UNIVERSITÀ DEGLI STUDI
DI TRENTO

My Software has a vulnerability, Should I worry?

An empirical validation of the CVSS industrial standard

Luca Allodi, Fabio Massacci
University of Trento - http://securitylab.disi.unitn.it

Research presented at

ACM TISSEC

Trento

Luca

- 3rd year Phd Student
- MsC Information Security
- Phd work: two bits
  1. CS Technical bit
  2. Economic modeling bit
     - In Durham working with Prof. Julian Williams for characterization of cybercrime markets
- http://disi.unitn.it/~allodi

SECONOMICS

TENACE

Outline of today

- Vulnerabilities: CIO perspective
  - Compliance and rules
- A medical equivalent of current practices
- Policy effectiveness measure:
  - Case control study for vulnerabilities and exploits
    - Results
    - Validation (according to available time)

Vulnerabilities: a CIO Perspective

- 50k+ vulnerabilities in NVD
- My Software has a vulnerability: should I worry?
  - Published somewhere at BlackHat, DefCon, Slashdot, whatever.
- The fanatical answer is "I should, for each and every one"
- The actual answer is "For this one, I just can't"
  - Technical Reasons
    - May not be technically fixable →integrated legacy sw may break
    - Even if expert to fix is there → she may have other tasks: relative priority?
    - Already planned upgrade in 3 months → why not just wait?
  - Budget Reasons
    - Money already allotted → again delay or stop other tasks
  - Compliance Issues
    - "It's the law" → zilions of competing laws (e.g. Internet crimes, building safety, health insurance contribution, etc. etc.)
    - Paying a fine (later) may be cheaper than deploying a fix (now)
- Need to Prioritize: "Worry now", "Worry later", "Life's too short"
  - Cannot tell CFO/CEO "I need extra money" → what is value for money?

SECONOMICS

TENACE

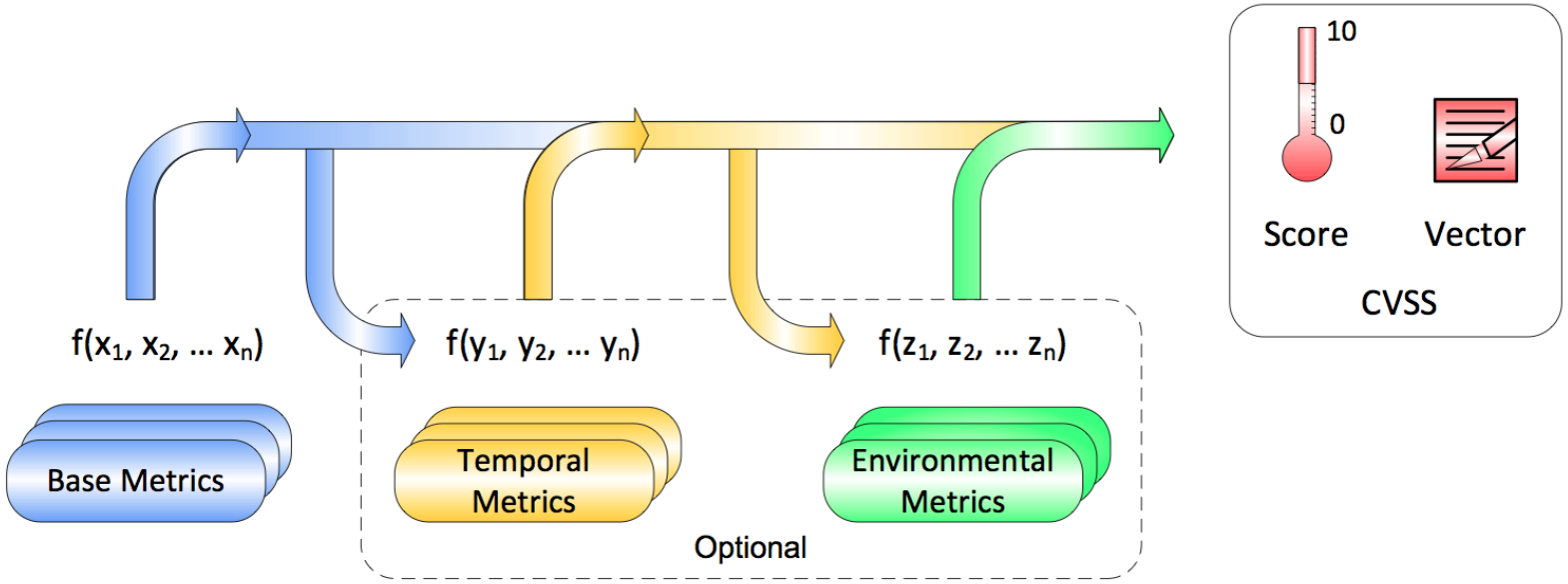
Vulnerabilities: a Compliance Perspective

