Securing Cyberspace: Identifying Key Actors in Hacker Communities

Victor Benjamin, Hsinchun Chen
Department of Management Information Systems
The University of Arizona
Tucson, AZ 85721, USA
vabenji@email.arizona.edu, hchen@eller.arizona.edu

***Abstract*— As the computer becomes more ubiquitous throughout society, the security of networks and information technologies is a growing concern. Recent research has found hackers making use of social media platforms to form communities where sharing of knowledge and tools that enable cybercriminal activity is common. However, past studies often report only generalized community behaviors and do not scrutinize individual members; in particular, current research has yet to explore the mechanisms in which some hackers become key actors within their communities. Here we explore two major hacker communities from the United States and China in order to identify potential cues for determining key actors. The relationships between various hacker posting behaviors and reputation are observed through the use of ordinary least squares regression. Results suggest that the hackers who contribute to the cognitive advance of their community are generally considered the most reputable and trustworthy among their peers. Conversely, the tenure of hackers and their discussion quality were not significantly correlated with reputation. Results are consistent across both forums, indicating the presence of a common hacker culture that spans multiple geopolitical regions.**

Keywords - Cyber security, Cybercrime, Hacker, Hacker community, Social media, Reputation, Key actor

I. INTRODUCTION

As the computer becomes more ubiquitous throughout society, the security of networks and information technologies is a growing concern. Critical infrastructures (e.g. power grids) are facing growing number of cyber-based threats that could result in service disruptions and physical damage. For example, the recent Stuxnet Worm was engineered to specifically target a country and destroy nuclear infrastructure. Further, keystroke logging software was found covertly spreading at the United States Drone Fleet Command. With growing security concerns, DARPA has emphasized that U.S. cyberspace is unsecure and more cyber security research is vital to national defense.

Advancing technology enables hackers to commit cybercrime much more easily now than in the past. Accessibility to technologies and methods to commit cybercrime has grown. Even legitimate tools, such as search engines, can be used to promote cybercriminal activity [1]. Malicious software, hacking tutorials, and other resources intended to help conduct cybercrime can be commonly found within hacker communities, often available for free or traded within black markets [2].

As a result, security researchers have taken interest in exploring hacker communities. Some research has found that community participants pool together skills and assets to form groups, often in order to accomplish operations much more advanced than any one individual could perform alone [3]. In particular, less skillful hackers often seek help from more experienced individuals, creating a meritocratic hierarchy within hacking culture [4]. Such skilled individuals could thus be considered as key actors within their communities, as they are heavily relied upon by peers. However, many participants of such communities aggressively protect their anonymity in order to minimize criminal evidence [5]. Given the importance of key actors in the formation of cybercriminal groups, as well as their pursuit of total anonymity, they make ideal subjects for further research aiming to secure cyberspace

II. LITERATURE REVIEW

To form the basis for this research, literature is reviewed from the following four areas:

- Past studies on hacker communities and cybercrime
- The significance of reputation within communities
- Control theory and criminal group leadership
- Previous literature on social media analytics

A. Hacker communities, and cybercrime

Hackers often congregate within underground communities, most commonly in the form of Internet Relay Chat (IRC) networks or online forums. Such communities are primarily used to share knowledge and resources [6]. Some communities evolve to include black markets where participants offer expertise, snippets of code, fully-developed applications, or stolen confidential data (e.g. credit card accounts) in exchange for other virtual goods or financial gain [2].

Due to the lucrative nature of cybercrime, some hackers use communities as a platform to organize into groups and launch sophisticated, financially-motivated cyber-attacks. Some past research has looked to develop methods for identifying collaboration between hackers, potentially revealing such groups [7]. Other research has studied group dynamics, revealing that peer approval and reputation are heavily integrated within cybercriminal group operations. This leads to a strong desire to increase reputation among hackers [8]. As the hierarchical chain of command within cybercriminal groups is heavily tied to member reputation, good reputation becomes an extremely valuable attribute hackers desire and pursue.

B. The significance of reputation

Reputation has a strong influence on the dynamics of social groups. Individuals with good reputations are more able to cooperate or receive help from others, as they are more trusted [9]. Their reputation can be affected by various factors in a group setting; for example, seniority or tenure within a social group can often amplify the effects of good reputation [10]. Within the context of criminal groups, trust and reputation between both members and leadership is necessary for a group to remain functional [11].

