Quantitative Cyber-Security

Colorado State University
Yashwant K Malaiya

CS559

L22

CSU Cybersecurity Center
Computer Science Dept

Peer Reviews

Each student needs to do two peer reviews by coming Sat Nov. 14. You will use the peer reviews to improve your presentation/final report.

The review process is somewhat similar to the review process for articles submitted to peer-reviewed conferences/journals. Do not include your name in the review. Use this format:

A: Comments: Include the following.

- What is the contribution and what is significant.
-Things you find positive.
-Things that can be improved including, technical, text, language, charts etc.
- Questions that you would like to see addressed in the presentation/final report.
-Additional references that the author should look at.

B.. Evaluate the following:

Novelty/Interest: [ ] Technical/ Research: [ ] Presentation: [ ] Overall: [ ]

Evaluate using E – Excellent G – Good B – Borderline U – Unacceptable.
Use no more than 25% Excellent in any of the four scores.

Presentations/Final Report

Slides should be ready by Wed 11/18/20, but ..

- Post 24 hours in advance of the presentation in the designated canvas forum.
- Schedule will be announced later
- Peer reviews will be needed.

Final report is due on Wed 12/9/20.

Topics

- Risk components
- Probability of a breach
- Gordon-Loeb Models
- Breach cost

Quantitative Cyber-Security

Colorado State University
Yashwant K Malaiya

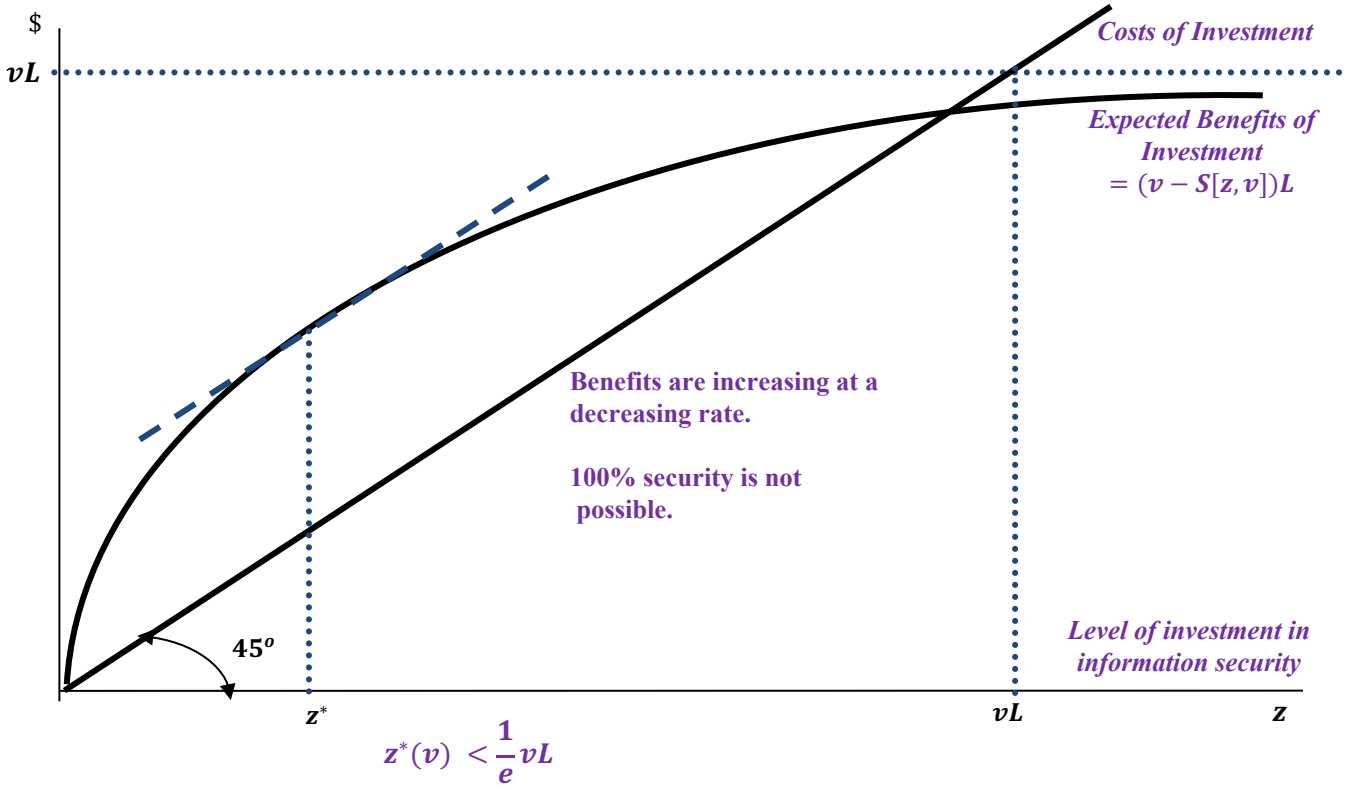
CS559

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.

