjanov and Alice Hutchings. 2025. SoK: Digging into the Digital Underworld of Stolen Data Markets. In *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE, 1–18.

[26] Per Håkon Meland, Yara Fareed Fahmy Bayoumy, and Guttorm Sindre. 2020. The Ransomware-as-a-Service economy within the darknet. *Computers & Security* 92 (2020), 101762.

[27] Dawei Meng, Lei Zhang, and Long Cheng. 2013. A Study of Human Flesh Search Based on SIR Flooding on Scale-Free Networks. In *International Conference in Swarm Intelligence*. Springer, 369–376.

[28] Artanti Tertia Mukti, Mochammad Tanzil Multazam, Emy Rosnawati, and Sarykulov Kurmangaly. 2024. Doxing Patterns Using Social Engineering in Cyberspace. In *2nd International Conference on Advance Research in Social and Economic Science (ICARSE 2023)*. Atlantis Press, 530–546.

[29] PCPD. 2025. A Female Arrested for Suspected Doxxing Arising from Online Shopping Dispute. https://www.pcpd.org.hk/english/news_events/media_statements/press_20250310.html.

[30] Reuters. 2025. China's Baidu denies data breach after executive's daughter leaks personal info. https://www.reuters.com/technology/cybersecurity/chinas-baidu-denies-data-breach-after-executives-daughter-leaks-personal-info-2025-03-20/.

[31] Dorsaf Sallami and Esma Aïmeur. 2024. Fndex: fake news and doxxing detection with explainable AI. *arXiv preprint arXiv:2410.22390* (2024).

[32] Guohou Shan, Wenxi Pu, Jason Bennet Thatcher, and Philip Roth. 2024. How Doxing on Social Media Leads to Social Stigma and Perceived Dignity. (2024).

[33] James Shires. 2024. Civil society in cyberwarfare: hack-and-leaks, attribution and mobilization. In *Research Handbook on Cyberwarfare*. Edward Elgar Publishing, 167–182.

[34] Peter Snyder, Periwinkle Doerfler, Chris Kanich, and Damon McCoy. 2017. Fifteen minutes of unwanted fame: Detecting and characterizing doxing. In *Proceedings of the 2017 internet measurement conference*. 432–444.

[35] SpyCloud. 2025. A Deep Dive Into the Intricate Chinese Cybercrime Ecosystem. https://spycloud.com/blog/deep-dive-chinese-cybercrime-ecosystem/#elementor-toc__heading-anchor-2.

[36] Telemetr.io. 2025. Telemetr.io: Telegram Channel Statistics. https://telemetr.io.

[37] TGStat. 2025. TGStat: Telegram Analytics. https://tgstat.com.

[38] Mattia Thibault et al. 2016. Trolls, hackers, anons: Conspiracy theories in the peripheries of the web. *Lexia* 23 (2016), 387–408.

[39] Six Tone. 2025. 'Zero Tolerance': China Cracks Down on Doxxing. https://www.sixthtone.com/news/1017157.

[40] Daniel Trottier. 2017. Digital vigilantism as weaponisation of visibility. *Philosophy & technology* 30, 1 (2017), 55–72.

[41] Daniel Trottier. 2020. Denunciation and doxing: Towards a conceptual model of digital vigilantism. *Global crime* 21, 3-4 (2020), 196–212.

Appendix

Figure 3 provides a visual representation of a typical DaaS channel homepage, showcasing the integration of multiple bot endpoints, customer service contacts, and pricing information.

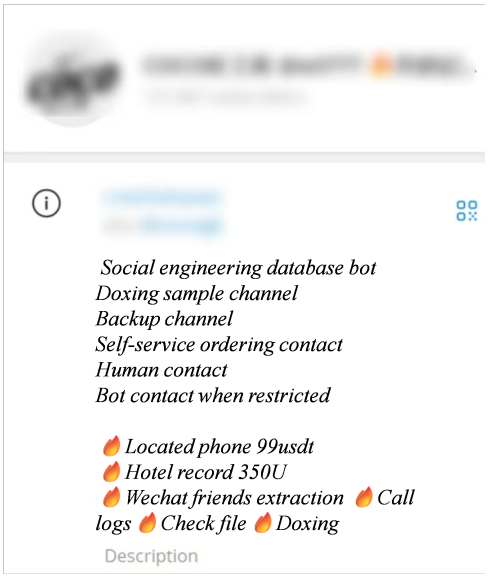

Figure 3: A typical DaaS channel homepage

Figure 4 shows a typical social-engineering bot pipeline. The interaction starts with the user sending the `/start` command and

joining required channels and groups. After gaining access to the bot interface, users earn query points through a points-based system (e.g., daily check-ins and referrals) and submit keywords via multiple query modes. The pipeline ends by either returning corresponding sensitive personal information or redirecting the user to a manual investigation channel.

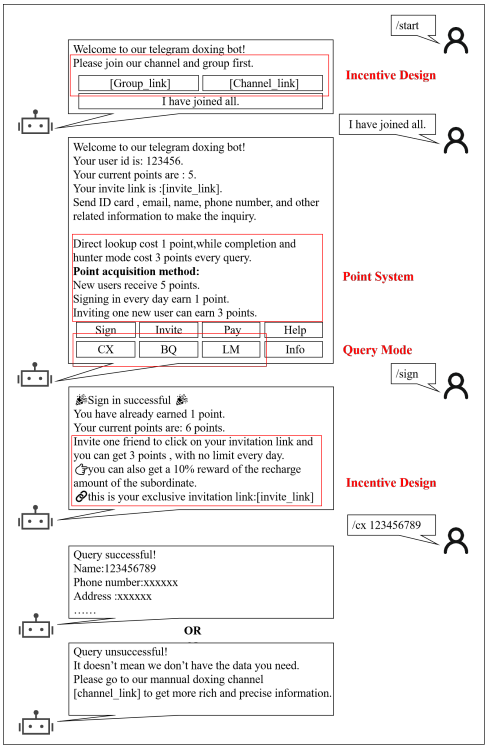

Figure 4: An illustration of the user-bot interaction pipeline.

Figure 5 illustrates a typical ad message format with an example of "WeChat-ID reverse lookup" service.

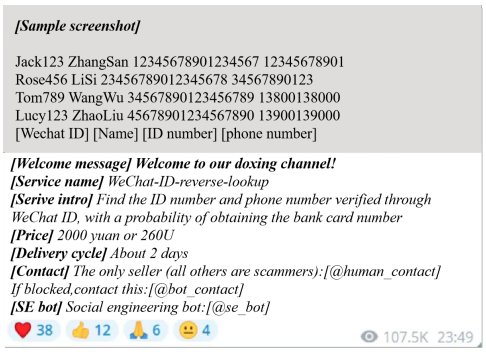

Figure 5: A typical ad message format.