Attacker economics for Internet-scale vulnerability risk assessment

Attacker economics for Internet-scale vulnerability risk assessment (Extended Abstract)

Luca Allodi
DISI - University of Trento, Italy
http://disi.unitn.it/~allodi

Abstract

Vulnerability risk assessment is a crucial process in security management, and the CVSS score is the standard-de-facto risk metric for software vulnerabilities. In this manuscript I show that current risk assessment methodologies do not fit real "in the wild" attack data. I also present my three-steps plan to identify an Internet-scale risk assessment methodology that accounts for attacker economics and opportunities. Eventually, I want to provide answers like the following: "If we deploy this security measure, the fraction of our users affected by this type of cyber attacks will be less than X%".

1 Motivations

Vulnerability exploitation is a major threat vector for cyber attacks [14], making vulnerability assessment a crucial point in the security management process. In my opinion, vulnerability assessment is currently undermined by two crucial problems.

Problem 1. Worrying about every vulnerability. Attacker models usually consider the attacker to be very powerful. The classic view of security is notoriously synthesised in Schneier's quote "security is only as strong as the weakest link"[1]: if a vulnerability is in my system, then an attacker will, sooner or later, exploit it.

However, according to Google, automated web attacks represent two thirds of the threats for the final user [14]. Indeed, in a previous joint work with Fabio Massacci [2], we show that only a small fraction of the population of available exploits is detected in the wild. For example, exploit kits are very popular automation tools [9], and yet they feature on average about 10 vulnerabilities each [2]. A foundational question remain therefore open: *How many vulnerabilities do attackers actually exploit? And, if not all vulnerabilities are exploited, what is the exploit selection rationale?*

Problem 2. Reliance on an empirically unverified assessment methodology. CVSS [12] is the *standard-de-facto* framework for vulnerability risk assessment. For example, the U.S Government recommends it as the reference assessment methodology for software security [13].

However, the score has never been properly validated against actual attack data, and may therefore be misleading as a risk measure. For example, the CVSS "Exploitability", used to measure the "likelihood" of exploitation [6, 12], is the same (and very high) for most vulnerabilities [6]: vendors and system administrators are therefore basically relying on "half" the metric only (the Impact assessment). This limitation may substantially affect security investment and management.

For example, Google built an ad-hoc reward program for vulnerability disclosure that rewards a fixed amount of money per vulnerability impact type[2]: consequently, a high-impact vulnerability that is not going to be exploited by anybody might be paid up to 200 times more than a low-impact but highly-exploited one.

2 Research objectives

We may identify two separate "levels" of security: The first sees the attacker as a dedicated and capable adversary that, with high motivation and skills, uses peculiar or previously unknown exploits (zero-day exploits) to attack the victim. This case is rather rare [5] and is hardly treatable because of lack of data. In the second, the attacker is not interested in a particular system [14], and relies on common knowledge and available exploits to attack a system by picking it "from the shoal".

My research goal is to modify the current models for attacks and attackers by looking at the security threat scenario from a macroscopic point of view: instead of trying to predict whether one individual will be attacked or not, I want to focus on the security of the population as a whole, to allow for realistic and Internet-scale security estimations. I aim at *enabling decision makers to estimate that deploying a security countermeasure will protect X% of the population of their users as a whole against cyber attacks*. In order to achieve this final goal, I will follow three research tracks:

T.1 *Characteristics of exploited vulnerabilities.* In this track I plan to identify (and partially have already identified) which features of a vulnerability are more susceptible of exploits. This will help in marking more likely to be exploited vulnerabilities as higher risk.

[1] http://www.schneier.com/blog/archives/2005/12/weakest_link_se.html

[2] http://www.google.com/about/appsecurity/reward-program/

T.2 *Context variables for exploitation.* Here I plan to explore and identify the influence of several contextual variables on the decision process of the attacker (exploit or not to exploit). As a result, this research path will delineate factors external to the vulnerability that favours (or disfavours) its exploitation.

T.3 *Trends of attacks enabled by attacker tools.* In the final track, I want to measure the influence of black-hat market trends on the final risk for users. The output of this track will identify temporal and volumetric trends of threats coming from the black markets.

3 Related works and Baseline

Frei et al. [8] were maybe the first to thoroughly analyse vulnerability and exploitation life-cycles. Part of their analysis was focused on data from NVD and OSVDB (equivalent to EDB). A similar approach, but using incident data from CERT/CC, is taken in [4]. These works have recently been extended by Shahzad et al. [15], which included vendors and software in Frei's analysis. Bozorgi et al. [6] were probably the first in looking at CVSS subscores against exploitation in EDB. However, all these studies looked at the same data, the reliability of which is not clear [2]. An investigation of how CVSS scores correlate to actual attacks in the wild can be found in [3]. Exploit kits infection dynamics are only covered in very recent studies [9, 14]. Vulnerability disclosure and discovery are often described as complex processes, that can be influenced by {black/white}-hat activities [8].

