Revealing, Characterizing, and Detecting Crowdsourcing Spammers: A Case Study in Community Q&A

Aifang Xu, Xiaonan Feng, Ye Tian*
Anhui Key Lab on High Performance Computing
School of Computer Science and Technology, University of Science and Technology of China
Email: {xaf, xnfeng}@mail.ustc.edu.cn, yetian@ustc.edu.cn

***Abstract*—Crowdsourcing services have emerged and become popular on the Internet in recent years. However, evidence shows that crowdsourcing can be maliciously manipulated. In this paper, we focus on the "dark side" of the crowdsourcing services. More specifically, we investigate the spam campaigns that are originated and orchestrated on a large Chinese-based crowdsourcing website, namely `ZhuBaJie.com`, and track the crowd workers to their spamming behaviors on `Baidu Zhidao`, the largest community-based question answering (QA) site in China. By linking the spam campaigns, workers, spammer accounts, and spamming behaviors together, we are able to reveal the entire ecosystem that underlies the crowdsourcing spam attacks. We present a comprehensive and insightful analysis of the ecosystem from multiple perspectives, including the scale and scope of the spam attacks, Sybil accounts and colluding strategy employed by the spammers, workers' efforts and monetary rewards, and quality control performed by the spam campaigners, etc. We also analyze the behavioral discrepancies between the spammer accounts and the legitimate users in community QA, and present methodologies for detecting the spammers based on our understandings on the crowdsourcing spam ecosystem.**

I. Introduction

Online crowdsourcing marketplaces, or Internet crowdsourcing systems, have emerged and become popular on the Internet in recent years. In such a system, a user can assemble massive manpower to accomplish a task by purchasing services from a large group of people. Many different jobs can be crowdsourced on the Internet, examples include not only the creative jobs such as designing a website or writing a blog article, but also the repetitive tedious jobs like copying Google search results or typing texts from images. In 2011, it is estimated that there were six million crowd workers around the world, who had made a total revenue of 375 million US dollars [1].

Although crowdsourcing facilitates large-scaled social collaborations, however, it can be maliciously manipulated, where a campaigner can pay a large number of workers to post biased comments, spam, or even virus URLs on the Internet. For example, it is reported that during the 360 vs. Tencent conflict between the two major Chinese IT companies in 2010, both sides were suspected of paying for postings [2]; in 2011, the online shopping website Taobao shut down over 200 "Internet water army" companies selling microblog followers [3]. A recent study [4] shows that malicious crowdsourcing campaigns have already become a concrete threat to the Internet, and exhibit a continuous growth all over the world.

In this paper, we study the spamming behaviors that are originated, orchestrated, and benefited from the Internet crowdsourcing systems. Unlike most of the previous works that focus only on the spam victims (e.g., YouTube [5], Twitter [6], Facebook [7], Renren [8], Foursquare [9], etc.), in this work we present a panoramic and insightful study on the entire ecosystem that underlies the spam attacks. More specifically, we analyze the malicious spam campaigns on a large Chinese crowdsourcing website, namely `ZhuBaJie.com` [10] (referred to as *ZBJ* for short), and track the spam workers[1] to their accounts and the posted spam on the target site of `Baidu Zhidao` [11] (referred to as *Zhidao* for short), which is the largest community-based question answering (QA) site in China. By linking the campaigns, workers, accounts, and spam contents together, we are able to uncover the entire crowdsourcing spam ecosystem, and study it from multiple perspectives. Our analysis is from not only the victim's, but also the campaigner's and the spammer's points of views. We also discuss the methodologies for detecting the crowdsourcing spammers based on our understandings on the ecosystem.

There are relatively a limited number of works focusing on the crowdsourced Internet mis-behaviors in the literature. Wang et al. [4][12] investigate two Chinese-based crowdsourcing websites and their spam campaigns targeted on microblogs, and point out that the threats from the malicious campaigns are non-trivial; Lee et al. [13] study the methodologies for detecting the worker accounts on Twitter by analyzing the spam campaigns from three English-based crowdsourcing websites. Unlike these studies that focus on microblogs, in this work we investigate the spamming behaviors targeted on the community-based QA systems. Our motivation is that unlike other spam targets such as forums, microblogs, and instant message groups, people usually consult a QA system for solutions of their daily-life problems [14], thus the spam in the form of answer or even "best answer" is generally more deceptive. Moreover, it is more demanding for the spammers to spam in community QA, for example, unlike simply tweeting and retweeting when spamming a microblog site, in our study we find that Sybil accounts and colluding are heavily employed by the spammers for bypassing Zhidao's anti-spam surveillance, thus enable us to obtain more insightful understandings on the crowdsourcing spam ecosystem.

We elaborate our contributions in this paper in four aspects.

*Correponding author

This work was supported by the National Natural Science Foundation of China (Nos. 61202405 and 61103228) and the sub task of the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA06010301).

[1] In this paper, we use the words "spam worker" and "spammer" interchangeably.

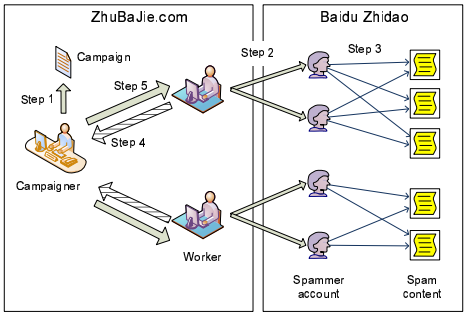


Fig. 1. Demonstration of the crowdsourcing spamming behaviors originated from ZhuBaJie.com on Baidu Zhidao.

- First, we study the organizational structures of the ZBJ and Zhidao websites, and employ crawlers to extract data from them in a sufficiently long time. The obtained datasets will serve as the basis of our analytical study.
- Second, we examine the requirements of the spam tasks targeted on Zhidao, and categorize them into four types; for each type, we develop method to track the spam workers to their accounts and the posted contents, thus link the two datasets which are independently collected from different sources together.
- Third, with the assistance of the measurement data, we carry out a detailed analysis on the crowdsourcing spam ecosystem from multiple perspectives; the topics under study include: scale and scope of the spam attacks, Sybil accounts and colluding employed by the spammers, workers' efforts and monetary rewards, and quality control performed by the campaigners, etc.
- Finally, we investigate the rationales and methodologies for detecting the Zhidao accounts that are employed by the spammers. By making use of the discrepancies between the spammer accounts and the legitimate users, we show that it is possible to detect the spammer accounts with high accuracies.

