UNIVERSITÀ DEGLI STUDI
DI TRENTO

SECONOMICS

A preliminary analysis of vulnerability scores for attacks in the wild

The SYM and EKITS datasets.

Luca Allodi, Fabio Massacci

lastname@disi.unitn.it

BADGERS 2012 CCS Workshop
15 October, Raleigh, NC, USA.

Outline

- Vulnerability research
  - Our datasets
  - What do bad guys actually need?
  - Threats to validity
- Distribution of CVSS scores
- Distribution of CVSS Exploitability
- Limitations
- New validation: case controlled experiment
- Preliminary conclusions

Vulnerability Research Today

- Much work relies on NVD, EDB and CVSS [1][2]
  - Software quality studies
  - Risk associated with software
  - Attack exposure windows
- NVD only tells us something about quality of software
- EDB tells us if a proof-of-concept exploit is released
- CVSS score assesses risk associated with the exploitation of the vulnerability
- Sum-up: NIST SCAP

[1] Frei, Stefan and May, Martin and Fiedler, Ulrich and Plattner, Bernhard. ***Large-scale vulnerability analysis. 2006.***
[2] Shahzad, Muhammad and Shafiq, Muhammad Zubair and Liu, Alex X. ***A large scale exploratory analysis of software vulnerability life cycles.2012***

Vulnerability Research Today

"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." [3]

[3] Quinn, Stephen D. and Scarfone, Karen A. and Barrett, Matthew and Johnson, Christopher S. ***SP 800-117. Guide to Adopting and Using the Security Content Automation Protocol (SCAP) Version 1.0. NIST 2010***

Vulnerability Research cnt'd

Our question is:

Is the data we're looking at by any means representative of actual attacks?

dataset	volume of CVEs
NVD	49.624
EDB	8.189
EKITS	103
Symantec (SYM)	1.289

What do bad guys actually need?

***Средний пробив на связке:* 10-25%**
* Пробив указывается приблизительный, может от

* Отстук стандартный, даже чуть выше станд
> Зевс = 50-60%
> Лоадер = 80-90%

Цена последней версии 1.6.x:
> Стоимость самой связки = 2000$
> Чистки от AB = от 50$
> Ребилд на другой домен/ИП = 50$
> Апдейты = от 100$
* Связка с привязкой к домену или IP .

Vulnerability	Affected sw	CVSS score
CVE-2006-0003	MDAC	5.1 (medium)
CVE-2006-4704	WMI Object Broke	6.8 (medium)
CVE-2008-2463	Snapshot	6.8 (medium)
CVE-2010-0806	IEpeers	9.3 (high)
CVE-2010-1885	HCP	9.3 (high)
CVE-2010-0188	PDF libtiff mod v1.0	9.3 (high)
CVE-2010-0886	Java Invoke	10.0 (high)
CVE-2010-4452	Java trust	10.0 (high)
CVE-2011-0558	Flash <10.2	9.3 (high)
CVE-2011-0611	Flash < 10.2.159	9.3 (high)

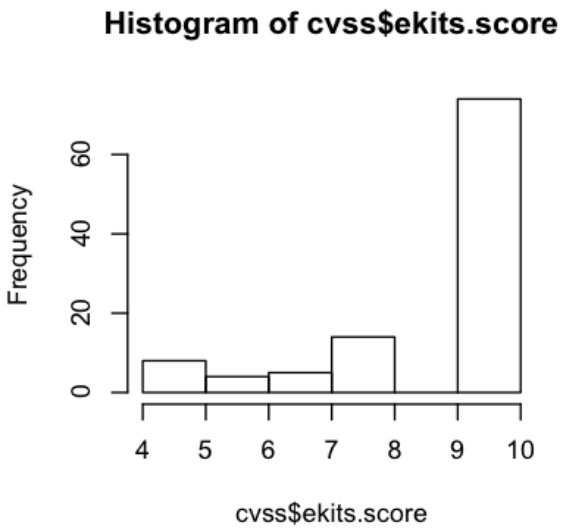
> CVE-2011-0611 (Flash <10.2.159)
> CVE-2010-0886 (Java Invoke)
> CVE-2010-4452 (Java trust)
*Виста и 7ка бьется

