Inf Syst Front (2015) 17:1239–1251
DOI 10.1007/s10796-015-9567-0

# The classification of hackers by knowledge exchange behaviors

**Xiong Zhang**[1] · **Alex Tsang**[1] · **Wei T. Yue**[1] · **Michael Chau**[2]

Published online: 10 June 2015


**Abstract** This paper examines messages posted in a hacker forum and constructs four user profiles based on the observed behavior patterns. It starts with the development of an automated forum post classification system to understand the knowledge transfer pattern exhibited by each user over time. Two patterns, knowledge acquisition and knowledge provision, are noted to be particularly informative. Based on these two and other user characteristics, user profiles are classified into four types: guru hackers, casual hackers, learning hackers, and novice hackers. Guru hackers are knowledgeable and respectable. They usually share ideas and advice with others. Casual hackers tend to act as observers. They can be skilled hackers who show interest mainly in deriving usable information from the forum. Learning hackers are also expert hackers who utilize the forum basically for learning. They actively seek knowledge and tend to share more of it over time. Novice hackers are new learners who typically join the forum for a short period. Overall, it is found that hacker communities very much represent learning communities where meritocracy is in place.

✉ Xiong Zhang
Xiong.Zhang@my.cityu.edu.hk

✉ Wei T. Yue
wei.t.yue@cityu.edu.hk

Alex Tsang
inuki.zx@gmail.com

Michael Chau
mchau@busienss.hku.hk

[1] Department of Information Systems, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong

[2] School of Business, Faculty of Business and Economics, The University of Hong Kong, Kowloon Tong, Hong Kong

**Keywords** Hacker · Hacker profile · Knowledge transfer · Online learning · Hacker ecosystem

## 1 Introduction

Ensuring information security is one of the major challenges facing enterprises in the contemporary net-based business world. In 2012, global financial losses due to information security attacks amounted to $114 billion (Albanesius 2011). According to a recent survey, this figure has continued to increasing in recent years (PricewaterhouseCoopers 2014). Studies have found that acquiring knowledge is a primary motivation many individuals to engage in hacking (Sarma and Lam 2013; Jordan and Taylor 1998; Holt and Kilger 2008). All this suggests that knowledge transfer in the form of knowledge provision or knowledge seeking plays an important role in identifying hackers. By contrast, most available hacker categorizations in the literature are based essentially on skills, intentions and motivations (Rogers 2006; Barber 2001; Pipkin 2003). As a result, the classifications have painted rather static pictures, where a hacker is someone possessing certain skillsets with either malicious or benign intentions.

In this study, we seek to extend previous works by incorporating the knowledge transfer aspect into hacker classification. With this intention, we collected data online from a well-known hacker forum and integrated the knowledge transfer framework with different hacker types. One of our main questions concerned whether a given hacker forum was indeed a place where junior hackers could acquire skill sets necessary for climbing up the membership ladder and become skilled hackers (Smith and Rupp 2002). We also examined whether there were hacker gurus in the forum providing assistance to junior hackers, signifying that the hacker ecosystem had involved apprenticeship in hacking knowledge transfer. Thus,

we could construct more complete profiles of hacker behaviors through the lens of "dynamic" knowledge exchanges seen invariably in online communities.

Our work follows up on a recent call for greater utilization of online data in understanding hacker behaviors (Holt 2010). Some of these studies have come up with a variety of frameworks suitable for mining and identifying expert contributors in online forums (e.g., Liu et al. 2012). Some noted that every hacker goes through a long learning process in developing the ability to conduct sophisticated attacks (Bratus 2007a). However, so far, none seems to have analyzed the particular type of learning involved in hacking. In contrast, our study has yielded several interesting and potentially important findings. For instance, during knowledge transfer, guru hackers act as knowledge providers by answering questions. They also tend to stay longer in the forums. Learner hackers seek to acquire knowledge continuously and, over time, become more and more inclined to answer questions. This suggests a dynamic scenario—knowledge transfer behaviors change continuously as knowledge is accumulated. Thus, for the active users, our analysis identified four general hacker profiles—guru hackers, casual hackers, learning hackers, and novice hackers—that, together, provide an overall representation of the hacker community.

Our work has deepened the understanding of hacker behaviors. Attacks on information security of systems often involve hackers determined to exploit system weaknesses with a view to compromising the system's assets. While we have increasing numbers of high-profile attacks over the years, hackers have continued to be as secretive as ever—their identities have always been elusive to the public (Heinzen and Picciano 2009). It is now generally recognized that we need to know much more than just about the techniques of fighting hackers; a better understanding of basic hacker behaviors is also critical. For instance, firms may adopt a very different strategy in security investment if the hacker is financially motivated and strategic (Cavusoglu et al. 2008). Our study seeks to shed light specifically on the knowledge transfer aspect of hackers, and demonstrates the evolutionary aspect of knowledge seeking and provision behaviors of certain hackers. Unlike prior studies on hackers which mainly examined hackers based on small sample interviews or surveys (Rogers 2006; Sarma and Lam 2013; Barber 2001), our study is the first to use online data to examine the knowledge transfer aspects of hackers.

The following sections are organized as follows. In Section 2, we briefly introduce important related literature. In Section 3, we summarize the characteristics of our dataset. In Section 4, we explain the methodology we used to construct user profiles—a multistep process involving message orientations, user knowledge transfer patterns, and, eventually, user profile constructions. In Section 5, we discuss different user profiles derivable on the basis of our dataset. Finally, Section 6 summarizes our conclusions.

## 2 Related works

The growing ubiquity of online communities has generated a lot of research interest in online forums in recent times. The present work continues this trend by seeking to understand hacker behaviors in online communities. One of the commonly studied issues concerning online forums is online sharing, a feature related to the willingness of individuals to generate or acquire knowledge in online forums (Bock et al. 2005). Knowledge sharing involves a very broad range of activities supporting mutually beneficial collaborations between people, organizations, and so forth (Minshall 2009). Knowledge provision refers mainly to the process of contributing knowledge, offering answers, and helping others in the community by knowledgeable or core users (Georgolios et al. 2007). By contrast, knowledge acquisition involves asking for knowledge while seeking answers and help from members of the community. For a knowledge sharing community to thrive, both knowledge provision and knowledge acquisition incentives have to be present (Phang et al. 2009). Our work classifies hacker types by making use of both components of knowledge sharing: knowledge contribution and knowledge acquisition.

Many studies directed at knowledge sharing within online communities have sought to understand the motivations behind voluntary contributions of knowledge. Reputation promotion (Wasko and Faraj 2005), identity confirmation or identity verification (Ma and Agarwal 2007), perceived belonging to a community or identity bond (Shih and Huang 2012), enhancing personal professional status (Hall and Graham 2004), reciprocity relationships (Bock et al. 2005) are among the motivations found to be behind users' knowledge contributions. In contrast our work analyzes the knowledge provision and knowledge acquisition behaviors of users over time and groups them on the basis of their knowledge sharing patterns and the corresponding forum evolution patterns. We use text-mining techniques to understand the knowledge sharing orientation of online posts. It is known that knowledge exchange can involve a transactional exchange of knowledge (declarative and procedural information exchange) as well as tacit knowledge sharing (transactive learning) (Desanctis, et al. 2003). The knowledge transferred in our forum exhibits two types of knowledge exchanges: knowledge provision and knowledge acquisition.

