# An Empirical Study of Web Vulnerability Discovery Ecosystems

Mingyi Zhao
Pennsylvania State University
muz127@ist.psu.edu

Jens Grossklags
Pennsylvania State University
jensg@ist.psu.edu

Peng Liu
Pennsylvania State University
pliu@ist.psu.edu

## ABSTRACT

In recent years, many organizations have established bounty programs that attract white hat hackers who contribute vulnerability reports of web systems. In this paper, we collect publicly available data of two representative web vulnerability discovery ecosystems (Wooyun and HackerOne) and study their characteristics, trajectory, and impact. We find that both ecosystems include large and continuously growing white hat communities which have provided significant contributions to organizations from a wide range of business sectors. We also analyze vulnerability trends, response and resolve behaviors, and reward structures of participating organizations. Our analysis based on the HackerOne dataset reveals that a considerable number of organizations exhibit decreasing trends for reported web vulnerabilities. We further conduct a regression study which shows that monetary incentives have a significantly positive correlation with the number of vulnerabilities reported. Finally, we make recommendations aimed at increasing participation by white hats and organizations in such ecosystems.

## Keywords

Bug Bounty; Vulnerability Discovery; Vulnerability Disclosure; Monetary Incentives

## 1. INTRODUCTION

Websites are critical pathways to facilitate e-commerce, customer service, input procurement, and employee connectivity, and they continue to reach significant penetration in various business sectors. Most large businesses are hosting web services, and over 50% of small businesses are now offering web accessibility [10]. As such, web security has become critically important for most organizations, and the prevention of security compromises enabled by web vulnerabilities is gaining increasingly the attention of company leadership and the broader security community. Nevertheless, web vulnerabilities are the likely causes of many recent security breaches contributing to massive disclosure of user data, leakage of business information, and other losses.

To reduce the number of web vulnerabilities, organizations can use automated web vulnerability scanners which however have been shown to only have limited coverage [16, 37]. In response, organizations more recently started to directly collaborate with or indirectly benefit from outside security researchers. These so-called *white hat* researchers spend time to analyze organizations' web systems and report vulnerabilities to self-run bug bounty programs of organizations such as Facebook, Github and PayPal, or to corresponding programs on third-party *bug bounty platforms* such as Wooyun, HackerOne, BugCrowd, Cobalt, etc.

White hats contribute in many positive ways to the discovery of web vulnerabilities. First, they can complement the limitations of automated scanners [16] by reaching deeper states of web applications, and may better understand the application logic. Second, with a mindset comparable to attackers, white hats are good at finding many exploitable vulnerabilities of high severity. Third, the large and diverse group of potential white hat contributors outnumbers internal security teams or penetration testing teams and could therefore cover a wider range of security issues.

White hats' considerable efforts are rewarded in different ways. Organizations or bug bounty platforms may provide monetary incentives based on severity and originality of the discovered issue, or publicize white hats' contributions to enhance their reputations. Previous studies and reports have shown that the cost of utilizing the white hat community may be lower compared with hiring internal security researchers [20] or using services from penetration testing companies [5].

The resulting interactions extend beyond organizational boundaries and form *web vulnerability discovery ecosystems* including businesses/organizations, white hats, and third-party vulnerability disclosure reward/bounty programs (Figure 1). These ecosystems have been growing rapidly and are becoming more prominent in the battle against malicious actors on the Internet. However, detailed studies of these web vulnerability ecosystems to understand their characteristics, trajectories, and impact are notably absent.

In this work, we conduct the first empirical study of two major web vulnerability discovery ecosystems. We base our analyses on publicly available data. The first dataset is collected from Wooyun[1], the predominant and likely the oldest web vulnerability discovery ecosystem in China. Our data contains 64,134 vulnerabilities affecting a total of 17,328 or-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
*CCS'15,* October 12–16, 2015, Denver, Colorado, USA.
© 2015 ACM. ISBN 978-1-4503-3832-5/15/10 ...$15.00.
http://dx.doi.org/10.1145/2810103.2813704.

[1] www.wooyun.org

| Platforms | Start | HQ | # Vuln. | # WHat | # Org. | Bounty Paid | Disclosure |
|---|---|---|---|---|---|---|---|
| **Wooyun** | **2010-07** | **China** | **64,134** | **7,744** | **17,328** | **Unknown** | **Full** |
| Facebook (2013) [4] | 2011-08 | US | 687 | 330 | 1 | $1.5M | No |
| BugCrowd [14] | 2012-09 | US | 7,958 | 566 | 166 | $0.7M | No |
| Loudong 360 | 2013-03 | China | 54,727 | 14,104 | 2,271 | $0.7M | Partial |
| Cobalt | 2013-07 | US | 8,119 | 2,600* | 230 | Unknown | Partial |
| **HackerOne** | **2013-11** | **US** | **10,997** | **1,653** | **99 (Public)** | **$3.64M** | **Partial** |
| Vulbox | 2014-05 | China | 10,000 | 20,000* | Unknown | Unknown | Partial |
| Sobug | 2014-05 | China | 3,270 | 8,611* | 285 | $0.8M (Budget) | Partial |

Table 1: Statistics for representative bug bounty platforms sorted by their start time. The two platforms studied in this paper are highlighted. Numbers were obtained from the cited references, or platforms' websites directly in early August of 2015. The exact definitions of each metric for different platforms may vary. For example, some platforms count registered white hats (marked with *), while others such as HackerOne count white hats that have made at least one valid contribution.

ganizations including almost all popular Chinese web companies. We additionally collect publicly available data from HackerOne[2], a US-based start-up company which hosts bug bounty programs for hundreds of organizations, such as Yahoo, Mail.ru and Twitter, from many parts of the world. The Wooyun dataset is larger due to its coercive participation model for involving organizations, and also contains more detailed vulnerability information due to its delayed full disclosure policy. The HackerOne dataset is smaller and not all of its reports can be accessed. However, it covers a different set of organizations and also contains monetary reward information that does not exist for the Wooyun dataset. By combining these two complementary datasets, we are able to explore a wide range of topics and gain a better understanding of the structure and dynamics of such ecosystems and their impact on Internet security. We anticipate that our study will be a valuable reference for organizations who want to create or optimize their existing bounty programs.

We make the following contributions:

- Our analysis shows how many white hats have been attracted by these ecosystems and how the number of contributing white hats evolves over time. We further assess their diversity in terms of productivity and breadth of vulnerability discovery (e.g., types of vulnerabilities and affected organizations) by studying individual contributions but also contributions by groups of white hats with high/medium/low productivity. We also analyze the potential (learning) value of disclosing vulnerabilities to the white hat community.

- We then quantitatively analyze participating organizations from several dimensions, including the vulnerability trends, the coverage of different business sectors, the response and resolve behaviors, and reward structures. We evaluate the trend of reported vulnerabilities for representative organizations.

- Our study further measures the impact of different factors on vulnerability discovery. In particular, we quantify the effect of offering monetary incentives for attracting white hats and reporting discovered vulnerabilities. Based on these analyses, we discuss the benefits of disclosing vulnerability information, offer suggestions on how to improve the effectiveness of the collaboration between white hats and organizations, discuss insights for relevant policy making (e.g., the Wassenaar Arrangement), and identify important research questions for future studies.

[2]`hackerone.com`

We proceed as follows. In Section 2, we discuss related work. In Section 3, we provide background information about Wooyun and HackerOne, and discuss the collection of the datasets. We present our data analysis results in Section 4, and provide a discussion in Section 5. We offer concluding remarks in Section 6.