3.1 Datasets

So far, in our research group we collected four datasets, all comprising vulnerability data from 1999 to 2012. For a thorough discussion on the datasets see [2]:

NVD: the universe of vulnerabilities. NVD contains all disclosed vulnerabilities and respective CVSS assessment. It contains 49599 vulnerabilities.

EDB: the white market for exploits. EDB reports the vulnerabilities for which a "proof-of-concept" exploit exists; this is *not* evidence of exploitation in the wild. References 8189 exploited vulnerabilities.

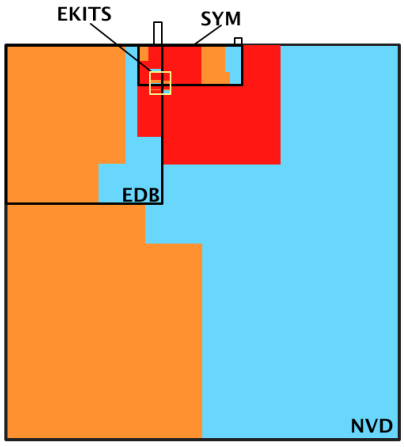
SYM: records of exploits in the wild. Symantec keeps two public datasets of signatures for local and network threats: AttackSignature[3] and ThreatExplorer[4]. If a CVE is reported in this dataset it means an exploit for it was observed in the wild. SYM reports 1417 exploits.

EKITS: the black markets for exploits. This dataset tracks more than 90 exploit kits alongside with market services, prices, and (126) exploited vulnerabilities[5].

[3] `http://www.symantec.com/security_response/attacksignatures/`

[4] `http://www.symantec.com/security_response/threatexplorer/`

[5] To better understand the markets, we are also testing these tools in a dedicated infrastructure [1].



Dimensions are proportional to data size. In red vulnerabilities with CVSS≥9 score. Medium score vulnerabilities are orange, and cyan represents vulnerability with CVSS lower than 6. The two rectangles outside of NVD space are vulnerabilities not present in NVD.

Figure 1: Relative Map of vulnerabilities per dataset

We are also collaborating with Symantec Worldwide Intelligence Network Environment (WINE) data sharing program [7] to access real attack data collected in the wild.

4 Preliminary results.

The first "show-stopper" for my research plan would be an empirical evidence that there is no difference among vulnerabilities: they are all going to be exploited. However, this is not the case. Figure 1 is a Venn diagram representation of our datasets. Areas are proportional to volume of vulnerabilities and colours represent HIGH, MEDIUM and LOW score vulnerabilities. As one can see the greatest majority of vulnerabilities in the NVD are not included nor in EDB nor in SYM. Moreover, EDB covers SYM for about 25% of its surface only: the attacker does not pick random exploits from EDB, but is involved in an autonomous vulnerability selection process. This is in contrast with previous assumptions [8, 15, 6]. EKITS overlaps SYM about 80% of the time.

Conclusion 1. Most vulnerabilities are unlikely to be exploited for the population of Internet users as a whole; moreover, the attacker seems involved in an independent exploit-selection process. If a vulnerability is traded in the black markets, it is most likely going to be attacked.

The second "show-stopper" would be the presence of a marker (CVSS) that could already well characterise exploited vulnerabilities. In order to understand if this is the case, at the University of Trento we tested the reliability of CVSS as a *marker* for exploitation risk [2]. In the medical domain, the sensitivity of a test is the conditional probability of the test giving positive results when the illness is present. Its specificity is the conditional probability of the test giving negative result when there is no illness. In this context, I assess to what degree the CVSS test predicts the illness ($v \in SYM$). To make statistically sound conclusions, I sampled the NVD, EDB and EKITS

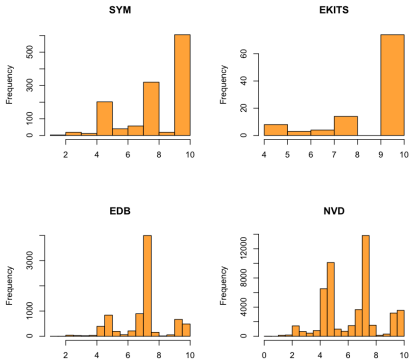


Figure 2: Distribution of CVSS scores per dataset.

datasets according to the distribution of the CVSS characteristics of the vulnerabilities in SYM (e.g. impact type, local or remote exploitability, etc.). For the experiment I consider CVSS scores higher than 6 to be HIGH, and those strictly lower than 6 to be LOW. In formulae, Sensitivity=$Pr(v.score \geq 6 \mid v \in$ SYM) while Specificity= $Pr(v.score < 6 \mid v \notin$ SYM). The sensitivity of all the samples is quite high ($> 89\%$), meaning that actually exploited vulnerabilities are most often marked with HIGH CVSS scores. On the other hand, the specificity is extremely low everywhere with a peak low in NVD and EDB at about 25% (EKITS settles at 49%). This means that 3 times out of 4, a vulnerability marked as HIGH risk is *not* going to be exploited. These observations have strong statistical evidence (Fisher exact test: $p < 2.2^{-16}$).

