Evaluating CVSS Base Score Using Vulnerability Rewards Programs

Awad Younis, Yashwant K. Malaiya, Indrajit Ray

Evaluating CVSS Base Score Using Vulnerability Rewards Programs

Awad Younis, Yashwant K. Malaiya, and Indrajit Ray

Colorado State University
Fort Collins, CO, 80523, USA
{awad,malaiya,indrajit}@CS.ColoState.EDU

Abstract. CVSS Base Score and the underlying metrics have been widely used. Recently there have been attempts to validate them. Some of the researchers have questioned the CVSS metrics based on a lack of correlation with the reported exploits and attacks. In this research, we use the independent scales used by the vulnerability reward programs (VRPs) to see if they correlate with the CVSS Base Score. We examine 1559 vulnerabilities of Mozilla Firefox and Google Chrome browsers. The results show that there is a significant correlation between the VRPs severity ratings and CVSS scores, when three level rankings are used. For both approaches, the sets of vulnerabilities identified as Critical or High severity vulnerabilities include a large number of shared vulnerabilities, again suggesting mutual conformation. The results suggest that the CVSS Base Score may be a useful metric for prioritizing vulnerabilities, and the notable lack of exploits for high severity vulnerabilities may be the result of prioritized fixing of vulnerabilities.

Keywords: CVSS Base Score. Vulnerability Reward Program. Bug Bounty Programs. Software Vulnerability Severity. Vulnerability Exploitation. Software Defensive Techniques. Exploit Mitigation Techniques.

1 Introduction

Assessing the risk associated with individual software vulnerabilities is accomplished by assessing their severity. CVSS Base score is the de facto standard that is currently used to measure the severity of vulnerabilities [1]. Evaluating the accuracy of the CVSS Base score is very important as they are intended to help decision makers in resource allocation, patch prioritization, and risk assessment. The lack of evaluation makes CVSS usability risky and that could lead to a waste of limited resources available or a breach with a high impact.

Vulnerability exploitability can be affected by the existence of defense techniques. The main objective of the defense techniques is to reduce the likelihood that the efforts of attackers will succeed [2]. As 92% of reported vulnerabilities are in software not in networks [3], recently vendors such as Microsoft, Cisco, Google, Mozilla, etc. have added new defensive layers at the software level. Among them

are secure development lifecycle, vulnerability mitigation techniques, sandbox, and VRPs. VRPs are programs adopted by software vendors to pay security researchers, ethical hackers and enthusiasts for exchange of discovering vulnerabilities in their software and responsibly disclosing the findings to the vendors [4]. Having more eyes on the code means that VRPs uncovered many more vulnerabilities and that makes finding vulnerabilities more difficult for malicious actors and hence ensure the security of software. Besides, vulnerabilities found by VRPs results in a coordinated disclosure and patch that minimizes the risk of vulnerabilities discovery and exploitation [5]. Our approach is inspired by the economics of exploitation model proposed by Miller et al. in [13]:

$$\begin{aligned} AttackerReturn = (Gain\ per\ use \times Opportunity\ to\ use) - \\ (Cost\ to\ acquire\ vulnerability + Cost\ to\ weaponize) \end{aligned}$$

The authors argue that an attacker must invest resources to acquire vulnerabilities and develop weaponized exploit for it. While mitigation techniques have shown to increase the cost and complexity of developing an exploit and hence cost of weaponize [14], we argue that VRPs can also be an important factor in this equation. As software vendors invest significantly to find vulnerabilities the cost of an attacker acquiring a vulnerability is going to be increased and that reduces the likelihood of vulnerabilities discovery and exploitation. Many software vendors such as Google, Mozilla, Facebook, PayPal, and recently Microsoft have adopted using VRPs. They realized that attackers are finding vulnerabilities faster and thus adapting VRPs will help put more sets of eyes looking for vulnerabilities and that makes all vulnerabilities shallow.

There are a number of VRPs and each one of them have their rules and criteria. Among these programs are Mozilla Firefox VRP [6] and Google Chrome VRP [7]. Mozilla Firefox and Google Chrome VRPs determine the reward amount of a vulnerability based on its severity and proof of its exploitation. Both VRPs classify the severity of vulnerabilities as critical, high, medium and low. The details about the description of every severity level for Firefox and Chrome VRPs can be found respectively in [8] and [9] respectively. While Firefox VRP rewards amount ranges from 500 - 10,000, Chrome VRP rewards ranges from 500 - 60000. Firefox VRP pays only for vulnerabilities that has been rated by VRP Committee as a critical or a high and some moderate vulnerabilities, while Chrome VRP rewards critical, high, medium, and some low vulnerabilities.