# 2. RELATED WORK

## 2.1 Software Vulnerability Datasets

Previous work has studied various *software* vulnerability datasets to understand vulnerability discovery, patching and exploitation. This research is relevant for the debate on whether vulnerability disclosure programs are beneficial to society [18]. That is, if the number of potential vulnerabilities is large with respect to the effort of white hats, and vulnerabilities are found in no particular order, then black hats could frequently discover and exploit vulnerabilities that are not covered by white hats' contributions; thereby questioning their effectiveness. On the one hand, Rescorla studied the ICAT dataset of 1,675 vulnerabilities and found very weak or no evidence of vulnerability depletion. He thus suggested that the vulnerability discovery efforts might not provide much social benefit [34]. On the other hand, this conclusion is challenged by Ozment and Schechter, who showed that the pool of vulnerabilities in the foundational code of OpenBSD is being depleted with strong statistical evidence [31, 32]. Ozment also found that vulnerability rediscovery is common in the OpenBSD vulnerability discovery history [31]. Therefore, they gave the opposite conclusion, i.e., vulnerability hunting by white hats is socially beneficial. More recently, Shahzad et al. [36] conducted a large-scale study of the evolution of the vulnerability life cycle using a combined dataset of NVD, OSVDB and FVDB. Their study provided three positive signs for increasing software security: (1) monthly vulnerability disclosures are decreasing since 2008, (2) exploitation difficulty of the identified vulnerabilities is increasing, and (3) software companies have become more agile in responding to discovered vulnerabilities. In another study, Frei et al. studied a security ecosystem including discovers, vulnerability markets, criminals, vendors, security information providers and the public, based on 27,000 publicly disclosed vulnera-

bilities [21]. They focus on vulnerability exploits and patching of native software, while we study the ecosystem around the discovery of web vulnerabilities, and our main focus are the behaviors and dynamics of white hats and organizations that compose such ecosystems.

## 2.2 Vulnerability Discoverers

Most of the existing research on software security focuses on vulnerabilities, affected software products or vulnerability discovery tools. More recently, researchers started to pay attention to the humans who make vulnerability discoveries. Edmundson et al. conducted a code review experiment for a small web application with 30 subjects [17]. One of their findings is that none of the participants was able to find all 7 Web vulnerabilities embedded in the test code, but a random sample of half of the participants could cover all vulnerabilities with a probability of about 95%, indicating that a sufficiently large group of white hats is required for finding vulnerabilities effectively. This is consistent with our analysis in Section 4.2.2 and Section 4.3.7. However, the code review process they focused on is mainly conducted inside an organization with source code available; while the vulnerability hunting focused on in this paper is conducted outside an organization. Finifter et al. provided contribution and payment statistics of participants in Google Chrome VRP and Mozilla Firefox VRP [20], and suggested that VRPs are more cost-effective compared to hiring full-time security researchers. Previous work has also reported that many discoverers primarily rely on their expertise and insights, and limited types of tools such as fuzzers and debuggers, rather than sophisticated automated vulnerability discovery tools [12, 19]. Zhao et al. conducted an initial exploratory study of white hats on Wooyun [38] and uncovered the diversity of white hat behaviors on productivity, vulnerability type specialization, and discovery transitions.

## 2.3 Vulnerability Markets

Böhme offers a terminology for organizational principles of vulnerability markets by comparing bug challenges, vulnerability brokers, exploit derivatives and cyber-insurance [13]. Algarni and Malaiya analyzed data of several existing vulnerability markets and showed that the black market offers much higher price for zero-day vulnerabilities, and government agencies make up a significant portion of the buyers [12]. Ozment proposed a vulnerability auction mechanism that allows a software company to measure its software quality based on the current bounty level, and to conduct vulnerability discovery at an acceptable cost [30]. This auction model can potentially be incorporated into today's vulnerability discovery ecosystems. A panel discussion at the New Security Paradigms Workshop examined ethics and implications for vulnerability markets [18]. Finally, Kannan and Telang showed that unregulated vulnerability markets almost always perform worse than regulated ones, or even no market at all [24]. They also found that it is socially beneficial to offer rewards for benign vulnerability discoverers.

# 3. METHODOLOGY

## 3.1 Analysis Overview

We organize our analysis around three components: the vulnerabilities disclosed, the white hats, and the involved businesses/organizations. Figure 1 outlines the structure of a representative web vulnerability discovery ecosystem. In the following, we describe our data collection efforts.

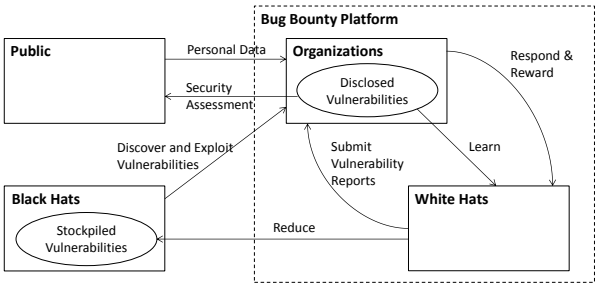


Figure 1: Structure of a web vulnerability discovery ecosystem.

## 3.2 Data Collection

We have collected publicly available data from Wooyun and HackerOne. The processed data and related Python scripts can be shared upon request in order to reproduce and extend our research.

### *3.2.1 Wooyun*

Wooyun is the predominant web vulnerability disclosure program in China launched in May 2010. It has attracted 7,744 white hats who contributed 64,134 vulnerability reports related to 17,328 organizations. In most cases, Wooyun does not offer monetary rewards.

We choose Wooyun as one of the data sources for our study for several reasons. First, Wooyun insists on a delayed full disclosure policy, which states that the vulnerability will be disclosed 45 days after the submission of the report, irrespective of whether the organization has addressed the issue or not. To the best of our knowledge, it is the only platform that has such a disclosure policy. We will focus on this aspect in Section 5.1. Second, Wooyun covers the longest period of time and the largest number of contributions compared with other platforms (Table 1). It also includes a large number of organizations from several different sectors, as we will discuss in Section 4.3.3. This is because Wooyun has a very relaxed submission rule compared with other US-based platforms: white hats can submit a vulnerability report to Wooyun for almost any organization, and Wooyun will publish it as long as the report is considered valid.

We crawled the vulnerability reports on Wooyun published from May 2010 to early August 2015. For each vulnerability report, we collected the following data fields: (1) white hat's registration name, (2) target organization, (3) vulnerability type, (4) severity and (5) submission time. We further explain key data types below.

*Vulnerability type*: Each vulnerability report on Wooyun has a vulnerability type from a predefined list. However, we also observe that for some reports, the vulnerability types used are not in the list, possibly due to mistakes. We manually corrected these instances. We also translated the types from Chinese into English and list them in Figure 5.

*Severity*: The severity level of a vulnerability reflects its impact on the target organization. There are three levels: high, medium and low. We mainly use the severity level assigned by the affected organization or by the Wooyun plat-

form. If this information is missing (e.g., when the organization does not respond to the report), we will use the severity level provided by the white hat reporter.

We have also collected the following data:

*Organization website's URL and Alexa rank*: To examine whether a website's popularity is related to vulnerability discovery, we collected the website's rank from the Alexa Top Sites service. Since Wooyun does not provide the URL for all organizations, we wrote a script that queries the organization's name on Google and takes the first result as the URL. We then retrieved the Alexa rank of all websites from the Alexa Top Sites service. Since most websites on Wooyun are Chinese, we use the Chinese Alexa rank, rather than the global rank.

*Organization sector*: We also categorized organizations into different sectors. The definition of sectors are based on previous studies [15, 6]. The categorization is initially based on patterns in the organization's name. For example, universities have names like "XX university" or "university of XX". After this step, we further manually categorized the remaining organizations that have received more than 40 vulnerabilities into different sectors.

Our dataset cannot contain all vulnerabilities discovered by white hats for organizations. First, due to the large volume of vulnerability reports received, Wooyun may ignore vulnerabilities that are considered irrelevant or of very low importance, such as many reflected XSS vulnerabilities [2]. The impact of this initial expert selection is ambiguous, but we expect that our analysis may benefit from a heightened focus on valuable contributions. Second, white hats are starting to use alternative platforms such as Vulbox which do not have a public disclosure policy. As Wooyun remains the dominant platform for Chinese website vulnerabilities, we anticipate that the latter effect is relatively small.

#### 3.2.2 HackerOne

HackerOne is a US-based bug bounty platform started in November 2013. As of early August 2015, it facilitates 99 public bug bounty programs for global companies such as Yahoo, Mail.ru and Twitter. Unlike Wooyun, white hats on HackerOne can only submit reports for these organizations. HackerOne also hosts invitation-only programs. To be eligible white hats must reach a reputation score threshold. Similar programs, such as BugCrowd also separate bounty programs into public and invitation-only [14]. Unfortunately, invitation-only programs cannot be accessed publicly, so our dataset only includes public programs.