The remainder part of this paper is organized as follows: In Section II, we introduce the crowdsourcing website of ZhuBaJie.com and the community-based QA site of Baidu Zhidao that are under study in this paper; in Section III, we describe our methodology for collecting data from the two websites and present the resulting datasets; we carry out a comprehensive and insightful analysis on the crowdsourcing spam ecosystem in Section IV; in Section V, we address the problem of detecting the spammer accounts and present our methodologies and the detecting results; we discuss the related works in Section VI and conclude this paper in Section VII.

II. Background

A. ZhuBaJie.com

In this paper, we select ZhuBaJie.com (or ZBJ for short), one of the most popular crowdsourcing websites in China, and study the spamming behaviors originated from it. As other crowdsourcing systems, many different jobs, such as translating an article and designing a logo can be crowdsourced on ZBJ. Moreover, on ZBJ there is a subsection called "Internet marketing", where a user, who wishes to promote his business through Internet spamming, can post his needs as the crowdsourcing campaigns on the website.

TABLE I. Rules of experience and wealth points on Zhidao

Activity	Experience pt.	Wealth pt.
First log-in each day	+2	
Ask a question		−5
Select a "best answer"		+5
Answer a question	+2	
Be selected as "best answer"	+20	+20
Be selected as "excellent answer"		+10
Question removed by administrator	−20	−20
Answer removed by administrator	−10	−10

We use Fig. 1 to demonstrate the major steps of launching a spam campaign on ZBJ. As shown in the figure, when initializing a campaign, the *campaigner* first needs to specify how many tasks are needed in the campaign and how much he will pay to the spammers. Then he detailedly specifies the task requirement such as the submission deadline, spam keywords, content and target of the spam, etc. (step 1). Usually the spam targets are the popular websites like forums, microblogs, and question answering (QA) sites that allow user-generated contents. When a spammer employs one or multiple *accounts* on the target site (step 2) to post the spam as required (step 3), he presents a *submission* to the campaigner (step 4); the submission could be the URL or a snapshot of the web page on which the spam is posted. After receiving the submissions, the campaigner evaluates them, selects the ones that are of high quality as the *qualified* submissions, and rewards the corresponding spammers with monetary payments (step 5).

B. Baidu Zhidao

Although the spam campaigns on ZBJ can be targeted on many different Internet services, in this paper, we focus on the campaigns that are targeted on the community-based question answering (QA) systems. More specifically, we select Baidu Zhidao, the largest community QA site in China, and study the crowdsourcing spamming behaviors on it.

Similar to other QA services like Yahoo Answers and Stack Overflow, Zhidao is based on user-generated QA contents: a Zhidao user can post a question for other users to answer; a questioner can select one answer as the "best answer" for his question; and a user can agree on another user's answer by clicking the thumb-up sign next to it. Zhidao is a typical community-based QA system, where each question falls into one of the 14 categories, and each category is further divided into many sub-categories. In each sub-category, Zhidao designates a few expert users as voluntary administrators, who can label the answers with high quality as the "excellent answers".

To encourage participating, Zhidao provides two virtual incentives, namely the *experience* points and the *wealth* points. The experience points indicate how active a user behaves on Zhidao, and a user can obtain the points from his activities such as logging-in, answering questions, etc. Unlike experience, the wealth points reflect how contributive a user is, and a user can gain his wealth points by answering questions and especially contributing the best and excellent answers. Table I summaries the rules of the two incentives on Zhidao.

Zhidao hires an anti-spam surveillance team that persistently inspects all the questions and answers, identifies and removes spam and any other inappropriate contents from the website. In addition, the voluntary administrators also have the privilege to remove or hide the spam contents.

III. Data Collection

A. Collecting the Zhidao Dataset

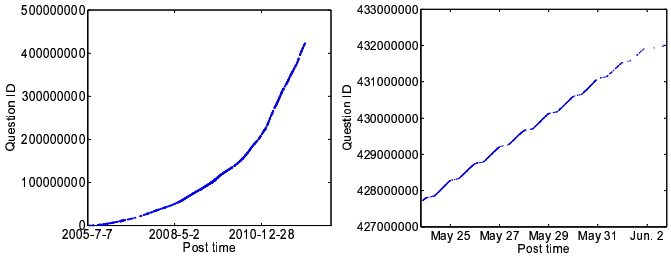
Our study is based on the data that we have extracted from ZBJ and Zhidao. We first describe our methodology for collecting the Zhidao dataset. On Zhidao there are two kinds of pages: *question page* and *profile page*. A question page contains a question and all its answers, and for the questioner and the answerers of this question, the page also contains the links to their profile pages. An account's profile page contains metadata of the account such as the account's experience points, wealth points, number of questions and answers, number of excellent answers, ratio of answers being selected as the "best answer", etc. In addition, the profile page lists all the questions the account has participated in, either as the questioner or as an answerer.

We carry out a random walk to crawl both the question and the profile pages on Zhidao. Twenty parallel crawling threads are employed in our measurement, where each thread starts at a random question page as seed. For each question page encountered, in addition to collecting the QA contents, the crawling thread selects one random account participating in this question, and collects all its metadata from the account's profile page; after that, the thread randomly selects one unencountered question that the account has engaged in, and crawls the question page, $\cdots$. The "question – profile – question" iteration repeats until the crawler can no longer proceed, then we select a new random question as seed and restart the thread.

We recognize that random walk would be biased towards the active users on Zhidao, however, study shows that when the size of the sampled dataset is large enough, as the one we have collected from Zhidao, the bias becomes negligible [15].

