Quantitative Cyber-Security

Colorado State University

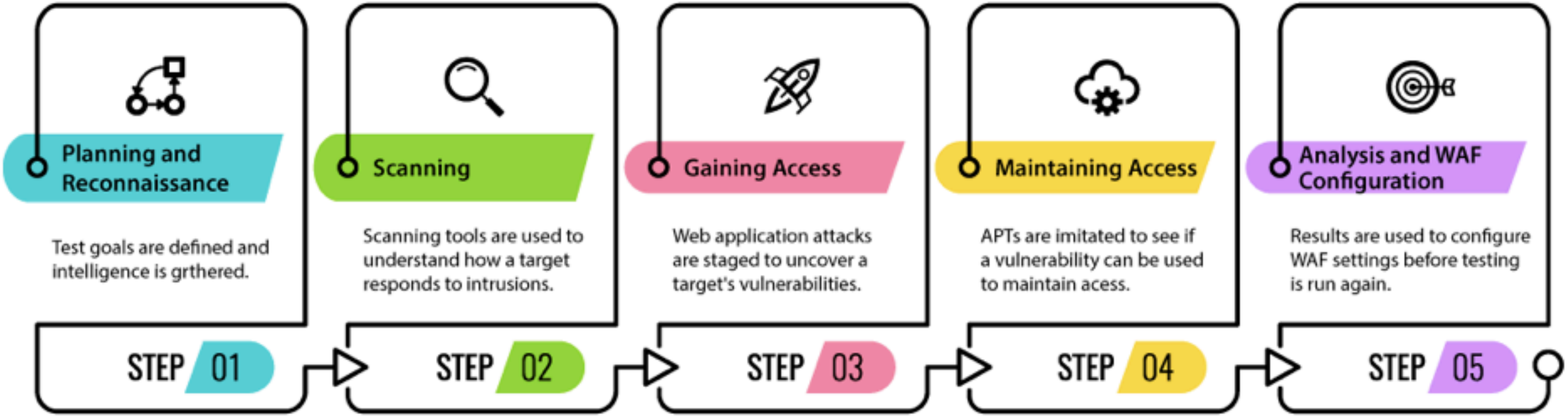
Yashwant K Malaiya

CS559

L21

CSU Cybersecurity Center
Computer Science Dept

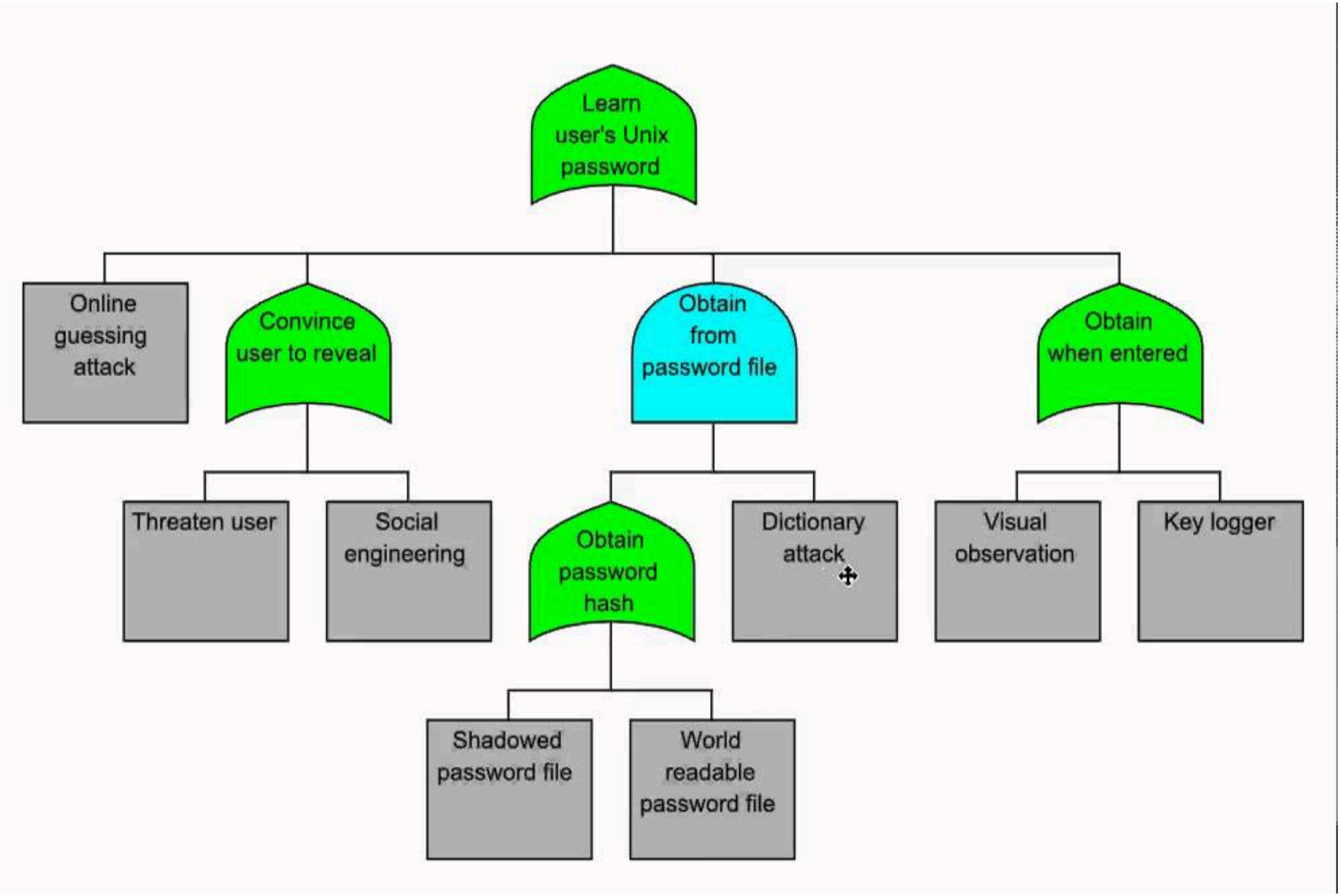
Pen Testing Stages


Planning and Reconnaissance
Test goals are defined and intelligence is grthered.
STEP 01
Scanning
Scanning tools are used to understand how a target responds to intrusions.
STEP 02
Gaining Access
Web application attacks are staged to uncover a target's vulnerabilities.
STEP 03
Maintaining Access
APTs are imitated to see if a vulnerability can be used to maintain acess.
STEP 04
Analysis and WAF Configuration
Results are used to configure WAF settings before testing is run again.
STEP 05

