Diversity or Concentration?
Hackers’ Strategy for Working Across Multiple Bug Bounty Programs
Keman Huang
Michael Siegel
Stuart Madnick
Xiaohong Li
Zhiyong Feng

Working Paper CISL# 2016-23
December 2016
Diversity or Concentration?
Hackers’ Strategy for Working Across Multiple Bug Bounty Programs

Keman Huang1,2,3, Michael Siegel3, Stuart Madnick3,4, Xiaohong Li1,2, Zhiyong Feng1,2
1Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin 300072, China
2School of Computer Science and Technology, Tianjin University, Tianjin 300072, China
3MIT Sloan School of Management, Cambridge, MA, USA
4MIT School of Engineering, Cambridge, MA, USA

keman.huang@tju.edu.cn, msiegel@mit.edu, smadnick@mit.edu, xiaohongli@tju.edu.cn, zfeng@tju.edu.cn

Abstract—Bug bounty programs have been proved effective in attracting external hackers to find and disclose potential flaws in a responsible way. There are many different bug bounty programs, so how do hackers balance diversity and concentration to effectively build their reputation in the vulnerability discovery ecosystem? In this paper, we present a novel methodology to understand how hackers spread their attention and earn bounties across different programs. The empirical result shows the relationship between diversity and concentration to effectively build their reputation in the ecosystem.

In this paper, we develop a methodology to understand how hackers spread their attentions across different bug bounty programs and how they gain their bounties. Based on the data collected from HackerOne, our empirical results show the different strategies for hackers with different reputation levels in the ecosystem. This opens a gateway for us to further study the hackers’ incentive and behaviors in the bug bounty programs.

II. METHODOLOGY

Generally speaking, in the vulnerability discovery ecosystem, each hacker will select a program and then devote time to work on vulnerability discovery in these programs. If a vulnerability is discovered, he/she will submit a report to the program, and the program will work with him/her to determine if the vulnerability is valid and first the first disclosure. Finally, the hacker can gain a bounty if the discovery meets the requirements of the program. By making more valid submissions and gaining bounties, the hackers also build their reputation in the ecosystem.

As the goal of this paper is to understand how hackers work across different programs. For a hacker \( h \) who participates in more than 1 program, we get all his/her valid submissions and then sort his/her programs based on the submission number:

\[
a(h) \leq s_{p_1}, s_{p_2}, \ldots, s_{p_N} > , \quad |s_{p_i}| \geq |s_{p_j}| \ldots \geq |s_{p_N}|
\]  

(1)

Here \( |s_{p_i}| \) refers to the number of valid reports \( h \) submits to the \( j \)th program.

Therefore, we can get the distribution representing how each hacker spread his/her attention to different programs:

\[
pa(h) = s_{r_{p_1}}, s_{r_{p_2}}, \ldots, s_{r_{p_N}}
\]  

(2)

\[
s_{r_{p_k}} = |s_{p_k}| \sum_{j=1}^{N} |s_{p_j}|
\]  

(3)

Then, the average attention rate \( aar_j \) is generated to evaluate how hackers in the ecosystem spread their attentions to their \( j \) priority programs:

\[
aar_j = \frac{1}{N} \sum_{i=1}^{N} s_{r_{p_i}}
\]  

(4)

1 In many cases, building a reputation (such as for future employment) may be as important as the amount of bounty collected.
Here $N$ refers to the number of hackers participating into more than 1 program.

Finally, we can calculate the entropy index [6] to evaluate the concentration of each hackers’ programs:

$$EI_w(h) = -\sum_{i=1}^{k} sr_{i,j} \log_{10}(sr_{i,j})$$

Similarly, we can calculate the average reward rate $arr_i$ and the entropy-based concentration rate $EI_r(h)$ to represent how hackers gain bounties from their priority programs:

$$arr_i = \frac{1}{M} \sum_{i=1}^{M} rr_{i,k}$$

$$EI_r(h) = -\sum_{i=1}^{k} rr_{i,j} \log_{10}(rr_{i,j})$$

$$rr_{i,k} = \frac{|r_{ps}|}{\sum_{j=1}^{n} |r_{ps}|}$$

Here $M$ refers to the number of hackers who gain bounties from more than 1 program. $|r_{ps}|$ refers to the number of rewarded submissions hacker $h$ gain from his/her $k$th top program.

