UNIVERSITY OF TRENTO

SECONOMICS

TENACE

My software has a vulnerability, should I worry?

An empirical validation of the CVSS industrial standard.

http://securitylab.disi.unitn.it

Luca Allodi, Fabio Massacci
University of Trento, Italy.

UNIVERSITY OF TRENTO

SECONOMICS

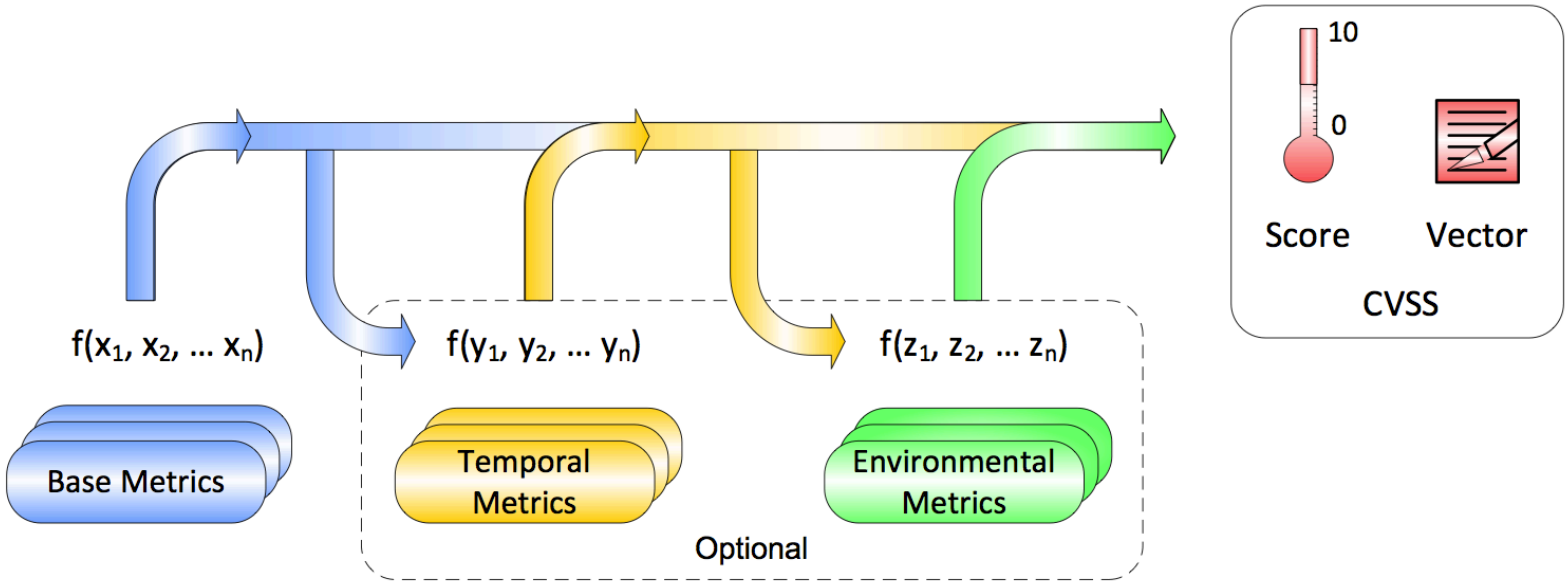
TENACE

Outline

1. What is CVSS, and who uses it
2. Research goals
3. Presentation of Datasets
4. Leveraging from cancer research
5. Threats to validity
6. Risk reduction with CVSS

What is CVSS, and who uses it

- CVSS is an assessment of vulnerability “system impact”
- Based on expert assessments to evaluate:
  - **Base Score *[0..10]***
  - Temporal Score
  - Environmental Score



How scientific is this?

- CVSS is a test by clinical expertise..

- I have a sw with a vulnerability…
- Is it easy to access?
- Is it high impact?
- Your CVSS doctor says HIGH Risk → patch
  - ✓ Of course please…

- I see double…
- Both eyes involved?
- Primary gaze impacted?
- Your CVSS doctor says HIGH risk → brain surgery
  - ? Ehm are you sure…

..But how informative is it?