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

Sources: 1, 2

Attacks and Attack trees



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

Risk as a composite measure

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

  $\text{Likelihood}_i = P\{\text{A security hole}_l \text{ is exploited}\}.$

  $= P\{\text{hole}_i \text{ present}\}.$

  $P\{\text{exploitation} | \text{hole}_i \text{ present}\}$

- $P\{\text{hole}_i \text{ present}\}$: an internal attribute of the system.
- $P\{\text{exploitation} | \text{hole}_i \text{ present}\}$: 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)

  ALE = SLE x ARO

  - Where ARO is Annualized rate of occurrence.
- A countermeasure reduces the ALE by reducing one of its factors.

COUNTERMEASURE_VALUE

= (ALE_PREVIOUS – ALE_NOW) – COUNTERMEASURE_COST

ALE_PREVIOUS: ALE before implementing the countermeasure.

ALE_NOW: ALE after implementing the countermeasure

COUTERMEASURE_COST: *annualized* cost of countermeasure

Colorado State University

Estimating 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**
- **Factors based on available data**
  - **Organization's Country $F_{country}$**
  - **Organization's Industry Classification $F_{industry}$**
  - **Sensitive Data Encryption $F_{encryption}$**
  - **Organization's Privacy $F_{privacy}$**
  - **Business Continuity Management Team F_{BCM}**
  - **Data Breach Causes F_{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

Breach Probability Model

A proposed model for the probability of a breach for the next

P {breach} $= F_{country} * FBCM * Findustry * Fbreach_{cause} * Fencryption * Fprivacy * \alpha\, exp\,(-\beta x)$

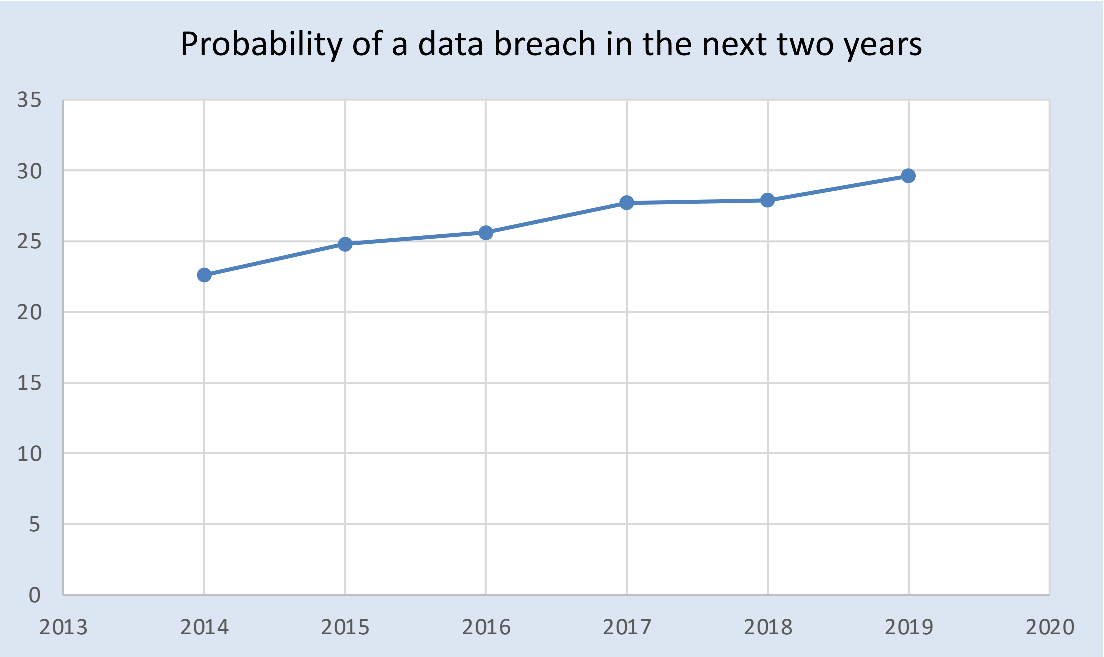
Where α = 0.4405, β = 4E-05, x the breach size 2015

Justification in the following slides.

Data Breach Probability

Cost of a Data Breach Report 2019, IBM Security, study by Ponemon Institute.