Benefits & Costs of an Investment in Cyber/Information Security



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)*

Colorado State University

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 $\alpha > 0$, $\beta \geq 1$ 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 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

$$z^*(v) < (1/e) vL .$$

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*

Quantitative Security

Colorado State University
Yashwant K Malaiya

CS 559

Breach probability

CSU Cybersecurity Center
Computer Science Dept

Modeling the Breach Probability

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

- Breach size
- Other factors:
- 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

Modeling the Breach Probability

- Multiplicative model for Breach probability
  - Factors largely orthogonal
  - Default value is 1.
    - If no known, value is not affected
    - Default value corresponds to the most common or average case
- Factors multiply
  - A factor may a mathematical function:
    - Can be linearly dependent on a measurable quantity or may be non-linear
  - May be specified using a table
    - Examples of tabular approach: CVSS metrics

Breach Probability Model

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

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

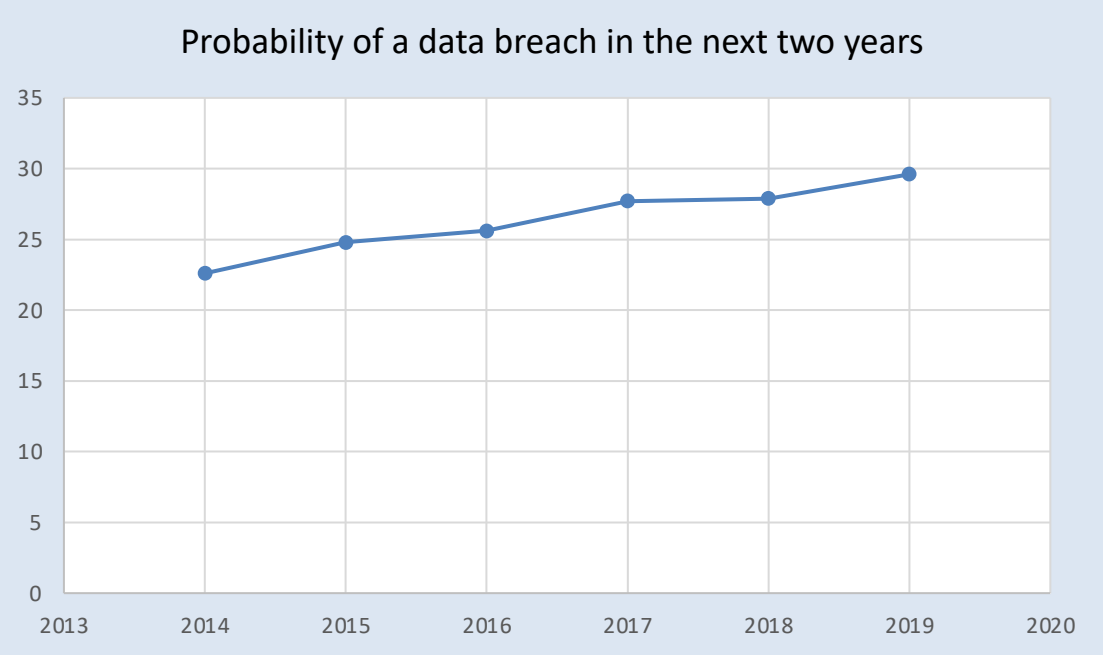
Where α = 0.4405, β = 4E-05, x the breach size 2015

- The values of the parameters may gradually change with time.
- 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

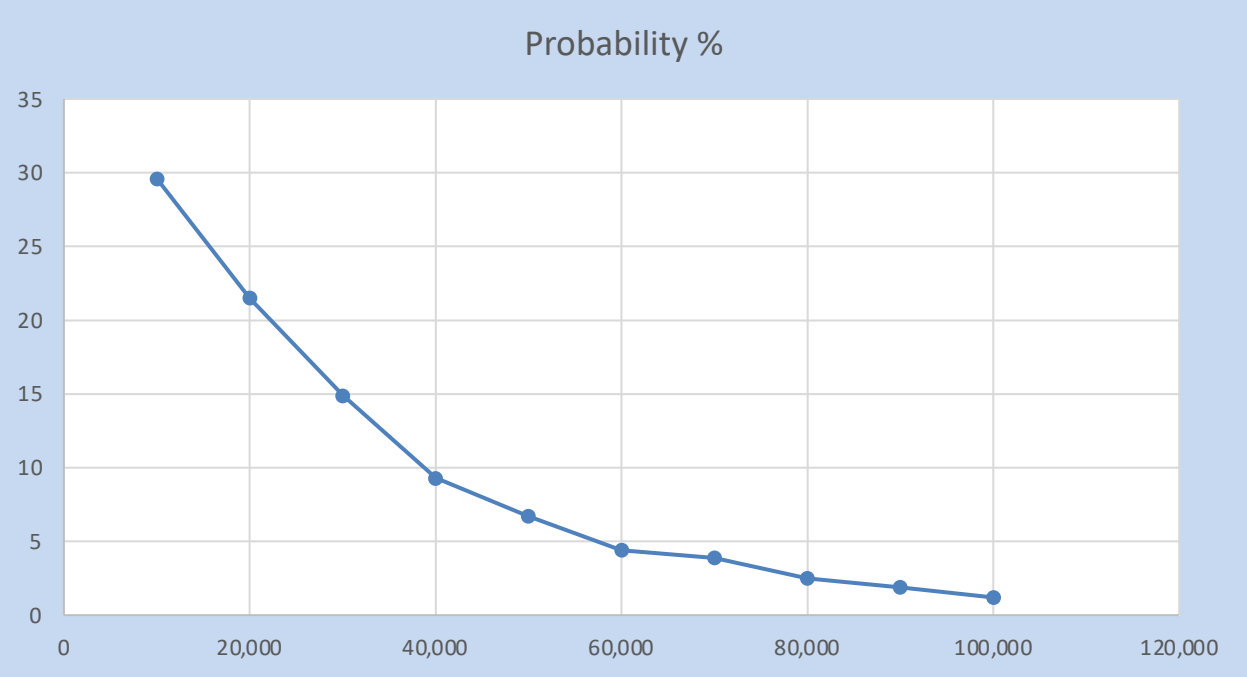


Colorado State University

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.