Establishing a good reputation within a community or group is also a vehicle to obtain leadership. Emergence of leaders is often considered the outcome of a self-organizing process involving reputation, as reputation constitutes a precursor to leadership and affects how much stakeholders trust a leader [12]. Leaders are considered among their social group to be experts and hold great tacit knowledge [13]. They also contribute to cognitive advance of a community, and individuals characterized by high levels of activity are often more reputable and able to obtain leadership positions [14]. Reputation appears to underlie many hacker activities, including the formation of group leadership.

C. Control theory and criminal group leadership

Control theory states that a system is controllable if it can be driven from its initial state to another final state given specific inputs over some amount of time [15]. The general objective of control theory is to calculate how a system's inputs should be manipulated in order to achieve desired output, e.g. a car's velocity can be controlled by how far one pushes a gas pedal. Two different factors are said to contribute towards a system's controllability: the system's overall structure and the rules by which the system's components interact [16]. Thus, systems with different structures or component interactions are controlled in vastly different means.

When control theory is applied to social networks, identification and manipulation of driver nodes, or nodes that are said to control a system, can be used to guide a network's dynamics [17]. When considering cars, the steering wheel, gas pedal, brake pedal, and shifter are all considered driver nodes as they manipulate the remainder of their system. Unfortunately, driver nodes are generally unknown in real systems and methods to identify driver nodes need to be developed for each application context. However, past research has identified that criminal networks may have leaders that command and steer their entire group [18]. Criminal leaders exhibit characteristics that match the operational definition of driver nodes, and thus they may be considered as such. When attempting to identify the driver nodes or key actors of online social networks, such as hacker communities, it becomes necessary to develop a further understanding of methods used in social media analytics.

D. Previous studies on social media analytics

With the advent of Web 2.0, Internet users can publish content through a variety of means. Blogs, Wikis, and web forums are just a few mediums individuals can use to publish information. Each platform contains unique contents and offers opportunities for analyses. For example, in the context of eLearning forums, each student's discussion quality was found to be related to the number of posts they created, the average length of their posts, and the average number of replies they made within a discussion started by a peer [19] In programming communities, user reputation has been found closely tied to the breadth and depth of knowledge an individual possesses; users that can produce thoughtful comments across a variety of topics have been considered as domain experts [20].

User reputation has also traditionally had a role in social media analytics. Reputation is crucial as a means of establishing trust, status, and for fostering social interactions [9]. It is subject to many factors of user behaviors, such as activity levels and discussion quality. For example, past work found that forum users who mostly contributed short messages were unable to successfully maintain communication with peers [21]. Other research has considered the effects of tenure, or length of membership within a community, as an influencer of reputation. Conclusions suggest tenure may amplify reputation [22]. Finally, content quality has been observed as an important contributor towards reputation. For example, attaching or linking to additional resources in a message can positively influence its perceived quality [23]. This is especially important in hacker communities, as reputation is regarded as extremely valuable and often influences social interactions [8].

III. RESEARCH GAPS AND QUESTIONS

Researchers have been interested in exploring hacker communities and cybercriminal groups, but few inquiries have focused on identifying key actors, or hackers considered very reputable and have leadership potential or are already leaders. Law enforcement would benefit by focusing resources on apprehending key actors, as they may be the driver nodes of their respective networks and thus potentially responsible for directing cybercriminal activity. Additionally, there is a lack of work studying mechanisms which individuals to become key actors within hacker environments. Security researchers would benefit from knowledge of the processes hackers use to gain status, power, and leadership within their communities

In this study, we analyze the relationships between aspects of hacker behavior and reputation. Analysis is performed across two diverse hacking communities from the United States and China. We pose the following research questions:

1. How does the behavior of an individual hacker affect how other hackers perceive them?
2. In what ways can one build their reputation within hacking communities?
3. Are hacker communities different across cultures? In what ways are they similar?

