2018 1st International Conference on Data Intelligence and Security

Mining Key-Hackers on Darkweb Forums

Ericsson Marin, Jana Shakarian and Paulo Shakarian
Arizona State University
Tempe, Arizona
Email: {ericsson.marin, jshak, shak}@asu.edu

***Abstract*—Recently, there is an interest in studying cyber crime from a hacker-centric perspective, whose insight is to locate key-hackers and use them to find credible threat intelligence. However, the great majority of users present in hacking environments seem to be unskilled or have fleeting interests, making the identification of key-hackers a complex problem. Moreover, as ground truth information is rare in this context, there is a lack of a method to validate the results. Thus, previous work neglected this validation step or had it done manually - by hiring qualified security specialists. In this work, we address the key-hacker identification problem including a systematic method based on reputation to validate the results. Particularly, we study how three different approaches - content, social network and seniority-based analysis - perform individually and combined to identify key-hackers on darkweb forums, aiming to confirm the following two hypotheses: 1) a hybridization of these approaches tends to produce better results when compared to the individual ones; 2) a model conceived to identify key-hackers in one forum can be generalized to other forums that lack a user reputation system or have a deficient one. We conduct our experiments using a carefully selected set of features, showing how an optimization metaheuristic obtains better performance when compared to machine learning algorithms that attempt to identify key-hackers.**

I. Introduction

Past research has found that hackers extensively share information regarding vulnerabilities in online communities, a fact strongly correlated to cyberattacks [1]. However, hacker communities usually have participants with different levels of knowledge, and those who want to identify emerging cyber threats need to scrutinize these individuals to find key-hackers [2]. Cybersecurity researchers claim that threats made by high-skilled hackers should be prioritized, since they are usually more successful in their goals [3]. Thus, an alternative path considered by those researchers to predict cyberattacks is: identify "key-hackers" first, and consequently, the emerging cyber threats. The challenge here is that key-hackers form just a small percentage of the hacking community members, making their identification a complex problem for cybersecurity.

In the literature, two categories of approaches have been considered for the identification of key-hackers in online communities: content [4], [2], [5] and social network analysis [6], [7]. In both approaches, the idea is to generate features and rank the community users based on their corresponding feature values, with the top rank users considered as key-hackers [4].

Additionally, there is another challenge which imposes more complications for this identification task: the lack of ground truth (the existing key-hackers of the communities to validate results). Usually, it is difficult to obtain this information, which made previous works neglect the validation step or have it done manually - e.g., by leveraging security consultancy companies [8]. Also, it is unclear if these methods can generalize, as training and testing are often done using the same forum data.

In this work, we address the key-hacker identification problem including a systematic method to validate the results. Particularly, we study how content, social network and seniority analysis perform individually and combined in this task. We conduct our experiments using a carefully selected set of features extracted from three highly ranked hacker forums on darkweb. Information related to activity, expertise, behavioral trend, structural position, influence and coverage are mined to develop a profile for each community member, in order to understand which features are unique to key cybercriminals.

To train and test our model, we use an optimization metaheuristic and compare its performance with machine learning algorithms. We leverage the users' reputation scores provided by the three forums analyzed to systematically cross-validate the results across those sites - models trained in one hacker forum are generalized to make predictions on different ones. Our work is novel since it offers researchers a solid strategy to find key-hackers on forums with no users' reputation scores, or with a deficient user reputation system. We observe this is the case of the vast majority of hacking forums, representing over 80% of the 36 that were scraped for this paper.

This work makes the following main contributions:

1) We show that a hybridization of features derived from content, social network and seniority analysis is able to identify key-hackers up to 17% more precisely than those derived from any of these strategies by itself;
2) We demonstrate how a model learned in a given hacker forum to identify its key-hackers, can be generalized to a different forum which was not used to train the model;
3) We compare the performance of different models when trying to identify key-hackers, showing how an optimization metaheuristic obtains up to 35% of predictive improvement over machine learning algorithms;
4) We show how users' reputation scores can be used to identify the characteristics of key-hackers.

The rest of the paper is organized as follows: Section II introduces darkweb forums, detailing our dataset and the user reputation score. Section III describes the features we derive to profile hackers. Section IV presents our experiments and Section V exhibits the corresponding results. Section VI shows some related work. Finally, Section VII concludes our work.

978-1-5386-5762-1/18/$31.00 ©2018 IEEE
DOI 10.1109/ICDIS.2018.00018

II. Darkweb/Deepweb Forums

Many people involved in malicious cyber activity rely on trustful online communities, among which, forums are the most prevalent [4]. For example, the recent "WannaCry" ransomware attack directed against hospitals in the UK and numerous other worldwide targets was discussed several weeks prior on a darkweb forum [9]. Hackers likely involved in this attack discussed the number of unpatched machines, the exploit to be used, the industry verticals, and the method of attack (ransomware). These forums provide user-oriented platforms that enable communication regardless geophysical location, facilitating the emergence of communities of hackers.

The World Wide Web is a vast network of linked hypertext files where forums are accessed via the Internet, being classified into 3 regions: surfaceweb, deepweb and darkweb [10]. Surfaceweb is the open portion of the Internet, where webpages are publicly accessible and indexed by search engines. On the other hand, deepweb refers to the websites hosted on surfaceweb but not indexed by search engines, usually because they require authentication. Finally, darkweb refers to a collection of websites that exist on encrypted networks of deepweb. It is a region intentionally and securely hidden from users, search engines and regular browsers. This is why darkweb forums constitute the most widespread hacking environment, and therefore the most commonly used data source for studies investigating key cybercriminals [4].

A. Dataset

In this work, we collect data provided by a commercial version of the system described in [10], from where we select three popular English hacker forums on darkweb. We anonymize these forums representing them as *Forum 1*, *Forum 2* and *Forum 3*, showing their statistics in Table I.

TABLE I
Darkweb Forums' Statistics.

	Forum 1	*Forum 2*	*Forum 3*
Time Period	2013-12-24 : 2016-03-16	2013-12-24 : 2016-08-16	2002-09-14 : 2016-03-15
Number of Users	4,380	2,495	2,802
Number of Topics	5,571	1,077	5,805
Number of Posts	36,453	25,115	49,078
Distinct Values of Users' Reputation	134	102	37

All these darkweb forums comprise discussions organized in a thread format, where a user initiates a topic (commonly referred to as the header), followed by many users' posts (replies). The discussions consist of a wide range of hacking-related messages posted by community members. For this paper, we collect post-centric information (topic author and content, reply author and content, topic and reply dates) and user-centric information (user ID, user reputation).