Problem Description. Recently, there have been efforts to validate CVSS Base score. Some of the researchers have evaluated CVSS Base score using reported exploits [10] and [11] and attacks [12]. The results show that CVSS Base score has a poor correlation with the reported exploits ([10] and [11]) and with the reported attacks [12]. Thus, CVSS Base scores have been considered not a good risk indicator [12] because the majority of vulnerabilities have high scores and have no reported exploits or attacks. Hence, it is hard to use those scores to prioritize among vulnerabilities. However, the lack of exploits or attacks may be a result of prioritized fixing of vulnerabilities.

Contribution. In this research, we propose using independent scales used by VRPs to evaluate CVSS Base score. VRPs use their own vulnerability severity rating systems that use a very thorough technical analysis and security experts opinions to assign a severity to vulnerabilities. The severity ratings are then used to pay money ranging from 500$ to 60,000$ or even more. Hence, comparing CVSS Base score with VRPs severity ratings could explain whether CVSS high scores are reasonable, and why having many high sever vulnerabilities with no reported exploit or attacks.

To conduct this study, we examine 1559 vulnerabilities of Mozilla Firefox and Google Chrome browsers. The two software has been selected because of their rewarding programs maturity and their rich history of publicly documented rewarded vulnerabilities. Besides, the examined vulnerabilities have been assessed by both VRPs rating systems and the CVSS Base score which makes their comparison feasible.

The paper is organized as follows. Section 2 presents the related work. In Section 3, the selected datasets are presented. In section 4, the validity of CVSS Base score is examined. Section 5 presents the discussion. In section 6, concluding comments are given and the issues that need further research are identified.

2 Related Work

Bozorgi et al. [10] have studied the exploitability metrics in CVSS Base metrics. They argued that the exploitability measures in CVSS Base metrics do not differentiate well between the exploited and not exploited vulnerabilities. They attributed that to the fact that many vulnerabilities with a CVSS high score have no reported know exploit and many vulnerabilities with low CVSS scores have a reported know exploit. However, in this paper, we evaluate the performance of CVSS Base score considering both the exploitability and the impact factors using vulnerability rewards programs and we provide an insight into why many vulnerabilities with a high CVSS score and have no exploits.

Allodi and Massacci in [12] have used a case-control study methodology to evaluate whether a high CVSS score or the existence of proof of concept exploit is a good indicator of risk. They use the attacks documented by Symantecs AttackSignature as the ground truth for the evaluation. Their results show that CVSS Base score performs no better than randomly picking vulnerabilities to fix. Besides, they also show that there are many vulnerabilities that have a high CVSS score and are not attacked. However, in this paper, we seek to find an explanation of why the majority of vulnerabilities have a high CVSS score and have no reported exploits. Thus, we use VRPs instead of an attack in the wild to conduct this study.

Younis and Malaiya in [11] have compared Microsoft rating systems with CVSS Base metrics using the availability of exploits as a ground truth for the evaluation. In addition to finding that both rating systems do not correlate very well with the availability of exploit, they also find that many vulnerabilities have a high CVSS score and have no reported exploits. However, in this study we try

to use different ground truth for the evaluation so that an explanation for why many vulnerabilities have a high CVSS scores and have no reported exploits may be provided.

Finifter et al. in [4] have examined the characteristics of Google Chrome and Mozilla Firefox VRPs. The authors find that using VRPs helps improving the likelihood of finding latent vulnerabilities. They also find that monetary rewards encourage security researchers not to sell their result to the underground economy. Besides, they find that patching vulnerabilities found by the VRPs increases the difficulties and thus the cost for the malicious actors to find zero-day vulnerabilities or exploit them. However, in this study, we examine using VRPs as ground truth to evaluate CVSS Base score.