Our HackerOne dataset includes contributions from 1,653 white hats. An organization can either reward white hats with reputation scores or monetarily compensate them. Unlike Wooyun, HackerOne does not have a delayed public disclosure policy. A vulnerability report can only be disclosed if both the white hat and the organization commit to its publication. As a result, only a small fraction (732 of 10,997) of all reports are publicly disclosed. For other reports, we only know limited metadata, including submission times, white hat identifiers, and the names of the affected organizations for each vulnerability. We are able to collect the metadata of 6,876 reports from public bounty programs in total. They constitute 62.5% of all resolved reports. We assume the remainder to be reports for invitation-only programs. In addition, HackerOne hosts bounty programs for several open source software projects, such as Perl, Python, OpenSSL. We exclude 69 reports for these bounty programs since they are not related to web vulnerabilities. Our data includes 3,886 reports with bounties paid during the study period. However, some organizations choose not to disclose the bounty amount; i.e., only 1,638 reports have exact monetary payment information. We calculate the average amount of monetary reward paid by an organization, and refer to this value as the *expected reward*.

# 4. RESULTS

## 4.1 Vulnerability Disclosure Trends

We first provide an overview of the disclosed vulnerabilities. Since the HackerOne dataset does not include data about the vulnerability type and severity, we will mainly focus on the Wooyun dataset.

#### 4.1.1 Number of Vulnerabilities

The number of vulnerabilities accepted by the bug bounty platforms provides an initial overview of the productivity of the web vulnerability discovery ecosystems, and also reflects the time trend of web security. Table 1 shows that each of the major bug bounty platforms has published a large number of vulnerability reports. Figure 2 further displays the number of vulnerabilities accepted by Wooyun and HackerOne every month. For Wooyun, the number of vulnerabilities accepted per month continues to grow rapidly in the 5-year span. After an initial growth, the number of vulnerabilities for HackerOne's public bounty programs is relatively stable at around 400 per month. We suspect that an inclusion of data for invitation-only programs would also result in an upward trend for the HackerOne trajectory.

#### 4.1.2 Severity Levels

We break down the overall vulnerability trend on Wooyun by severity in Figure 3. While the percentage of low severity vulnerabilities is decreasing, the percentage of published high severity reports is increasing over time. One known reason is the intentional omission of certain low severity reports, as we have discussed in Section 3.2.1. It is also possible that white hats are becoming more skilled in finding severe vulnerabilities over time. Another hypothesis is that low severity vulnerabilities are easier to discover and thus are usually reported well before more severe problems. Further investigation of these possible causes would be an interesting research question. Overall, the displayed trend indicates that organizations inside this ecosystem are still at risk, and more efforts from both the white hat community and the involved organizations are required.

#### 4.1.3 Vulnerability Types

We next examine vulnerability reports on Wooyun according to their types. Figure 4 shows the trend for the top 3 most common vulnerability types. While the percentage of XSS reports is decreasing (possibly due to filtering as mentioned previously), we observe a small relative increase of SQL injection reports. The high amount of XSS is expected for web applications; other platforms, such as BugCrowd, have also reported that XSS is the most common vulnerability type (17.9%) [14]. In contrast, the high amount of SQL injection vulnerabilities on Wooyun is particularly surprising, since SQL injection vulnerabilities are not common on other platforms such as BugCrowd (only 1.3%) [14]. A

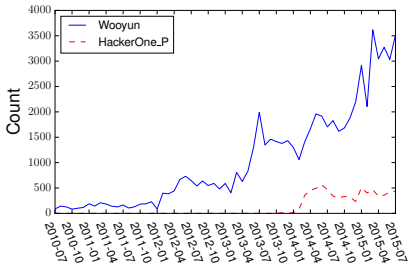


Figure 2: Number of vulnerabilities reported per month on Wooyun and HackerOne (public data).

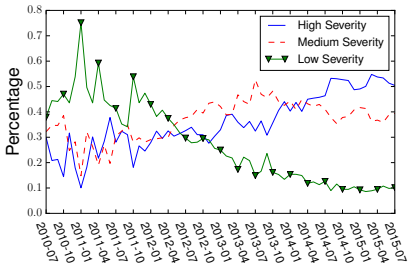


Figure 3: Trend of vulnerabilities with different severity on Wooyun.

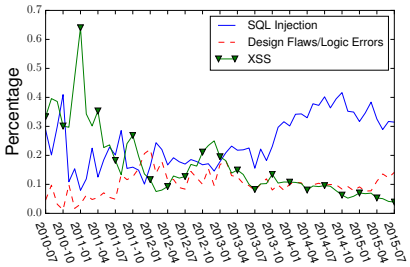


Figure 4: Trend of top 3 vulnerability types on Wooyun.

recent study also reveals that many Chinese websites are generally less secure [15]. However, the observed differences could also be caused by the particular organization participation model of Wooyun, which is able to cover much more poorly secured websites. We will discuss more on this in Section 5.4.

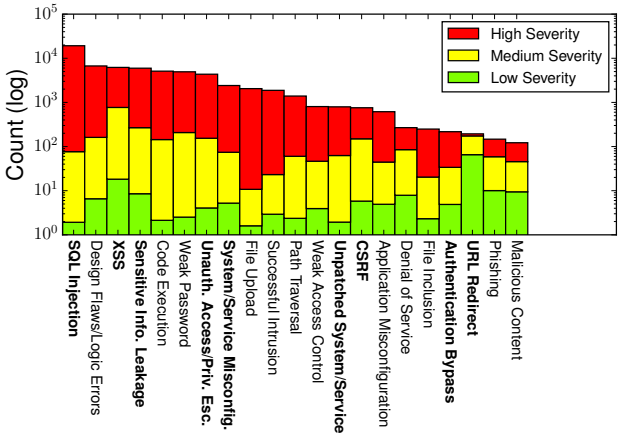


Figure 5: Number of reports for each vulnerability type on Wooyun (log scale). The compact visualization uses three colors to represent the percentage of three severity levels. Note that percentages are not affected by the log scale. Types in bold font also appear in OWASP's 2013 top 10 [1].

Figure 5 further shows the number of published reports, and the breakdown in severity categories for all vulnerability types on Wooyun. The distribution across vulnerability types is comparable to other sources [14, 1]. We also observe that some types have a larger proportion of high severity vulnerabilities; for example, SQL injection attacks and malicious file uploads may frequently open up a direct pathway to sensitive data.

In summary, data from bug bounty platforms can be used to meaningfully aggregate valuable security information. Disclosing such information, even at the aggregate level, can help the defense side to update its strategies and to allocate resources against different types of threats.

## 4.2 The White Hat Community

In this section, we first look at the size and growth of the white hat communities on Wooyun and HackerOne. Then, we discuss significant differences regarding productivity and accuracy among white hats using the two datasets. Next, we investigate different skills and strategies of white hats. Finally, we analyze how disclosure of reports can have positive effects on the white hat community.

### 4.2.1 *Size and Growth*

The outcome of a web vulnerability discovery ecosystem is closely related to the size of the white hat community, who is the "supplier" of vulnerability reports. Table 1 shows that these ecosystems have accumulated large white hat communities with tens of thousands of contributors, who may come from all over the world [14, 4]. Later, we will analyze how the size and the diversity within the white hat community correlate with vulnerability discovery outcomes.

We first examine how the size of the white hat community changes over time, using two metrics: the number of white hats who reported at least one vulnerability in each month (*active white hats*), and the number of white hats who submitted their first vulnerability in each month (*new white hats*). The difference between the number of active white hats and the number of new white hats is the number of repeat contributors. We report these two metrics for Wooyun and HackerOne in Figure 6. For Wooyun, the number of active white hats per month gradually grows to 700 per month. The number of new white hats per month is about 200 in the past 2 years, which means that there is a relatively constant flow of newcomers joining the ecosystem. The trend for the public programs of HackerOne is similar. In summary, both platforms attract a relatively constant number of white hats who contribute in a given month, while the overall size of the white hat community keeps increasing.

### 4.2.2 *Productivity and Accuracy*

While the size of the community matters, we also care about the individual productivity of a white hat, i.e., the number of vulnerabilities found by each white hat. In Figure 7, we plot the distribution of vulnerabilities found by individual white hats on both Wooyun and HackerOne. We observe that the distributions on both platforms are very skewed. Of 7,744 white hats on Wooyun, the top 1 has found 521 vulnerabilities, the top 100 have published more than 147 reports per person on average, but 3725 of the white hats have contributed only once. Similar observations can be made for white hats on HackerOne. Such long-tail pattern has also been found in other domains, such as scientific productivity [26].