Previous studies have shown that hacker communities are heterogeneous (Chantler 1997), which has prompted many to categorize hacker communities. Some classified on the basis of their skillsets, motivations, and intentions (Rogers 2006; Jordan and Taylor 1998; Pipkin 2003). For example, Rogers (2006) created a hacker taxonomy consisting of nine categories where expert hackers with malicious and benign intentions are called professional criminals and old guard hackers. In general, low skill hackers are considered to be novice, or script kiddies, while high skill hackers are considered to be

elite hackers, professionals, and so forth (Whitman and Mattord 2012). Pipkin (2003) classified hackers based on whether they are internal employees or external hackers. In contrast, as knowledge has been found to be a critical element in hacking activities (Sarma and Lam 2013; Jordan and Taylor 1998; Holt and Kilger 2008), our hacker classification does not focus on hacker intentions but rather on the knowledge transfer patterns observed and the reputation levels of the hackers.

Studies on hacker behaviors have found that the numbers of skillful and knowledgeable hackers are highly limited (Holt et al. 2012), hackers utilize both online and offline channels to acquire knowledge (Holt 2007), hacker subcultures are shaped by their virtual and real world interactions, hackers tend to interact with fellow hackers and use different cyberspace channels to share knowledge with others (Olson 2012). Holt and Kilger (2008) distinguished between hackers who produce hacking knowledge and hacking tools/materials (makecrafters) and hackers who consume hacking knowledge and hacking tools/materials (techcrafters). Holt and Bossler (2008) found that computer-based deviance increases the odds of being victimized online. Also, females are more likely to be a harassment target in the cyberworld. Studies have also found that an important trait of a successful hacker is the ability to conduct a "hacker way of thinking" which is having the necessary strategic mindset (Bratus 2007b). Since knowledge is critical in hacking activities, in contrast to many previous studies, we seek to understand a larger field—hacker behaviors—from the knowledge transfer perspective. Thus, the present study contributes to the broader field that aims to understand hacker behaviors.

## 3 Data

We collected raw data for our analysis from a highly ranked hacker forum. According to Alexa, a well-known web traffic data provider, this forum was ranked No. 1 in the subcategory of hacking. Although this forum is a website based in India, visitors to the forum included hackers from India (16.7 % of total), US (21 %), UK (9.1 %), followed by many other countries. Therefore, this specified hacker forum could be taken as a representative hacker forum.

The forum in question is a semi-closed forum, which means a user needs to go through a log-in process to enter the forum. The forum setting is very similar to those of other forums where discussions are organized in a thread format. In such a format, a user initiates a thread with a post, which is commonly referred to as the header. Based on the header post, other users post comments within the thread which are called the replies. In other words, a discussion thread has a header post and can have numerous replies. In this paper, we refer to all the users in the online hacker forum as hackers; we do not look at the intention of the hackers for using forum information except that they share the same interests. During data analysis, we just chose users who are noticeably active in terms of posting in the forum as the research subjects.

We downloaded forum postings starting from February 2007 to August 2010. The forum discussions consisted of a wide range of hacking-related postings as well as otherwise. More specifically, two types of information were included in our dataset:

i. Post-centric information: post ID, post title, post content, post author, post category to which the thread belongs to, and post date—the date on which each post (either header or replies) was written.
ii. User-centric information, i.e., user ID, user name, user level, date on which the user registered with the forum, and user reputation.

Our final dataset included information of 26,691 users who had initiated 90,054 threads while 47,257 users participated in those threads and posted a total of 749,955 posts. Following the format suggested by Desanctis et al. (2003), Table 1 provides descriptive summary of the dataset.

For each user, the duration since registration with the forum (the time between the date when the user registered in the forum and the date on which the user had posted his last post) was noted. It was observed that most users had stayed in the forum for about 4 months. Surely, many users would have continued to write posts subsequent to the data downloading period. However, we assumed in our dataset that users had stayed until their last post. On average, there were more than 1736 unique users who posted in the forum every month. Out of the users who had posted, 45 % (about 781) had made at least one post within two consecutive months. A user was considered to be active in a month if he/she had posted in that month.

On average, a user posted close to 8 postings per month. Each post was about 33 words long. The average user initiated less than two discussion threads and posted less than 14 replies. The discussion density (which reflects the level of user involvement in a thread) and the posting intensity were noted. In general, a thread is considered dense if it involves fewer unique contributors and the number of the posts or conversations made by each contributor increases. Thus, a higher-density discussion occurs when there are fewer unique users each writing a large number of postings. On average, threads in our dataset had more than 5 users with more than 8 postings. Over time, we see an increase in values for the different forum attributes, indicating the forum has become larger, both in terms of the number of users and the number of posts. In terms of communication behaviors, the numbers do not show any particular trend.

The forum studied by us had a specific scoring system that allowed users to give positive, neutral or negative score to a

**Table 1** Statistics of data set

| | | |
|---|---|---|
| Forum attributes | Time length since registered (in days) | 32 |
| | Number of unique contributors per month[a] | 1736 |
| | Monthly retention of users[a,b] | 781.7 |
| | Post number in each thread | 8.33 |
| | Unique user number in each thread | 5.46 |
| Communication behaviors | Message production per user | 7.7 |
| | Word count per message | 33.35 |
| | Number of involved threads per contributor[a,c] | 1.91 |
| | Replies per user[a] | 13.97 |
| | Discussion density[a,d] | 22.62 |

[a] Based on average values per 30-day period

[b] The number of users who posted in the community in two consecutive months

[c] The number of threads participated per user

[d] For each discussion thread, density is measured as (1-[number of unique contributors]/total messages)*100

given post. The reputation score of each user could therefore be calculated based on the scores from the posts that a user had authored during the study period. Typically, hackers who posted a lot had higher reputation scores. However, reputation also depended on post contents. Sometimes a hacker received penalties from others if what they had posted was not approved by others. Therefore hackers in general needed to be perceived to be posting broadly useful posts to earn higher reputation.

## 4 Model and implementation

A user-centric post content analysis system was built to investigate the changes in posting behaviors. With respect to the hacker community, we were interested in the nature of knowledge transfer exhibited in the posts and how, based on the exchanges, the posting behaviors of the users evolved and how, based on the posting behaviors, we could generate different user profiles. In view of the very large number of posts contained in our dataset, we used a text-mining approach to analyze and classify post orientations in terms of knowledge transfer in our dataset, i.e., knowledge provision, knowledge acquisition or neither. Specifically, our automatic post content classifier helped us to determine whether the given post had demonstrated knowledge transfer orientations. We then classified how the user posting behaviors changed over time in terms of knowledge transfer orientations. Finally, from the knowledge transfer patterns and other user characteristics, we constructed user profiles.

Figure 1 provides a technical overview of our user profile identification system. The process started with an automated post classification exercise. Firstly, 10,000 posts were randomly extracted from the dataset and coded manually in terms of the knowledge transfer orientation observed by three student research assistants. Following the 80/20 rule, the 80 % of the coded posts were assigned to the training set and the remaining 20 % to the testing set. For each knowledge transfer orientation, one support vector machine (SVM) was trained using the training set and validated using the testing set accordingly. Next the validated classifiers were applied to classify all the posts in the dataset. To achieve user categorization, the knowledge transfer numbers were summed up for all the messages posted by all users. On the basis of different posting patterns, users could then be categorized so as to arrive at generalized user profile categories.

