Quantitative Analyses of Software Vulnerabilities

HyunChul Joh, Ph.D.
dean2026@cs.colostate.edu
CS530 Guest Speech
2012-MAR-01
Main idea of the Work

“If you cannot measure it, you cannot improve it.”
- Lord Kelvin (William Thomson) –

Measuring Vulnerability Activities

- Comparison of software systems
- Optimal software deployment
- Patch management
- Risk remediation
- Etc.
Qualitative v.s. Quantitative

Qualitative approach

- Heavily depends on the experts’ opinions.
- Tends to be subjective.
- More like an “Art” which does not depend on definitions[1].

Quantitative approach

- Actual data-driven empirical analysis followed by statistical tests.
- Requires enough datasets.
- More like a “Science” which depends on definitions[1].

Software Vulnerabilities

Definition

Defects in the security system which might be exploited by malicious users causing loss or harm[2]

Standards

Public online databases

Vulnerability Discovery Model (VDM)

• Vulnerability Discovery Model (VDM):
  – Modeling discovery process of software vulnerabilities
  – Estimating number of vulnerabilities
  – To be used to assess risk & to evaluate possible mitigation approaches

• Existing models:
  – Linear: constant discovery rate
  – Exponential: exponentially declining rate
  – Alhazmi Malaiya Logistic (AML) model: First S-shaped
  – Effort based model
  – Etc.
Publications

Vulnerability discovery process


Risk assessment


Periodic behavior in Vulnerability activities


Under review / In preparation

Motivation

AML performs well ... but,

It is based on Logistic PDF

How about other S-shaped PDFs:
Normal, Weibull, Beta, Gamma, ...

Performance test:
Model fitting & Prediction error
Skewness

left/negative skewness

Zero skewness (symmetrical)

right/positive skewness

\[ Skewness = \frac{n}{(n - 1)(n - 2)} \sum \left( \frac{x_i - \bar{x}}{s} \right)^3 \]
The five distributions

- 3-phases for S-shaped models
- Normal Distribution
- Logistic Distribution
- Weibull Distribution
- Gamma Distribution
- Beta Distribution
Datasets - Run charts

+ 0.4539
- 7.064
- 0.1394
+ 0.2636

+ 0.342
+ 1.0034
- 0.5436
+ 0.0236
Model fitting

Red Hat Linux

MAC OSX

Windows XP

Windows Server 2003

Apache Web Server

Internet Information Server

Internet Explorer

Firefox
Prediction error

Red Hat Linux

MAC OSX

Windows XP

Windows Server 2003

Apache Web Server

Internet Information Server

Internet Explorer

Firefox
Average Bias & Average Error

\[ AB = \frac{1}{n} \sum_{t=1}^{n} \frac{\Omega_t - \bar{\Omega}}{\bar{\Omega}} \]

\[ AE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{\Omega_t - \bar{\Omega}}{\bar{\Omega}} \right| \]
Section summary

Model Fitting
- Hard to observe differences
- All VDMs perform well

Prediction Capability
- Gamma VDM performs better w/ right skew
- Symmetrical VDMs perform better w/ left skew
- Good model fitting       Good Prediction capability
Outline

1. Introduction
2. Modeling Skewness
3. Risk Assessment
4. Periodic Behavior
5. Summary
Risk

Definition\textsuperscript{[3]}

- 

So... it is

- Probability: Internal vulnerability & External threat
- Impact: Value of loss

Common Vulnerability Scoring System (CVSS)

- *De facto* industry standard for assessing the severity of computer security vulnerabilities
- Ranges 0.0 – 10.0
- Effort could be prioritized
- Consists three metric groups:
  - Base (mandatory) : intrinsic and fundamental
  - Temporal : change over time
  - Environmental : particular user’s environment
CVSS Base metric

**Exploitability sub-score**
- captures how a vulnerability is accessed and whether or not extra conditions are required to exploit it in terms of Access Vector, Access Complexity, and Authentication. [0.0 10.0]