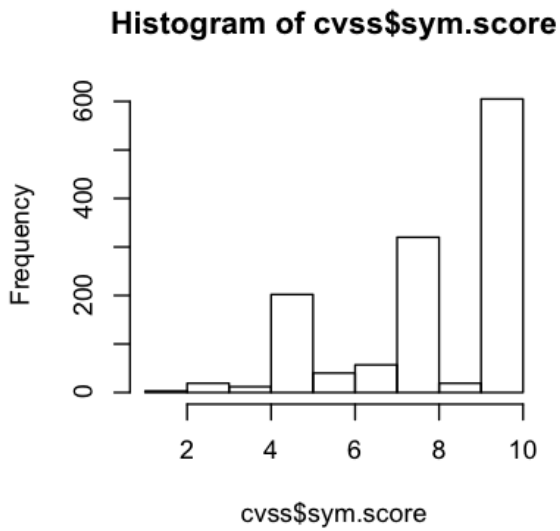
Threats to Validity

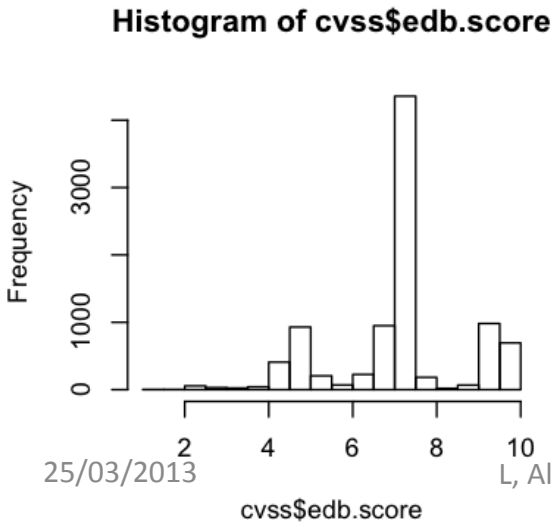
- CVE entry mentioned in NVD
  - That's just hearsay (good for witch hunt and government compliance): too much noise
- Its exploit code appears in the Exploit-DB
  - It proves researcher is skilled (hire him!) but why bad guys should be using it?
- Mentioned in SYM
  - E.g.: Does it report only client-side vulnerabilities?
    - 200+ server vulnerabilities,
  - 50+ non-windows, 60+ dev tools, 100+ browser, ..
- Advertised in an Exploit Kit
  - Maybe bad guys are just selling junk (remember IRC credit card numbers?) [4]

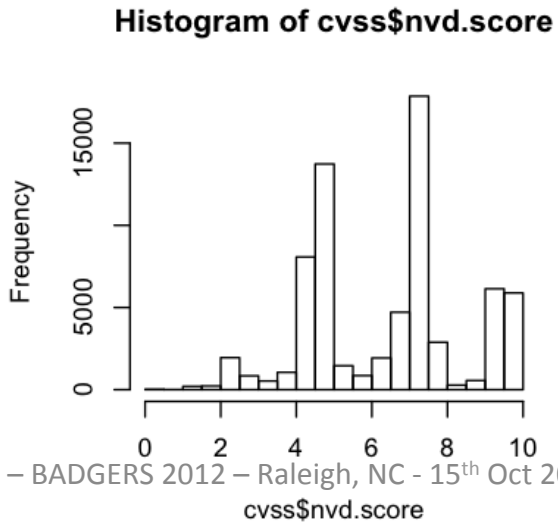
[4] Herley, C. and Florencio, D. ***Nobody sells gold for the price of silver: Dishonesty, uncertainty and the underground economy*** . SPRINGER-EISP 2010

Distribution of CVSS Scores (HIST)



