Quantitative Cyber-Security
Colorado State University
Yashwant K Malaiya
CS559
L21

CSU Cybersecurity Center
Computer Science Dept
Pen Testing Stages

1. Planning and reconnaissance
2. Scanning
3. Gaining access
4. Maintaining access:
5. Analysis and remediation

Sources: 1, 2
Attacks and Attack trees

- Learn user's Unix password
  - Online guessing attack
  - Convince user to reveal
  - Obtain from password file
    - Obtain password hash
      - Threaten user
      - Social engineering
      - Dictionary attack
    - Dictionary attack +
      - Shadowed password file
      - World readable password file
  - Obtain when entered
    - Visual observation
    - Key logger
Topics

- Risk components
- Probability of a breach
- Gordon-Loeb Models
Quantitative Security

Colorado State University
Yashwant K Malaiya
CS 559
Breach probability

CSU Cybersecurity Center
Computer Science Dept
Formal definition:

- **Risk** due to an adverse event $e_i$
  \[
  \text{Risk}_i = \text{Likelihood}_i \times \text{Impact}_i
  \]
- Likelyhood$_i$ may be replaced by frequency$_i$, when it may happen multiple times a year.
- This yields the expected value. Sometimes a worst-case evaluation is needed.

In classical risk literature, the internal component of Likelihood is termed “Vulnerability” and external “Threat”. Both are probabilities. There the term “vulnerability” does not mean a security bug, as in computer security.
Risk as a composite measure

- Likelihood can be split in two factors
  \[ \text{Likelihood}_i = \text{P}\{\text{A security hole}_i \text{ is exploited}\}. \]
  \[ = \text{P}\{\text{hole}_i \text{ present}\}. \]
  \[ \text{P}\{\text{exploitation} | \text{hole}_i \text{ present}\} \]
- \(\text{P}\{\text{hole}_i \text{ present}\}\): an internal attribute of the system.
- \(\text{P}\{\text{exploitation} | \text{hole}_i \text{ present}\}\): depends on circumstances outside the system, including the adversary capabilities and motivation.
- In the literature, the terminology can be inconsistent.

Caution: In classical risk literature, the internal component of Likelihood is termed “Vulnerability” and external “Threat”. Both are probabilities. There the term “vulnerability” does not mean a security bug, as in computer security.
Annual Loss Expectancy (ALE)

Note the terminology is from the Risk literature.

- Annual loss expectancy (ALE). *(It is a risk measure)*
  
  \[ \text{ALE} = \text{SLE} \times \text{ARO} \]
  
  Where ARO is Annualized rate of occurrence.

- A countermeasure reduces the ALE by reducing one of its factors.

\[
\text{COUNTERMEASURE}_\text{VALUE} = (\text{ALE}_{\text{PREVIOUS}} - \text{ALE}_{\text{NOW}}) - \text{COUNTERMEASURE}_\text{COST}
\]

- \text{ALE}_{\text{PREVIOUS}}: ALE before implementing the countermeasure.
- \text{ALE}_{\text{NOW}}: ALE after implementing the countermeasure
- \text{COUNTERMEASURE}_\text{COST}: *annualized* cost of countermeasure
What factors impact the probability of an organization to be breached?

- Breach size
- Other factors:
  - Default value of factor = 1
    - Specific value relative to the default value
- Factors based on available data
  - Organization’s Country $F_{\text{country}}$
  - Organization’s Industry Classification $F_{\text{industry}}$
  - Sensitive Data Encryption $F_{\text{encryption}}$
  - Organization’s Privacy $F_{\text{privacy}}$
  - Business Continuity Management Team $F_{\text{BCM}}$
  - Data Breach Causes $F_{\text{breach_cause}}$
Modeling the Breach Probability

What factors impact the probability of an organization to be breached?

