Using Social Network Analysis to Identify Key Hackers for Keylogging Tools in Hacker Forums

Sagar Samtani and Hsinchun Chen
Department of Management Information Systems
The University of Arizona
Tucson, AZ 85721
sagars@email.arizona.edu, hchen@eller.arizona.edu

Abstract— Cyber-attacks are critical cybersecurity concerns across the world. Catching malicious hackers prior to a cyber-attack can save significant financial cost as well as avoid devastating cyber-attacks. Current methods of identifying and reprimanding hackers generally occurs after an attack and is reactive in nature. This research aims to proactively identify key hackers who are creating and disseminating malicious tools within hacker forums. Specifically, we utilize social network analysis techniques to systematically identify key hackers for keylogging tools within a large English hacker forum. Results of this study indicate that many key hackers are the most senior, longest tenured participants of their community.

Keywords—social network analysis; keylogging; hacker forums

I. INTRODUCTION

Recent years has seen an alarming growth in cyber-attacks. To mitigate such attacks, governing bodies around the world implement laws and regulations punishing malicious hackers causing cyber damage. Catching such individuals prior to a cyber-attack can save significant financial cost as well as avoid devastating cyber-attacks. However, the current methods of identifying and reprimanding hackers generally occurs after an attack, rather than catching them prior to attack [1].

Online hacker communities can provide a valuable data source for law enforcement to proactively identify and catch malicious individuals prior to any cyber-attack. Online hacker communities span various geo-political regions including the United States, Russia, China, and the Middle East. Each region tends to specialize in various aspects of cybersecurity such as cyberwarfare, underground economies, malware creation, etc. Various online platforms (e.g., IRC, carding shops, and forums) help facilitate the dissemination of hacking related knowledge. Among these platforms, hacker forums provide the ability for hackers to easily share malicious tools with each other. These tools can be easily downloaded and used by hackers for malicious purposes. Figure 1 illustrates an example of a forum post with an attached botnet tool for download.

Overall, there are hundreds of forums with millions of members posting tens of thousands of malicious tools. Studying key hackers creating and disseminating these tools would help to discover individuals who may conduct malicious cyber-attacks. Given this value, this research utilizes social network analysis techniques to identify key hackers for selected tool types within a large, English hacker forum.

The remainder of the paper is structured as follows. First we review literature on hacker community research and social network analysis techniques. Second, we detail our research testbed and design. Subsequently, we summarize key findings and results. Finally, we highlight future directions and conclude this research.

II. LITERATURE REVIEW

A. Hacker Community Research

Past hacker community research has focused primarily on understanding the overall social network of hacker community members. The majority of this work has utilized manual explorations on Russian, English, and German hacker forums to identify that the majority of forum participants are unskilled, a medium sized group is semi-skilled, and a small percentage is highly skilled [2-5]. Many of these studies noted that the highly skilled individuals were often the ones which created and disseminated many of the malicious tools available in the forums. This sharing often leads to the advancement of their social status within the forums [4].
B. Social Network Analysis

Social network analysis utilizes graph theory to study relationships between social entities. Social networks have two major components: nodes and edges. Nodes represent social entities (e.g., individuals) while edges represent the relationships between nodes. Edges can be directed (one-way) or undirected (mutual, two-way). While useful for studying various phenomena, it is less applicable in determining the relationships between two different types of entities (e.g., hackers and tools). However, a bipartite network is a special network configuration designed to model such relationships [5]. Bipartite networks are used extensively in a range of virtual community context ranging from Wikipedia co-authorship [6] to knowledge translation in health forums [7]. Unlike traditional monopartite networks, bipartite networks partition nodes into two sets. One node category (individuals) is “affiliated” with the other category (events). Nodes representing individuals can only have relationships with events, not other individuals. While useful for modeling different context they can only generate a subset of the standard network metrics. As such, bipartite networks are often projected as monopartite networks to calculate all metrics (e.g., distance, diameter, closeness, betweenness, eigenvector, degree, etc.).

C. Research Gaps and Questions

We identified two major research gaps from our literature review. First, most prior hacker forum literature has focused on general understanding general behavior of hackers within forums. While useful, such analyses do not provide a granular view of key hackers for specific categories of tools. Additionally, the approaches have been manual, and thus not scalable to larger datasets. Second, well-developed social network analysis techniques (i.e., monopartite and bipartite networks) have been used extensively in traditional virtual community settings, but not in hacker forums. Given these gaps, the following research questions are posed for study:

- How can monopartite and bipartite network analysis techniques be utilized to identify key hackers within forums?
- Who are the key hackers for selected groups of malicious tools?