Tests and Risks: a practical question

- A clinical test must be matched to the risk
  - Binocular diplopia AND intracranial lesion → 0% recovered without treatment
  - Binocular diplopia and no additional evidence → 42% recovered ***without*** treatment
    - Nolan "Diplopia" B. J. Ophtalm. 1966

- What the CIO would like to know:
  - IF HIGH CVSS listed by Sec. Config. Manager and Metasploit finds it → fix it and decrease risk of successful attacks by +15%
  - IF fix all remaining HIGH listed by Sec. Config. Manager but no additional evidence → risk decreses only by 3%
  - → Is +3% worth the extra money?

Related works

- Academia
  - [Frei et al.] ***Large-scale vulnerability analysis.*** *LSAD 2006.*
  - [Bozorgi et al.] ***Beyond Heuristics: Learning to Classify Vulnerabilities and Predict Exploits.*** *SIGKDD 2010.*
  - *[Gallon]* ***Vulnerability Discrimination Using CVSS Framework.*** *NTMS 2011.*
  - *[Edwards et al.]* ***An historical examination of open source releases and their vulnerabilities.*** *CCS 2012.*
  - [Shahzad et al.] ***A large scale exploratory analysis of software vulnerability life cycles.*** *ICSE 2012.*
- *Industry*
  - *[Scarfone et al.]* ***SP 800-117. Guide to Adopting and Using the Security Content Automation Protocol (SCAP) Version 1.0.*** *NIST, 2010*
  - *[Oracle]* ***Use of Common Vulnerability Scoring System (CVSS).*** *Oracle, 2010.*
  - *Vuln Assessment/Management tools: Qualys QualyGuard, NOPSEC PCI Compliance, BeyondTrust Rentina, ….*

Research goal

- *"If we fix this group of vulnerabilities risk of attacks for our costumers decreases by 85%"*

- Think of car accidents:
  - You **can't prove** that a particular technology will save your life
  - You **can't prove** that if you wear a safety belt you will not die
  - But still, **you want *statistical evidence*** that a particular car or using a belt ***improves*** your chances of surviving in a car accident [Evans 1986]
- Same with vulnerabilities:
  - Fixing a vulnerability will **not** assure you you will not be hacked
  - But this **improves** your chances of **not being** hacked
- → **you can't prove that you are safe, but you can show statistical evidence that you are x% better off**