III. DATASET AND RESULT

A. Data Set

HackerOne is a well-known US bug bounty platform which hosts many different programs (132 public programs, as of April 1, 2016) offered by different companies including Yahoo!, Twitter, Adobe, Uber, etc. Data was collected from November 29, 2013 to October 28, 2015 on the 567 hackers who participated in more than 1 program for attention concentration analysis and the 214 hackers who gained rewards from at least 2 programs for reward concentration analysis. Detail about the dataset is in the support material.

B. Empirical Result

Due to the space limitation, the figures are presented as support material. Based on the data we collected from HackerOne, we can observe the well-known power-law distributions [7] both for the average attention rate and the average reward rate. This means that overall, hackers in the ecosystem pay most of their attentions to the prioritized programs and gain most of their earnings from them. Actually, it can be seen that 90.47% of submissions are from the first 3 priority programs for these hackers. The first 3 prioritized programs contribute 95.68% bounties for hackers and only 0.77% are from the programs with a priority less than 5. Additionally, the gap between the average attention rate and the average reward rate is increasing with the reduction in priority, which reveals that reports submitted to the less priority programs gain negligible reward.

Furthermore, in order to compare the different strategies for hackers with different levels, we separate the hackers into 10 groups based on their effectiveness, which we will refer to as reputation. For each group, we calculate their average entropy index for both the attention rate and the reward rate. From Figure 2, we can observe a significant transform between the diversity and concentration for hackers: for the hackers with low reputations, they submit to different programs and get a relatively higher diversity. However, this distraction limits their ability to discover important and valuable vulnerabilities and build reputation; The ones with medium reputations, focus on their priority programs; The top hackers have a higher diversity than the overall ecosystem which means that they spread their vulnerability discovery ability to more programs. Therefore, for the hackers, it is a reasonable strategy to focus on few programs to gain professional recognition and then diversify to different programs to build up ones reputation in the community.

IV. DISCUSSION AND CONCLUSION

Bug bounty programs have been launched by many companies, attracting external hackers to discover potential vulnerabilities through responsive disclosure. It is important for hackers to balance between diversity and concentration. Our empirical study shows that most hackers concentrate on few programs, empirically less than 5, and earn most of their bounties from these programs. Additionally, the entropy-based concentration reveals the strategy between concentration and diversity: it is a good choice for hackers to initially focus on few programs and then diversify to multi-programs to build reputation in the community.

This preliminary result opens a gateway for us to further dig deeper to understand the hackers’ behavior and investigate the vulnerability discovery ecosystem.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China grant 61502333 and the MIT (IC)3 project. Zhiyong Feng is the corresponding author.

REFERENCES

Support Material

A. Data Set

Table I reports the basic statistic of our dataset. “Attention” refers to dataset in which only hackers who submit valid reports to at least 2 programs are included. “Reward” refers to the dataset containing only hackers who gain bounty from at least 2 programs.

<table>
<thead>
<tr>
<th></th>
<th>Attention</th>
<th>Reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Hackers</td>
<td>567</td>
<td>214</td>
</tr>
<tr>
<td>#Programs</td>
<td>98</td>
<td>48</td>
</tr>
<tr>
<td>Maximum #Programs per Hacker</td>
<td>44</td>
<td>15</td>
</tr>
<tr>
<td>#Valid Submissions</td>
<td>7,095</td>
<td>/</td>
</tr>
<tr>
<td>#Rewarded Bounty</td>
<td>/</td>
<td>$678,504.25</td>
</tr>
</tbody>
</table>

B. Empirical Result

How do hackers spread their labor & earn reward from multiple programs

![Graph](image)

Figure 1. Average Attention Rate and Average Reward Rate for Hackers across Different Programs.
Figure 2. Average Attention Rate and Average Reward Rate for Hackers across Different Programs.