We have collected about 5 million question pages and 7 million profile pages during a crawling period between May 3 and Jun. 2, 2012. Table II presents a summary of the dataset. We note that each question on Zhidao is assigned with a numerical ID, whose value increases over time. Fig. 2(a) shows how the question ID increases, and we can see that the number of the questions on Zhidao increased very fast after the website was initially setup in 2005, and the increasing rate becomes stable after 2011. Moreover, if we use the ID of the largest numerical value in the dataset (which is 432014465) to estimate the total number of the questions on Zhidao, then our dataset covers approximately 1.16% of the entire question set.

We also plot all the questions that were posted between May 23 and Jun. 2, 2012 in our dataset in Fig. 2(b). From the figure one can see that the question ID increases at a stable rate, only slightly slows down during the mid-nights. From the figure we can estimate that about 420K new questions were posted daily on Zhidao since 2011.



(a) (b)

Fig. 2. (a) Question ID increasing over time and (b) linear increasing of the question ID between May 23 and Jun. 2, 2012 on Zhidao.

B. Collecting the ZBJ Dataset

We develop crawlers and collect data of the spam campaigns that were posted on ZBJ between Jan. 6, 2011 and Mar. 19, 2012. A summary of the ZBJ dataset is given in Table II. We have collected a total number of $2,293$ campaigns targeted on Zhidao, and among them, $1,730$ have received at least one qualified submission. We have found $5,943$ distinct spam workers participating in at least one campaign, and $1,958$ of them have presented at least one qualified submission. A total number of $106,604$ submissions have been collected, among them, $12,069$ were labeled as qualified by the spam campaigners. We have also extracted $12,470$ question page URLs from the submissions, and $9,500$ of them contain qualified spam contents. Note that a question page URL can appear in multiple task submissions. By applying the method that will be described in the next subsection, we have successfully identified $9,218$ accounts on Zhidao that are employed by the workers for spamming.

C. Tracking Spammers to their Accounts

The two datasets of ZBJ and Zhidao were collected independently. To enable a behavioral analysis, we need to link the two datasets by tracking the ZBJ spam workers to their accounts on Zhidao, for this purpose, we carefully examine the task requirements of the spam campaigns on ZBJ, and find that there are four task types, which we describe in the following:

- **Self-answering**: in a "self-answering" task, a spammer needs to be able to access at least two accounts on Zhidao, supposedly account A and account B. The spammer first posts a question using account A, then he uses account B to answer the question, and selects the answer as the "best answer" using account A again. By self-answering, the spammer can post the spam content as the "best answer" on Zhidao.
- **Free-riding**: in such a task, a spammer is required to find an existing question on Zhidao that is revelent to the spam content and is available to answer, then answers the question with the spam content.
- **Given-question**: in such a task, the campaigner specifies the questions on Zhidao for the spammers to answer using the spam content.
- **Question-only**: in a "question-only" task, the campaigner requires the spammers to post questions con-

TABLE II. SUMMARIES OF THE ZHIDAO AND ZBJ DATASETS

Zhidao		ZBJ	
Questions	4,992,511	Campaigns (with qualified submissions)	1,730
Answers	24,600,297	Spammers (with at least one qualified submission / all)	1,958 / 5,943
Best answers	3,412,714	Submissions (qualified / all)	12,069 / 106,604
Accounts	6,979,013	Submitted question page URLs (qualified / all)	9,500 / 12,470
		Spammer accounts on Zhidao	9,218

TABLE III. SUMMARIES OF THE DIFFERENT SPAM CAMPAIGN CATEGORIES

	Self-answering	Free-riding	Given-question	Question-only	Total
Campaigns	1,614	89	16	11	1,730
Qualified question pages	8,722	619	53	106	9,500
Spammers	1,788	191	51	63	1,958
Spammer accounts on Zhidao	8,906	200	98	87	9,218

taining the spam contents without answering them.

Based on the classification, we manually categorize the $1,730$ campaigns that have at least one qualified submission into the four types, and apply the following method to identify the Zhidao accounts that were employed by the spam workers from their submissions.

- From each qualified submission of a "self-answering" task, we visit the question page in the submission, label the questioner account and the account posting the "best answer" as the spammer accounts.
- From each qualified submission of a "question-only" task, we only label the questioner on the submitted question page as the spammer account.
- For each task falling into the "free-riding" and the "given-question" categories, we manually identify the spammer account from the question page in each qualified submission.

Using the above method, we have successfully identified $9,218$ spammer accounts on Zhidao, and associated them with the $1,958$ spammers who have presented at least one qualified submission on ZBJ. We consider a spammer as a real-world person or organization, as on ZBJ, each user is registered with a unique bank account for receiving the task reward.

In Table III, we list the spam campaigns with tasks of each type, and we also list the participating spammers on ZBJ, the employed spammer accounts on Zhidao, and the qualified spam question pages for each type in the table. One can see that the campaigners prefer the "self-answering" tasks most, as they constitute 93.3% of the campaigns and 91.8% of the submissions. One possible reason is that by posting spam as the "best answer", the campaigners believe that it can be more deceptive and draw more attentions from the Internet users than the non-best answers.

IV. CHARACTERIZING THE CROWDSOURCING SPAM ECOSYSTEM

Assisted with the datasets collected from ZBJ and Zhidao, in this section, we present an insightful analysis on the *crowdsourcing spam ecosystem*. Unlike the previous studies on the Internet spamming that focus mostly on victims, our analysis covers the entire ecosystem from the spam campaigners and workers to the spammer accounts and spam contents in the forms of questions and answers.

A. Scale and Scope of the Spam Attacks

We first look at the scale and scope of the spam attacks originated from ZBJ. Fig. 3 presents the spam question pages posted by all the spammer accounts on Zhidao in each day in our measurement. From the figure we can see that the spamming behaviors are non-trivial: it is observed that as many as 118 spam pages were posted in one day, constituting nearly 0.03% of the new questions daily posted on Zhidao. Note that the spam contents are originated from just one crowdsourcing site of ZBJ. If we consider the other crowdsourcing websites with spam campaigns targeted on Zhidao, such as `sandaha.com` and `vikecn.com`, more spam pages should be observed.