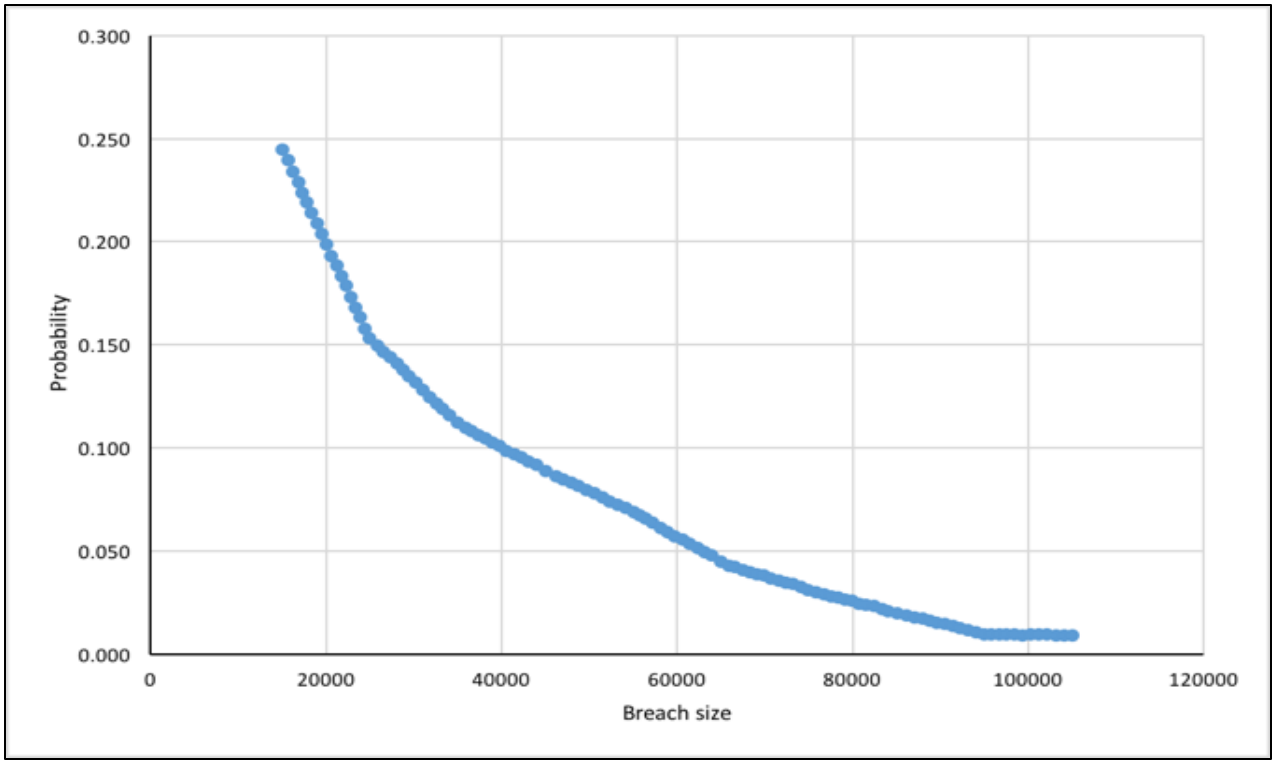
Cost of a Data Breach Report 2019, IBM Security, study conducted by Ponemon Institute.