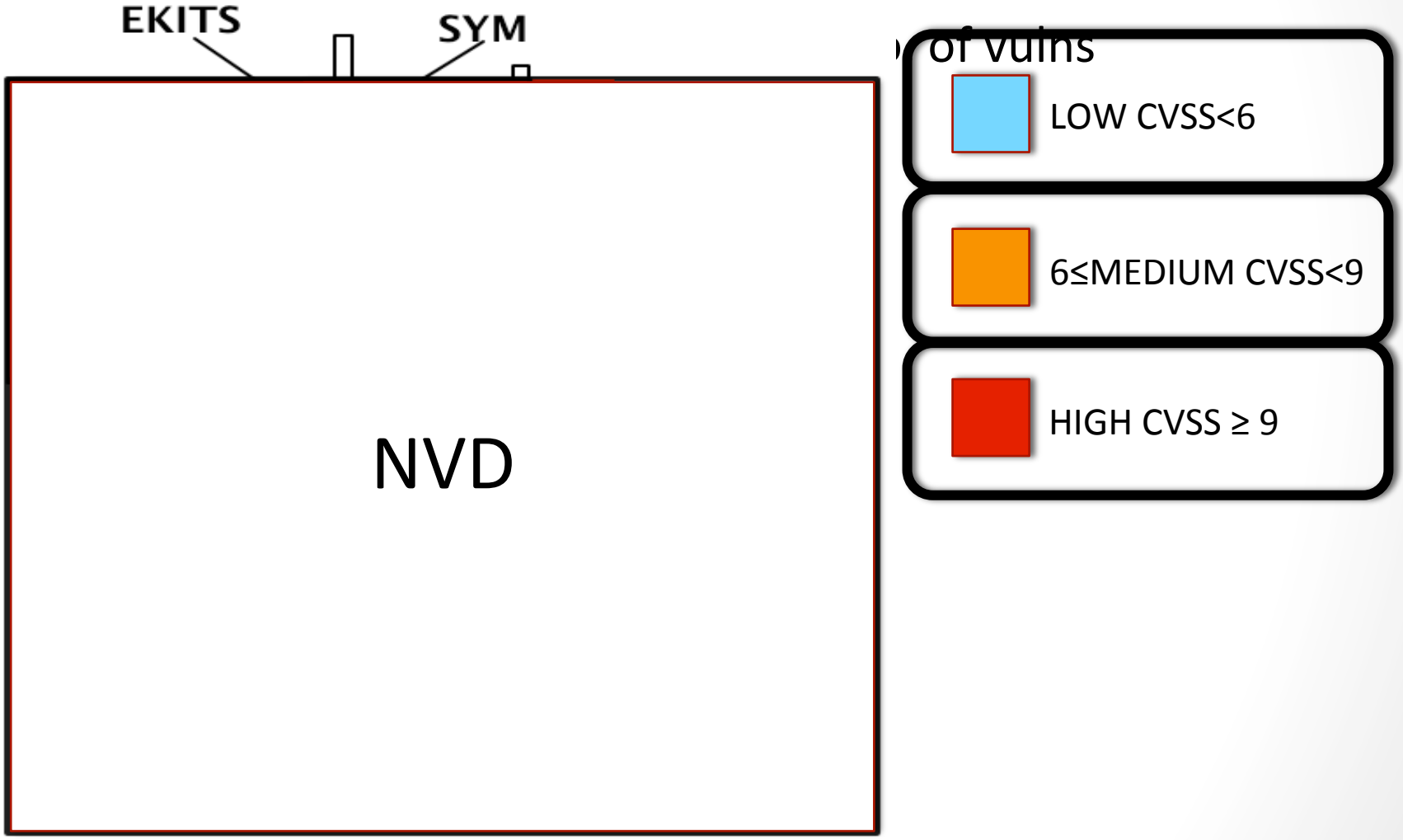
Attack scenarios: scope of work

- Victim is THE Target
  - Can mitigate this risk (IDSs, DLP, other Remediation strategies, insurance, etc.)
  - But cannot control it
  - → speaking of "risk decrease by X%" doesn't make sense
- Victim is only ONE of the Targets
  - Automated exploitation, phishing sites etc.
  - GOOGLE: 80% of attacks are of this nature
    - M. Rajab et al., Google Tech Report 2011
  - For these threats → "risk decrease by x%" makes sense
- We do not focus on Black Swan events
- → We focus on the most common threats