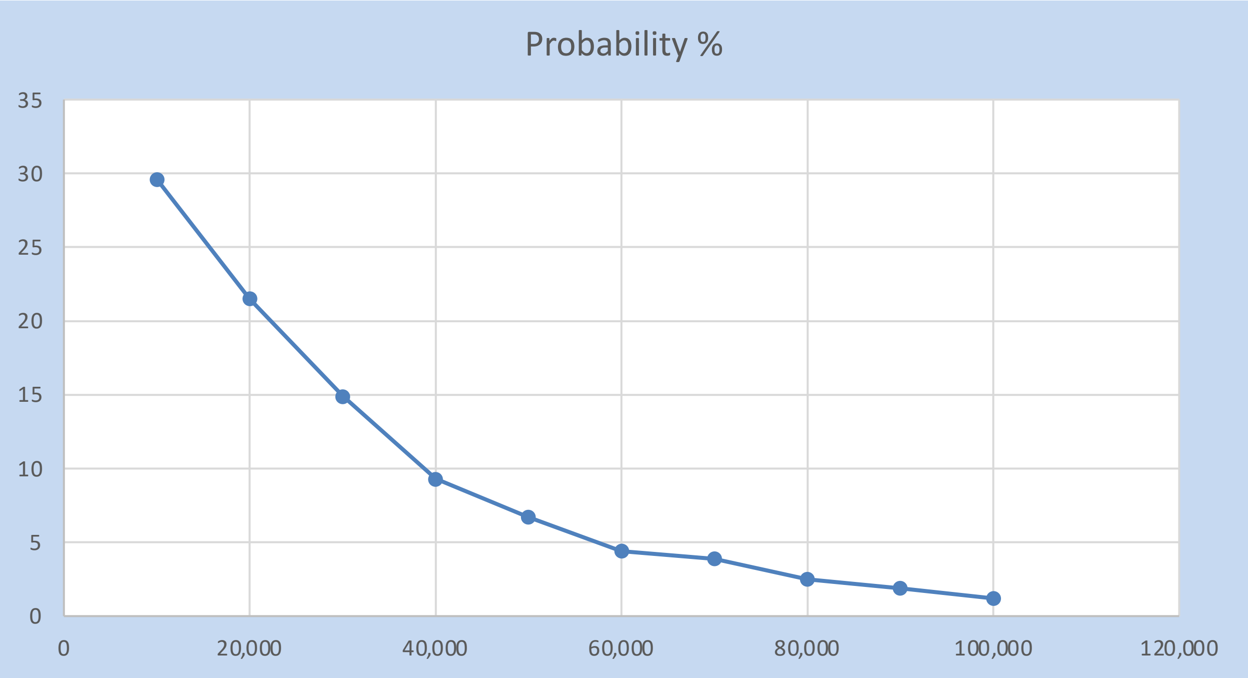
- 507 participating companies, with a minimum of 10,000 records
- United States, India, the United Kingdom, Germany, Brazil, Japan, France, the Middle East, Canada, Italy, South Korea, Australia, Turkey, ASEAN, South Africa, Scandinavia



Probability of a data breach by number of records lost

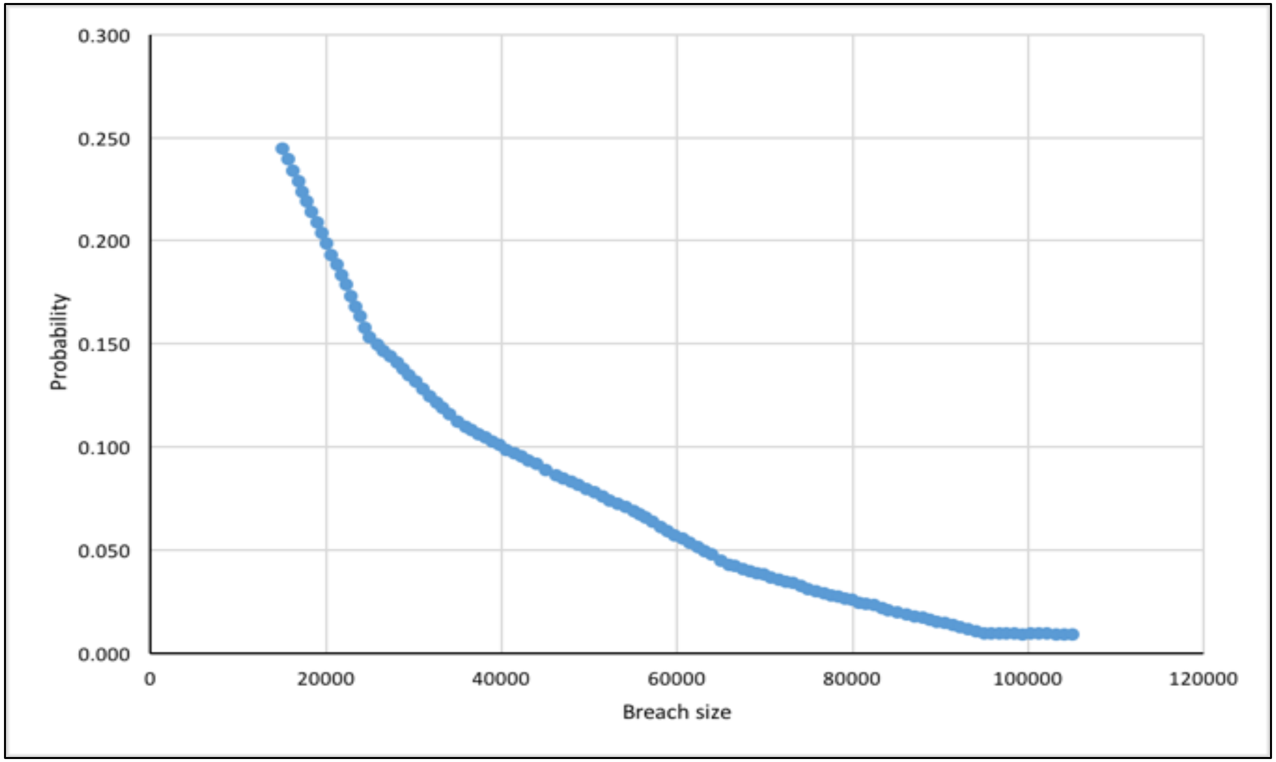
Over the next two years, involving minimum of 10,000 and maximum of 100,000 records.

Cost of a Data Breach Report 2019, IBM Security, study conducted by Ponemon Institute.