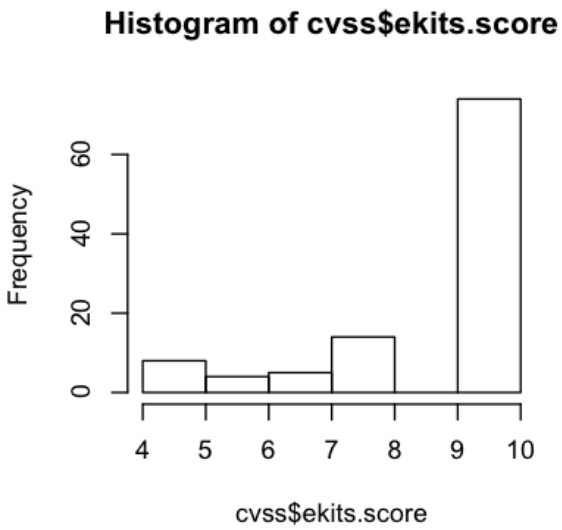


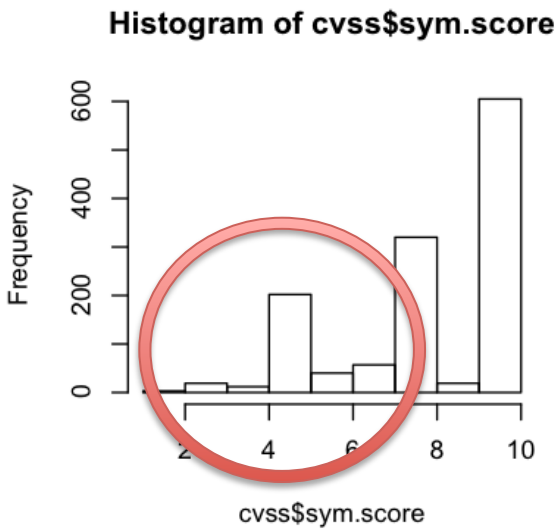


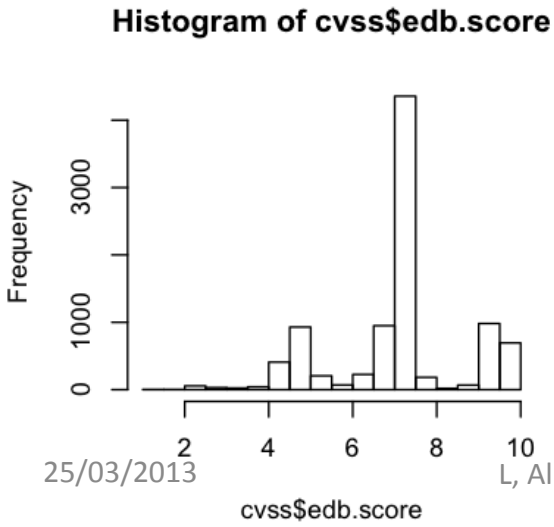


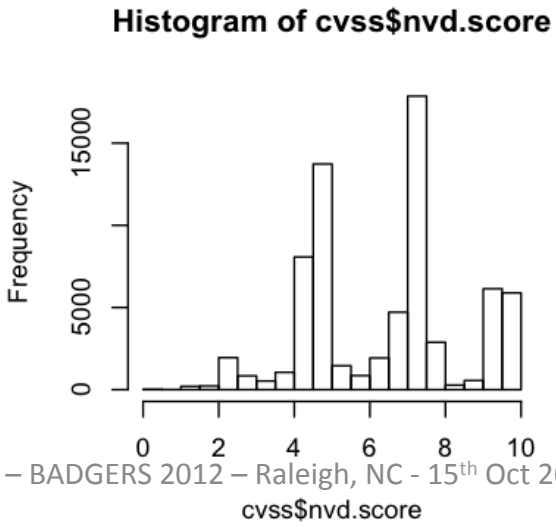
- LOW: CVSS <6
- MEDIUM: 6<CVSS<9
- HIGH: CVSS > 9

Distribution of CVSS Scores (HIST)





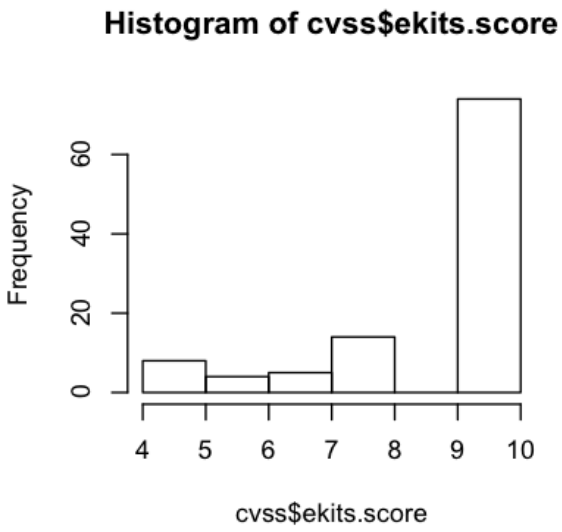


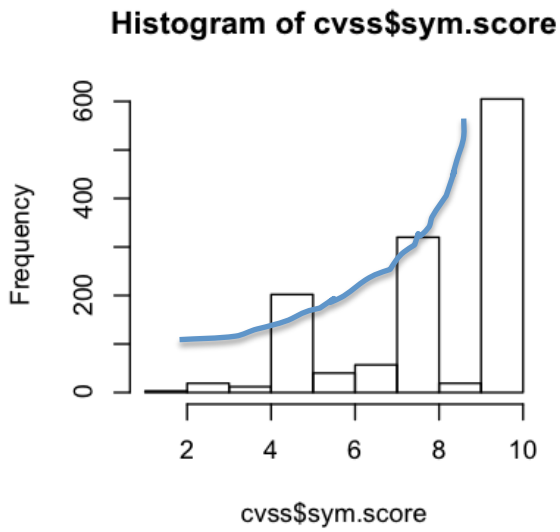


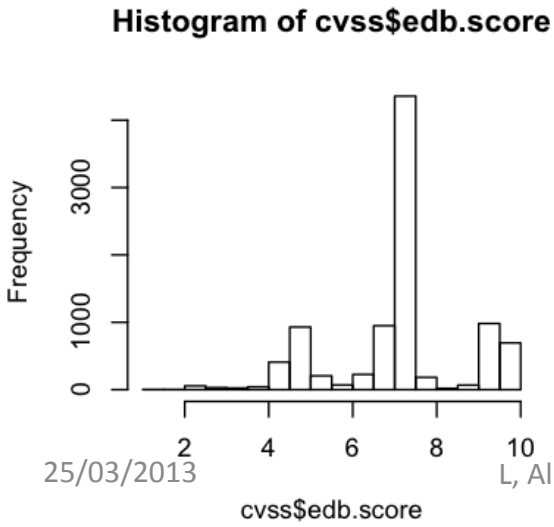
20% of vulnerabilities in SYM are scored LOW

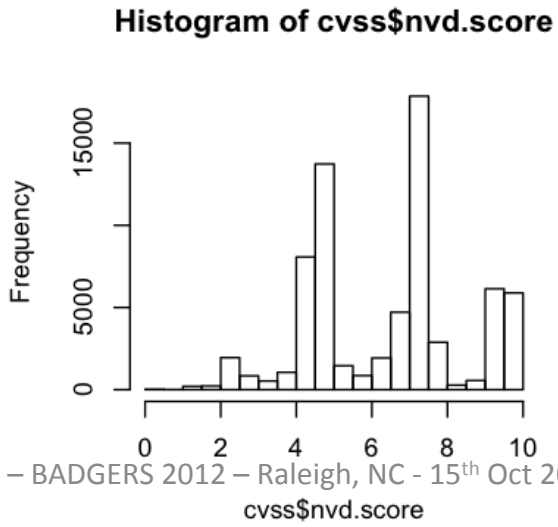
Are we sure CVSS and risk correlate?