• Breach size

• Other factors:

• Default value of factor = 1
  – Specific value relative to the default value

• Do factors add or multiply?
  – Factors largely orthogonal: multiplicative
  – Factors overlap: additive

• Examples of multiplicative models
  – COCOMO Cost estimation model
  – RADC software defect density model
  – VLSI failure rate models
A proposed model for the probability of a breach for the next

\[ P \{\text{breach}\} = F_{\text{country}} \times F_{\text{BCM}} \times F_{\text{industry}} \times F_{\text{breach cause}} \times F_{\text{encryption}} \times F_{\text{privacy}} \times \alpha \exp\left(-\beta x\right) \]

Where \( \alpha = 0.4405, \ \beta = 4E-05, \ x \) the breach size 2015

Justification in the following slides.
Data Breach Probability


- 507 participating companies, with a minimum of 10,000 records
- United States, India, the United Kingdom, Germany, Brazil, Japan, France, the Middle East, Canada, Italy, South Korea, Australia, Turkey, ASEAN, South Africa, Scandinavia
Over the next two years, involving minimum of 10,000 and maximum of 100,000 records.

Data breach probability for the next two years based on the breach size (Ponemon data 2015)
Data breach probability by country (Ponemon data 2015)

A minimum of 10,000 compromised records
## Data breach probability by country

**Country Name** | **Multiplier**  
--- | ---  
USA | 1 (default)  
Germany | 0.72  
Canada | 0.74  
France | 1.62  
UK | 0.97  
Italy | 1.02  
Japan | 1.03  
Australia | 0.79  
Arabian Cluster (Saudi Arabia and United Arab Emirates) | 1.33  
Brazil | 1.68  
India | 1.35
Model proposed:

<table>
<thead>
<tr>
<th>Industry classification</th>
<th>Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communications</td>
<td>1.09</td>
</tr>
<tr>
<td>Consumer Products</td>
<td>1.24</td>
</tr>
<tr>
<td>Education</td>
<td>1.30</td>
</tr>
<tr>
<td>Financial Services</td>
<td>0.98</td>
</tr>
<tr>
<td>Government Services</td>
<td>1.63</td>
</tr>
<tr>
<td>Healthcare and Pharmaceuticals</td>
<td>1.26</td>
</tr>
<tr>
<td>Industrial</td>
<td>0.77</td>
</tr>
<tr>
<td>Retail</td>
<td>1.69</td>
</tr>
<tr>
<td>Services: professional and general services</td>
<td>1.48</td>
</tr>
<tr>
<td>Technology and software</td>
<td>1.26</td>
</tr>
<tr>
<td>Transportation</td>
<td>0.86</td>
</tr>
<tr>
<td>All others</td>
<td>1 (default)</td>
</tr>
</tbody>
</table>
Model proposed:

<table>
<thead>
<tr>
<th>BCM involved in incident response plan</th>
<th>Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>0.84</td>
</tr>
<tr>
<td>No</td>
<td>1.1</td>
</tr>
<tr>
<td>Not sure</td>
<td>1 (default)</td>
</tr>
</tbody>
</table>

Model proposed:

<table>
<thead>
<tr>
<th>Encryption of sensitive data</th>
<th>Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>0.64</td>
</tr>
<tr>
<td>No</td>
<td>1.03</td>
</tr>
<tr>
<td>Not sure</td>
<td>1 (default)</td>
</tr>
</tbody>
</table>
Model proposed:

<table>
<thead>
<tr>
<th>Organization’s Privacy</th>
<th>Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>A formal privacy and data protection program that is enterprise-wide</td>
<td>0.89</td>
</tr>
<tr>
<td>A formal privacy and data protection program that is not enterprise-wide</td>
<td>0.92</td>
</tr>
<tr>
<td>An informal privacy and data protection program that is enterprise-wide</td>
<td>1 (default)</td>
</tr>
<tr>
<td>An informal privacy and data protection program that is not enterprise-wide</td>
<td>1.19</td>
</tr>
<tr>
<td>No privacy or data protection program in place</td>
<td>1.42</td>
</tr>
</tbody>
</table>
Model proposed:

<table>
<thead>
<tr>
<th>Data breach causes</th>
<th>Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malicious or criminal attack</td>
<td>1.32</td>
</tr>
<tr>
<td>Negligence or mistakes (Human error)</td>
<td>0.82</td>
</tr>
<tr>
<td>System glitch</td>
<td>0.75</td>
</tr>
<tr>
<td>don’t know</td>
<td>1 (default)</td>
</tr>
</tbody>
</table>
Quantitative Cyber-Security

Colorado State University
Yashwant K Malaiya
CS559
Gordon-Loeb Models

Gorden Loeb models


• Model for the impact of a security investment on the probability of a breach.
  – $S(z,v)$
  – $S$: probability of a breach after an investment $z$
  – $v$: probability of a breach before investment

• Derived using concepts from economics, without using any data.