IV. RESEARCH TESTBED AND DESIGN

The two communities used for this study are web forums based out of the United States and China. Communities were chosen for their size relative to other hacker communities, and

This work was supported in part by the National Science Foundation under Grant No. CBET-0730908, and by the Defense Threat Reduction Agency, Award No HDTRA1-09-1-0058.

also for the growing relevance of cyber security in both countries. Other hacker communities suitable for this research could possess differing geopolitical origins, such as the Middle-East or Russia. Additionally, such communities may exist within entirely different mediums (e.g. IRC channels). However, web forums are particularly useful to study as they naturally provide access to data spanning multiple years into the past. Both communities we explored in this research also feature unique mechanisms allowing hackers to attach files or source code to their messages, promoting collaboration. We also perform analysis of hacker communities from differing geopolitical regions, as this allows for further observation of an expanding global phenomenon.

TABLE I. RESEARCH TESTBED

Forum Name	Language	# of Messages	# of Users	Forum Start Date
Hackhound.org	English	77,061	5,794	October 9, 2008
Unpack.cn	Chinese	646,494	22,743	October 12, 2004

written to extract relevant content. Text parsers were written to extract messages and user information from both of the communities.

Many previous studies have frequently used regression models to determine individuals' reputations. For example, ordinary least squares (OLS) regression was used to model leadership in business group settings [25]. The reputations and trustworthiness of online multiplayer gamers was modeled using a regression approach [26]. Additionally, to help find quality content on Yahoo! Answers, user reputations have been assessed through a series of regression models using message content features [27].

For this study, we borrow six features from past research and observe them within our hacker community dataset. As previous studies often aimed to observe user forum involvement and user discussion quality within virtual communities, we select related features. Two of our selected features aim to capture hacker discussion quality while the other four features are intended to model the extent of a hacker's involvement in their community.