Distribution of CVSS Scores (HIST)

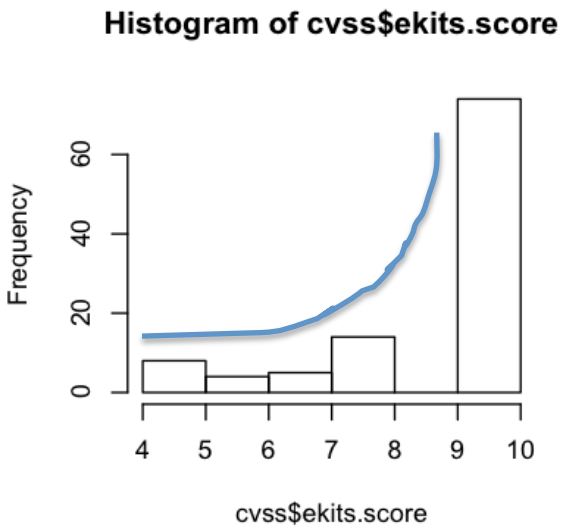


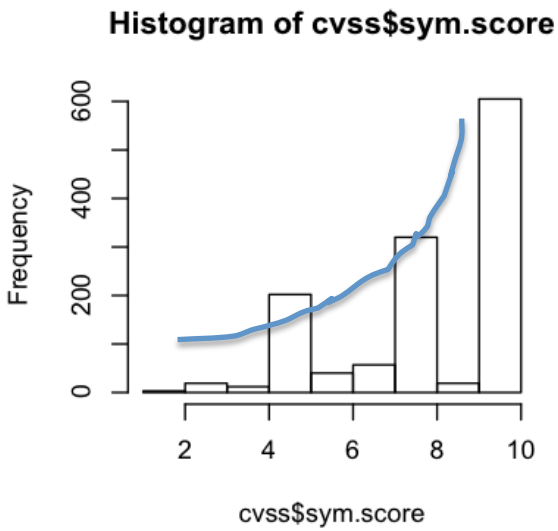


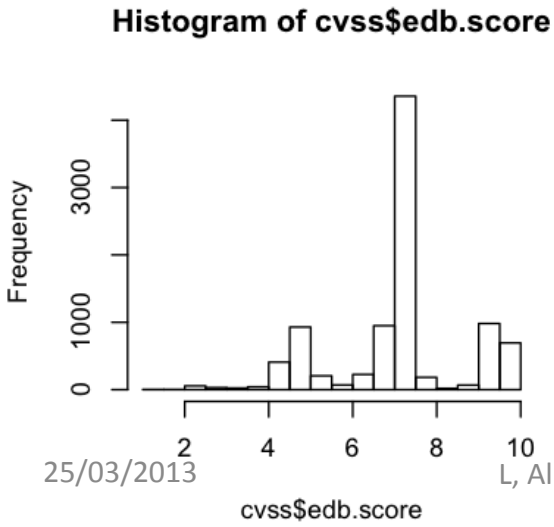


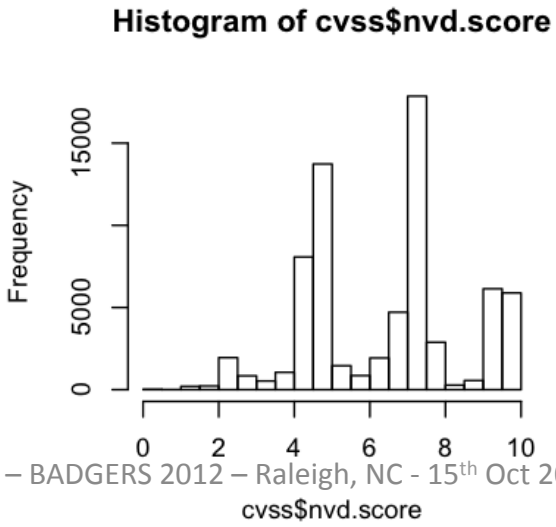


Distribution of CVSS Scores (HIST)