Swamy et al. in [14] at the Microsoft Security Response Center examine the impact of using exploit mitigation techniques that Microsoft has implemented to address software vulnerabilities. One of their result shows that stack corruption vulnerabilities that were historically the most commonly exploited vulnerability class are now rarely exploited. However, in this research, we focus on the impact of using VRPs on the availability of exploits and on the relationship between VRPs measures and CVSS Base score.

3 Datasets

In this section, we first provide the source of the data. Then we show how the data were collected and analyzed. In this research, the data about vulnerabilities, exploits, and vulnerabilities rewards program data have been collected from the National Vulnerability Database (NVD) [15], Exploit Database (EDB) [16], and Mozilla Firefox [17] and Google Chrome bug databases [18] receptively. Table 1 shows the number of the examined vulnerabilities and their exploits. It should

Table 1. Firefox and Chrome Vulnerabilities

Software	Vulnerabilities	Exploit Exist
Firfox	547	22
Google Chrome	1012	5

be noted that the total number of the Firefox vulnerabilities is 742. Out of this number, a 195 vulnerabilities were not examined because we could not find information about them and that is explained as follows. First, 71 vulnerabilities have no direct mapping between the Common Vulnerabilities and Exposures (CVE) number and the Firefox Bug ID. Second, a 122 vulnerabilities could not be accessed due to the unauthorized access permission (You are not authorized to access this data); Third, two vulnerabilities have no data recorded in the Firefox bug database. We found that the VRP data in the Firefox bug database have started to be recorded starting 2009. Thus, all vulnerabilities and exploits

of Firefox during the period 2009 to October 2015 were collected. On the other hand, it should also be noted that the total number of vulnerabilities of Chrome is 1084. A 72 vulnerabilities were not examined because we could not find information about them because of the unauthorized access permission"You are not authorized to access this data". We also found that the VRP data in the Chrome bug database have started to be recorded starting 2010. Therefore, all vulnerabilities and exploits of Chrome during the period 2010 to October 2015 were collected. The data of every examined vulnerability of Firefox and Chrome were collected using the following steps. First, from NVD, the vulnerability is first identified. Next, for every existing link in NVD to vendors' bug database, we collected the vulnerability's severity rating and rewards data assigned by the VRPs. After that, for every vulnerabilitys CVE number found in the vendors bug database, the CVSS scores and severity values were collected. Lastly, for every examined vulnerability we used the CVE number to verify whether it has an exploit reported in the EDB or not.

3.1 Firefox Vulnerabilities Analysis.

Table 2 shows only three of the Firefox vulnerabilities because showing the whole vulnerabilities is limited by the number of pages allowed. Firefox VRP does not provide data about the amount of the reward paid and rather it uses: 1) + symbol to indicate the bug has been accepted and payment will be made, 2) - symbol to indicate the bug does not meet the criteria and payment will not be paid, and 3) ? symbol to indicate the bug is nominated for review by the bounty committee [8].

Table 2. The obtained measures of Firefox and CVSS Base Score

CVE	Mozilla Firefox VRP		CVSS Base Score		Exploit Existence
	Reward	VRP Severity	Severity	Score	
CVE-2011-2371	3000-7500	sec-critical	High	10	EE
CVE-2013-1727	500-2500	sec-moderate	Medium	4	NEE
CVE-2015-0833	3000-5000	sec-high	Medium	6.9	NEE

The CVSS Base score assigns a score in the range [0.0, 10.0]. This score represents the intrinsic and fundamental characteristic of a vulnerability and thus the score does not change over time. CVSS score from 0.0 to 3.9 corresponds to Low severity, 4.0 to 6.9 to Medium severity and 7.0 to 10.0 to High severity. Mozillas security ratings are see-critical: vulnerabilities allow arbitrary code execution, sec-high: vulnerabilities allow obtain confidential data, sec-moderate: vulnerabilities which can provide an attacker additional information, sec-low: minor security vulnerabilities such as leaks or spoofs of non-sensitive information. We have found that 13 vulnerabilities did not meet the criteria for rewarding and hence have been assigned "–" symbol. We have also found that 11 of them

have a low and a moderate severity (five are low and six are moderate) and two are high and critical.