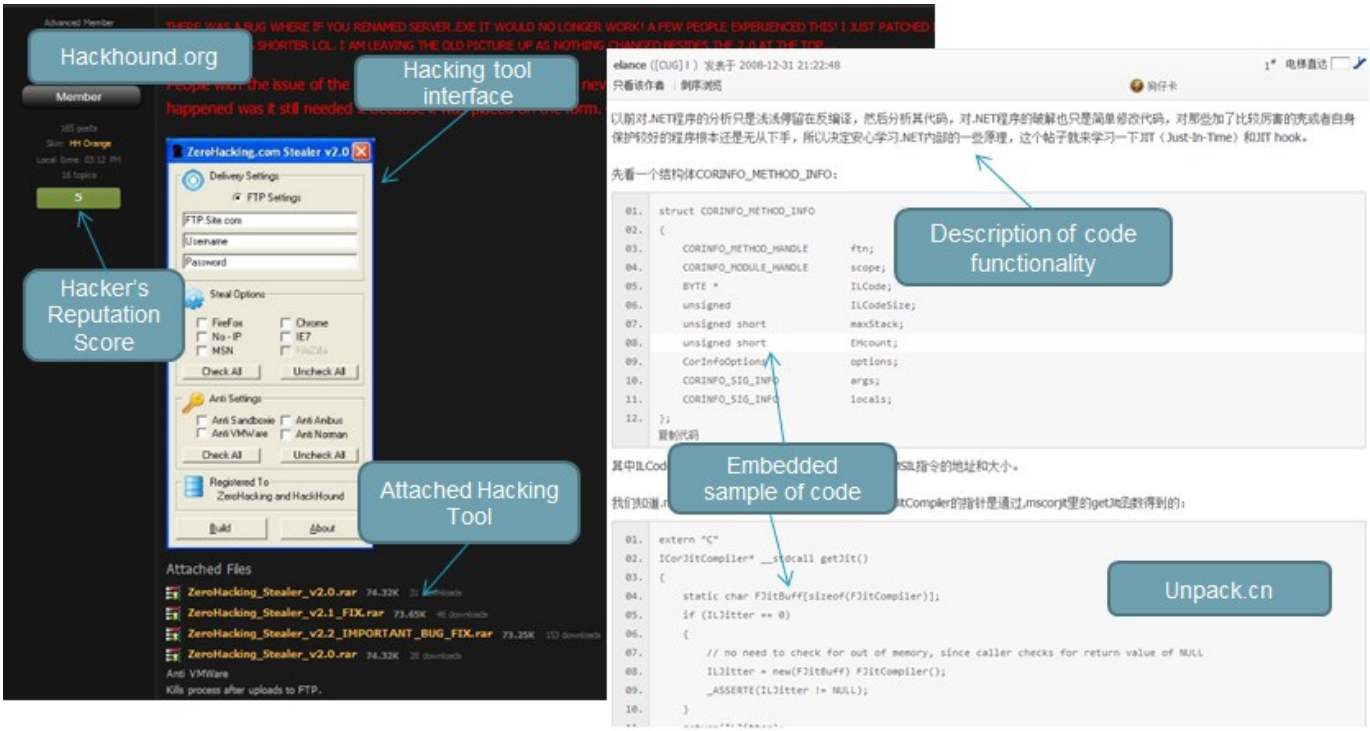

Figure 1. **Left**: A member of Hackhound.org publishes the latest version of his hacking tool meant to help others steal cached passwords on victims' computers **Right:** A hacker of the Chinese community Unpack.cn posts sample code for reverse engineerimg software written in the Microsoft .NET framework

Both communities allow for the unique feature for hackers to attach hacking tools and program source code to their messages for others to use (Figure 1). Additionally, both communities allow hackers to assign each other a reputation score in order to rate one another's usefulness and trustworthiness. Communities were identified through keyword searches similar to past research [24] Keywords such as "malware detection avoidance" were searched on Google.com and Bing.com to find relevant communities. Automated spiders were used to collect the communities and text parsers were

To conduct analysis of what features are related to hacker reputation, we employ an OLS regression similar to previous work. Each forum allows for hackers to assign each other reputation scores through forum mechanisms, resulting in reputation as a method for peers to evaluate each other. We observe the relationships between all selected features and reputation in attempt to reveal why certain hackers emerge as most reputable and trustworthy among their peers. Such hackers often have leadership potential or are already commanding less skilled individuals, and thus can be considered as key actors.

A. *Discussion Quality Features*

Discussion quality features have been widely studied in past research to scrutinize the members of virtual communities. In particular, previous work has studied how user discussion quality can affect social behaviors and interactions. We observe two related features within this study:

1. *Average_Message_Length* – This feature is simply the average length of an individual hacker's messages. Past research has used average message length to determine the amount of content a user generally contributes within a single post, with the assumption that longer posts are generally more valuable to the cognitive advance of a community [21].
2. *Number_Of_Replies_Per_Thread* - The average number of replies a user posts within each discussion participated in. This feature has previously been used in research to measure the amount of dialogue a user contributes to a particular discussion. Users who contribute more dialogue to a discussion are thought to have a greater impact on the cognitive advance of a community [19].

B. *Community Involvement Features*

Along with observing the styles in which users communicate, previous research has also investigated the implications that differing levels of community involvement have on user reputation. Specifically, the breadth of involvement and the type of involvement have typically been considered important features used for analysis in social media analytics. We observe four features related to community involvement in this study:

1. *Number_Of_Threads_Involved* - User expertise is closely related to reputation in online communities; expert users are often involved in many discussions and are able to discuss various concepts and topics [20]. Users who post in many different threads take the initiative to often express their opinions and share their knowledge.
2. *Tenure* - Previous investigation has found that member tenure can amplify their reputation and the trust that others place into them [10].We measure tenure by observing the date that each hacker posted their first message within their community, as it marks their initial participation. To be able to quantify tenure and compare hackers, we use days as our unit of measurement.
3. *Sum_Of_Attachments* - The number of times a hacker attaches a file or code example to their post, a measure of cognitive advancement towards a hacker's community [14].
4. *Total_Messages* - Total number of messages a hacker has posted, relevant as previously highlighted literature states reputation may be a result of higher levels of activity [14].

C. *Regression Model*

$$\begin{aligned} Reputation = {} & \beta_1 Average\_Message\_Length \\ & + \beta_2 Number\_Of\_Replies \\ & + \beta_3 Number\_Of\_Threads\_Involved \\ & + \beta_4 Tenure \\ & + \beta_5 Sum\_Of\_Attachments \\ & + \beta_6 Total\_Messages \\ & + \varepsilon \end{aligned}$$

Regression analysis was performed on both forums. Only users with at least 1 reputation point were included in analysis in order to reduce statistical skewness that would result from members who join a community for only very briefly before becoming inactive. Additionally, in some instances, users were given negative reputations as a result of harmful or offensive postings. For example, one hacker was caught distributing a tool that contained a hidden level of malware used to infect other hackers. These users were excluded to more accurately depict the mechanisms in which reputation is gained. Lastly some community administrators disqualified from analysis as they were witnessed to assign themselves reputation points arbitrarily, which would skew results.

