Understanding and Analyzing Appraisal Systems in the Underground Marketplaces

Zhengyi Li, Xiaojing Liao
Indiana University Bloomington
{zl11, xliao}@indiana.edu

Abstract—An appraisal system is a feedback mechanism that has gained popularity in underground marketplaces. This system allows appraisers, who receive free samples from vendors, to provide assessments (i.e., appraisal reviews) for products in underground marketplaces. In this paper, we present the first measurement study on the appraisal system within underground marketplaces. Specifically, from 17M communication traces from eight marketplaces spanning from Feb 2006 to Mar 2023, we discover 56,229 appraisal reviews posted by 18,701 unique appraisers. We look into the appraisal review ecosystem, revealing five commonly used requirements and merits in the appraiser selection process. These findings indicate that the appraisal system is a well-established and structured process within the underground marketplace ecosystem. Furthermore, we reveal the presence of high-quality and unique cyber threat intelligence (CTI) in appraisal reviews. For example, we identify the ge-locations of followers for a social booster and programming languages used for malware. Leveraging our extraction model, which integrates 41 distinct types of CTI, we capture 23,978 artifacts associated with 16,668 (50.2%) appraisal reviews. In contrast, artifacts are found in only 8.9% of listings and 2.7% of non-appraisal reviews. Our study provides valuable insights into this under-explored source of CTI, complementing existing research on threat intelligence gathering.

I. INTRODUCTION

The last decade has seen an upsurge of underground online marketplaces that offer a wide range of malicious and illegal products, ranging from malware (i.e., ransomware and RATs) [105], [110] to some newly-emerged scam services [101]. These marketplaces operate in anonymity and are structured like conventional e-commerce platforms such as eBay or Amazon. Similar to conventional e-commerce platforms, most underground marketplaces have introduced a feedback system to maintain a “healthy” community environment. This system helps users assess product quality and decrease scam activities in underground marketplaces [59].

Appraisal system. With the rapid expansion of underground marketplaces, the feedback system has also evolved to provide better product assessment and more valuable reviews. A new type of feedback system – the appraisal system – has emerged as a trend in the underground marketplaces. In the appraisal system, vendors provide free samples, also known as vouch copies, to qualified members, or appraisers, in exchange for their detailed and in-depth reviews. An example of an appraisal review is shown in Figure 1. Similar appraisal systems are also observed in legitimate sites such as Amazon Vine [26], Influenster [63], and BzzAgent [41]. Customers with good reputations are given the opportunity to participate in these systems. Members of these systems can request free copies of products and post opinions about items to help their fellow customers to make educated purchasing decisions. Compared to traditional feedback systems such as review-based or rating-based feedback systems, the appraisal system is generally considered more trustworthy. This is because vendors who offer vouch copies typically set certain member criteria (e.g., having 100 posts or being a VIP) to select qualified appraisers. Additionally, the appraisal system can provide expert comments and feedback. Unlike the review content, which may only contain simple comments such as "Excellent! Thanks," appraisers provide a comprehensive evaluation of the product from different aspects such as price, functionality, unique features, potential drawbacks, and more. So far, little has been done to systematically discover and analyze this new feedback system, not to mention any effort to understand the ecosystem behind it and the potential cyber threat intelligence that can be mined from this ecosystem.

Finding appraisers and appraisal reviews. In this paper, we report the first measurement study on the appraisal system. The study relies on the identification of appraisers and appraisal reviews from underground marketplaces, which is challenging. Specifically, unlike e-commerce platforms like Amazon or eBay, where customer reviews are easily accessible in the Customer Reviews section, underground marketplaces typically operate in a forum-like format where different types of traces, including discussion, dispute, and reviews, are mixed without any labeling. Moreover, there has been no prior work on identifying features that differentiate appraisal reviews from non-appraisal reviews.

To address these challenges, we propose a method, that identifies appraiser and appraisal reviews in the wild. Our
method was bootstrapped by a set of “groundtruth” appraisers and their appraisal reviews. We identified official appraiser groups, led by marketplace administrators or reputable members, through which appraisal services were offered, such as the “Official Reviewers Group” and “Official Appraisers” in Hack Forums. By comparing them with non-appraisal reviews, we found that appraisal reviews typically declare the vouch copies and sometimes follow a review template (§ III-B). These features make it possible to identify appraisers and their appraisal reviews with high confidence. From 17M communication traces from 8 underground marketplaces spanning from Feb 2006 to Mar 2023, our study reported 479 “groundtruth” appraisers associated with 4,054 appraisal reviews, and used our tool to flag 18,222 appraisers associated with 52,175 appraisal reviews.

Measurement and discoveries. Looking into the 18,701 appraisers and 56,229 appraisal reviews reported in this study, we observed that the appraisal system has been widely deployed, with a significant impact on today’s underground marketplaces. More specifically, our analysis revealed that the practice of providing vouch copies for appraisers was first mentioned in the listings on BlackHatWorld and Hack Forums as far back as Dec. 2008 and Oct. 2011, respectively. Starting from February 2014, we observed that the appraisal system was consistently implemented in newly-launched underground marketplaces, such as Evolution (launched in January 2014) and Nulled Marketplace (launched in February 2015 and reloaded in Jan 2018).

Also interesting is the ecosystem of the appraisal system, as discovered in our study. It includes building official appraiser groups, vendors’ appraiser recruitment, etc. For example, in an official group’s appraiser recruitment, every applicant is required to submit a sample review as a test of their capability to appraise a product. In vendors’ appraiser hiring, they prefer appraisers who have purchased a VIP membership in the marketplaces, or have at least 500 forum posts (§ IV and V). These findings indicate that the appraisal system is a well-established and structured process within the underground marketplace ecosystem. When investigating communication between appraisers and vendors via a leaked dataset from Nulled, an underground marketplace in our study, (un)surprisingly, we found that vendors sometimes interfere with appraiser’s reviews in order to manipulate the review content to better promote their products (§ V-B).

Furthermore, the analysis of cyber threat intelligence (CTI) from appraisal reviews revealed a new source of valuable and unique threat information. Particularly, in our study, we found that appraisers offer rich and detailed technical information, providing new insights into the evolving threat landscape. As an instance, in the appraisal review of Helix Crypter, the appraiser provided extensive information, including the malicious file hash (MD5 and SHA1), filename, file size, and scan results from 34 antivirus products. Notably, the appraiser observed that the malware could be identified and labeled as “suspicious”, despite the vendor’s assertion that it is fully undetectable (FUD). Interestingly, other intelligence sources, such as VirusTotal [112] and DigitalSide [52], did not offer any information on the malware using the provided hash or filename, nor did any industry white papers. Our study revealed that a significant proportion of the appraisal review (50.2%) contained valuable and diverse CTI. Such CTI can supplement the existing understanding of cyber threats.

Contributions. The contributions of the paper are as follows.

- We report the first in-depth measurement study on the appraisal system in underground marketplaces. Our study investigates the ecosystem of the appraisal system and the actors involved.
- We demystify the characteristics of appraisers (e.g., profile, credibility, merits for appraiser selection, etc.) and appraisal reviews (e.g., assessment merits, quality comparing to non-appraisal reviews, etc.) via a large-scale analysis on 18,701 appraisers and 56,229 appraisal reviews from 8 marketplaces spanning 15 years.
- We shed light on an under-explored source of cyber threat intelligence – appraisal reviews, which supplements current studies on threat intelligence gathering.
- We build a taxonomy of IOCs including 41 types of valuable threat intelligence specific to the underground marketplaces and appraisal reviews.

II. BACKGROUND

A. Appraiser System in Underground Marketplace

Underground online marketplaces are virtual platforms that facilitate transactions between sellers and buyers. These marketplaces typically include forums where buyers and sellers can share information, promote their products, leave feedback, and discuss their experiences with purchases. Prior study [46], has noted that, underground marketplaces, which offer anonymity guarantees, provide little to no legal recourse against scammers, indicating the potential for deceptive behavior. Some of those marketplaces (e.g., Silk Road, Evolution, Agora) have implemented feedback systems to monitor vendors and reduce fraudulent activities. These feedback systems also assess product quality and offer guidance to vendors, providing buyers with useful information to make purchasing decisions.

In our study, we have discovered the appraiser system, a new type of feedback system, is widely deployed in underground marketplaces in recent years. In the appraiser system, a vendor will offer a free trial sample to a trustful member, named appraiser. After testing the sample, known as vouch copy, the appraiser will post a detailed and insightful review, called appraisal review, under the vendor’s listing. Prior to 2010, we had only observed two underground marketplaces (Hack Forums and BlackHatWorld) with appraisal systems, which had a low number of active appraisers. However, over the next decade, as more marketplaces emerged (e.g., Evolution, Nulled, V3million), the appraisal system was adopted by these platforms. We also found that the number of appraisers in these marketplaces has been gradually increasing. We will elaborate on the measurement study of the appraisal system in Section V-A.

B. Cyber Threat Intelligence Gathering

CTI. Cyber Threat Intelligence (CTI) is defined as “evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, about an existing or
emerging menace or hazard to assets that can be used to inform decisions regarding the subject’s response to that menace or hazard” [18]. This knowledge is essential for an organization to gain visibility into the fast-evolving threat landscape, identify early signs of an attack and the adversary’s strategies, tactics, and techniques, and effectively contain the attack with proper means. Given its importance, CTI has been aggressively collected and increasingly exchanged across organizations, often in the form of Indicators of Compromise (IOC) [86], which are forensic artifacts of an intrusion such as virus signatures, IPs/domains of botnets, MD5 hashes of attack files, etc.

Sources of CTI. One of the essential problems in CTI gathering is where to find the sources of CTI. The common sources of CTI include structured attack artifact datasets (e.g., PhishTank [14], CleanMX [3], and OpenPhish [13]), technical blogs and articles (e.g., research paper, white papers, etc.) [55], [73]. Recent years have witnessed underground marketplaces and forums becoming important sources of CTI. In particular, the CTI from underground marketplaces and forums can create an integrated and accurate picture of the threat landscape [35], due to the diverse roles and communications which exist between cybercriminals in the underground marketplaces. Such data includes valuable information for understanding the ecosystem of malicious activities, which can be automated to enable timely identification of the adversary’s capabilities, strategies, and infrastructure [31], [77], [85].