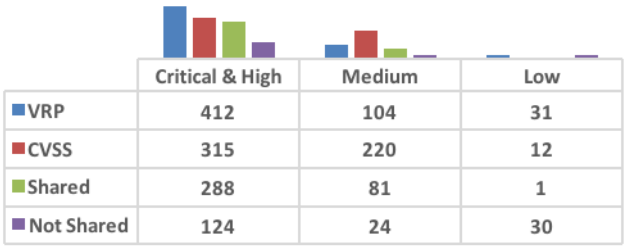
Table 3 shows the number of the rewarded and not rewarded vulnerabilities and their severity values for Firefox dataset. It should be noted that the Not Rewarded vulnerabilities are most likely have been discovered by internal discoverers (41.13%) whereas Rewarded vulnerabilities have been discovered by external discoverers (58.87%) [4]. While the Firefox bug database does not clearly provide information about whether the vulnerabilities have been discovered internally or externally, this was very clear in the Chrome bug database where the name and the team the discoverer works with is provided. Table 3 also shows the severity values and their frequency for rewarded and on rewarded data. It should be noted that the majority of the medium severity vulnerabilities and all low severity vulnerabilities were discovered internally.

Table 3. Firefox Dataset

Vulnerabilities	**Rewarded**	**Not Rewarded**
547	225	322
VRP Severity	**Rewarded**	**Not Rewarded**
Critical & High	210	202
Medium	15	89
Low	0	31

Fig. 1 shows the vulnerabilities severity values of CVSS Base score and Firefox VRP ratings of Firefox dataset. There are 412 vulnerabilities that have been assessed as critical or high severity by VRP rating system whereas there are 312 vulnerabilities that have been assessed as high severity by CVSS Base score. It should be notated that Firefox VRP severity rating is the baseline that we are comparing CVSS scores with. It should also be noted that Shared means the same vulnerabilities, which have the same CVE number, that have been assigned the same severity value by Firefox VRP rating system and CVSS Base score. On the other hand, Not Shared means the same vulnerabilities, but have been assigned different severity values by CVSS Base score. Almost 70% (69.9) of the vulnerabilities that have been assessed by Firefox VRP as critical or high severity have also been assessed as high severity by CVSS. Using Common Weakness Exposure (CWE) [19], which is used to identify vulnerabilities types, we have found that the majority of the Shared vulnerabilities are of the vulnerabilities that execute code. However, the 124 vulnerabilities that are Not Shared have all been assigned a high or critical severity by VRPs rating system, whereas CVSS Base score has assigned to 7 of them a low severity and to 117 of them a moderate severity.

On the other hand, almost 78% (77.88) of the Shared vulnerabilities that have been assessed by Firefox VRP as a medium severity have also been assessed as a medium severity by CVSS. However, there are 24 Not Shared vulnerabilities that have been assigned a medium severity by the VRP rating system. Out of these,



	Critical & High	Medium	Low
VRP	412	104	31
CVSS	315	220	12
Shared	288	81	1
Not Shared	124	24	30

Fig. 1. Comparing Firefox VRP and CVSS Severity Values

19 vulnerabilities have been assigned a high severity and five have been assigned a low severity. However, only one vulnerability that has been assessed as a low severity by CVSS base score and VRP rating system. While 30 vulnerabilities that have been assessed as a low severity by the VRP rating system, eight have been assessed as a high severity and 22 as a medium severity.

Table 4 shows the vulnerabilities that have been mismatched by CVSS Base score. As can be seen, seven vulnerabilities have been assessed as low severity by CVSS whereas three of them have been assessed as critical and four of them has been assessed as high severity by the VRP. It has noticed that those seven vulnerabilities have been assigned critical and high severity values by the Firefox VRP during the debate time, but the vulnerabilities severity first assignments were later changed [20]. We have found that the majority of those vulnerabilities requires unusual users interactions. We have also noticed that CVSS version 3, which has not been used yet, have consider using user interaction factor when assessing exploitability factor [21]. It is clear that the medium range of CVSS scores makes the main part of the mismatch compared to the high and low ranges.

Table 4. Vulnerabilites Mismatched by CVSS Base score

VRP	**Crtical**		**High**		**Moderate**	**Low**	
CVSS	Low	Medium	Low	Medium	Low	Medium	High
Total	3	**24**	4	**93**	4	**22**	8

3.2 Chrome Vulnerabilities Analysis.

The Chrome vulnerabilities have been examined similar to the Firefox vulnerabilities as shown in Table3. The only difference is that Chrome bug database provides the amount rewarded. Chrome security ratings are similar to that of Mozilla. Unlike Firefox where low severity vulnerabilities are not rewarded, seven

low severity vulnerabilities have been rewarded by Chrome VRP. The seven low severity vulnerabilities have been found to effect non-critical browser features, crash inside the sandbox, or hang the browser.