We also examine the categories that the spam pages fall in, and present the result in Fig. 4. One can see that there are considerable spam contents in all the 14 categories on Zhidao, and 43.7% of them are posted in the categories of "life" and "local", in which many people seek advices for their daily-life problems. By posting in these categories, the spam contents are likely to be more deceptive than in the other forms such as email or microblog.

B. Spammer Accounts

We then look at the spammer accounts that we have identified in Section III-C. Fig. 5 presents the distribution of the spam pages posted by each account for all the spammer accounts in the ZBJ dataset. From the figure one can see that an individual account is not productive, since as few as 1.45 pages are posted by an account on average, and over 70% of the accounts have posted only one spam page.

On the other hand, we find that the accounts are dedicating, that is, a spammer account always posts spam on Zhidao. To show this, we randomly select 100 spammer accounts and manually examine all the questions and answers they have posted. We find that 93.8% of their posts are obvious spam, suggesting that once a spammer account is identified, we can safely remove all its posts without worrying removing the legitimate contents.

C. Sybil accounts and Colluding

As we have seen in Section III-C, to accomplish a "self-answering" task, a spammer needs to access at least two Zhidao accounts. In this paper, we refer to the accounts that are controlled by a same spammer as the spammer's *Sybil accounts*, and we consider any question-answer interactions between two Sybil accounts as *colluding*.

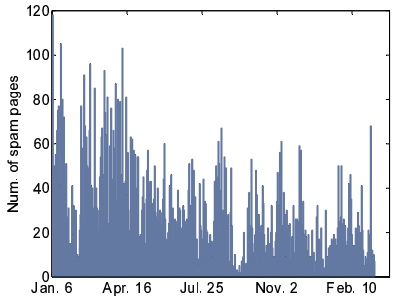


Fig. 3. Spam pages originated from ZBJ and posted on Zhidao in each day in our measurement.

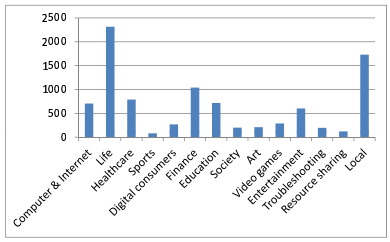


Fig. 4. Num. of the spam pages in each of the 14 categories on Zhidao.

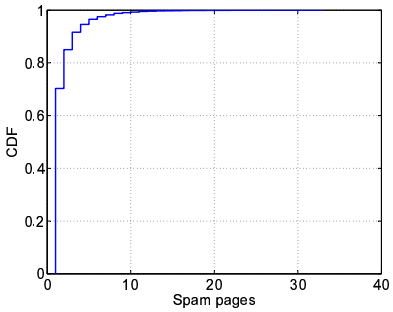


Fig. 5. Distribution of the spam pages posted by the spammer accounts on Zhidao.

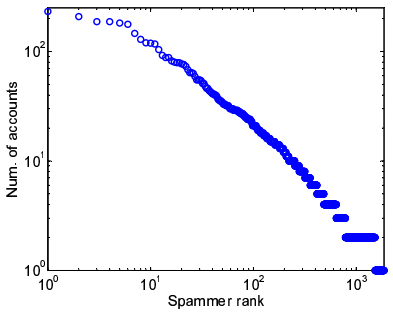


Fig. 6. Num. of the Sybil accounts employed by a spammer for all the spammers on ZBJ.

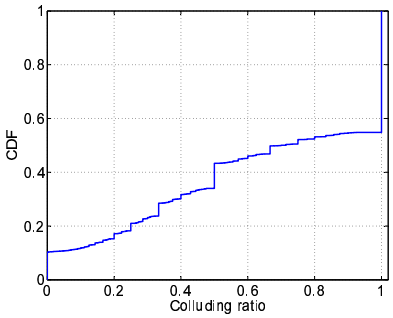


Fig. 7. Distribution of the colluding ratios for the spammer accounts.

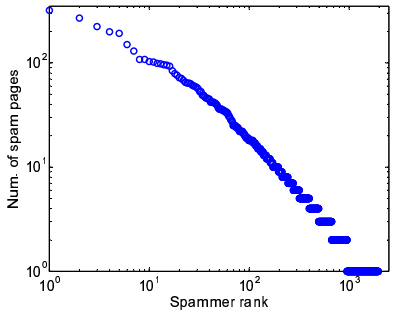


Fig. 8. Num. of the spam pages posted by a spammer for all the spammers on ZBJ.

Fig. 6 presents the Sybil accounts that one spammer has employed for all the spammers in our dataset. From the figure we can see that the distribution is Zipf-like, which is widely observed in many wealth distributions. We find that 81.8% of the spammers engage more than one account, and the average number of the accounts used by a spammer is 6.4; we find that a few spammers possess a large number of accounts, for example, the "richest" spammer employs as many as 232 Sybil accounts in our dataset. We also observe some collaborations among the spammers by sharing their accounts. For example, the most shared account in our dataset has been used by 14 different spammers. But for 79% of the accounts, each has been used by only one spammer.

We then consider the colluding among the Sybil accounts. For each spammer account with at least one Sybil account under control by a same spammer, we compute its *colluding ratio* as the ratio of the colluding interactions in all the account's question-answer interactions. Fig. 7 plots the distribution of the colluding ratios for all the spammer accounts. From the figure we can see that for nearly half of the accounts, they only interact with their colluding Sybil accounts (i.e., with a colluding ratio of one). Clearly, our observations from Fig. 6-7 suggest that Sybil accounts and colluding are heavily employed by the spammers when accomplishing their spam tasks targeted on community QA, which is very different from their behaviors when spamming on the microblog sites [4].

D. Efforts and Rewards of the Spammers

In this subsection, we measure the efforts a spammer has made in posting the spam on Zhidao, and how he is economically benefitted from his work. Fig. 8 presents the spam pages posted by a spammer for all the spammers in the ZBJ dataset, and in Fig. 9, we present the relationship between the Sybil accounts one spammer has employed and the spam pages he has posted on Zhidao. We can see that there is a strong correlation between the two, with a Person's correlation coefficient value as high as 0.9394. We also note that an individual spammer could be very productive by making use of many Sybil accounts on Zhidao, for example, the spammer who employs most accounts has posted as many as 319 qualified spam pages, making him also the most productive spammer in our dataset.