Exponential form

Colorado State University

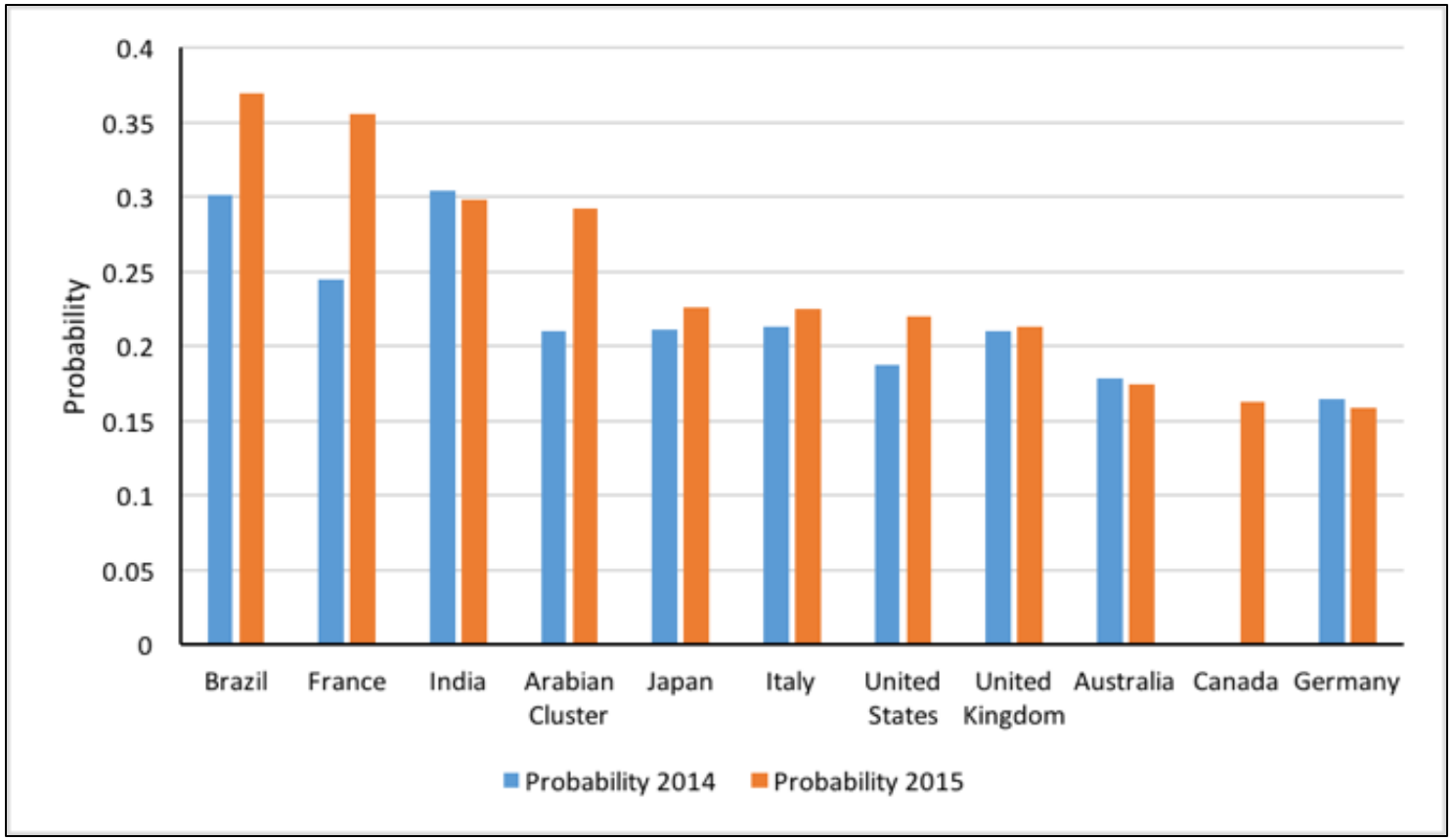
Breach probability -Breach size



Data breach probability based on the breach size (Ponemon data 2015)

Colorado State University

Data breach probability by country



Data breach probability by country (Ponemon data 2015)

Colorado State University

Data breach probability by country

Data breach probability by country Fcountry (Ponemon data 2015)

Default value: USA

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)

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 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 =

Direct costs + Indirect costs – Recovered costs

Direct costs: funds spent directly

= Incident investigation cost

+ Customer Notification/crisis management cost

+ Regulatory and industry sanctions cost*

+ Class action lawsuit cost*

Indirect costs: lost business opportunity

= loss of goodwill, customer churn#

Recovered costs = Insurance recovery + tax break

* Post data breach response

Measured by the stock-market?

Cost Metrics

Total Cost of a Breach

= fixed costs + variable costs – recovered costs

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

- Fixed cost: regardless of the size of breach
- Variable costs depend on the number of records.
  - May not be linear because of economy of scale

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.
  - Not yet fully published.
- **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 Int. Conf. on Information Management (ICIM), 26-39
Quantitative economics of security: software vulnerabilities and data breaches, Algarni, Abdullah M., PhD Dissertation, 2016

Significant Factors impacting Cost and Probability

Classification	Significant Factors	Source
Total number of affected records	Total number of affected records?	All
Type of data breaches	What is your organization's industry classification?	Symantec & IBM
	What types of information do your employees handle?	Symantec & IBM
Incident investigation cost	Data is in a centralized system/location?	Hub Int'l
	Actual fraud is expected already?	Hub Int'l
	Federal class action lawsuit filed?	Hub Int'l
	What do you think is the most likely cause of a data breach?	Symantec & IBM
	Is sensitive data encrypted on all laptops or removable storage?	Symantec & IBM
	How long does the business keep/retain sensitive information pertaining to employees, customers, and patients?	IDT911
	What best describes your organization's privacy and data protection program?	Symantec & IBM
Crisis management cost	Number of Years for credit monitoring?	Hub Int'l
	What is the global headcount of your organization?	Symantec & IBM
	Is your organization's business continuity management team involved in the data breach incident response process?	IBM
Regulatory and sanction cost	Is PCI compliance an issue?	Hub Int'l
Lawsuit cost	Actual fraud is expected already?	Hub Int'l
	Federal class action lawsuit filed?	Hub Int'l