### 4.1 Ground truth generation

Firstly, we coded the posts in the training and testing sets based on the following three dimensions. We extracted 10,000 posts randomly from the full data set and manually labeled them based on knowledge transfer orientations. Basing on its contents, each post was assigned to one of the following three classes:

*Knowledge Acquisition*—includes questions, doubts, requests, advice seeking, and anything else reflecting the user's request for information.
*Knowledge Provision*—includes answers, tutorials, teachings, troubleshooting, guidelines, demonstrations, and anything reflecting information offering.
*Neither*—the post does not exhibit either of the above two orientations, meaning it does not have clear knowledge acquisition and knowledge provision orientation.

We conducted some preliminary investigations before we started coding the knowledge transfer orientations in our dataset manually. Some of the posts showed very clear knowledge acquisition or knowledge provision orientations. For example, some posts clearly possessed post titles addressing or asking

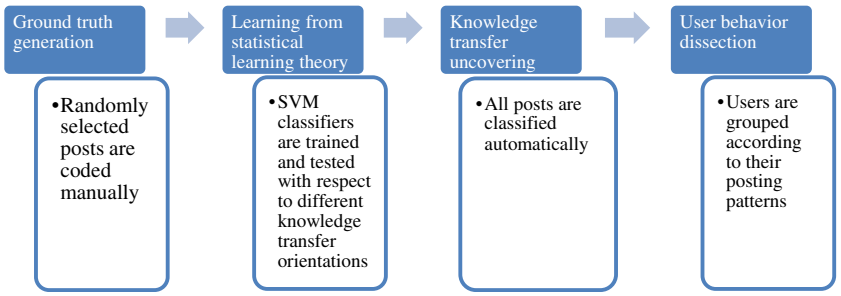


**Fig. 1** An overview of our automated post classification and user categorization system

questions. There were also other types of posts consisting of no more than greetings, expressions of gratefulness, and the like. We classified such posts as being unrelated to knowledge acquisition or provision. Sometimes, there were discussions unrelated to the first post in the thread. Based on these observations, we label the header posts and repliers manually in terms of three categories: knowledge acquisition, knowledge provision, and neither.

For each post, two student research assistants with computer science / engineering knowledge background were assigned to do the coding. In cases of inconsistent coding outcomes, a third student research assistant also coded the post. The final coding outcome was determined according to the majority rule. Finally, 80 % of the coded posts were assigned to the training set and the rest as the test set.

### 4.2 Application of statistical learning theory

We used a Support Vector Machine (SVM) to automatically classify all the posts in our dataset. An SVM is a classical classifier which demonstrates promising classification performance on applications in different domains (Abbasi et al. 2010). An SVM is a supervised learning algorithm which recognizes patterns for classification and can engage in regression analysis. It is a non-probabilistic binary linear classifier that takes a set of data as input and predicts one of two possible classes to each given point as output (Cortes and Vapnik 1995). Given a training set in which each point is assigned to one of two categories, the SVM training algorithm can train a classifier to assign new examples into one category. If we consider each post as a point in feature space, then the points of separate categories are divided by a gap between the points of two different categories. The testing examples are then mapped into this feature space so it can be predicted which side of the gap these examples fall into. Using the well-known kernel trick, an SVM can also efficiently perform a non-linear classification, which implicitly maps inputs into high-dimensional feature spaces. The mathematical definition of SVM is as follows:

In a training set $D = \{(x_i, y_i) | x_i \in \mathbb{R}^p, y_i \in \{-1, 1\}\}$, $i = 1 : n$ including $n$ points in $p$ dimensional space, $y_i$ indicates the class to which point $x_i$ belongs. The objective is to find a hyperplane $wx - \mathrm{b} = 0$ which can divide points from different classes as wide as possible. Here, $w$ is the normal vector to this hyperplane. The optimization problem can be defined as follows:

$$\begin{aligned} &\text{Minimize } \|\mathrm{w}\| \\ &\text{subject to } y_i(wx_i - b) \geq 1,\ i = 1 : n \end{aligned}$$

Keyword extraction is a very important procedure in text-mining tasks. For all the posts in the full data set, the stop words, underlines, punctuations, numbers, and spaces in the content of each post are removed first. Then the $N$-gram stemming procedure is performed on each post's content, where $N = \{1, 2, 3, 4, 5\}$. All of the stemmed words are reassembled as a space-delimited text for keyword extraction (Porter 1980). Words appearing at least five times in all the posts are recognized as keywords. In total, 63,277 keywords were extracted from our full dataset. In each post category, the keywords recognized were assigned consecutive feature IDs.

To measure the importance of a keyword drawn from a post in the dataset, we chose the *terms frequency* (*tf*) and *inverse document frequency* (*idf*) as numerical statistics. Mathematically,

$$tf{*}idf(t, d, D) = tf(t, d) \times idf(t,\ D)$$

where the value of $tf{*}idf$ is the product of the two statistics. *Term frequency* was defined as the number of times that term $t$ appears in post $d$. *Inverse document frequency* is defined as $idf(T, D) = log \frac{|D|}{|\{d \in D : t \in d\}|}$, where $|D|$ is the total number of posts in the data set and $|\{d \in D : t \in d\}|$ is the number of posts where term $t$ appears. Each $tf{*}idf$ value for a keyword found in a post is included as a feature prefixed with its ID.

Finally, the training set was used to train the SVM classifier. The SVM is trained using a radial basis function kernel because of its promising performance. Since an SVM is by nature a binary algorithm, a "one versus all" strategy was applied, where a single classifier is trained per class to distinguish one class from others (Joachims 2002). For each post, class prediction is performed using each trained classifier and choosing the prediction with the highest predicted value.

Applying the above classification process, we could learn that the header post and replies were associated with different keywords when it came to the two knowledge transfer orientations. The keywords identified in header posts were found usually to be shorter and neater. At the same time, there were many deformed keywords contained in replies; some even combined of different words. Thus, we could train classifiers

separately for headers and replies. Sample keywords for both headers and replies are shown in Table 2.

The classification results in the test set are reported in Table 3.

In general, header posts tend to exhibit more distinct knowledge orientations whereas, in replies, users often show greater diversity in terms of the things that they want to express. For example, suppose a thread was initiated asking about how to hack a forum in Cyprus but, in the replies, other users happened to focus on the language used in Cyprus. Obviously, the subsequent discussions concerned a totally different topic than that addressed by the header post. In general, given the complex and noisy nature of the post content (such as large numbers of misspellings, a variety of linguistic patterns, posted URL links, etc.), it is expected that our classification results would not be as high in terms of accuracy and precision as is found in classification results associated with more structured forums studied in the literature.

### 4.3 Uncovering knowledge transfer

The classifiers generated were next applied to classify all the headers and replies in the dataset respectively. Based on the classified posts, the changing trends of posting behaviors with respect to the three post classes for each were observed so as to further reveal users' behavioral patterns.