Colorado State University

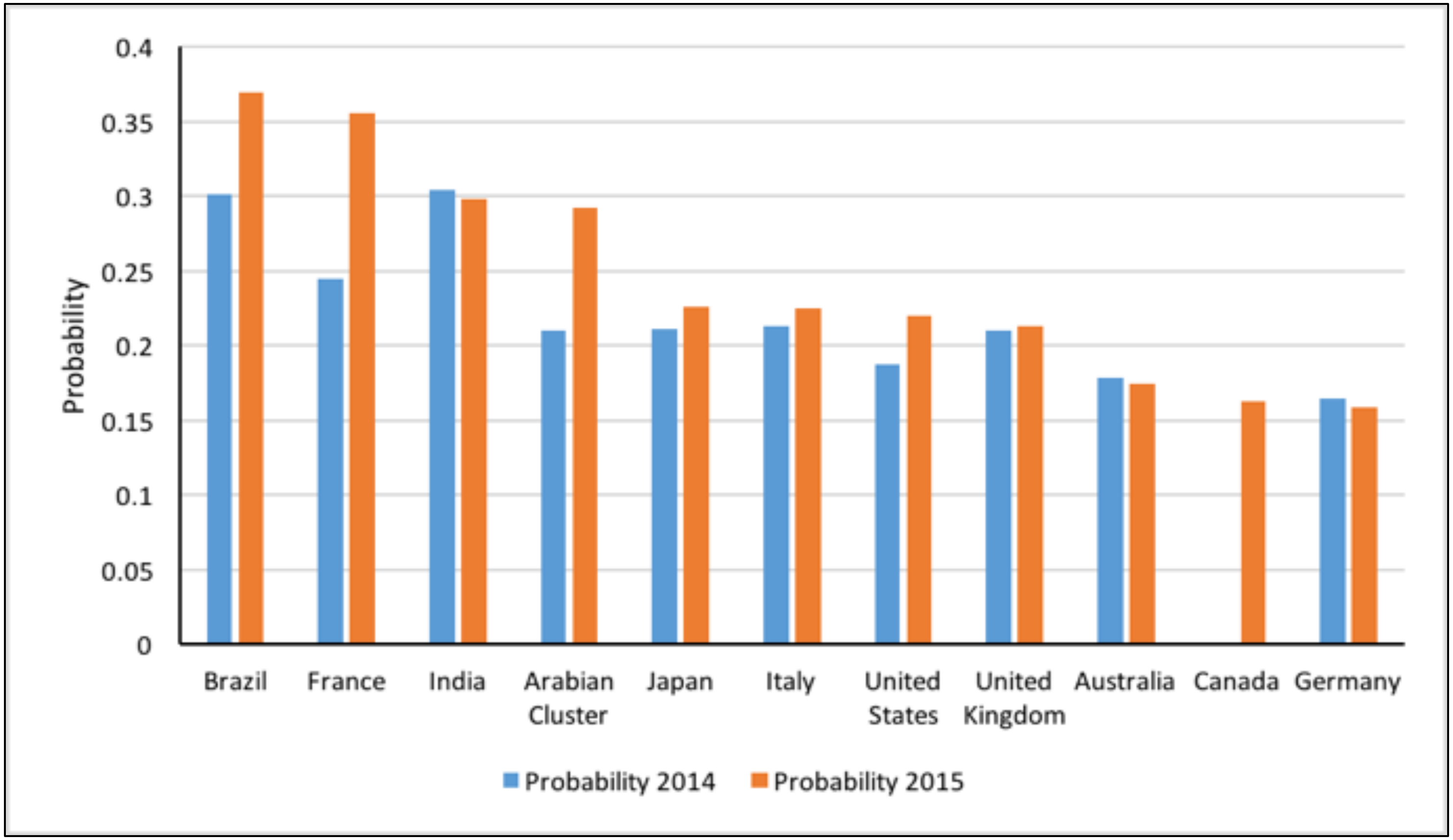
Breach probability -Breach size



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

Colorado State University

Data breach probability by country

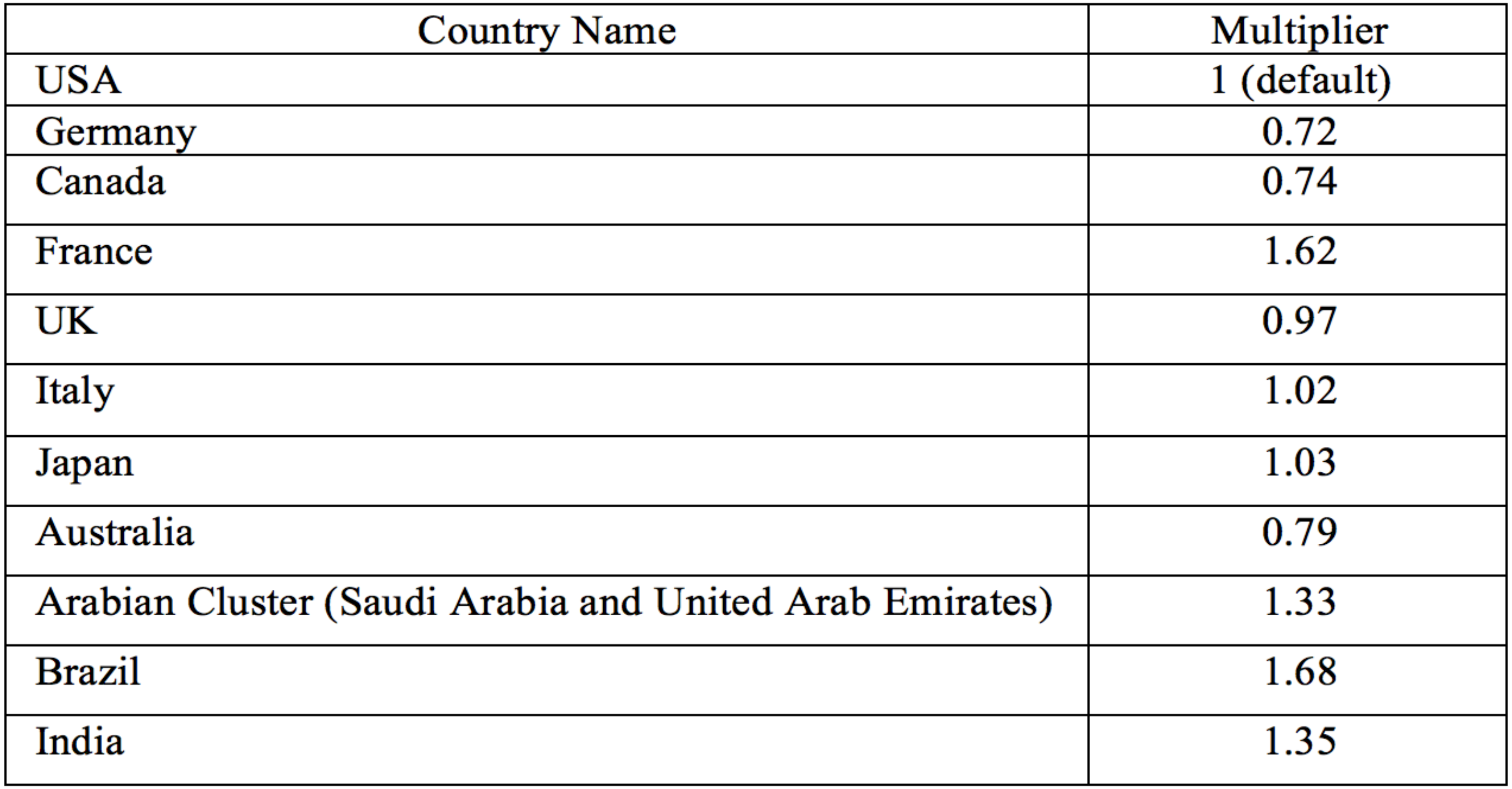


Data breach probability by country (Ponemon data 2015)

A minimum of 10,000 compromised records

Data breach probability by country

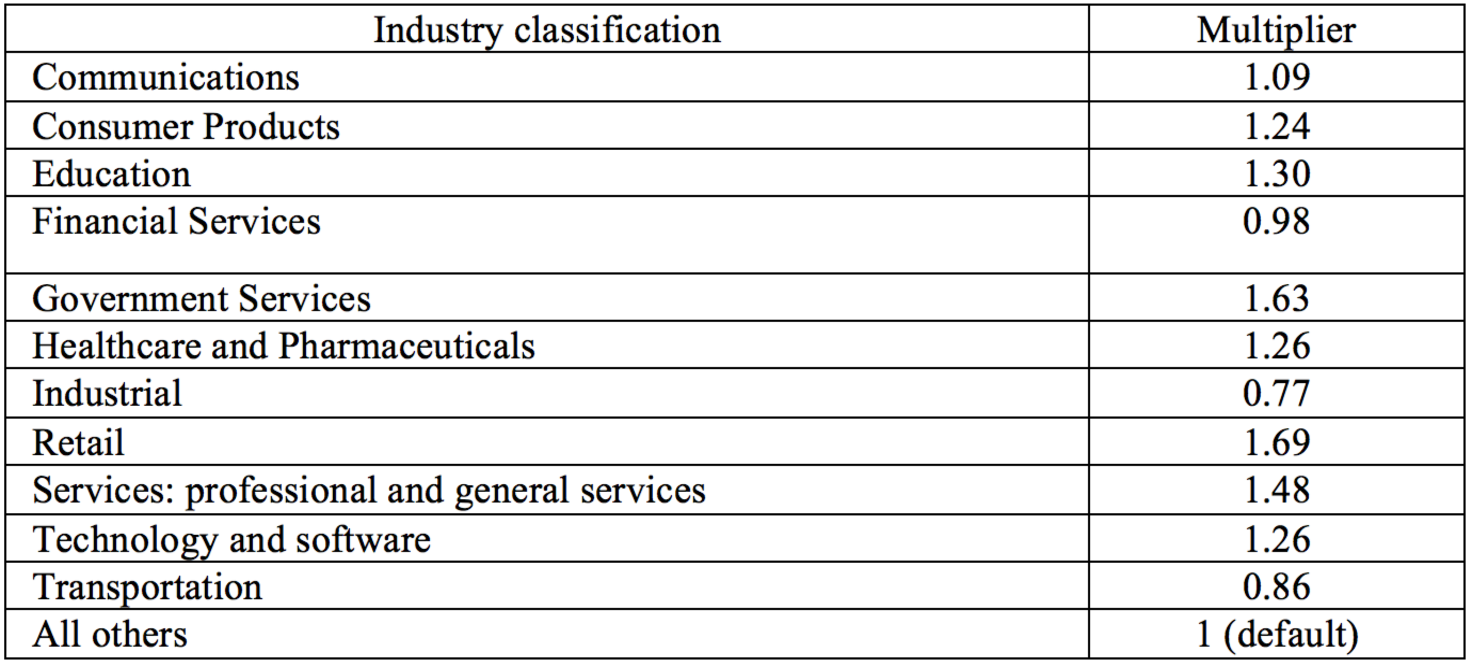
Data breach probability by country Fcountry (Ponemon data 2015)