Conclusion 2. The CVSS score is **not** *a good predictor for exploitation. By prioritising patches following the US Government SCAP guidelines, the economic effort may be much higher than an optimal policy.*

Figure 2 reports the CVSS histogram distribution of vulnerabilities per dataset. The distribution of vulnerabilities reported in EDB is obviously different from that of SYM: legal vulnerability markets generate a population of vulnerabilities *different* from those *actually exploited by the bad guys*.

Conclusion 3. Current databases for vulnerabilities and exploits are **not representative** *of what the bad guys are actually doing. Conclusions on the security of software or networks using these datasets, as previously done in literature [15, 8], can therefore be misleading.*

5 Research plan

Because of the cross-field scope of these research tracks, I will collaborate with Prof. Fabio Massacci and Shim Woohyun from the University of Trento for the data analysis and regression part, and with Prof. Julian Williams from the University of Aberdeen for the modelling of the economic attacker.

Characteristics of exploited vulnerabilities. I will focus on *what vulnerabilities are of actual interest for the attacker*. I want to identify interesting CVSS characteristics for exploitation. To this purpose I will rely on our SYM and WINE-DB datasets. Using a logistic regression model in the form:

$$P(SYM = 1) = \frac{\exp(\alpha + x_i\beta)}{1 + \exp(\alpha + x_i\beta)} \quad (1)$$

I plan to evaluate evidence for statistical significance of a (set of) CVSS characteristic(s) x_i for the inclusion of a vulnerability in SYM[6]. This way we can isolate statistically significant CVSS sub-factors for exploitation.

On a second step, I plan to extend the analysis to the WINE-DB dataset, in order to look for possible correlations between vulnerability characteristics and volume of attacks. For example, attackers may prefer high complexity, high impact vulnerabilities (CVSS Access complexity=High, Impact=High) over easy but low impact ones.

Context variables for exploitation. In this research track, I want to develop an *expected utility* model for cybercrime, as done in more "traditional" criminology studies [11]: an attacker needs to evaluate expected returns from various criminal activities $EU_{C_1}, EU_{C_2}, .., EU_{C_n}$. In this case, the criminal activity is the exploitation of a vulnerability: *what external factors will trigger the decision of exploiting vulnerability v_i instead of v_j?*. I formulate three hypotheses for possibly interesting contextual factors.

H.1 *Presence of alternate exploits with better pay-offs for the attacker.* If the system is affected with more than one vulnerability, the attacker will try to exploit the one with the highest pay-off. The trade-off may account for effort required for exploitation, cost of buying an exploit from the black market, or exploitation code availability.

H.2 *Existence of an alternate exploit.* If the system is affected by more than one vulnerability, and an exploit for one vulnerability already exists (e.g. as recorded in the wild), than the remaining vulnerabilities represent lower risk.

H.3 *Expected persistence of the vulnerability on victims machines.* Auto-update policies and shorter software lifecycles (e.g. as in Google Chrome, Mozilla Firefox) may reduce the potential return for the attacker, who may be less interested in developing exploits for that software.

The SYM and WINE-DB datasets are central to this track. Similarly to the approach adopted for the CVSS score, I plan to use logistic regression to model the data. I also plan to test correlation with: a. inclusion in SYM. This first test will highlight whether evidence for correlation between the factor and the exploitation (SYM=$\{1,0\}$) of the vulnerability exists. b. Volume of attacks recorded in WINE-DB. The second test will highlight possible relative weights for the identified factors in the model (i.e. factor correlation with exploit "popularity").

Trends in attacks enabled by attacker tools. The third and last part of my research is the more explorative one. As exploit kits represent the majority of threats for the final user [14], I argue that black market trends may be of major importance for risk assessment. Because in the past IRC black markets

[6]Logit regression is the natural choice for analysis of categorical variables. However, other regression models may be applied.