### 4.4 Dissecting user behavior

Having determined post orientations with respect to knowledge transfer for the entire dataset we analyzed users' posting behaviors at the individual level. For each user, we determined the post orientation of posts of the user in terms of three knowledge transfer patterns: knowledge provision, knowledge acquisition, and neither. For each knowledge transfer pattern, we examined how knowledge transfer orientation changed as we looked at the posting in a sequential order. For example, does the number of a user's knowledge provision posts increase or decrease over time? Specifically, we chose five posts as a stamp and counted the post number of each knowledge transfer pattern within these five posts. The counting generated a data point for each knowledge orientation for a user. We then examined five posts of the same user to generate the next data point. In other words, for every five posts, we count the number of knowledge acquisition post, knowledge provision post, and 'neither' posts.

We choose five posts as a basic unit as a balance among random noise and post number of each hacker. We can obtain similar results when we chose two posts as a unit, but that introduced an average of 5 % more standard error in regression. On the other hand, when we chose eight posts as a unit, there were fewer data points observe user behavior and the average standard error regression increased to almost 30 %. Hence, in general, we wanted to capture the trend in hackers' behaviors and while try to minimize random noise. We then observed how the numbers with respect to each orientation changed as the user posted further messages. Based on the data points, we analyzed the knowledge transfer pattern using linear regression. Finally we examined the knowledge transfer patterns of the user, i.e., we asked whether the user had posted more or fewer messages related to knowledge acquisition and knowledge provision following the post sequence and the rate at which the changes occurred.

In this process, we disregarded users who had s fewer than 10 posts. We regarded such users to be inactive because, according to our observation, most knowledge in online communities was contributed by a small number of active users. Further, we were unable to examine the behaviors of inactive users because not enough records for these users were included in our data set. Inactive users wrote just a few posts in almost 3 years. In particular, we removed 36,184 inactive users with 2.77 posts per user on average. In the end there were 11,073 active users with 58.68 posts per user and 649,714 posts in total. Aside from that, users exhibiting standard errors larger than the average value were also excluded. Finally, 2053 users revealed clear posting behavioral patterns. For each user, we observed the posting patterns over time with regard to knowledge acquisition and knowledge provision. For example, a user could exhibit a knowledge transfer pattern characterized by increasing knowledge provision and decreasing knowledge acquisition.

After analyzing the knowledge transfer behaviors of other user characteristics as well, we could group together users

**Table 2** Keyword samples for headers and replies

| Knowledge acquisition | | Knowledge provision | |
|---|---|---|---|
| Header | Reply | Header | Reply |
| Request | howto | demonstr | advice/suggest |
| Ask | need/want | answeri | teachyourself |
| Doubt | troublesom | recommend | guide/tutori |
| Why'd | commentsthanks | follow | easy-to-follow |

**Table 3** Classification of headers and replies in test sets

| | Knowledge acquisition | Knowledge provision | Neither |
|---|---|---|---|
| Header | | | |
| Accuracy | 79 % | 69 % | 70 % |
| Precision | 79 % | 63.24 % | 67.24 % |
| Recall | 98.73 % | 87.86 % | 78 % |
| Reply | | | |
| Accuracy | 75 % | 68.7 % | 73 % |
| Precision | 77 % | 68.2 % | 70.1 % |
| Recall | 83.12 % | 73.22 % | 79 % |

who showed similar patterns. The grouping involves computing the average value of the knowledge patterns of all the users in a group. Figure 2 shows the corresponding knowledge transfer patterns – the average number knowledge provision, acquisition, and no orientation posts authored by the same type of hackers as authored more posts in the forum. Overall, based on active users, four user groups were generated and associated eventually to four user profiles. For obvious reasons, we label the four types as guru hackers, casual hackers, learning hackers, and novice hackers. The next section engages in further discussions on these four types of users.

Using an approach similar to generating the knowledge transfer patterns of the users, we also computed the changes in reputation following the post sequence for all users. For each user group, we computed the average reputation change following post sequence. Figure 3 shows the reputation changes for different groups of users.

## 5 Results and analyses

Four user groups were identified using the system described in the last section. The groups were constructed based on their respective knowledge transfer patterns. Next we studied the characteristics of the four groups in greater detail and, eventually, profiled the four groups as guru hackers, casual hackers, learning