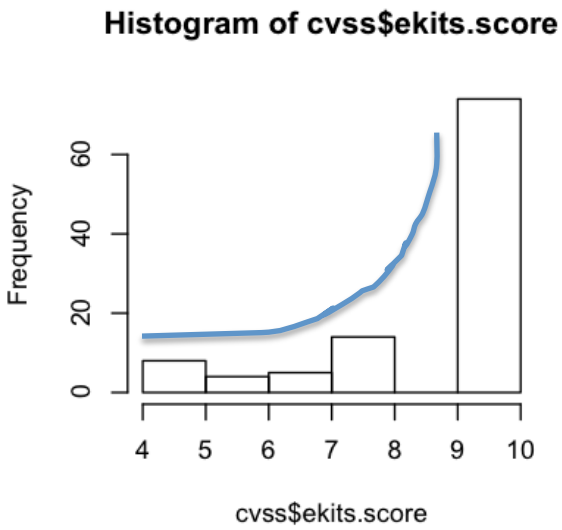


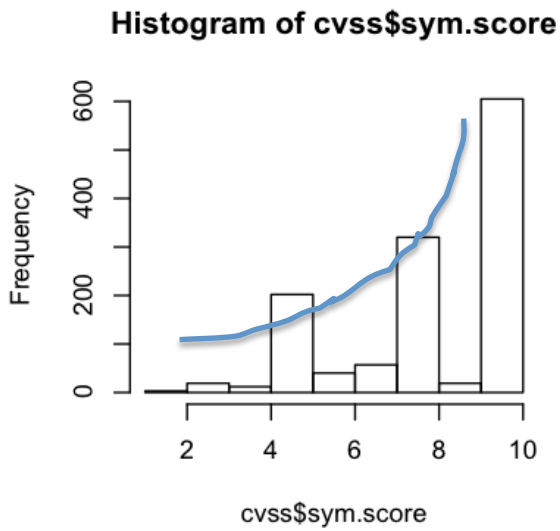


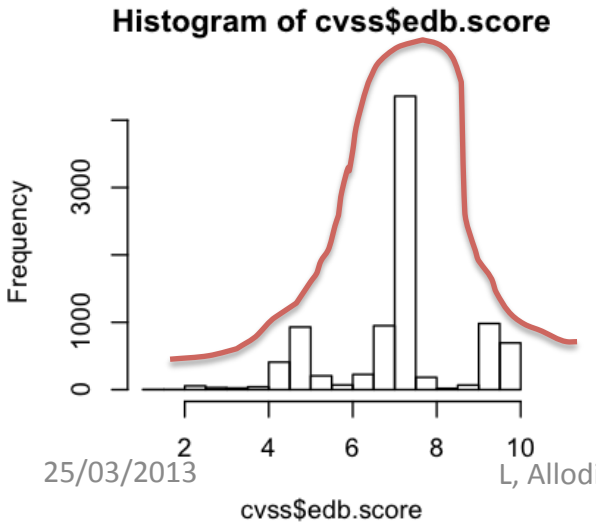


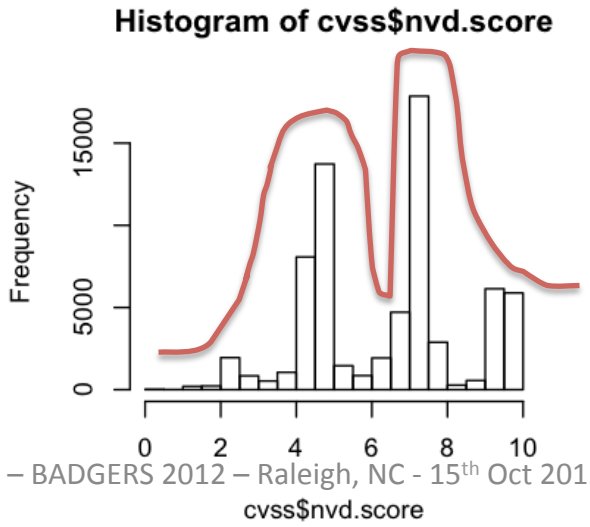


Distribution of CVSS Scores (HIST)

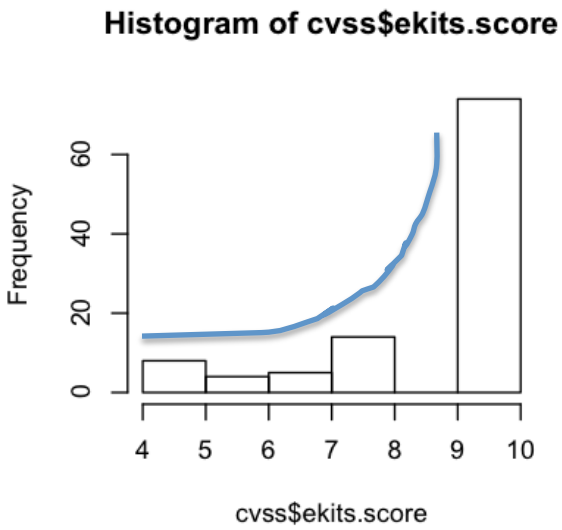


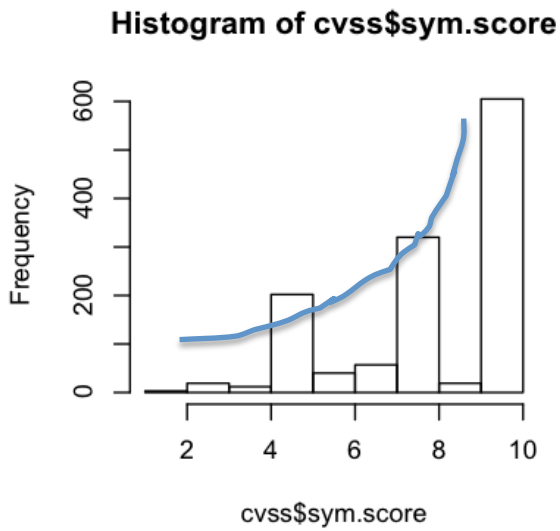


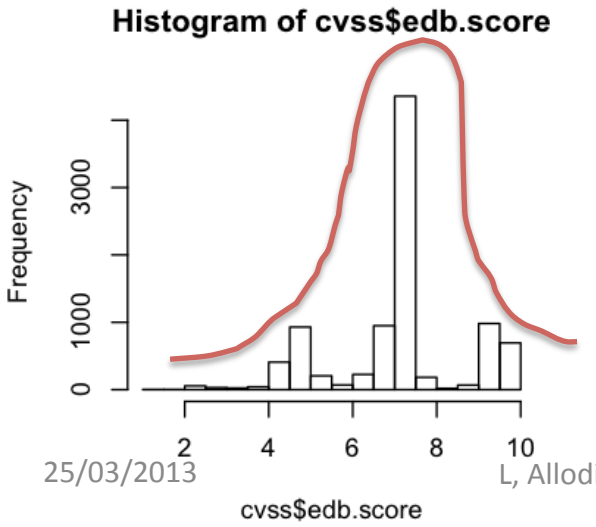


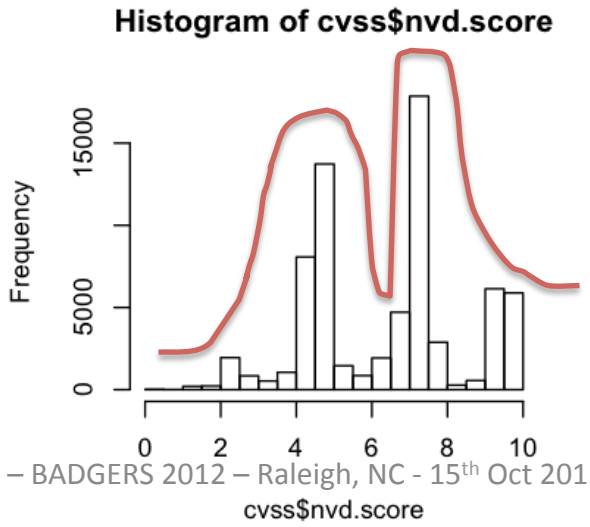


Distribution of CVSS Scores (HIST)