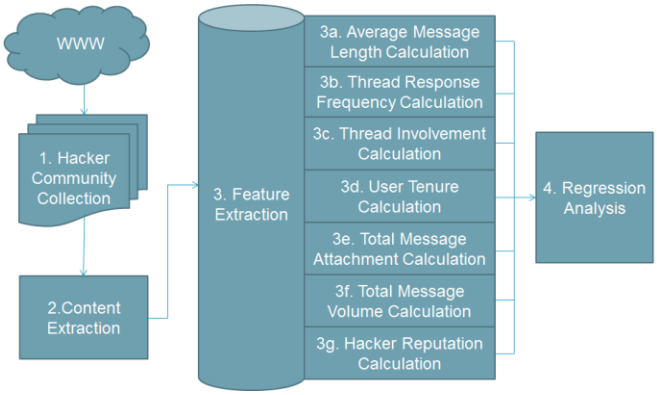

Figure 2. The research design, visualized. Two hacker communities are downloaded from the Internet and their contents extracted into a SQL database. Features identified through literature review are selected and calculated using forum contents. Finally, a regression analysis is run to identify relationships between features and hacker reputation.

Our first set of hypotheses delineates the relationship between a hacker's discussion quality and their reputation. Specifically, we observe how much content a hacker typically includes in a single message, as well as the amount of dialogue a hacker contributes to one discussion. Discussion quality has been previously been shown to affect social behaviors and interactions:

H1: Discussion quality will be a significant contributor towards hacker reputation as it relates to the cognitive advance of a community.

H1a: The average message length of a hacker will be a significant contributor of a hacker's reputation, as past research has highlighted that average message length is related to user discussion quality.

H1b: The number of replies a hacker posts will affect their reputation. Number of replies has previously been used in research to measure the amount of dialogue a user contributes to a particular discussion.

Our second set of hypotheses is related to the relationship between a hacker's community involvement and their reputation among peers. Community involvement encompasses a hacker's activity level and their contributions to the cognitive advance of the community. Both have been previously found to influence user reputation within virtual communities:

H2: Community involvement will be a significant contributor influencer of hacker reputation as it encompasses user activity levels and contributions towards the cognitive advancement of the community.

H2a: The number of threads a hacker is involved in will impact their reputation; previous research has demonstrated that the ability to discuss a wide variety of topics and concepts is related to expertise.

H2b: The seniority of a hacker will be a significant contributor of a hacker's reputation, as previous literature states tenure generally amplifies an individual's expertise and trustworthiness.

H2c: The sum of attachments a hacker posts throughout their community activities will be an influencer of reputation as hackers are contributing to the cognitive advancement of their community.

H2d: Total message volume will be a significant contributor towards hacker reputation, as activity level is a previously observed contributor towards reputation.

Lastly, we expect to observe similar results in both hacking communities. Previous research exploring hacking communities and cybercriminal activity from the United States and China have unveiled similarities between hackers from both countries; for example, observations of hacker communities across multiple geopolitical regions has found black market trading to be universally popular [2,20].

H3: Results of the regression model will be consistent across both the English and Chinese forums, as both forums share aspects of hacker culture.

VI. RESULTS AND DISCUSSION

TABLE II. HACKHOUND.ORG RESULTS

Feature	Estimate	Std. Error	T Value
Average_Message_Length	-0.0083	0.0025	-0.968
Number_Of_Replies_Per_Thread	0.0188	0.0616	0.305
Number_Of_Threads_Involved	0.1689	0.0538	2.822***
Tenure	0.0041	0.0123	0.526
Sum_Of_Attachments	0.2786	0.1437	5.323***
Total_Messages	0.3396	0/0379	6.55 ***

Adjusted R-squared: 83.14%; ***p ≤0.001 ** p ≤0.01 *p ≤0.05)