In order to prepare the data for feature extraction, we retrieve the interactions between hackers over time, generating a directed network according to their posts. We denote a set of users V and a set of connections E, as the nodes and edges in a directed graph $G = (V, E)$, a set of topics Θ, a set of messages $\mathcal{M}$ and a set of discrete time points T. We will use the symbols v, θ, m, t to represent a specific node, topic, message and time point. We denote an activity log $\mathcal{A}$ containing all posts (topics and replies) as a set of tuples of the form $\langle v, \theta, m, t \rangle$, where $v \in V$, $\theta \in \Theta$, $m \in \mathcal{M}$ and $t \in T$. It describes that "v posted in topic θ a message m at time t". A directed edge (v, v') is created when users v and v' post together in a given topic, so that the posting time of v is greater than of v'. We formalize the set of direct edges E in equation (1) bellow:

$$E = \{(v, v') \,|\, \exists \langle v, \theta, m, t \rangle \in \mathcal{A}, \exists \langle v', \theta, m', t' \rangle \in \mathcal{A}, \quad s.t.\ v \neq v', t > t'\} \tag{1}$$

The intuition here is to make visible to users that are posting in θ at time t, all other users that have already posted in θ prior to t, but not vice-versa. We believe this strategy can better reproduce the interaction process in online forums (compared to the general strategy of creating a complete undirected graph, including all users that post together in a topic [7]), since users will be aware about only previous posts. Figure 1 illustrates this process by showing (a) the original users' posts, and (b) the corresponding directed social network generated.

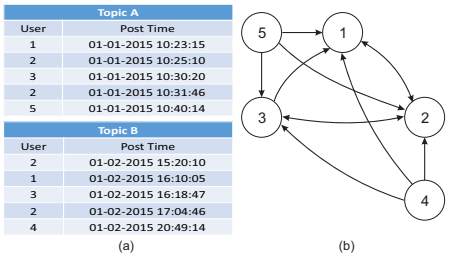

Fig. 1. Directed graph generated from the users' posts.

B. Data Cleaning

After generating the social networks, we conduct a data cleaning process. We remove all users that do not belong to the giant component of their corresponding forum, since they can produce misleading centrality values. The issue happens because some centralities are computed and normalized for each component, which tends to produce high values for users in small parts of the networks. Because of their few connections, these individuals would hardly be considered as key-hackers. Table II shows the size of all components of each forum, detailing that 67 users from *Forum 1*, 78 users from *Forum 2* and 65 users from *Forum 3* were removed.

TABLE II
Network Component Analysis.

	Forum 1	*Forum 2*	*Forum 3*
Giant Component Size	4,313	2,417	2,737
Component Size = 11	0	1	0
Component Size = 3	1	0	0
Component Size = 2	19	6	7
Component Size = 1	26	55	51

C. Users' Reputation

As hacker communities form meritocracies [11], [12], members own different levels of capability, expertise and influence (to mention a few human factors). According to [13], those factors are organically consolidated in the user reputation score, which is a metric that codifies users' standing, driving engagement by measuring participation, activity, content quality, content rating, etc. Zhang et al. [2] showed that hackers

who own high reputation are usually linked to emerging cyber threats, becoming a strong indicator of key cybercriminals.

A case study in our own data supports these findings well. In 2016, Anna Senpai had a high reputation score when he released the source code of the Mirai Botnet on a popular hacker forum [14]. His posts generated a high number of responses, though none more than the post containing the code. Since user reputation on these forums is peer-assigned, the reputation score is mirroring how other forum members evaluated the usefulness of the user's contributions.

Consider the analysis of posting patterns of high and low reputation hackers done in *Forum 1*, *Forum 2* and *Forum 3*, and presented in Figure 2. For all these forums we observe a similar pattern that corroborates with the Zhang's assumption.

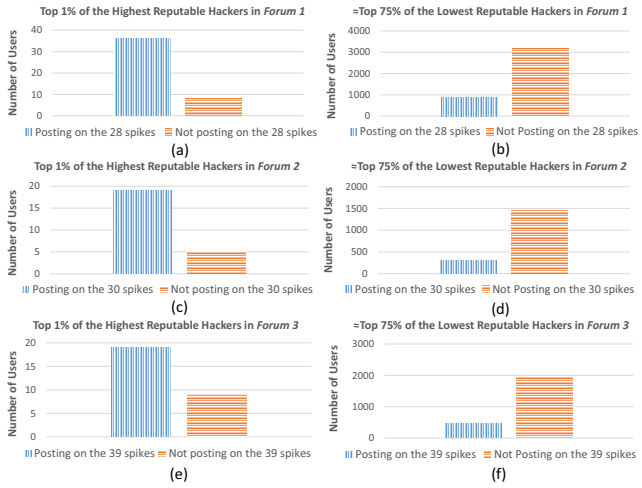

Fig. 2. Posting pattern analysis of the highest and the lowest reputable users in (a),(b) *Forum 1*, (c),(d) *Forum 2* and (e),(f) *Forum 3*.

Figure 2(a) shows that 36 out of the 44 hackers with the highest reputation ($Top_{1\%}$ of all users) are actually posting on the 28 spikes of activity (minimal of $4 \times \sigma$ above average) observed in *Forum 1*, while only 8 of them are not engaged in these conversations. On the other hand, Figure 2(b) shows that 870 out of the 4.039 hackers with the lowest reputation ($\approx Top_{75\%}$ of all users) are posting on those spikes, while 3.169 are not engaged in these conversations. Using the same analysis in *Forum 2*, we observe 19 out of the 24 hackers with the highest reputation posting on the 30 existing spikes (Figure 2(c)), while 1.450 out of the 1.761 hackers with the lowest reputation are not engaged in these conversations (Figure 2(d)). Finally, for *Forum 3*, we observe 19 out of the 28 hackers with the highest reputation posting on the 39 existing spikes (Figure 2(e)), while 1.806 out of the 2.105 hackers with the lowest reputation are not engaged in these conversations (Figure 2(f)).

We analyze those spikes, since they offer some intuition about possible interesting topics promoted by skilled and influential hackers on the forums, confirming the user reputation score as a strong indicator for key-hacker identification.

In this paper, we also rely on the assumption that users with high reputation form our set of key-hackers, using this metric as our ground truth. This way, we deliberately select forums that explicitly provide this information, so that the corresponding key-hackers can be easily identified. Figure 3 shows the distribution of the reputation score in (a) *Forum 1*, (b) *Forum 2* and (c) *Forum 3*, pointing out the existence of a hacking meritocracy. As it is observed, only a few number of users (key-hackers) own high reputation in all the analyzed forums, although their corresponding score vary in magnitude.

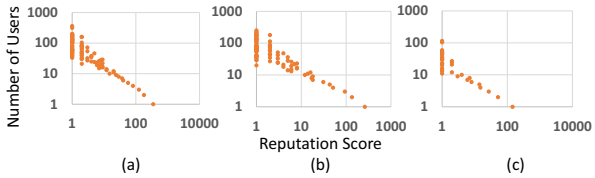

Fig. 3. Distribution of reputation score in (a) *Forum 1*, (b) *Forum 2* and (c) *Forum 3* (log-log scale). These curves fit power-laws with $p_k \approx k^{-0.92}$, $p_k \approx k^{-0.86}$ and $p_k \approx k^{-0.78}$ respectively, where k is the number of users.