• Further work needed.
Security breach probability function. $S(z, v)$

- where $z > 0$ denote the monetary (e.g., dollar) investment in security to protect the given information set.
- $v$= “vulnerability” (probability of a security breach before investment)

Assumptions concerning $S(z, v)$:

**A1.** $S(z, 0) = 0$ for all $z$. If the information is completely invulnerable, then it will remain perfectly protected for with a zero investment.

**A2.** For all $v$, $S(0,v)=v$. That is, if there is no investment in information security, the probability of a security breach, conditioned on the realization of a threat, is the inherent vulnerability, $v$.

**A3.** For all $v \in (0, 1)$, and all $z$, $Sz(z, v) < 0$ and $Szz(z, v)>0$, where $Sz$ denotes the partial derivative with respect to $z$ and $Szz$ denotes the partial derivative of $Sz$ with respect to $z$.

That is, as the investment in security increases, the information is made more secure, but at a decreasing rate. Furthermore, we assume that for all $v \in (0,1)$, $\lim S(z,v) \rightarrow 0$, as $z \rightarrow \infty$, so by investing sufficiently in security, the probability of a security breach, $t$ times $S(z, v)$, can be made to be arbitrarily close to zero.
Impact of investment \( z \):

The **expected benefits of an investment in information security**, \( EBIS \), are equal to the reduction in the firm’s expected loss attributable to the extra security.

\[
EBIS(z) = [v - S(z, v)] L
\]

The **expected net benefits from an investment in information security**, \( ENBIS \), equal \( EBIS \) less the cost of the investment, or:

\[
ENBIS(z) = [v - S(z, v)] L - z
\]

- \( v \) – Probability of security breach
- \( L \) – Potential Loss.
- \( vL \) – Expected Loss
- \( z \) – Level of Investment
- \( S[z, v] \) – Revised probability of breach
Benefits & Costs of an Investment in Cyber/Information Security

\[ \text{Expected Benefits of Investment} = (v - S[z, v])L \]

\[ z^*(v) < \frac{1}{e} vL \]

\( v \) – Vulnerability (Probability of security breach)
\( L \) – Potential Loss
\( vL \) – Expected Loss
\( z \) – Level of Investment
\( z^* \) – Optimal Investment Level
\( S[z, v] \) – Revised v after z (Revised probability of breach)
They proposed two broad classes of security breach probability functions that satisfy A1-A3.

- The first class of security breach probability functions, denoted by $SI(z, v)$, is given by:

$$SI(z, v) = \frac{v}{(\alpha z + 1)^\beta}$$

where the parameters $\alpha > 0$, $\beta \geq 1$ are measures of the productivity of information security (i.e., for a given $(v, z)$, the probability of a security breach is decreasing in both $\alpha$ and $\beta$).

Solving for optimal $z^*$

$$z^{I*}(v) = \frac{(v \beta \alpha L)^{1/(\beta + 1)} - 1}{\alpha}.$$
The second class of security breach probability functions is given by:

\[ S_{II}(z, v) = v^{az+1} \]

Optimal value can be found as

\[ z_{II^*}(v) = \frac{\ln(1/-\alpha vL(\ln v))}{\alpha \ln v} \]

For both functions they have shown that

\[ z^*(v) < (1/e)vL. \]

Note that \( 1/e = 0.3679 \)

\( v \) – Probability of security breach
\( L \) – Potential Loss
\( vL \) – Expected Loss
\( z \) – Level of Investment
\( S[z, v] \) – Revised probability of breach
Propositions

• Proposition 1. For all security breach probability functions for which A1–A3 hold, there exists a loss, L, and a range of v in which increases in vulnerability result in an increase in the optimal investment in information security.