We also investigate the monetary payments a spammer has received. Fig. 10 presents a spammer's income in US dollars[2] for all the spammers in our dataset, and we present the relationship between a spammer's Sybil accounts and his income in Fig. 11. Again a strong correlation can be observed, with a Person's correlation coefficient value of 0.9086. From the figures, we can see that although most spammers only make small amount of money, however, there are a few spammers whose income is close to one hundred dollars, which conforms to the observation in [4] that most spammers are casual but there are a few professional ones. In addition, by using the linear regression technique, it is estimated that on average, a spammer can make a revenue of $0.23 from one Zhidao account per year. From Fig. 8-11, we can draw the conclusion that *the Zhidao accounts are the most important asset for a spammer: the more Sybil accounts a spammer controls, the more capable he is in spamming, and the more money he can make from the spam campaigns*.

[2]Since the payments are made in RMB, in this paper, we use a fixed exchange rate of 1 USD = 6.254 RMB.

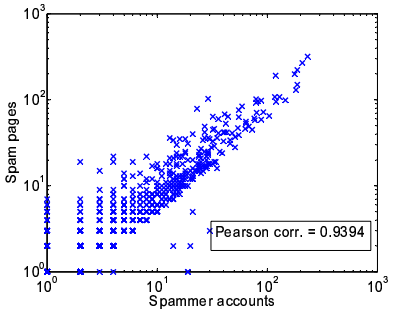


Fig. 9. Correlation between the accounts employed by a spammer and the spam pages he has posted.

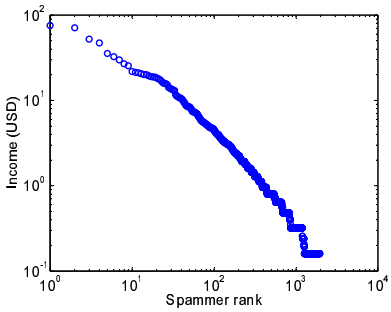


Fig. 10. Income of a spammer for all the spammers on ZBJ.

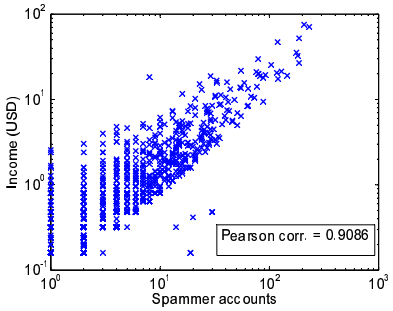


Fig. 11. Correlation between the accounts employed by a spammer and his monetary income.

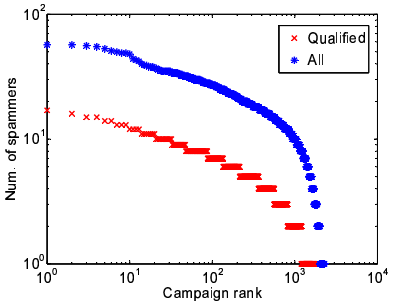


Fig. 12. All spammers and the spammers presenting at least one qualified submission in the campaigns.

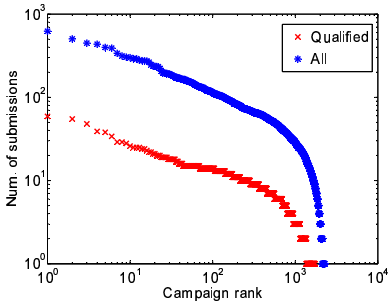


Fig. 13. All and qualified submissions in the campaigns.

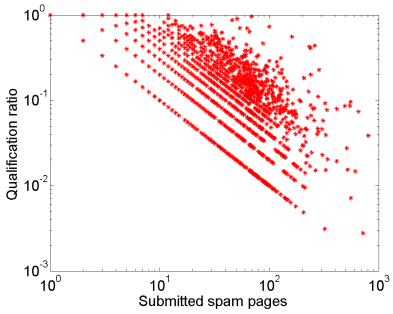


Fig. 14. Correlation between the campaign size and the campaign's qualification ratio.

E. Spam Campaigns and Quality Control

Finally, we study the crowdsourcing spamming ecosystem from a campaigner's perspective. Our question is: is the crowdsourcing system like ZBJ an effective platform for the campaigners to spread spam contents in community QA?

To answer this question, in Fig. 12 we present the participating spammers, as well as the spammers submitting at least one qualified page, in each campaign in the ZBJ dataset; and in Fig. 13, we plot all and the qualified submissions in each campaign. From the two figures, we can see that it is quite effective for a campaigner to spread spam by posting crowdsourcing campaigns on ZBJ, as on average, a campaign attracts 10.2 spammers, and incurs as many as 36.3 submissions from them.

From Fig. 12-13, we note that the spam campaigners generally hold a selective altitude towards the submissions, as on average, only 11.4% of the submissions were considered as qualified by the campaigners, and only 22.6% of the spammers actually received the monetary payments (by submitting at least one qualified page). Obviously, such a selective attitude makes the spammers to improve the quality of their posted spam, that is, they must make the spam contents more difficult to be distinguished from the legitimate ones, so as to pass the campaigner's qualification review and get the task reward.

To testify our point, we carry out an active experiment on Zhidao in Feb. 2012. In the experiment we used 10 different accounts to post 20 benign spam questions and answers composed of random meaningless sentences. We find that all our spam got removed by Zhidao's administration within one day. On the contrary, in Aug. 2013, 14 months after the data collection, we revisited all the qualified spam pages, and find that only 2.5% of them were removed.