- Listen to the U.S. Government….
  - US Cyber Security Order (Press release Feb'2013)
    - "NIST will work collaboratively with critical infrastructure stakeholders to develop the framework relying on existing international standards, practices, and procedures that have **proven to be effective**"
  - U.S. NIST SCAP Protocol v1.2(Draft Jan 2012)
    - "Organizations should use **CVSS base scores** to assist in **prioritizing** the remediation of known security-related software flaws based on the **relative severity of the flaws**."
  - PCI-DSS v2 (June 2012)
    - "**Risk rankings** should be based on industry best practices. For example, criteria for ranking —Highllrisk vulnerabilities may include a **CVSS base score of 4.0 or above**"
  - U.S. Government Configuration Baseline (USGCB)
    - Supported by the industry→ Rapid7, Telos, VmWare, Symantec, Qualys, Retina etc. etc.

- Conclusion: fix all vulnerabilities with high or medium CVSS score
  - But how this is "proven to be effective"?

What is CVSS

- CVSS (2.0) is an assessment of how the vulnerability can impact the system
- Based on expert assessments to evaluate:



Zooming in on Base Metrics

CVSS.base = f(Exploitability) x g(Impact)

- Exploitability
  - Access Vector (local, adj, network)
  - Access Complexity (high,medium,low)
  - Authentication (multiple,single,none)
- Impact (High,Medium,Low)
  - Confidentiality, Integrity, Availability
- Basically it is a "Clinical Evaluation"
  - "clinical examination is the process by which a medical professional investigates the body of a patient for signs of disease" (Wikipedia)

Comparing Clinical Tests

I HAVE A VULNERABILITY

- Is it of high impact?
  - Confidentiality affected?
  - Integrity?
  - Availability?
- Locally or from the network?
- …
- Overall score HIGH → your CVSS doctor says "patch your system"

I SEE DOUBLE

- Is it of high impact?
  - Primary gaze affected?
  - Left and right?
  - Downward and upward?
- Is it permanent or transient?
- …
- Overall score HIGH → your CVSS doctor says "brain surgery" → Ehm.. Sure..?

Tests and Risks: a practical question

- A clinical test must be matched to the risk
  - Binocular diplopia and no additional evidence → 42% recovered ***without*** treatment
  - Binocular diplopia AND intracranial lesion → 0% 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?

Research goal

- A methodology and practical criteria to prioritize security activities
  - "IF we mitigate vulnerabilities with feature X THEN risk of attacks decreases by x%"
- Think of car accidents:
  - You can't prove that if you wear a safety belt you will not die
  - But still, you want statistical evidence that using a belt improves your chances of surviving in a car accident
- Same with vulnerabilities:
  - Fixing a vulnerability will not assure you will not be hacked
  - But it improves your chances of not being hacked
- An important criterion is only "foresight" features
  - Vulnerabilities should be characterized by features that can be checked *before* an attack takes place
    - CVSS is ok → clinical expert assessment
    - Presence of Proof of Concept in Exploit DB → symptom
    - Among "Attacked vulns" in AV report → hindsight
  - Hindsight information should only be used to validate foresight prediction

