Examining the Social Networks of Malware Writers and Hackers

Thomas J. Holt
Michigan State University, USA

Deborah Strumsky
University of North Carolina at Charlotte, USA

Olga Smirnova
East Carolina University, USA

Max Kilger
The Honeynet Project, USA

Abstract
A substantive body of research has emerged exploring the social dynamics and subculture of computer hacking. Few, however, have considered the structure of social networks in the hacker community due in part to the lack of visible information about active hackers or malware writers. Our research focuses on the rarely studied subject of underground networks of computer hackers. Thus, this study explores the social networks of a group of Russian hackers using publicly accessible data to understand the nature of social relationships and the ways that they affect information sharing and action. The findings demonstrate that there are a limited number of highly skilled hackers relative to those with some knowledge of computers. Additionally, those hackers with substantive technical skills are centrally located within friendship networks and are the focus of greater attention overall. The impact of these findings for our understanding of computer hacking, and peer networks generally are considered in detail.

Keywords: Cyber crime, computer hackers, malware, malicious software, network analysis, geography of networks.

1Associate Professor, 434 Baker Hall, School of Criminal Justice, Michigan State University, Lansing, MI 48824, USA. Email: holtt@msu.edu
2Department of Geography and Earth Sciences, University of North Carolina at Charlotte, University City Boulevard Charlotte, NC 28223, USA. Email: dstrumsk@uncc.edu
3Department of Political Science, East Carolina University, East 5th Street Greenville, NC 27858, USA. Email: smirnovao@ecu.edu
4Spartan Devils Honeynet Chapter, The Honeynet Project, 1425 Broadway #438, Seattle, WA, 98122-03854, USA. Email: maxk@smrb.com
Introduction

The proliferation and societal dependence on computer technology and the Internet over the last three decades has dramatically increased the risks posed by computer hackers and malicious actors online (Bachmann, 2010; Holt, 2007; Holt & Graves, 2007; Newman & Clarke, 2003; Schell & Dodge, 2002; Wall, 2001, 2007). Though there is no single definition for hacker activity, many recognize hackers as individuals with a profound interest in computers and technology who use their knowledge to access computer systems with or without permission from the system owner (Holt, 2007; Schell & Dodge, 2002). The issue of unauthorized access is particularly costly for both individuals and corporations alike, as electronic databases of sensitive information can be accessed and compromised quickly and efficiently steal massive amounts of information (Newman & Clarke, 2003; Peretti, 2009; Wall, 2007). In fact, the number of high profile data breaches has increased significantly, with millions and in some cases billions of customer records stolen by small groups of hackers (Chu, Holt, & Ahn, 2010; Holt & Lampke, 2010; Peretti, 2009). Additionally, hackers utilize malicious software, or malware, to facilitate their attacks which can cause hundreds of thousands of dollars in damage across corporate settings (Computer Security Institute, 2012).

A small body of research has explored the subculture and norms of computer hackers and malware writers, finding that hackers value learning on their own, but share information with others about attacks and successful practices through on-line and off-line networks (Gordon & Ma, 2003; Holt, 2007; Holt, Soles, & Leslie, 2008; Jordan & Taylor, 1998; Taylor, 1999; Thomas, 2002). The hacker community is also a meritocracy where individuals are judged based on their skill and ability (Holt, 2007; Jordan & Taylor, 1998; Kilger et al., 2004). It is thought that there are a small proportion of highly skilled actors with the capacity to engage in sophisticated attacks relative to the larger populace of low skilled actors (Holt et al., 2008; Jordan & Taylor, 1998). In addition, a limited number of studies demonstrating that hackers maintain small intimate peer networks but participate in larger collegial networks through web forums and other on-line communications venues (Holt, 2009; Holt & Kilger, 2008; Holt et al., 2008; Meyer, 1989; Schell & Dodge, 2002).

It is unknown, however, how skill and ability impact individual position within hacker social networks, and the general distribution of actors within larger on-line networks. In addition, it is unclear how information may flow within social networks online due to variations in individual capability. This study attempts to address these gaps in the literature by examining the structure of the members of multiple Russian hacker and malware writing groups. The research utilizes multiple analysis techniques to analyze the data obtained from individual social networking profiles posted on-line. The findings consider the demographic backgrounds of participants, their relationships with one another, and techniques to predict the likelihood that an individual will participate in dangerous hacking activity based on the information provided in their networking profiles.