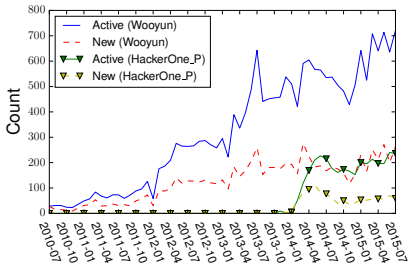


Figure 6: Number of new white hats and active white hats per month on Wooyun and HackerOne.

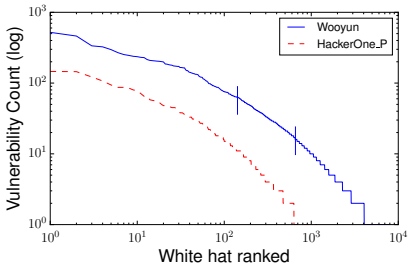


Figure 7: Contribution count of white hats on Wooyun and HackerOne (log-log). Vertical bars: thresholds for different productivity groups on Wooyun.

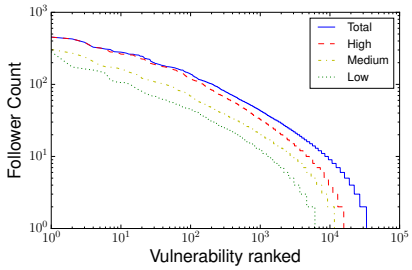


Figure 8: Distribution of follower count for vulnerabilities with different severity (log-log).

Another important aspect associated with productivity is accuracy. Many existing public bounty programs have complained about the low signal-to-noise ratio and the effort required to deal with a large amount of invalid reports, which generally include duplications, non-security issues, out-of-scope, false positives, or even spam [14, 9, 4, 8]. The signal-to-noise ratio is roughly 20% for platforms such as HackerOne and BugCrowd, and even lower for individually hosted bounty programs by Facebook and Github [14, 8]. In addition, HackerOne has reported that in general more productive researchers have a higher signal-to-noise ratio [8]. Based on [8], we estimate that the top 1% researchers on HackerOne have an average ratio of 0.54, while the bottom 50% only have an average ratio of 0.03, indicating that approximately among 100 reports submitted by them, only 3 are expected to be valid vulnerability reports. Bug bounty platforms have introduced various data-driven approaches, including reputation systems and rate limiting, to improve the signal-to-noise ratio [8]. This partly explains the higher signal-to-noise ratio of bounty platforms over individually hosted bounty programs. However, the low signal-to-noise ratio remains a key challenge for effective vulnerability discovery and requires more research effort.

The long-tailed distribution of contribution levels as well as concerns about accuracy lead to an increased focus on the top contributors in today's bug bounty programs, since they are on average much more productive, and more accurate. As a result, existing bounty platforms such as HackerOne and BugCrowd have created private bounty programs that only invite a small number of top contributors [14, 9, 8]. In some cases, the top contributors were directly hired by organizations or bounty platforms [4, 7].

Less attention is given to white hats with lower productivity. However, taken as a group, they contribute a sizable number of accepted reports. As such, the question arises how to evaluate their contributions. To do an initial comparative assessment, we split the white hat community on Wooyun into three groups of different levels of productivity. The two thresholds, displayed in Figure 7, are chosen so that the three groups have approximately the same number of reports, thus allowing us to compare other dimensions of their contributions.

We report the results in Table 2. Unsurprisingly, the average number of accepted reports differs substantially across these groups. In contrast, an interesting observation is that the less productive groups have contributed reports for a considerably larger number of organizations. There could be multiple reasons to explain this difference. First, the less productive groups have many more white hats, leading to more "manpower" and more diverse interests covering a wider range of websites. Meanwhile, white hats in the highly productive group have more limited attention or may benefit from an increased focus on a specific set of websites. Second, some websites may have been particularly popular targets for white hats, and easy-to-be-found vulnerabilities are already removed. For many low productive white hats who may also have limited expertise, spending effort on such websites might not be cost-effective. Thus, they shift their attention to other websites, which are more likely to yield discoveries.

| | Productivity Groups | | |
|---|---|---|---|
| **Variable** | **High** | **Medium** | **Low** |
| # white hats | 142 | 658 | 6,972 |
| Total # vuln. | 17,611 | 17,586 | 17,595 |
| Average # vuln. | 124 | 27 | 2.5 |
| # contributed org. | 4,727 | 5,686 | 7,247 |
| Alexa 1-200 (%) | 32.5 | 34.4 | 33.1 |
| Alexa 201-2000 (%) | 32.4 | 33.6 | 34.0 |
| Alexa > 2000 (%) | 33.7 | 32.9 | 33.3 |
| Severity High (%) | 38.4 | 33.5 | 28.1 |
| Severity Medium (%) | 31.3 | 32.5 | 36.1 |
| Severity Low (%) | 25.1 | 34.5 | 40.4 |

Table 2: Comparison across three white hat groups of different productivity levels on Wooyun.

The broader coverage of websites by less productive white hats has a positive impact on the security of the Internet, since even less popular sites still receive a considerable amount of visitors every day. In addition, the security of organizations is rather connected in many ways [22, 33]. For example, a user could use the same username and password across multiple sites, and the compromise of one of them will jeopardize others. Therefore, by complementing the limited attention of top white hats, the less productive white hat groups make different but important contributions.

We further break down the contributions of each group by target websites' popularity and by vulnerability severity. Rows 5 - 7 of Table 2 show that for the different popularity categories the contributions (in %) across the three productivity groups are remarkably consistent. In particular, the

least productive group also reports a significant percentage of discoveries for popular websites. Row 8 shows that more productive white hats have a larger percentage of contributions with high severity vulnerabilities, but 28.1% of high severity vulnerabilities were still discovered by the least productive white hats.

In summary, the results support the existence of a substantial expertise and productivity gap on an individual level, but from a collective perspective the difference is smaller than perhaps expected. How to better utilize the potential of these different groups of white hats is an interesting challenge. In particular, it would be useful to think about how to boost the productivity of less productive white hats through better incentives, training, and other measures.

### 4.2.3 Skills and Strategies

Next to productivity, we measure two additional metrics: the number of different organizations an individual white hat investigated, and the number of different vulnerability types an individual white hat reported. These two metrics partly reflect the skills, experiences and strategies of white hats. Figure 9 shows the distribution of these two metrics for white hats on Wooyun with more than 5 discoveries. The average number of organizations investigated by a white hat of this group is 18, while the average number of vulnerability types found is 7. The most productive individuals (i.e., red triangles in the figure) generally surpass others in both metrics which partially explains the productivity difference. First, top white hats' broad knowledge of different types of vulnerabilities may enable them to discover more vulnerabilities. Second, they may find more vulnerabilities because their strategy is to investigate a larger number of websites. Furthermore, we hypothesize that there is a trade-off between exploration vs. exploitation: to find more vulnerabilities, a white hat must develop a good balance between spending effort at one particular website and exploring opportunities on other sites. However, different successful strategies coexist. For example, our dataset includes several white hats in the bottom left corner of Figure 9 that is much more focused on exploitation. Similarly, a white hat named 'meals' ranked 4th on HackerOne only focuses on Yahoo's bounty platform, and has to-date found 155 vulnerabilities.

Investigating the optimal degree of strategy diversification during web vulnerability hunting is an interesting area for future work.

### 4.2.4 Disclosure and Learning

A primary consideration of previous research was to understand how vulnerability disclosure pushes software vendors to fix flaws in their products [35, 36]. However, when considering the whole ecosystem, we question whether vulnerability disclosures also have positive effects on the white hat community itself. One possible effect is to enable white hats to learn valuable technical insights and skills from others' findings. Another effect is to obtain valuable strategic information for their own vulnerability discovery activities, such as which organizations to investigate. Both effects likely improve white hats' productivity and accuracy. In addition, the software engineering community and peer organizations can also learn valuable lessons from vulnerability reports to avoid making similar mistakes in the future. While the latter factor may be of high practical relevance,

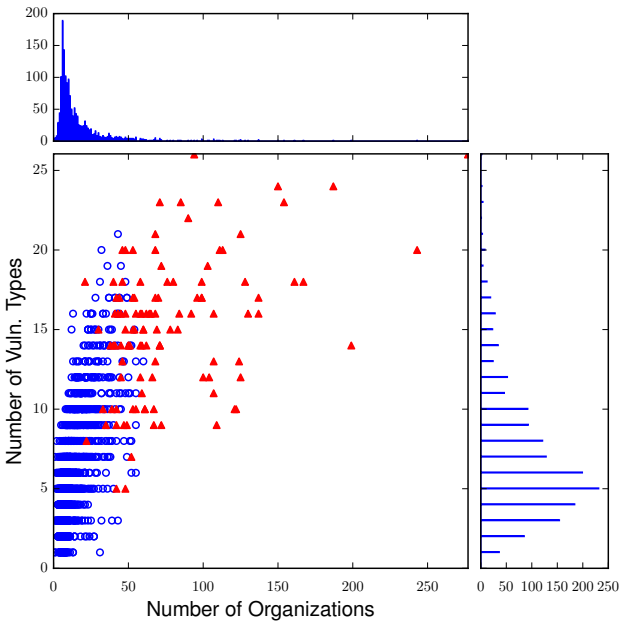


Figure 9: Scatter plot of white hats' vulnerability type count and targeted organization count on Wooyun. Each dot represents a white hat who has found more than 5 vulnerabilities in total. The red triangle dots are white hats of the high productivity group defined in Section 4.2.2.

we are unaware of related research. In this paper, we investigate the first effect using data from Wooyun.