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

SECONOMICS

TENACE

Learning from Medicine

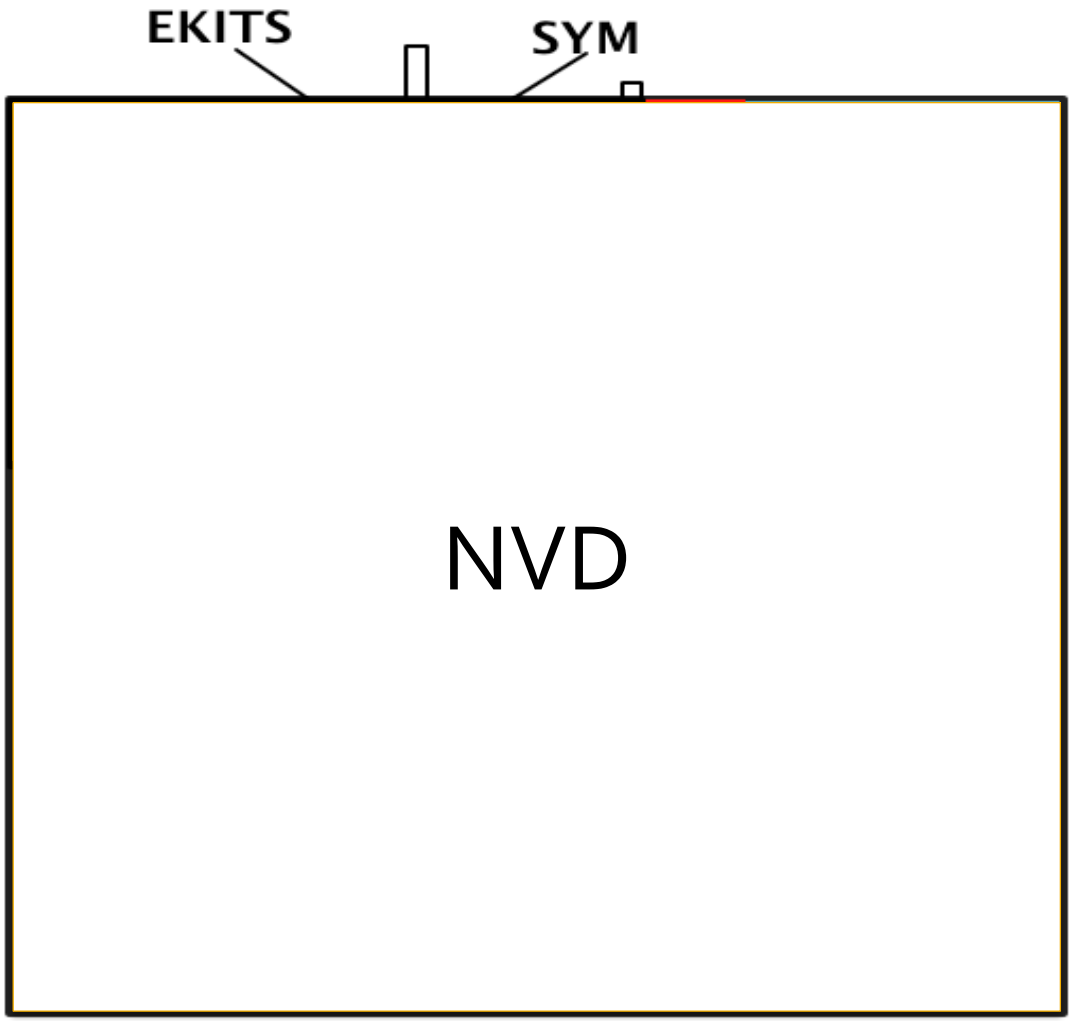
- How do you "prove" that
  - Giving up smoking reduces risk of lung cancer?
  - Safety belts reduce risks of deaths?
- You can't run a "controlled" experiments
  - Can't ask people to start smoking and see if they die
  - Can't ask people to run vulnerable software and see if they get hacked
- So… you do a "case-controlled" study
  - Doll & Bradfor Hill, British Medical Journal 1950 (&1970) (Smoking → Lung Cancer)
  1. Explanatory variable: Smoking habit
  2. Cases: people with lung cancer
  3. Possible confounding variables: Age, Sex, Social Status, Location
  4. Controls: random people with same characteristics of confounding variables
- Is there a (statistical) difference between your cases and a control population with the same characteristics?