The structure and nature of the Hacker Community

Though there is a common perception in the general public that hackers are loners and anti-social (Furnell, 2002; Schell & Dodge, 2002; Wall, 2007), social science research indicates hackers operate within a collegial subculture where individuals share information with one another (Holt, 2007, 2009; Holt & Kilger, 2008; Jordan & Taylor, 1998; Meyer, 1989; Taylor, 1999). In fact, a relatively large community has formed around computer
hacking, existing in both on and off-line environments (Holt, 2009). An inordinate number of web forums, IRC channels, blogs, and other on-line resources exist providing social links to other hackers, as well as resources to facilitate hacking activities. Businesses and conventions also cater to hackers, including clothing companies, magazines such as the 2600 and Blacklisted, and technology producers (Holt, 2009; Furnell, 2002).

Despite the communal nature of hacking, most hackers indicate that they operate within loose social networks with limited numbers of people on and off-line (Holt, 2007; 2009; Meyer, 1989; Schell & Dodge, 2002). These collegial relationships enable the exchange of information, tools, and normative values and goals, though they largely hack alone. This is due to the fact that the hacker community is a meritocracy, where individuals are judged based on their skill and abilities (Holt, 2007; Jordan & Taylor, 1998; Kilger et al., 2004; Meyer, 1989; Taylor, 1999). Sharing information that benefits other people demonstrates an individual’s mastery of subject matter, and their overall capacity and commitment to the broader hacker community. Though knowledge is acquired in part through social relationships, individuals must also come to understand computer technology through trial and error, and self-learning (Bachmann, 2010; Holt, 2007, 2009; Jordan & Taylor, 1998; Taylor, 1999). As a result, individuals may be more likely to hack alone in order to gain an understanding of technology and further develop their skills with technology.

At the same time, the hacker community is particularly secretive, and individuals who engage in illicit hacks attempt to minimize their likelihood of detection from law enforcement (Holt, 2007; Taylor, 1999). Successful hackers do, however, feel a need to brag and share their accumulated knowledge (Holt, 2007; Jordan & Taylor, 1998). This can help individuals to gain notoriety and status within the hacker community, but increases their risk of detection if they have engaged in illegal activities (Holt, 2007; Furnell, 2002). Similarly, individuals use their handles when advertising the malware that they develop and sell on open black markets operating on-line (Chu et al., 2010; Holt & Lampke, 2010; Honeynet Research Alliance, 2003; Thomas & Martin, 2006). This information is helpful to generate a reputation of skill and ability for a malware writer, as well as make a profit. The handle can, however, be used to directly tie an on-line identity to malicious activity. Thus, some malware writers and sellers face an increased risk of exposure in the course of their activities. This risk can be significantly diminished through the use of multiple handles and on-line identities as a means of obfuscation. As a result, generating attention for one’s activities may increase their centrality in larger hacker social networks but increase their risk of detection from law enforcement or investigators (Holt et al., 2008; Jordan & Taylor, 1998; Taylor, 1999).

A final issue affecting an individual’s network position is the stratification of skill within the hacker community. While there is little to no demographic research on hackers, there is sufficient evidence to demonstrate variations in the ability of hackers (Furnell, 2002; Holt, 2007; Jordan & Taylor, 1998; Taylor, 1999). A sizeable percentage of the hacker community appears to have minimal skill with computer technology. These individuals may be developing hackers who have an interest in technology, but lack a deep or profound level of knowledge about the ways in which computer software and hardware function (Furnell, 2002; Holt, 2007; Jordan & Taylor, 1998; Schell & Dodge, 2002; Taylor, 1999). A smaller proportion comprises of individuals with some skill, who have the capacity to engage in attacks with or without authorization from the system owners and understand how their attacks function (Furnell, 2002; Holt, 2007; Jordan & Taylor,
Finally, the smallest percentage of the population consists of hackers who not only have the ability to engage in attacks, but also create the software and tools necessary to facilitate complex automated attacks against a variety of systems (Holt, 2007; Holt et al., 2008; Jordan & Taylor, 1998; Schell & Dodge, 2002).