The Wooyun platform allows white hats to mark and follow a particular report. Therefore, we can use the number of followers of a vulnerability report as an approximate indicator of its learning value to white hats. In Figure 8, we plot the distribution of this follower count for all vulnerabilities on Wooyun, and also break down the data by different severity levels. We observe that the distribution is very skewed. There are 9,489 reports that have at least 10 followers, indicating that white hats have been actively learning from a broad portion of reports. On average, high severity vulnerabilities have more followers, which is not surprising, as more severe vulnerabilities tend to have a more significant security impact, and higher discovery and exploit complexity. What might be counter-intuitive is that some low severity vulnerabilities still receive more than 100 followers.

To examine why some vulnerabilities have received much more attention than others, and why some low severity vulnerabilities are followed by many, we selected the 30 most followed vulnerabilities from each severity level. We then manually examined these 90 vulnerabilities. We find that these vulnerabilities mostly belong to one or more of the following categories: (1) Vulnerabilities with significant impact (e.g., with a potential for massive user data leakage, or an XSS inside the site statistics javascript code from a major search engine company); (2) Vulnerabilities that are associated with novel discovery or exploitation techniques; (3) Vulnerabilities of widely used web applications, such as CMS; (4) Vulnerabilities that are explicitly organized as tutorials. We found 21 such tutorial-style reports belonging to a series about XSS, which are all of low severity, yet they

still receive a lot of attention because of the emphasis on learning. We also examined a subset of disclosed reports from HackerOne and have discovered that some organizations make disclosures[3] to teach the writing of concise reports.

In summary, our analysis provides evidence of how white hats are learning from vulnerability reports; a typically overlooked benefit of vulnerability disclosure to the white hat community. We will discuss additional facets of disclosure in Section 5.1.

## 4.3 Organizations

We now shift our focus to the organizations who have participated in vulnerability discovery ecosystems. These organizations harvest vulnerability reports from the white hat community, fix security flaws, and thereby ultimately improve the security of the whole Internet (e.g., by reducing the impact of security interdependencies [22, 33]). However, collecting data about them is non-trivial because many organizations, such as banks, are still reluctant to collaborate with white hats due to various concerns [3]. In addition, for many organizations who joined platforms such as HackerOne, data about discovered vulnerabilities, monetary rewards and other important factors is often not publicly disclosed.

Wooyun provides a valuable opportunity to study the impact of such ecosystems on organizations; and not only because of the existence of the delayed public full disclosure policy. More importantly, an organization is rather coerced to join this ecosystem once a white hat publishes a vulnerability on Wooyun affecting the organization. This *coercive model* is different from most other platforms which only host bounty programs for organizations that agree to participate (i.e., *voluntary model*). Due to the diversity of the large white hat community, Wooyun covers a broad range of organizations from many sectors, as we will show in Section 4.3.3. As a result, observations made from this dataset do not only help us understand the web vulnerability discovery ecosystem in China, and the general security status of the Chinese web, but also help us to envision the impact of the bug bounty model for organizations in other parts of the world.

### 4.3.1 Size and Growth

Table 1 lists the number of organizations participating in representative vulnerability discovery ecosystems. We observe that Wooyun affects a larger number of organizations compared with US-based platforms, who typically have tens or hundreds of participating organizations. The difference is partly due to the coercive versus voluntary ways of involving organizations. Therefore, the Wooyun ecosystem roughly represents an upper bound of coverage (growth) for other ecosystems. We also investigate the trajectory of the growth of the number of organizations covered on Wooyun. Figure 10 shows that in every month, there are about 300 organizations benefiting from white hats' efforts. Around 150 of them are new organizations, which implies that the white hat community is continuously broadening its horizon. It would be interesting to understand whether this effect relies on the fact that new businesses are founded (or new websites become public), or that white hats are moving to already established but previously unresearched websites.

[3]For example: `https://hackerone.com/reports/32825`.

### 4.3.2 Vulnerability Distribution

For both Wooyun and HackerOne, Figure 11 shows that only few organizations receive a high number of vulnerability reports, while most organizations receive very few vulnerability reports. We hypothesize that the number of vulnerabilities received by organizations is related to multiple factors, such as the complexity of the web system, the existence of monetary incentives, the popularity of the website, etc. We will further investigate the relation between these factors and the number of published vulnerability reports in Section 4.3.7.

### 4.3.3 Impact on Different Sectors

To investigate the diversity within participating organizations, we have manually tagged organizations on HackerOne based on their business types. We find that *all* participating companies are IT-focused and cater to different business/consumer needs which are shown in Table 3.

| |
|---|
| social network (13), security (9), content sharing (9) |
| payment(8), communication (8), bitcoin (6), |
| cloud (5), customer management (5), |
| site builder (5), finance (4), ecommerce (4) |

Table 3: Frequency of IT-business types within the group of publicly available bounty programs on HackerOne. Only tags with frequency greater than 3 are shown.

Due to its coercive model for involving organizations, the Wooyun dataset includes a larger and more diverse set of organizations (see Figure 12). Further, it shows that white hats do not exclusively focus on certain business sectors.

For non-IT organizations, two sectors with many vulnerability reports are government and finance. We consider this finding surprising since these sectors have robust incentives for security investments. While the finance sector, and possibly the government sector as well, are often not willing to collaborate with non-commercial white hats [3], we infer from the Wooyun data that they can disproportionally benefit from the involvement of the white hat community. Participating in disclosure programs may also reduce the likelihood that vulnerabilities flow into the black market [21].

Portal sites, telecommunication and e-commerce organizations have the highest number of vulnerability reports in the IT-sector. A possible explanation is that web services and systems from these domains are large and complex which increases the amount of latent vulnerabilities. Further, these companies serve substantial user populations which increases their desirability for vulnerability researchers.

### 4.3.4 Response and Resolution

After the initial submission of a vulnerability report, the typical follow-up process on most bounty platforms contains the following steps: triage/confirm, resolve, and disclose. During this process, the white hat and the security/development team of the organization may collaborate together to address the identified problem. A delayed response likely increases the risk of a security breach, since it increases the time frame for rediscovery of the vulnerability, and stealthy exploitation of stockpiled vulnerabilities by malicious agents. Given its full disclosure policy, a delayed response to a submission on Wooyun may be even more serious because details will be disclosed publicly after 45 days.

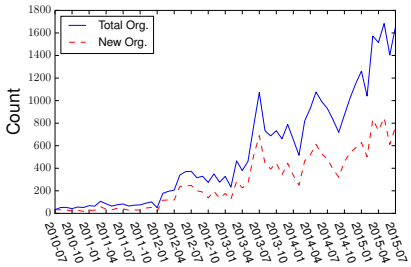


Figure 10: Count for new and total number of organizations with vulnerability reports (per month) on Wooyun.

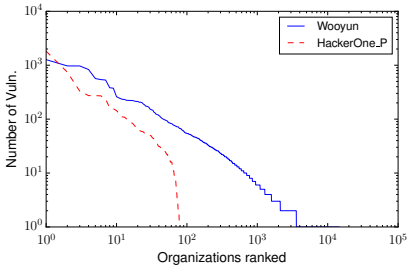


Figure 11: Number of vulnerabilities by organizations on Wooyun and HackerOne.