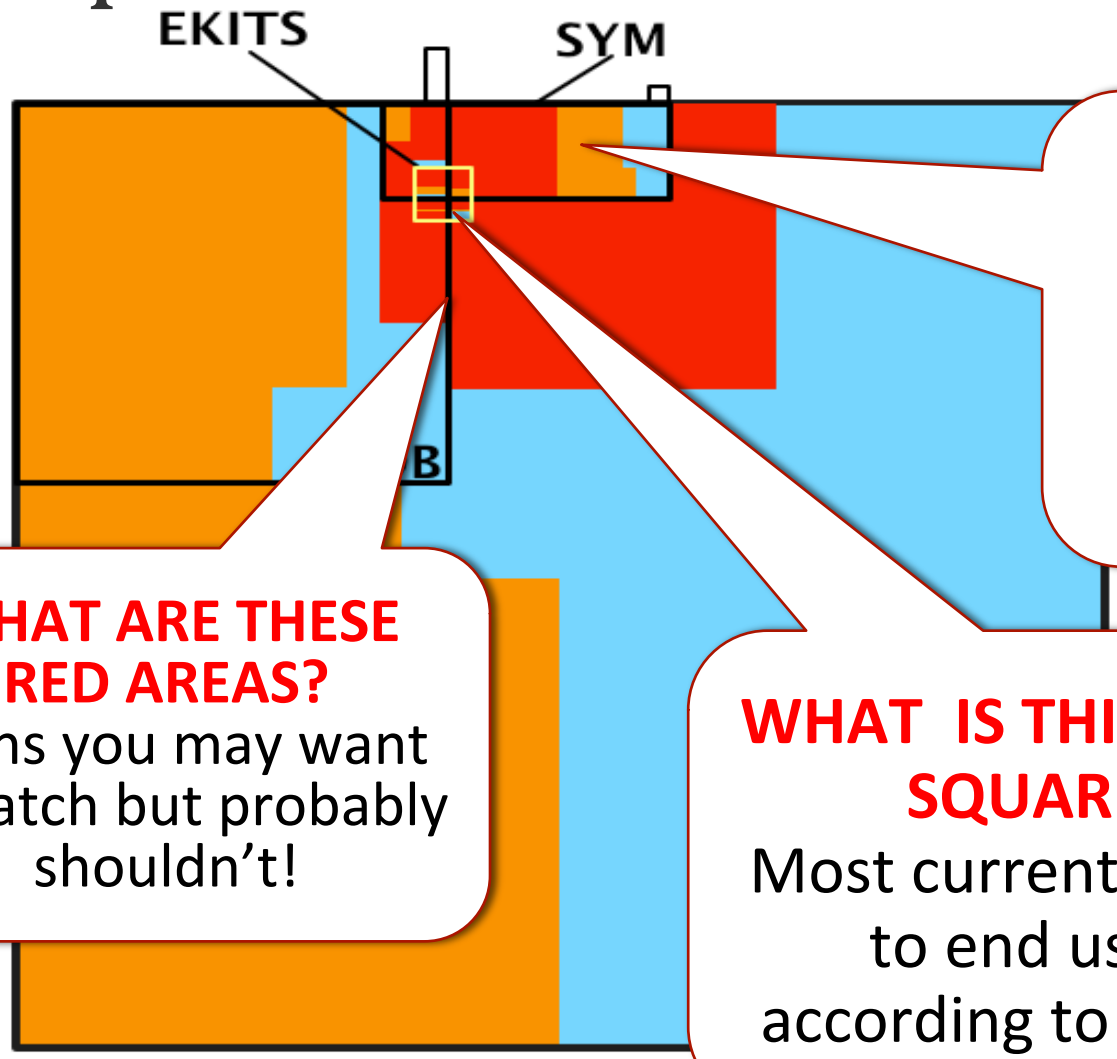
Vulnerabilities: data collection

- NATIONAL VULNERABILITY DATABASE: NVD – 49.624 vulns
  - The universe of vulnerabilities
- WHITE MARKETS OF EXPLOITS: EXPLOIT-DB – 8.189 vulns
  - Proof-of-Concept exploits published by security researchers
- ACTUAL EXPLOITS IN THE WILD: SYM – 1.274 vulns
  - Symantec / Kaspersky Threat reports
  - Vulnerabilities actually exploited in the wild
  - Conservative approach: SYM represents the existence of an attack
  - Browser/Plugins 14% – Server 22% – App. 17% - Windows 13%
  - Other OS 5% - Developer 5% - Business 7% - Unclassified 17%
- BLACK MARKETS FOR EXPLOITS: EKITS – 114 vulns
  - 2/3 of client threaths according Google (2011)
  - Exploit advert from the bad guys in an exploit kit
  - 90+ exploit kits from the black markets expanding Contagio's exploit pack table