**Impact sub-score**
- measures how a vulnerability, if exploited, will directly affect an IT asset in terms of Confidentiality, Integrity, and Availability. [0.0 10.0]

**Base score formula[4]**
- Base score

Distribution of Base score

By Base score formula

Impact x Exploitability

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
<th>Combinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>0</td>
<td>5</td>
<td>6.8</td>
<td>6.341</td>
<td>7.5</td>
<td>10</td>
<td>63</td>
</tr>
<tr>
<td>(b)</td>
<td>0</td>
<td>29</td>
<td>49</td>
<td>48.59</td>
<td>64</td>
<td>100</td>
<td>112</td>
</tr>
</tbody>
</table>

NVD on Jan 2011 (44615 vuln.)
Evaluating risk level for known unpatched vulnerabilities

Datasets: Discovery Trend

Dataset gathered on Jan. 2011
Simulated patch dates

Average patch time [6]

<table>
<thead>
<tr>
<th></th>
<th>0-day</th>
<th>30-day</th>
<th>90-day</th>
<th>180-day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft</td>
<td>61%</td>
<td>75%</td>
<td>88%</td>
<td>94%</td>
</tr>
<tr>
<td>Apple</td>
<td>32%</td>
<td>49%</td>
<td>71%</td>
<td>88%</td>
</tr>
</tbody>
</table>

Simulated patch date

<table>
<thead>
<tr>
<th></th>
<th>OS 1</th>
<th>OS 2</th>
<th>Browser 1</th>
<th>Browser 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 day</td>
<td>289</td>
<td>33</td>
<td>54</td>
<td>14</td>
</tr>
<tr>
<td>1-30</td>
<td>66</td>
<td>18</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>31-90</td>
<td>61</td>
<td>23</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>91-180</td>
<td>28</td>
<td>18</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>No patch</td>
<td>30</td>
<td>14</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>474</td>
<td>106</td>
<td>89</td>
<td>44</td>
</tr>
</tbody>
</table>

Risk evaluation

OS 1

OS 2

OS 1 v.s. OS 2

Browser 1

Browser 2

Browser 1 v.s. Browser 2
Section summary

Possibility of revising CVSS equation

- More rational and simpler way

Formal measures of software security risk

- Based on known vulnerabilities
- CVSS base score
  - Exploitability sub score as probability
  - Impact sub score as impact
Outline

1. Introduction
2. Modeling Skewness
3. Risk Assessment
4. Periodic Behavior
5. Summary
Seasonality

- Predictable & periodic behavior based on calendar time
- Well known statistical approach
- Common in other fields

Seasonality in IT & Security area

- Recognized by some experts
- No scientific studies have been provided in vulnerability activities

This study

- Tries to provide some of the statistical evidences for seasonal/periodic behavior in vulnerability related to activities
Vulnerability activities considered

Vulnerability discovery process

- Long term (Annual)
- Eighteen software systems (NVD; Analysis only for six Windows OSes in this presentation)

Exploitation pattern and % of unpatched vuln.

- Short term (weekly)
- 104 million global vuln. Scans during 2008 (Qualys [7])

Statistical methodologies

- Seasonal index
- Autocorrelation Function (ACF) analysis
- ANOVA w/ LSD test

Seasonal index

Means...

• How much the average for a particular period tends to be above (or below) the expected value

Seasonal index equation

• We will evaluate it using the monthly seasonal index values given by [8]:
  - $s_i$: seasonal index for $i^{th}$ month
  - $d_i$: mean value of $i^{th}$ month
  - $d$: grand average

Autocorrelation function (ACF)

Mathematically

• With time series values of \( z_b, z_{b+1}, ..., z_n \), the ACF at lag \( k \), denoted by \( r_k \), is \([9]\):

\[
\hat{r}_k = \frac{\sum (z_i - \bar{z})(z_{i+k} - \bar{z})}{\sum (z_i - \bar{z})^2}, \quad \text{where}
\]

In plain English

• Measures the linear relationship between time series observations separated by a lag of time units

For the analysis

• When an ACF value is located outside of confidence intervals at a lag \( t \), it can be thought that every lag \( t \), there is a relationships along with the time line

