Like, Comment, Get Scammed: Characterizing Comment Scams on Media Platforms

Xigao Li
Stony Brook University
xigli@cs.stonybrook.edu

Amir Rahmati
Stony Brook University
amir@cs.stonybrook.edu

Nick Nikiforakis
Stony Brook University
nick@cs.stonybrook.edu

***Abstract*—Given the meteoric rise of large media platforms (such as YouTube) on the web, it is no surprise that attackers seek to abuse them in order to easily reach hundreds of millions of users. Among other social-engineering attacks perpetrated on these platforms, comment scams have increased in popularity despite the presence of mechanisms that purportedly give content creators control over their channel comments. In a comment scam, attackers set up script-controlled accounts that automatically post or reply to comments on media platforms, enticing users to contact them. Through the promise of free prizes and investment opportunities, attackers aim to steal financial assets from the end users who contact them.**

In this paper, we present the first systematic, large-scale study of comment scams. We design and implement an infrastructure to collect a dataset of 8.8 million comments from 20 different YouTube channels over a 6-month period. We develop filters based on textual, graphical, and temporal features of comments and identify 206K scam comments from 10K unique accounts. Using this dataset, we present our analysis of scam campaigns, comment dynamics, and evasion techniques used by scammers. Lastly, through an IRB-approved study, we interact with 50 scammers to gain insights into their social-engineering tactics and payment preferences. Using transaction records on public blockchains, we perform a quantitative analysis of the financial assets stolen by scammers, finding that just the scammers that were part of our user study have stolen funds equivalent to millions of dollars. Our study demonstrates that existing scam-detection mechanisms are insufficient for curbing abuse, pointing to the need for better comment-moderation tools as well as other changes that would make it difficult for attackers to obtain tens of thousands of accounts on these large platforms.

I. Introduction

Media platforms such as YouTube have been extremely popular among Internet users, attracting hundreds of millions of viewers per year [31]. Naturally, because of the large concentration of users on a small number of platforms, these platforms have become targets of cyber criminals who try to expose users to a variety of social-engineering attacks, including phishing and scams [15], [35], [2].

One of the most recent attacks targeting the users who read and post comments are so-called comment scams [38], [6]. In comment scams, scammers use script-controlled accounts to create comments or replies on media platforms, enticing users to contact them through text messages for the chance to receive a gift or participate in an investment opportunity. The comments can vary in length, from a simple phone number to multiple paragraphs that extol the virtues of specific (and fictitious) investment advisers. Once users engage with scammers, they will reach a person conducting the next step of scam campaigns that convince them to participate in a fake investment or pay shipping charges for their free prize.

Despite the growing prevalence of comment scams, there has been no systematic study of the ecosystem surrounding these attacks by the research community. While YouTubers and media outlets have drawn attention to specific instances of comment scams [5], the larger picture of the tactics and techniques used by scammers remains unknown. Information about comment scams is mostly limited to the reports of individual victims [16], [17], and the creators targeted by impersonation activities [44]. This lack of systematic investigation of comment scams presents a significant challenge to those seeking to detect and stop these attacks.

In this paper, we perform a three-part study to understand and characterize comment scams. Given its market share, we focus on YouTube, where we aim to gain insights into the tactics and payment channels used by scammers who target its users. First, we design and implement a system to longitudinally collect comments posted under videos. Our infrastructure periodically collects comments posted under monitored videos as "snapshots", allowing us to monitor the comment dynamics, *i.e.,* the creation and deletion of comments, as well as the status of the accounts that post them (*e.g.,* detect account-deactivation events). This enables us to gain a clear understanding of the evolving nature of comment scams on YouTube. Using our infrastructure, we monitor 20 different channels over 6 months, capturing a total of 8.8 million comments.

Second, we design three filters that take advantage of scam comments' textual, graphical, and temporal features. We identified 206,306 comments that exhibit scam behavior by applying these filters to our dataset. We find that scammers evade existing scam-detection mechanisms through various tactics, such as utilizing visually similar symbols to obfuscate their text, abusing account names, and splitting text into multiple comments posted by multiple accounts. Through detecting similar profile images, we also find evidence that scammers

Network and Distributed System Security (NDSS) Symposium 2024
26 February - 1 March 2024, San Diego, CA, USA
ISBN 1-891562-93-2
https://dx.doi.org/10.14722/ndss.2024.24060
www.ndss-symposium.org

are impersonating channel owners by abusing their names and profile images. Moreover, we merge scammer accounts through their common contact information and present the resulting campaigns of scam activity. We found large campaigns that used more than a hundred YouTube accounts to promote WhatsApp phone numbers, as well as campaigns that promote phone numbers across 8 different channels. From the perspective of comment dynamics, we find only 31.42% accounts were deactivated in our monitored 6-month period, indicating their evasion tactics effectively avoid existing detection mechanisms.

Finally, we perform an IRB-approved study where we directly interact with scammers to gain insights into tactics and payment channels used by scammers. We collect a total of 50 conversations with scammers through WhatsApp and Telegram. By pretending to be unaware victims, we let scammers progress through their scams while recording the conversations. Our study uncovered two major scam activities from the collected conversations: cryptocurrency investment scams and fake prize scams. We find that scammer conversations can last for days, as well as evidence indicating that scammers using a United States phone number work in different time zones than those in the United States. Furthermore, we observe that current online blocklists contain almost none of the websites used by comment scammers. We find scammers are polite but cautious as they frequently request screenshots as proof and prefer cryptocurrencies as their main payment instrument. While the anonymous nature of cryptocurrencies makes it non-trivial to identify them, we take advantage of publicly accessible blockchains to track their transactions. Unlike prior cyber-crime research that can only estimate the amount of funds stolen based on several assumptions, we are able to calculate the exact amount of funds stolen by these 50 scammers, which is worth approximately 2 million US Dollars. Moreover, we show that the scammers are highly convincing in their scams, finding that 93.5% of scammer-provided cryptocurrency wallet addresses had at least one incoming transaction.

Overall, we make the following contributions:

- We design infrastructure that can dynamically capture comments on the YouTube platform, including deleted ones. We collected a large dataset of scam comments over a 6-month period of study.
- We design filters to uncover the textual, graphical, and temporal features of comment scams, producing a dataset of scam comments and analyzing scam-comment dynamics and behavior.
- We perform an IRB-approved study directly interacting with scammers, gaining valuable insights into their tactics and payment channels.

To enable future research in the weaponization of large media platforms (particularly as it relates to scams), we make our code that captures scam comments on the YouTube video platform available at https://like-comment-get-scammed.github.io/.

II. Background

Comment scams are initiated through script-controlled programs that create comments and replies under media

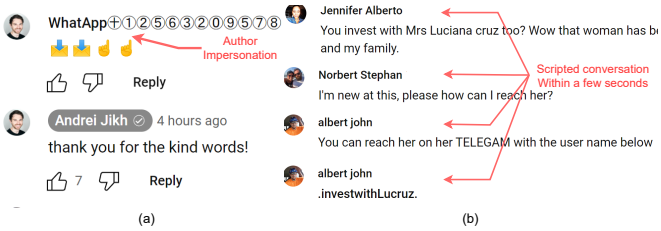

Fig. 1: *Examples of comment scams in media platforms. (a) Shows an example of a scam by impersonation. The comment shown on the bottom with a dark background and a checkmark belongs to the channel owner, whereas the comment on the top is a scammer impersonating the channel owner, advertising a WhatsApp account. (b) Shows a made-up conversation from multiple scammer accounts, advertising a Telegram account.*

platforms such as YouTube, enticing users to send a text message to claim a prize or add a user on other platforms in order to participate in a lucrative investment opportunity. These comments can be as brief as one emoji or as long as multiple paragraphs that recount a fabricated story. Scammers use social engineering techniques, such as pretending to be popular channel owners advertising fake personal contact information, or posing as regular users recommending a person that helps people invest with unrealistically high returns. Figure 1 provides two examples of comment scams.

By abusing publicly available APIs to post comments under popular videos and lure users, a single scammer can target users across thousands of videos. While YouTube has implemented certain comment regulation mechanisms, scammers use various tactics to work around them, such as using visually similar symbols (VSS) to evade text-based checks or splitting scripts into multiple segments and sending them with multiple accounts. For example, instead of using ASCII characters, scammers abuse visually similar Unicode symbols (VSS). As shown in Figure 1(a), the top user with the name “WhatApp” uses a VSS “Mathematical Sans-Serif Bold Small” instead of ASCII Latin letters. Although keyword-based detection cannot capture such characters, users can easily interpret these symbols as numbers and contact the scammer. In Figure 1(b), scammers omit letters or replace keywords such as telegram usernames with visually similar symbols and keep the non-sensitive conversations as regular Latin/English characters.

In addition to the obfuscation of text, impersonating channel owners in scammers’ comments is also a common practice used by scammers. To achieve this, scammers simply need to save the channel owner’s profile image and apply it to their own account. As shown in Figure 1(a), YouTube has implemented new features to combat impersonation, such as a checkmark next to the authenticated channel owner’s username with a highlighted background. Despite these efforts, inexperienced users may still have difficulty distinguishing between authentic and fake comments, particularly when there are no nearby comments from the channel owner to compare them to.