EKITS and SYM distributions are completely different from EDB and NVD.

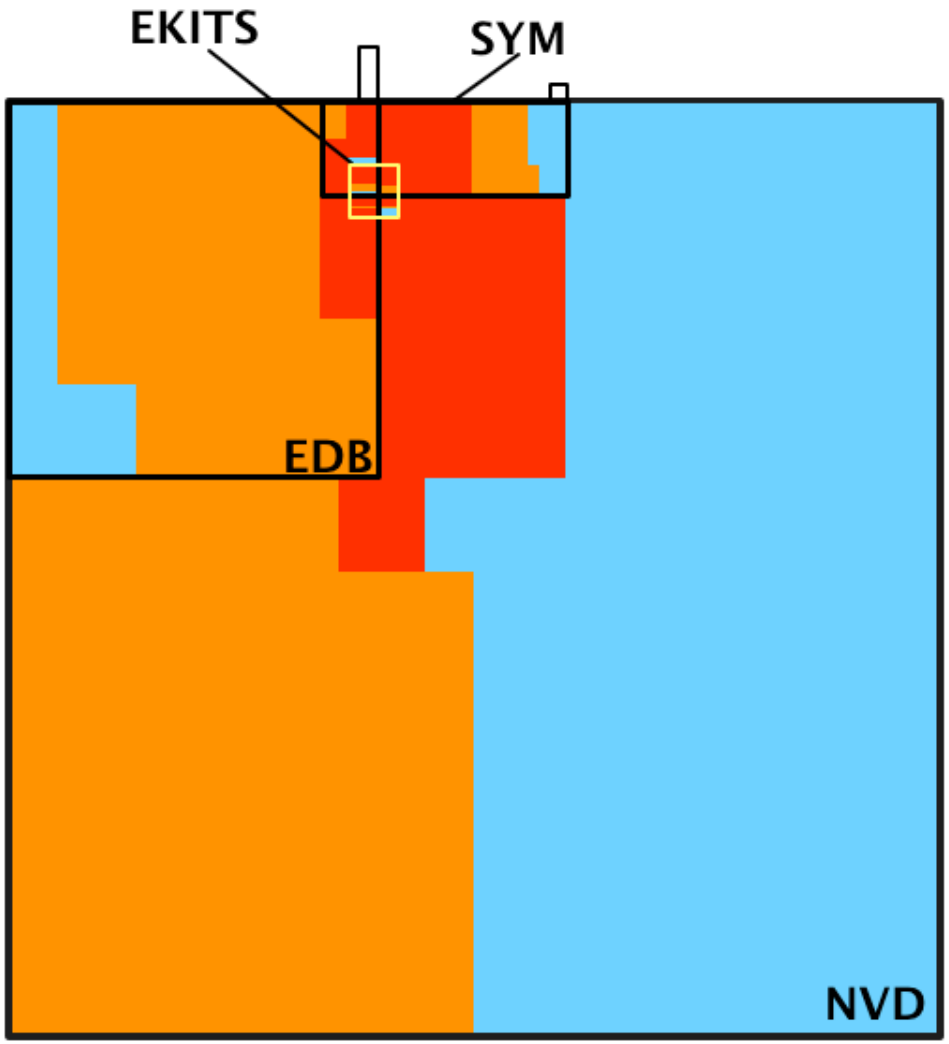
We need to be careful with statistical analyses on different populations. Sampling correctly might be tricky.

Distribution of CVSS Scores (Table)

CVSS Score	EKITS	SYM	EDB	NVD
HIGH	74	612	1.209	7.026
MEDIUM	19	393	5.324	20.858
LOW	10	272	1.589	21.715
Tot	**103**	**1.277**	**8.122**	**49.599**

- 20% of vulnerabilities in SYM are scored LOW
- EKITS and SYM distributions are completely different from EDB and NVD

Distribution of CVSS Scores (VENN)



LOW CVSS

MEDIUM CVSS

HIGH CVSS

Distribution of CVSS Scores (VENN)

Distribution of CVSS Score (ENG)

- SYM sees only some vulns with high and medium scores
  - Recall vuln in SYM → vuln used by bad guys
- EKITs sell mostly vulns with high scores
- NVD and EDB have lots but really lots of totally uninteresting vulns
  - If you are using the NVD or EDB to assess your company status (eg SCAP) → Maybe you're worrying too much
- CVSS scores tell something, but not enough
  - It's good for witch hunt - "Kill them all, God will recognize its brethren"
  - Maybe we can tell something more by looking at metrics for likelihood-of-exploitation (Exploitability) [5]

[5] Mehran Bozorgi and Lawrence~K. Saul and Stefan Savage and Geoffrey~M. Voelker. ***Beyond Heuristics: Learning to Classify Vulnerabilities and Predict Exploits. SIGKDD-10***