ANOVA with LSD - Long term

ANOVA table for Seasonal Index

<table>
<thead>
<tr>
<th>Windows</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F_{crit}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>13.5128</td>
<td>11</td>
<td>1.228436</td>
<td>8.259795</td>
<td>1.3712-08</td>
<td>1.952212</td>
</tr>
<tr>
<td>Within Groups</td>
<td>8.923489</td>
<td>60</td>
<td>0.148725</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>22.43629</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LSD test for Seasonal Index; LSD = 0.4453

<table>
<thead>
<tr>
<th>Month</th>
<th>JAN</th>
<th>FEB</th>
<th>MAR</th>
<th>APR</th>
<th>MAY</th>
<th>JUN</th>
<th>JUL</th>
<th>AUG</th>
<th>SEP</th>
<th>OCT</th>
<th>NOV</th>
<th>DEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.9869</td>
<td>0.7756</td>
<td>0.6327</td>
<td>0.5827</td>
<td>1.2827</td>
<td>1.1370</td>
<td>1.1837</td>
<td>0.7260</td>
<td>0.9383</td>
<td>1.1820</td>
<td>0.9233</td>
<td>2.1678</td>
</tr>
<tr>
<td>JAN</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FEB</td>
<td>0.7756</td>
<td></td>
<td>0</td>
<td>0.1506</td>
<td>0.1800</td>
<td>0.5343</td>
<td>0.7916</td>
<td>0.1232</td>
<td>0.3825</td>
<td>0.1933</td>
<td>0.5293</td>
<td>0.3205</td>
</tr>
<tr>
<td>MAR</td>
<td>0.6327</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5343</td>
<td>0.6116</td>
<td>0.0567</td>
<td>0.2026</td>
<td>0.3738</td>
<td>0.3493</td>
</tr>
<tr>
<td>APR</td>
<td>0.5827</td>
<td>0.1506</td>
<td>0</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0.3543</td>
<td>0.6116</td>
<td>0.0567</td>
<td>0.2026</td>
<td>0.3738</td>
<td>0.3493</td>
</tr>
<tr>
<td>MAY</td>
<td>1.2827</td>
<td>0.1800</td>
<td>0.5343</td>
<td>0.7916</td>
<td>0.1232</td>
<td>0.3825</td>
<td>1.2827</td>
<td>0.7282</td>
<td>0.1518</td>
<td>0.0650</td>
<td>0.2138</td>
<td>1.0308</td>
</tr>
<tr>
<td>JUN</td>
<td>1.1370</td>
<td>1.1837</td>
<td>0.7916</td>
<td>0.1232</td>
<td>0.3825</td>
<td>0.1933</td>
<td>0.5293</td>
<td>0.3205</td>
<td>1.5651</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JUL</td>
<td>1.1837</td>
<td>0.7282</td>
<td>0.5343</td>
<td>0.7916</td>
<td>0.1232</td>
<td>0.3825</td>
<td>0.1933</td>
<td>0.5293</td>
<td>0.3205</td>
<td>1.5651</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUG</td>
<td>0.9869</td>
<td>0.3543</td>
<td>0.6116</td>
<td>0.5343</td>
<td>0.7916</td>
<td>0.1232</td>
<td>0.3825</td>
<td>0.1933</td>
<td>0.5293</td>
<td>0.3205</td>
<td>1.5651</td>
<td></td>
</tr>
<tr>
<td>SEP</td>
<td>0.7260</td>
<td>0.2573</td>
<td>0.4111</td>
<td>0.6883</td>
<td>0.4090</td>
<td>0.2623</td>
<td>0.7282</td>
<td>0.4710</td>
<td>0.7735</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCT</td>
<td>1.1320</td>
<td>0.2573</td>
<td>0.4111</td>
<td>0.6883</td>
<td>0.4090</td>
<td>0.2623</td>
<td>0.7282</td>
<td>0.4710</td>
<td>0.7735</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOV</td>
<td>0.9233</td>
<td>0.2573</td>
<td>0.4111</td>
<td>0.6883</td>
<td>0.4090</td>
<td>0.2623</td>
<td>0.7282</td>
<td>0.4710</td>
<td>0.7735</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEC</td>
<td>2.1678</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