Once users contact the provided phone numbers or user-names, they are directed to a person (scammer) conducting the second phase of the scam. Scammers defraud users through

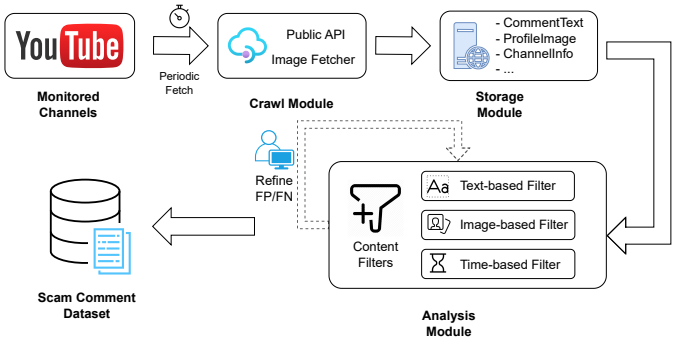

Fig. 2: *Crawler Design.*

various schemes, such as pretending to be celebrities, popular channel owners, or investment experts. They try to persuade users to invest their funds in cryptocurrencies by promising high returns, or convincing them that they have won a prize and must pay for shipping. Additionally, scammers may show fake evidence of other users earning large sums of money, urging users to participate while the opportunity is still available to them. Once the user is convinced, scammers proceed to the final phase of the scam campaign by providing payment options for users to deposit their funds. The scenario outlined above often results in financial loss for the victim, ranging from a few hundred dollars to tens of thousands of dollars. Unlike traditional bank transfers or credit card charges, the decentralized nature of cryptocurrency transactions makes it impossible to reverse these transactions once they are completed.

III. Crawler Design

To capture scammer activities, we developed an infrastructure to capture and analyze YouTube comments to detect and analyze scam comments, which is shown in Figure 2. The system is composed of three modules: crawl module, storage module, and analysis module.

A. Threat Model

Our threat model is focused on capturing and detecting scam comments on YouTube. Scam comments are typically posted by malicious actors who seek to defraud or deceive other users on the platform. While it is possible to conduct some of the scams manually, all evidence points to automated scripts that control various accounts. The scammers post unsolicited contacts such as WhatsApp or Telegram to lure unaware users to get in touch with them. Then, scammers may conduct the second phase of the scam campaign, which defrauds users by using social engineering techniques such as introducing fraudulent investments, soliciting personal information, and persuading users to send funds.

YouTube currently has regulations in place for posting comments that violate the community guidelines [43], [45], for example, comments that offer cash gifts or "get rich quick" schemes [42]. The platform uses automated, machine-learning-based systems to detect and remove comments that violate these policies. As we will show in this paper, despite these measures, there are still thousands of accounts that post comments

MESSAGE ME ON TELEGRAM +1234	(ASCII latin)
MESSAGE ME ON TELEGRAM ⁺¹²³⁴	(Latin Letter Small Capital Unicode)
whatsapp 1234	(ASCII latin letters)
whatsapp 1234	(Mathematical Sans-Serif Small Unicode)

Fig. 3: *Example of VSS frequently used by scammers. Some VSS are difficult to distinguish in the eyes of inexperienced users.*

with contact information. Scammers are constantly adapting their tactics and finding new ways to bypass automated detection systems, such as using subtle character variations and misspellings. Our goal is to capture these scam comments, differentiate them from the comments of regular users, and further investigate the patterns of these scam activities.

B. Module Design

Crawling module: The crawling module is responsible for periodically capturing comments from different YouTube channels using the Google YouTube API.

One of the challenges in capturing YouTube comments is the interactive nature of the media platform. Users can delete their own comments, and channel owners can regulate or remove comments posted under their videos. As a result, a single crawl of a video's comments may not be sufficient to capture the complete set of comments or reflect their growth over time. To address this challenge, we captured periodic "snapshots" of the comments every hour. By taking hourly snapshots, we capture the comments that have been posted since the last crawl and retain deleted comments in the database for future analysis. The snapshot mechanism allows us to build a comprehensive dataset that reflects the dynamic nature of YouTube comments and ultimately provides a more accurate representation of scammer activities over time.

Storage module: The storage module stores the comments in a database for further analysis. The database is designed to store comments from different channels separately, enabling efficient processing of comments from a specific channel. The storage module also provides an interface to retrieve comments from the database for further analysis. We store both comments and their corresponding metadata, including comment creation time, user channel ID, and profile images.

Analysis module: We designed the analysis module to detect and filter scam comments from captured data. The module is designed to analyze the comments to detect patterns such as phone numbers, conversations, and account names associated with scam activities. We use three different filters to capture behaviors from scam comments: Text-based filter, Image-based filter, and Time-based filter. To capture the scam comments with higher accuracy, we applied a snowball refining approach by manually refining the criteria of filters [12], which helps the filter to include more scam comment variations as well as to exclude false positives.

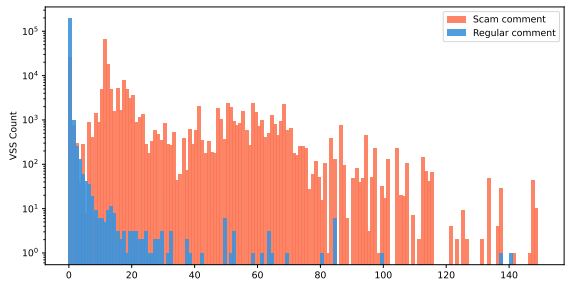

Fig. 4: *Distribution of Visually Similar Symbol (VSS) in scam comments and regular user comments.*

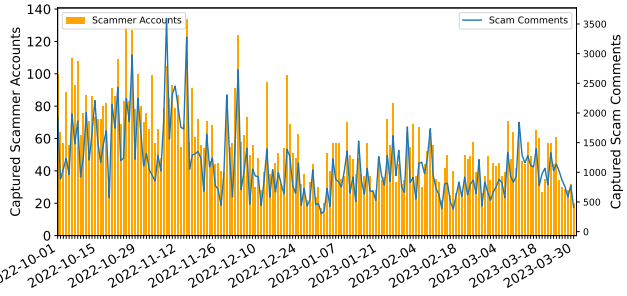

Fig. 5: *Captured scam comment traffic. Our study captured a total of* 206,306 *scam comments posted from* 10,541 *scammer-controlled accounts (an average of* 1,140 *scam comments posted each day).*

C. Filter Design

Text-based filters: The text-based filter targets the presence of visually similar symbols (VSS) and specific keywords, as well as the inclusion of phone numbers. As shown in Figure 3, instead of directly posting comments with ASCII letters, scammers can replace those letters with VSS, such as replacing the "Latin capital letter M" (U+004D) with "Unicode Latin letter small capital M" (U+1D0D). The replacements are meant to throw off automated keyword detection algorithms but be fully legible to human users. We start with an initial set of scam comments and extract common VSS alphabets, then expand and refine them using a snowball approach with a larger sample until we reach a saturation point of scam comments and false positive rates. Using this approach, we identified at least 40 different variations of VSS alphabets that can be used by scammers. Furthermore, scammers may combine different Unicode letters from each alphabet to compose more complicated messages. Figure 4 shows the distribution of VSS between scam comments and benign user comments that our system discovered.

Image-based features: In addition to text-based features, we also utilized image-based filtering to detect scam comments conducting impersonation. We employed perceptual hashing techniques [27] to generate image hashes of both the comment author and channel owner and compared the two to determine if they were highly similar.

Time-based features: In addition to obscuring the text and impersonating the channel owners, we discovered that scammers use advanced tactics to evade platform detection and improve their credibility with potential victims. One such tactic involves splitting text into multiple paragraphs and sending it with multiple accounts under their control, simulating a conversation between different users. To effectively capture these comments, we designed a time-based filter. We first analyze the time period of scam comments captured in text-based filters and image-based filters and calculate their reply time interval from the thread or previous reply. We apply the concept of "session" to group comments under a thread, dividing the comments if there were more than 15 seconds of inactivity between two consecutive comments. The filtered comments contain false positives, such as multiple users replying to a thread simultaneously. To refine our results, we apply a keyword-based check, which includes searching for the presence of phone numbers or Telegram usernames.

D. Data Collection

We deployed our system to monitor a total of 20 YouTube channels based on their popularity, which includes the number of views, subscribers, and comments posted on their videos. Among those channels, 10 are related to financial topics such as investments and wealth management. To make a comparison, we selected another 10 channels as the baseline group. Our hypothesis was that financial-related channels are more prone to being targeted by scammers as viewers of such channels are more likely to respond to an opportunity to increase their income. To establish a reliable comparison group, we selected baseline channels that we considered less likely to be targeted by scammers. These channels cater to viewers who are generally not seeking financial-related content and, hence, are less likely to fall for financial scams in the comments.

We selected a diverse range of channels from five different categories to form the baseline group. These categories include sports (2 channels), cooking (2 channels), politics/news (2 channels), education (2 channels), and gambling (2 channels). By using baseline channels, we aim to gain a deeper understanding of the unique characteristics of scam campaign targets. We have included a comprehensive overview of YouTube channels that we monitored in Table VI in Appendix A.

IV. Dataset Analysis

In this section, we provide a detailed analysis of our captured dataset, which we collected over a 6-month period from October 1^{st}, 2022 to March 31^{st}, 2023. The dataset includes a total of 8,801,224 individual comments from 20 YouTube channels, captured across 428,350 snapshots of 8,226 videos. Among those comments, we identified 206,306 (2.34%) comments that clearly belong to scammers.