In addition, Table I shows that *Forum 1* and *Forum 2* are closer in terms of reputation distinctness, since this score is formed by 134 and 102 different values respectively, while *Forum 3* owns only 37. Nevertheless, the similar distribution pattern of all forums informs us that characteristics of the highest reputation hackers should be somehow shared between them. Based on this intuition, we explain now how we learn those characteristics in the forums that provide the user reputation score, to use them in the forums that do not provide or have a deficient user reputation system.

III. Feature Engineering

In order to mine relevant characteristics and behaviors of key-hackers, we design 25 features to estimate users' reputation on darkweb forums. Table III shows a subset of 17 features extracted using *Content Analysis*, subdivided in *Activity* (3), *Expertise* (10) and *Behavioral Trend* (4). We further subdivide the 10 *Expertise* related features into *Involvement Quality* (7), *Cybercriminal Assets* (1) and *Specialty Lexicons* (2). In general, content analysis related features examine conversations among participants, so that the magnitude and quality of the corresponding contributions can be mined.

In addition, Table IV shows 5 features extracted using *Social Network Analysis (SNA)*, subdivided into *Structural Position* (3) and *Influence* (2). Social network analysis related features study group structure and participant interactions, constructing a network of hackers to focus on those with high centralities. Finally, Table V shows the last 3 features related to *Coverage*, which are extracted using what we call *Seniority Analysis*. We analyze seniority, since it generates indicators of forum involvement over time, with the features measuring how long and how much the hackers continuously post [15], [3], [16].

We will show this set of features comprises a variety of information that can differentiate key-hackers from the standard ones, helping us to better estimate the real users' reputation. All features have their values normalized to avoid problems with different scales [17]. In the following set of tables (Table III, Table IV, Table V), we present the features with their categories, description, reference and formalization.