Our study reveals that the appraisal reviews written by appraisers offer a comprehensive analysis of malicious products and contain a dense amount of cyber threat intelligence (CTI). These reviews can be considered as a novel source of CTI that requires particular attention (see §VII).

III. METHODOLOGY

In this section, we present the design and implementation for identifying appraisal reviews from eight underground marketplaces. We first elaborate on the process of data collection and validation, and then discuss how we identify appraisal reviews.

A. Datasets

Data collection. Our study collected traces (including the initial post and the following replies of a thread [89]) from eight underground marketplaces and forums (Hack Forums [9], Blackhat World [2], V3million [24], MPGH [11], Nulled [12], OGUsers, Evolution, and Raid) from 2006 to 2023 to identify appraisers and appraisal reviews. Note that we focus on underground marketplaces of malware and other cyber products/service, instead of illegal drugs in our research. Specifically, we used the following four data sources and also elaborated on ethics discussion around these datasets in §VIII. The three public data sources: CrimeBB [89], Nulled database [87], and dark net markets (DNM) archive [49] are reputable and widely-recognized resources extensively used in cybercrime research [93], [102]. This enables us to fully reproduce the entire dataset and make comparisons on dataset volume with other studies.

- CrimeBB underground marketplaces dataset [89]: CrimeBB dataset consists of communication traces and user account information of seven underground marketplaces of our interest (BlackHatWorld, HackForums, MPGH, V3million, OGUsers, Raid, and Nulled) from 2006 to 2020. In total, we collected 12,752,742 communication traces and 812,080 user account information (see Table I). Moreover, to investigate the credibility of appraisers (§V), we collected 753,933 traces from the scam reporting sub-forum of each marketplace. These sub-forums function as a platform for users to report scams and disputes between vendors and reviewers.

- Nulled database [87]: The Nulled database consists of 121,499 traces spanning from Feb 2015 to May 2016 and includes 599,085 user information. In our study, we combined the Nulled database (02/15-05/16) and the Nulled traces from CrimeBB (01/18-07/19), as shown in Table I. Also, the Nulled dataset includes 800,593 private messages exchanged among 36,606 users. We incorporated these messages into our study to examine the potential collusion between appraisers and vendors (§V). We discuss the ethics of using this dataset in §VIII.

- dark net markets (DNM) archive [49]: We obtained 9,385 traces of the Tor-based marketplaces Evoluion from the dark net markets (DNM) archive. These traces cover a period from February 2014 to November 2014. Note that in our study of the DNM dataset, which includes traces from 89 DNMs and 37 related forums spanning from 2013 to 2015, we have excluded marketplaces with fewer than 50 listings or that do not primarily focus on malware and other cyber products/service.

- Self-scraped dataset: To address the data gap of CrimeBB from 2020 to 2023, we built our scrapers on the top of Selenium [19] to conduct site crawling and content parsing on five marketplaces (BlackHatWorld, Hackforum, MPGH, Nulled, V3million), which are still active. Specifically, given the target listing URL to be crawled, our scraper performed a direct HTTP request from our client server to the target URL, scraping its contents, storing the raw HTML, and parsing the raw page data. In our study, we ensured a complete scrape by checking the HTTP status code and the returned page size, monitoring session expiry, handling unsuccessful sites with CAPTCHA-solving mechanisms, and addressing other intermittent failure modes. We retried requests when necessary to ensure that all relevant data was captured. We validate the data completeness as elaborated below.

Altogether, we gathered 17,340,789 traces, i.e., $D_{all}$ (spanning from Dec 2006 to Mar 2023) from the eight underground forums as shown in Table I. Note that the measurement date indicated the earliest and the last listings/traces we have seen for each marketplace.

Validation of data completeness. Before using the Self-scraped dataset, we validated the data completeness by checking their over-time consistency and by comparing them to the statistics reported by other studies.

Inspired by work [48], [114], we present the cumulative number of unique listings on the five markets collected by our scrapers in Figure 2. The curves exhibit general smooth upward trends, indicating good data completeness. However, we still observed a plateau on V3million and HackForums during the middle of 2020. The decrease of newly-appeared listings started from then might be caused by the effects of
B. Groundtruth Appraisal Reviews

In our study, we found two sources of official appraisers as our groundtruth dataset. We then identified their appraisal reviews using a set of heuristics we devised.

Specifically, we observed official appraiser groups led by marketplace administrators or reputable members through which to offer appraisal services; for example, the “Official Appraisers” in Hack Forums. Those appraiser members were publicly selected through an application process that included a background check (i.e., length of membership, number of posts, and no scam reports) and their capability of appraising product testing (e.g., writing a public review on an assigned product). To request appraisal services, vendors can reach out to these official appraiser groups and send out vouch copies. Then appraisers will be randomly assigned to provide appraisal reviews. We observed that those appraisal reviews follow a template that is specified by each group. For instance, the “Official Appraisers Group” requires all appraiser members to use the following review template: “Review (Ratings on a scale of 1-5), Quality of Information: ***, Ease of Use: ***, Layout: ***, Grammar: ***, Originality: ***.” when posting an appraisal review. We manually investigated those review templates and carefully designed a set of regular expressions (regex) to find appraisal reviews from all traces in our dataset $D_{all}$. In total, our regex matched 1,927 groundtruth appraisal reviews, which belong to 379 appraiser members and 3 appraiser groups in the Hack Forums, as shown in Table I.

We also used a sub-forum called Vouch Copy Profiles [25] in MPH to identify additional groundtruth appraisers. This sub-forum serves as a platform for vendors to locate reliable appraisers and establish guidelines to regulate appraisers’ behavior. Specifically, only appraisers who meet specific requirements (i.e., 1,500 posts, 3-month membership and who communicate using clear English) are allowed to post a thread

---

**TABLE I: Dataset summary. The statistics of appraisers and non-appraisers, as well as their reviews across eight marketplaces**

<table>
<thead>
<tr>
<th>Type</th>
<th>Marketplace</th>
<th>Data source</th>
<th>Measurement date</th>
<th># traces</th>
<th># appraisal review (%)</th>
<th># appraisal (%)</th>
<th># appraisal listing (%)</th>
<th># non-appraisal review (%)</th>
<th># non-appraisal (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groundtruth</td>
<td>Hack Forums</td>
<td>Our scrape</td>
<td>02/07 – 03/23</td>
<td>9,312,519</td>
<td>1,927 (4.3%)</td>
<td>379 (2.0%)</td>
<td>1,256 (4.3%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>BlackHatWorld</td>
<td>Our scrape</td>
<td>03/08 – 03/23</td>
<td>13,529,041</td>
<td>2,127 (15.3%)</td>
<td>108 (0.02%)</td>
<td>960 (7.3%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MPH</td>
<td>CrimeBB</td>
<td>12/06 – 03/23</td>
<td>3,433,465</td>
<td>26,304 (46.8%)</td>
<td>6,505 (34.8%)</td>
<td>4,285 (14.6%)</td>
<td>366,844 (21.6%)</td>
<td>63,003 (17.6%)</td>
</tr>
<tr>
<td></td>
<td>HackForums</td>
<td>Our scrape</td>
<td>02/07 – 03/23</td>
<td>9,312,519</td>
<td>19,414 (34.3%)</td>
<td>8,006 (43.1%)</td>
<td>13,678 (22.8%)</td>
<td>829,301 (48.9%)</td>
<td>204,199 (27.1%)</td>
</tr>
<tr>
<td></td>
<td>MPH</td>
<td>CrimeBB</td>
<td>12/06 – 03/23</td>
<td>1,522,961</td>
<td>2,127 (13.8%)</td>
<td>881 (4.7%)</td>
<td>1,374 (6.7%)</td>
<td>147,911 (8.7%)</td>
<td>41,389 (11.6%)</td>
</tr>
<tr>
<td></td>
<td>V3million</td>
<td>02/16 – 03/23</td>
<td>1,332,279</td>
<td>3,300 (4.5%)</td>
<td>1,997 (9.6%)</td>
<td>3,149 (9.1%)</td>
<td>56,229 (3.6%)</td>
<td>8,067 (43.1%)</td>
<td>19,414 (34.3%)</td>
</tr>
<tr>
<td></td>
<td>OGUsers</td>
<td>04/17 – 02/19</td>
<td>1,665,800</td>
<td>442 (8.0%)</td>
<td>219 (1.2%)</td>
<td>355 (1.4%)</td>
<td>51,851 (3.1%)</td>
<td>56,229 (3.6%)</td>
<td>46 (0.01%)</td>
</tr>
<tr>
<td>Detected</td>
<td>Nulled</td>
<td>Our scrape</td>
<td>01/18 – 03/23</td>
<td>1,053,825</td>
<td>1,504 (2.7%)</td>
<td>42,720 (2.5%)</td>
<td>24,368 (6.8%)</td>
<td>100 (0.00%)</td>
<td>881 (4.7%)</td>
</tr>
<tr>
<td></td>
<td>Evolution</td>
<td>DNM Archives</td>
<td>02/14 – 11/14</td>
<td>9,384</td>
<td>69 (0.1%)</td>
<td>69 (0.4%)</td>
<td>63 (0.2%)</td>
<td>979 (0.06%)</td>
<td>979 (0.3%)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>-</td>
<td>12/06 – 03/23</td>
<td>17,340,789</td>
<td>56,229 (3.8%)</td>
<td>1,927 (3.8%)</td>
<td>46 (0.02%)</td>
<td>979 (0.06%)</td>
<td>46 (0.01%)</td>
</tr>
</tbody>
</table>

* OGUsers was hacked in May 2019. † Raid marketplace was seized by FBI on April 19, 2022. ** Evolution marketplace was shut down by its administrators in mid-March 2015.