A. Captured Scam Comment Activities

Daily Captures: Figure 5 displays the number of scam comments we captured during our experiment, which indicates the overall activity of these scammers' accounts. On average, we identified 1,140 new scam comments being posted each day. In terms of new scammer account creation, we identified a total

TABLE I: *Scam comments captured in each channel.*

Category	Monitored Channels	Total Scam Comments	Average Scam Comments Per Channel
Finance	10	148,070	14,807
Sports	2	2,050	1,025
Cooking	2	29,271	14,635.5
News/Politics	2	25,126	12,563
Education	2	1,486	743
Gambling	2	303	151.5
Total	20	206,306	-

of 10,541 YouTube accounts controlled by scammers, which means an average of 58 new accounts were observed each day.

Phone numbers: Most scam comments we captured during our experiment included contact information. As mentioned in Section II, scammers use visually similar symbols to bypass existing detection mechanisms, so it is difficult to extract them from the text. The most common cases are phone numbers that can be used as WhatsApp numbers. We successfully extracted 2,814 different phone numbers from 5,594 scam comment accounts. We also successfully extracted 272 Telegram usernames from 2,643 scam comment accounts. In addition to comments with contact information, we also identified scammer accounts that did not include any contact information but were still used by scammers to split their script and convince users, such as recommending a financial advisor or posting a fake story.

We used the country code of phone numbers to infer their origin. Out of 2,814 phone numbers that we collected from scammer comments, we successfully identified the origin country for 2,227 phone numbers. A majority (88.37%) of phone numbers originated in the United States, followed by Canada (2.92%) and the Dominican Republic (1.84%). Based on our interaction with scammers (Section V), we conclude that the majority of them reside outside the United States. As such, US phone numbers merely lend an air of legitimacy to their scams.

Captured comments by Category: Our study analyzed scam comments across 20 channels from 6 different categories. A brief summary of scam comments is shown in Table I. The results showed that the Finance category received the highest number of scam comments, followed by Cooking and News/Politics. In contrast to those highly active scam comment channels, the Sports, Education, and Gambling categories received significantly less attention from scammers. We also observed targeted impersonation tactics, such as using the image and name of the channel owner to lure unsuspecting users, which we report in Section IV-C. The scams in the cooking and politics categories are unexpected, as these channels are unrelated to financial topics. This indicates that scammers tend to conduct their campaigns over popular channels (*i.e.,* large numbers of subscribers, views, and comments). Despite the fact that the impersonation activity is usually targeted toward celebrities, we found that scammer activity was not necessarily correlated with the number of subscribers. For example, two

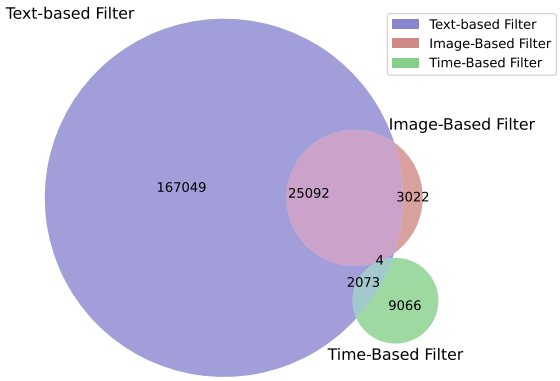

Fig. 6: *Venn diagram of captured comments with different filters. There are intersections with each filter, which means a comment may use multiple techniques to evade platform regulations.*

education channels that attracted the fewest scam comments had 7.52M and 14.6M subscribers, respectively, higher than the average number of subscribers in the Finance category.

Figure 6 shows the contribution of each of our three filters to the detection of scam comments. The Text-based filter captured the majority of 194,218 (94.14%) comments. Following that, the Image-based filter captured a total of 28,118 (13.63%) comments. Finally, the Time-based filter captured 11,143 (5.4%) comments, which were included in 2,957 groups of conversations. It is important to note that a single comment can be labeled as scam by multiple filters, resulting in intersections between filters. For example, a total of 25,096 comments in our dataset were captured by both the Text-based and Image-based filters.

We also evaluate the effectiveness of filters by video channel category and present the result in Figure 7. Across all categories, the text-based filter was most effective. We discovered that the image-based filter was more effective in the Cooking and Gambling channels, indicating that scammers make more impersonation attempts in these channels. Conversely, the conversation-based filter was more effective in the Sports (13.46%) and News/Politics channels (26.27%). This suggests that scammers are targeting a more general audience in these channels rather than impersonating channel owners.

B. False Positives and False Negatives

After designing our filters, we refined them through multiple iterations to minimize false positives. We randomly sampled 300 scam comments and found 2 (0.67%) false positives (*i.e.,* user comments were flagged as scam comments). The false positives could be due to users accidentally introducing phone numbers or using symbols to form a series of numbers. Furthermore, we investigated the false positive rate and false negative rate among deleted comments. We randomly sampled 300 comments from deleted comments that were marked as scam comments, as well as 300 comments from deleted comments that were marked as regular comments. While we find all deleted scam comments are true positives, we find 3 false negatives from deleted regular comments. From our manual inspection, deleted regular comments are either in violation of YouTube's community

guidelines [45], or removed by corresponding channel owners. Lastly, we randomly sampled another 300 comments from all deleted comments, where we found 4 (1.33%) true positives, 295 (98.33%) true negatives, 0 false positives, and 1 (0.33%) false negative. Overall, although the filters have a very low false positive rate and false negative rate, we do not recommend using the filters for automatic comment deletion. Instead, we suggest flagging the comments and prioritizing them for further investigation. This would allow for human review and reduce the risk of mistakenly deleting legitimate comments.

C. Filtered Comments

Textual features: We discovered 168,938 (81.89%) scam comments contained at least one VSS, which was likely used to circumvent automated checks on the YouTube platform. Additionally, we observed that 166,988 (80.94%) scam comments included at least one emoji. In terms of "likes" (i.e. up-votes) received by scam comments, we observed that 3,851 (1.87%) of scam comments received at least one "like". Assuming that the vast majority of users are not going to "like" scam comments and since one account cannot like a comment twice, with 2,536 likes to a single scam comment, this means that scammers have thousands of accounts available to artificially increase a comment's popularity. We also observed that scammers abuse account usernames for advertising contacts, such as WhatsApp phone numbers and Telegram accounts. We found at least 4,802 (45.56%) scammer accounts abusing the username for scam activity. By including contact information in the username with VSS obfuscation and posting different comments each time, scammers can evade existing comment text detection algorithms and maximize exposure to their potential victims. Overall, our text-based features made scam comments distinguishable from those of regular users.

Graphical features: Through our Image-based filter, we discovered that 28,118 (13.63%) scammer accounts used the same or similar profile images as the corresponding channel owners. This suggests that these scammers are not simply employing accounts by sending messages across random popular videos but rather targeting specific channels by customizing these accounts to increase their success rate.

Temporal features: As shown in Figure 8, most replies happened within 15 seconds from the previous comment, so we set the "session" split time as 15 seconds to capture the majority of scam comments. We ensured the filtered conversations were from scammers by randomly sampling and manually analyzing the contents. In total, we captured 2,957 groups of conversations from scammers, containing a total of 11,143 comments. From the content of these conversations, we discovered that most scammers used a group of accounts that pretended to have succeeded in investments, and they then asked for recommendations for a financial advisor or broker before finally using another account to post the scammer's contact information.

D. Scam Campaigns

Contact information such as phone numbers and Telegram user names play a crucial role throughout the entire scamming process. To maximize exposure to potential victims and avoid the risk of account deactivation, scammers often utilize multiple accounts to advertise their contact information in various locations. To capture this relationship, we merged the scammer accounts that advertised the same contact, then calculated their posted comments and targeted channels. Table II summarizes the 10 largest campaigns we identified. The largest campaign included 112 different accounts, and the most widespread campaign targeted 324 videos in four different channels. Scammers also targeted as many as eight different channels in a single campaign. It is worth noting that a single scammer can purchase multiple phone numbers and advertise multiple scam campaigns. However, tracking these phone numbers is difficult, so we were unable to connect them in our analysis. Later in Section V, we interact with scammers directly and present evidence of similar chat scripts, indicating that the same person may be controlling different phone numbers and Telegram accounts. Our results highlighted the evidence that scammers utilize multiple accounts to maximize their exposure to potential victims, and we connect seemingly unrelated account behavior and provide insight into the tactics used by scammers to reach their targets.