III. RESEARCH TESTBED AND DESIGN

For the purposes of this study, we identified one English hacker forum, OpenSC, for collection and analysis. Our overall analytical framework is highlighted in Figure 2. OpenSC promotes itself as “The Largest Malware Forum Known to Man” and has been in use since 2005. By utilizing a variety of data collection and analytical techniques highlighted in [10], we were able to crawl and extract 124,993 made by 6,796 authors from 2005-2016. From these forum posts, we identified a variety of different tools such as Remote Administration Tools (RATs), crypters, rootkits, and browser exploits, etc.

Although it would be possible to model the entire social network of key hackers for all malicious tools, such an analysis would not provide us with a granular look for key hackers for specific hackers. For the purposes of this study, we focus on modeling key hackers for keylogging tools. Keyloggers are designed to monitor an individuals’ keystrokes. They can steal a variety of personal information such as credit card information and bank login credentials. In our framework, we constructed bipartite networks connecting hackers (node type 1) to threads with keylogging tools (node type 2). While it would be ideal to build networks where we can directly model hacker relationships with specific keylogging posts, such information is limited to forum administrators. Additionally, modeling user to thread relationships has been used in other bipartite virtual community studies [8]. To maximize our analytics, we calculate a variety of network metrics, identify key hackers for keylogging assets, and project the bipartite network as monopartite.

IV. RESULTS AND DISCUSSION

For space considerations, we only visually depict monopartite networks (Figure 3). However, a variety of node and topological network measures are calculated and presented for both networks in Table 1. As our goal is to understand key hackers for keylogger tools, we focus our discussion on the monopartite network projection.

![Fig. 2. Research Design](image1)

![Fig. 3. Monopartite hacker network](image2)

<p>| TABLE I. | TOPOLOGICAL AND NODE METRICS FOR NETWORKS |</p>
<table>
<thead>
<tr>
<th>Category</th>
<th>Metric</th>
<th>Bipartite Network</th>
<th>Monopartite Hacker Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topological Metrics</td>
<td># of nodes</td>
<td>82</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td># of edges</td>
<td>117</td>
<td>510</td>
</tr>
<tr>
<td></td>
<td>Network Diameter</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Graph Density</td>
<td>0.018</td>
<td>0.131</td>
</tr>
</tbody>
</table>
Overall, there are 63 nodes (i.e., hackers) in the monopartite network and 82 nodes in the bipartite network. Hackers have to take a minimal amount of steps to contact other keylogger hackers (indicated by a network diameter of 5). This is possibly due to the giant component, which makes up 28.78% of the network. However, despite the minimal number of steps to reach other hackers, the low graph density (0.131) indicates that many hackers do not take advantage of knowledge from all other hackers, but only a select few. This phenomena is reinforced by prior literature as well as the degree distribution [3]. Only a select few hackers have high (30 or above) degree. The majority of these top hackers are senior members in the forum. They have been part of their communities for extended periods of time, and contribute knowledge beyond keylogging tools. We summarize the top eight hackers based on degree and betweenness centrality in Table 2. Both of these metrics are well accepted in identifying key members within social networks [6].

V. CONCLUSION AND FUTURE DIRECTIONS

Cyber-attacks are critical cybersecurity concerns across the world. This research aimed to identify key hackers for specific groups of malicious tools. By utilizing social network analysis approaches, discovered many key hackers responsible for the dissemination of keylogging assets. Overall, the results of our study show that many of these key hackers are senior members within the forum and provide a large amount of value to their overall community.

There are many promising avenues for future expansion. First, we only analyzed a single tool type within one forum. However, this can easily be expanded to include multiple tool types or multiple forums. Such an analysis would be useful to discover whether some of the identified phenomena appear in multiple contexts. Second, additional network analysis can be utilized. For example, including community detection metrics may help to identify specific groups of hackers. Finally, incorporating dynamic network modeling techniques would be valuable in identifying key hackers at specific points in time. All of the aforementioned analysis would aid in the detection of malicious, key hackers within hacker forums and potentially reduce future cyber-attacks.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. NSF-DUE-1303362 and NSF SES-1314631.

REFERENCES