We then calculate the *qualification ratio* for each spam campaign, and correlate it with the size of the campaign in terms of the submissions in Fig. 14. From the figure we can see that in general, the qualification ratio decreases as the campaign becomes larger. One possible interpretation is that as the campaigner receives many submissions, he tends to be more selective by qualifying only the submissions of the highest quality. For example in our dataset, the most selective campaigner, who qualifies only two out of 719 submissions, is also the owner of the second largest campaign. From the above analysis, we conclude that *it is very effective for a campaigner to spread spam on Zhidao through the ZBJ crowdsourcing campaigns, regarding both the quality and the quantity of the spam contents*.

V. Detecting the Spammer Accounts

The previous analysis suggests that the spamming behaviors originated from the crowdsourcing systems impose a considerable threat to the question answering services on the Internet. Although in Section III-C, we have successfully identified the spammer accounts on Zhidao, based on the information of the task submissions collected from the ZBJ crowdsourcing website, however, we can't assume that such information will be available all the time. In fact, the "Internet water army" businesses of the crowdsourcing websites have already received attentions from the law-enforcement agencies [16][17], and are likely to be shutdown in future; on the other hand, distributed crowdsourcing systems, which make use of the private communication channels such as instant

message groups and chat rooms to distribute spam campaigns and deliver task submissions, appear and thrive in recent years [4]. Obviously, on such a system, it will be difficult to harvest the task submissions in a large scale.

Motivated by the observation, in this section, we investigate the rationales and methodologies for detecting the accounts that are employed by the crowdsourcing spammers in community QA. By analyzing their behaviors using the dataset collected from Baidu Zhidao, we find that comparing with the legitimate users, the spammer accounts exhibit different behavioral patterns. Furthermore, by making use of the discrepancies, we show that the spammer accounts can be detected from the legitimate accounts with high accuracies.

A. Comparing User Behaviors

1) Preparation: For analyzing the behaviors of the accounts on Zhidao, we first construct a *social network* that is denoted as a directed weighed graph $\vec{G} = (V, E)$ from our Zhidao dataset. Each node on the graph represents a Zhidao account, and if one account $v_1 \in V$ has a question that is answered by another account $v_2 \in V$, we add a directed edge $\vec{e} = (v_1, v_2) \in E$ on the graph. We randomly select $117,546$ accounts from the Zhidao dataset, and construct the graph. The resulting social network contains $134,310$ edges.

Based on the social network, we build a test collection for studying the behavioral characteristics of the spammer and the legitimate accounts. We first include the $5,820$ spammer accounts on the network $\vec{G}$ in the collection, then we randomly select a number of the non-spammer accounts, and for each account, we manually check all its questions and answers in our dataset to decide if it is legitimate. With the help of a few volunteers, we have located $5,559$ legitimate accounts and include them in the test collection.

2) Attributes selection: Since the spammers and the legitimate users have different objectives on Zhidao, they should exhibit different behaviors. We examine dozens of the accounts' attributes to capture their behavioral characteristics. These attributes can be grouped into three sets, namely the *profile attributes*, the *question / answer (QA) attributes*, and the *social network (SN) attributes*, which we describe in the following.

Profile attributes. The profile attributes are the properties of an account that reflect how the account behaves on Zhidao, and can be extracted from the account's profile page. We first select 6 basic attributes, which are the account's experience points, wealth points, number of questions, number of answers, number of excellent answers, and the best answer ratio.

Besides the basic attributes, we also consider 5 additional attributes, which capture an account's efficiency in acquiring the experience and wealth points. The attributes are: mean experience points per answer, mean experience points per question plus answer, mean wealth points per answer, mean wealth points per question plus answer, and the ratio between the wealth and experience points.

QA attributes. The question / answer attributes of an account are the properties of the questions and answers that the account has posted on Zhidao. We consider 4 attributes, namely the mean question length, the mean answer length, the mean time difference between a question posted by the account and its answers, and the mean times answers being agreed. We summarize these attributes from the account's question pages.

TABLE IV. TOP TEN ATTRIBUTES

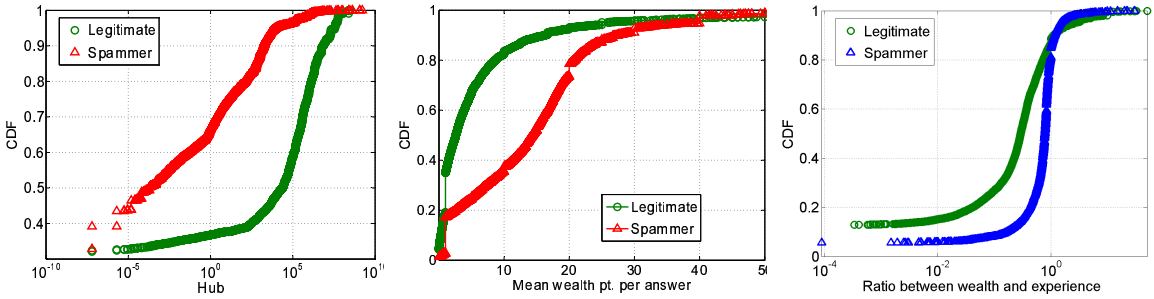
Type	Rank	Attributes
Profile attribute	2	Mean wealth pt. per answer
	3	Ratio between wealth pt. and experience pt.
	4	Mean wealth pt. per question plus answer
	8	Mean experience pt. per answer
	10	Wealth pt.
SN attribute	1	Hub [19]
	5	Authority [19]
	6	Inbound closeness centrality [18]
	7	Betweenness centrality [18]
	9	Outbound closeness centrality [18]

SN attributes. The social network attributes are the properties derived from the social network $\vec{G} = (V, E)$ that we have constructed. Specifically, we consider the following 5 attributes, namely the inbound and outbound closeness centralities [18], the betweenness centrality [18], the hub and authority [19] for each account corresponding to a node on the network $\vec{G}$.

3) Attributes comparison: We evaluate the 20 attributes by computing their information gains (i.e., the Kullback-Leibler divergence [20]), using the test collection containing $5,559$ legitimate accounts and $5,820$ spammer accounts. Table IV lists the top 10 most useful attributes. From the table, one can see that none of the QA attributes are included, suggesting that it is difficult to distinguish a spammer account by examining only its content statistics. Meanwhile, among the top ten most useful attributes, five are profile attributes and five are derived from the social network, indicating that the behavioral discrepancies of the two different accounts can be captured in either the accounts' profiles or from the social network. In the following we compare the three most useful attributes, namely the hub, the mean wealth points per answer, and the wealth-experience ratio for the spammer and the legitimate accounts, and investigate the rationales of the observed discrepancies.