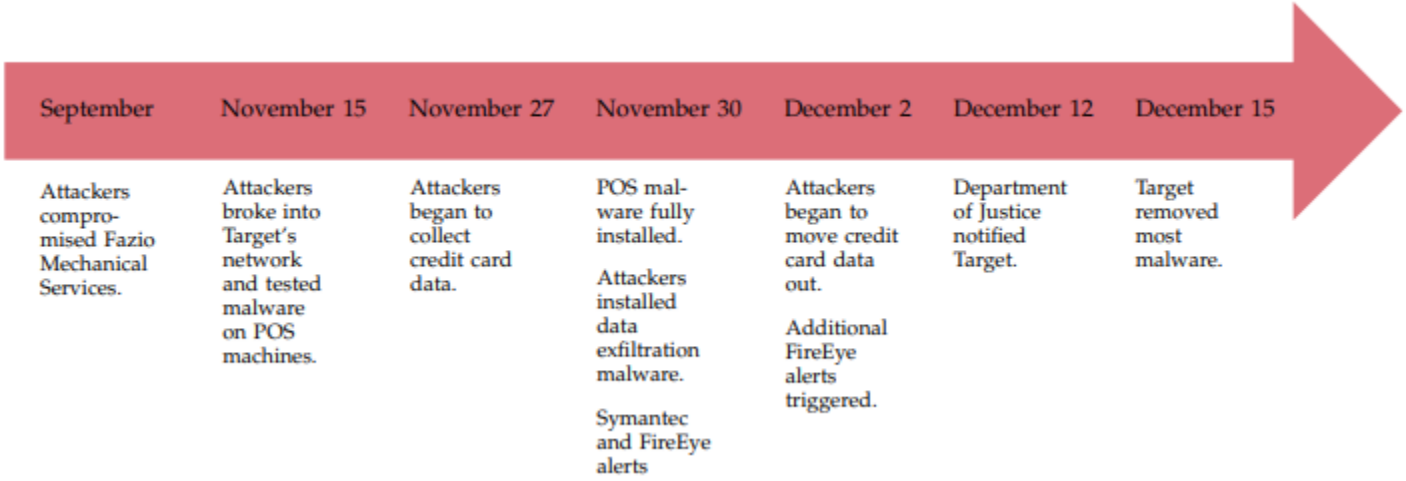
A consolidated approach for estimation of data security breach costs, AM Algarni, YK Malaiya 2016 2nd International Conference on Information Management (ICIM), 26-39

Examples

- Target Data Breach 2013
- Home Deport Data Breach, 2014

Target data breach (2013)

- Target Corporation's network
- Breach Dates: Between November 27 and December 18, 2013
  - Announced Dec 19, 2013 to media (Dec 18 KrebsOnSecurity, WSJ)
  - second largest credit and debit card breach after the TJX breach in 2007.
  - 40 million credit and debit card numbers and 70 million records of personal information were stolen.
  - It cost credit card unions over two hundred million dollars for just reissuing cards.
  - Wildly different cost estimates by experts, up to a billion.



Xiaokui Shu, Ke Tian, Andrew Ciambrone, and Danfeng Yao. Breaking the Target: An Analysis of Target Data Breach and Lessons Learned. CoRR, abs/1701.04940, 2017

Target data breach (2013)

- TGT Price chart (Yahoo Finance)

Note:

Colorado State University

TARGET DATA BREACH ACTUAL REPORTED COSTS

Years	Gross Expenses	Insurance receivable	Net Expenses (before tax deductions)	Net Expenses (after tax deductions)
2013	$61m	$44m	$17m	$11m
2014	$191m	$46m	$145 m	$94m
2015	N/A	N/A	$39	$28
Total	$252m	$90m	$201m	$133m
Raw cost per card= $6.30 (40 million cards affected)				

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

Home Depot Data Breach Actual reported Costs

- September 8th, 2014, Home Depot released a statement indicating that its payment card systems were breached.
- The data breach occurred from a sophisticated custom-built malware program installed on Home Depot's payment system network using a third-party vendor's login credentials.

Case Study: The Home Depot Data Breach, Brett Hawkins, 2015

Home Depot Data Breach Actual reported Costs

Years	Gross Expenses	Insurance receivable	Net Expenses (before tax deductions)	Net Expenses (after tax deductions)
Q3, 2014	$43m	$15m	$28m	N/A
Q4, 2014	$20m	$15m	$5m	N/A
Total	$63m	$30m	$33m	N/A
Raw cost per card= $1.13 (56 million cards affected)				

NA: not available

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

Cost per record

- Cost per record metric
- Partial costs
- Average costs?
- Available data
- Proposed model for Cost per record

Is there an average cost per record?

- Using averages make sense, at least for initial estimates
- The **law of large numbers**:
  - sample size grows, its mean gets closer to the average of the whole population.
- The **Flaw of Averages**:
  - $2 billion in property damage in North Dakota.
  - In 1997, the U.S. Weather Service forecast that North Dakota's rising Red River would crest at 49 feet.
  - Officials in Grand Forks made flood management plans based on this single figure.
  - The river crested above 50 feet, breaching the dikes, and unleashing a flood that forced 50,000 people from their homes.

The Flaw of Averages, Sam Savage, Harvard Business Review, Nov. 2002

Ponemon: 2015 Cost of Data Breach in US

Partial costs	Avg. cost per breach	Avg. cost per record
Detection & escalation (includes investigation and crisis management)	$610,000	$21.73
Notification (includes notification and determination of regulatory)	$560,000	$19.95
Ex-post response (includes regulatory and lawsuit)	$1,640,000	$58.43
Lost business (includes reputation loss)	$3,720,000	$132.53
Total costs	$6,530,000	$217
Average number of records= 28070		

Average Cost per record: Hub Int.