---

**TABLE II: Comparisons on dataset volume between our study and other works**

<table>
<thead>
<tr>
<th>Market</th>
<th>Author / Measurement Date</th>
<th>Traces</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>HackForums</td>
<td>Zhang et al. [124] (01/09/2018)</td>
<td>238,212</td>
<td>74,909</td>
</tr>
<tr>
<td>Our work (09/2018)</td>
<td>12,916,668</td>
<td>480,101</td>
<td></td>
</tr>
<tr>
<td>Nulled</td>
<td>Zhang et al. [124] (09/2018)</td>
<td>356,605</td>
<td>118,738</td>
</tr>
<tr>
<td>Our work (09/2018)</td>
<td>525,169</td>
<td>76,668</td>
<td></td>
</tr>
<tr>
<td>MPH</td>
<td>Sun et al. [101] (09/2015 – 05/2016)</td>
<td>357,883</td>
<td>599,085</td>
</tr>
<tr>
<td>Our work (01/2015 – 05/2016)</td>
<td>121,486</td>
<td>599,085</td>
<td></td>
</tr>
<tr>
<td>BlackHatWorld</td>
<td>Portnoff et al. [91] (01/09/2005 – 03/2008)</td>
<td>7,270</td>
<td>8,718</td>
</tr>
<tr>
<td>Our work (01/09/2005 – 03/2008)</td>
<td>599,085</td>
<td>8,718</td>
<td></td>
</tr>
</tbody>
</table>

The Covid-19 pandemic, which aligns with the results of [40], [113] that reported a decrease in trading activities on the underground marketplaces after April 2020.

We also compared four markets in our dataset with the information included in three other works: Zhang et al. [124], Sun et al. [101] and Portnoff et al. [91], as shown in Table II. For traces, we directly compared the trace counts by dropping those whose creation dates did not fall into the measurement date mentioned by the authors. In terms of users, we counted all users whose account registration dates are prior to the end date of the measurement date, which is actually an upper bound. The comparison results show that the trace counts and the number of users in these marketplaces covered in our dataset mostly matches or surpasses those of earlier work.

**Fig. 2: Cumulative listing count of our scrape over time**

We also used a sub-forum called Vouch Copy Profiles [25] in MPH to identify additional groundtruth appraisers. This sub-forum serves as a platform for vendors to locate reliable appraisers and establish guidelines to regulate appraisers' behavior. Specifically, only appraisers who meet specific requirements (i.e., 1,500 posts, 3-month membership and who communicate using clear English) are allowed to post a thread.

---

4
with his/her profile information (e.g., number of previous transactions, number of appraisal reviews, contact method, and professional product categories). It is mandatory for appraisers to provide a link to their profile thread when requesting a vouch copy from the vendor. Once receiving the vouch copies, appraisers must post a “detailed, completely free of charge and strictly neutral review within 72 hours” under vendors’ listings. Vendors can also make comments on appraisal reviews to show appreciation or dissatisfaction. According to the rules, any misleading or copying of others’ reviews will be deleted or even lead to account suspension. We consider the thread authors in this sub-forum as groundwork appraisers. To glean their appraisal reviews, we first retrieved all their previous posts, then applied a review classifier (see details in III-C) to filter out non-review posts. We next removed the reviews which contained the words “buy” or “bought” and utilized other rules to find appraisal reviews with low false negative: the review must either contain the word “vouch” (to ensure it is a “vouch review” or a review respond to “vouch copy”) or the length of the review must have at least 50 words (based on our observation of appraisal reviews). In this way, we recognized 100 appraisers and 2,127 appraisal reviews from MPGH.

Afterward, we retrieved the content of each appraisal review by using the unique thread ID assigned to each thread, which is shared by the author’s listing and all other traces. We next extracted the thread’s first trace as the listing. Altogether, we identified 4,054 unique appraisal reviews, i.e., \( D_{gt} \) and 479 appraisers belonging to Hack Forums and MPGH marketplaces, as shown in Table I. We manually validated all appraisal reviews in \( D_{gt} \).

C. Appraisal Review Identification

We next moved to retrieve appraisal reviews posted by “unofficial” appraisers. In our study, we first trained a review classifier to identify all reviews in \( D_{all} \), then use a set of pre-defined keywords to match appraisal reviews with high confidence.

We first randomly sampled and annotated 2,400 reviews and 2,400 non-reviews groundtruth (300 in each forum for every class) from \( D_{all} \). Our annotation guideline is compatible with [36], [82]. Specifically, if a trace includes an evaluation of a product that shares opinions regarding its objective attributes, features, performance, quality, overall value, or other common characteristics based on usage experience, it is classified as a review. Note that for data annotation, each sample was labeled only when two annotators (both cybersecurity graduate students) agreed with each other. The inter-coder reliability, measured using Cohen’s kappa coefficients [80], was 0.79. Finally, we obtained 2,093 annotated reviews and 2,213 non-review instances that were mutually agreed on by both annotators. Next, we built a review classifier. In our implementation, we compare the performance of three DNN-based models (TextCNN [68], LSTM [60], BiLSTM [65]) and six statistical ML models (Support Vector Machine (SVM), Naïve Bayesian, Logistic Regression, K-Nearest Neighbors, Multi-Layer Perceptron, and Random Forest), which are widely used in review classification tasks [50], [56], [58] on our dataset. Specifically, in DNN models, feature engineering was implemented by utilizing the embedding layer which translated each word into a 256-dimension vector from scratch. In ML models, we adopted the approaches of previous work [50] and computed word-count-based vectors as inputs to the model. We evaluated the performance of our 9 classifiers by randomly splitting groundtruth into a training (90%) and a testing set (10%), and computed the recall and precision score for each model (see Table III). Ultimately the LSTM model was chosen as it outperformed other models on our dataset, having a recall score of 96.4% and precision scores of 93.1%. By applying our LSTM-based review classifier on \( D_{all} \), we identify 1,753,413 review traces for further filtering.

We then applied a list of keywords (in Table IV) along with their plural to all classified reviews to match appraisal reviews. Those keywords were decided by manually checking groundtruth appraisal reviews and 1000 random reviews from the classification results. We found that appraisers will indicate that a particular review is an appraisal review by clarifying she received a vouch copy, for example, “I got a vouch copy, here is my honest review...”. Our method is highly conservative and may result in false negatives, but this step was necessary to ensure that our subsequent studies focused only on the appraisal system. We also tried automated approaches, such as tf-idf and Word2Vec, to identify keywords. However, neither of them performed well due to the prevalent informal writing style in underground markets and the extensive usage of dark jargon, which often carry crime-related meanings [122]. For instance, dark jargons “rat” and “Illusion” mean “remote access trojan”, and “lôn” represent crypter, respectively. We evaluate our keyword match method by randomly selecting and labeling 200 positive samples from each market (all reviews for Evolution and Raid). It yields a precision of 97.8%. Note that determining the number of missed appraisal reviews is challenging. We elaborate on the discussion of potential false negatives in §VIII. Table I shows statistics of our results.

In addition, to study the effectiveness of the keyword-based approach, compared with an appraisal review classifier, we trained another LSTM classifier using 2,400 appraisal reviews from §III-B and 2,400 non-appraisal reviews. Then repeat the evaluation process above. The appraisal review classifier yields a precision of 93.7%. When running on \( D_{all} \setminus D_{gt} \) and randomly selecting and labeling 1,273 positive samples, this model reports a precision of 67.1%, which is far lower than 97.8%, in the keyword match method. It indicates that simply using \( D_{gt} \) to train a classifier is not as effective as the combination of review identification and keyword matching approach.

IV. CHARACTERISTICS OF APPRAISAL SYSTEM

A. Workflow of Appraisal System

Before coming to the details of our measurement findings, we first summarize the workflow of a typical appraisal system discovered in our research.

As shown in Figure 3 and Table V, our study outlines the workflow of an appraisal system in underground marketplaces, as well as a representative set of communication traces involved in the process. Specifically, the workflow consists of two steps: appraiser recruitment (➊-➋) and vouch copy appraisal (➋-➌). In our study, we observe two channels of appraiser recruitment: one is via vendors’ product listings (see Section V-B), where a vendor will include the requirements
of expected appraisers in his/her product listing to select the qualified appraisers (①-③); the other is via official appraiser groups of the marketplaces (see Section III-B), where a vendor will select an appraiser from the official appraiser group formed by the marketplaces (④-⑥). Moreover, we observed that both types of appraisers will post requests under vendors’ listings to promote themselves. Additionally, we noted that some non-official appraisers have begun offering their services to review products in underground marketplaces as a separate business. In the vouch copy appraisal step, the vendor will send out vouch copies to the selected appraisers in private (e.g., providing a download link in private messages, emails, or social messaging apps) (⑦) and then wait for their assessment (⑧). The appraisers are expected to post an in-depth appraisal review under the product listing (⑨). The vendor can also show appreciation or disagreement based on the content of these appraisal reviews (⑩). As a result, the appraisal reviews provided by these specialized appraisers offer valuable insights to other customers regarding the quality of the products, helping them make informed purchasing decisions. Consequently, the appraisal system has become a significant element in transactions, with some users choosing to wait for appraisers to review a product before making a purchase.

B. Scope and magnitude

Altogether, we identified 18,701 unique appraisers from 8 underground marketplaces during 2008 – 2023. Those appraisers posted 56,229 reviews in total under 25,836 unique listings. In particular, Hack Forums has the greatest number of appraisers (43.1%, 8,067), followed by BlackHatWorld marketplace (34.8%, 6,050) and V3million markets (9.6%, 1,797). Appraisers first appeared in three marketplaces before 2011: BlackHatWorld, Hack Forums, and MPGH. We consider this as a trial period because the monthly number of appraisers either fluctuated tremendously or increased at a low rate. During this time, the first appraiser group - Official Reviewers Group - started its recruitment in Jan 2011. Moreover, the earliest listings that mentioned providing vouch copies went back to Dec. 2008 and Oct. 2011 in BlackHatWorld and Hack Forums, respectively. After Feb. 2014, appraisers were observed once other newly-appeared marketplaces were launched; for instance, the Evolution launched in Jan 2014 and Nulled marketplace launched in Feb. 2015 and reloaded in April 2018. We provide a detailed analysis of the evolution of the number of appraisers in §V-A.