We first compare the hub attribute in Fig. 15(a). Note that according to its definition [19], a node with a large hub value indicates that many other nodes can be reached by following the edges from this node, while a node with a small value suggests that it is poorly connected to the majority part of the network. From Fig. 15(a), one can see that most of the spammer accounts' hub values are very small, on the other hand, considerable legitimate accounts have moderate and large hub values. In other words, from a spammer account it is difficult to reach many other accounts, as such an account's questions are answered by the accounts with low authorities, that is, the accounts who only answer the questions from few other accounts. This observation conforms to our observation in Section IV, where a spammer account tends to answer the questions posted by its colluding accounts in the "self-answering" tasks. On the other hand, we find that some legitimate accounts have large hub values, suggesting that their questions have attracted some expert users, who tend to answer the questions asked by many different users on Zhidao.

In Fig. 15(b), we compare the attribute of the mean wealth points per answer for the two different accounts. From the figure, one can see that comparing with the legitimate



(a) Hub (b) Mean wealth pt. per answer (c) Ratio between wealth and experience

Fig. 15. Comparison of the three most useful attributes between the spammer and the legitimate accounts.

TABLE V. CLASSIFICATION RESULTS USING ALL ATTRIBUTES

Classifier		True positive	False positive
J48	Spammer account	90.4%	11.3%
	Legitimate account	88.7%	9.6%
PART	Spammer account	90.3%	10.6%
	Legitimate account	89.4%	9.7%

TABLE VI. CLASSIFICATION RESULTS USING ONLY SN ATTRIBUTES

Classifier		True positive	False positive
J48	Spammer account	88.9%	19.1%
	Legitimate account	80.9%	11.1%
PART	Spammer account	87.0%	18.4%
	Legitimate account	81.6%	13.0%

accounts, the spammer accounts acquire the wealth points more efficiently. This can be explained with the fact that in the majority "self-answering" tasks, an answer posted by a spammer account must be selected as the "best answer" by its colluding account that asks the question, and from Table I, we can see that 20 wealth points are granted by Zhidao to the "best answer" account. Meanwhile, an answer posted by a legitimate user is unlikely to be selected as the "best answer" each time, so the legitimate users have much lower rates in acquiring the wealth points as compared to the spammer accounts.

Fig. 15(c) compares the wealth-experience ratios between the spammer and the legitimate accounts. Similar to Fig. 15(b), we have observed faster rates for the spammer accounts to acquire the wealth points. As each time a spammer logs in on Zhidao, he has a deterministic objective to post spam questions or answers, and most of the spam answers are selected as the "best answers" by its colluding accounts; meanwhile, a legitimate user does not necessarily contribute a "best answer" each time he logs in and posts an answer.

B. Classification Methods and Results

The problem of detecting the spammer accounts from the legitimate accounts is a classification problem, for which we evaluate various supervised classification algorithms. We settle down with the C4.5 algorithm because of its accuracy and efficiency. In particular, we employ two implementations of the algorithm, namely the tree-based J48 [21] and rule-based PART [20]. We apply the 10-fold cross-validation to avoid over-fitting.

Table V presents our classification matrix using J48 and PART. From the table one can see that by using the attributes that capture the behavioral patterns of the different accounts, both classifiers can detect the spammer accounts from the legitimate ones accurately.

C. Detecting Spammer Accounts with SN Attributes

In the previous study, we employ the profile attributes as well as the SN attributes for detecting the spammer accounts. Note that an account's profile attributes can only be collected from its profile page, while the SN attributes are derived from the social network that is formed by the question-answer interactions among the users. Such interactions exist and can be publicly accessed on Zhidao and nearly all the other community-based QA systems.

We have two concerns regarding the profile attributes: First, different QA sites may have different incentive mechanisms, so the attributes such as the wealth-experience ratio that are useful on Zhidao may not be applicable on other QA systems. Moreover, although all the accounts' profile pages are publicly accessible on Zhidao, it is not the case on some other QA systems. For example, on Yahoo Answers, a user can choose to hide his profile from the public.

Motivated by the observation, we consider the case that only the information of the question-answer interactions among the accounts are available, and use only the SN attributes to detect the spammer accounts. We carry out a classification experiment with our test collection using the J48 and PART classifiers, and present the result in Table VI. Comparing with Table V, we find that the accuracies degrade about 2% for detecting the spammer accounts, and degrade about 8% for the legitimate accounts, but the classifiers still achieve an overall accuracy around 85%. Despite the accuracy degradation, we stress that the trained classifiers can be applied on detecting crowdsourcing spammers in other community-based QA systems, regardless of the system's incentive rules and users' privacy settings, as long as the social network structure that is shaped by users' question-answer interactions can be constructed.

VI. RELATED WORK

Spammer analysis and detection. Recently, researchers have studied spammers and their behaviors in various Internet services such as YouTube [5], Twitter [6], Facebook[7], Renren[8], Foursqure[9], etc., and have developed methodologies based on machine learning for detecting the spammers in these systems. Although with different sources of the ground-truth information, these works are focused on analyzing and

detecting the spamming behaviors on the target system of spam. Different from these works, in this paper we reveal the entire spam ecosystem consisting not only the spam target website of Baidu Zhidao, but also the crowdsourcing system of ZhuBaJie.com where the spam campaigns are initialized and the spammers are organized and rewarded.

Crowdsourcing systems. The crowdsourcing systems such as Mechanical Turk and Microworkers have attracted attentions from academic researchers in recent years (e.g., [22][23]). However, there are relatively few works revealing the "dark side" of the Internet crowdsourcing systems. Wang et al. [4] investigate two Chinese-based crowdsourcing websites and their spam campaigns targeted on microblogs, and point out that the threats from the malicious campaigns are non-trivial; Lee et al. [13] study the methodologies for detecting the worker accounts on Twitter by analyzing the spam campaigns from English-based crowdsourcing websites; they also propose classifiers for detecting the spam tasks on the crowd marketplaces [24]. Unlike the previous studies focusing on microblogs, in this work we investigate the spam campaigns targeted on the community-based QA systems, which require more collaborations among the spammer accounts, thus enable us to have more insights on the spam ecosystem.