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

Organization’s Industry Classification Findustry

Model proposed:

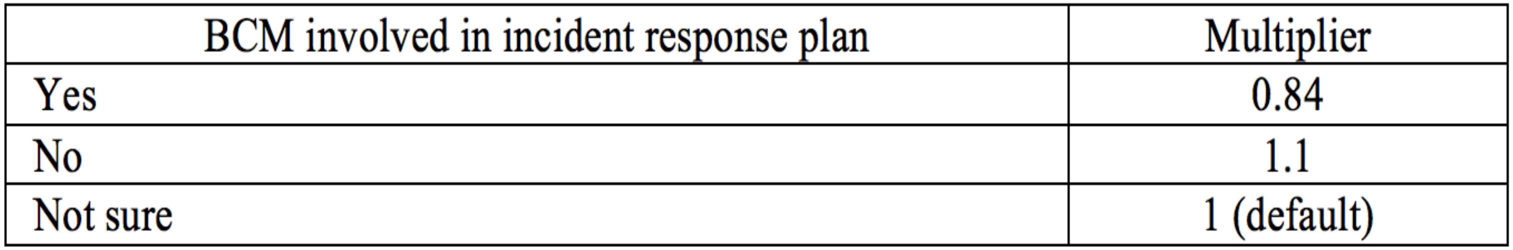


Industry classification	Multiplier
Communications	1.09
Consumer Products	1.24
Education	1.30
Financial Services	0.98
Government Services	1.63
Healthcare and Pharmaceuticals	1.26
Industrial	0.77
Retail	1.69
Services: professional and general services	1.48
Technology and software	1.26
Transportation	0.86
All others	1 (default)