Regarding appraisal reviews, BlackHatWorld has the most reviews (26,304, 46.8%), followed by Hack Forums (19,414, 34.5%) and V3million (3,330, 5.9%). In our study, we found that official appraisers were more active than non-official appraisers. On average, each official appraiser in $D_{gt}$ contributed 6 reviews, while detected appraisers in $D_{det}$ wrote an average of three reviews. For example, one of the most active appraisers is A***a*† (also known as [MPGH]A***a*†), who has an appraiser profile in the Vouch Copy Profile sub-forum of the

### TABLE III: The evaluation of review classifier

<table>
<thead>
<tr>
<th>Review classifier</th>
<th>Evaluation metrics (Recall, Precision)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM</td>
<td>96.4/93.3</td>
</tr>
<tr>
<td>BiLSTM</td>
<td>94.4/91.9</td>
</tr>
<tr>
<td>TextCNN</td>
<td>93.1/92.3</td>
</tr>
<tr>
<td>SVM</td>
<td>93.2/89.0</td>
</tr>
<tr>
<td>Naive Bayesian</td>
<td>95.1/79.6</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>87.5/90.7</td>
</tr>
<tr>
<td>K-Nearest Neighbors (neighbors=3)</td>
<td>3.7/87.3</td>
</tr>
<tr>
<td>Multi-Layer Perceptron</td>
<td>89.3/89.8</td>
</tr>
<tr>
<td>Random Forest (max depth=2))</td>
<td>67.5/95.4</td>
</tr>
</tbody>
</table>

### TABLE IV: Keywords used to match appraisal reviews

| List of keywords | Vouch copy, review copy, trial copy, preview copy, free copy, free sample, free trial, free review, free service, vouch review, sample copy |

### TABLE V: Quotes from participants involved in appraisal system

<table>
<thead>
<tr>
<th>Step</th>
<th>Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quote</td>
<td>Requirements. If you are interested in applying to Reviewers, you must reach the following requirements: *** Application. If you'd like to become an Official Reviewer you will need to post a fake review in this thread.</td>
</tr>
<tr>
<td>Quote</td>
<td>How will I know if I am accepted? If you are accepted, you will be sent an acceptance PM. The PM will contain everything you need to know to help you get started.</td>
</tr>
<tr>
<td>Quote</td>
<td>Selling cracked NordVPN accounts from 15. Accounts up to 2024. I am going to give out 2-3 vouch copies with accounts that expire in a week or less to 133/ab/b36gar.</td>
</tr>
<tr>
<td>Quote</td>
<td>Hi, add me on Skype: x/x/f/f for vouch copy, tha.</td>
</tr>
<tr>
<td>Quote</td>
<td>Hi, was wondering if Fodooget A vouch copy of Homa*?</td>
</tr>
<tr>
<td>Quote</td>
<td>Vouch copies have been given out to T* and S* in **.</td>
</tr>
<tr>
<td>Quote</td>
<td>Th3shold: 4** Plagiarism: 5** + Additional Information: th.</td>
</tr>
<tr>
<td>Quote</td>
<td>E-book Name: *** Download Link: *** Reviewers required: 2</td>
</tr>
<tr>
<td>Quote</td>
<td>Pending review by the official Review Team. This post will be updated with a full review within 48 hours.</td>
</tr>
<tr>
<td>Quote</td>
<td>Hi, added to Vouchcopy, thx.</td>
</tr>
<tr>
<td>Quote</td>
<td>Continuing this I have received a vouch copy of the Platinum Package and will be posting my review soon.</td>
</tr>
<tr>
<td>Quote</td>
<td>I have received a vouch copy and I'm not really satisfied with this user, this is my HONEST review on S***N. Originality.</td>
</tr>
<tr>
<td>Quote</td>
<td>10/10 **. Ease of use. 9/10 **. Grammar. 10/10 **. Overall. 9/10. This eBook is no bullshit. The method in it works 100% and is autopilot. **. Get the book while its cheap!</td>
</tr>
<tr>
<td>Quote</td>
<td>clear instructions, and provided answers that I need. If you need someone to Vouch you, this guy can do it perfect.</td>
</tr>
</tbody>
</table>

---

†: Also known as [MPGH]A***a*†
from 5 to 257, a staggering growth rate of 5.040%. Similarly, in BlackHatWorld, the number of appraisers increased by 3,800% (from 5 to 195).

Moreover, we found that the marketplaces launched after 2014 immediately had appraisers appearing as they started off, including the Tor-based marketplace Evolution. With one year right after the launch of Nulled, V3rmillion and MPGH, the number of appraisers increased by 170%, 97% and 600%, respectively. For Nulled, the appraisal system was integrated when the marketplace was relaunched in January 2018. As a result, the ratio of appraisers to vendors was relatively high, since there were only a limited number of vendors on the platform at that time. Notably, even though Evolution was short-lived, the number of appraisers increased from 1 in March 2014 to 22 in June 2014. We also observed that the Covid-19 pandemic had a significant impact on all marketplaces in terms of the number of listings and vendors. However, the ratio of appraisers to vendors remained stable. As an example, the number of vendors in Hack Forums decreased from 890 to 111 (a decrease of 88%), while the appraisers decreased from 68 to 34 (a decrease of 50%) between Mar 2020 and Dec 2022. The significant drop in vendors caused a slight increase in the ratio of appraisers to vendors in Hack Forums starting from 2020, as shown in Figure 4. The main takeaway is that the appraisal system of marketplaces requires a certain number of active appraisers to support its operation, regardless of the popularity of the marketplaces.

We found it interesting that in BlackHatWorld each vendor on average offers 13 vouch copies for appraisal, likely because SEO and social media-related products make up the majority of this marketplace. More specifically, vendors in this marketplace typically hold a large number of backlinks and social media promotions, such as followers, likes, and views. They provide several vouch copies, each containing only a small number of backlinks or followers, to multiple appraisers. In this way, vendors can receive vouches from multiple appraisers and boost their listings with the help of appraisal reviews.

B. Characterizing the Appraiser Role

We characterized the appraiser role based on the feedback of appraisers. First, we categorize appraisers into three groups: Official Appraisers, Official Reviewers Group and Private Reviewers. Official Appraisers are members of the marketplace who have been approved by the marketplace administrators. Official Reviewers Group is a group of appraisers who are active in the forum. Private Reviewers are appraisers who are not active in the forum.

Finding II: Vendors adhere to strict rules when selecting appraisers, as they aim to ensure that only trustworthy appraisers are involved in the process.

Merits for appraiser selection. As mentioned earlier, appraisers can be recruited either by official appraisal groups or by vendors who provide appraiser requirements in their listings. To examine the criteria used to select appraisers and the features that distinguish them from general users, we analyzed the requirements listed in recruitment traces of appraisal groups as well as those specified by vendors in their listings. Specifically, we identified appraisal selection criteria such as VIP status, minimum number of posts, length of membership, and reputation score from the recruitment posts of two appraisal groups: the Official Reviewers Group and the Official Appraisers. We then used these criteria to match the appraiser requirements specified in vendors’ listings (if they existed). Additionally, we manually checked the matched vendors’ listings to identify any other selection criteria, such as whether the appraiser was a staff member of the marketplace.

A. Prevalence of active appraisers

Finding I: Today a certain number of appraisers exist in most underground marketplaces to support their operation.

Figure 4 shows the evolution of the proportion of active appraisers to active vendors in eight marketplaces. This ratio was used instead of directly showing the number of monthly active appraisers, as it allows for analysis together with natural changes in the products and user population of respective communities. We adopted the definition of active sellers in [99] to describe active appraisers. Specifically, we first choose a time period $\tau$ (we set $\tau$ to three months here). An appraiser is considered active at time $T$ if she posted (1) at least one review during $T - \tau \sim T$, and (2) at least one review during $T \sim T + \tau$. This definition considers the time that appraisers spend on evaluating the product before posting reviews.

As shown in Figure 4, the proportion of appraisers in BlackHatWorld and Hack Forums, the two marketplaces that implemented the appraisal system as early as 2008, experienced a rapid increase before reaching a level of stability around 2015. In particular, during the time period of 2009 and 2014, the average number of monthly listings and vendors in Hack Forums increased from 1,734 to 5,161 with a growth rate of 198%, and from 850 to 2,673 with a rate of 214%, respectively. Meanwhile, the number of appraisers increased...

1We also calculate the “survivability” of those appraisers: a given appraiser is still active and posting reviews after a certain number of days. On average, appraisers are active for 227 days.
In total, we identified five merits that are commonly-used to determine whether a user can be selected as an appraiser and recognized 782 unique listings which mentioned appraiser requirements. The most frequently used criterion was VIP status, with 67.3% of the listings requiring it. The second most common criterion was the minimum number of posts a user had made, with 29.2% of the listings requiring it. Additionally, some vendors preferred to rely on marketplace staff (12%) to perform appraisals. It is worth noting that some criteria may be specific to certain marketplaces and may not be available in a user’s profile in other marketplaces. Vendors may also use a combination of these criteria to select qualified appraisers. Table VI summarizes these merits. Below we further examine whether the appraisers selected by vendors meet the identified criteria, and compare their merits with those of non-appraisers, such as buyers who leave reviews after making a purchase.

- **Minimum number of posts.** The total number of posts can indicate the activity of a user in a marketplace. It is also one of the mostly-used rules by vendors to find an appraiser. We observed that 106 listings (46.5%) required appraisers to have at least 500 posts to apply for review copies. After checking the profiles of appraisers, we found that 93.9% of them met this requirement. Furthermore, appraisers had an average number of posts over 1,000, which is significantly higher than that of non-appraisers (242), as shown in Table VI.

- **Whether VIP.** At some marketplaces such as BlackHatWorld, Hack Forums, and Nulled, users can upgrade their accounts by purchasing VIP memberships to gain access to additional features and privileges (e.g., access to a VIP forum). For instance, Hack Forums offers a L33t membership that costs $25 for a 6-month upgrade. We identified 526 listings across the three marketplaces of BlackHatWorld, Hack Forums, and Nulled that required a VIP membership before a user could apply for a vouch copy. This is the most commonly used requirement by vendors. For instance, a listing may state that "2 Free Review Copies of STARTER Package are Only Available for Jr.VIP and Higher Members."