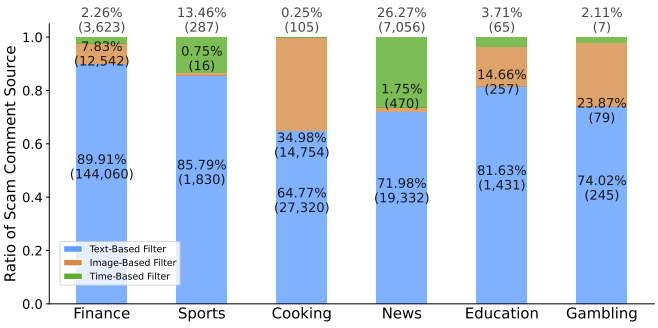

Fig. 7: *Filter effectiveness in each category. The X-axis represents different video categories, and the numbers on each bar represent the number and the ratio of comments being captured through that filter. Generally, the text-based filter is the most effective approach, but the number of comments captured by the image-based filter and time-based filter varies between channels.*

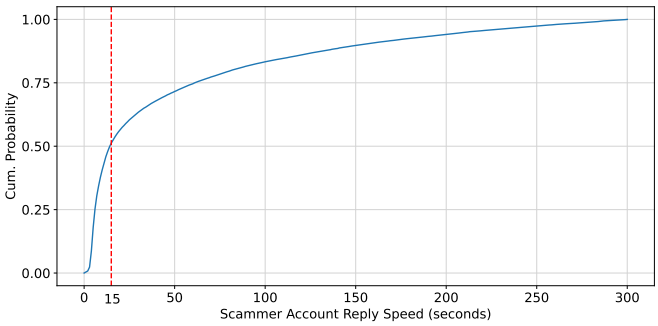

Fig. 8: *CDF of reply interval of scam comments. Most scammers post replies within 15 seconds of the thread (parent) comments.*

TABLE II: *Top 10 scam campaigns. Scammers can control hundreds of accounts to promote a single phone number or promote the same phone number across multiple channels.*

Campaign ID	Accounts	Comments Posted	Affected Videos	Targeted Channels	Affected Categories
1	112	4,045	92	1	Finance
2	59	703	324	4	News/Politics, Finance
3	46	5,405	66	2	Finance
4	45	692	321	4	News/Politics, Finance
5	44	5,662	76	2	Finance
6	39	4,435	46	2	Finance
7	39	3,880	57	2	Finance
8	35	67	40	6	Cooking, News/Politics, Finance
9	33	4,573	68	2	Finance
10	33	2,951	35	2	Finance

E. Comment Dynamics

Frequently used contents: In Figure 9, we present a word cloud that displays the most frequent words used by scammers in their comment scams. We use the unigrams, bigrams, and trigrams of the text to capture the most popular words/phrases, replacing the space character with an underscore to better differentiate the individual words with 2-3 words of a phrase. The size of each word in the cloud represents the frequency with which it appeared in the comments. As previously reported in Section IV-C, scammers frequently abuse the account username to increase the likelihood that victim users will interact with them. As a result, we also included the usernames of scam comments in the text corpus used to generate the word cloud. In terms of content, we observed three major patterns. The first pattern simply spreads the contact information and urge viewers to send a message to scammers. The second pattern includes a short story intimating the potential high profit of an investment, whereas the last pattern promotes the fake pre-sale of a specific NFT. Scammers often use specific words to urge and convince users to participate in their scam. For instance, they may use words that make users believe that the channel owner is responding to their comments or solicit users to participate in a financial investment project. Figure 10 displays the number of scam comments that were posted during different hours of the day. Interestingly, we observed that most scam comments appeared between the hours of 22:00 to 0:00 Eastern Time. This may have to do with the scammers' timezone or could be attempts to use two days' worth of API quotas in a single burst.

Fig. 9: *Word cloud based on the text of scam comments. The spaces between words have been replaced to underscore for better visibility. Scammers prefer to leave short and obscure messages to attract victims for contact.*

Deleted comments: Our periodic snapshots capture the creation and deletion of comments. To identify deleted comments, we compare the timestamps of comments to their related video snapshot timestamps. If comments appeared in one or more snapshots but were not present in subsequent snapshots, we mark them as deleted. Throughout our dataset, we identified 123,506 (59.87%) scam comments that were deleted. Figure 11 displays the cumulative distribution function (CDF) of the deleted scam comments. As shown in the figure, most deletions happen within one day. Scam comments can be deleted by three different parties: the comment author, the channel owner, or the video platform (*i.e.,* YouTube). It is important to note that scammers often choose to reply to a thread from a regular user. In this case, even if the comments were deleted and other viewers cannot see them, the user who received the reply will still get a permanent notification, which does not disappear after the deletion of the comment.

F. Scammer Account Activities

Username Change: During our study of YouTube comment scams, we observed the dynamic behavior of scammer account activities. Out of 10,541 accounts we captured, we found that 1,298 (12.31%) of them changed their usernames over time. Some accounts changed their names as many as 17 times during the study period and posted different comments each time. We believe that those accounts change their IDs to better match the channels they are targeting.

Account Deactivation: Given the long lifespan of scammer-controlled accounts, we also investigated the deactivation of those accounts. Despite posting scam comments frequently, only

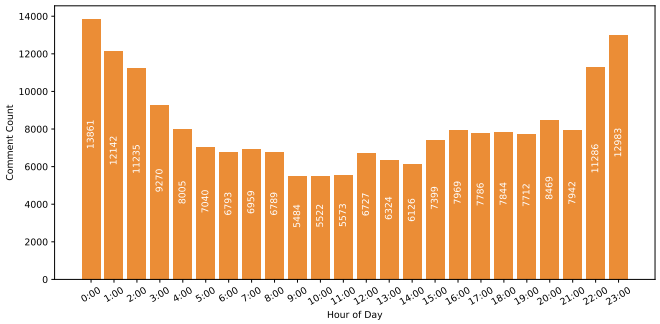

Fig. 10: *Scam comment activity throughout the day (UTC -5). Scammers prefer time at 0:00 to send messages, which may be due to the API quota reset or their true geographical location.*

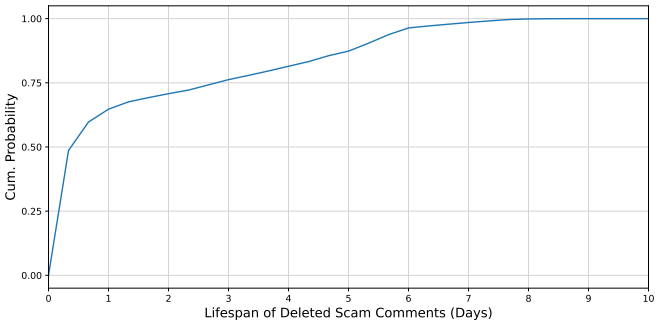

Fig. 11: *CDF of the lifespan of deleted scam comments. Most scam comments were deleted within a day after they were published.*

3,312 (31.42%) scam accounts out of the 10,541 scam accounts we captured during the six-month period were deactivated. This suggests that scammers are able to evade platform detection techniques and can keep these scammer accounts active for a considerable amount of time. The longevity of scammers' accounts is a key factor in the success of scam campaigns, as it reduces the cost of registering new accounts.

V. Interacting with Scammers

So far, our collected dataset has enabled us to gain insights into the prevalence and characteristics of scam comments on video platforms such as YouTube. Yet, by itself, our comment dataset does not shed light on what will happen to victim users who engage with these scammers by contacting them, believing that they are channel owners or financial advisors.

To gain a deeper understanding of the tactics used by comment scammers, we conduct an IRB-approved experiment where we directly interact with scammers. In this experiment, we posed as unsuspecting victims and conversed with scammers via messaging applications, such as, WhatsApp and Telegram. These conversations were recorded and analyzed to identify common patterns and payment channels utilized by scammers. Through these interactions, we discover common text scripts and scam schemes that scammers use to persuade users to send their funds.

Our research has yielded an important discovery regarding the payment channels favored by scammers. Specifically, our research reveals that scammers prefer cryptocurrencies as a payment channel. This has also enabled us to accurately calculate the amount of funds obtained by scammers without making assumptions or estimations. By analyzing publicly accessible blockchain networks such as Bitcoin and Ethereum, we can straightforwardly calculate the total number of transactions and determine the exact funds stolen by scammers, along with their corresponding value in US dollars.

A. Interaction Setup

Data Source: The data source of our study comes from the filtered result of our dataset. We randomly select publicly available contact information of scammers and ensure that we do not select them more than once. We then manually verified that these selected comments violated YouTube's Term of Service that defines "Spam, deceptive practices, & scams policies" [42] and "Impersonation policy" [13]. Specifically, we verify that the comment is from a scammer-controlled account, which either impersonated the channel owner using a similar image and username or the comment is advertising false financial advice that promised unrealistic profits. By selecting only these types of comments, we are confident that we did not inadvertently interact with legitimate users.

Based on the sampled contact information we selected, we use two mobile applications that are preferred by scammers: WhatsApp [40] and Telegram [34]. We exclusively utilize the text message function within the mobile application to interact with scammers while refraining from using voice or video calling functions as well as SMS messaging functions from the mobile phone. In the event that a scammer attempts to contact us via SMS or initiates a voice/video call, we do not respond or engage with these forms of communication.

IRB Approval: Because we interacted with users (the scammers), we sought and secured IRB approval from our institution. Our experiment does not involve any risky methods or physical contact with scammers; Instead, we focus solely on observing their methods to defraud average individuals. We applied to our institute's IRB and acquired permission to perform and record text conversations with scammers while pretending to be unsuspecting users. Additionally, we do not want scammers to be aware of our study and share the information with other scammers, which may interfere with our future conversations and impact the ecological validity of the study. Hence, we obtained a waiver from IRB regarding the debriefing of scammers at the end of our conversations. The IRB approved our approach of deception and waived the requirement of consent, *i.e.,* we do not need to reveal our true identities or intent to scammers in our entire interaction with scammers, and we do not need to ask scammers if they agree to participate our study.

The deceptive nature of the study will avoid measurement artifacts of our own intervention by tampering with the scam ecosystem while studying it. Since we limit our study to polite, text-only conversations that do not seek to uncover private user information, our study does not incur any risk to the subjects' (*i.e.,* scammers') emotional, psychological, or physical well-being since they already have these conversations with real victims on a daily basis.