Colorado State University

Business Continuity Management Team FBCM

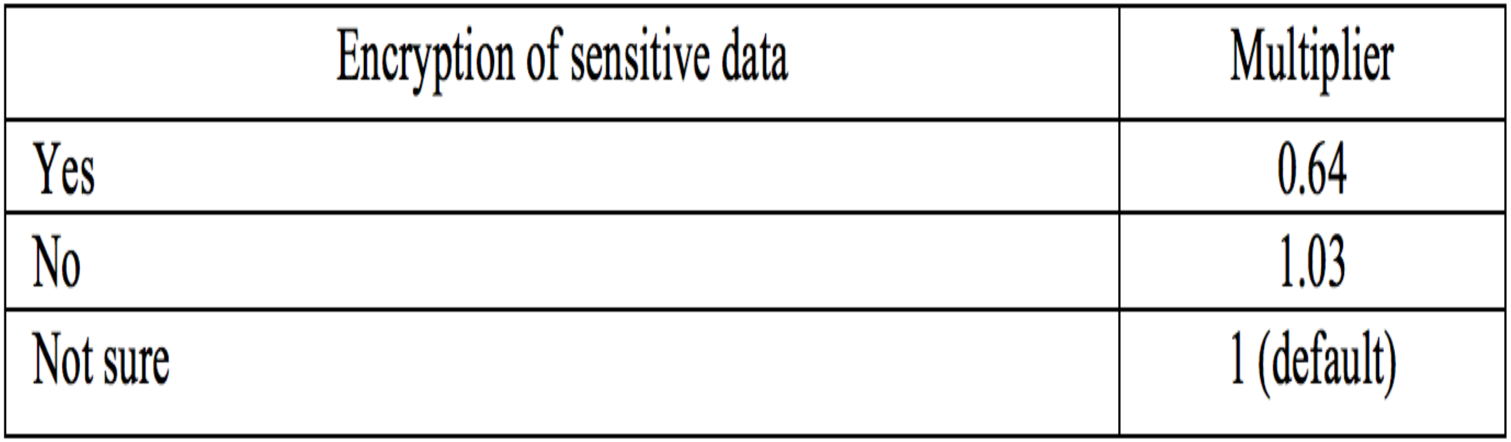
Model proposed:



BCM involved in incident response plan	Multiplier
Yes	0.84
No	1.1
Not sure	1 (default)

Sensitive Data Encryption Fencryption

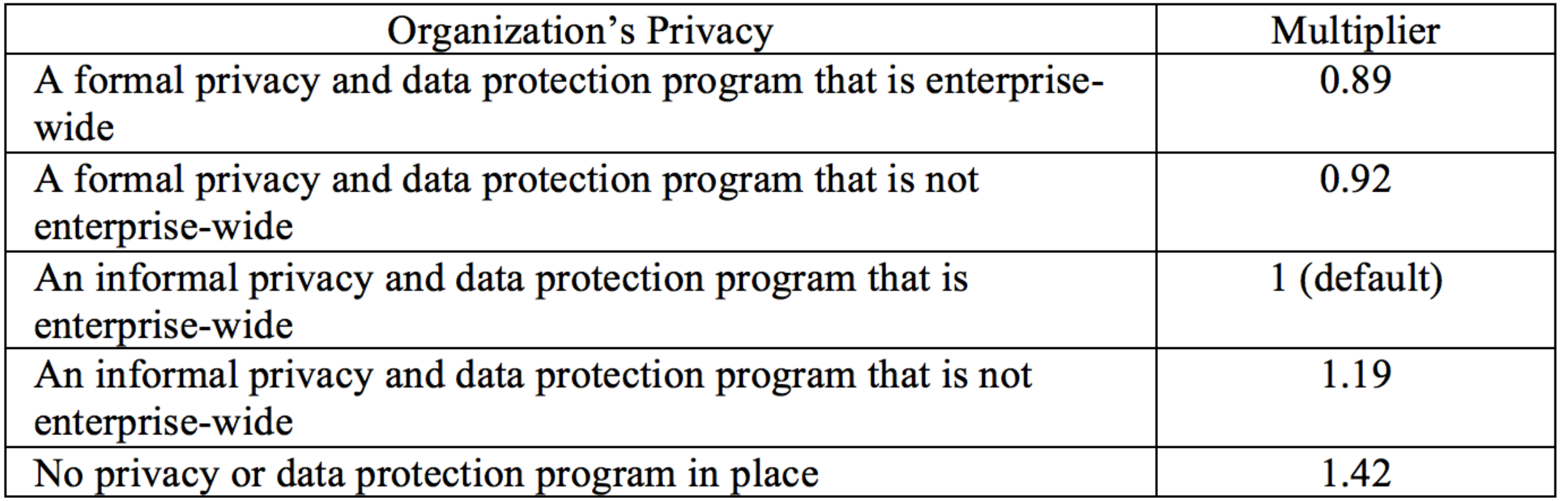
Model proposed:



Encryption of sensitive data	Multiplier
Yes	0.64
No	1.03
Not sure	1 (default)

