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

[2] Shahzad, Muhammad and Shafiq, Muhammad Zubair and Liu, Alex X. *A large scale exploratory analysis of software vulnerability life cycles*. 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.” [3]

Vulnerability Research cnt’d

Our question is:
Is the data we’re looking at by any means representative of actual attacks?

<table>
<thead>
<tr>
<th>dataset</th>
<th>volume of CVEs</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVD</td>
<td>49.624</td>
</tr>
<tr>
<td>EDB</td>
<td>8.189</td>
</tr>
<tr>
<td>EKITS</td>
<td>103</td>
</tr>
<tr>
<td>Symantec (SYM)</td>
<td>1.289</td>
</tr>
</tbody>
</table>
What do bad guys actually need?

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

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

*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]

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)

Histogram of cvss$ekits.score

Histogram of cvss$sym.score

Histogram of cvss$edb.score

Histogram of cvss$nvd.score

25/03/2013

Distribution of CVSS Scores (HIST)

Histogram of cvss$ekits.score

Histogram of cvss$sym.score

Histogram of cvss$edb.score

Histogram of cvss$nvd.score
Distribution of CVSS Scores (HIST)

Histogram of cvss$ekits.score

Histogram of cvss$sym.score

Histogram of cvss$edb.score

Histogram of cvss$nvd.score
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)

<table>
<thead>
<tr>
<th>CVSS Score</th>
<th>EKITS</th>
<th>SYM</th>
<th>EDB</th>
<th>NVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH</td>
<td>74</td>
<td>612</td>
<td>1.209</td>
<td>7.026</td>
</tr>
<tr>
<td>MEDIUM</td>
<td>19</td>
<td>393</td>
<td>5.324</td>
<td>20.858</td>
</tr>
<tr>
<td>LOW</td>
<td>10</td>
<td>272</td>
<td>1.589</td>
<td>21.715</td>
</tr>
<tr>
<td>Tot</td>
<td>103</td>
<td>1.277</td>
<td>8.122</td>
<td>49.599</td>
</tr>
</tbody>
</table>

- 20% of vulnerabilities in SYM are scored LOW
- EKITS and SYM distributions are completely different from EDB and NVD
Distribution of CVSS Scores (VENN)

EQUITIS

SYM

EDB

NVD

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]

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]

Are datasets representative for exploits?

<table>
<thead>
<tr>
<th>P(x Threat</th>
<th>in DB)</th>
<th>EKITS</th>
<th>EDB</th>
<th>NVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYM</td>
<td>75.73%</td>
<td>4.08%</td>
<td>2.10%</td>
<td></td>
</tr>
<tr>
<td>NOT SYM</td>
<td>24.27%</td>
<td>95.92%</td>
<td>97.90%</td>
<td></td>
</tr>
</tbody>
</table>

- 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

Histogram of cvss$ekits.expl

Histogram of cvss$sym.expl

Histogram of cvss$edb.expl

Histogram of cvss$nvd.expl
## Distribution of CVSS Exploitability Subscores

<table>
<thead>
<tr>
<th>Authentication</th>
<th>Access vector</th>
<th>Access complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Network</td>
<td>High</td>
</tr>
<tr>
<td>Single</td>
<td>Adjacent Network</td>
<td>Medium</td>
</tr>
<tr>
<td>Multiple</td>
<td>Local</td>
<td>Low</td>
</tr>
</tbody>
</table>
Distribution of CVSS Exploitability Subscores

<table>
<thead>
<tr>
<th>Authentication</th>
<th>Access vector</th>
<th>Access complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Network</td>
<td>High</td>
</tr>
<tr>
<td>Single</td>
<td>Adjacent Network</td>
<td>Medium</td>
</tr>
<tr>
<td>Multiple</td>
<td>Local</td>
<td>Low</td>
</tr>
</tbody>
</table>