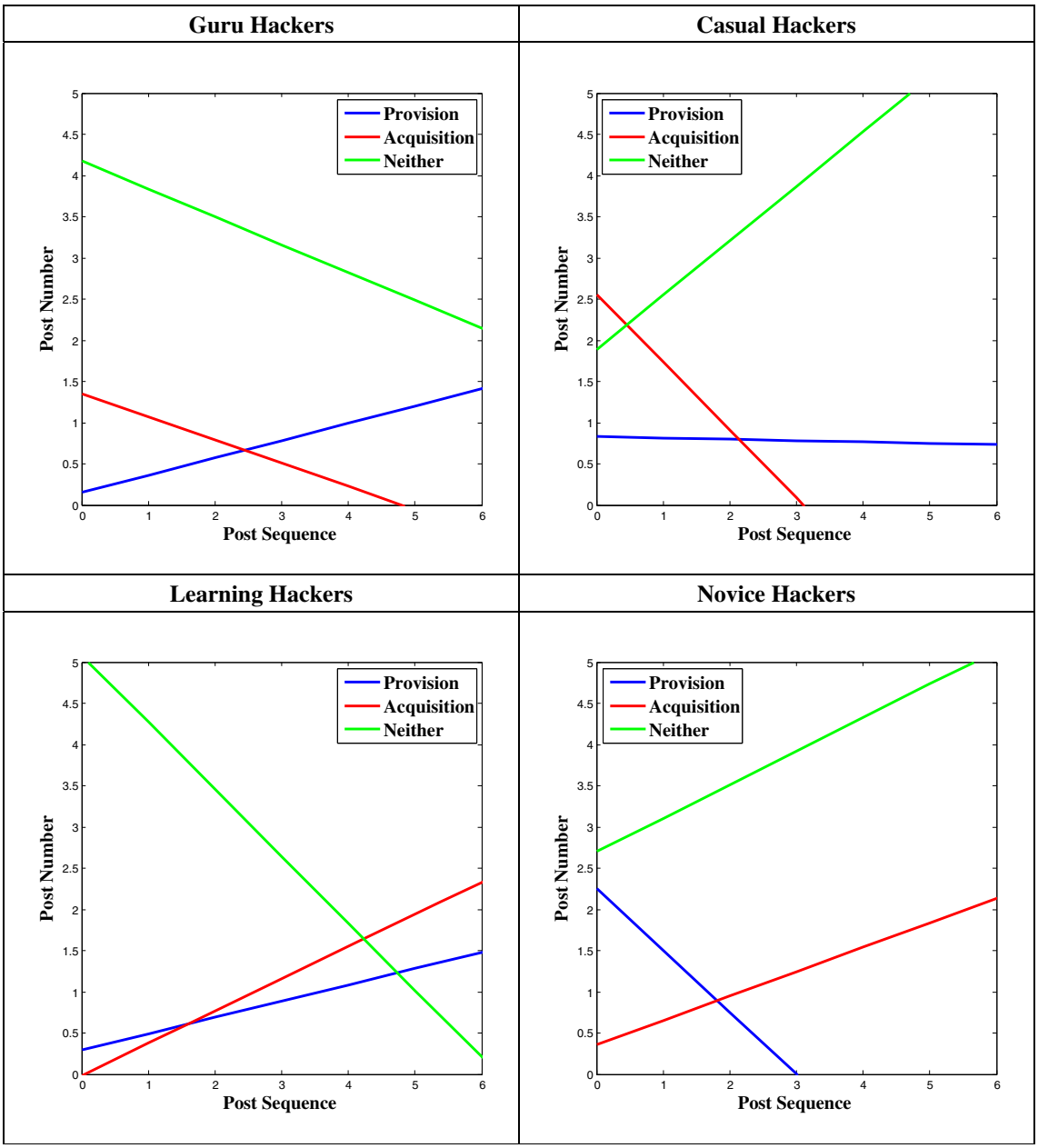


**Fig. 2** Knowledge transfer patterns of different types of users

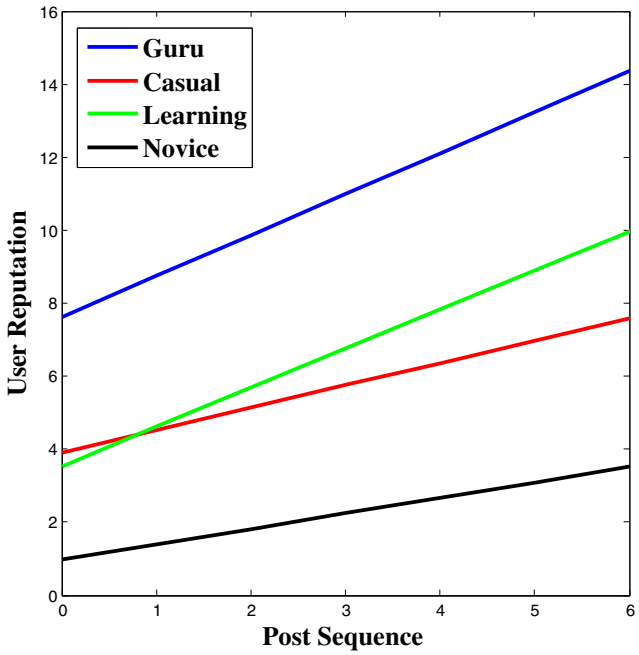


**Fig. 3** Reputation variation trend

hackers, and novice hackers. Our classification scheme is consistent with those in (Rogers 2006), although it focuses more on knowledge transferring and sharing. In this section, we provide further discussion and justification of the four user profiles. In Table 4 we present more user characteristics based on different groups. Note that regression results on the knowledge transfer patterns and reputation trends discussed in the previous section are presented first. We also present the reputation scores, user time lengths since registered, lengths of headers and replies, and thread densities for the four groups of users. From a social networking perspective, note the average local centrality for each category of hackers. The interactions among hackers at the group level will also be discussed.

Table 4 lists the descriptive statistics for the four types of hackers. For each hacker type, we use both slope and intercept to capture the variation trends of knowledge transfer behaviors and reputation. The slopes measure how hackers' posting behaviors evolve over time. For example, a positive slope for the category of knowledge acquisition means the given type of hackers asks increasingly frequently over time. The intercept measures their initial percentage of knowledge acquisition/provision posts after they had joined the forum. In general, the intercept measures the initial status while the slope measures the behavior trend over time for hackers. To zoom into the detailed behaviors for each type of hackers, we also examined how each had initiated and/or replied to posts in different sub-boards.

In online communities, some sub-boards are usually more popular than others because different sub-boards focus on different topics. In Table 5, we looked further at the average number of headers and replies posted by different types of users. We are interested in which categories the different hacker types make their postings. In our dataset, more than 90 % of posts belonged to the sub-boards of Beginner Hacking (33 %), Hacking Tools and Programs (23.8 %), Website and Forum Hacking (15.7 %), and Botnets, IRC Bots, and Zombies (18.6 %). Only 6.1 % of posts belonged to Request for Hacking, and 2.6 % into Proxies and Socks. In the section, we present a more detailed description of each type of users.

### 5.1 Guru hackers

Among the 2053 users, we found that 288 users (or 14.03 % of users) had exhibited the pattern of increasing knowledge provision combined with decreasing knowledge acquisition over time. We found evidence that users of this type continuously offer more help and share more knowledge. Such users are likely to be expert hackers because their reputation scores are the highest among the four types of users. Furthermore, the increase in reputation is the largest among the four groups. Next, we drilled down by looking at some of the sample posts written by this type of users and noted evidence that they were mainly engaging in discussions involving sophisticated hacking techniques. Some sample messages written by one of the users profiled to be a guru are reproduced below this paragraph. This hacker seems to know numerous hacking techniques and tools. He would also like to share some hacking software (which he probably made by himself) with others. More importantly, he pays attention by offering suggestions to others. The average local centrality of this type of hacker is the highest among others. This type also interacts with more users in the forum initiating posts or replying to others' posts (as in Table 4). This explains why guru hackers enjoy such high reputations. Although different classification schemes were applied, similar groups of hackers (termed elsewhere as old guard hackers, virus writers or professional criminals depending on hacking intention) with deep technique skills have been identified in the literature (Rogers 2006).

"*Nice share, did you copy from MilWorm or is this new?*"

"*You could start with a R.A.T such as Spy-Net or Cybergate. Both of them have a keylogger in them i think.*"

"*Nope, He just took Swarm Bot because it got released on opensc and tried to make it look like his. But SwarmBot is stable from what i hear and it's released now.*"

"*Schwaze Sonne Code Crypter 0.3. In this new Version you have NO restrictions anymore: you can use multiple markers, you can place the markers on ALL places now and so on. Source + Tutorial included! Code:* http://www.mediafire.com/?m54bdo6hhek505l"

Note that the users in this group had stuck with the forum for a relatively longer time. They tend to be very active in most sub-boards. In particular, guru hackers posted more in beginner hacker group as replies (on average, one guru hacker posts more than 14 replies in Beginner Hacking and almost 3 replies in Request for Hacking). These high reply numbers serve as an indicator that

**Table 4** Descriptive statistics of different hackers

| | | Guru hackers | Casual hackers | Learning hackers | Novice hackers |
|---|---|---|---|---|---|
| Number of users | | 288 | 676 | 704 | 385 |
| Provision trend[a] | Slope | >0 (0.209) | <0 (−0.016) | >0 (0.197) | <0 (−0.749) |
| | Intercept | 0.156 | 0.829 | 0.295 | 2.009 |
| Acquisition trend[b] | Slope | <0 (−0.280) | <0 (−0.822) | >0 (0.391) | >0 (0.296) |
| | Intercept | 1.348 | 2.553 | −0.012 | 0.360 |
| Neither[c] | Slope | <0 (−0.338) | >0 (0.661) | <0 (−0.813) | >0 (0.407) |
| | Intercept | 4.174 | 1.888 | 5.082 | 2.700 |
| Reputation trend[d] | Slope | 1.120 | 0.612 | 1.071 | 0.425 |
| | Intercept | 7.634 | 3.916 | 3.540 | 0.970 |
| Average reputation score | | 56.58 | 20.32 | 22.85 | 4.86 |
| Time length since registered (in days) | | 192.81 | 175.75 | 177.87 | 137.97 |
| Length of header | | 57.04 | 52.06 | 55.96 | 68.63 |
| Length of reply | | 17.03 | 15.01 | 16.72 | 21.03 |
| Density[e] | | 27.33 | 26.62 | 27.24 | 27.23 |
| Degree centrality | | 50.37 | 36.96 | 34.06 | 16.25 |
| Number of interactors | | 844 | 607 | 736 | 396 |

[a] R-squared value is between 0.32 and 0.55

[b] R-squared value is between 0.26 and 0.69

[c] R-squared value is between 0.32 and 0.64

[d] R-squared value is between 0.02 and 0.14

[e] Density is measured based on the threads initiated by the users in the different respective group

they like to answer questions in these sub-boards to help others. They also tend to start threads in other sub-boards focusing on highly technical topics, such as Website and Forum Hacking and Hacking Tools and Programs. They tend to post tutorials in sub-boards. The discussion density of the threads initialized by guru hackers is also the highest among all types of hackers. This phenomenon indicates that, in threads initiated by guru hackers, other hackers not only participate but also engage in deep discussions. This could be partly because of the reputation status of the guru hackers. It could also be because guru hackers usually initiate threads that are most relevant to others. Table 4 also shows that guru hackers have the highest value in the local centrality measurement, indicating that they are the focal points of the hacker forum.