- **Minimum reputation score.** In underground marketplaces, a user’s reputation score is determined by the ratings they receive from other users, which can be positive (increasing the score) or negative (decreasing the score). The value of the reputation score varies from 25 to 273 on average, depending on the difficulty of acquiring a reputation score. From Table VI, it can be seen that both ground truth and detected appraisers have a significantly higher reputation score than non-appraisers.

- **Length of membership.** The length of membership is a measure of how long a user has been registered as a member on a marketplace. In our study, we found the requested length of membership for appraisers varies from one month to one year. We calculated the membership length for each user by measuring the time between their registration date and their last post. Our results show that there is a small difference in the average membership length between appraisers (816 days) and non-appraisers (630 days).

- **Other requirements.** After conducting manual checks on vendors’ listings, we discovered that some vendors choose marketplace staff members (administrators or moderators) as appraisers. In total, we found 94 listings where review copies were only offered to staff members. Some vendors explained that they chose staff members due to their high prestige and popularity in the marketplace, and because they are considered trustworthy and “out of the concerns of leaking out.”

**Finding III:** Less-trusted appraisers may post fake reviews to promote vendor’s sale, but they are likely to be reported once victims of the scam appear.

**Less-trusted appraisers.** To understand the credibility of non-official appraisers, we looked into the scam reporting subforum of each marketplace to determine if there were any scam reports linked to the appraisers we had discovered. Note that those sub-forums also help marketplace administrators identify potentially untrustworthy appraisers who may be falsifying product reviews. Specifically, we first utilized the dependency parser in spaCy [21] to extract all subjects (with the label of nsubj), objects (with the label of dobj), and object of a preposition (with the label of pobj) in each sentence of these reports. Then we filtered the reports that contained at least one appraiser’s username in any of the positions above (i.e., nsubj, dobj, and pobj), before we used a list of keywords related to a vouch copy or an appraisal review to identify the posts related to appraisers’ vouch activities. For instance, our method will flag the report with title “L***p Posted Fake Vouch for D***e” along with the appraiser, i.e., L***p. This process identified 82 unique reporting threads and 22 (0.2%) less-trusted appraisers. The average length of membership of those appraisers is 327 days and the number of previous posts is 2,461.

After manually examining 22 scam reports, we identified three categories of reported suspicious appraisers, who might post fake vouch reviews. First, suspicious appraisers have a personal relationship with the vouch copy provider. The appraiser will post a positive but fake review per the vendor’s request, even though he might not receive any free vouch copy. For instance, in report “Scam Assist by P***k - $75” in Hack Forums, the appraiser with user ID P***k was blamed for posting a fake appraisal review, leading numerous users to get scammed by the vendor who sold Motorola Xoom tablets. In the report, P***k admitted that he actually did not receive the tablets and the vendor is one of his friends. Surprisingly, the report author also found evidence that this appraiser had received $100 for writing appraisal reviews for the vendor.

The second type of suspicious appraisers are those who use multiple accounts created by the same vendor. In other words, vendors use different aliases to pretend to be appraisers and post appraisal reviews for themselves. For example, in the report with the title “A***x is a Scammer” in MPGH, the author requested the marketplace administrator to check the IP address of the appraiser, which confirmed that A***x and L***n were the same person.

The third type is a fake appraiser group. The goal of this kind of group is to bump the sale thread to the top pages by posting fake appraisal reviews. For instance, in MPGH, a user posted a report with the title: “Apology Regarding Jades Expose (Full Story)” and mentioned his experience in the fake appraiser group. Specifically, a vendor called U*** invited him to join the so-called scamming group. After joining, he was told “that all he needed to do was vouch or ask for a vouch copy then type something like “AWESOME” or “100%
TABLE VI: Merits for appraiser selection

<table>
<thead>
<tr>
<th>Merit (non-exhaustive list)</th>
<th># of listing (%)</th>
<th>Value set by vendor in average (if applicable)</th>
<th># of groundtruth appraiser meet requirements (%)</th>
<th># of detected appraiser meet requirements (%)</th>
<th>Groundtruth appraiser’s merit value in average</th>
<th>Detected appraiser’s merit value in average</th>
<th>Non-appraiser’s merit value in average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min # of posts</td>
<td>228 (29.2%)</td>
<td>618</td>
<td>25 (100%)</td>
<td>408 (92.9%)</td>
<td>3,788</td>
<td>1,138</td>
<td>242</td>
</tr>
<tr>
<td>Whether VIP†</td>
<td>526 (67.3%)</td>
<td>BlackHatWorld: Jr. VIP</td>
<td>Hack Forums: UBRR/L33T</td>
<td>MPGH: Premium Member</td>
<td>Nulled: VIP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min reputation score</td>
<td>69 (8.9%)</td>
<td>BlackHatWorld: 80</td>
<td>Hack Forums: 273</td>
<td>MPGH: 106</td>
<td>Nulled: 35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hack Forums: Jr. VIP</td>
<td>MPGH: Premium Member</td>
<td>Nulled: VIP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of member (days)</td>
<td>26 (1.3%)</td>
<td>103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whether staff†</td>
<td>94 (12%)</td>
<td>Administrator/moderator/staff</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The users’ VIP status and type are only available in the following two marketplaces in our dataset: Hack Forums and Nulled.
† Users’ reputation score system in Hack Forum was deleted in June, 2018 and replaced with contracts system.
‡ Whether a user is a staff can be found from the following two marketplaces in our dataset: Hack Forums (users with staff badge) and Nulled (users’ group information).

LEGIT” and more to review”. With such experience, the report author realized that “no wonder why it got so many vouches & stuff quickly”.

Finding IV: Some appraisers may write biased reviews that collude with vendors to promote their products or threaten vendors to obtain a vouch copy.

Appraiser’s private interactions with vendors. In this section, we investigated the private messages exchanged between appraisers and vendors on the underground marketplace Nulled. Specifically, we analyzed a leaked dataset from Nulled and retrieved all messages involving both appraisers and vendors in the same listing. Next we used a set of keywords in Table IV to find all communications related to appraisal activities. In total, we found 26 conversations containing 275 (0.3%) communication traces related to appraisals. Manually checking each communication yielded three observations.

- **Review template.** We observed three cases in which vendors interfered with appraisers’ reviews to manipulate the content and better promote their sales. For example, one vendor provided a template for a positive review that the appraiser could use:

  “Vendor: Hey, could you maybe vouch for my sales here: [link]. If you want say smth like: The seller gave me an vouch copy got the scraper in a rar, works fast and doesn’t crash just like the seller said! Would recommend this to everyone.”

Other vendors may also instruct appraisers on what should not be included in their appraisal review to prevent sensitive information about the product from being leaked:

  “Vendor: Don’t post any screenshots please, event names are secret and should not be shared to public.”

- **Negative review in exchange for a vouch copy.** We came across an appraiser who intentionally posted a negative review about the vendor’s sale listing, despite not having received a vouch copy. This was done with the intention of bargaining with the vendor to receive a vouch copy:

  “Vendor: Hi, I seen your left negative rating in my thread. Sir I am newbie here. Looking to sell my new product. If you need Vouch copy I will give you now. Please remove that negative Rep. Please I am selling legit 100% working coins. [link]. Thanks Appraiser: if you give a vouch copy, I will change my opinion.”

- **Technical support from the vendor.** The technical support provided by sellers helps appraisers to successfully test their products, which can prevent negative reviews to some extent. One of the five cases we observed looks like this:

  “Appraiser: Hey it didn’t work unfortunately :P Vendor: You need to use the account’s username if you want to use on PSN, mobile, etc. Try this:[username]. If it works, please leave some positive feedback on the original thread so people know it’s legitimate. Appraiser: My bad it worked, thanks! will leave a review!”

VI. Appraisal Merits

In this section, we present an analysis of appraisal reviews. Starting with an overview of product categories being reviewed, we then elaborate on the assessment standards that appraisers use when evaluating a product.

A. Items to appraisal

Finding V: Most appraisers prefer appraising products in the same category.

Appraisal item analysis. In this section, we analyze the product categories in appraisal reviews. Specifically, we started by identifying eleven categories of items that were appraised. Next, we used a state-of-the-art illicit product classifier [99], [110] trained on underground forum corpus to categorize the listings of vendors. Finally, we linked each appraisal review to its corresponding listing category.

As shown in Table VII, we determined eleven product categories based on our dataset and previous cybercrime studies [37], [105], [110]: account, social booster services, email, video game, malware, RAT, botnet, website, hosting, making
TABLE VII: Appraiser and appraisal review per category

<table>
<thead>
<tr>
<th>Category</th>
<th># appraisal review (%)</th>
<th># appraiser (%)</th>
<th># appraisal listing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Website</td>
<td>19,765 (35.2%)</td>
<td>4,002 (21.4%)</td>
<td>3,903 (15.1%)</td>
</tr>
<tr>
<td>Making money</td>
<td>14,819 (26.4%)</td>
<td>4,189 (22.4%)</td>
<td>8,881 (35.3%)</td>
</tr>
<tr>
<td>Account</td>
<td>5,940 (10.6%)</td>
<td>2,618 (14.0%)</td>
<td>3,978 (15.1%)</td>
</tr>
<tr>
<td>Other</td>
<td>5,374 (9.6%)</td>
<td>2,431 (13.0%)</td>
<td>2,581 (10.3%)</td>
</tr>
<tr>
<td>Social booster</td>
<td>3,372 (6.6%)</td>
<td>1,851 (9.9%)</td>
<td>1,838 (7.3%)</td>
</tr>
<tr>
<td>Malware</td>
<td>2,443 (4.3%)</td>
<td>1,246 (2.2%)</td>
<td>1,768 (7.0%)</td>
</tr>
<tr>
<td>Game</td>
<td>1,246 (2.2%)</td>
<td>649 (1.2%)</td>
<td>1,000 (2.8%)</td>
</tr>
<tr>
<td>Hosting</td>
<td>847 (1.5%)</td>
<td>467 (2.5%)</td>
<td>517 (2.1%)</td>
</tr>
<tr>
<td>Botnet</td>
<td>655 (1.2%)</td>
<td>355 (2.0%)</td>
<td>503 (2.0%)</td>
</tr>
<tr>
<td>Trojan</td>
<td>649 (1.2%)</td>
<td>355 (2.0%)</td>
<td>498 (2.0%)</td>
</tr>
<tr>
<td>Email</td>
<td>759 (1.3%)</td>
<td>430 (2.3%)</td>
<td>241 (1.0%)</td>
</tr>
<tr>
<td>Total</td>
<td>56,629</td>
<td>18,701</td>
<td>25,836</td>
</tr>
</tbody>
</table>

money guide and others. We present the detailed description for each category below:

- **Account.** The listings in this category mainly contain two types of products: 1) vendors who sell either single or many accounts in social media (i.e., Twitter and Instagram), video games and streaming (i.e., Netflix and Spotify), and 2) the tools for generating or cracking accounts in bulk.