This sort of tiered hierarchical structure suggests that information and tools flow down from hackers with the greatest amount of skill to those with the least significant capacity for attacks. In fact, the emergence of markets where tools can be bought and sold reflect this process, as skilled hackers can make a significant profit by offering their services to those with less capability to complete an attack (Chu et al., 2010; Holt & Lampke, 2010; Holt et al., 2008; Honeynet Research Alliance, 2003; Thomas & Martin, 2006). These markets reduce the skills needed to complete an attack, and create a dependence on skilled hackers to engender the capabilities of the larger malicious hacker community (see Chu et al., 2010; Holt & Lampke, 2010). As a consequence, it is thought that those hackers with the greatest skill play a central role in hacker social networks and are looked to by others.

The existing body of knowledge on the hacking and malware community suggests that there is a limited proportion of skilled hackers, though it is unknown how they interact with and are positioned relative to others in the community. Few studies have attempted to develop models to accomplish this effort, thus this study attempts to address this gap in the literature by examining the social networks within a population of Russian malware writers, hackers, and actors. We utilize a unique dataset collected from social networking sites of active Russian hackers and malware writers to identify the nature and structure of relationships between different hackers and groups overall.

**Methods and Data**

In order to assess the social networks of the malware and hacking community in Russia and Eastern Europe, this study utilizes a sample of 336 individuals who maintain accounts via a commonly used social networking site that is extremely popular in Russia. The individual account profiles acquired appear to be distinct identities and do not appear to be duplicates. Blogs provide substantive information on individuals in the hacker community, including their relationships, behaviors, interests, attitudes, beliefs, and location (Holt, 2010; Holt et al., 2008). Information provided by bloggers is, however, self-reported making it difficult to determine the veracity of each claim.

Biographies sometimes provide useful information on psychological status of the user or whether the journal is friends-only. The friends-only status indicates that the user shares all his/her entries only with the people whom s/he has added as ‘friends’. We usually cannot observe those entries, but can estimate the user’s activity through the information on the number comments posted, and blog entries created. The ‘friends’ term does not necessarily reflect the nature of relationships between the people, especially if they belong to the same community. At least one user in the data set complained that his ‘enemy’ has be-friended him. Thus, the LJ allows users to be ‘friends’ in a similar fashion to Facebook, though one can have unidirectional connections in this site. The general network of relationships can be structured based on mutual friends (both have added each other), friends (a user has added another user without a reciprocal tie from that user), and also friend of (a user has been added by another user without adding them in return). Using this information, we then create social networking structures of hacker groups.

This study uses a sample of the members of multiple hacker groups that have connected forums known to sell and trade malicious software and stolen data. The community
membership variable is not mutually exclusive; 25 individuals are members of multiple groups. The groups also vary in their overall representation in the data, from 2 members in Hell Knights to 117 members in Hackzona. The following table shows the composure of our dataset. The content of each blog was downloaded and translated by a native speaker.

Table 1. The data by group membership

<table>
<thead>
<tr>
<th>Community</th>
<th>Members</th>
<th>Members in Multiple Groups</th>
<th>Percentage of group</th>
</tr>
</thead>
<tbody>
<tr>
<td>BH Crew (BH)</td>
<td>104</td>
<td>15</td>
<td>14.4%</td>
</tr>
<tr>
<td>CUP (CU)</td>
<td>16</td>
<td>7</td>
<td>4.37%</td>
</tr>
<tr>
<td>Damage Lab (DL)</td>
<td>27</td>
<td>5</td>
<td>8.5%</td>
</tr>
<tr>
<td>Hell Knights Crew (HN)</td>
<td>2</td>
<td>2</td>
<td>100%</td>
</tr>
<tr>
<td>Hackzona (HZ)</td>
<td>117</td>
<td>23</td>
<td>19.7%</td>
</tr>
<tr>
<td>Mazafaka (MF)</td>
<td>14</td>
<td>3</td>
<td>21.5%</td>
</tr>
<tr>
<td>RU Hack (RU)</td>
<td>9</td>
<td>5</td>
<td>55.6%</td>
</tr>
<tr>
<td>Zloy (ZL)</td>
<td>75</td>
<td>6</td>
<td>8%</td>
</tr>
</tbody>
</table>