Map of Vulnerabilities



Map of Vulnerabilities: Summary

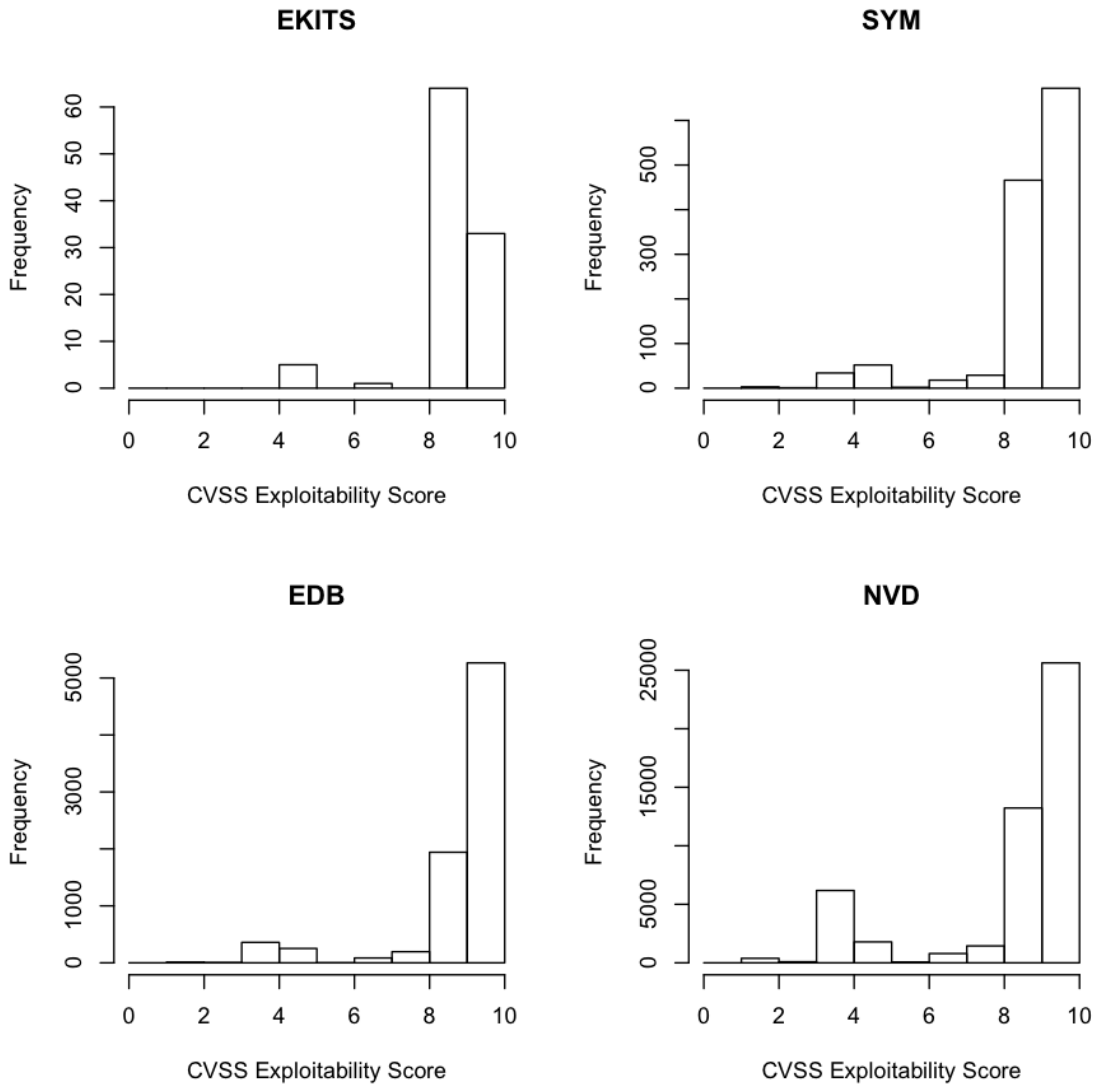


WHAT IS THIS?
50% of attacked vulns you did not patch

WHAT ARE THESE RED AREAS?
Vulns you may want to patch but probably shouldn't!

WHAT IS THIS LITTLE SQUARE?
Most current threats to end users according to Google

What makes CVSS so inaccurate?



- Risk (CVSS)= Impact x Likelihood
  - CVSS Likelihood = Exploitability
- Everything is exploitable → CVSS lacks of a real characterization of likelihood of exploitation

We need something more precise

- These conclusions **are a reflection of Symantec's view of security and the properties of our data:**
  - Maybe NVD and EDB contain many vulnerabilities not comparable with SYM
    - "Obscure" EDB/NVD software entries
      - E.g. Joomla expansion modules only in EDB/NVD
    - SYM*antec* detects primarily particular types of vulnerabilities:
      - Windows vulnerabilities in SYM are more frequent than in NVD
      - Vulnerabilities in SYM tend to have "complete" impacts on Confidentiality and Integrity
- How to go beyond these limitations?

Leveraging on cancer research

- Do High CVSS scores predict exploitation?
- Do smoking habits predict cancer?
  - → You can't ask people to start smoking so you can't run a controlled experiment → same here
- Case controlled study
  - Cases: people with lung cancer
  - Controls:
    - People with characteristics similar to the cases (Confounding factors)
    - Age, Sex, Social Status, Location
  - Explanatory variable
    - Smoking habit
- For each of the cases select another person with the same values of the control variables
  - Doll & Bradfor Hill, British Medial Journal 1950

CVSS Case Controlled Experiment I

- We control for
  - Year of Vulnerability
    - Date of exploit may condition the probability of being detected by Symantec
  - Software Type
    - Symantec sells technology to protect software typically used by its costumers
  - Type of Confidentiality, Integrity, Availability Impact
    - Symantec may detect mainly vulnerabilities that, for example
      - Allow for execution of arbitrary code
      - Allow privilege escalation/Data Leakage
    - While certain type may remain largely ignored
      - E.g. attacks against Availability

CVSS Case Controlled Experiment II

Observed effect	Cases	Controls	Explanatory variable
Lung Cancer	Same Hospital Patients	• Age • Sex • Location	• Smoke a lot • Smoke • Don't smoke
Exploitation	Same kind of exploitable vulns	• Confidentiality • Integrity • Avail • Year • Affected software	• CVSS is HIGH • CVSS is LOW • Vuln is in EDB • Vuln is in EKITS