TABLE III
CONTENT ANALYSIS FEATURES

Categories/Sub-Categories		Features
Activity		**Topics Created *(TOC)*:** Number of topics created by hackers. According to [6], key-hackers create few topics with high relevance. $TOC_{(v)} = \sum_\theta g(\langle v,\theta,m,t\rangle)$, with $g(\langle v,\theta,m,t\rangle) = \begin{cases} 1, \ if \ \langle v,\theta,m,t\rangle \in \mathcal{A} \wedge \nexists \langle v',\theta,m',t'\rangle \in \mathcal{A}, \ s.t. \ t' < t \\ 0, \ otherwise \end{cases}$ (2)
		Replies Created *(REC)*: Amount of times each user replies the topics. According to [6], [5], [2], [4], key-hackers create a huge quantity of quality replies offering help to the low-skilled community members. $REC_{(v)} = \sum_\theta \sum_m \sum_t |\langle v,\theta,m,t\rangle| - TOC_{(v)}, \ s.t. \ \langle v,\theta,m,t\rangle \in \mathcal{A}$ (3)
		Replies by Month *(REM)*: Monthly average of replies created by hackers. Previous work [4] claims key-hackers have a minimum frequency of posting over time to keep their acquired status. $REM_{(v)} = \frac{REC_{(v)}}{\lvert Months_{(v)}\rvert}$, where $Months_{(v)}$ is the set of distinct months in which v posted. (4)
Expertise	**Involvement Quality**	**Length of Topics *(LET)*:** Average number of words used by hackers to create a topic. Previous works [4] state that key-hackers do not create extensive topics, since this is a characteristic of low-skilled users. $LET_{(v)} = \frac{\sum_\theta h(\langle v,\theta,m,t\rangle)}{TOC_{(v)}}$, with $h(\langle v,\theta,m,t\rangle) = \begin{cases} Length(m), \ if \ \langle v,\theta,m,t\rangle \in \mathcal{A} \ \wedge \ \nexists \langle v',\theta,m',t'\rangle \in \mathcal{A}, \ s.t. \ t' < t \\ 0, \ otherwise \end{cases}$ (5)
		Length of Replies *(LER)*: Average number of words used by hackers to create a reply. According to [6], key-hackers produce detailed answers, as they have interest and knowledge to teach the other users. $LER_{(v)} = \frac{\sum_\theta \sum_m \sum_t i(\langle v,\theta,m,t\rangle)}{REC_{(v)}}$, with $i(\langle v,\theta,m,t\rangle) = \begin{cases} Length(m), \ if \ \langle v,\theta,m,t\rangle \in \mathcal{A} \wedge \\ \qquad \exists \langle v',\theta,m',t'\rangle \in \mathcal{A}, \ s.t. \ t' < t \\ 0, \ otherwise \end{cases}$ (6)
		Topics Density *(TOD)*: Average number of users posting in topics created by a given hacker [2]. As discussions started by key-hackers are relevant to the community, they often promote a higher engagement. $TOD_{(v)} = \frac{\sum_\theta j(\langle v,\theta,m,t\rangle)}{TOC_{(v)}}$, with $j(\langle v,\theta,m,t\rangle) = \begin{cases} \sum_{v'} \sum_{m'} \sum_{t'} |\langle v',\theta,m',t'\rangle|, \ if \ \langle v,\theta,m,t\rangle \in \mathcal{A} \wedge \\ \qquad \nexists \langle v',\theta,m',t'\rangle \in \mathcal{A}, \ s.t. \ t' < t \\ 0, \ otherwise \end{cases}$ (7)
		Replies with Knowledge Provision *(RKP)*: Number of replies including knowledge provision keywords *(kpk)*, such as "suggest", "check", "recommend", "guide" and "follow". Zhang et al. [2] observed key-hackers produce more replies providing than requesting information. $RKP_{(v)} = \sum_\theta \sum_m \sum_t k(\langle v,\theta,m,t\rangle)$, with $k(\langle v,\theta,m,t\rangle) = \begin{cases} 1, if \ \exists w \in m \wedge w \in Key_{\{kp\}} \wedge \\ \ \langle v,\theta,m,t\rangle \in \mathcal{A} \ \wedge \ \exists \langle v',\theta,m',t'\rangle \in \mathcal{A}, \ s.t. \ t' < t \\ 0, \ otherwise \end{cases}$ (8) where $Key_{\{kp\}}$ is the predefined set of knowledge provision keywords.
		Replies with Knowledge Acquisition *(RKA)*: Number of replies including knowledge acquisition keywords *(kak)*, such as "doubt", "fail", "struggling", "request" and "need". $RKA_{(v)} = \sum_\theta \sum_m \sum_t l(\langle v,\theta,m,t\rangle)$, with $l(\langle v,\theta,m,t\rangle) = \begin{cases} 1, if \ \exists w \in m \wedge w \in Key_{\{ka\}} \wedge \\ \ \langle v,\theta,m,t\rangle \in \mathcal{A} \ \wedge \ \exists \langle v',\theta,m',t'\rangle \in \mathcal{A}, \ s.t. \ t' < t \\ 0, \ otherwise \end{cases}$ (9) where $Key_{\{ka\}}$ is the predefined set of knowledge acquisition keywords.
		Topics with Knowledge Provision *(TKP)*: Number of topics including *kpk*. It has the same *RKP* pattern, but with other magnitude [2]. $TKP_{(v)} = \sum_\theta n(\langle v,\theta,m,t\rangle)$, with $n(\langle v,\theta,m,t\rangle) = \begin{cases} 1, if \ \exists w \in m \wedge w \in Key_{\{kp\}} \wedge \ \langle v,\theta,m,t\rangle \in \mathcal{A} \wedge \\ \qquad \nexists \langle v',\theta,m',t'\rangle \in \mathcal{A}, \ s.t. \ t' < t \\ 0, \ otherwise \end{cases}$ (10)
		Topics with Knowledge Acquisition *(TKA)*: Number of topics including *kak*. It has the same *RKA* pattern, but with other magnitude [2]. $TKA_{(v)} = \sum_\theta o(\langle v,\theta,m,t\rangle)$, with $o(\langle v,\theta,m,t\rangle) = \begin{cases} 1, if \ \exists w \in m \wedge w \in Key_{\{ka\}} \wedge \ \langle v,\theta,m,t\rangle \in \mathcal{A} \wedge \\ \qquad \nexists \langle v',\theta,m',t'\rangle \in \mathcal{A}, \ s.t. \ t' < t \\ 0, \ otherwise \end{cases}$ (11)
	Cybercriminal Assets	**Attachments *(ATH)*:** Search for attachments in the posts (topics and replies). According to [4], key-hackers provide relevant cybercriminal assets in form of attachments. $ATH_{(v)} = \sum_\theta \sum_m \sum_t p(\langle v,\theta,m,t\rangle)$, with $p(\langle v,\theta,m,t\rangle) = \begin{cases} 1, if \ \exists w \in m \wedge w \in Key_{\{at\}} \wedge \ \langle v,\theta,m,t\rangle \in \mathcal{A} \\ 0, \ otherwise \end{cases}$ (12) where $Key_{\{at\}}$ is the predefined set of attachment keywords.
	Specialty Lexicons	**Technical Jargon *(TEJ)*:** Number of posts containing technical keywords. According to [4], key-hackers often use technical jargon to reference hacking techniques/tools. $TEJ_{(v)} = \sum_\theta \sum_m \sum_t q(\langle v,\theta,m,t\rangle)$, with $q(\langle v,\theta,m,t\rangle) = \begin{cases} 1, if \ \exists w \in m \wedge w \in Key_{\{ja\}} \wedge \ \langle v,\theta,m,t\rangle \in \mathcal{A} \\ 0, \ otherwise \end{cases}$ (13) where $Key_{\{ja\}}$ is the predefined set of technical jargon keywords.
		Darkweb Jargon *(DWJ)*: Number of posts containing darkweb keywords. Based on to [4], key-hackers often use these jargons to reference hacking environments on darkweb. $DWJ_{(v)} = \sum_\theta \sum_m \sum_t r(\langle v,\theta,m,t\rangle)$, with $r(\langle v,\theta,m,t\rangle) = \begin{cases} 1, if \ \exists w \in m \wedge w \in Key_{\{dw\}} \wedge \ \langle v,\theta,m,t\rangle \in \mathcal{A} \\ 0, \ otherwise \end{cases}$ (14) where $Key_{\{dw\}}$ is the predefined set of darkweb keywords.
Behavioral Trend		**Velocity of Knowledge Provision Topics *(VPT)*:** Verifies how fast the knowledge provision pattern increases or decreases in the topics. According to [2], key-hackers present increasing knowledge provision and this feature measures its velocity in the topics. For every sequential 10 topics, we check the presence of *kpk* (totalized in y_i) to create a corresponding data point x_i. Then, we analyze all points created using linear regression, checking the slope a of the line generated ($y = ax$). For the next 3 features, we will use the same strategy. $VPT_{(v)} = \frac{y_2(v)-y_1(v)}{x_2(v)-x_1(v)} = \frac{y_2(v)-y_1(v)}{2-1} = y_2(v) - y_1(v) = a$, with $y_i(v) = \sum_{Top_{(v)}=[(i-1)*10]+1}^{[(i-1)*10]+10} n(\langle v,\theta,m,t\rangle)$ (15) where $Top_{(v)}$ is the set of sorted topics created by v and n is defined in equation (10).
		Velocity of Knowledge Acquisition Topics *(VAT)*: Verifies how fast the knowledge acquisition pattern increases or decreases in the topics. According to [2], key-hackers present decreasing knowledge acquisition and this feature measures its velocity in the topics. $VAT_{(v)} = \frac{y_2(v)-y_1(v)}{x_2(v)-x_1(v)} = \frac{y_2(v)-y_1(v)}{2-1} = y_2(v) - y_1(v) = a$, with $y_i(v) = \sum_{Top_{(v)}=[(i-1)*10]+1}^{[(i-1)*10]+10} o(\langle v,\theta,m,t\rangle)$ (16) where $Top_{(v)}$ is the set of sorted topics created by v and o is defined in equation (11).
		Velocity of Knowledge Provision Replies *(VPR)*: Verifies how fast the knowledge provision pattern increases or decreases in the replies. According to [2], key-hackers present increasing knowledge provision and this feature measures its velocity in the replies. $VPR_{(v)} = \frac{y_2(v)-y_1(v)}{x_2(v)-x_1(v)} = \frac{y_2(v)-y_1(v)}{2-1} = y_2(v) - y_1(v) = a$, with $y_i(v) = \sum_{Rep_{(v)}=[(i-1)*10]+1}^{[(i-1)*10]+10} p(\langle v,\theta,m,t\rangle)$ (17) where $Rep_{(v)}$ is the set of sorted replies created by v and p is defined in equation (12).
		Velocity of Knowledge Acquisition Replies *(VAR)*: Verifies how fast the knowledge acquisition pattern increases or decreases in the replies. According to [2], key-hackers present decreasing knowledge acquisition and this feature measures its velocity in the replies. $VAR_{(v)} = \frac{y_2(v)-y_1(v)}{x_2(v)-x_1(v)} = \frac{y_2(v)-y_1(v)}{2-1} = y_2(v) - y_1(v) = a$, with $y_i(v) = \sum_{Rep_{(v)}=[(i-1)*10]+1}^{[(i-1)*10]+10} q(\langle v,\theta,m,t\rangle)$ (18) where $Rep_{(v)}$ is the set of sorted replies created by v and q is defined in equation (13).