### 5.2 Casual hackers

Six hundred seventy-six users in our dataset exhibited behavior pattern similar to a casual hacker. We reached this conclusion because of the behavioral patterns of decreasing knowledge acquisition and almost constantly little knowledge provision. Users of this type tend to ask a limited number of questions and write shorter messages. While casual hackers seem to be less active, they tend to hang on to the forum. Their average local centrality is much smaller than that of guru hackers. They also tend to interact with fewer hackers in the forum through posting. As indicated by their reputation levels, they possess knowledge skills which are slightly lower than those of guru hackers. Some sample messages posted by a

**Table 5** Posting location

| | Guru hackers | | Casual hackers | | Learning hackers | | Novice hackers | |
|---|---|---|---|---|---|---|---|---|
| Category | Header | Reply | Header | Reply | Header | Reply | Header | Reply |
| Beginner hacking | 1.71 | 14.41 | 1.80 | 9.26 | 1.03 | 7.94 | 0.83 | 4.61 |
| Request for hacking | 0.42 | 2.75 | 0.39 | 1.72 | 0.28 | 1.49 | 0.18 | 1.11 |
| Website and forum hacking | 0.89 | 6.90 | 0.73 | 6.02 | 0.64 | 5.96 | 0.38 | 2.11 |
| Proxies and socks | 0.10 | 1.32 | 0.65 | 0.87 | 0.11 | 0.85 | 0.10 | 0.49 |
| Botnets, IRC Bots, and Zombies | 0.56 | 4.77 | 0.58 | 4.34 | 0.43 | 4.44 | 0.31 | 1.44 |
| Hacking tools and programs | 0.68 | 15.86 | 0.66 | 9.94 | 0.43 | 10.46 | 0.26 | 4.43 |

casual hacker are reproduced below. It seems that casual hackers are skilled hackers since they wrote posts related to hacking techniques. They also seem to pay more attention to learning the newest information about Internet communications. Other studies have also found similar classes of hackers (Rogers 2006; Landreth 1985; Hollinger 1988). For example, Rogers (2006) uses the term "cyber punk" while referring to hackers who have the capability to write simple scripts and other software in a limited fashion. Yet other researchers classify a hacker with a moderate technical ability as "the browser" since his/her since knowledge transfer behavior is less effective compared with guru hackers (Hollinger 1988).

*495 fresh & fast US Socks. Checked and filtered with ProxyFire (Timeout: 5):? Download Proxies ?*

"*Keep us updated please :) If it works for you, it should work for me, and I will be very happy.*"

"*I'm gonna call up now actually, what do I call, AT&T, if so can you provide me the number (if you want.)*"

"*Dear members, Kindly note that the blogs "Proxiesking" and "Proxy Hunter" of extremeboy are complete rip-offs of my blogs Socks24, Proxy Heaven and Elite Proxies. Thank you for reading.*"

In general, except in the sub-board of Proxies and Socks, casual hackers are not as active as expert hackers. Casual hackers tend to post significantly more headers in this sub-board than other hackers. This implies that casual hackers typically spend more time than other types of hackers in proxy-related and sock-related techniques, which enable them to stay on the Internet anonymously and communicate with other computers. This type of information is different from typical hacking knowledge. They are more related to accessories information needed for maneuvering in the online world. For example, which proxies in different countries could be used at the given moment? What are the newest socks available for target? Such information is usually valid only for a few hours. The casual hackers' posts are relatively shorter compared those of expert hackers. As they may be less interested in knowledge transfer, they also interact with fewer hackers through posting (only higher than novice hackers)—see Table 4. Besides, as shown in Fig. 2, their posts tend not to have any orientations. The discussion density of 26.62 in threads is the lowest compared with other groups. This implies that casual hackers are mainly observers of posts and are not willing to engage in deep discussions with others. Although skilled, casual hackers do not have as high reputations as guru hackers.

### 5.3 Learning hackers

We identified 706 users in our study to be learning hackers as they exhibited the patterns of increasing knowledge acquisition and increasing knowledge provision. They tended to ask more and more questions over time while sharing ever more knowledge. These patterns show that the forum had become an important place for the learning hackers to interact with others in knowledge transfer. This finding is consistent with literature (Desanctis et al. 2003) where online communities have been shown to be e-learning venues too. We also see that their reputation levels are inferior to those of casual hackers in their initial posts but, after the knowledge exchange behaviors exhibited in their subsequent posts, their reputation levels go up noticeably. Clearly, such knowledge exchange patterns and reputation changes indicate the kinds of learning behaviors being exhibited. Users of this type become more comfortable about sharing information appreciated by the hacker community. As we can see from the sample posts (see below), hackers of this type usually learn by asking detailed questions. Their posts usually request for or talk about specific techniques, targets, and the like. This is similar to findings by researchers on offline hacker communities that some members are just students who spend much time learning what they are interested in and challenge themselves (Landreth 1985).

"*I have an anon emailer site, private of course.*"

"*Any botnet lower than 3 k won't do anything to home connections. Invest in shells.*"

"*can someone help me take down the RT forums : D stupid admins need a lesson, not to be so dumb.*"

"*I have tried numerous times to open my ports, but I have only managed to open TCP at max. Would appreciate any help from those people on 2wire.*"

Learning hackers in our study seemed to utilize the forum mainly for learning. On average, most learning hackers stay in this forum for sufficiently long periods. They spend most their online time in sub-boards such as Beginner Hacking, Website and Forum Hacking, and Hacking Tools and Programs. As shown in Table 5, they posted almost 9 posts in Beginner Hacking, almost 7 posts in Website and Forum Hacking, and almost 11 posts in Hacking Tools and Programs, compared with less than 8 posts in all other sub-boards. These subgroups typically cover topics that are more hacking-centric. The local centrality of learning hackers was also quite high. Furthermore, these hackers were pretty active in the forum (only lower than guru hackers). They also initiated threads yielding high densities, indicating that they use forums as venues for learning.