TABLE III: *Summary of our interaction with scammers. Scammers prefer cryptocurrency investment scams and conducting scam activities over the WhatsApp platform.*

Scam Scheme	WhatsApp	Telegram	(Total)
Cryptocurrency Investments	31	7	38
Fake Prize	0	11	11
Other	1	0	1
(Total)	32	18	50

B. Interaction Scripts

In our interaction process with scammers, we pretended to be unaware YouTube users with limited financial knowledge and an inability to differentiate between legitimate and fake investment advice. To initiate contact, we send a text message to the scammer expressing an interest. Apart from the initial greeting, during our text conversation with scammers, we positively responded to the questions asked by scammers, allowing the scam to progress as much as possible. We acted as interested / shocked after scammers presented their investment profits to make scammers believe that we were potential victims. Once the scammer asked for payment, such as registering a financial investment account in the scammer-provided URL or asking us to transfer funds to a specific wallet, we used polite excuses to leave the conversation and eventually stopped responding to messages.

To maintain anonymity, we used an online random name and address generator to prepare a list of fictional names and addresses. In the event that a scammer asked for our name and location, we assigned a specific name and address for that conversation. At the end of each conversation, we use the chat export function built within WhatsApp and Telegram to export the conversation text for further analysis.

We never sent any funds or provided financial asset information to scammers due to ethical considerations. On the other hand, our study discovered that scammers prefer cryptocurrency payments due to their anonymous nature. The transactions of those wallets are publicly accessible on their respective blockchains, which allows us to track and calculate the actual funds stolen by the scammers.

C. Data Analysis

Scam campaign distribution: Overall, we reached out to 74 scammers and captured 50 complete conversations, *i.e.,* scammers responded and asked for payment. Because of the asynchronous response of text conversations, we completed all 50 conversations in a total of 7 days. The detailed breakdown of conversations is shown in Table III.

Scam schemes: From our interactions with scammers, we identified common schemes. Table III shows the major scam schemes scammers used: cryptocurrency investments and fake prizes. Cryptocurrency trading schemes usually start with a brief introduction, followed by scammers advertising that they have trading systems that earn quick and high profits while guaranteeing success. If the victim is convinced, the scammer

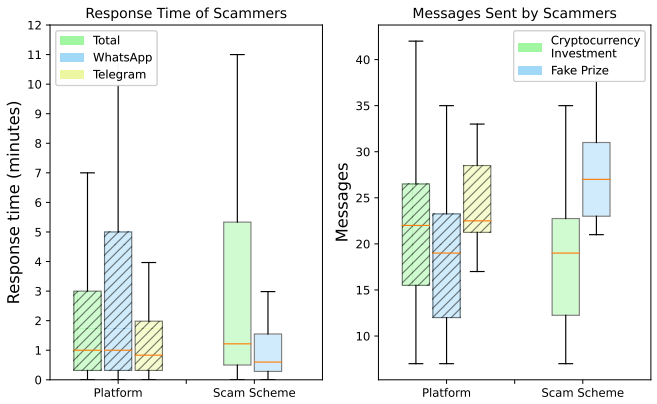

Fig. 12: *Duration of scammer conversations. Scammers are actively monitoring all their communication tools but present different verbosity in different scam campaigns.*

then guides the victim to deposit funds using various payment channels. A total of 38 (76%) scammers conducted this scheme. Fake prize scams involve scammers convincing victims that they have won some sort of an expensive prize, then asking them to fill out forms for gift shipping and requesting payment for shipping. 11 (22%) scammers conducted a fake prize scheme. Interestingly, we found there are no WhatsApp scammers conducting fake prize scams despite more scammers using WhatsApp overall. Apart from those two scam schemes, we encountered a scammer pretending to be an expert in programming languages and trying to sell us access to programming courses.

Scammers primarily use two types of identities when engaging with victims: impersonating the channel owner or pretending to be a broker / investment advisor. This behavior correlates with the type of comment script used to advertise their contact. For instance, if the original scam comment impersonates a YouTube channel owner, the scammer's WhatsApp / Telegram account will also be impersonating the channel owner. For all fake prize scam schemes on Telegram, we discovered 100% of scammers are impersonating (*i.e.,* using the same profile image and similar names) the channel owner. This is because a fake prize scheme requires convincing the victim that the channel owner is offering the prize. In contrast to the fake prize scam, cryptocurrency investment scammers do not adopt such a strategy. Instead, they utilize images of trading or investments and present themselves as investment advisors or brokers in their WhatsApp or Telegram accounts. During one of our interaction experiments, we greeted the scammer with a name that was different from the one used in the original scam comment. Surprisingly, the scammer did not notice and continued to pretend to be the person we had addressed. Combining this observation with the findings presented in Table II, we can conclude that scammers target multiple channels and use multiple identities to increase their exposure to victims, thereby being flexible with their greetings.

Scam campaign with multiple accounts: Our analysis revealed that while most scam campaigns were conducted using a single account, there were several instances where scammers used multiple accounts to defraud victims. Out of the

50 conversations we had with scammers, 16% conversations involved the use of two or more accounts to interact with us, often providing multiple reasons for doing so. For example, in the case of cryptocurrency investment scams, scammers asked us to contact their "business VIP account", while in fake prize scams, they asked us to contact a "delivery agent". The strategy of using multiple accounts can separate the scam conversation and payment channel, giving scammers an advantage by increasing credibility and protecting the payment account from being reported through the initial screening. While most scammers used the same platform for their multiple accounts, we encountered one case where a WhatsApp scammer asked us to join a Telegram group chat. From the chat history, we found more than 150 scammer-controlled accounts posting fake investment profit information, as well as signs of other victims. Additionally, greeting bots were used to redirect victims to a scammer for making payments.

Deceptive Rewards: We report the rewards provided by scammers. Scammers that conduct cryptocurrency investment scams offer unrealistic high yields of weekly returns, ranging from 15% to 1300%, with an average of 494.92%. Scammers also provide multiple options to participate, such as "Basic", "Gold", and "Diamond" plans, each with different minimum investment amounts and return rates. Meanwhile, fake prize scams offer expensive items such as iPhones or Macbook Pros. We also observe that scammers tend to tailor their social-engineering attacks to the audience of that channel by providing prizes that correlate with their impersonated identity. For example, a scammer impersonating a food channel owner provides high-end refrigerators and ovens as prizes, and a scammer impersonating a sports channel owner provides sports gear and luxury vehicles.

Temporal features of our conversations: We analyzed the response times of scammers to the messages we sent during our interactions with them and present the result in Figure 12. The leftmost figure shown in Figure 12 illustrates the distribution of these response times. Overall, 50% scammers responded to our messages within a minute, indicating that they were actively monitoring and engaging with potential victims. We found that the scammers' response times did not vary significantly between the platforms or applications they used to conduct their scams.

Our analysis of the messages and words sent by scammers revealed that they prefer short, quick campaigns, with a median of 22 messages sent per conversation. Interestingly, we found that cryptocurrency investment scams had fewer messages (with a median of 19) than fake prize scams (with a median of 27). These findings could indicate that different scammer groups are operating these two types of scams with different levels of verbosity. Scammers tend to use fewer variants in their fake prize scripts, which allows them to quickly persuade victims to send funds in exchange for these false promises. This often leads to more direct interactions between scammers and their victims. Figure 13 shows the active time of scammers. Our analysis revealed that scammers were most active between 2 PM and 3 AM Eastern Time (UTC -5) and least active between 4 AM and 12 PM Eastern

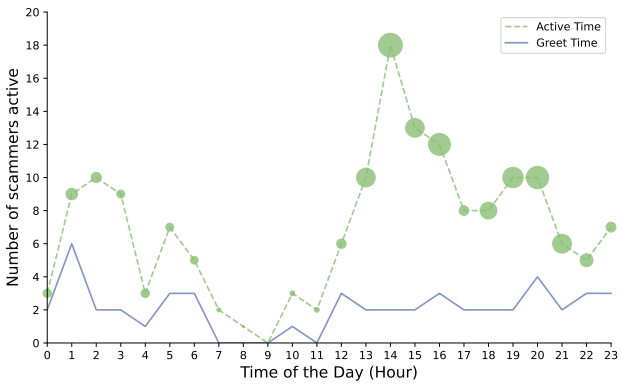

Fig. 13: *Active time of scammers. The X-axis represents the time of day in Eastern Time (UTC -5), The green dashed line represents the number of scammers that reply to us at a specific time frame (Active Time), the blue solid line represents the time scammers first respond to our messages (Greet Time). The size of the green dots represents the number of messages scammers sent at a specific time. The data suggests scammers may work in a different time zone than Eastern Time.*

Time. This pattern suggests that scammers may be working part-time, or they may be operating from a different time zone.

Payment channels: All 50 scammers provided at least one payment channel and asked us to make a deposit. Of these scammers, 38 (76%) preferred cryptocurrency payment, while 11 (22%) preferred digital payment platforms such as PayPal, CashApp, or Zelle. Among the scammers who preferred cryptocurrency payments, 23 provided a website for us to register and deposit funds, 11 directly provided a cryptocurrency wallet address in chat, and 4 asked us to download a legitimate cryptocurrency mobile wallet app and send funds through the app. In all cases, the scammers requested screenshot verification from victims at the time of payment. This common pattern may suggest a division of labor where front-end scammers cannot independently confirm backend payments.