Partial costs	Avg. cost per record for CC	Avg. cost per record for PHI&SSN
Incident investigation	$1.15	$1.64
Crisis management	$3.52	$4.57
Sanctions	$0.81	$0.81
Lawsuit	$7.09	$1.56
Total costs	$12.57	$8.58

Credit cards, Personal Health Information, SSN

From Hub International web site

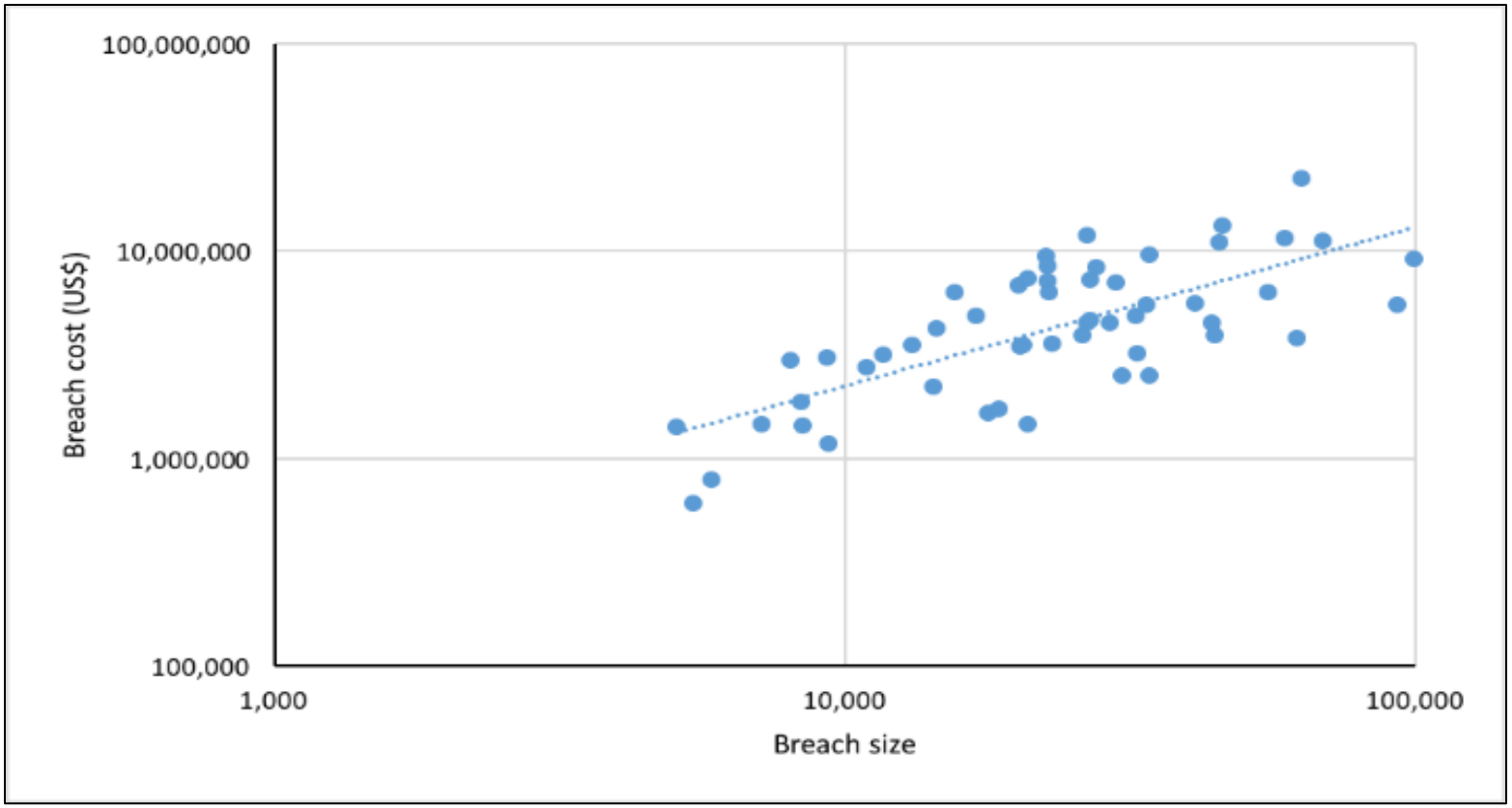
Average Cost per record

- What is the right number for *average cost per record*?
  - $217 Ponemon?
  - $8-$13 Hub International?
  - $0.58 Verizon?
- Controversy

Ken Spinner, Data breach cost estimates get it wrong: What you need to know.

"Why Ponemon Institute's Cost of Data Breach Methodology Is Sound and Endures". Ponemon Institue. 2015.