Our cases and controls

Population to build the control groups	What's there
NATIONAL VULNERABILITY DB (the "universe")	45K+ vulns, 16K types of sw/versions etc.
EXPLOIT DB (Proof-of-Concept exploits is published by security researchers)	8K+ vulns, (~6k sw)
EKITS (our info on 90+ exploit kits adverts from the black markets expanding Contagio's table) 2/3 of End Users Threats are from there according to Google (2011)	101 vulns (46 sw)

Our Cases (the lung cancer patients, deads in accidents)	What's there
SYMANTEC's Threat Explorer Browser/Plugins 14% – Server 22% – App. 17% - Windows 13% - Other OS 5% - Developer 5% - Business 7% - Unclassified 17%	1K+ vulns with at least 1 attack in the wild (~600 sw)

Map of Vulnerabilities



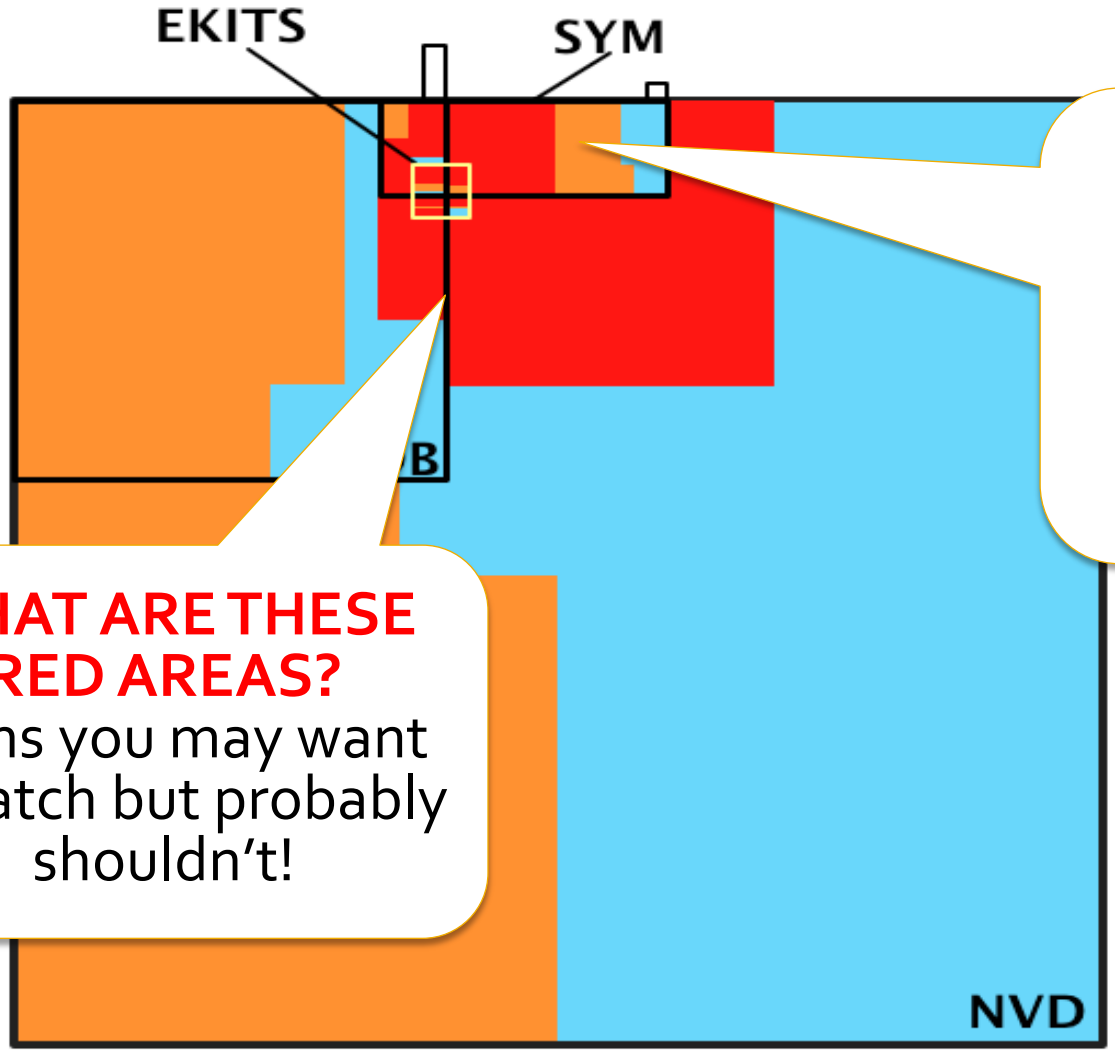
no. of vulns

LOW CVSS<6

6≤MEDIUM CVSS<9

HIGH CVSS ≥ 9

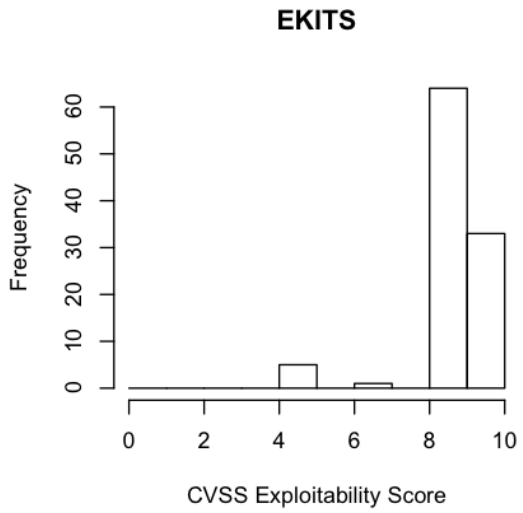
Glimpse of the problem

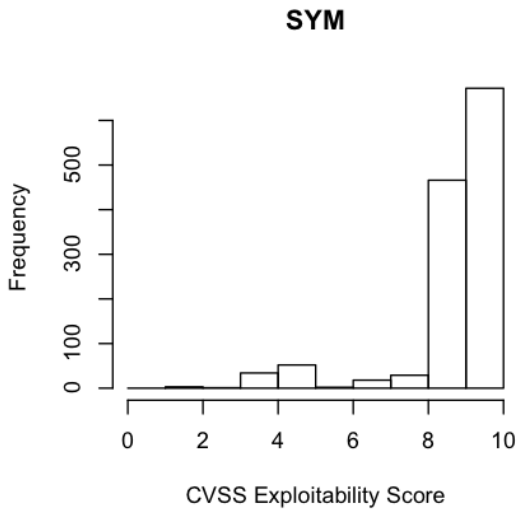


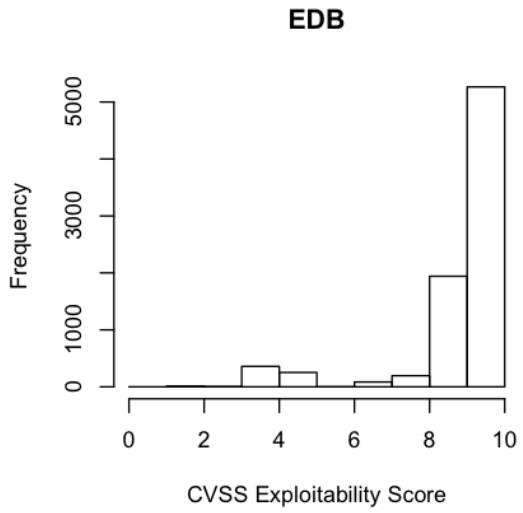
WHAT IS THIS?
50% of attacked vulns you did not patch

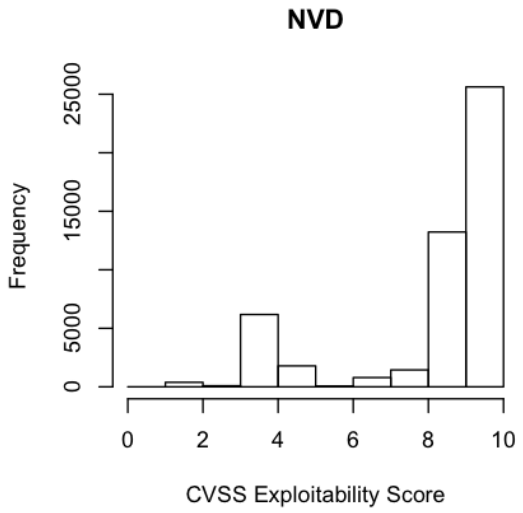
WHAT ARE THESE RED AREAS?
Vulns you may want to patch but probably shouldn't!