Interestingly, only one scammer asked us to purchase gift cards for a deposit, despite gift cards being the most common method of payment requested by criminals in 2021 [25]. When we analyzed payment channels by scam scheme, we found that all fake prize scams (100%) preferred digital payment platforms, while the vast majority (96.8%) of cryptocurrency investment scammers preferred cryptocurrency transfer. Only one scammer in the cryptocurrency investment scheme provided a Zelle account. During our study, we also observed that scammers sometimes accidentally copy-pasted the payment option with numbered details. This suggests that scammers may possess multiple payment accounts, and they choose one of them when asking victims to pay. Note that due to ethical reasons, we did not actually pay any funds to the scammers and instead used polite excuses to leave the conversation after they asked us to make a payment.

Investment websites: During our interactions with scammers, 24 of them provided us with different URLs of websites pretending to be investment platforms. We investigated these URLs to gain a better understanding of the infrastructure of scam activities. By analyzing the domain registrar information,

TABLE IV: *Transaction summary of scammers. The funds stolen from 31 scammers are worth as high as 1.99 million USD.*

Crypto-currency	# of Wallets	Total Amount of Cryptocurrency	USD Value (Min. - Max.)
Bitcoin (BTC)	31	67.64	$1.07M - $1.92M
Ethereum (ETH)	16	36.49	$0.04M - $0.07M
(Total)	47	-	$1.11M - $1.99M

TABLE V: *Transaction summary of victims. The amount of funds is calculated based on the maximum USD value during the study period.*

Crypto-currency	Victims	Min.	Max.	Avg.	Median
Bitcoin (BTC)	1,901	$0.09	$226,485.6	$2,444.95	$305.08
Ethereum (ETH)	85	$1.33	$9,252.15	$780.63	$274

we found that while 29.17% of the domains were registered through NameCheap, the rest were registered with less popular hosting providers in the United States or India, such as OwnRegistrar, Inc. or PublicDomainRegistry.

As part of our data analysis, we evaluated the security aspects of the websites used by scammers. We found that all of the websites adopted HTTPS connections, presumably to increase the credibility of the sites by using SSL/TLS certificates. On the other hand, we observed unexpected behaviors in the registration processes used by these websites. While all of them required at least one Email address for registration, only 20.83% implemented an Email verification function. The remaining websites either did not require Email verification at all (allowing users to register with any Email), or they requested email verification but did not make it mandatory (allowing users to ignore the verification request and log in anyway). We registered two accounts for each site and found that the cryptocurrency wallet address provided by the website was *always* the same. This means scammers are simplifying the scam by using a single cryptocurrency wallet address as the client's wallet address. It is not possible for an investment website to provide the same cryptocurrency wallet address for different clients, as they could have no way of tracking which client invested which funds. The lack of attention to detail in website construction and business logic are clear signs that these websites are conducting fraudulent activity rather than legitimate business.

In terms of the registration date of the domain, we discovered 20 (83.33%) websites were registered in 2022. Interestingly, 100% of websites were still active at the time we ended our study, resulting in a median lifetime of 186.5 days. This indicates that scammers are able to keep these websites active for a prolonged period of time.

In order to evaluate the current online blocklist for those websites, we utilize VirusTotal API, which is an online service that integrates more than 90 antivirus scanners and URL/domain blocklisting services [39]. Following standard VirusTotal labeling practices[47], We define a domain as malicious if at least 3 of the 90 tools that are integrated into VirusTotal labeled it as either "Suspicious" or "Malicious". In total, out of the 24 URLs we submitted, only 1 URL is marked as suspicious according to the online blocklist provided by VirusTotal. The low coverage of these URLs by online blocklists underscores the limited ability of current systems to detect comment scam domains, even when these domains remain accessible for a significant period of time. This result stands in contrast to other cryptocurrency-related scams [20], [3], [21], which were characterized by short-lived infrastructure. Our findings suggest this difference is largely due to the look and feel of the scam websites, which appear to be legitimate investment platforms unless someone is aware of the full context. Our study found that most scammers are patient and guide victims through the registration process, asking for screenshots to ensure that the victim has actually paid the funds. The longevity of the scam websites in this scam campaign is also due to the fact that scammers only provide payment channels *after* they are confident that users will submit the payment.

Transaction Analysis: So far, we have presented statistics about our conversations with scammers and their preferred payment methods. As mentioned earlier, we find that scammers generally prefer cryptocurrencies as a payment channel, with 38 out of 50 scammers requesting payment via this method; the rest of 12 scammers prefer digital payment platforms such as CashApp, as mentioned previously. While the preference for cryptocurrency provides scammers with anonymity, transactions made by scammers are publicly accessible on their respective blockchains, providing us with an accurate record of the funds stolen by scammers. By analyzing these transactions, we are able to report the precise amount of funds scammers have stolen.

We successfully extracted 47 cryptocurrency wallet addresses from 31 scammers. There were 7 scammers who provided a registration URL but did not provide a cryptocurrency address at the time we ended our conversation. While all 31 scammers prefer Bitcoin (BTC) as their payment channel, 16 of them also offered Ethereum (ETH) wallet addresses as an alternative. To track the transactions made to these wallet addresses, we used publicly available API services to query the blockchains supporting these cryptocurrencies [4], [9]. The results of our investigation are presented in Table IV. Overall, scammers received a total of 67.64 BTC and 36.49 ETH, which is equivalent to $1.11M - $1.99M in USD value, calculated based on the minimum and maximum USD price of BTC/ETH during our research period. We recognize that, beyond this straightforward approach, there are other ways of calculating the USD equivalent of stolen funds and whether all transactions sent to a scammer's wallet can be characterized as part of the same campaign [11].

Our analysis revealed that out of the 31 BTC wallet addresses obtained, 29 (93.5%) had at least one successful transaction. Similarly, 11 (68.8%) out of 16 ETH wallet addresses had at least one successful transaction. Additionally, we calculated the age of the cryptocurrency wallets, defined as the number of days since the first transaction appeared in the

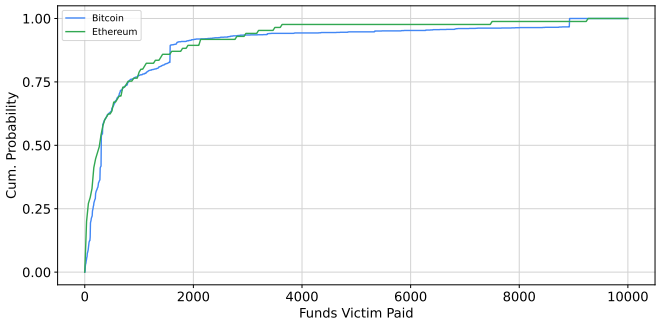

Fig. 14: *Funds sent from victims. While most victims deposit funds within $5,000, there are victims that send funds worth tens of thousands of US Dollars to scammers.*

wallet. Our results showed that most scammers' wallets were not freshly generated, with an average of 227 days for BTC wallets and 319 days for ETH wallets. We list the most active websites by their transactions in Table VII in Appendix B.

The high success rate suggests that once victims have reached out to scammers, there are limited ways to protect them from being defrauded. These findings underscore the importance of early detection and removal of these scam campaigns on media platforms, which can help eliminate victims' contact with scammers at the source. Furthermore, the profit we discovered from 31 scammers is merely a small sample of our entire dataset. Based on the success rate (31 out of 72 scammers we contacted) and their total earnings, we can estimate that a total of 10,541 scammer-controlled accounts in our dataset have potentially stolen more than 100 million US dollars from victims.

Victims: With a high success rate of scammers, we investigated the statistics from the view of victims and present the result in Table V. Overall, we identified 1,901 Bitcoin wallet addresses and 85 Ethereum wallet addresses that send funds to scammers. The small number of victims in Ethereum is due to scammers only using them as an alternative payment channel, as shown in Table IV. While 50% of the victims paid hundreds of dollars to scammers, there are victims who sent funds that are equivalent to 220K USD to the scammer. Figure 14 shows the CDF of funds that individual victims send to scammers. The large amount of funds sent from victims is due to the complicated tactics used by scammers. We present a common pattern from the web pages and scripts from scammers. Scammers entice their potential victims to invest as much as possible by providing multiple tiers of investment plans (*e.g.,* A Silver plan has a return profit of 400%, and a Gold plan has a return profit of 600%). We analyzed the corpus of text in the conversation and report on the profit promised by scammers. The average profit promised by a scammer (if provided) is 494.92%, and one scammer promised a 1300% profit. In addition, scammers often provide a minimum amount of funds that are required for participating in the investment. These limits are present in both the conversation script and scam websites. We discover that while scammers require an average of $1,820 as a minimum investment deposit, some scammers allow as small as $50 to participate. Fake prize scammers require a fixed amount of funds for the shipping cost, ranging from $50 to $500, with an average of $178.54.

Scammer Demeanor: During our interactions with scammers, we found that most exhibit a polite attitude. Scammers begin by kindly greeting us and patiently introducing their investment plan, often providing long and detailed descriptions to account for our "inexperience". When we asked questions about the safety of the investment, such as "Is this investment safe?" or "Will I lose my money?", they immediately provided kind and confident answers to persuade us to deposit funds. Scammers also frequently asked us to provide screenshots, such as a screenshot of funds at a cryptocurrency exchange, so that they could guide us through the entire process and ensure we had sufficient funds in our account. However, the scammers' attitudes changed at the end of our conversations.

