Improved Prediction of Vulnerability Exploitation using   
CVSS Base-Score with Optimized Equation Parameters

Ben Church – 10006197

**Abstract**

# Introduction

Software vulnerability addressment is an important problem for organizations since exploitation of such vulnerabilities causes preventable losses. This problem tends to grow with time because vulnerabilities are often reported more frequently than they can be fixed. [Find example from VRP?] To minimize losses in spite of the increasing number of vulnerabilities becoming known, and with limited available resources, organizations must prioritize their vulnerability addressment strategy. Organizations minimize real losses by prioritizing vulnerabilities based on the cost each one is expected to impart if left unaddressed. Probabilistically, this expected cost, commonly called risk, is equal to the real cost in the event of exploitation, times the probability of that exploitation occurring.

The Common Vulnerability Scoring System (CVSS) was created to provide a single, objective measure of the risk presented by any software vulnerability [Scarfone2007]. The CVSS offers three scores in the range of 0-10, measuring the risk presented by the vulnerability. They are the base, temporal, and environmental scores. The base score coveys the intrinsic vulnerability risk, or time and context invariant risk, by considering six mostly objective vulnerability metrics, which will be examined shortly. The temporal and environmental scores essentially provide context specific re-evaluations of the base score. The temporal score takes into consideration changes in public or expert vulnerability knowledge, or vulnerability patch deployment, one might achieve a more accurate measure of the risk currently posed by a vulnerability. The environmental score considers the distribution of the vulnerability in the computer system or organization of interest, as well as the subjective importances of various computer system services which could be compromised. Because of their (mostly) objective and unchanging nature, CVSS base scores are widely available from vulnerability databases and are often used in research.

The complete formula for the CVSS base score equation v2 is shown in Figure 1. To the end of providing an objective scoring system, most of the metrics in the equation are objective and where possible, quantitative. Variation in a particular vulnerabilities scoring will not come from subjectivity in whether zero, one, or multiple instances of user authentication are required for exploitation. Likewise, whether a system’s data confidentiality is not at all compromised, partially compromised, or totally compromised, is objective and quantitative. A notable exception to the objective metrics is the AccessComplexity metric, which can be high, medium, or low. The CVSS has in fact received criticisms of subjectivity [Khazaei2016], and been the subject of several critical evaluation studies.

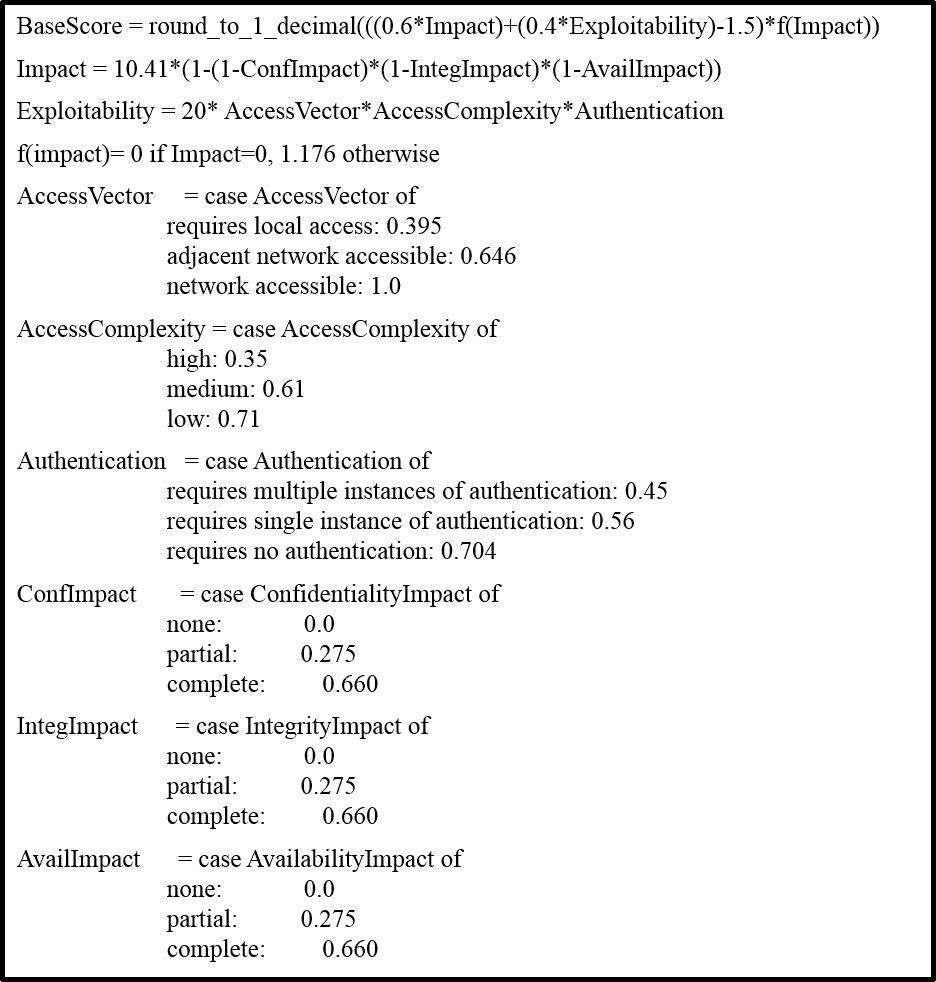


Figure : CVSS base score equation, v2 [Scarfone2007]

Bozorgi *et al.* [Bozorgi2010] and Khazaei *et al.* [Khazaei2016] proposed machine learning methods for automated base score prediction using additional information from vulnerability database entries. They proposed these methods in response to questions of the ability of the CVSS base score to predict actual exploitation of vulnerabilities. While machine learning may be useful for correlating vulnerability information with base scores, or even predicting exploitation, it probably cannot be used to achieve the single scoring system intended by the CVSS. If machine learning was integrated into the CVSS, different users of the system could not take advantage of online learning from new vulnerabilities. System specific online learning will cause the system’s model to change, potentially scoring a vulnerability differently than another system. Otherwise, it cannot improve itself from new information. As such, the need remains for a single, standardized vulnerability scoring system.