Subsequently, each username and the contact information provided were used to conduct Google search queries to determine their involvement in the hacker community. The results were then developed to create a risk index for each user based on their involvement in the creation, distribution, and use of malicious software and hacking techniques. Specifically, a four frame risk typology was created, where those with no information that could be found through open source searches were categorized as zero. Those whose searches indicated that they posted in or ran computer security blogs or participated in on-line discussions about security issues were given a rank of one. Individuals, whose profile searches indicated that they participated in hacking and malicious software forums, including posting tutorials or other information to engage in attacks, were ranked at two. Users in this group posed a prospective risk to other computer users because they may engage in attacks against various targets. Finally, those whose searches indicated that they created and sold malicious software or hacking tools and/or served as hacker forum moderators or managers were given a score of three.

These individuals pose the highest risk because of the preponderance of information that supports their participation in the facilitation of attacks or malicious activity on-line. Overall, 70 percent of the user population had no perceived risk, though a substantive proportion of users were in the second risk category (19%) and a smaller proportion were in the high risk category (6.3%). This is in keeping with the evidence that very small proportion of high skilled actors operate within the hacker community, relative to a larger proportion of semi-skilled actors (Holt, 2010).
User Demographics

In examining the content of all posts, it is possible to extract information concerning the demographic composition of this sample. It appears that the majority of respondents lived within the Russian Federation (see Table 2), and former Soviet republics. A small proportion also lived in parts of Europe and the United States, though a substantive proportion either did not include physical location information in their profiles, or listed locations that do not exist such as Hogwarts School of Magic.

Table 2. Location of Blog User

<table>
<thead>
<tr>
<th>Country</th>
<th>Members</th>
<th>Percent*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belarus</td>
<td>3</td>
<td>2.8</td>
</tr>
<tr>
<td>China</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>Estonia</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>Germany</td>
<td>3</td>
<td>2.8</td>
</tr>
<tr>
<td>Jamaica</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>Kyrgyzstan</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>Laos</td>
<td>2</td>
<td>1.9</td>
</tr>
<tr>
<td>Moldova</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>78</td>
<td>73.6</td>
</tr>
<tr>
<td>USA</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>Ukraine</td>
<td>13</td>
<td>12.3</td>
</tr>
</tbody>
</table>

*Percent from all none missing entries (106 total). Number of entries missing geographic location equals 235.

One of the most interesting aspects of our analysis is that many of the higher threat actors indicate their current location and general personal interests. For instance, nine individuals self-identified living in Moscow, Chelyabinsk, Novosibirsk, and Murmansk. Their posts also suggest they genuinely live in these cities, which goes against the larger emphasis on secrecy within the hacker community (Jordan & Taylor, 1998; Taylor, 1999). There may, however, be value in revealing one’s location in the blogosphere in order to connect with people from the same location who operate outside of the hacker subculture.

Individual users were, however, less likely to indicate their former or current educational institution. This could be due to the fact that many of the users are students, and revealing their hacking skills and university information could lead to the identification of their real identity. It may also be that those with higher skill do not necessarily have a two or four year degree, but rather honed their skills on their own in keeping with the importance of self learning within the hacker subculture (Bachmann, 2010; Holt, 2007; Jordan & Taylor, 1998; Schell & Dodge, 2002). Taking information from the profiles about education, it appears that a small proportion of the respondents are currently students attending formal universities within Russia. In terms of age, the majority of respondents were between the ages of 21 and 29 in keeping with the general age distribution of the hacker community (Holt, 2007; Jordan & Taylor, 1999; Schell & Dodge, 2002). Furthermore, the language used by posters suggests that only seven individual profiles were females, again conforming to the general evidence on gender in the hacker community (Bachmann, 2010; Gilboa, 1996; Holt & Kilger, 2008; Jordan & Taylor, 1998; Schell & Dodge, 2002).
Data Analysis