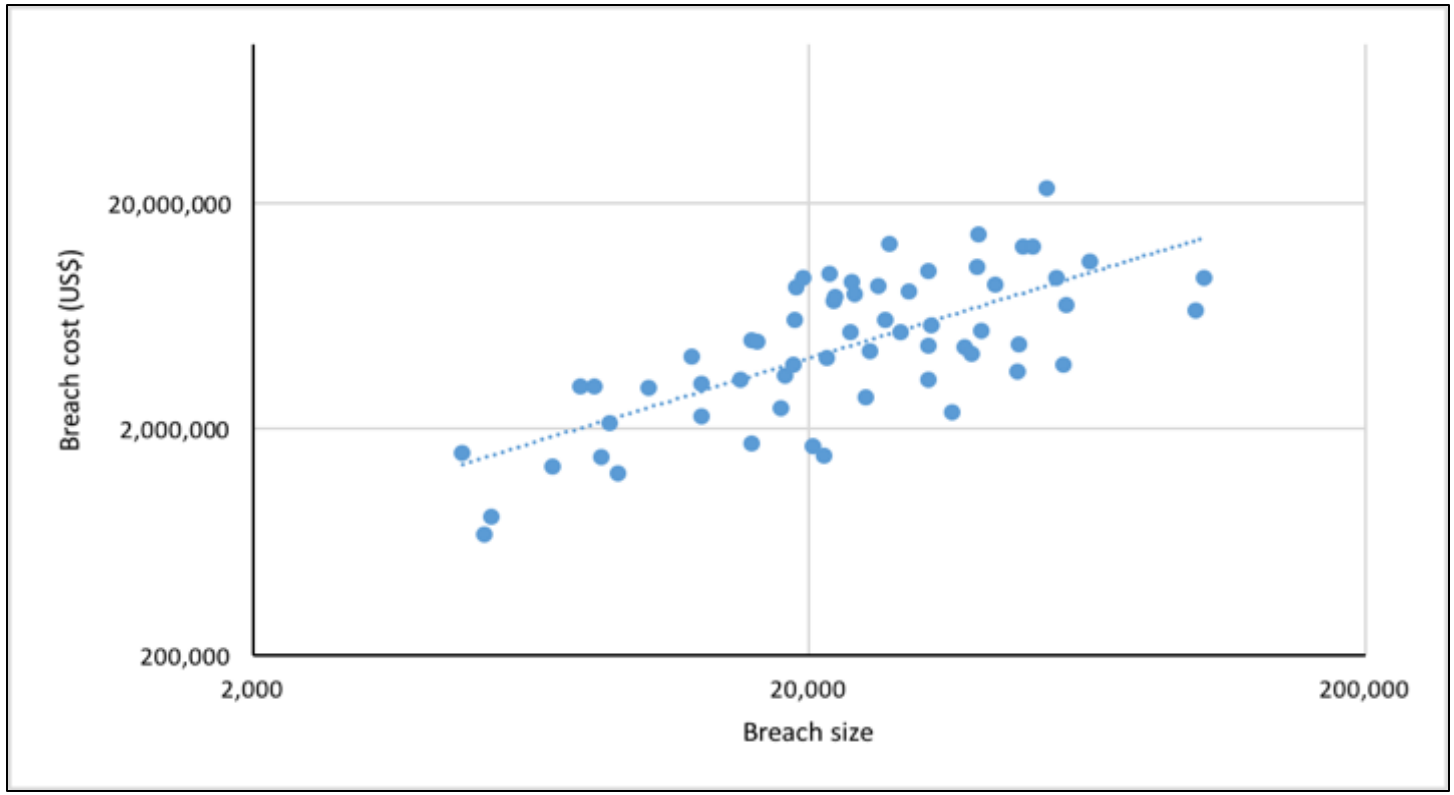
The breach cost vs. breach size



Ponemon 2013 data, the breach cost vs. breach size. Note log-log scate. (ranges from 5,000 to 100,000 records)