Distribution of CVSS Score (ENG)

- If you are using the NVD or EDB to assess your company status (eg SCAP) → Maybe you're worrying too much
- CVSS scores tell something, but not enough
  - Looks like witch hunt - "Kill them all, God will recognize its brethren"
- Maybe we can tell something more by looking at metrics for likelihood-of-exploitation (Exploitability) [5]

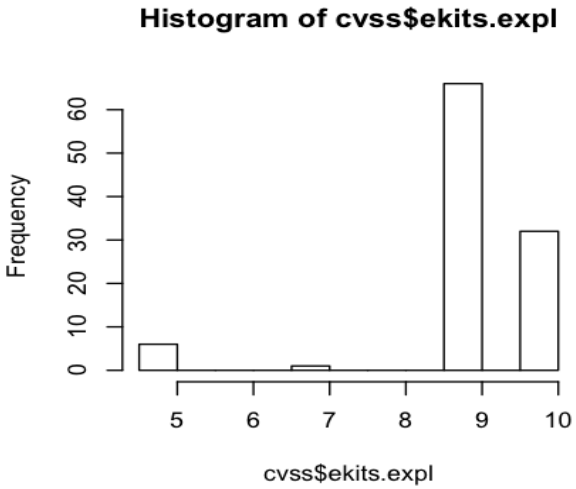
[5] Mehran Bozorgi and Lawrence~K. Saul and Stefan Savage and Geoffrey~M. Voelker. ***Beyond Heuristics: Learning to Classify Vulnerabilities and Predict Exploits. SIGKDD-10***

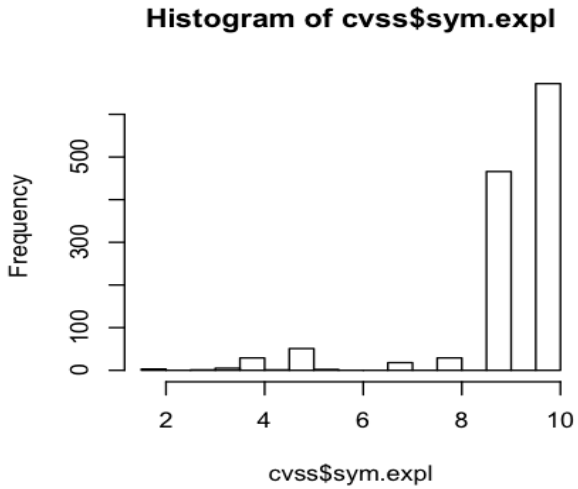
Are datasets representative for exploits?

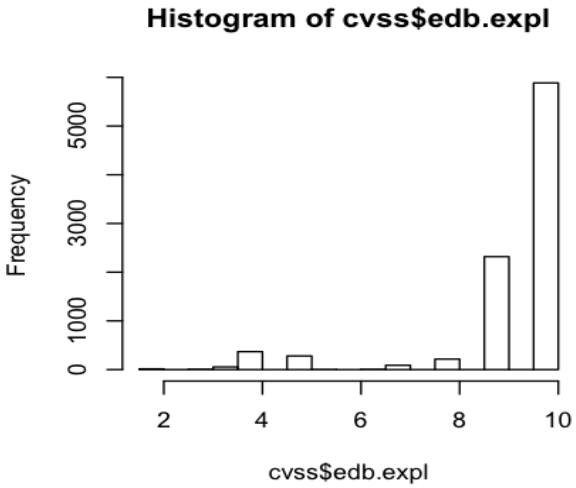
P(x Threat | in DB)	EKITS	EDB	NVD
SYM	75.73%	4.08%	2.10%
NOT SYM	24.27%	95.92%	97.90%

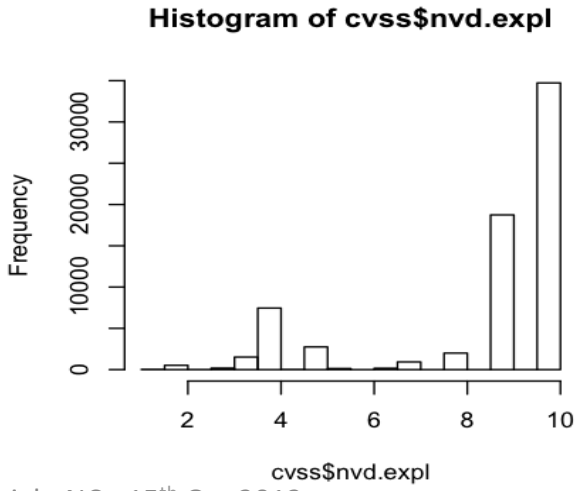
- If exploit is traded in the black market it is likely in the wild
- EDB and NVD report way too much data?
- Two possibilities with available data and metrics:
  - SYM is widely incomplete
  - Only high-CVSS vulns in EDB and NVD are to be expected to be in SYM
- To rule these options out we extended the study: population sampling

Distribution of CVSS Exploitability









Distribution of CVSS Exploitability Subscores

Authentication	Access vector	Access complexity
None	Network	High
Single	Adjacent Network	Medium
Multiple	Local	Low

Distribution of CVSS Exploitability Subscores

Authentication	Access vector	Access complexity
None	Network	High
Single	Adjacent Network	Medium
Multiple	Local	Low