After the scammer provided their payment channel, we terminated the conversation by not answering any subsequent messages. 96% of scammers sent at least one follow-up message asking us to deposit funds. Although most scammers gave up after four to five additional messages, two scammers persistently sent us messages for nine days in an angry and blaming tone, and six scammers initiated a phone call or video chat in an attempt to reach us. After we ended our interaction, we also observed that scammers reached out to us through new accounts and identities that we did not recognize. This could potentially indicate that there are multiple scammers inside a single campaign, and they share their victims' information. One scammer, after realizing that we were not providing any funds, deleted all previous conversations.

VI. Limitations

Our work has two limitations that we want to address: i) the omission of comments between snapshots, and ii) the potential evasion of scam comment filters.

Data incompleteness: We recognize that our system may not capture all comments of a video. Due to API limitations, our crawler periodically captures comments once an hour to maximize the breadth of measurement. While this approach captures most comments, comments that are created and deleted within the one-hour timeframe are not captured. This could occur, for example, if a user posts a comment and deletes it to correct it. In such cases, the deleted comments from users are not part of our research subject. Alternatively, early deletion of comments could be due to the platform detecting scam comments and deleting them. In both cases, we argue that the interval between snapshots is not rigid and can be adjusted if necessary.

Scammer Evasions: We discovered evidence that scammers employ various adversarial techniques to evade regulations, such as manipulating textual, graphical, and temporal aspects. Although we have designed and refined several filters based on our observations, it is possible that some techniques, such as introducing typos or using AI-generated scripts, may be evading our filters. In such cases, additional filters can be added as part of the analysis module without disrupting the rest of our pipeline's detection logic.

VII. Related work

Concurrently with our work, Na et al. performed a large-scale study of social-bot activity on YouTube [24]. Prior to our work and that of Na et al., the only investigation of this phenomenon was done by blog posts and reports by cyber-security companies investigating specific case studies [38], [10], [33]. Contrastingly, through a three-pronged approach, we were able to collect over 100K scam comments, map their velocity and underlying infrastructure, and ultimately interact with the scammers that left these comments. Older papers on YouTube spam were largely focused on studying unwanted self-promotion and links to external (possibly malicious) websites [7], [1], [46], all of which are nowadays turned off by large channel operators.

Technical support scams. Comment scams are mostly related to technical support scams in terms of scammer tactics and used communication mediums. Technical support scams are social-engineering attacks that lie in the intersection of scareware [28], [32] and the telephone scams [37]. In such scams, malicious actors set up websites that impersonate globally recognized brands and pretend to find issues (like malware) on user devices with fake diagnostics, before asking users to pay funds to repair their machines [30], [14], [29]. In both types of scams, phone numbers play an important role as the communication medium between scammers and their potential victims [36], making the fraud appear more convincing to potential victims [8]. A previous study by Miramirkhani *et al.* [22] systematically investigated technical support scams by crawling a large dataset of malicious websites and providing insights into scammers' prevalence and profits. In technical support scams, scammers communicate with their potential victims through phone calls, which require increased "bandwidth" from scammers (*i.e.,* one scammer cannot talk with multiple users at the same time). In contrast, once users contact a comment scammer, they face a different set of tactics including fake prizes and "get-rich-quick" schemes, enticing users to voluntarily send funds to scammers. Our study also finds that scammers prefer text communication instead of voice calls, allowing them to better impersonate celebrities, have more time to prepare responses (*i.e.,* one scammer can chat with multiple users at the same time), and thus lower their cost in conducting scams. Additionally, our results indicate that scammers prefer cryptocurrencies as a payment method instead of traditional payments (e.g. credit cards or gift cards) used in prior social-engineering attacks.

Cryptocurrency scams. Comment scams are related to cryptocurrency scams in terms of payment channels preferred by scammers, *i.e.,* cryptocurrency transactions. Due to their decentralized nature, cryptocurrencies provide a layer of anonymity in payments and have thus been extensively abused by fraudsters, with reported losses of $2.57 billion in 2022 [26]. In cryptocurrency scams, scammers set up schemes like ransomware [18], [19], [23], [41], advance-fee scams [3] and cryptocurrency giveaway scams [20]. Among recent rising scam schemes that utilize cryptocurrency, one of them is known as cryptocurrency giveaway scams. In this scheme, scammers set up websites that abuse the names and images of celebrities, then advertise fake giveaway events that promise to double or triple the funds victims send to them. Scammers use various channels to promote the event, like setting up fake YouTube live broadcasts that abuse past celebrity talks or interviews and advertising the giveaway event or using compromised Twitter accounts to promote cryptocurrency giveaway scams. Unlike cryptocurrency giveaway scams that simply ask users to deposit funds to a wallet, scammers conducting comment scams directly interact with their potential victims and provide them with payment instruments only *after* the scammer has confidence that a victim is likely to send them funds. This difference dramatically increases the lifespan of a scam campaign and the validity of URLs and wallet addresses, thereby further lowering the costs of successfully operating comment scams.

VIII. Conclusion

In this paper, we report on the first systematic study that taps into the ecosystem of comment scams on media platforms, where scammers attempt to defraud users by enticing them with impersonation or fabricated "get-rich-quick" schemes. Over a period of six months, our YouTube-focused infrastructure captured a total of $8{,}801{,}224$ comments from 20 channels, in which $206{,}306$ were posted by comment scammers. We discovered that scammers constantly create new accounts and employ various evasion tactics to evade platform regulations. By analyzing our collected dataset, we discovered that scammer comments differ greatly from regular comments in terms of textual, graphical, and temporal features, with 81.89% scammer accounts utilizing visually similar symbols and 45.56% scammer accounts abusing a channel's username in scam activities. Furthermore, we grouped scams into campaigns, discovering a single scammer using more than 100 accounts in a single campaign. We also reported scam-activity dynamics by tapping into the snapshots of comments and highlighting the low account-deactivation rate of scammer accounts. Finally, we presented a seven-day experiment where we directly interacted with 50 scammers to better understand the last phase of their social engineering attacks. Among others, we took advantage of the scammers' reliance on cryptocurrencies to identify that even a small number of scammers can steal millions of dollars from unsuspecting victims. Our study shows that existing mechanisms put in place by media platforms are clearly insufficient, setting the stage for future research for fast takedowns of comment scams as well as client-side protections for the users who end up contacting the scammers.

Availability. One of our core contributions is the construction of an infrastructure along with a set of filters to identify and capture scam comments which we make available: https://like-comment-get-scammed.github.io/.

IX. Acknowledgements

This work was supported by the Office of Naval Research (ONR) under grants N00014-20-1-2720 and N00014-22-1-2001 as well as by the National Science Foundation (NSF) under grants CNS-1813974, CNS-2126654, and CNS-2211575.

REFERENCES

[1] T. C. Alberto, J. V. Lochter, and T. A. Almeida, "Tubespam: Comment spam filtering on youtube," in *2015 IEEE 14th international conference on machine learning and applications (ICMLA)*. IEEE, 2015, pp. 138–143.

[2] S. Alshamrani, A. Abusnaina, and D. Mohaisen, "Hiding in plain sight: A measurement and analysis of kids' exposure to malicious urls on youtube," in *2020 IEEE/ACM Symposium on Edge Computing (SEC)*, 2020, pp. 321–326.

[3] M. Bartoletti, S. Lande, A. Loddo, L. Pompianu, and S. Serusi, "Cryptocurrency scams: analysis and perspectives," *IEEE Access*, vol. 9, pp. 148 353–148 373, 2021.

[4] Blockstream, "Blockstream explorer api," "https://blockstream.info/", 2023.

[5] M. Brownlee, "Youtube needs to fix this," https://www.youtube.com/watch?v=1Cw-vODp-8Y, 2023.

[6] Chicoer, "Watch out for comment reply scams on youtube, other social media — scam of the week," https://www.chicoer.com/2023/03/14/watch-out-for-comment-reply-scams-on-youtube-other-social-media-scam-of-the-week/, 2023.

[7] R. Chowdury, M. N. M. Adnan, G. Mahmud, and R. M. Rahman, "A data mining based spam detection system for youtube," in *Eighth international conference on digital information management (ICDIM 2013)*. IEEE, 2013, pp. 373–378.

[8] A. Costin, J. Isacenkova, M. Balduzzi, A. Francillon, and D. Balzarotti, "The role of phone numbers in understanding cyber-crime schemes," in *Proceedings of the 11th International Conference on Privacy, Security and Trust (PST)*. PST, 2013.

[9] Etherscan, "Etherscan api," "https://etherscan.io/", 2023.

[10] Express, "Youtube superstars issue scam warning: Avoid this trick trying to steal your money," https://www.express.co.uk/life-style/science-technology/1590429/youtube-comment-spam-scam-warning-trick-trying-to-steal-your-money, 2023.

[11] G. Gomez, K. van Liebergen, and J. Caballero, "Cybercrime bitcoin revenue estimations: Quantifying the impact of methodology and coverage," in *ACM Conference on Computer and Communications Security (CCS)*, 2023.

[12] L. A. Goodman, "Snowball sampling," *The annals of mathematical statistics*, pp. 148–170, 1961.

[13] Google, "Impersonation policy (youtube)," https://support.google.com/youtube/answer/2801947, 2023.

[14] D. Harley, M. Grooten, S. Burn, C. Johnston *et al.*, "My pc has 32,539 errors: how telephone support scams really work," *Virus Bulletin*, 2012.