Organization’s Privacy Fprivacy

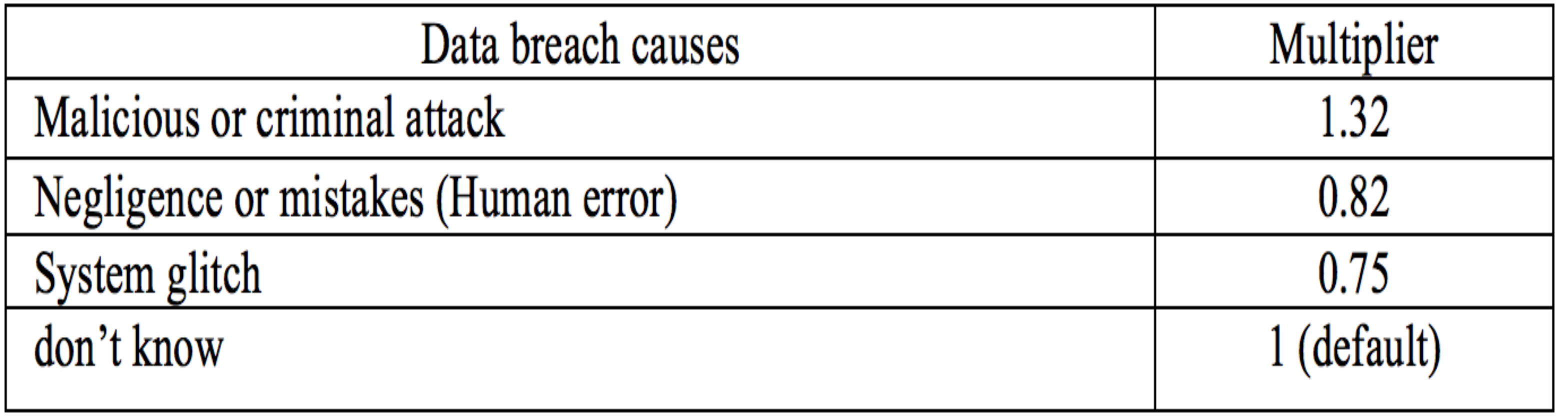
Model proposed:



Organization’s Privacy	Multiplier
A formal privacy and data protection program that is enterprise-wide	0.89
A formal privacy and data protection program that is not enterprise-wide	0.92
An informal privacy and data protection program that is enterprise-wide	1 (default)
An informal privacy and data protection program that is not enterprise-wide	1.19
No privacy or data protection program in place	1.42

Data Breach Causes Fbreach_cause

Model proposed:



Data breach causes	Multiplier
Malicious or criminal attack	1.32
Negligence or mistakes (Human error)	0.82
System glitch	0.75
don’t know	1 (default)

Quantitative Cyber-Security

Colorado State University

Yashwant K Malaiya

CS559

Gordon-Loeb Models

CSU Cybersecurity Center
Computer Science Dept

L. A. Gordon and M. P. Loeb, “The economics of information security investment,” *ACM Trans. Inf. Syst. Secur.*, vol. 5, no. 4, pp. 438–457, 2002.

Gorden Loeb models

- L. A. Gordon and M. P. Loeb, “The economics of information security investment,” *ACM Trans. Inf. Syst. Secur.*, vol. 5, no. 4, pp. 438–457, 2002.
- 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

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 $v \in (0,1)$, $\mathit{lim}\ S(z,v) \to 0$, as $z \to \infty$, 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.

Expected benefits of an investment in information security

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.

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

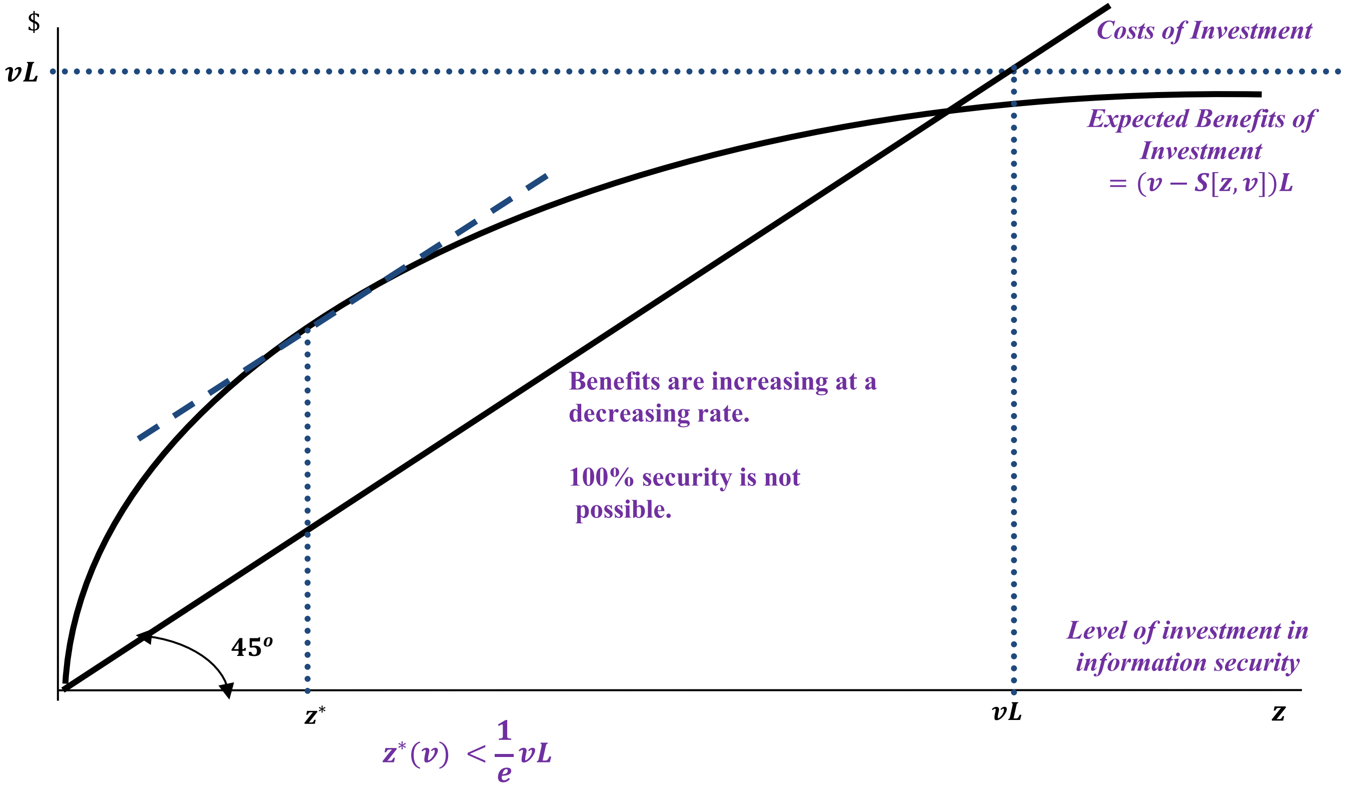
The **expected net benefits from an investment in information security**, *ENBIS* equal *EBIS* less the cost of the investment, or:

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

v – *Probability of security breach*
L – *Potential Loss.* vL – *Expected Loss*
z – *Level of Investment*
$S[z, v]$ – *Revised probability of breach*

Colorado State University

Benefits & Costs of an Investment in Cyber/Information Security



Benefits are increasing at a decreasing rate.