TABLE IV
SOCIAL NETWORK ANALYSIS FEATURES

Categories/Sub-Categories	Features
Structural Position	**Degree Centrality *(DEC)*:** Analyzes the number of direct neighbors connected to a given node [18]. Here, we define *DEC* using the out-going edges. According to [7], key-hackers present high degree centralities, since each of their many replies produces an outgoing edge. $DEC_{(v_i)} = d_i^{out}$, where d_i^{out} is the out-degree of v_i. (19)
	Betweenness Centrality *(BEC)*: Analyzes the number of shortest paths that pass through a given node [18], indicating its importance for the information flow. According to [6], key-hackers present high betweenness centralities, since they often appear in those shortest paths. $BEC_{(v_i)} = \sum_{s \neq t \neq v_i} \frac{\sigma_{st}(v_i)}{\sigma_{st}}$ (20) where σ_{st} is the number of shortest paths from node s to t and $\sigma_{st}(v_i)$ is the number of shortest paths from node s to t that pass through v_i.
	Closeness Centrality *(CLC)*: Analyzes how an individual is near all other individuals in the networks [18]. As key-hackers have central positions in the networks, which it is a fact that shrinks their distance to the others, these individuals present high closeness centralities. $CLC_{(v_i)} = \frac{1}{\bar{l}_{v_i}}$, where $\bar{l}_{v_i} = \frac{1}{n-1} \sum_{v_j \neq v_i} l_{i,j}$ is node v_i's average shortest path length to other nodes and n is the number of nodes. (21)
Influence	**Eigenvector Centrality *(EIC)*:** Assigns importance to a node if other important nodes are linked to it [18]. According to [8], key-hackers usually present high eigenvector centralities, since they form connections among themselves. $EIC_{(v_i)} = \frac{1}{\lambda} \sum_{j=1}^{n} A_{j,i} EIC(v_j)$ (22) where λ is a fixed constant, $A_{j,i}$ is the adjacent matrix of the directed graph and n is the number of nodes.
	Page Rank *(PAR)*: Analyzes the number and quality of links of a given hacker in order to estimate its importance [18]. According to [8], key-hackers present high page rank, since they are likely to receive more links from other key-hackers. $PAR_{(v_i)} = \alpha \sum_{j=1}^{n} A_{j,i} \frac{PAR(v_j)}{d_j^{out}} + \beta$ (23) where λ, α, and β are fixed constants, $A_{j,i}$ is the adjacent matrix of the directed graph and n is the number of nodes.

TABLE V
SENIORITY ANALYSIS FEATURES

Categories/Sub-Categories	Features
Coverage	**Interval btw User's & Forum's First Posts *(IFP)*:** Checks the difference between the first post date of a forum and of a given hacker in this forum. Previous works [19], [4] argue that founding members are usually key-hackers, being in the discussions since the beginning. $IFP(v) = \sum_{\theta} \sum_{m} \sum_{t} s(\langle v, \theta, m, t \rangle) - \sum_{v'} \sum_{\theta'} \sum_{m'} \sum_{t'} u(\langle v', \theta', m', t' \rangle)$, with $s(.)$ and $u(.)$ defined as: (24) $s(\langle v, \theta, m, t \rangle) = \begin{cases} t, \ if \ \langle v, \theta, m, t \rangle \in \mathcal{A} \wedge \nexists \langle v, \theta'', m'', t'' \rangle \in \mathcal{A}, \ s.t. \ t'' < t \\ 0, \ otherwise \end{cases}$ $u(\langle v', \theta', m', t' \rangle) = \begin{cases} t, \ if \ \langle v', \theta', m', t' \rangle \in \mathcal{A} \wedge \nexists \langle v'', \theta'', m'', t'' \rangle \in \mathcal{A} \ , \ s.t. \ t'' < t' \\ 0, \ otherwise \end{cases}$
	Distinct Days of Postings *(DDP)*: Checks the continuity of posts created by hackers. Previous works [19], [2] argue that continuous participants are more likely to be key-hackers, since they are often contributing to the communities. $DDP_{(v)} = \lvert Days_{(v)} \rvert$, where $Days_{(v)}$ returns the set of distinct days in which v posted. (25)
	Interval btw User's & Forum's Last Posts *(ILP)*: Checks the difference between the last post date of a forum and of a given hacker in this forum. Previous works [19], [4] argue that long-term community members are usually key-hackers, with no fleeting interests. $ILP(v) = \sum_{\theta} \sum_{m} \sum_{t} x(\langle v, \theta, m, t \rangle) - \sum_{v'} \sum_{\theta'} \sum_{m'} \sum_{t'} y(\langle v', \theta', m', t' \rangle)$, with $x(.)$ and $y(.)$ defined as: (26) $x(\langle v, \theta, m, t \rangle) = \begin{cases} t, \ if \ \langle v, \theta, m, t \rangle \in \mathcal{A} \wedge \nexists \langle v, \theta', m', t' \rangle \in \mathcal{A}, \ s.t. \ t' > t \\ 0, \ otherwise \end{cases}$ $y(\langle v', \theta', m', t' \rangle) = \begin{cases} t, \ if \ \langle v', \theta', m', t' \rangle \in \mathcal{A} \wedge \nexists \langle v'', \theta'', m'', t'' \rangle \in \mathcal{A} \ , \ s.t. \ t'' > t' \\ 0, \ otherwise \end{cases}$

IV. SUPERVISED LEARNING EXPERIMENTS

This section presents our supervised learning experiments and their results for the key-hacker identification problem. We first introduce how we perform training and testing using four different algorithms, including an optimization metaheuristic and three machine learning methods. We want to verify which algorithms generalize better, using the hackers' characteristics learned in one forum to test on another one. Then, we compare the performance of our model under two conditions: 1) when it is trained/tested using the features related to each approach individually and combined; 2) when it is trained/tested using different ranges for the definition of key-hackers.

A. Training and Testing

Our 25 features are used here for supervised learning algorithms to mine characteristics of key-hackers. We compare the algorithms when they use features of content, social network and seniority analysis individually and combined. To perform that, we leverage the user reputation score to maximize the overlap of two distinct sets of hackers: the $Top_{10\%}$ found with their *Estimated Reputation Score (ERS)* and the $Top_{10\%}$ found with their *Actual Reputation Score (ARS)*.