Table 5 shows the number of the rewarded and not rewarded vulnerabilities and their severity values for Chrome dataset. We have found nine vulnerabilities have been classified as TBD (To Be Determined) and thus considered them as not rewarded. It should be noted that the majority of the Not Rewarded vulnerabilities have been discovered by Google internal discoverers and they represent around 41.4%, whereas the Rewarded vulnerabilities have been discovered by external discoverers and represents around 57.7%. Table 5 also shows the severity values and their frequency for the rewarded and not rewarded data. It should be noted that while the majority of the critical and high vulnerabilities have been discovered externally, the majority of the low vulnerabilities have been discovered internally. The frequency of the amount paid is shown in Fig. 2. As can be

Table 5. Chrome Dataset

Vulnerabilities	**Rewarded**	**Not Rewarded**
1012	584	428
VRP Severity	**Rewarded**	**Not Rewarded**
Critical & High	441	175
Medium	136	137
Low	7	116

seen, the majority of the rewarded vulnerabilities (404) have been paid either 500$ or 1000$. We have noticed that 70.79% (286) of those vulnerabilities have

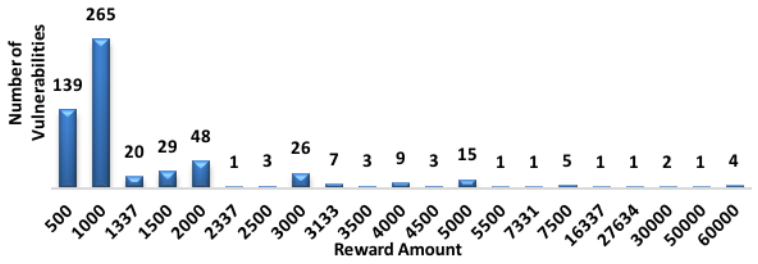


Fig. 2. Rewarded Amount of Chrome Rewarded Vulnerabilities

been assigned a high severity, 27.47% (111) have been assigned a medium severity, and only 1.73% (7) have been assigned a low severity by VRP. Looking at the point number 3-5 under the Reward amounts section in [7], we can see that establishing exploitability or providing a Proof of Concept (PoC) or with a poor quality of PoC could be the reason for paying less for many severe vulnerabilities.

Fig. 3 shows the vulnerabilities severity of CVSS Base score and Chrome VRP rating system of Chrome dataset. Almost 82% (81.65) of the vulnerabilities that have been assessed by Chrome VRP as critical or high severity have also been assessed as high severity by CVSS. However, the 113 vulnerabilities that are Not Shared have all been assigned a high severity by CVSS, whereas VRPs rating system have assigned to 35 of them a low severity and to 78 of them a medium severity. On the other hand, almost 73% (72.89) of the Shared vulnerabilities

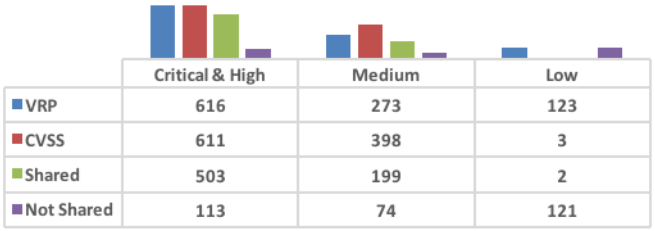


Fig. 3. Rewarded Amount of Chrome Rewarded Vulnerabilities

that have been assessed by Chrome VRP as a medium severity have also been assessed as a medium severity by CVSS. However, there are 74 Not Shared vulnerabilities that have been assigned a medium severity by CVSS. Out of these, 73 vulnerabilities have been assigned a high severity and only one has been assigned a low severity. However, only two vulnerabilities that have been assessed as a low severity by CVSS Base score and VRP rating system, whereas out of the 121 that have been assigned a low severity by VRP, 35 have been assigned a high severity and 86 have been assigned medium severity by CVSS Base score. Table 6 shows the vulnerabilities that have been mismatched by CVSS Base score. We have noticed that out of the 113 vulnerabilities, 75 vulnerabilities have been assigned a high medium score 6.8.