Likelihood of exploitation
# Distribution of CVSS Exploitability Subscores

## Table: CVSS Exploitability Subscores

<table>
<thead>
<tr>
<th>Authentication</th>
<th>Access vector</th>
<th>Access complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Network</td>
<td>High</td>
</tr>
<tr>
<td>Single</td>
<td>Adjacent Network</td>
<td>Medium</td>
</tr>
<tr>
<td>Multiple</td>
<td>Local</td>
<td>Low</td>
</tr>
</tbody>
</table>

## Exploitability SubScore Distribution

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Auth: none</th>
<th>Acc. Vector: Network</th>
<th>Acc. Complexity: Low or Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVD</td>
<td>95.45%</td>
<td>87.31%</td>
<td>95.46%</td>
</tr>
<tr>
<td>EDB</td>
<td>96.27%</td>
<td>95.31%</td>
<td>96.73%</td>
</tr>
<tr>
<td>EKITS</td>
<td>99.03%</td>
<td>100.00%</td>
<td>95.15%</td>
</tr>
<tr>
<td>SYM</td>
<td>96.08%</td>
<td>96.79%</td>
<td>95.77%</td>
</tr>
</tbody>
</table>
Limitations

• Everything is exploitable $\rightarrow$ Exploitability score is not an interesting variable at all!

• Looking at Bozorgi et al. SIGKDD’10[5]
  – Confirm finding $\rightarrow$ CVSS exploitability score does not correlate well to “exploits”

• Still, Exploitability $\rightarrow$ Pr(v in SYM)
  – We use CVSS Exploitability submetrics to sample the populations

## CVSS subfactors: Exploitability

<table>
<thead>
<tr>
<th>access vector</th>
<th>SYM</th>
<th>EKITS</th>
<th>EDB</th>
<th>NVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>local</td>
<td>2.98%</td>
<td>0.00%</td>
<td>4.57%</td>
<td>13.18%</td>
</tr>
<tr>
<td>adj</td>
<td>0.23%</td>
<td>0.00%</td>
<td>0.12%</td>
<td>0.35%</td>
</tr>
<tr>
<td>net</td>
<td>96.79%</td>
<td>100.00%</td>
<td>95.31%</td>
<td>87.31%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>access complexity</th>
<th>SYM</th>
<th>EKITS</th>
<th>EDB</th>
<th>NVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>4.23%</td>
<td>4.85%</td>
<td>3.37%</td>
<td>4.54%</td>
</tr>
<tr>
<td>medium</td>
<td>38.53%</td>
<td>63.11%</td>
<td>25.49%</td>
<td>30.42%</td>
</tr>
<tr>
<td>low</td>
<td>57.24%</td>
<td>32.04%</td>
<td>71.14%</td>
<td>65.68%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>authentication</th>
<th>SYM</th>
<th>EKITS</th>
<th>EDB</th>
<th>NVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>multiple</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.02%</td>
<td>0.05%</td>
</tr>
<tr>
<td>single</td>
<td>3.92%</td>
<td>0.97%</td>
<td>3.71%</td>
<td>5.35%</td>
</tr>
<tr>
<td>none</td>
<td>96.08%</td>
<td>99.03%</td>
<td>96.27%</td>
<td>95.45%</td>
</tr>
</tbody>
</table>

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: $\Pr(\text{vuln.score} \geq 6 \mid \text{vuln in SYM})$
• Specificity: $\Pr(\text{vuln.score} < 6 \mid \text{vuln in} \neg \text{SYM})$

<table>
<thead>
<tr>
<th>DB</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>EKITS</td>
<td>96.30%</td>
<td>36.19%</td>
</tr>
<tr>
<td>EDB</td>
<td>93.85%</td>
<td>18.69%</td>
</tr>
<tr>
<td>NVD</td>
<td>76.92%</td>
<td>43.24%</td>
</tr>
</tbody>
</table>
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)
FAQ

• Do you think SYM and EKITS are representative of actual attacks?
  – This is a start at looking at real attack data. We are constantly 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”.

25/03/2013
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 constantly 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.