The *ARS* represents the user's reputation informed by the forums, and we want to use this information as our ground truth. This way, we sort the users according to their reputation in descending order, and the $Top_{10\%}$ will represent the key-hackers - for instance, the $Top_{10\%}$ of *Forum 1* are the 431 hackers with highest *ARS*. The *ERS* is the reputation to be estimated by our algorithms based on the features extracted, and we also want to use the $Top_{10\%}$ to infer who are the key-hackers. With these both metrics in hands, our goal is to maximize the value of $Overlap_{10\%}$ presented in equation (27), which provides a measure for user ranking consistency [20].

$$Overlap_{10\%} = \frac{|ERS_{10\%} \cap ARS_{10\%}|}{|ERS_{10\%} \cup ARS_{10\%}|} \quad (27)$$

We train and test four supervised learning algorithms that use different approaches. The first one comprises an optimization metaheuristic inspired by the natural selection process: Genetic Algorithms (GA) [21]. In the training phase, we use this algorithm to perform a linear combination of our 25 features, calibrating the *ERS*'s feature weights in equation (28) so that $Overlap_{10\%}$ is maximized. As this approach relies on genetic operators such as *selection*, *crossover* and *mutation* (we apply the elitist, two-points and order-changing methods respectively) to produce high-quality solutions to optimization problems [21], we expect it can search through a huge combination of feature weights to find the ones

generating the highest value for $Overlap_{10\%}$. Then, we use the calibrated linear system trained on a particular forum to test its performance on a different one, also using the $Overlap_{10\%}$.

$$ERS(v) = \sum_{i=1}^{n} w_i * v_{x_i} \quad (28)$$

where w_i is the weight of feature i, v_{x_i} is the value of the feature i for user v, and n is the set of considered features.

The second algorithm uses a multiple Linear Regression approach (LR) [17]. In the training phase, we want to model the relationship between our scalar dependent variable (reputation) and our 25 independent variables (features). This relationship is modeled using linear predictor functions, whose parameters are estimated from the data to fit a curve that produces the highest value for $Overlap_{10\%}$. Then, we want to use this curve fitted to a particular forum to test its performance on a different one, also using the $Overlap_{10\%}$. Note that the correct order of hackers based on reputation is not required here, only the presence of the correct individuals in the $Top_{10\%}$.

The next two algorithms comprise classifiers: Random Forests (RF) and Support Vector Machines (SVM) [17]. Here, we define a binary classification problem to identify the individuals belonging to the positive class (key-hackers). Random Forests are an ensemble method that use multiple decision trees for training, outputting the class that is the mode of the classes. Support Vector Machines (SVM) are a discriminative classifier formally defined by a separating hyperplane that gives the largest minimum distance to the training examples. For both classifiers in the training phase, we want to learn the feature values of the $Top_{10\%}$ hackers of a given forum, in order to apply this knowledge to another forum (testing phase), maximizing the value of $Overlap_{10\%}$.

B. Results

Figure 4 presents the performance of the algorithms when they are trained using *Forum 1* and tested using *Forum 2* and *Forum 3*. We detail the performances when the algorithms learn only the features of individual approaches, and when they learn all the features combined (hybrid approach).

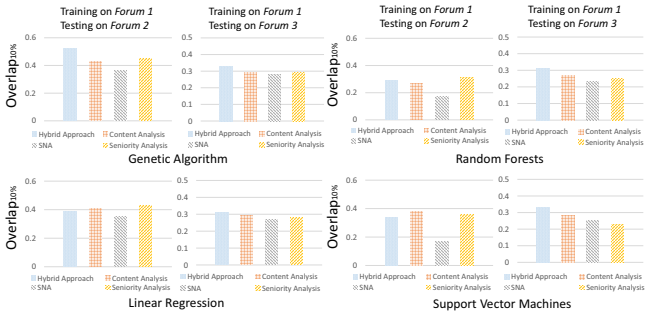

Fig. 4. $Overlap_{10\%}$ performance when algorithms are trained on *Forum 1* and tested on *Forum 2* and on *Forum 3*.

We observe 5 cases when the hybrid approach obtains the best performance, while 1 and 2 cases are verified for content and seniority analysis respectively. The highest value for $Overlap_{10\%}$ when testing on *Forum 2* (0.52) is achieved by Genetic Algorithms using the hybrid approach. This result implies that more than half of the $Top_{10\%}$ hackers were identified, which for this forum represents around 121 users. Also, the highest value for $Overlap_{10\%}$ when testing on *Forum 3* (0.33) is achieved by using Genetic Algorithms and SVM using the hybrid approach. This result implies that more than one third of the $Top_{10\%}$ hackers were identified, which for this forum represents around 92 users. Note these performances correspond to find only 10% of the hackers (those with the highest reputation), which represents a strict filter of users.

Figure 5 presents the performance of the algorithms when they are trained using *Forum 2* and tested using *Forum 1* and *Forum 3*. We observe 5 cases when the hybrid approach obtains the best performance, while 1 case is verified for each one of the other three approaches. The highest value for $Overlap_{10\%}$ when testing on *Forum 1* (0.43) is achieved by Genetic Algorithms using the hybrid approach. This result implies that almost half of the $Top_{10\%}$ hackers were identified, which for this forum represents around 186 users. In addition, the highest value for $Overlap_{10\%}$ when testing on *Forum 3* (0.32) is achieved by Genetic Algorithms and Random Forests using the hybrid approach. This result implies that almost one third of the $Top_{10\%}$ hackers were identified, which for this forum represents around 88 users.

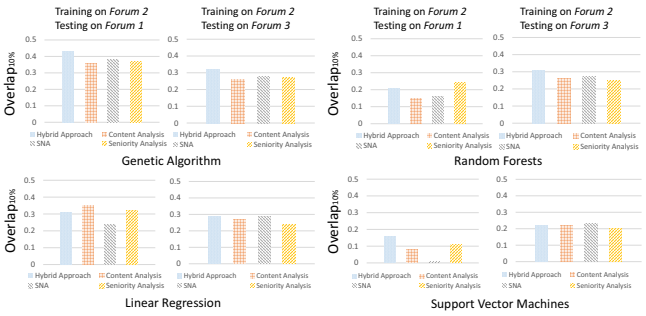

Fig. 5. $Overlap_{10\%}$ performance when algorithms are trained on *Forum 2* and tested on *Forum 1* and on *Forum 3*.

Figure 6 presents the algorithms' performance when they are trained using *Forum 3* and tested using *Forum 1* and *Forum 2*. We observe there are 5 cases when the hybrid approach obtains the best performance, while 1,1 and 2 cases are verified for content, seniority (tied with the hybrid approach) and social network analysis respectively. The highest value for $Overlap_{10\%}$ when testing on *Forum 1* (0.45) is achieved by Genetic Algorithms using the hybrid approach. This result implies that almost half of the $Top_{10\%}$ hackers were identified, which for this forum represents around 194 users. Also, the highest value for $Overlap_{10\%}$ when testing on *Forum 2* (0.5) is achieved by Genetic Algorithms using the hybrid approach. This result implies that half of the $Top_{10\%}$ hackers were identified, which for this forum represents around 121 users.