Likelihood of exploitation

Distribution of CVSS Exploitability Subscores

Authentication	Access vector	Access complexity
None	Network	High
Single	Adjacent Network	Medium
Multiple	Local	Low

dataset	Exploitability SubScore		
	Auth: none	Acc. Vector: Network	Acc. Complexity: Low or Medium
NVD	95.45%	87.31%	95.46%
EDB	96.27%	95.31%	96.73%
EKITS	99.03%	100.00%	95.15%
SYM	96.08%	96.79%	95.77%

Limitations

- Everything is exploitable → Exploitability score is not an interesting variable at all!
- Looking at Bozorgi et al. SIGKDD'10[5]
  - Confirm finding → CVSS exploitability score does not correlate well to "exploits"
- Still, Exploitability → Pr(v in SYM)
  - We use CVSS Exploitability submetrics to sample the populations

[5] Mehran Bozorgi and Lawrence~K. Saul and Stefan Savage and Geoffrey~M. Voelker. ***Beyond Heuristics: Learning to Classify Vulnerabilities and Predict Exploits. SIGKDD-10***

CVSS subfactors: Exploitability

		SYM	EKITS	EDB	NVD
access vector	local	2.98%	0.00%	4.57%	13.18%
	adj	0.23%	0.00%	0.12%	0.35%
	net	96.79%	100.00%	95.31%	87.31%
access complexity	high	4.23%	4.85%	3.37%	4.54%
	medium	38.53%	63.11%	25.49%	30.42%
	low	57.24%	32.04%	71.14%	65.68%
authentication	multiple	0.00%	0.00%	0.02%	0.05%
	single	3.92%	0.97%	3.71%	5.35%
	none	96.08%	99.03%	96.27%	95.45%

These are our control variables to sample populations identically distributed to SYM

A few observations

- Everything is exploitable → Exploitability is not an interesting variable at all!
- Looking at Bozorgi et al. SIGKDD'10
  - Took OVSDB (basically Exploit-DB) and compared SVM machine learning vs CVSS exploitability
  - Confirm finding → CVSS exploitability score does not correlate well to "exploits"

New validation: controlled experiment (medical fashion)

- Do smoking habits predict cancer?
  - R Doll & A Bradford Hill, BMJ
  - You can't ask people to start smoking so you can't run a controlled experiment
- Case controlled study
  - Cases: people with lung cancer
  - Controls (Possible confounding variables)
    - 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
- Do high risk vulns (smoking) predict attacks (cancer)?

New validation: controlled experiment (CS fashion)

- Our case controlled study
  - Cases: vulns with exploits in the wild
  - Controls (Possible confounding variables)
    - Access vector, access complexity, authentication
  - Explanatory variables
    - CVSS Score, Database
- CVSS Score is a "test". How to evaluate it?
  - Sensitivity → ability to identify true positives
  - Specificity → ability to rule-out true negatives

CVSS as "should I worry" test

- Tests on a random population identically distributed to SYM
  - Conclusions with *p<2.2E-16*
- Sensitivity: Pr(vuln.score >= 6 | vuln in SYM)
- Specificity: Pr(vuln.score < 6 | vuln in ! SYM)

DB	Sensitivity	Specificity
EKITS	96.30%	36.19%
EDB	93.85%	18.69%
NVD	76.92%	43.24%

Preliminary Conclusions

- Where should we look for "real" exploits?
  - EDB, NVD are the wrong datasets.
- The CVSS score is a good predictor for exploitation only occasionally (Sensitivity)
  - Not for the NVD dataset
- No datasets show high Specificity:
  - CVSS doesn't rule out "un-interesting" vulns
  - Vendors, Policy Makers, Researchers are doomed to treat 60-80% false positives (thousands of vulnerabilities)
- *"Don't quote me on that":*
  - *Big European Vendor estimates 100$ cost just to start addressing a bug (i.e. acknowledge the problem should be addressed and allocating human/IT resources)*

Future Work

- Address limits of our study
  - Not all control variables considered
  - For example TIME (= Age in Smoking), GeoLocation? Affected Software or platform?
- WINE Symantec Database
  - Correlate actual temporal occurrences and frequencies of exploits with temporal discovery of vulnerabilities and presence in the EKITS and EDB
  - Control experiments with data on system configuration
- Final goal:
  - Identify the explanatory variables for exploitation and tune a better Risk Test
  - Evaluate black market dynamics and its influence in attack trends (i.e. risk metrics)

Questions?

FAQ

- Do you think SYM and EKITS are representative of actual attacks?
  - This is a start at looking at real attack data. We are costantly working to replenish the dataset and data is growing fast.
- Nobody ever said NVD and EDB had to represent attacks.
  - True. Still, most studies rely on them to assess software risk. Here we simply checked if these datasets are meaningful for THAT purpose – not in general. We found that real risks do not map nicely in neither of them.
- The CVSS score is meant as a static metric, not as a representation of dynamic internet usage.
  - One exploitation is sufficient for us to be considered. We are not measuring volumes of attacks against vulnerabilities. We are looking for stronger correlations between exploit markets and vulnerability characteristics to build a meaningful "exploitation metric".

Questions?

FAQ

- Do you think SYM and EKITS are representative of actual attacks?
  - This is a start at looking at real attack data. We are costantly working to replenish the dataset and data is growing fast.
- Nobody ever said NVD and EDB had to represent attacks.
  - True. Still, most studies rely on them to assess software risk. Here we simply checked if these datasets are meaningful for *that*purpose – not in general. We found that real risks do not map nicely in neither of them.