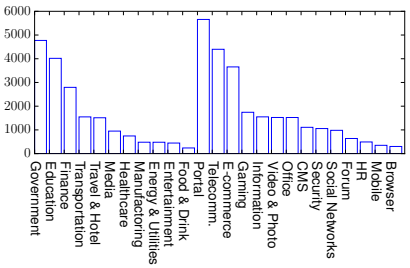


Figure 12: Number of vulnerability reports for non-IT businesses (left), and IT-sector businesses (right) on Wooyun.

Our data allows us to examine how organizations respond and resolve vulnerability reports in the studied ecosystems. HackerOne maintains a detailed handling history for each vulnerability report. Unfortunately, only a small portion of all resolved reports (732 of 10,997) are publicly disclosed. For these disclosed reports, we determined the time distribution for three types of response activities (see Figure 13). The median time for the first response (e.g., a confirmation of receiving the report) is 0.18 days, and the median time for triage is 0.88 days. The median resolve time is 6.49 days, and 75% of the disclosed reports are resolved in 25 days. However, one should be cautious when generalizing from these observations since the data is possibly biased. Particularly, the analysis likely underestimates the time required for triaging and resolving vulnerabilities, since the organizations that are willing to disclose vulnerabilities may be more efficient in handling reports and may have more experience in running bounty programs.

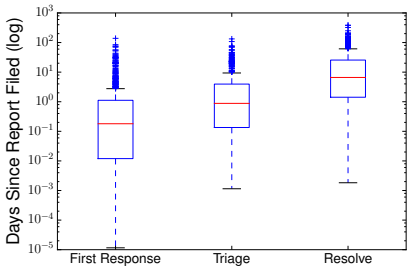


Figure 13: Boxplots for the time of three types of response activities based on publicly disclosed reports on HackerOne.

Wooyun shows four types of responses by organizations: confirmed by organization (CO), confirmed and handled by a third party such as CNCERT (CT), ignored by the organization (IG), and no response (NO). Since all reports are classified in this way, the Wooyun response data is considerably larger, but provides less details. For example, it is difficult to discern whether the organization eventually fixed the vulnerability (or not), but the first two types of response can serve as an indication that the organizations recognizes the problem. The third type of response suggests that the organization considers the vulnerability report invalid. The fourth type means that the organization did not respond to the report at all. We use the count of the fourth type as a rough estimate for the number of cases when an organization fails to address a vulnerability report, and consider the other three types of responses as situations when the vulnerability is likely being handled.

| | CO | CT | IG | NO |
|---|---|---|---|---|
| Overall (%) | 40 | 34 | 3 | 23 |
| Organizations: | | | | |
| - Alexa 1 - 200 (%) | 71 | 13 | 5 | 12 |
| - Alexa 201 - 2000 (%) | 57 | 18 | 4 | 20 |
| - Alexa > 2000 (%) | 28 | 44 | 1 | 26 |

Table 4: Percentages of different types of responses by organizations on Wooyun.

Table 4 shows the percentages for the different types of response as a breakdown by the popularity of the websites. We observe that overall, the majority (77%) of the vulnerability reports have been handled. Popular websites address more vulnerabilities by themselves, while less popular websites rely more often on third parties. In addition, less popular websites have a higher rate of no response, possibly due to limited resources for vulnerability management.

### 4.3.5 *Monetary Rewards*

We also examine the role of monetary rewards offered by some organizations. We observe that in their absence, white hats still make contributions to Wooyun and HackerOne for the purpose of making the Internet safer and for reputation gains. For example, 33 of the public programs on HackerOne do not provide monetary rewards, yet they still have received 1201 valid reports from the white hat community. But as Table 1 shows, most platforms offer monetary rewards as an additional incentive for white hats to contribute their time and expertise.

We conduct a preliminary analysis based on the disclosed bounties for public programs on HackerOne. Given a total of 3886 bounties, 1638 have the amount information disclosed. The maximum bounty is $7560, paid by Twitter, and the average bounty amount is $424 which varies considerably by organizations. Yahoo pays $800 on average, followed by Dropbox ($702) and Twitter ($611). We hypothesize that the current reward level is attractive to many white hats, and we explore this topic in more detail with a regression study in Section 4.3.7.

### 4.3.6 *Improvements to Organizations' Web Security*

The participation in a bug bounty program should over time improve the web security of an organization in a noticeable way. In particular, it is reasonable to expect that the

number of latent vulnerabilities in an average organization's web systems (and the stockpile of web vulnerabilities held by black hats) would gradually diminish. Our data allows us to investigate whether the number of vulnerability reports per month is changing over time which is a relevant metric in this context. Moreover, it is a type of analysis that can be conducted by external evaluators if the bug bounty program is public, and provides stakeholders an indication of the web security of an organization. For example, the Cobalt bounty platform offers security seals for organizations that use their services, which is expected to improve the public perception of the organization's security [37]. Other services (e.g., cyber-insurance companies), can also benefit from such security assessments (e.g., for the determination of insurance premiums) [23, 25].

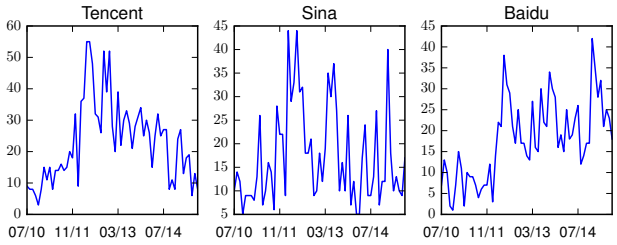


Figure 14: Trend of vulnerability report count for three organizations on Wooyun.

To initially explore this question, we show the vulnerability report trends for three large organizations on Wooyun in Figure 14. While one notices a slight decreasing trend for Tencent, it is hard to observe a clear tendency for the other two organizations. More importantly, Wooyun may not exclusively host these organizations' bounty programs which could influence the analysis.

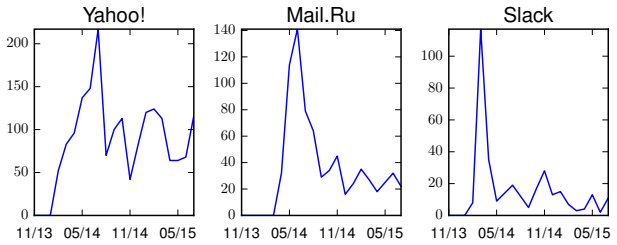


Figure 15: Trend of vulnerability report count for three organizations on HackerOne.

In contrast, HackerOne is tasked to exclusively host bounty programs for participating organizations which ensures a more reliable analysis. We show the vulnerability trends for the three organizations with the most vulnerabilities on HackerOne (Figure 15). Interestingly, these organizations have received a large volume of vulnerability reports right after the launch of their bounty programs. We propose three possible explanations. First, the monetary compensation offered by HackerOne provides stronger incentives for white hats to compete for vulnerability discoveries in the early stage of a bounty program since the bounty program only rewards the first discoverer. Second, the target range for white hats on HackerOne is much more limited compared to Wooyun, thus concentrating white hats' focus. Third, some white hats might have stockpiled vulnerabilities to offload them for reward in anticipation of the opening of new reward programs. After these initial spikes, the number of vulnerability reports on HackerOne drops significantly, possibly because the difficulty of finding new vulnerabilities is increasing. However, even though we observe decreasing trends, these organizations still receive a positive number of vulnerability reports every month. These additional discoveries may either be related to further latent vulnerabilities in existing code or stem from new code. Therefore, we suggest that organizations continuously collaborate with white hats.

To further examine the vulnerability trends for organizations, we apply the Laplace test [32] to the vulnerability history of organizations who have received at least 50 reports and have a bounty program for more than 4 months. We also excluded data before 2012-02 and 2014-02 (the initial growth periods), for Wooyun data and HackerOne data, respectively. This test indicates whether there is an increasing trend, a decreasing trend, or no trend for the number of reported vulnerabilities for a given organization (Table 5).

| Platform | Decrease | Increase | No Trend |
|---|---|---|---|
| Wooyun | 11 | 81 | 17 |
| HackerOne | 32 | 8 | 9 |

Table 5: Trend test results for organizations on Wooyun and HackerOne. The confidence level is 0.95.

Only 11 of the 109 organizations on Wooyun (which match the criteria) fit a decreasing trend, while most selected organizations have an increasing trend for the number of vulnerability reports. The data omission bias discussed previously could be one reason of the result. A sufficiently large pool of latent vulnerabilities in combination with increasing activity on Wooyun could serve as an alternative explanation. For organizations on HackerOne, 32 of 49 have a decreasing trend indicating a positive effect of the vulnerability discovery ecosystem.