### 5.4 Novice hackers

Finally, 18.75 % users exhibited patterns with increasing information acquisition and decreasing information provision (see Table 4). Over time, users of this type asked more questions and raise more topics in the forum. However, in the process, they tended not to share knowledge in the forum. They also possessed less knowledge about hacking, as shown by their reputation level. Hence, the users in this group are labeled novice hackers. Some sample posts from such hackers

are reproduced below. Note that the posts of novice hackers are relatively longer compared with those of other hackers. This may seem counterintuitive at first glance. Further investigation showed that longer posts did not necessarily mean that novice hackers had knowledge to share. On the contrary, longer posts are the results that they not only requested knowledge but also introduced themselves often as newcomers and expressed their needs for hacking techniques. The long messages could also be one of the ways that they wanted to increase the chances of getting help from others. It is also possible that this type of users could be younger in age. In some sense, the category of novice hackers in our paper is similar to the class of novice in (Rogers 2006; Landreth 1985). These novice hackers usually know little about computer hacking techniques and they rely more on available resources, like tools, codes from others, to do hacking-related activities.

"*Hey i have some screenshots from an Nmap scan of cookingbynumbers.com. I was wondering what do i do at this point. The Nmap scan textSpoiler (Click to View)*"

"*Hello i'm new here. There're some very nice posts. But how do you hack with this test? I can hack somegames, i'm using for games Cheat Engine 5.6. A password ? Please help me !*"

"*Im new to this, and i would like to find out how to hack a server by bruteforcing it. Any bruteforce wordlists for the passwords ect would be helpfull thanks H4XX'R.*"

"*im kind of new to hacking and i wanted to learn how to hack xbox live accounts a couple days back my friend was hacked and i wanted to get it back as well as get back at my enemies also i wanted to know some good host booting apps and how to host boot if you would like to add my gt is ice 3old killa c=3rd letter of alphabet i would also like to know how can i steal accounts using cain and abel i also use net tools any info will help plz help another thing i would like to be invited to a 10th prestige modded lobby*"

Novice hackers are beginners who have joined the forum in question for a relatively short time. Basically, they are still in the initial stage of learning. They post the highest number of posts in the Beginner Hacking sub-board. Aside from that, as newcomers, they show great interest in learning hacking tools and programs. They also write the longest headers and replies compared with other types of hackers. They tend to exhibit their naivity in hacking so as to increase their chances of receiving replies and help from others. One possible reason for this is that, during their early learning stages, novice hackers are unfamiliar with most hacking topics. They do not understand the larger hacker culture in the community. As shown in Table 4, they interact with the fewest numbers of hackers in the forum through posting, have the smallest local centrality and the smallest number of interactors. This suggests that they spend most of their time browsing different sub-boards while spreading their posts in different sub-boards. Their density (27.23) is also the lowest among all types of hackers, indicating that their discussions are perhaps not as engaging for similar numbers of users.

Finally, based on our analyses and discussions, we summarize the profiles of hackers in terms of some important characteristics in Table 6.

## 6 Conclusion

Online communities are venues where users can learn, share, give advice, help others, and even show off. In this study, a very popular hacker forum is chosen to download raw data for more than 3 years. Specifically, we examine how users' posting behaviors evolve over time. In view of the huge number of posts, an automated post classification system is built and used to classify posts in terms of knowledge transfer orientations. For each user, we analyze the post number variation trend in three dimensions: knowledge provision, knowledge acquisition, and neither. On the basis of different post number variation trends, active users are categorized into four types: guru hackers, casual hackers, learning hackers, and novice hackers. Online communities have been regarded as a place for knowledge transfer. In recent years, business organizations often

**Table 6** Profiles of different hacker types

| | Guru hackers | Casual hackers | Learning hackers | Novice hackers |
|---|---|---|---|---|
| Time length since registered | Long | Medium | Medium | Short |
| Density | High | Low | Medium | Medium |
| Reputation | Highest | High | High | Low |
| Activity | Active in posting replies in the Beginner Hacking and Request for Hacking subgroups | Active in Proxies and Socks subgroup | Active in beginning hacking and general hacking by posting headers and replies | Active in beginning hacking and general hacking, but with much lower activity intensity |
| Knowledge provision | Increasing | Flat | Increasing | Decreasing |
| Knowledge acquisition | Decreasing | Decreasing | Increasing | Increasing |
| Post length | Long | Medium | Medium | Longest |
| Interaction | High | Medium | Medium | Low |

adopt online communities as a means for customer support, product development, customer interactivity creation, and products reviews. The mechanism of knowledge conversion and transfer is the same in online communities (Abuhamdieh 2006). In our study, we found that people learn from hacker communities mainly to acquire skills although the same knowledge can be garnered from security professionals to protect their enterprises.

Guru hackers stay in the forum for the longest time and write reply posts in beginner hacker sub-boards. As knowledgeable hackers, they request less and offer more in the forum. Therefore, these users enjoy the highest reputation and the highest increases in reputation when compared with other types of hackers. Casual hackers usually pay more attention to some specific sub-boards such as those related to proxy and socks. As relatively skilled hackers, they spend less time in hacking-technique-related sub-boards. On the other hand, they need updated proxy and sock information. Learning hackers write most of their posts in general-hacking-related sub-boards. Users of this type have a relatively high reputation since they continuously request for more knowledge and offer more in the forum. Novice hackers are in the early learning stage. Compared with other types of hackers, they write the fewest posts, whether headers and replies, in different sub-boards. However, their posts are usually the longest. This may be because newcomers usually take time to get used to the community such as introducing themselves to others and explaining to them that they are not familiar with hacking techniques. It may also just be that they are younger.

The key implication of our study is that the hacker community is a diverse group, where there are different types of users playing different knowledge sharing roles in the community. Prior work has classified hackers based on skillsets, intentions and motivations. In this work, we have presented the hacker profiles from the knowledge transfer perspective. Our findings are consistent with prior work in that expert and novice hackers are found in the community and the hacker ecosystem involves apprenticeship in hacking knowledge transfer. The present study can shed more light on the knowledge exchange behaviors of these users using the hacker forum data. We have illustrated the dynamic properties of the changing reputation, together with the changing knowledge sharing patterns which enables us to understand the possible changes of behaviors of a hacker as knowledge is accumulated.

Our work has the following limitations: The performance of the trained classifiers in our study is not as impressive as those from studies on well-structured forums. Future researchers could try semi-supervised techniques while training classifiers so as to better exploit the huge amount of uncoded data. There could also be many information security professionals in the forum. We do not actually know the true intentions of the users with knowledge or whether they are black-hat hackers (malicious hackers acting on their own initiative for personal gain) or white-hat hackers (ethical hackers seeking under contractual agreements to expose weaknesses in specified enterprise systems). There are many directions into which one can extend the present study. First, social network analyses could be performed to further observe different interaction groups. Second, the nature of the posts in relation to actual attacks could be analyzed further. For instance, the impact of a hacker forum on actual attack or protection can be further explored using other secondary data sets.

**Acknowledgments** The work described in this paper was partial supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 152611)

## References

Abbasi, A., et al. (2010). Detecting fake websites: the contribution of statistical learning theory. *MIS Quarterly, 34*(3), 1–28.

Abuhamdieh, A. (2006). Knowledge transfer in virtual communities. *Review of Business Information Systems, 10*(4), 23–36.

Albanesius, C. (2011). Cyber crime costs $114B per year, mobile attacks on the rise. Retrieved August 2013, from http://www.pcmag.com/article2/0,2817,2392570,00.asp.

Barber, R. (2001). Hackers profiled – who are they and what are their motivations. *Computer Fraud & Security, 2001*(2), 14–17.