100% security is not possible.

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

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)*

Security breach probability functions

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:

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

where the parameters α > 0, β ≥ 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 α and β).

Solving for optimal z∗

$$z^{I*}(v) = \frac{(v\beta\alpha L)^{1/(\beta+1)} - 1}{\alpha}.$$

v – Probability of security breach
L – Potential Loss. *vL – Expected Loss*
z – Level of Investment
S[*z*, *v*] *– Revised probability of breach*

Security breach probability functions

- The second class of security breach probability functions is given by:

$$S^{II}(z,v) = v^{\alpha z+1}$$

- Optimal value can be found as

$$z^{II*}(v) = \frac{\ln(1/-\alpha v L(\ln v))}{\alpha \ln v}$$

- For both functions they have shown that

$$z^{*}(v) < (1/e)vL\,.$$

v – *Probability of security breach*
L – *Potential Loss.* vL – *Expected Loss*
z – *Level of Investment*
$S[z,v]$ – *Revised probability of breach*

Note that 1/e = 0.3679

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)$^{\beta}$ for some α>0, β≥1) or to class II (i.e., it can be expressed as $S^{II}(z, v) = v^{\alpha z+1}$ 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

Colorado State University

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.

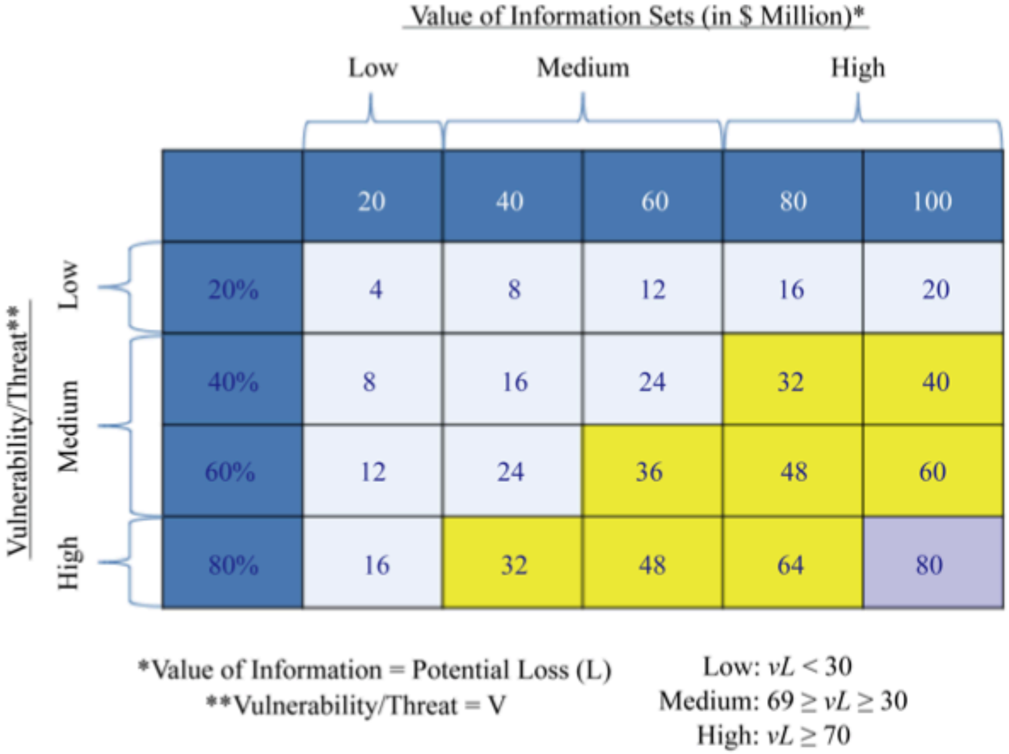
2. Estimate the probability that an information set will be breached, by examining its vulnerability (v) to attack.

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."
- Cybersecurity Investment Guidance: Extensions of the Gordon and Loeb Model, **S**. Farrow, J. Szanton, 2016
- Calibration of the Gordon-Loeb Models for the Probability of Security Breaches, M. Naldi, M. Flamini, 2017.
  - Values used: v = 0.5-0.9, L = 1 million, $\alpha = 4\text{x}10^{-5}$, $\beta = 1$
    - Optimal about 0.2 v
- Table based investment distribution: based on risk values of each component.



Value of Information Sets (in $ Million)*

Vulnerability/Threat**		Low	Medium		High	
		20	40	60	80	100
Low	20%	4	8	12	16	20
Medium	40%	8	16	24	32	40
	60%	12	24	36	48	60
High	80%	16	32	48	64	80

*Value of Information = Potential Loss (L)
**Vulnerability/Threat = V

Low: $vL < 30$
Medium: $69 \geq vL \geq 30$
High: $vL \geq 70$

Gordon, L.A., Loeb, M.P., Zhou, L.: Investing in cybersecurity: insights from the Gordon-Loeb model. J. Inf. Secur. **7**(02), 49 (2016)

Quantitative Security

Colorado State University

Yashwant K Malaiya

CS 559

Costs of security breaches

CSU Cybersecurity Center
Computer Science Dept

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**
  - Symantac (2010),
  - Megapath (2013), and
  - IBM (2014)
- **NetDiligence Model**
  - Hub International calculator (2012) and
  - contributed to the Verizon report

Cost Metrics

Total Cost of a Breach =

Incident investigation cost

+ Customer Notification/crisis management cost

+ Regulatory and industry sanctions cost

+ Class action lawsuit cost

$$\boldsymbol{Cost\ per\ Record} = \frac{Total\ cost\ of\ breach}{number\ of\ affected\ records}$$

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