TABLE III. UNPACK.CN RESULTS

Feature	Estimate	Std. Error	T Value
Average_Message_Length	0.0052	0.0027	0.125
Number_Of_Replies_Per_Thread	0.0372	0.0040	0.528
Number_Of_Threads_Involved	0.1403	0.0033	1.914*
Tenure	-0.0086	0.0135	-0.144
Sum_Of_Attachments	0.3805	0.1991	4.757***
Total_Messages	0.2838	0.0252	3.714**

Adjusted R-squared: 57.38 %; ***p ≤0.001 ** p ≤0.01 *p ≤0.05)

Involvement in various threads, the sum of attachments, and total message volume all appear to be significant contributors of reputation, supporting H2a, H2c, and H2d. This demonstrates support for theories tying higher user reputations to individuals who are active and contribute to cognitive advancement of their communities. Correlations are stronger in Hackhound.org, perhaps due to forum age; as Hackhound.org is younger and has a much lighter total message volume than U npack.cn, each message on Hackhound.org holds more potential weight for contributing towards a user's reputation. In hacker communities, knowledge appears to be power in the form of reputation and influence

Average post length, number of replies per thread, and tenure do not appear to be significant contributors towards reputation, thus not supporting H1a, H1b, and H2b. Discussion quality and seniority seem to be mostly irrelevant when considering reputation in the context of hacker communities. Results suggest reputation may be more heavily influenced by the content and diversity or novelty of information posted, rather than how information is delivered or who presents it.

Both communities share similar patterns in regards to how reputation is built by members, supporting H3. Despite obvious cultural differences between the English and Chinese forums, an overriding form of hacker culture appears to be experienced by both communities. A slight variation in R-squared is experienced when performing analyses on these two forums. The analysis of Hackhound.org has a higher R-squared than that of the Unpack.cn forum. Unpack.cn is a much older community and thus may have an accumulation of more variation among data over time. The differences in R-squared may also be related to cultural phenomena not yet identified.

VII. CONCLUSION AND CONTRIBUTIONS

As the computer becomes more ubiquitous within society, the security of networks and information technologies is a growing concern. Critical infrastructures and military interests are at an increasing risk. Tools to commit cybercrime are now more accessible and easy to use than any previous point in the Internet's history. In particular, key actors responsible for leading hacker communities and cybercriminal groups are of interest.

In this research, two major hacker communities from the United States and China are examined to identify the mechanisms in which key actors arise. These communities were selected due to unique characteristics of allowing the attachment of tools and code to messages, as well as a peer-

evaluated reputation system. Relationships between message volume, message lengths, community seniority, the inclusion of tools or code in messages, and reputation were studied. Hackers that contributed to cognitive advance of their community or were considerably active had the highest reputations.

Our experiment contributes to research in social media analytics, as results revealed correlations between user reputation and various user behaviors and social media features within web forums. Specifically, we confirm previous findings features reported to interact with increasing reputation. This study was unique from other social media research because it investigated the mechanisms built into hacker communities that allow users to share code and files as part of their message postings.

Further, this research contributes to advancing our understanding of hacker communities. Security researchers would benefit from knowledge of the processes hackers use to gain status, power, and leadership within their communities. Law enforcement would benefit by more efficiently focusing resources on cybercriminal pursuits. If key actors contributing towards the cognitive advance of their communities or groups are apprehended, novice hackers relying on such individuals could not as easily develop malicious capabilities. Lastly, the results of this study can also be used to identify key actors in other hacker forums where reputation-ranking systems are not provided.

This research can be expanded by additional investigation of hacker communities. Analytical tools developed from control theory can be used to provide additional scrutiny when attempting to identify driver nodes within communities of different types and sizes [17]. These tools are important when attempting to study hacker communities with unknown origins and context. Further, the perspectives of other social network theories can be used to extend current research on hacker communities, potentially revealing interesting phenomena. Deeper exploration is necessary to more fully understand hacker behaviors and methodologies.

REFERENCES

[1] T. Moore and R. Clayton, "Evil searching: compromise and recompromise of Internet hosts for phishing," in Financial Cryptography and Data Security, pp. 256-272, 2009.

[2] J. Radianti and J.J. Gonzalez, "A preliminary model of the vulnerability black market," Society, 2007.

[3] R. McCusker, "Transnational organised cyber crime: distinguishing threat from reality," in Crime, Law and Social Change, vol. 46, no. 4-5, pp. 257-273, December 2006.