• Proposition 2. Suppose a security breach probability function meets conditions A1–A3, then it is not necessarily the case that the optimal level of investment in information security, z*(v), is weakly increasing in vulnerability, v.

• Proposition 3. Suppose the security breach probability function belongs to class I (i.e., it can be expressed as SI(z,v)=v/(αz+1)β for some α>0, β≥1) or to class II (i.e., it can be expressed as SII(z,v) = vαz+1 for some α > 0), then z*(v) < (1/e) vL. (See their Appendix for proof. )

  – The optimal investment in information security is always less than or equal to 36.79% of the loss that would be expected in 20 absence of any investment in security
How Can Organizations Use the Gordon-Loeb Model?

1. Estimate the potential loss (L) from a cybersecurity breach for each set of information
   – information segmentation is important.
2. Estimate the probability that an information set will be breached, by examining its vulnerability (v) to attack.
3. Create a grid with all the possible combinations of the first two steps, from low value, low vulnerability, to high value, high vulnerability.
4. Focus spending where it should reap the largest net benefits based on productivity of investments.
Recent Developments

• Widely citable ed in economic/financial fields.

• **Main impact:** 2017 U.S. Better Business Bureau (BBB) report recommends the Gordon-Loeb Model as "...a useful guide for organizations trying to find the right level of cybersecurity investment."


  
  — Values used: \( v = 0.5-0.9, L = 1 \text{ million}, \alpha = 4 \times 10^{-5}, \beta = 1 \)
  
  • Optimal about 0.2 \( v \)

• Table based investment distribution: based on risk values of each component.
Quantitative Security

Colorado State University
Yashwant K Malaiya
CS 559
Costs of security breaches
Cost Models

- **Ponemon Institute**
  - Founded in 2002 by Larry Ponemon and Susan Jayson
  - Conducts independent research on data protection
  - Collaborates with several large organizations and publishes annual reports

- **NetDiligence**
  - Privately-held cyber risk assessment and data breach services company.
  - Since 2001, NetDiligence has conducted thousands of enterprise-level cyber risk assessments for a broad variety of organizations
  - NetDiligence services are used by leading cyber liability insurers in the U.S. and U.K.

- **Ponemon assisted models, sponsored by**
  - Symantec (2010),
  - Megapath (2013), and
  - IBM (2014)

- **NetDiligence Model**
  - Hub International calculator (2012) and
  - Contributed to the Verizon report
Total Cost of a Breach =
- Incident investigation cost
- + Customer Notification/crisis management cost
- + Regulatory and industry sanctions cost
- + Class action lawsuit cost

\[
\text{Cost per Record} = \frac{\text{Total cost of breach}}{\text{number of affected records}}
\]
Cost Models: Investigations

• **The Ponemon Institute and NetDiligence data/models**
  – They used proprietary data available to them.
  – They derived computational models based on their data (“calculators”).
  – Large number of factors, considerable variation in factors considered.

• **Objective of study by Algarni and Malaiya**
  – Identify the major factors that are significant
  – Build models for the factors identified.

• **Approach**
  – regenerate data using the computational engines by providing a large number of input combinations.
  – Identified and removed the factors that emerged as non-significant.
  – Developed systematic computational models.
Cost Models: Investigations

• **The Ponemon Institute and NetDiligence data/models**
  – They used proprietary data available to them.
  – They derived computational models based on their data ("calculators").
  – Large number of factors, considerable variation in factors considered.

• **Objective of study by Algarni and Malaiya**
  – Identify the major factors that are significant
  – Build models for the factors identified.

• **Approach**
  – regenerate data using the computational engines by providing a large number of input combinations.
  – Identified and removed the factors that emerged as non-significant.
  – Developed systematic computational models.

A consolidated approach for estimation of data security breach costs, AM Algarni, YK Malaiya 2016 2nd International Conference on Information Management (ICIM), 26-39