[15] M. N. Hussain, S. Tokdemir, N. Agarwal, and S. Al-Khateeb, "Analyzing disinformation and crowd manipulation tactics on youtube," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2018, pp. 1092–1095.

[16] J. Jani, "I confronted the youtube comment scammers," https://www.youtube.com/watch?v=ypYQKlyoHwU, 2023.

[17] Kayla, "Individual report: Common comment scam on youtube," https://twitter.com/lilsimsie/status/1569361526670098432, 2023.

[18] A. Kharraz, W. Robertson, D. Balzarotti, L. Bilge, and E. Kirda, "Cutting the gordian knot: A look under the hood of ransomware attacks," in *International conference on detection of intrusions and malware, and vulnerability assessment*. Springer, 2015, pp. 3–24.

[19] N. Kshetri and J. Voas, "Do crypto-currencies fuel ransomware?" *IT professional*, vol. 19, no. 5, pp. 11–15, 2017.

[20] X. Li, A. Yepuri, and N. Nikiforakis, "Double and nothing: Understanding and detecting cryptocurrency giveaway scams," *Network and Distributed Systems Security (NDSS) Symposium*, 2023.

[21] S. Mackenzie, "Criminology towards the metaverse: Cryptocurrency scams, grey economy and the technosocial," *The British Journal of Criminology*, vol. 62, no. 6, pp. 1537–1552, 2022.

[22] N. Miramirkhani, O. Starov, and N. Nikiforakis, "Dial One for Scam: A Large-Scale Analysis of Technical Support Scams," in *Proceedings of the 24th Network and Distributed System Security Symposium (NDSS)*, 2017.

[23] S. Mohurle and M. Patil, "A brief study of wannacry threat: Ransomware attack 2017," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 5, pp. 1938–1940, 2017.

[24] S. H. Na, S. Cho, and S. Shin, "Evolving bots: The new generation of comment bots and their underlying scam campaigns in youtube," in *Proceedings of the 2023 ACM on Internet Measurement Conference*, ser. IMC '23, 2023, p. 297–312.

[25] NIH, "Protecting retail customers from gift card payment scams," https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9771048/, 2022.

[26] U. D. of Justice, "Justice dept. seizes over $112m in funds linked to cryptocurrency investment schemes, with over half seized in los angeles case," https://www.justice.gov/usao-cdca/pr/justice-dept-seizes-over-112m-funds-linked-cryptocurrency-investment-schemes-over-half, 2023.

[27] PyPI, "Imagehash 4.3.1 - pypi," https://pypi.org/project/ImageHash/, 2023.

[28] M. A. Rajab, L. Ballard, P. Mavrommatis, N. Provos, and X. Zhao, "The nocebo effect on the web: an analysis of fake anti-virus distribution," in *USENIX workshop on large-scale exploits and emergent threats (LEET)*, 2010.

[29] S. Rauti and V. Leppänen, ""you have a potential hacker's infection": A study on technical support scams," in *2017 IEEE International Conference on Computer and Information Technology (CIT)*. IEEE, 2017, pp. 197–203.

[30] B. Srinivasan, A. Kountouras, N. Miramirkhani, M. Alam, N. Nikiforakis, M. Antonakakis, and M. Ahamad, "Exposing Search and Advertisement Abuse Tactics and Infrastructure of Technical Support Scammers," in *Proceedings of the Web Conference (WWW)*, 2018.

[31] Statista, "Youtube superstars issue scam warning: Avoid this trick trying to steal your money," https://www.statista.com/statistics/469152/number-youtube-viewers-united-states/, 2023.

[32] B. Stone-Gross, R. Abman, R. A. Kemmerer, C. Kruegel, D. G. Steigerwald, and G. Vigna, "The underground economy of fake antivirus software," in *Proceedings of the 10th Workshop on Economics of Information Security (WEIS)*, 2011.

[33] TeamPassword, "Youtubers are sick of comment spam, so youtube is testing a stricter moderation system," https://teampassword.com/blog/the-youtube-telegram-giveaway-scam, 2023.

[34] Telegram, "Telegram faq: What is telegram," https://telegram.org/faq, 2023.

[35] A. Tripathi, M. Ghosh, and K. Bharti, "Analyzing the uncharted territory of monetizing scam videos on youtube," *Social Network Analysis and Mining*, vol. 12, no. 1, p. 119, 2022.

[36] H. Tu, A. Doupé, Z. Zhao, and G.-J. Ahn, "Sok: Everyone hates robocalls: A survey of techniques against telephone spam," in *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2016, pp. 320–338.

[37] ——, "Users really do answer telephone scams." in *USENIX Security Symposium*, 2019, pp. 1327–1340.

[38] T. Verge, "Youtubers are sick of comment spam, so youtube is testing a stricter moderation system," https://www.theverge.com/2022/4/8/23016861/youtube-comment-spam-testing-moderation, 2023.

[39] VirusTotal, "Fake name generator: Generate u.s., u.k., etc. fake person," https://developers.virustotal.com/reference/overview, 2023.

[40] WhatsApp, "Whatsapp," https://www.whatsapp.com/, 2023.

[41] P. Xia, H. Wang, X. Luo, L. Wu, Y. Zhou, G. Bai, G. Xu, G. Huang, and X. Liu, "Don't fish in troubled waters! characterizing coronavirus-themed cryptocurrency scams," in *2020 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 2020, pp. 1–14.

[42] YouTube, "Spam, deceptive practices, & scams policies (youtube)," https://support.google.com/youtube/answer/2801973, 2023.

[43] ——, "Updates on comment spam & abuse," https://support.google.com/youtube/thread/192701791, 2023.

[44] ——, "Youtube scams to watch out for," https://www.youtube.com/watch?v=wfrhlG_Xns8, 2023.

[45] ——, "Youtube's community guidelines," https://support.google.com/youtube/answer/9288567, 2023.

[46] Y. Yusof and O. H. Sadoon, "Detecting video spammers in youtube social media," 2017.

[47] S. Zhu, J. Shi, L. Yang, B. Qin, Z. Zhang, L. Song, and G. Wang, "Measuring and modeling the label dynamics of online anti-malware engines." in *USENIX Security Symposium*, 2020, pp. 2361–2378.

APPENDIX

A. YouTube Channels

TABLE VI: *Summary of YouTube channels used in our study.*

Category	Channel Name	Channel ID	Subscribers (As of March 31,2023)	Scam comments
Finance	Graham Stephan	UCV6KDgJskWaEckne5aPA0aQ	4.28M	63,198
Finance	Grant Cardone	UCdlNK1xcy-Sn8liq7feNxWw	2.37M	9,186
Finance	Andrei Jikh	UCGy7SkBjcIAgTiwkXEtPnYg	2.21M	49,378
Finance	Brian Jung	UCQglaVhGOBI0BR5S6IJnQPg	1.29M	6,080
Finance	BiggerPockets	UCVWDbXqQ8cupuVpotWNt2eg	1.05M	6,162
Finance	The Financial Diet	UCSPYNpQ2fHv9HJ-q6MIMaPw	1M	6,726
Finance	Marko - WhiteBoard Finance	UCL_v4tC26PvOFytV1_eEVSg	948K	405
Finance	Sebastian Ghiorghiu	UCZ59iKBmGRfQlnl73sOX0Lw	895K	6,414
Finance	Ryan Scribner	UC3mjMoJuFnjYRBLon_6njbQ	808K	381
Finance	His And Her Money	UCCnXqVJZq_cD9wpycpml9LQ	236K	140
Sports	MLB	UCoLrcjPV5PbUrUyXq5mjc_A	4.22M	1,299
Sports	CoshReport	UCOIaQxIvcpPjxgSaYmqaE6g	394K	751
Cooking	Gordon Ramsay	UCIEv3lZ_tNXHzL3ox-_uUGQ	19.8M	8,916
Cooking	Joshua Weissman	UChBEbMKI1eCcejTtmI32UEw	7.68M	20,355
News/Politics	Fox News	UCXIJgqnII2ZOINSWNOGFThA	10.5M	18,483
News/Politics	MSNBC	UCaXkIU1QidjPwiAYu6GcHjg	5.58M	6,643
Education	CrashCourse	UCX6b17PVsYBQ0ip5gyeme-Q	14.6M	3
Education	freeCodeCamp.org	UC8butISFwT-Wl7EV0hUK0BQ	7.52M	1,483
Gambling	CardMechanic	UCusqjyLtS_hrcnfF98QbG2A	36.3K	0
Gambling	The Jackpot Gents	UC-x4b_ZpsSUS3ICfwHWkO2A	191K	303

B. Scammer websites

TABLE VII: *Top 5 scammer websites that have the most transactions.*

Domain	BTC Wallet Address	BTC Amount	ETH Wallet Address	ETH Amount	# Transactions
globalmarkettrade[.]online	bc1qj0 ... g5he46	2.4664	N/A	N/A	321
moonchoiceassets[.]com	bc1qu8 ... lnq0rx	1.5289	0x22d7...5B8221	0	219
extrademarket[.]com	bc1qnl ... 7kxe3a	36.2337	N/A	N/A	202
tradeprogression[.]net	bc1qxs ... wqf0c0	8.0264	0x4b8d...4B8B5B	0.7737	116
bi-investments[.]com	3PJjJJ ... UgJyhC	1.0044	0x2428...6dafa8	4.1336	98