The trend test, however, cannot completely assess the web security status of an organization for several reasons which we have partly discussed above. Further, as a possible part of their vulnerability discovery strategy (see Section 4.2.3), white hats might switch to new organizations or newly deployed web systems which are expected to have more low hanging fruits. In general, we suggest that a reliable assessment requires careful modeling and statistical analysis of the whole ecosystem which is an important area for future work.

### 4.3.7 *Attracting Vulnerability Reports*

How can an organization harvest more vulnerability reports from the white hat community to improve its web security? To address this question, we first study the correlation between the number of vulnerability reports per organization and the number of contributing white hats.

Figure 16 plots the number of white hats that have made at least one discovery, and the number of vulnerabilities, for each organizations (with at least 20 vulnerability reports) on Wooyun and HackerOne. We observe very strong positive linear (Pearson) correlations for these measures (as shown in the figures). Therefore, the following strategies are likely

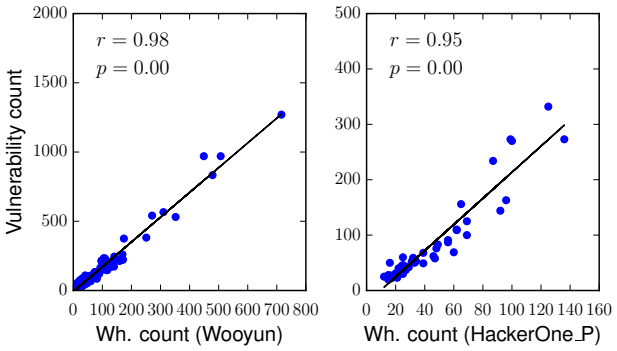


Figure 16: Scatter plots of organizations' white hat count and vulnerability count for Wooyun and HackerOne public programs (excluded Yahoo and Mail.ru as outliers).

| VARIABLES | (1) # Vuln. | (2) # Vuln. | (3) # Vuln. |
|---|---|---|---|
| Expected Reward ($R_i$) | 0.04*** | 0.03*** | 0.03*** |
| | (0.01) | (0.01) | (0.01) |
| Alexa [log] ($A_i$) | | -2.52* | -2.70** |
| | | (1.20) | (1.21) |
| Platform Manpower ($M_i$) | | | 10.54 |
| | | | (10.14) |
| Constant | 3.21* | 16.12** | -133.05 |
| | (1.88) | (6.39) | (143.66) |
| R-squared | 0.35 | 0.39 | 0.40 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Results of regression analysis. There are 60 observations (HackerOne).

beneficial: (1) While paying special attention to top contributors is a useful strategy, it is also important to increase the total number of contributors. A possible reason to explain the observed effect is that vulnerability discovery requires diversity, i.e., investigators with different expertise using different tools may find different vulnerabilities; (2) It is important to incentivize new participation, for example, by offering an extra bonus (e.g., badge or money) for the first valid submission of a white hat to a platform or specific program.

Other factors such as the popularity of the target, the expected bounty amount, and the number of alternative choices are all related to a bounty program's attractiveness to white hats. To better understand these factors, we conduct a linear regression by taking the number of vulnerability reports as the dependent variable and other factors as independent variables, as the following equation shows:

$$V_i = \beta_0 + \beta_1 R_i + \beta_2 A_i + \beta_3 M_i + \epsilon_i$$

where for each organization, $V_i$ is the average number of vulnerabilities per month, $R_i$ is the expected reward, $A_i$ is the log Alexa rank of $i$'s website, and $M_i$ is the average platform manpower during the lifetime of organization $i$'s bounty program. $M_i$ is defined as the time-weighted number of white hats divided by the time-weighted number of peer organizations during the lifetime of $i$'s bounty program:

$$M_i = \frac{NW_1 T_i + \sum_{k=2}^{T_i}(NW_k - NW_{k-1})(T_i - k + 1)}{NO_1 T_i + \sum_{k=2}^{T_i}(NO_k - NO_{k-1})(T_i - k + 1)}$$

Here, $T_i$ is the number of months for $i$'bounty program. $NW_k$ and $NO_k$ are the accumulated number of white hats and the number of peer organizations on the whole platform at the $k$th month for organization $i$, respectively.

Table 6 shows three variations of the regression model. In all three models, we find a highly significant positive correlation between the expected reward offered and the number of vulnerabilities received by that organization per month. Roughly speaking, a $100 increase in the expected vulnerability reward is associated with an additional 3 vulnerabilities reported per month. We also find a significant negative correlation between the Alexa rank and the number of vulnerabilities in models (2) and (3) suggesting that rank determines the attractiveness of a website to white hats. However, it is also possible that less popular websites are in general less complex in design and implementation, and thus contain less vulnerabilities. For model (3), we expect that with higher average platform manpower, an organization will receive more attention from white hats and thus will have more vulnerability reports. However, the analysis does not yield a conclusive answer, possibly due to the omission of invitation-only programs and limited sample size.

The quantified model can be used by organizations when determining their bug bounty policies and attracting an effective white hat following. In particular, offering higher rewards and running the program for a longer time contributes to a higher number of reports. The model also contributes to the security assessment question in Section 4.3.6. Nevertheless, our regression model is only a first step towards modeling the dynamics of the web vulnerability discovery ecosystem. It could be extended with more independent variables, such as the business type of organizations (see Section 4.3.3), or the expected rewards from peer organizations in the ecosystem.

# 5. DISCUSSION

## 5.1 Importance of Disclosure

Based on our analysis, we believe that disclosing important information about vulnerability discovery (such as the resolve time for each vulnerability, bounty amounts, and even the detailed reports) is important for the success of a web vulnerability discovery ecosystem. For the white hat community, disclosing more vulnerability information not only enables them to learn and improve, but also potentially allows to make better decisions on target selection, as we have discussed in Section 4.2.4. The transparency associated with disclosure could also reduce conflicts between organizations and white hats on issues like the validity of a report or the reasonableness of a bounty amount. For organizations, disclosing more information enables the public (e.g., Internet users, or cyber-insurance providers) to better assess the security of an organization (Section 4.3.6). Disclosure is also vital for the research community to tackle some of the challenging issues and future research questions we have discussed. In addition, a platform such as Wooyun

with a delayed full disclosure policy also pushes organizations to fix their reports sooner.

However, there are also potentially less desirable consequences of disclosing vulnerabilities about organizations' web systems, such as the leakage of critical information that can be utilized by black hats. An ideal disclosure policy has to balance the potential benefits and disadvantages to the ecosystem or a specific organization. Several disclosure programs are moving towards this direction. For example, some programs on HackerOne disclose only a subset of their vulnerabilities to the public. The Github bounty program discloses data about every vulnerability discovered by white hats, yet intentionally redacts certain details. Further analyzing the benefits and risks of disclosing vulnerability information, and designing improved disclosure policies is important future work.

## 5.2 Potential Incentive Structure Evolution

Our study shows that monetary incentives increase the number of vulnerability reports (Section 4.3.7). We anticipate that more organizations will start paying bounties, as more organizations are joining vulnerability disclosure ecosystems and are competing for the limited attention of white hats. The amount of an average bounty will likely rise not only for the purpose of attracting more white hats, but also for compensating the increasing cost incurred by white hats to discover vulnerabilities (e.g., to compete with black hats). In addition, high reward amounts can also be a positive signal of an organization's security practices to the public, similar to the proposal in [30].

Many organizations will continue to not offer bounties. For small organizations with limited revenues, maintaining a competitive bounty level could be challenging. It has also been suggested that an organization could start with no bounty first, and gradually increase the reward level, to alleviate the initial surge of reports (including many invalid submissions) [27], as we have shown in Figure 15. Organizations that cannot afford paying bounties can resort to other forms of incentives, such as reputation scores, hall-of-fame memberships, or even public disclosure.

## 5.3 Encouraging White Hat Participation

Increasing the size of the white hat community allows more organizations to be covered and more vulnerabilities to be found (see Section 4.2.2). A larger white hat community might also decrease the cost of running bounty programs for organizations, similar to the increase of supply in any economic market. Therefore, potential regulations that hinder the collaboration between white hats and organizations are likely detrimental. One such example is the proposed update of the Wassenaar Arrangement, which aims to control the export of intrusion software. The utilized overly broad definition of intrusion software could easily limit the participation of white hats [29], particularly considering the global nature of the white hat community [4, 14].