Table 6. Vulnerabilites Mismatched by CVSS Base score

VRP	**Crtical**		**High**		**Moderate**	**Low**	
CVSS	Low	Medium	Low	Medium	Low	Medium	High
Total	0	**1**	0	**112**	1	**86**	35

4 Validation of CVSS Base Score

In this section, we compare CVSS Base score severity with VRPs severity ratings. We assume that VRPs severity rating values are the ground truth because of the

through technical analysis and security experts opinions used and the fact that the severity rating are used to pay money. To evaluate the performance of CVSS Base score, we describe when a condition (true or false) is positive or negative as follows:

True Positive (TP)	When the CVSS Base score assigns a high severity and VRPs assign critical or high, or, When CVSS Base assigns medium and VRPs assign medium.
True Negative (TN)	When the CVSS Base score assigns low severity and VRPs assign low.
False Negative (FN)	When the CVSS Base score assigns low and VRPs assign critical or high, or, When the CVSS Base score assigns medium and VRPs assign critical or high, or, When the CVSS Base score assigns low and VRPs assign medium.
False Positive (FP)	When the CVSS Base score assigns and VRPs assign medium or low. Or when the CVSS Base score assign medium and VRPs assign low.

Since CVSS Base score uses an ordinal range: 0-3.9 = Low, 4 - 6.9 = Medium, and 7-10 = High, the possibility of overlapping between the ranges could make high Low vulnerability such as 3.9 close to medium and high Medium such as 6.9 close to high. To take this into consideration, we used a cluster algorithm to group severity ranges based on the distance between their values. We implemented the K-Means clustering algorithm provided by R language [23] to cluster CVSS Base score for Firefox and Chrome dataset. The result for Firefox vulnerabilities show that Low is in the range from 1.9-5.4, Medium from 5.8-7.6, and High from 8.3-10, whereas the results for Chrome vulnerabilities show that Low is in the range from 2.6-5.1, Medium from 5.8-7.1, and High from 7.5-10.

We used statistical measures, termed sensitivity, precision, and F-measure to evaluate the performance of CVSS Base score severity. Sensitivity, which also termed recall, is defined as the ratio of the number of vulnerabilities correctly assessed as high or medium to the number of vulnerabilities that are actually high or medium as shown by the following: Sensitivity = TP / TP+FN. Precision, which is also known as the correctness, is defined as the ratio of the number of vulnerabilities correctly assessed as high or medium to the total number of vulnerabilities assessed as high or medium as shown by the following: Precision = TP / TP+FP. For convenient interpretation, we express these two measures in terms of percentage, where a 100% is the best value and 0% is the worst value. Both precision and sensitivity should be as close to the value 100 as possible (no false positives and no false negatives). However, such ideal values are difficult to obtain because sensitivity and precision often change in opposite directions. Therefore, a measure that combines sensitivity and precision in a single measure is needed. F-measure can be interpreted as the weighted average of sensitivity

and precision. It measures the effectiveness of a prediction with respect to a user attached β times as much importance to sensitivity as precision. The general formula for the F-measure is shown by the following:

$$F_\beta - Measure = \frac{(1+\beta^2) \times Precision \times Senetivity}{(\beta^2 \times Precision) + Senetivity}$$

β is a parameter that controls a balance between sensitivity and precision. When $\beta = 1$, F-measure becomes to be equivalent to the harmonic mean, whereas when β <1 it becomes more precision oriented. However, when β >1, F-measure becomes more sensitivity oriented. In this paper β has been chosen to be 2. Due to their importance, we have also used the FP rate measure: FP rate = FP / FP+ TN and the FN rate measure: FN rate = FN / TP+FN.