CVSS Case Controlled Experiment III

- Case (attacked vulnerability):
  - CVE-2010-3962 (use-after-free vulnerability in MS IE 6,7,8)
  - Year=2010
  - Confidentiality =C, Integrity=C, Availability=C
  - Vendor=Microsoft, Software = ie
- Control (vulnerabilities similar to attacked ones):
  - Select 1 out of:
    - 5 from EKITS
    - 7 from EDB
    - 37 from NVD
- Repeat for all cases of attacked vulnerabilities
  - See what values of CVSS we get
  - See how many times you find an attacked vulnerability

How to measure a test

- Sensitivity → true positives vs all sick people
  - HIGH→ the test correctly identifies exploited vulns
  - LOW → lots of “sick people” undetected

- Specificity → true negatives vs all healthy people
  - HIGH → the test correctly identifies non exploited vulns
  - LOW → lots of “healthy people” flagged

Results & Statistical validation

- By experiment design, output of the experiment is a Latin Square of the type:

	In SYM	Not in SYM
CVSS Med + High	X	Y
CVSS Low	K	J

- Sensitivity=Pr(X|SYM) = X/(X+K)
- Specificity=Pr(J|not SYM) = J/(J+Y)
- However, even after controlling for confounding variables and sampling, we cannot know by looking solely at the results if (SYM) and (CVSS High) are independent one from the other:
  - If they are, the results (and conclusions from the data) would be generated by a random process → need of a statistical test to strengthen validity of conclusions
- X,Y,K,J may be small (<5) → Chi Square and other tests not suitable
  - We use Fisher's Exact test

A "Generate Panic" test

- Sensitivity: is High/Med CVSS good marker for v∈SYM?
- Specificity: is Low CVSS good marker for v∉SYM?
- Fisher test: significance with **p< 0.05(*) p<0.01(**)**

Test for Patching	Sensitivity	Specificity
Patch Everything	100%	0%
CVSS High+Med	91%(**)	23%(**)
CVSS + PoC in EDB	97%(*)	22%(*)
CVSS + EKITS	94%(**)	50%(**)
3BT: Down Syndrome	69%	95%
PSA: Prostate Cancer	81%	90%

Effects of low specificity

- Sensitivity, specificity are a-priori probabilities tied to the performance of CVSS as a test
- But what is the effect on the **population of vulnerabilities?**
- For example one has patched HIGH and MEDIUM score vulnerabilities?
  - What is then the a **posteriori** probability that a **patched vulnerability** would get an **attack**?
- → Apply Bayes theorem:

$$\Pr(v \in SYM \mid patch) = \frac{\Pr(v \in SYM) * sensitivity}{\Pr(v \in SYM) * sensitivity + \Pr(v \notin SYM) * (1 - specificity)}$$

= 6%

Threats to validity

- **External validity:** Generalization of conclusions
  - We assume that exploits in SYM are representative of attacks against final (Windows) users
    - Collecting multiple data sources to validate findings with other data (Reply, TrendMicro)
  - Maybe attacks against Linux/MacOS target different type of vulnerabilities
- **Internal validity:** Case control study can be further refined:
  - More statistical accuracy (e.g. Bootstrapping)
    - This would also allow us to derive confidence intervals for our conclusions
  - Different definition of “case”
    - Webkit vulnerability can also be a Chrome, Safari, Opera vulnerability
    - One vulnerability affecting three different versions may represent three different cases
  - Different/more controls?
    - E.g. Category of software (e.g. Browsers, Plugins, Windows..)

CVSS Risk reduction: answer to the CIO

- Is wearing a seat belt any useful?
  - Pr(Death x Safety Belt on) – Pr(Death x Safety Belt off)
  - Yes it is→ 43% improvement of chances of survival
    - L. Evans, Accident Analysis and Prevention 1986

CVSS Risk reduction: answer to the CIO

- Is patching HIGH score any useful?
  - Pr(Attack x CVSS High) – Pr(Attack x CVSS Low)

- Finally the figures the CIO wants
  - Patching HIGH/MED and exploit sold in Exploit Kits →improves by +62.81% (**Buckle up!**)
  - Patching fix HIGH/MED and PoC exploit by white hats → improves by +19.64% (**Up to you**)
  - Patching just HIGH/MED → improves by +3.2% (**Life is too short**)