- **Social booster services.** In the category of social booster services, the anonymous merchants aim to supply a range of synthetic followers, views, likes and subscribers.

- **Email.** Within the email category, the two main themes are 1) spamming services or tools for emails or SMS; and 2) an email list for spamming.

- **Video game.** The prominent product in this category is game cheat for different purposes, such as power-leveling, rank boosting, auto-run game bots and unlimited resources in video games.

- **Malware.** The malware category is comprised of various malicious apps such as crypter, exploit kit, ransomware, worm, keylogger, cryptojacking (miner) and virus spreader. The keylogger and cryptojacking are prominent products in this category.

- **RAT.** Generic RATs (e.g., Blackshades and jRAT) or malicious RDP services are the prominent products in this category.

- **Botnet.** The listings in this category are selling slaves/bots, tutorials, and platforms related to botnet or DDoS services.

- **Website.** This category is composed of two types of products: 1) blackhat/greyhat search engine optimization (SEO) techniques; and 2) VPN connections and proxies.

- **Hosting.** This category contains listings that sell web hosting services through VPS.

- **Making money guide.** These listings sell tutorials about how to earn money such as eWhoring and gambling.

- **Others.** Others contain listings that do not fall into the previous categories. It contains products related to physical items, gift cards, coupons and so on.

To train the illicit product classifier, we manually labeled 1,500 (5% of all) unique listings as groundtruth while ensuring each category contained at least 50 samples. Each listing sample is a concatenation of its title and product description, after we cleaned the data by removing stop words, punctuations, numbers and URLs. We tokenized each sample and calculated a tf-idf value for each word for each sample, and used these values as inputs to a Linear SVM under L2-Loss classifier implemented with scikit-learn. We also utilized SMOTE [44] to mitigate the imbalance in the distribution of the listing categories. The performance of the classifier was evaluated using another 300 labeled samples. The average precision is 87% and the average recall is 82%. The confusion matrix is shown in Figure 5.

**Findings.** As shown in Table VII, we observed that website-related products have appeared in most appraisal reviews (35.2%, 19,765), followed by making money guides and accounts. This is because website-related products make up the majority of the BlackHatWorld marketplace, which has the most appraisal reviews.

We also studied the appraiser’s preferences for appraising either a specific product category or across multiple categories. Figure 6 shows the distribution of appraisers across various product categories. The value on the diagonal indicates the number of appraisers who had previously posted at least one appraisal review in the same category. From the figure, we found that most appraisers prefer appraising products in the same category, which could be an attempt to establish themselves as experts in a specific product category and increase their chances of being selected by vendors.

**B. Assessment merits**

We further studied the merits that appraisers use to assess products/services. To summarize those merits in each product category, we leveraged the review template of two official appraiser groups (as illustrated in Section III-B) and individual appraisers (see an example in Table V), to understand how appraisers provide product assessment from different
making money 955 5 6 0 0 19 0 0 0 0 0
account 5 400 10 0 0 36 0 4 0 0 0
social booster 6 10 270 0 0 90 0 2 0 0 0
game 0 0 0 61 0 0 0 0 0 0 0
trojan 0 0 0 0 0 0 0 0 0 0 0
website 19 36 90 0 0 1213 0 12 0 1 0
malware 0 0 0 0 0 211 0 0 0 0 0
hosting 0 4 2 0 0 12 0 34 0 0 0
botnet 0 0 0 0 0 0 0 0 31 0 0
email 0 0 0 0 0 1 0 0 0 8 0

Fig. 6: Appraisers across multiple product categories

TABLE VIII: Assessment metrics

<table>
<thead>
<tr>
<th>Product category</th>
<th>Assessment merit (non-exhaustive list)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Making money guide</td>
<td>Grammar/language, method originality/uniqueness, ease of use, design/layout, content, price, method, support, compatibility, effectiveness, profit, installation time, functionality</td>
</tr>
<tr>
<td>Social booster services</td>
<td>Support, followers, views, likes, communication, retweets, subscribers, price</td>
</tr>
<tr>
<td>Website</td>
<td>Turnaround time (TAT), backlink features (types, number of received backlinks, ranking scores and domain age), communication, price, support, SERP boosting results, keyword features (# of searches, competition and KD value)</td>
</tr>
<tr>
<td>Malware</td>
<td>Ease of use, file size, price, GUI/panel design, support, features, detection, stability, installation, functionality, compatibility, performance</td>
</tr>
<tr>
<td>Hosting</td>
<td>Storage, speed, memory, support, uptime, bandwidth, databases</td>
</tr>
<tr>
<td>Video game</td>
<td>Aminbot, triggerbot, visuals, support, radar, bunnyhop, RCS (recoil control system)</td>
</tr>
</tbody>
</table>

perspectives. More specifically, we designed a regex, ([a-z]+):\s*\d*([a-z]+), to extract the merits that appraisers mentioned in their reviews (i.e., Support: 10/10). Next, we ranked those merits according to their frequency and selected the top 20 for each product category. We then manually identified and combined the merits in the same category. In total, we found 4,363 appraisal reviews using templates and extracted 49 unique merits. Table VIII shows the result.

We further summarized the assessment merits into 4 categories:

- **Delivery speed.** The turnaround time (TAT) and speed are merits used by appraisers to describe the product delivery speed. Appraisers often use adjectives like "fast" or quantitative terms, such as "I received the review confirmation on the 23rd and received the report on the 26th," to accurately describe the time span.

- **Product characteristics.** The merits in this category vary across different product categories. For instance, when appraising malware, appraisers consider factors such as the design of the control panel (including its theme color, functionality description, and button design), detectability from scan results, stability (e.g., whether the malware crashes or freezes), installation time, and functionality (including features such as keylogger, file explorer, webcam viewer, mutex, and BSoD on termination of the client). Regarding making money guides, the appraiser will emphasize grammar (checking for errors, typos, jargon, and confusing sentences), uniqueness of the method (determining whether it's a common or saturated approach), design (e.g., evaluating the layout, organization, and inclusion of visuals), easy to use (time cost to set up and begin to make a profit), and investment (the amount of initial capital required).

- **Customer service.** Appraisers commonly used two merits – communication and support – to describe their interactions with vendors. They noted whether vendors responded quickly and provided detailed explanations for frequently asked questions (FAQs).

- **Value.** Appraisal will evaluate whether the listing price is reasonable or not, compared to the product's quality and functionality. Such information is usually recorded in the "price" merit.

VII. THREAT INTELLIGENCE IN APPRAISAL REVIEW

In this section, we shed light on appraisal reviews as a new source of CTIs. We first identified CTI taxonomy for products in the underground marketplaces through a literature review as detailed in §VII-A. Further §VII-B details how we retrieve CTIs. In §VII-C, we compare appraisal reviews with other sources of CTIs to reveal how they supplement other sources and provide new insights into underground illicit products.

A. A Taxonomy of CTI

Finding VI: The CTI obtained from appraisal reviews can enhance our understanding of underground illicit products.

As described in §VI, we observed that appraisers often discuss the usage and quality of products in their reviews, which indicates that appraisal reviews can provide valuable CTI. To identify useful CTI that can enhance the detection of cybercriminal activities, we defined a taxonomy of CTI associated with products in underground marketplaces. Note that existing CTI taxonomies (e.g., OpenIOC [104], STIX [100] and yara [120]) mainly focus on malware, but not other products (e.g., social boosting services, or blackhat SEO tools) in the underground marketplaces. However, the CTI associated with those products also provides valuable information for detecting and mitigating cybercrime activities (e.g., fraudulent accounts and malicious website detection).

We developed our taxonomy through a literature review. In particular, we manually examined the last eleven years of research from IEEE S&P, USENIX Security, CCS, NDSS, WWW and APWG eCrime. Our team focused on topics related to blackhat SEO campaign detection, malware detection, fraudulent account detection, cybercrime, and underground marketplace. We then manually searched through the related work of these papers for relevant research. In total, we reviewed 94 articles and papers on the topic of cybercrime ranging from 2011 - 2022. We next conducted a thorough analysis of the features used in previous studies to identify and classify
cybercriminal activities. For instance, some studies leveraged the number of followers to detect fake social media accounts. We consider these features as valuable CTI from underground marketplaces. Altogether, we identified 41 unique types of CTI among 3 product categories (website, account, and malware) shown in Table IX. We also elaborated on our systematized work below.

**Website.** CTIs in the website category are mostly related to black-hat SEO techniques. In on-page SEO optimization, keyword selection is the most important part in terms of generating relevant terms and articles [54]. Previous work shows that attackers preferred targeting low competition keywords [53], [76] such as long-tail keywords [72] but with high search volume [118] and cost per click (CPC) [64]. In article spinning which aims to create the deceitful appearance of what appears to be new content so as to distinguish it from what already exists, Shahid et al. [95] utilized some basic lexical features (i.e., word count and sentence count) of generated articles to detect spun content and its seed without needing the text spinner’s dictionary. The length of an article is also used to identify a malicious website for phishing and spam [74], [84], [90].