To encourage more white hats to join the ecosystem, organizations can try to offer a first time bonus (see Section 4.3.7), organize capture-the-flag activities, etc. In addition, by analyzing the behavioral patterns and dynamics of white hats (e.g., Section 4.2.3), bounty platforms can design customized services for white hats, such as target selection or recommender systems, which match white hats' skills and organizations' requirements. For this purpose, it would be helpful to further investigate vulnerability discovery by white hats (e.g., tool usage) through interview or survey studies [19].

## 5.4 Stimulate Participation by Organizations

Our study results provide incentives for organizations to join vulnerability discovery ecosystems and to benefit from white hats' efforts. In addition, government agencies such as the Federal Trade Commission also encourage organizations to have a process of receiving and addressing vulnerability reports [11], which can be achieved by running a bounty program.

In our work, we contrast two participation models from the organizations' perspective. The first one is the *coercive participation model*, represented by China-based platforms such as Wooyun. That is, an organization is coerced to join the ecosystem once a white hat has submitted a vulnerability for that organization. The second participation model, represented by US-based platforms including HackerOne, is voluntary, i.e., companies explicitly authorize external researchers to study the security of their web systems. Both models have their advantages and disadvantages. Our results show that the first model is capable of covering a wider range of organizations, although varying legal conditions in different countries might not allow for such an approach (see, for example, [28]). Also, this model might allow many severe vulnerabilities to be found earlier in websites that are not willing to participate bug bounty and are poorly secured. This could partly explains the high percentage of SQL injection on Wooyun in Section 1. The coercive model might be more attractive to white hats, for example, since they may feel more in control. The second model clearly grows more slowly when considering the number of participating organizations. However, voluntary participation likely encourages a better response behavior to vulnerability reports, as we have discussed in Section 4.1.3. To encourage organizations to participate in the voluntary model, future work is needed to identify and address organizations' concerns including the perceived lack of trustworthiness of the white hat population [3], misuse of automated vulnerability scanners, and time wasted due to false reports [8].

# 6. CONCLUSION

In this paper, we have studied emerging web vulnerability discovery ecosystems, which include white hats, organizations and bug bounty platforms, based on publicly available data from Wooyun and HackerOne. The data shows that white hat security researchers have been making significant contributions to the security of tens of thousands of organizations on the Internet.

We conducted quantitative analyses for different aspects of the web vulnerability discovery ecosystem. Based on our results, we suggest that organizations should continuously collaborate with white hats, actively seek to enlarge the contributor base, and design their recognition and reward structure based on multiple factors. We have also proposed future work directions to help to increase the impact and coverage of these ecosystems.

## Acknowledgments

We thank the anonymous reviewers for their helpful comments. The authors would also like to thank Yue Zhang, Aron Laszka, Kai Chen and Zhaohui Wu for their valuable comments on earlier drafts of this paper. Peng Liu was supported in part by ARO W911NF-09-1-0525 (MURI), CNS-1422594, and ARO W911NF-13-1-0421 (MURI).

## 7. REFERENCES

[1] OWASP 2013 Top 10. www.owasp.org/index.php/Top_10_2013-Top_10.

[2] Updates on vulnerability handling process. www.wooyun.org/notice.php?action=view&id=28, 2013.

[3] Banks reluctant to use 'white hat' hackers to spot security flaws. NPR, 2014.

[4] Bug bounty highlights and updates. Facebook, 2014.

[5] How Bugcrowd uses crowdsourcing to uncover security flaws faster than the bad guys do (Interview). VentureBeat, 2014.

[6] Website security statistics report. White Hat Security, 2014.

[7] CSUS student hunts for computer bugs as a 'white hat'. www.sacbee.com/news/business/article5014716.html, 2015.

[8] Improving signal over 10,000 bugs. https://hackerone.com/blog, 2015.

[9] LinkedIn's private bug bounty program: Reducing vulnerabilities by leveraging expert crowds. security.linkedin.com, 2015.

[10] Small business website statistics. www.statisticbrain.com/small-business-website-statistics/, 2015.

[11] Start with security: A guide for business. FTC, 2015.

[12] A. Algarni and Y. Malaiya. Software vulnerability markets: Discoverers and buyers. *International Journal of Computer, Information Science and Engineering*, 8(3):71–81, 2014.

[13] R. Böhme. A comparison of market approaches to software vulnerability disclosure. In *Emerging Trends in Information and Communication Security*. 2006.

[14] BugCrowd. The state of bug bounty, 2015.

[15] P. Chen, N. Nikiforakis, L. Desmet, and C. Huygens. Security analysis of the Chinese web: How well is it protected? In *Workshop on Cyber Security Analytics, Intelligence and Automation*, 2014.

[16] A. Doupé, M. Cova, and G. Vigna. Why Johnny can't pentest: An analysis of black-box web vulnerability scanners. In *Detection of Intrusions and Malware, and Vulnerability Assessment*, 2010.

[17] A. Edmundson, B. Holtkamp, E. Rivera, M. Finifter, A. Mettler, and D. Wagner. An empirical study on the effectiveness of security code review. In *Engineering Secure Software and Systems*, 2013.

[18] S. Egelman, C. Herley, and P. van Oorschot. Markets for zero-day exploits: Ethics and implications. In *New Security Paradigms Workshop*, 2013.

[19] M. Fang and M. Hafiz. Discovering buffer overflow vulnerabilities in the wild: An empirical study. In *8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 2014.

[20] M. Finifter, D. Akhawe, and D. Wagner. An empirical study of vulnerability rewards programs. In *USENIX Security Symposium*, 2013.

[21] S. Frei, D. Schatzmann, B. Plattner, and B. Trammell. Modeling the security ecosystem - The dynamics of (in)security. In *Economics of Information Security and Privacy*, 2009.

[22] J. Grossklags, N. Christin, and J. Chuang. Secure or insure? A game-theoretic analysis of information security games. In *17th International Conference on World Wide Web*, 2008.

[23] B. Johnson, R. Böhme, and J. Grossklags. Security games with market insurance. In *Decision and Game Theory for Security*, 2011.

[24] K. Kannan and R. Telang. Market for software vulnerabilities? Think again. *Management Science*, 51(5):726–740, 2005.

[25] A. Laszka and J. Grossklags. Should cyber-insurance providers invest in software security? In *20th European Symposium on Research in Computer Security*, 2015.

[26] A. Lotka. The frequency distribution of scientific productivity. *Journal of Washington Academy Sciences*, 16(12):317–323, 1926.

[27] R. McGeehan and L. Honeywell. Bounty launch lessons. medium.com/@magoo/bounty-launch-lessons-c7c3be3f5b, 2015.

[28] E. Messmer. Hacker group defies U.S. law, defends exposing McAfee website vulnerabilities. *Network World*, 2011.

[29] K. Moussouris. You need to speak up for internet security. Right now. Wired, 2015.

[30] A. Ozment. Bug auctions: Vulnerability markets reconsidered. In *Workshop on the Economics of Information Security*, 2004.

[31] A. Ozment. The likelihood of vulnerability rediscovery and the social utility of vulnerability hunting. In *Workshop on the Econ. of Information Security*, 2005.

[32] A. Ozment and S. Schechter. Milk or wine: Does software security improve with age? In *USENIX Security Symposium*, 2006.

[33] S. Preibusch and J. Bonneau. The password game: Negative externalities from weak password practices. In *International Conference on Decision and Game Theory for Security*, 2010.

[34] E. Rescorla. Is finding security holes a good idea? *IEEE Security & Privacy*, 3(1):14–19, 2005.

[35] G. Schryen. Is open source security a myth? *Communications of the ACM*, 54(5):130–140, 2011.

[36] M. Shahzad, M. Shafiq, and A. Liu. A large scale exploratory analysis of software vulnerability life cycles. In *International Conference on Software Engineering*, 2012.

[37] T. Van Goethem, F. Piessens, W. Joosen, and N. Nikiforakis. Clubbing seals: Exploring the ecosystem of third-party security seals. In *ACM Conference on Computer and Communications Security*, 2014.

[38] M. Zhao, J. Grossklags, and K. Chen. An exploratory study of white hat behaviors in a web vulnerability disclosure program. In *Proceedings of the 2014 ACM Workshop on Security Information Workers*, 2014.