The overall results clearly show that the hybrid approach is preferable comparing to the individual ones, specially if used by Genetic Algorithms. In general, these generalization results are satisfying, since we are able to retrieve a considerable part of the key-hackers in the different situations analyzed.

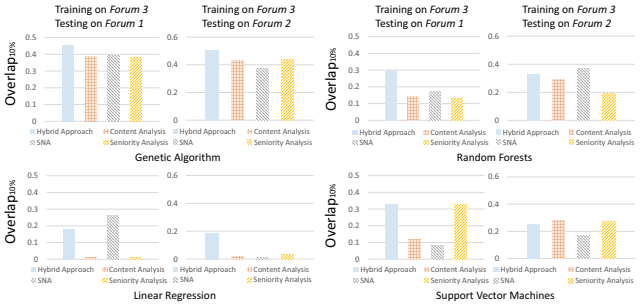

Fig. 6. $Overlap_{10\%}$ performance when algorithms are trained on *Forum 3* and tested on *Forum 1* and on *Forum 2*.

Varying the Overlap. Here, in order to observe how the results change according to the fraction of users that represent the key-hackers, we decide to try different values for $Overlap_{X\%}$. The idea is to cover different zones for the identification of key-hackers, including $Overlap_{1\%}$, $Overlap_{5\%}$, $Overlap_{10\%}$ and also what we denominate $Overlap_{10}$, which means only the top 10 hackers in the forums. Figure 7 presents the performance of the four algorithms with the exact same previous setting, except that we only consider the hybrid approach now. We also include a random key-hacker identification approach for comparison purposes.

The highlighted cells correspond to the highest values of $Overlap_{X\%}$ computed among all the algorithms analyzed. We observe that Genetic Algorithms have the best performances in 87.5% of the cases. The reason for this behavior is a peculiarity of this optimization metaheuristic for our specific problem. Genetic Algorithms are able to work under very strict search conditions, for instance, when the number of individuals to be filtered is considerably low and the search space is very large.

Finally, we show in Figure 8 the curves of $Overlap_{X\%}$ for all implemented algorithms, comparing the performances according to the forums used to train and test. As verified, Genetic Algorithms produce a superior fit. We believe the characteristic of this strategy - population of candidates searching in multiple directions simultaneously - helps it to avoid potential local optima and provide more adapted solutions. This condition makes this metaheuristic suitable to the key-hacker identification problem, leading opportunities for evolutionary algorithms be more considered in cybersecurity research.

V. Related Work

Different works have addressed the key-hacker identification problem in the last years. For instance, Abbasi et al. [4] proposed a framework to identify expert hackers in web forums, based on content-mining. First, the authors represented each user with three categories of features: cybercriminal assets, specialty lexicons, and forum involvement. Then, they profiled the users into four groups based on their specialties: black market activists, founding members, technical enthusiasts, and average users. Analyzing the interactions among hackers, they noted that average users (86% of the total) were participants that did not engage in the community enough, being the other three groups constituted by key-hackers.

Later, Zhang et al. [2] also used a content-mining approach in a hacker forum to analyze post orientations regarding knowledge transfer. Knowledge acquisition and knowledge provision were noted as the patterns to construct user profiles, classified by the authors into four ordinal types: guru, casual, learning, and novice hackers. They found that guru hackers act as key knowledgeable and respectable members in the communities, increasingly acting as knowledge providers.

In a sequence, Fang et al. [5] developed a framework with a set of topic models for extracting popular topics, tracking topic evolution and identifying key-hackers with their specialties. Using Latent Dirichlet Allocation (LDA), Dynamic Topic Model (DTM) and Author Topic Model (ATM), they identified five major popular topics, trends related to new communication channels, and key-hackers in each expertize area.

Using a different approach, Seebruck proposed a weighted arc circumplex model to capture the motivations of hackers with different level of expertise [22]. The author created a hacker typology based on 5 motivations: recreation, prestige, revenge, profit and ideology, and also based on 8 levels of expertise: novices, crowdsourcers, punks, hacktivists, insiders, criminals, coders, and cyber warriors. Then, the model should determine (as no real experiments were performed) the likelihood of an organization be targeted by a certain type of hacker.

Another distinct approach was used by Samtani and Chen [7]. They performed social network analysis to identify key-hackers in hacker forums. The authors analyzed the interactions between users by leveraging metrics such as network diameter and average path length, and found the importance of each user to the community using centrality measures.

Additionally, Zhang and Li performed survival analysis using Cox proportional hazard regression model to examine what makes a user a high-reputation hacker in online forums [6]. According to the results, users should reply detailed posts and also broaden their interests in multiple topics.

In all these works, we noted the authors did not fully explore a hybrid model to find key-hackers online, considering the

Train	Test	Genetic Algorithms				Linear Regression				Random Forests				SVM				Random			
		Top 10	Top 1%	Top 5%	Top 10%	Top 10	Top 1%	Top 5%	Top 10%	Top 10	Top 1%	Top 5%	Top 10%	Top 10	Top 1%	Top 5%	Top 10%	Top 10	Top 1%	Top 5%	Top 10%
Forum 1	Forum2	0.33	0.39	0.49	0.52	0.33	0.31	0.34	0.36	0.11	0.31	0.37	0.28	0.17	0.22	0.19	0.26	0.00	0.00	0.00	0.02
	Forum3	0.11	0.20	0.31	0.33	0.11	0.22	0.25	0.31	0.11	0.22	0.20	0.31	0.11	0.20	0.25	0.33	0.00	0.00	0.00	0.01
Forum 2	Forum 1	0.25	0.28	0.37	0.43	0.17	0.24	0.35	0.24	0.11	0.06	0.12	0.19	0.05	0.03	0.05	0.16	0.00	0.00	0.00	0.02
	Forum3	0.05	0.20	0.24	0.32	0.05	0.17	0.24	0.29	0.11	0.17	0.25	0.31	0.11	0.14	0.19	0.22	0.00	0.00	0.00	0.01
Forum 3	Forum 1	0.17	0.22	0.26	0.45	0.17	0.16	0.17	0.12	0.00	0.06	0.18	0.25	0.00	0.03	0.16	0.29	0.00	0.00	0.00	0.00
	Forum2	0.33	0.29	0.44	0.50	0.11	0.05	0.02	0.18	0.00	0.17	0.17	0.29	0.05	0.08	0.34	0.23	0.00	0.00	0.00	0.00

Fig. 7. Analysis of algorithms' performance using different values for $Overlap_{X\%}$.