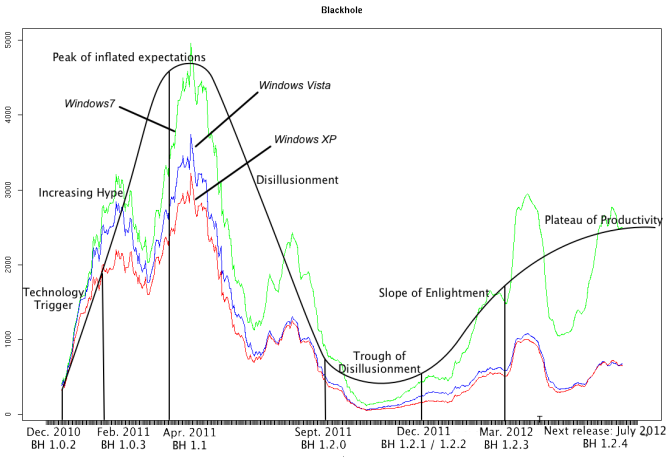


Figure 3: Monthly moving average for attacks delivered by Blackhole against the large-scale population of users, superimposed to the typical hype curve for new technology trends by Gartner.

have been shown to feature the typical characteristics of unfair markets [10], I will focus on the quality assessment process of the traded goods. To this aim, in our WINE-DB dataset I collected attack data for the most popular exploit kits [9]. I am particularly interested in evaluating:

1. Resiliency of attacks driven by exploit kits against different platforms (see [1]).

2. Predictability of trends of attacks for single exploit kits as opposed to the competition. An example of an exploit kit's attack trend from WINE-DB is given in Figure 3.

3. Assessment of exploit kit attacks against unique machines worldwide (from WINE-DB).

6 Conclusions

In this research proposal I underlined current problems with vulnerability risk assessment and management. I presented my three-tracks research plan to find a "general law of security" based on attacker economics. My work aims at the "internet-scale" level of security rather than to individuals. As an example, I expect that my methodology will allow people to draw conclusion like "*The number of our users affected by this family of cyber attacks will be below 10%*".

Acknowledgments

I thank Fabio Massacci, to whom goes the credit for the formulation of the idea of a "general law" for IT security; I also thank Julian Williams, Woohyun Shim, Vadim Kotov and Viet Hung Nguyen for our many constructive discussions upon which part of this work is based. This work is partly supported by the EU-SEC-CP-SECONOMICS and MIUR-PRIN-TENACE Projects.

References

[1] L. Allodi, V. Kotov, and F. Massacci. Malwarelab: Experimentation with cybercrime attack tools. In *Proc. of CSET'13*, 2013.

[2] L. Allodi and F. Massacci. A preliminary analysis of vulnerability scores for attacks in wild. In *ACM Proc. of CCS BADGERS'12*, 2012.

[3] L. Allodi and F. Massacci. How cvss is dossing your patching policy (and wasting your money). *BlackHat USA'13*, 2013.

[4] W. A. Arbaugh, W. L. Fithen, and J. McHugh. Windows of vulnerability: A case study analysis. *Computer*, 33(12):52–59, 2000.

[5] L. Bilge and T. Dumitras. Before we knew it: an empirical study of zero-day attacks in the real world. In *Proc. of CCS'12*, pages 833–844. ACM, 2012.

[6] M. Bozorgi, L. K. Saul, S. Savage, and G. M. Voelker. Beyond heuristics: Learning to classify vulnerabilities and predict exploits. In *Proc. of SIGKDD'10*, July 2010.

[7] T. Dumitras and D. Shou. Toward a standard benchmark for computer security research: The worldwide intelligence network environment (wine). In *Proc. of BADEGRS'11*, pages 89–96. ACM, 2011.

[8] S. Frei, M. May, U. Fiedler, and B. Plattner. Large-scale vulnerability analysis. In *Proc. of LSAD'06*, pages 131–138. ACM, 2006.

[9] C. Grier, L. Ballard, et al. Manufacturing compromise: the emergence of exploit-as-a-service. In *Proc. of CCS'12*, pages 821–832. ACM, 2012.

[10] C. Herley and D. Florencio. Nobody sells gold for the price of silver: Dishonesty, uncertainty and the underground economy. *Springer Econ. of Inf. Sec. and Priv.*, 2010.

[11] C. P. Krebs, M. Costelloe, and D. Jenks. Drug control policy and smuggling innovation: a game-theoretic analysis. *Journal of Drug Issues*, 33(1):133–160, 2003.

[12] P. Mell and K. Scarfone. *A Complete Guide to the Common Vulnerability Scoring System Version 2.0*. CMU, 2007.

[13] S. D. Quinn, K. A. Scarfone, M. Barrett, and C. S. Johnson. Sp 800-117. guide to adopting and using the security content automation protocol (scap) version 1.0. Technical report, NIST, 2010.

[14] M. Rajab, L. Ballard, N. Jagpal, P. Mavrommatis, D. Nojiri, N. Provos, and L. Schmidt. Trends in circumventing web-malware detection. Technical report, Google, 2011.

[15] M. Shahzad, M. Z. Shafiq, and A. X. Liu. A large scale exploratory analysis of software vulnerability life cycles. In *Proc. of ICSE'12*, pages 771–781. IEEE Press, 2012.