Community-based QA systems. Community-based QA systems such as Yahoo Answers and Stack Overflow have been studied in many perspectives, such as evaluating the quality of the user-generated contents [25][26], classifying factual questions from conversational questions [27], and discovering users with high expertise [28]. As far as we know, our work is the first one to study a community-based QA system under deliberately and organized spam attacks, and present solutions for detecting the spammer accounts.

VII. Conclusion

In this paper, we study the underlying crowdsourcing ecosystem that initializes, orchestrates, and finances Internet spamming. We develop crawlers for collecting data from the crowdsourcing website of ZhuBaJie.com, where the spam campaigns are initialized and spam workers are organized and rewarded, and extract millions of questions, answers, and user profiles from Baidu Zhidao, which is the largest Chinese-based question answering site and the target site of the spam campaigns under study in this paper. With the assistance of the measurement data, we present a comprehensive and insightful analysis on the crowdsourcing spam ecosystem. The topics under study include the scale and scope of the spam attacks, Sybil accounts and colluding employed by the spammers, spammers' efforts and monetary rewards, and quality control performed by the campaigners, etc. Finally, we investigate the rationales and methodologies for spammer detecting in community QA. By making use of the discrepancies between the spammer accounts and legitimate users, we show that the spammer accounts can be detected with high accuracies.

References

[1] P. May, "Crowd labor matches quirky jobs with micro-workers," June 2012, Los Angeles Times.

[2] C. Chen, K. Wu, V. Srinivasan, and X. Zhang, "Battling the Internet water army: detection of hidden paid posters," in *Proc. of the IEEE/ACM Conference on Advances in Social Networks Analysis and Mining (ASONAM'13)*, Aug. 2013.

[3] "Taobao takes aim at 'Internet Army'," Jan. 2011, Shanghai Daily.

[4] G. Wang, C. Wilson, X. Zhao, Y. Zhu, M. Mohanlal, H. Zheng, and B. Y. Zhao, "Serf and turf: Crowdturfing for fun and profit," in *Proc. of WWW'12*, Apr. 2012.

[5] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, and M. Gonçalves, "Detecting spammers and content promoters in online video social networks," in *Proc. of SIGIR'09*, Jul. 2009.

[6] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu, "Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter," in *Proc. of WWW'12*, Apr. 2012.

[7] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, "Detecting and characterizing social spam campaigns," in *Proc. of IMC'10*, Nov. 2010.

[8] Z. Yang, C. Wilson, X. Wang, T. Gao, B. Y. Zhao, and Y. Dai, "Uncovering social network sybils in the wild," in *Proc. of IMC'11*, Nov. 2011.

[9] A. Aggarwal, J. Almeida, and P. Kumaraguru, "Detection of spam tipping behaviour on foursquare," in *Proc. of WWW'13 Companion*, May 2013.

[10] "ZhuBaJie.com," http://www.zhubajie.com/.

[11] "Baidu Zhidao," http://zhidao.baidu.com/.

[12] T. Wang, G. Wang, X. Li, H. Zheng, and B. Y. Zhao, "Characterizing and detecting malicious crowdsourcing," in *Proc. of SIGCOMM'13 Poster*, Aug. 2013.

[13] K. Lee, P. Tamilarasan, and J. Caverlee, "Crowdturfers, campaigns, and social media: Tracking and revealing crowdsourced manipulation of social media," in *Proc. of the International Conference on Weblogs and Social Media*, June 2013.

[14] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman, "Knowledge sharing and yahoo answers: everyone knows something," in *Proc. of WWW'08*, Apr. 2008.

[15] S. Ye, J. Lang, and F. Wu, "Crawling online social graphs," in *Proc. of APWEB'10*, Apr. 2010.

[16] M. Liu, "Fierce movie 'water army'?" Feb. 2013, Business Value.

[17] W. Tan, "Fake reviewers boost clicks on fake goods and other products," May 2014, Shanghai Daily.

[18] M. Newman, *Networks: An Introduction*. Oxford University Press, 2010.

[19] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604 – 632, 1999.

[20] E. Frank and I. H. Witten, "Generating accurate rule sets without global optimization," in *Proc. of ICML'98*, 1998.

[21] K. Karimi and H. Hamilton, "Timesleuth: a tool for discovering causal and temporal rules," in *Proc. of Tools with Artificial Intelligence (ICTAI'02)*, 2002.

[22] P. Venetis and H. Garcia-Molina, "Quality control for comparison microtasks," in *Proc. of the International Workshop on Crowdsourcing and Data Mining*, Aug. 2012.

[23] T. Xia, C. Zhang, J. Xie, and T. Li, "Real-time quality control for crowdsourcing relevance evaluation," in *Proc. of IEEE International Conference on Network Infrastructure and Digital Content*, Sep. 2012.

[24] K. Lee, S. Webb, and H. Ge, "The dark side of micro-task marketplaces: Characterizing fiverr and automatically detecting crowdturfing," in *Proc. of the International Conference on Weblogs and Social Media*, Jun. 2014.

[25] Y. R. Tausczik and J. W. Pennebaker, "Predicting the perceived quality of online mathematics contributions from users' reputations," in *Proc. of SIGIR'10*, May 2011.

[26] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Discovering value from community activity on focused question answering sites: a case study of stack overflow," in *Proc. of KDD'12*, Aug. 2012.

[27] F. M. Harper, D. Moy, and J. A. Konstan, "Facts or friends?: distinguishing informational and conversational questions in social q&a sites," in *Proc. of CHI'09*, Apr. 2009.

[28] B. Li and I. King, "Routing questions to appropriate answerers in community question answering services," in *Proc. of CIKM'10*, Oct. 2010.