Bock, G.-W., Zmud, R. W., Kim, Y.-G., & Lee, J.-N. (2005). Behavioral intention formation in knowledge sharing: examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *MIS Quarterly, 29*(1), 87–111.

Bratus, S. (2007a). Hacker curriculum: how hackers learn networking. *IEEE Distributed Systems Online, 8*(10), 2.

Bratus, S. (2007b). What hackers learn that the rest of us don't. *IEEE Security and Piracy, 5*(4), 72–75.

Cavusoglu, H., Raghunathan, S., & Yue, W. T. (2008). Decision-theoretic and game theoretic approaches to IT security investment. *Journal of Management Information Systems, 25*(2), 281–304.

Chantler, N. (1997). *Profile of a computer hacker*. Seminole: Inter. Pact Press.

Cortes, C., & Vapnik, V. (1995). Support - vector networks. *Machine Learning, 20*(3), 273–297.

Desanctis, G., Fayard, A.-L., Roach, M., & Jiang, L. (2003). Learning in online forums. *European Management Journal, 21*(5), 565–577.

Georgolios, P., Kafentzis, K., & Mentzas, G. (2007). Knowledge provision with intelligent E-services. *International Journal of Intelligent Systems, 22*(5), 501–518.

Hall, H., & Graham, D. (2004). Creation and recreation: motivating collaboration to generate knowledge capital in online communities. *International Journal of Information Management, 24*(3), 235–246.

Heinzen, T. E., & Picciano, L. M. (2009). Ilk hunting: Newbies, cyberpunks, coders and the search for elusive, ego-twisted, talented computer hackers. In L. V. Shavinina (Ed.), *International handbook on giftedness* (pp. 809–823). Netherlands: Springer.

Hollinger, R. C. (1988). Computer hackers follow a guttman-like progression. *Social Sciences Review, 72*(3), 199–200.

Holt, T. J. (2007). Subcultural evolution? Examining the influence of on- and off-line experiences on deviant subcultures. *Deviant Behavior, 28*(2), 171–198.

Holt, T. J. (2010). Exploring strategies for qualitative criminological and criminal justice inquiry using on-line data. *Journal of Criminal Justice Education, 21*(4), 466–487.

Holt, T. J., & Bossler, A. M. (2008). Examining the applicability of lifestyle-routine activities theory for cybercrime victimization. *Deviant Behavior, 30*(1), 1–25.

Holt, T. J. & Kilger, M. (2008). *Techcrafters and makecrafters: A comparison of two populations of hackers*. Paper presented at WOMBAT Workshop on Information Security Threats Data Collect and Sharing (WISTDCS2008), Amsterdam, Netherlands.

Holt, T. J., Strumsky, D., Smirnova, O., & Kilger, M. (2012). Examining the social networks of malware writers and hackers. *International Journal of Cyber Criminology, 6*(1), 891–903.

Joachims, T. (2002). *Learning to classify text using support vector machines-methods, theory and algorithms, The Springer International Series in Engineering and Computer Science*. Springer.

Jordan, T., & Taylor, P. (1998). A sociology of hackers. *The Sociological Review, 46*(4), 757–780.

Landreth, B. (1985). *Out of the inner circle: A hacker's guide to computer security*. Redmond: Microsoft Books.

Liu, X. et al. (2012). Harnessing global expertise: a comparative study of expertise profiling methods for online communities. *Information Systems Frontiers*, 1–13. doi:10.1007/s10796-012-9385-6.

Ma, M., & Agarwal, R. (2007). Through a glass darkly: information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research, 18*(1), 42–67.

Minshall, T. (2009). What is knowledge transfer?. Retrieved February 2014, from:http://www.cam.ac.uk/research/news/what-is-knowledge-transfer.

Olson, P. (2012). *We are anonymous: Inside the hacker world of LulzSec, anonymous, and the global cyber insurgency*. New York: Back Bay Books.

Phang, C. W., Kankanhalli, A., & Sabherwal, R. (2009). Usability and sociability in online communities: a comparative study of knowledge seeking and contribution. *Journal of the Association for Information Systems, 10*(10), 721–747.

Pipkin, D. L. (2003). *Halting the hacker: A practical guide to computer security* (2nd ed.). Upper Saddle River: Prentice Hall Professional.

Porter, M. F. (1980). An algorithm for suffix stripping. *Readings in information retrieval* (pp. 313–316).

PricewaterhouseCoopers (2014). *The Global State of Information Security Survey 2014*, Retrieved March 2014, from: www.pwc.com.

Rogers, M. K. (2006). A two-dimensional circumplex approach to the development of a hacker taxonomy. *Digital Investigation, 3*(2), 97–102.

Sarma, M. & Lam, A. (2013). *Knowledge creation and innovation in the virtual community? Exploring structure, values and identity in hacker groups*. Paper presented at 35th DRUID Celebration Conference 2013, Barcelona, Spain.

Shih, H.-P. & Huang, E. (2012). Influences of Web interactivity and social identity and bonds on the quality of online discussion in a virtual community. *Information Systems Frontiers*, 1–15. doi:10.1007/s10796-012-9376-7.

Smith, A. D., & Rupp, W. T. (2002). Issues in cybersecurity: understanding the potential risks associated with hackers/crackers. *Information Management & Computer Security, 10*(4), 178–183.

Wasko, M. M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly, 29*(1), 35–57.

Whitman, M. E., & Mattord, H. J. (2012). *Principles of information security* (4th ed.). Boston: Course Technology Cengage Learning.

**Xiong Zhang** is a PhD candidate in the Department of Information Systems at City University of Hong Kong. His research interests includes online piracy, information security, product bundling, economics of IS, and data analytics. His work has appeared in the proceedings of International Conference on Information Systems, Workshop on Information Technologies and Systems, Pacific Asia Conference on Information Systems, among others.

**Alex Tsang** is a mobile solution specialist. He previously worked as a research assistant at City University of Hong Kong. He received his Master's degree in Information System Management from City University of Hong Kong, and Bachelor's degree in Computer Science and Technology from Sun Yat-sen University in Guangzhou, China.

**Wei T. Yue** is an Associate Professor of Information Systems at City University of Hong Kong. He received his Ph.D. in Management Information Systems from Purdue University. Prior to joining City University of Hong Kong, he was a faculty member at University of Texas, Dallas. His research interests focus on the economic and operational aspects of information security and information systems. His work has appeared in journals including *Management Science*, *Information Systems Research*, *Journal of Management Information Systems*, *Decision Support Systems*, and others.

**Michael Chau** is an Associate Professor in the School of Business, Faculty of Business and Economics at the University of Hong Kong. He received a Ph.D. degree in management information systems from the University of Arizona and a B.Sc. degree in computer science and information systems from the University of Hong Kong. His current research interests include web mining, data mining, business analytics, social media, electronic commerce, and security informatics. His research has appeared in top-tier journals including *ACM TMIS*, *ACM TOIS*, *CACM*, *DSS*, *IEEE TKDE*, *JAIS*, *JMIS*, *JASIST*, and *MISQ*. He is the author of more than 100 articles and has been ranked as the #14 most productive researcher in the field of information science in the period 1998–2007 in a research productivity study. He is also the program co-chair of the International Conference on Information Systems 2013. More information can be found at http://www.business.hku.hk/~mchau/.