What makes CVSS so inaccurate?









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

Need for the case control study

- Cannot use data as-is to draw conclusions on CVSS
  - NVD/EDB may list software almost nobody use
    - Sw in SYM → 568, sw in EDB→ 5.819, sw in NVD → 16.399
    - E.g. a Joomla expansion module
  - SYMantec may focus on a subset of vulns
    - E.g. Windows vulnerabilities in SYM more frequent than in NVD
    - E.g Vulnerabilities in SYM usually have complete impacts on CIA
- So we run a case-controlled experiment
  - Cases → 1266 vulnerabilities with attacks in the wild
  - Controls → Random population of same size from EDB,NVD or EKITS with the same control variables
  - Bootstrapping → repeat 400 times and see the results

Our controls

- Smoking study
  - Controls for Age, Sex, Social Status, Location
- We control for
  - **Year of Vulnerability** →must be in SYM
    - Date of exploit may condition the probability of being detected by Symantec
  - **Software Type** → must be in SYM
    - Symantec sells technology to protect software typically used by its costumers
  - **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

Control implementation

- 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 randomly 1 out of:
    - 5 from EKITS
    - 7 from EDB
    - 37 from NVD

Repeat 400 times (bootstrapping)

- Repeat for all 1266 cases of attacked vulnerabilities
  - See what values of CVSS we get
  - See how many times you find an attacked vulnerability
- Repeat all above for N times to select different samples

SECONOMICS

TENACE

Result of the Experiment

- Result of each Nth sample is a latin square

	In SYM	Not in SYM
Value of Marker for **Risky Condition** (e.g. HIGH CVSS and vuln in EKITS)	Sick people correctly detected	Healthy people wrongly flagged
Value of Marker for **Not Risky Condition**	Sick people not detected by the test	Healthy people marked as such by the test

- We are interested in 3 things
  - Sensitivity and specificity → assess the quality of the test
  - Risk reduction → tells the CIO what to do
  - Variability due to randomness → confidence intervals

How to evaluate the marker

- 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

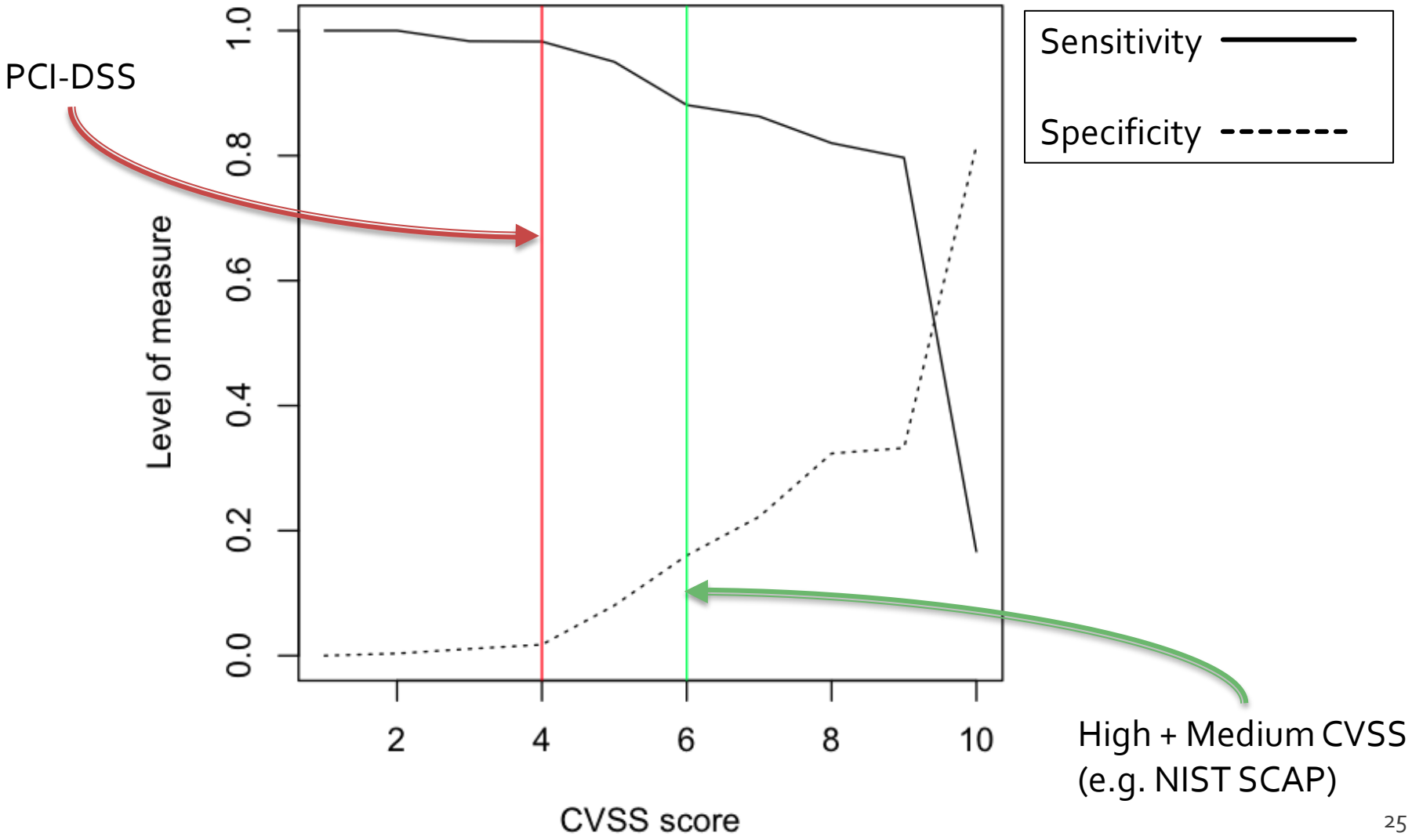
- Output of experiment:

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

- Sensitivity=Pr(X|SYM) = X/(X+K) <- SYM by column
- Specificity=Pr(J|not SYM) = J/(J+Y) <- Not SYM by column

- 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



A"Generate Panic" test

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

Test's Risk factors	Sensitivity	Specificity
None (Patch Everything)	100%	0%
CVSS High+Med	88%	16%
CVSS + **PoC (EDB)**	97%(**)	20%(**)
CVSS + **Bmar (EKITS)**	100%(*)	23%(*)
3BT: Down Syndrome	69%	95%
PSA: Prostate Cancer	81%	90%

From Experiment to Advice

- All this is very nice but… what about the CIO?
  - "If I patch vulnerabilities with features X would this reduce my risk of getting attack?"
- Compute answer from same table but by row
  - How good is our Assessment (CVSS etc) in predicting the future (Bayes Theorem)

$$Risk(MarkedHigh) = \frac{MarkedHigh \in SYM}{MarkedHigh \in SYM + MarkedHigh \notin SYM}$$

	in SYM	Not in SYM
Marked HIGH by CVSS+other information	Vuln marked for a patch that were attacked	Vuln marked for a patch that were **not** attacked
Marked LOW by CVSS+other information	Dangerous vuln not marked for a patch	Not Dangerous and not marked

CVSS Risk reduction: answer to the CIO

Risk Factor	RR	95% C.I.
CVSS >= 6	4%	-5% ; 12%
CVSS >= 6 + PoC	**42%**	**38% ; 48%**
CVSS >= 6 + BMar	**80%**	**80% ; 81%**
CVSS >= 9	8%	1% - 15%
CVSS >= 9 + PoC	**42%**	**36% ; 49%**
CVSS >= 9 + Bmar	24%	23% ; 29%

Validation in the wild: examples

- **WINE** -> Symantec dataset reporting **actual** attacks in wild:
  - count of exploitation attempts worldwide
  - PA (Potential of Attack) = log(attacks)

Risk factor	BROWSER vulns		WINDOWS vulns	
	%Vulns	PA red.	%Vulns	PA red.
None	100%	5	100%	6.1
CVSS >= 4	98.8%	5	97.3%	6.1
CVSS >=4 + PoC	4.1%	5	**16.7%**	**6.1**
CVSS >=4 + BMarket	**1%**	**4.8**	1.2%	4.8