In off-page SEO techniques, previous studies have examined how to use the merits of backlinks to detect malicious activities, for example, using the number of backlinks to detect websites created through private blog networks (PBNs) [109] or link spam [33], [116]. Other work leveraged the search engine ranking scores (i.e., page ranking (PR), domain authority (DA) and page authority (PA)). For instance, Du et al. [53] infiltrated the spider pool - a new type of blackhat SEO infrastructure which constructs link networks using cheap domains with low PR. Others [47], [97], [98] made use of those scores to detect spam links.

**Account.** Previous research has investigated how to leverage the account features to detect fake accounts. Prominent examples include using the age of the account (in days) [32], [125]; or the number of characters in the profile description [45], [96]; or whether an account has a profile picture [38], [96]; or whether it has been phone verified at registration time [121] to identify spammers in a social network. In addition, the verification status of an account was utilized as a useful indicator to detect fake news spread through social media [75], [107], [119].

We also took social booster service into consideration here, as it is related to social media account. Those features include the number of followers [29], [30], tweets [42], friends [107], and YouTube video views [108], [111] which are used to evaluate the credibility of an account. Jang et al. [66] leveraged the geographical location and distance of followers in an online social graph to detect fake followers in Twitter.

**Malware.** CTIs in the malware category occur primarily in the form of Indicators of Compromise (IOC) which are forensic artifacts of an intrusion such as SHA256 hashes of attack files and malicious file sizes [73]. We adopted the CTIs pre-defined in existing IOC frameworks including OpenIOC [104], STIX [100] and yara [120] used for identifying a known malware, an attacker’s methodology, or other evidence of a compromise.

### B. CTI Extraction

To collect CTIs automatically from the appraisal reviews within the corresponding category, we adopted both regex and the state-of-the-art CTI extraction approaches (i.e., named-entity recognition (NER)-based methods) [62], [69], [83], [115]. Specifically, for CTIs that have fixed patterns (i.e., MD5 hash: 9e***d4), we designed a set of regex to retrieve their value with high accuracy. We show four CTIs along with their values within the corresponding category, we adopted both regex and the state-of-the-art CTI extraction approaches (i.e., named-entity recognition (NER)-based methods) [62], [69], [83], [115].

For the remaining CTIs, we used the state-of-the-art CTI extraction approaches (i.e., named-entity recognition (NER)-based methods) [62], [69], [115] to label CTIs and their values within sentences. In our implementation, we adopted spaCy’s NER engine [62] – an existing NER model which uses deep learning.
convolutional neural networks – and adapted the pre-trained en_core_web_sm model to the threat intelligence domain. To train the model, we first randomly selected and annotated 100 unique reviews for each CTI in Table XI. Then we adapted the pre-trained model to the CTI domain by performing 1,500 iterations over the annotated training data, shuffling at each epoch, and using minibatch training with a batch size of 4, to update weights in the neural network, while preserving all other components (i.e., tokenization, word2vec, etc.) in the pipeline. We evaluated the performance of the model using an additional 50 randomly selected samples for each entity. Table XI shows the results of the test dataset.

After applying the CTI extraction model to 33,184 appraisal reviews in website, account and malware product categories, we extracted 23,978 CTIs belonging to 16,668 (50.2%) reviews. The website category has the highest number of CTI, with 15,508 (64.7%) instances from 10,258 (61.5%) appraisal reviews by 2,914 appraisers, and the malware category has 1,371 (5.7%) instances from 5,056 (30.3%) reviews by 1,354 (8.1%) appraisers. We also observed that 8.9% (105,020) listings and 2.7% (45,727) non-appraisals contain CTIs, compared to 50.2% (16,668) appraisal reviews. In addition, when comparing three product categories (website, account, and malware), appraisal reviews containing CTIs consistently exhibit a higher proportion than listings and non-appraisal reviews. This is particularly noticeable in the website category, where almost 78.5% of appraisal reviews contain CTIs, in contrast to only 0.2% in listings and 0.7% in non-appraisal reviews. The main takeaway here is the appraisal reviews exhibit a higher CTI density compared to other sources that have a much larger scale in underground marketplaces.

We next compared the overlap of CTIs in appraisal reviews, listings, and non-appraisal reviews. More specifically, we selected 5 CTI categories under the malware category: file hash, filename, whether need dependency, whether scan time or run time undetectable, and detection result. Unlike other features from the account or website product category, these malware-related CTI values are unique to specific malware and can be easily distinguished. In total, we identified 526 CTIs within these 5 CTI categories from 735 appraisal reviews. Only 34 CTIs are shared with listings and non-appraisal reviews. All the remaining 492 (93.5%) CTIs can only be found from appraisal reviews. For instance, in the “V ox Office Builder” listing, the appraiser provided comprehensive information about the builder, which included the names of three files, their respective file sizes, hashes (MD5 and SHA1), and their antivirus scan results. It shows that the appraisal reviews can serve as a supplementary source of CTI, adding substantial value to the comprehension of these illicit products.

Interestingly, in addition to the CTI values that were either shared with or uncovered by appraisal reviews, we observed six cases where appraisal reviews provided different CTI values than those offered by listings. Specifically, those appraisers found that the malware sold by the vendor is actually not either scan-time undetectable or run-time undetectable, despite the vendor asserting in listings that the malware is fully undetectable (FUD). For example, in the listing that sold “Hyper Downloader”, the vendor showed its clean antivirus (AV) run-time results. However, the appraiser said “When I try to run it, my Chinese version 360 AV picked it up immediately”.

We also observed some CTI categories that are not covered by the listings. The appraisers may express their personal opinions regarding the price set by the vendor. For instance, “I received a vouch copy. I feel like this ebook is all public information regarding the price set by the vendor. For instance, “I received a vouch copy. I feel like this ebook is all public

<table>
<thead>
<tr>
<th>Category</th>
<th>CTI</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Followers</td>
<td>84.4%</td>
<td>87.1%</td>
<td>85.7%</td>
<td></td>
</tr>
<tr>
<td>Views</td>
<td>83.2%</td>
<td>82.1%</td>
<td>82.6%</td>
<td></td>
</tr>
<tr>
<td>Likes</td>
<td>84.3%</td>
<td>85.1%</td>
<td>84.6%</td>
<td></td>
</tr>
<tr>
<td>Subscribers</td>
<td>85.1%</td>
<td>83.2%</td>
<td>84.2%</td>
<td></td>
</tr>
</tbody>
</table>

We then study how CTIs extracted from appraisal reviews supplement public CTI sources (i.e., VirusTotal [112], DigitalSide [52] and industrial white papers).

Specifically, we submitted 493 malware hashes extracted from appraisal reviews to both VirusTotal and DigitalSide. Surprisingly, only 2 (0.4%) hashes are labeled as malicious.
(1.8%) are labeled as benign, and the remaining 482 (97.8%) do not have a linked record in either platform. It is possible that those attackers use local AV versions and do not submit samples to VT to avoid sharing.

We also compared CTIs extracted from appraisal reviews with CTIs in white papers. Specifically, we identified 14 popular organizations who collected CTIs from underground marketplaces and forums, namely: Trellix by FireEye [23], Recorded Future [16], ZeroFox [27], Digital Shadows [6], Cyjax [5], Red Sky Alliance by Wapack Labs [17], Eclec-
ticIQ [7], ThreatStream by Anomali [1], CTI League [4], ThreatConnect [22], Qualys [15], Skurio [20], F5 Labs [8], and Intel 471 [10]. We next collected white papers from each company’s official site that discussed threat intelligence from underground markets and forums. Here we took advantage of search engines embedded within the site and applied a set of keywords (i.e., underground marketplace/forum, darknet, dark web and threat intelligence) to it if applicable. In total, we collected 218 white papers, reports and blogs dating from between 2016 to 2022. We then manually investigated each of them to find if any of the CTIs defined in Table IX were mentioned. Table XIII shows the results. Altogether, we found 9 (64%) cybersecurity companies whose 58 (27%) unique white papers mentioned the CTIs defined by us.

We observed that 1) Only 2 (1%) white papers cover the topic in the website and account category. Most of those white papers only focused on malware (especially ransomware, exploit and trojan). Specifically, out of the total 218 white papers, 166 (76%) included a discussion on ransomware, detailing their initial appearance date, version, target victims, attack procedure, CVE, hashes, and more. The appraisers focused on other aspects such as the encryption algorithm, price on underground marketplaces, and detectability by antivirus scanners. Another 42 (19.3%) white papers provided analysis on exploits and trojans, covering their variants, IP addresses, domain names, targeted systems, and more. Appraisers, on the other hand, evaluated these illicit products based on their ability to bypass User Account Control (UAC), detection rates, filenames and sizes. Such thorough assessments by appraisers have significantly enhanced our understanding of malware by providing valuable additional information. Furthermore, our manual analysis revealed that appraisers evaluate a broad range of illicit products, such as crypters, miners, keyloggers, botnets, builders, worms, stealers, exploit kits, and more. This diverse range that appraisers target is expected to greatly supplement the intelligence collection efforts, providing a more comprehensive understanding of the threat landscape; 2) The threat intelligence collected by those companies is mainly

<table>
<thead>
<tr>
<th>Category</th>
<th>CTI</th>
<th>Company (# of white paper mentioned)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating system</td>
<td>FireEye (4), ZeroFox (3), Wapack Labs (3)</td>
<td></td>
</tr>
<tr>
<td>Version</td>
<td>FireEye (6), Recorded Future (5), ZeroFox (2), Qualy (3), F5 Labs (3)</td>
<td></td>
</tr>
</tbody>
</table>
| Fingerprint         | FireEye (13), Eclec-
ticIQ (1), ThreatConnect (1), Qualy (3) |
| Hash                | FireEye (8), Eclec-
ticIQ (1), ThreatConnect (1), Qualy (3) |
| Programming language| FireEye (3), ZeroFox (4), Cyjax (1)          |

Social media platforms are risky environments that encourage users to share content, sometimes leading to the unintentional dissemination of harmful content. Therefore, the threat intelligence we present is still non-exhaustive in its current form. Moreover, as mentioned earlier, to better support our measurement analysis, appraisal review identification was tuned toward high precision, the method for appraisal review identification could yield some potential false negatives. For example, a review may not consist of any keywords listed in the Table IV but simply mentioned testing the product for free: “...I have tested the hosting with the free version and the web server seems fast. There is a decent amount of features...”. Furthermore, we only analyzed the untrustworthy appraisers but did not examine the effectiveness of the appraisal system in enhancing trust, mitigating fraud, or influencing user behavior. We will leave the study on a more efficient appraisal review detector as our future work.