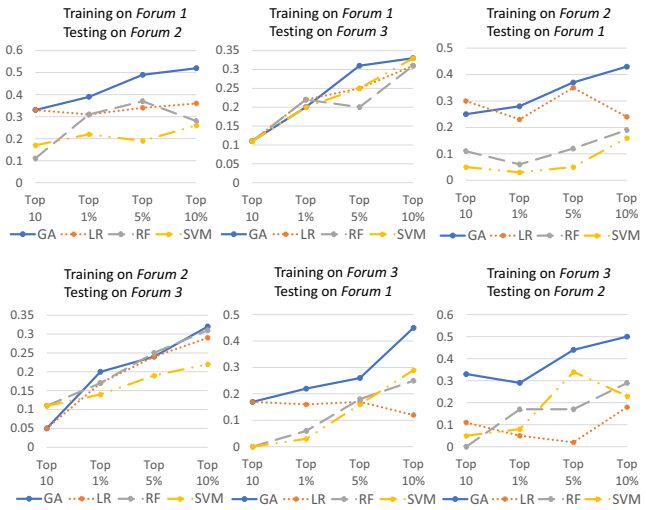

Fig. 8. Comparison between the $Overlap_{X\%}$ curves.

advantages of different approaches combined. Also, there is still a lack of a method to validate the key-hackers identified, which makes the results of these previous works not comparable. Here, we take the next steps to fill these two gaps, proposing a hybrid model to identify key-hackers on darkweb forums and a systematic method to validate the results.

VI. Conclusion

In this work, we address the key-hacker identification problem on darkweb forums using a hybrid approach that combines content, social network and seniority analysis. Extracting 25 carefully selected features, we confirm two hypotheses of great value for cybersecurity. First, we show how different algorithms (specially the Genetic Algorithms) perform better when all the features are used together, highlighting that a hybridization of approaches improve the results. Second, we show our model is able to generalize, learning features in one particular forum that can be applied to another one. This generalization is evaluated by leveraging the user reputation score to systematically cross-validate the key-hackers identified, providing a strategy to find these users in environments that do not offer a reputation system or offer a deficient one.

Although improvements are necessary in this research area, we explore in this work some ideas to pavement the road - including a comparison between genetic and machine learning algorithms when they use our predictive model to identify key-hackers. These insights offer to researchers an alternative strategy to find the emerging threats using a hacker-centric perspective, which can lead to the prediction of cyberattacks.

Acknowledgment

Some of the authors were supported by the Office of Naval Research (ONR) contract N00014-15-1-2742, the Office of Naval Research (ONR) Neptune program, the ASU Global Security Initiative (GSI) and the National Council for Scientific and Technological Development (CNPq-Brazil). Paulo Shakarian and Jana Shakarian are supported by the Office of the Director of National Intelligence (ODNI) and the Intelligence Advanced Research Projects Activity (IARPA) via the Air Force Research Laboratory (AFRL) contract number FA8750-16-C-0112. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, AFRL, or the U.S. Government.

References

[1] S. Goel, "Cyberwarfare: Connecting the dots in cyber intelligence," *Commun. ACM*, vol. 54, no. 8, pp. 132–140, Aug. 2011.

[2] X. Zhang, A. Tsang, W. Yue, and M. Chau, "The classification of hackers by knowledge exchange behaviors," *Inform. Systems Frontiers*, vol. 17, no. 6, pp. 1239–1251, 2015.

[3] M. Motoyama, D. McCoy, K. Levchenko, S. Savage, and G. Voelker, "An analysis of underground forums," in *Proceedings of the ACM SIGCOMM*, ser. IMC '11. ACM, 2011, pp. 71–80.

[4] A. Abbasi, W. Li, V. Benjamin, S. Hu, and H. Chen, "Descriptive analytics: Examining expert hackers in web forums," in *Proceeding of ISI 2014*. IEEE, Sept 2014, pp. 56–63.

[5] Z. Fang, X. Zhao, Q. Wei, G. Chen, Y. Zhang, C. Xing, W. Li, and H. Chen, "Exploring key hackers and cybersecurity threats in chinese hacker communities," in *Proceeding of ISI 2016*. IEEE, 2016.

[6] X. Zhang and L. C., "Survival analysis on hacker forums," *2013 SIGBPS Workshop on Business Processes and Service*, pp. 106–2013, 2013.

[7] S. Samtani and H. Chen, "Using social network analysis to identify key hackers for keylogging tools in hacker forums," in *Proceeding of ISI 2016*, Sept 2016, pp. 319–321.

[8] M. Nouh and J. Nurse, "Identifying key-players in online activist groups on the facebook social network," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, Nov 2015, pp. 969–978.

[9] J. Swarner, "Before WannaCry was Unleashed, Hackers Plotted About it on the Dark Web." 2017, http://slate.me/2xQvscu (accessed on 07-03-2017).

[10] E. Nunes, A. Diab, A. Gunn, E. Marin, V. Mishra, V. Paliath, J. Robertson, J. Shakarian, A. Thart, and P. Shakarian, "Darknet and deepnet mining for proactive cybersecurity threat intelligence," in *Proceeding of ISI 2016*. IEEE, 2016, pp. 7–12.

[11] J. Robertson, A. Diab, E. Marin, E. Nunes, V. Paliath, J. Shakarian, and P. Shakarian, *Darkweb Cyber Threat Intelligence Mining*. Cambridge University Press, 2017.

[12] J. Shakarian, A. T. Gunn, and P. Shakarian, *Exploring Malicious Hacker Forums*. Cham: Springer International Publishing, 2016, pp. 259–282.

[13] D. Décary-Hétu and B. Dupont, "Reputation in a dark network of online criminals," *Global Crime*, vol. 14, no. 2-3, pp. 175–196, 2013.

[14] J. Swearingen, "The Creator of the Mirai Botnet Is Probably a Rutgers Student With the Bad Habit of Bragging." 2017, http://slct.al/2wpr54I (accessed on 08-12-2017).

[15] T. Holt, D. Strumsky, O. Smirnova, and M. Kilger, "Examining the social networks of malware writers and hackers," *International Journal of Cyber Criminology*, vol. 6, no. 1, pp. 891–903, 1 2012.

[16] J. Radianti, "A study of a social behavior inside the online black markets," in *2010 Fourth International Conference on Emerging Security Information, Systems and Technologies*, July 2010, pp. 189–194.

[17] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Addison-Wesley, 2013.

[18] R. Zafarani, M. Abbasi, and H. Liu, *Social Media Mining: An Introduction*. Cambridge University Press, 2014.

[19] V. Benjamin, B. Zhang, J. Jr., and H. Chen, "Examining hacker participation length in cybercriminal internet-relay-chat communities," *Journal of Management Information Systems*, vol. 33, no. 2, pp. 482–510, 2016.

[20] S. P. Borgatti, K. M. Carley, and D. Krackhardt, "On the robustness of centrality measures under conditions of imperfect data," *Social Networks*, vol. 28, no. 2, pp. 124–136, 2006.

[21] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press, 1996.

[22] R. Seebruck, "A typology of hackers: Classifying cyber malfeasance using a weighted arc circumplex model," *Digital Investigation*, vol. 14, pp. 36–45, 2015.