Allodi and Massacci [Allodi2014] examined the risk assessment capabilities of the CVSS base score by retrospectively correlating base scores with both proof-of-concept and real black market exploits. To measure exploit prediction, they computed confusion matrices where the CVSS base score’s relation to a threshold serves as the predictor for exploitability, and appearance of the vulnerability in Symantec’s AttackSignature[[1]](#footnote-1) and ThreatExplorer[[2]](#footnote-2) datasets as true exploitation in the wild. They measured prediction system’s risk reduction as:

Where P(e|f) denotes the probability of event e given event f, EV is the event that vulnerability V is exploited (appearing in Symantec’s data, in this case), score(V) is V’s CVSS base score, and T is some threshold value. They found that using CVSS base scores alone provided an optimal risk reduction of about 0.15 at a threshold of 2, little better than guessing.

Younis *et al.* [Younis2015] performed a study on the effectiveness of the CVSS base score and Microsoft’s vulnerability rating systems for predicting vulnerability exploitation. Similarly to Allodi and Massacci, they correlated predictions based on CVSS base scores and Microsoft system scores with real exploitation as determined by the existence of Exploit Database[[3]](#footnote-3) (EDB) entries. Both systems performed similarly, in that they were both sensitive enough to detect most exploitable vulnerabilities, but they had false positive rates above 90% for vulnerabilities found in Internet Explorer.

The failure of the CVSS to reliably predict exploitation in the wild is illustrated both in literature such as [Allodi2014, Younis2015], and by the adoption of organization-specific vulnerability rewards programs (VRPs) such as Mozilla’s[[4]](#footnote-4) or Google’s[[5]](#footnote-5). Nonetheless, the system’s open availability, the amount of published research, and the official recognition from organizations such as the National Institute of Standards and Technology under the Security Content Automation Protocol[[6]](#footnote-6) (SCAP) make the CVSS an attractive starting point for the development of a better scoring system. Improvements in the CVSS’ ability to predict vulnerability exploitation might make it a viable alternative to VRPs.

The work presented in the remainder of this paper explores the possibility of improving CVSS exploit prediction by optimizing the scalar parameters in its base score equation (Figure 1). Details about the optimization process including parameters, data, and objective functions, are given in the Methods section. Prediction system performance metrics are presented in the Results section. Finally, implications and limitations of the results, as well as potential improvements are covered in the Discussion & Conclusions section.

# Methods

To optimize the CVSS for exploit prediction, both vulnerability exploit existence and base score data were required. The National Vulnerability Database[[7]](#footnote-7) (NVD) was used as the source of this data. The NVD was chosen for the prospect of a large, unbiased set of vulnerability entries, by virtue of its 80,000 plus entries and governmental operation. Vulnerabilities were classified into ‘exploited’ or ‘unexploited’ based on whether the NVD entry had a reference to the EDB. The six element base score vectors, containing the vulnerability’s metric cases, rather than the final numerical score, were extracted from the database entries so base scores could be recomputed to drive the equation optimization.

# References

[Scarfone2007] - Mell P, Scarfone K, and Romanosky S. “A complete guide to the common vulnerability scoring system version 2.0” Published by FIRST-Forum of Incident Response and Security Teams. 2007.

[Khazaei2016] - Khazeai A, Ghasemzadeh M, and Derhami V. “An automatic method for CVSS score prediction using vulnerabilities description” Journal of Intelligent & Fuzzy Systems. 2016; 30:89-96.

[Bozorgi2010] - Bozorgi M. Lawrence KS, Savage S, and Voelker GM. 2010 “Beyond Heuristics: Learning to Classify Vulnerabilities and Predict Exploits” Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. 2010; 105-114.

[Allodi2014] - Allodi L, and Massacci F. “Comparing Vulnerability Severity and Exploits Using Case-Control Studies” ACM Transactions on Information and System Security. 2014; 17(1):1- 20.

[Younis2015] - Younis AA, and Malaiya YK. “Comparing and Evaluating CVSS Base Metrics and Microsoft Rating System” IEEE International Conference on Software Quality, Reliability and Security. 2015; 252-261.

[NVD] - https://nvd.nist.gov/

1. <http://www.symantec.com/security_response/attacksignatures/> [↑](#footnote-ref-1)
2. <http://www.symantec.com/security_response/threatexplorer/> [↑](#footnote-ref-2)
3. <https://www.exploit-db.com> [↑](#footnote-ref-3)
4. <https://www.mozilla.org/en-US/security/bug-bounty/> [↑](#footnote-ref-4)
5. <https://www.google.com/about/appsecurity/reward-program/> [↑](#footnote-ref-5)
6. <https://scap.nist.gov/index.html> [↑](#footnote-ref-6)
7. <https://nvd.nist.gov/> [↑](#footnote-ref-7)