Ethics of data collection. This study is guided by our institute’s IRB. We effectively took action to address a variety of ethical concerns that arise from gathering and analyzing data from underground forums. More specifically, to restrict the burden we added to the network and marketplace servers in our data collection, and we set parameters such as sleeping time to limit the speed of crawling. We also avoided censorship of site administrators by registering as users and providing as input to the scraper a session cookie that we obtained by manually logging into the marketplace, plus the usage of proxies across the world. While we acknowledge the ethical implications of using cookies, which enables us to bypass the CAPTCHA and may not follow the policy set by site administrators, the benefit of ensuring data integrity and completeness weighed in favor of this design choice. Our data scrapes did not require us to establish reputable accounts or interact with harmful content, since all the content within these markets is publicly accessible to all registered users. The registration is open to all market visitors. During this process, we only need to create a username and password, along with providing a usable email address to receive a verification code. There is no need to provide any Personally Identifiable Information (PII). Aligned with the choice of previous cybercrime studies [101], [124], we disclosed the links to underground forums and we believe the research benefits (i.e., reproducibility and transparency) weighed in favor of providing links. In addition, using publicly available or leaked datasets has been acceptable in previous studies on the underground ecosystems [28], [81], [102], [122]. The CrimeBB dataset [89] we used obtained approval from

VIII. DISCUSSION

Limitations. Although our analysis of the appraisal systems in 8 underground marketplaces provides insights into understanding product merits and cyber threat intelligence, still, we only take into consideration the marketplaces whose communication traces are mostly written in English and leave out the non-English marketplaces which also contribute to a large number of cybercriminals. In addition, we only analyzed CTIs from three major categories related to malware and malicious services. Therefore, the threat intelligence we present is still non-exhaustive in its current form. Moreover, as mentioned earlier, to better support our measurement analysis, appraisal review identification was tuned toward high precision, the method for appraisal review identification could yield some potential false negatives. For example, a review may not consist of any keywords listed in the Table IV but simply mentioned testing the product for free: “...I have tested the hosting with the free version and the web server seems fast. There is a decent amount of features...”. Furthermore, we only analyzed the untrustworthy appraisers but did not examine the effectiveness of the appraisal system in enhancing trust, mitigating fraud, or influencing user behavior. We will leave the study on a more efficient appraisal review detector as our future work.
their Research Ethics Board (REB) and we followed the data sharing agreements from the Cambridge Cybercrime Centre. Similar to previous work [88], [102], our study also used the Nulled database [87] which consists of private messages among Nulled users. Note that our analysis of private messages was focused on identifying the interactions between appraisers and vendors, and not analyzing user identities or message content. Similar to previous work [88], [102], we did not find PII in the dataset during our manual analysis. We did not identify particular members from the marketplace, nor did we publish their usernames. We also replaced their sensitive information (i.e., email, Skype, and phone number) with anonymous expressions for further protection of private information. Our research work did not focus on studying any particular individual as well.

IX. RELATED WORK

Study on feedback/review system in underground marketplaces. Christin [46] performed a measurement analysis on feedback ratings in Silk Road to understand sellers’ reliability. Li et al. [70] chose a card hacking market and used a recursive neural tensor network to classify customer’s review texts to a five-point Likert scale which can reflect the sellers’ product quality. Li et al. [71] used a binary classifier on review posts related to opioid transactions to understand the customers’ satisfaction. Vu et al. [113] analyzed the trading activity of vouch copies, which was adopted as one of the contract types in HackForums. Other works studied the effect of reviews on vendors and underground markets, for instance, the impact on sellers’ reputations, sales and prices of goods [59], [61].

Different from previous work, our study focuses on a new feedback system, i.e., appraisal system. The findings of this study uncovered the ecosystem of the appraisal system and the characteristics of appraisers and appraisal reviews.

CTI gathering. Liao et al. [73] proposed iACE to automatically extract IOCs from technical articles by using graph mining techniques. Zhu et al. [126] leveraged text mining tools to extract malware behaviors from scientific papers. Catakoglu et al. [43] gleaned Web Indicators of Compromise (WIOCs) from compromised or malicious web pages and web applications by making use of the attackers’ JavaScript files. Khandpur et al. [67] detected cybersecurity events (i.e., data leak, DDoS attack, etc) from online social media (i.e., Twitter). Other works leveraged machine learning techniques to find exploited vulnerabilities from hacking forums and marketplaces on the Dark Web [31], [34], [35], [103], [117]. Some underground market-related properties are also retrieved by previous work, for instance, price [91], key actors [78], [93], [94], [123] and product types [51], [57]. Other works studied malware-related intelligence such as malware download channel [92], geolocation of malware campaigns [39], malware sample feed [106] and the value chain of Ransomware-as-a-Service economy [79].

Different from previous work, our study is not focusing on proposing a new CTI extraction method. Instead, we shed light on a new CTI source and highlight the high-quality threat information it can provide.

X. CONCLUSION

In this paper, we present the first measurement on a previously-unexplored feedback system, i.e., appraisal system, in underground marketplaces. Specifically, we conducted a large-scale analysis on 18,701 appraisers and 56,229 appraisal reviews from 8 marketplaces spanning 15 years to demystify the ecosystem behind the appraisal system, as well as the characteristics of appraisers (e.g., profile, credibility, merits for appraiser selection, etc.) and appraisal reviews (e.g., assessment merits, quality comparing to non-appraisal reviews, etc.). Moving forward, we further investigate appraisal reviews as a new source of cyber threat intelligence, which supplements current studies on threat intelligence gathering.

ACKNOWLEDGMENT

We thank the shepherd and anonymous reviewers for their insightful comments. This work is supported in part by the NSF CNS-1850725 and Indiana University Institute for Advanced Study (IAS). Zhengyi Li was supported by the graduate teaching assistantship at Indiana University (IU)’s department of Intelligent System Engineering. Xiaojing Liao was also partially supported by the Grant Thornton Institute and Indiana University Institute for Advanced Study (IAS). We also thank Xiangyu Du for his efforts in data annotation.

REFERENCES


[58] Tommi Gröndahl, Luca Pajola, Mika Juuti, Mauro Conti, and N. Asokan. All you need is "love": Evading hate speech detection. ASlEc ’18, page 2–12, New York, NY, USA, 2018. Association for Computing Machinery.


[88] Rebekah Overdorf, Carmela Troncoso, Rachel Greenstadt, and Damon McCoy. Under the underground: Predicting private interactions in underground forums. 05 2018.


[115] Xinyu Wang, Yong Jiang, Nguyen Bach, Tao Wang, Zhongqiang Huang, Fei Huang, and Kewe Tu. Improving named entity recognition by external context retrieving and cooperative learning, 2022.


APPENDIX A

We show the full regex list in Table XIV.

TABLE XIV: Full list of regexes used to extract CTIs

<table>
<thead>
<tr>
<th>Category</th>
<th>CTI</th>
<th>Regexp</th>
<th># review</th>
</tr>
</thead>
<tbody>
<tr>
<td>KW search volume</td>
<td>(?&lt;keyword&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*search)</td>
<td>236</td>
<td></td>
</tr>
<tr>
<td>KW competition</td>
<td>(?&lt;keyword&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*competition)</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>KW CPC</td>
<td>(?&lt;keyword&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*CPC)</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td># directory link</td>
<td>(?&lt;keyword&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*directory)</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td># profile link</td>
<td>(?&lt;keyword&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*profile)</td>
<td>577</td>
<td></td>
</tr>
<tr>
<td># blog link</td>
<td>(?&lt;keyword&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*blog)</td>
<td>1,686</td>
<td></td>
</tr>
<tr>
<td># bookmark link</td>
<td>(?&lt;keyword&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*bookmarks)</td>
<td>349</td>
<td></td>
</tr>
<tr>
<td># web2.0 link</td>
<td>(?&lt;keyword&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*web)</td>
<td>467</td>
<td></td>
</tr>
<tr>
<td>Domain authority</td>
<td>(?&lt;keyword&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*domain)</td>
<td>514</td>
<td></td>
</tr>
<tr>
<td>Domain name</td>
<td>(?&lt;domain&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*name)</td>
<td>593</td>
<td></td>
</tr>
<tr>
<td>Domain age</td>
<td>(?&lt;domain&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*age)</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>(?&lt;location&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*location)</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>Proxy type</td>
<td>(?&lt;proxy&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*proxy)</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td># of accounts</td>
<td>(?&lt;account&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*account)</td>
<td>549</td>
<td></td>
</tr>
<tr>
<td>Account age</td>
<td>(?&lt;account&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*age)</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Verified method</td>
<td>(?&lt;verified&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*verified)</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>Whether has profile</td>
<td>(?&lt;has_profile&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*profile)</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td>Social promotion</td>
<td>(?&lt;promotion&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*promotion)</td>
<td>1,574</td>
<td></td>
</tr>
<tr>
<td>PUD type</td>
<td>(?&lt;type&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*type)</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>PE File</td>
<td>(?&lt;file&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*file)</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>(?&lt;os&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*os)</td>
<td>312</td>
<td></td>
</tr>
<tr>
<td>Version</td>
<td>(?&lt;version&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*version)</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Hash</td>
<td>(?&lt;hash&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*hash)</td>
<td>487</td>
<td></td>
</tr>
<tr>
<td>Programming language</td>
<td>(?&lt;language&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*language)</td>
<td>198</td>
<td></td>
</tr>
<tr>
<td>Whether metadependencies</td>
<td>(?&lt;dependencies&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*dependencies)</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Anti-virus detection result</td>
<td>(?&lt;result&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*result)</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Whether auto-update</td>
<td>(?&lt;auto-update&gt;[a-zA-Z0-9]*)(?=[a-zA-Z0-9\s]*auto-update)</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>