1) DEC > all others
2) JUN > FEB, MAR, APR, JUL, NOV, SEP
3) SEP < JAN, MAY, JUN, AUG, OCT, NOV, DEC
4) OCT > MAR, SEP
5) MAY > MAR
ACF – Long term

<table>
<thead>
<tr>
<th>Windows NT: 95% confidence interval = (-.152145, .152145)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Windows XP: 95% confidence interval = (-.1885976, .1885976)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Windows 2K: 95% confidence interval = (-.1633303, .1633303)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Windows Server 2003: 95% confidence interval = (-.2138496, .2138496)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Windows 95: 95% confidence interval = (-.1633303, .1633303)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Windows 98: 95% confidence interval = (-.1789194, .1789194)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Half-life (Critical vulnerabilities in 2008 / 29.5 days)
• Time interval required to cut its occurrence by half

Patch level (Prevalence / Adobe Reader & Acrobat)
• How much machines in the scan have patched their vulnerable software system

Exploitation (MS08-067: Windows Server Service Vulnerability)
• Detected number of exploitations
**Weekly Seasonal Index Values**

<table>
<thead>
<tr>
<th>Day</th>
<th>Half-life (2008)</th>
<th>Patch level (APSA09-01)</th>
<th>Exploitation (MS08-067)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mon / day1</td>
<td>1.0495</td>
<td>1.3091</td>
<td>1.0069</td>
</tr>
<tr>
<td>Tue / day2</td>
<td>1.4100</td>
<td>1.2945</td>
<td>1.2973</td>
</tr>
<tr>
<td>Wed / day3</td>
<td>1.3600</td>
<td>1.2570</td>
<td>1.0203</td>
</tr>
<tr>
<td>Thu / day4</td>
<td>0.7211</td>
<td>1.0805</td>
<td>1.0354</td>
</tr>
<tr>
<td>Fri / day5</td>
<td>0.5426</td>
<td>0.7046</td>
<td>0.9534</td>
</tr>
<tr>
<td>Sat / day6</td>
<td>0.9424</td>
<td>0.6784</td>
<td>0.7307</td>
</tr>
<tr>
<td>Sun / day7</td>
<td>0.9745</td>
<td>0.6759</td>
<td>0.9560</td>
</tr>
<tr>
<td>Chi statistic</td>
<td>12.5916</td>
<td>12.5916</td>
<td>12.5916</td>
</tr>
<tr>
<td>Chi critical</td>
<td>165.6114</td>
<td>236.8411</td>
<td>119.9789</td>
</tr>
<tr>
<td>p-value</td>
<td>3.83E-33</td>
<td>2.65E-48</td>
<td>1.65E-23</td>
</tr>
</tbody>
</table>
ACF – Short term

Very strong seven day periodic pattern!
Section summary

Statistically significant periodic behaviors

- Vulnerability discovery process (peak in mid-year & year-end)
- Percentage of patch level (Peak in Weekdays)
- Exploitation pattern (Peak in Weekdays)
Summary

S-shaped discovery process w/ Skew

- Four new S-shaped VDMs
- Suggest circumstances when to apply Gamma and AML
- Independence between model fitting and prediction ability

Software risk assessment

- Proposed a simple & handy method to assess software risk
- Possibility to revise CVSS base score

Seasonality

- Provides statistical evidences for the seasonality (recognized but no scientific studies so far)
Future works

In general

• Continually observe software vulnerability related activities

Specifically

• Mathematical analysis on constant discovery growth rates influenced by elements such as market share, software age and evolution, etc.
• Correlations between model parameters & software types
• Applying accurate patch information for assessing the risk model
• Non-homogeneous Semi-Markov process in risk assessment
• Possible enhancement of VDMs’ predicting ability by the longer and shorter seasonal fluctuations
• Etc.
Thank you.

• Questions?
• Comments?
• Suggestions?