[4] N. Kshetri, "The Simple Economics of Cybercrime," in IEEE Security & Privacy, vol. 4, no.1, pp. 33-39, January-February 2006.

[5] T. Holz, M. Engelberth, and F. Freiling, "Learning more about the underground economy: a case study of keyloggers and dropzones," in European Symposium on Research in Computer Security (ESORICS), 2009.

[6] D. Wall, "Cybercrime: the transformation of crime in the information age," Polity, September 2007.

[7] C. J. Mielke and H. Chen, "Botnet and the cybercriminal underground," in IEEE International Conference on Intelligence and Security Informatics 2008, pp. 206-211. June 2008.

[8] S. Gosh and E. Turrini, "Cybercrimes: a multidisciplinary analysis," Springer, October 2010.

[9] D. Semmans, H. Krambeck, and M. Milinksi, "Reputation is valuable within and outside one's own social group," in Behavioral Ecology and Sociobiology, vol.57, no. 6, April 2005.

[10] R. Lines, M. Selart, B. Espedal, S. Johansen, and T. Svein, "The production of trust during organizational change," in Journal of Change Management, vol. 5, no. 2, February 2005.

[11] K. V. Lampe and P. O. Johansen, "Organized crime and trust: on the conceptualization and empirical relevance of trust in the context of criminal networks," in Global Crime, vol. 6. no. 2, 2004.

[12] A. T. Hall, F. R. Blass, G. R. Ferris, and R. Massengale, "Leader reputation and accountability in organizations: implications for dysfunctional leader behavior," in The Leadership Quarterly, vol. 15, no. 4, August, 2004.

[13] S. J. Zaccaro, C. Kemp, and P. Bader, "Leader traits and attributes," in The Nature of Leadership, pp. 101-123, 2004.

[14] P. Muller, "Reputation, trust and the dynamics of leadership in communities of practice," in Journal of Management and Governance. Vol. 10, no. 4, November, 2006.

[15] R. E. Kalman, "Mathematical description of linear dynamical systems," in Society for Industrial and Applied Mathematics, 1963.

[16] D. G. Luenberger, "Introduction to dynamic systems: theory, models, & applications," Wiley, May 1979.

[17] Y.Y. Liu, J.J. Slotine, and A.L. Barabási, "Controllability of complex networks," in Nature, no. 473, May 2011.

[18] J. Xu, and H. Chen, "Criminal network analysis and visualization: a data mining perspective," in Communications of the ACM, vol. 48, no. 6, 2005.

[19] J. Kim, E. Shaw, D. Feng, C. Beal, and E. Hovy, "Modeling and assessing student activities in on-line discussions," in Proceedings of the Workshop on Educational Data Mining, 2006.

[20] J. Zhang, M. S. Ackerman, and L. A. Adamic, "Expertise networks in online communities: structure and algorithms," in World Wide Web Conference., May 2007.

[21] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman, "Knowledge sharing and Yahoo Answers: everyone knows something," in World Wide Conference, April 2008.

[22] O. Nov, M. Naaman, and C. Ye, "Motivational, structural and tenure factors that impact online community photo sharing," in Proceedings of AAAI International Conference on Weblogs and Social Media, May 2009.

[23] K. Chai, C. Wu, V. Potdar, and P. Hayati, "Automatically measuring the quality of user generated content in forums," in Proceedings of Australasian Conference on Artificial Intelligence, pp. 51-60, 2011.

[24] T. J. Holt, E. Lampke, "Exploring stolen data markets online: products and market forces," in Criminal Justice Studies, vol. 23, no. 1, March 2010.

[25] A. Mehra, A. L. Dixon, D. J. Brass, and B. Robertson, "The social network ties of group leaders; implications for group performance and leader reputation," in Organization Science, vol. 17, no. 1, February 2006.

[26] G.Y. Huang, S.Y. Hu, and J.R. Jiang, "Scalable reputation management with trustworthy user selection for P2P MMOGs," in International Journal of Advanced Media and Communication, vol. 2, no. 4, 2008.

[27] J. Bian, Y. Liu, D. Zhou, E. Agichtein, and H. Zha, "Learning to recognize reliable users and content in social media with coupled mutual reinforcement," in World Wide Web Conference, May 2009.