Network analyses were conducted to indicate the centrality of users with high perceived threat level. We have used communities as one way to pool users together, along with peer relationships, group membership, and risk levels. Social network analysis techniques allow for the visualization of large networks through multiple data points. In this data set, individual blog posters become network vertices, while their connections (membership and friendship ties) establish the relationships between different vertices. Specifically, each hacker (U) has relationships with another hacker through the membership in the same network. The networks are connected through the users belonging to multiple networks. The relationship (R) or set of arcs (connections between the users) and users (U) identify a given network N(U,R) (where $R \subseteq U \times U$). The following figure shows an example of network relationships between three different actors (e.g. hackers). The color of nodes can identify their threat levels; the weight of the lines shows the frequency of their communication.

Figure 1. Example of Network Relationships

Since there are no specific studies on Russian hackers’ networks, we use exploratory techniques to study the structure of networks. In particular, we attempt to identify where higher threat hackers reside within social networks relative to other lower skilled actors. The users with mutual membership in different hacker networks might serve as liaisons between the groups or major hackers with the highest risk level. Thus, we examined networks based on risk levels. The majority of high risk actors (23) belonged to two groups, while one individual belonged to three groups and another to four respectively. Thus, the question arises whether there is a relationship between the number of groups an individual belongs to and their risk level.

A t-test with unequal variances was conducted to test whether there are significant differences in mean risk level according to multiple memberships. The findings indicated that those belonging to multiple groups are users with the highest risk levels. For example, the Hell Nights Crew (N=2) has overlapping memberships with many other groups and both members are in the high risk category. Similarly, RU Hack had over half of all members with multiple memberships, and CUP had 46% belonging to other groups as well. Interestingly, groups whose members have multiple affiliations tend to have the smallest size. ANOVA analyses demonstrate that there are significant differences between the groups based on risk levels, suggesting that group membership is a key predictor of overall risk.
The overlapping members have been allocated to the groups with the smaller sizes to conduct network analysis presented in Figure 2. For the network model, an edge exists between individuals if they belong to the same group. Since the members of the same group are connected to one another, the visualization in Figure 2 provides an easier way to depict connections between groups. This figure presents the ties between groups, particularly the dense ties of the Hell Nights Crew with the BH Crew and Mazafaka. It is important to note that the strongest ties are between the Hell Nights Crew and Damage Lab.

**Figure 2: Network Structure**

A slightly different visualization of the network sociogram is represented by the connection between the groups (see Figure 3). Line weights indicate the strength or scale of connection between groups. The strongest connections are between RU Hack, Hack Zona, and Mazafaka. Similarly, there are dense connections between HackZona, the Hell Nights Crew and Damage Lab. Users belonging to these groups have a higher propensity to distribute information into these groups through their overlapping memberships. This finding suggests that individuals possessing overlapping group memberships may play an important role in communicating information across groups, as noted in research on the connections between hacker groups generally (Holt, 2009; Meyer, 1989; Schell & Dodge, 2002). Also, this implies there are redundant relationships in hacker networks which may drive the continued use of older malicious software and exploits in the hacker community (Chu et al., 2010; Holt & Lampke, 2010; Peretti, 2009).
In order to clarify the role of group membership on individual behavior, a sociograph for “popularity” as measured by the number of mutual friends by group was created (see Figure 4). The size of the node indicates the number of mutual friends of each individual member in each group. The color coding represents the network groups as indicated by the legend. The edges in the following graph indicate that two users are mutual friends. Despite having a smaller membership, RU Hack is at the center of connectivity because of the number of mutual friends. Similarly, members of MazaFaka and Zloy are centrally connected, while the members of the .Hell Nights Crew since these two individuals are not being closely followed.
In order to further refine these models, we explored whether risk levels predict popularity (see Figure 5). Those individuals with high risk scores are clearly involved in the creation and dissemination of attack tools, and may serve as innovative attackers that serve as models or sources of imitation for other actors (see Chu et al., 2010; Holt, 2009; Holt & Kilger, 2008). Thus, Figure 5 provides a sociograph for connectivity and centrality based on risk score. This figure indicates that users with higher risk level are actually more centrally located in the network. Similarly, high risk actors are connected with 2 degrees of separation, so if they are not friends with one another, they will have a mutual friend who knows both hackers. This demonstrates that risk is an important factor in popularity, and that high threat actors are likely to know one another due in part to their involvement in multiple networks (Holt, 2009; Meyer, 1989; Schell & Dodge, 2002). In addition, this graph provides visual support for the notion that there are a higher number of no or low-skill actors within the hacker community than high skilled actors (see Holt & Kilger, 2008; Jordan & Taylor, 1998). The overall network density is 0.275 based on the number of actual dyadic connections divided by the number of total possible connections. The higher the number, the more connected the members are on the network.
Discussion and Conclusion