4.1 Result

To calculate the above mentioned performance measures, we need to obtain the confusion matrix for the two datasets. Using the severity ratings assigned by VRPs and CVSS Base score, the confusion matrix was determined as shown in Table 7. We have also determined the CVSS Base score ranges obtained by the clustering algorithm and due to the limited pages allowed we only show the results for the CVSS Base score original ranges. It should be noted that we add up the number of every condition, for instance True Positive for Fire fox = 369. As can be seen, CVSS scores before the clustering have a very high FP rate. Using

Table 7. CVSS Base score compared to VRPs Rating Systems

Condition	CVSS Vs. Actual VRPs	Firefox	Chrome
True Positive	When the CVSS High and VRPs Critical or High	288	503
	When CVSS Medium and VRPs Medium	81	199
True Negative	When CVSS Low and VRPs Low	1	2
False Negative	When CVSS Low and VRPs Critical or High	7	0
	When CVSS Says Medium and VRPs Critical or High	117	113
	When CVSS Low and VRPs Medium	4	1
False Positive	When CVSS High and VRPs Low or Medium	27	108
	When CVSS M and VRPs Low	22	86

the values in Table 7, the performance measures for CVSS Base score original ranges, clustering ranges and our mismatching analysis ranges have been calculated as shown in Table 8. We also used Spearman correlation measure to assess the correlation between CVSS scores before clustering and after clustering as shown in Table 9. As can be seen, CVSS score correlate with VRPs rating values with p-value less than 0.0001. Clustering score and using mismatching analysis

Table 8. Performance Measures for CVSS before and after clustering

Software	Performance Measures	CVSS Scores before clustering (%)	CVSS Scores after clustering (%)
Firefox	Sensitivity	74.25	56
	Precision	88.25	93
	F1-Measure	80.66	70.21
	F2-Measure	57.99	46.94
	False Positive Rate	98	50
	False Negative Rate	25.75	43.54
Chrome	Sensitivity	86	66
	Precision	78	82
	F1-Measure	82	73
	F2-Measure	63	52
	False Positive Rate	99	60
	False Negative Rate	14	34

have shown a slight improvement on the correlation value for the Chrome vulnerabilities whereas no effect have been noticed on the correlation value for Firefox vulnerabilities. However, we also looked at the percentage of the vulnerabilities

Table 9. Spearman Correlation between CVSS Base score and VRPs Rating System

Software	Correlation	CVSS Scores before clustering	CVSS Scores after clustering
Firefox	Value	0.65	0.47
	P-value	0.0001	0.0001
Chrome	Value	0.53	0.59
	P-value	0.0001	0.0001

that have been assigned high and medium severity by CVSS scores and VRPS ratings to verify which measure is more aggressive. For the whole dataset, VRPs have assigned 66% of the vulnerabilities a high severity, whereas CVSS have assigned 59% of the vulnerabilities a high severity. On the other hand, VRPs have assigned 24% of the vulnerabilities a medium severity, whereas CVSS have assigned 46% a medium severity.

4.2 Threats to Validity

In this research, we have considered two datasets of two software of the same domain, internet browsers. We consider extending our analysis as long as the data about software from different domains are publicly available and accessible. We are also aware that there are other factors that can affect the vulnerabilities exploitation. Thus, we in no way imply that VRPs should be the only consideration when trying to assess CVSS Base score.

5 Discussion

Results have shown that CVSS scores have a higher FP rate. This is mainly because of the number of True Negatives. Out of the 131 vulnerabilities that have been assigned as Low by chrome VRP only two (True Negative) vulnerabilities have been assessed as Low by CVSS. We have found that 86 of these vulnerabilities have been assigned medium and 35 have been assigned high. On the other hand, out of the 31 vulnerabilities that have been assessed as Low by Firefox VRP, only one (True Negative) vulnerabilities have assessed as Low by CVSS. We have found that 22 of these vulnerabilities have been assigned medium and 8 have been assigned high.

There are more Execute Code vulnerabilities in Firefox than in Chrome. This could be explained by the effect of defensive mechanism, Sandbox, used by Chrome. Furthermore, based on the amount paid, the data from Chrome show that proving exploitability is more valuable than discovering vulnerabilities.

6 Conclusion and Future work

This study evaluates CVSS Base Scores as a prioritization metric by comparing it with VRP reward levels, which are arguably more direct measures. We used 1559 vulnerabilities from Mozilla Firefox and Google Chrome browsers to conduct this study. The performance measures and the correlation results show that CVSS Base Score is suitable for prioritization. The fact that there are more vulnerabilities with a high CVSS scores and have no exploits or attacks have been explained by the effect of VRPs on vulnerabilities exploitation. Besides, considering that CVSS score assess most of the vulnerabilities as severer, data show that VRPs have assessed even more vulnerabilities as severe more than CVSS Base score.

Still, there appears to be a need for continued updating of the CVSS metrics and measures. CVSS should highly consider including the Likelihood of Exploit factor (not only the availability of exploit, but also how likely it is that functioning exploit code will be developed) as CWSS [24] and Microsoft [25] rating systems did. Besides. The two chosen VRPs rating systems have shown that Likelihood of Exploit is the main factor that determine the amount of the reward paid for the discoverers and that was very evident in Chrome dataset. As the two datasets considered represent two software of the same domain, examining data from different domains can be valuable as long as their data is publicly available and accessible.

References

1. Mell, P., Scarfone, K., Romanosky, S.: A complete guide to the common vulnerability scoring system version 2.0. Published by FIRST-Forum of Incident Response and Security Teams, pp.123. (2007)

2. Defense in Depth, http://www.nsa.gov/ia/_files/support/defenseindepth.pdf. Accessed on 08 January 2016.
3. Pescatore, J.: Application Security: Tools for Getting Management Support and Funding. White Paper, SANS Institute (2013)
4. Finifter, M., Devdatta, A., David, W.: An Empirical Study of Vulnerability Rewards Programs. In: Proceedings of the 22nd USENIX Security Symposium, pp. 273288. Washington (2013)
5. Dark Reading: Connecting The Information Security Community, http://www.darkreading.com/coordinated-disclosure-bug-bounties-help-speed-patches/d/d-id/1139551
6. The Mozilla Security Bug Bounty Program, https://www.mozilla.org/en-US/security/bug-bounty/ Accessed on 08 January 2016.
7. Chrome Reward Program Rules, https://www.google.com/about/appsecurity/chrome-rewards/index.html. Accessed on 08 January 2016.
8. Security Severity Ratings, https://wiki.mozilla.org. SecuritySeverity—Ratings
9. Severity Guidelines for Security Issues, https://www.chromium.org/developers/severity-guidelines. Accessed on 08 January 2016.
10. Younis, A. A., Malaiya,Y. K.: Comparing and Evaluating CVSS Base Metrics and Microsoft Rating System. In: The 2015 IEEE International Conference on Software Quality, Reliability and Security, pp. 252-261. Vancouver, BC (2015)
11. Allodi, L., Massacci, F.: Comparing Vulnerability Severity and Exploits Using Case-Control Studies. J. Tra. Info. Syst. Secu. 17, 1 (2014)
12. Miller, M., Burrell, T., Howard, M.: Mitigating Software Vulnerabilities. Technical report, Microsoft Security Engineering Center (2011)
13. Nagaraju, S.S., Craioveanu, G., Florio, E.: Software Vulnerability Exploitation Trends. Technical Report, Microsoft Trustworthy Computing Security (2013)
14. Bozorgi, M., Saul, L. K., Savage, S., & Voelker, G. M.: Beyond heuristics: learning to classify vulnerabilities and predict exploits. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 105114. New York (2010)
15. National Vulnerability Database, https://nvd.nist.gov/. Accessed on 08 January 2016.
16. Exploit Database, https://www.exploit-db.com/. Accessed on 08 January 2016.
17. Security Advisories for Firefox, https://www.mozilla.org/en-US/security/known-vulnerabilities/firefox/. Accessed on 08 January 2016.
18. Chromium, https://code.google.com/p/chromium/issues/list. Accessed on 08 January 2016.
19. Common Weakness Enumeration (CWE), http://cwe.mitre.org/. Accessed on 08 January 2016.
20. Mozilla Bugzilla,https://bugzilla.mozilla.org/. Accessed on 08 January 2016.
21. Common Vulnerability Scoring System, V3 Development Update, https://www.first.org/cvss. Accessed on 08 January 2016.
22. Point-Biserial, https://www.andrews.edu/~calkins/math/edrm611/edrm13.htm. Accessed on 08 January 2016.
23. R: A Language and Environment for Statistical Computing, https://www.r-project.org/
24. Common Weakness Scoring System, https://cwe.mitre.org/cwss/cwss_v1.0.1.html. Accessed on 08 January 2016.
25. Using Exploitability Index, https://technet.microsoft.com/en-us/security/ff943560.aspx. Accessed on 08 January 2016.