Though the threat posed by computer hackers and malware writers has increased dramatically, few researchers have attempted to examine this community in depth. The existing research literature provides some insights on the structure and organization of hacker groups (Chu et al., 2010; Holt, 2009; Holt et al., 2008; Meyer, 1989), though few have attempted to address these issues with quantitative data (Bachmann, 2010; Holt & Kilger, 2008; Schell & Dodge, 2002). This study attempted to explore these issues using on-line data from social networking sites for a population of individuals interested in hacking and malware.

Our analysis underscores that a significant amount of information can be generated from social networking data. A great deal of demographic data was derived from the content of their blogs, such as the fact that many participants were located in Moscow and St. Petersburg. Furthermore, the data suggests that this community is composed of mostly young males with minimal higher education in keeping with various other studies of hacker populations around the world (Bachmann, 2010; Holt, 2007; Holt et al., 2008; Jordan & Taylor, 1998; Schell & Dodge, 2002). In addition, there is an extremely small proportion of this community who may be deemed a highly skilled computer hacker or malware writer. A substantive proportion of the actors had generally less skill, or no
discernable skill whatsoever, in keeping with general assertions from qualitative (Holt, 2007; Jordan & Taylor, 1998; Taylor, 1999) and quantitative (Holt & Kilger, 2008; Schell & Dodge, 2002) studies of the hacker subculture. It is possible that this may reflect hackers carefully managing the information they share through social networking sites and other on-line resources to reduce their detection by law enforcement and the general public (Holt et al., 2008; Meyer, 1989). Thus, additional research is needed to replicate and validate the distribution of hacker skill identified in this study. Such findings could increase our understanding of hacker activity, and the hacker subculture generally.

Examining the sociographs and networks within this community also demonstrate that hacker groups are well connected to one another. This is sensible, as individuals with a significant interest in hacking may participate in multiple on-line communities in order to gather more information and increase the density of their social networks (Holt 2009; Holt et al., 2008; Meyer 1989). Hackers who pose the greatest threat to the broad population of computer users through the creation and distribution of malware and hacker tools were also densely connected in the center of networks. As a consequence, the tools they develop may be easily acquired by those with less skill through the substantive number of redundant networks. In addition, the insulation noted within this network may account for the constant recycling of tools and attacks from skilled to unskilled hackers within the community (Chu et al., 2010). Thus, the network structures identified in this study generally reflect the assertion that hackers operate within a collegial subculture that encourages information sharing and values innovation and skill (Bachmann, 2010; Chu et al., 2010; Holt, 2009; Holt & Kilger, 2008; Holt et al., 2008; Jordan & Taylor, 1998; Meyer, 1989; Schell & Dodge, 2002; Taylor, 1999).

This study is, however, limited by a variety of factors requiring greater exposition. Specifically, this sample was generated from one social networking site limiting its representativeness and generalizability to other hacker communities. In fact, the regional variations evident in communications channels and information sharing techniques (see Chu et al., 2010; Holt, 2010; Holt et al., 2008) may create substantive variations in the network structures of hacker communities across the world. These networks also included several well known hackers and forums operating on-line. As a result, the findings may not be representative of more sensitive and hidden social networks in the hacker community. Thus, additional research is needed using both qualitative and quantitative data techniques to shed light on the social organization and composition of hacker groups around the world. In turn, we can vastly improve our knowledge of the hacker subculture, and its participants.

References