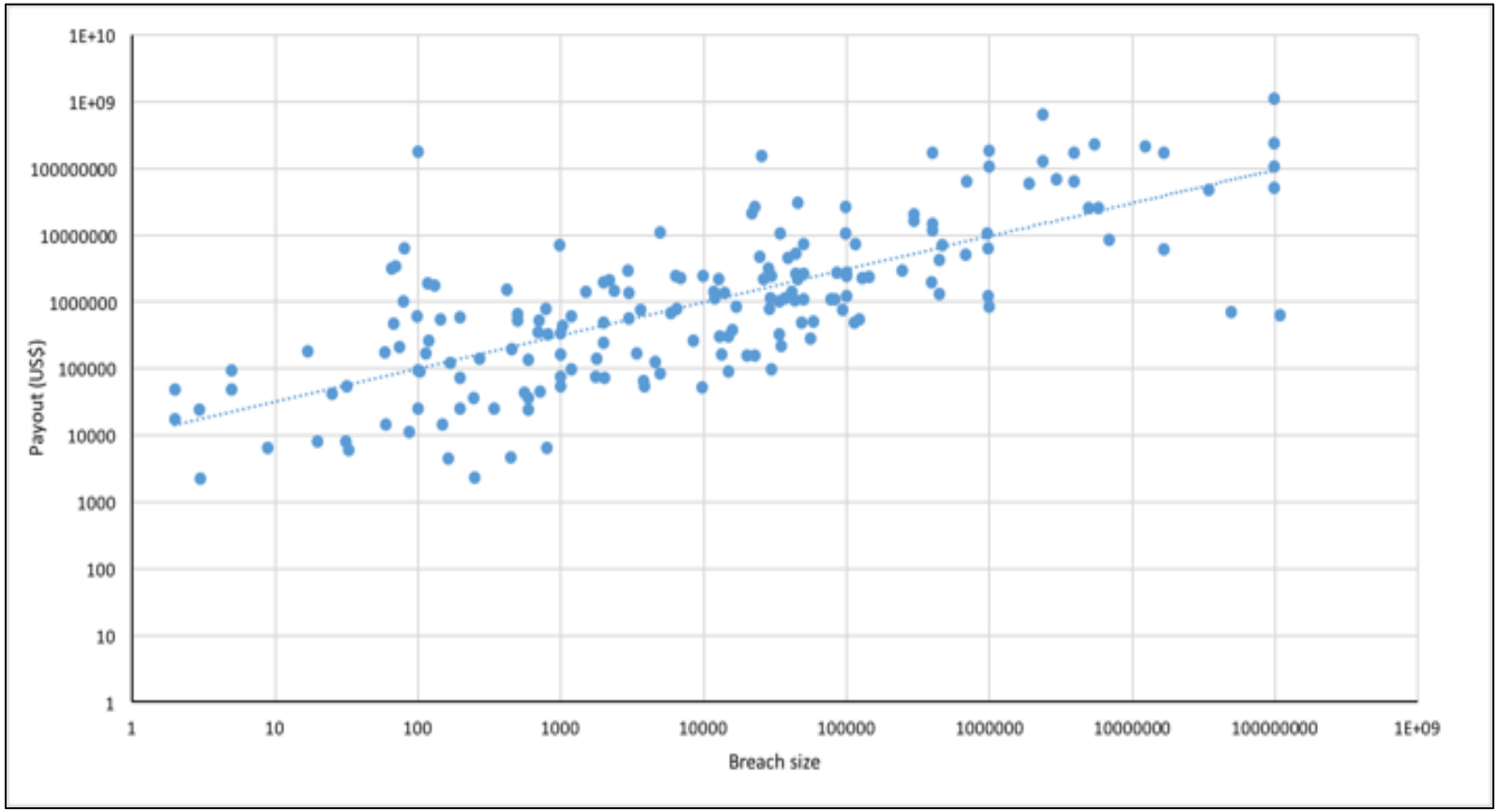
Colorado State University

The breach cost vs. breach size



Ponemon 2014 data, the breach cost vs. breach size
(ranges from 4,700 to 103,000 records)

The breach cost vs. breach size



Verizon 2015 data, the claim amount vs. breach size
(ranges from single digits to 108 million records)

Colorado State University

The breach cost vs. breach size

- Our proposed model

$$\boldsymbol{Total\ breach\ cost} = a\ *\ size\ \hat{}\ b$$

for breach sizes bigger than or equal to 1000 records

- Nonlinearity caused by ***economy of scale,*** *thus b should be < 1.*
- Thus

$$\boldsymbol{Cost\ per\ record} = a\ *\ (size)\ \hat{}\ (b-1)$$

Breach Cost/Payout Regression Models

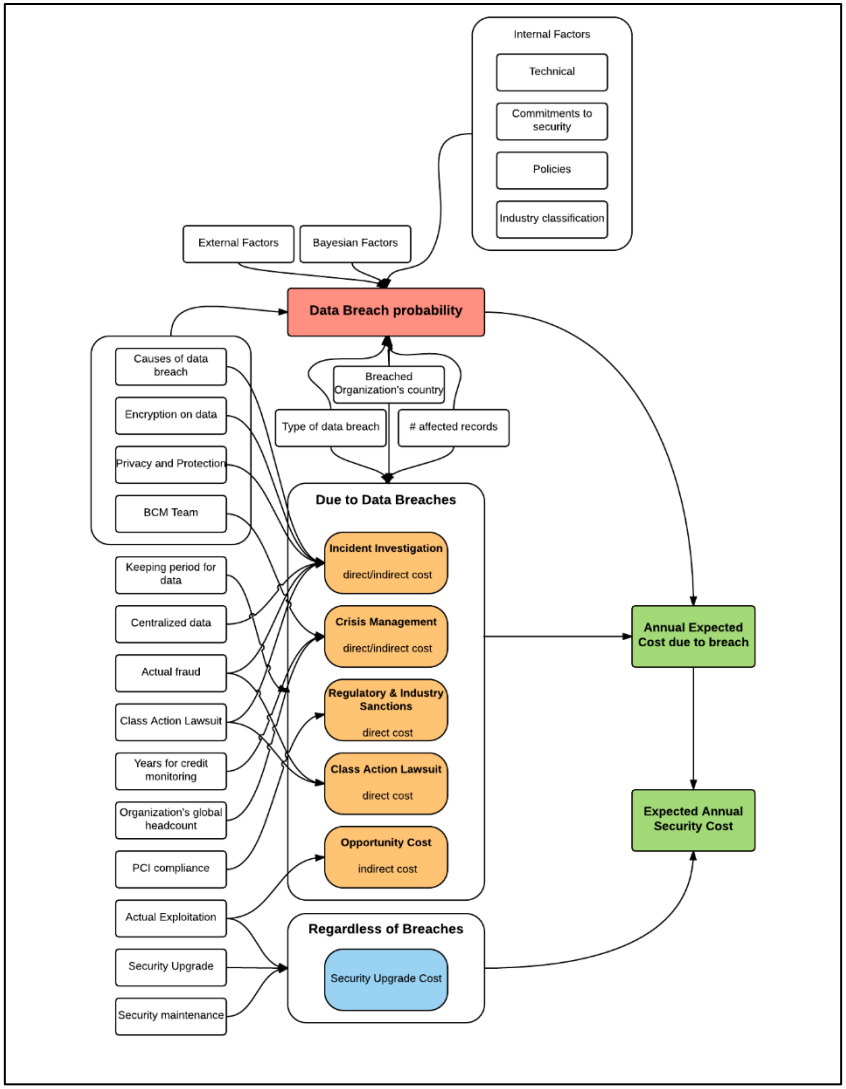
Data sets	Size of breaches	Data points	Regression	
			Breach Cost Model	R^2
Ponemon 2013	5000 - 100,000	54	$y = 1924.2x^{0.7662}$	0.52
Ponemon 2014	4700 - 103,000	61	$y = 2439.9x^{0.7499}$	0.50
NetDiligence (Verizon report)	2 -108 million	183	$y = 10002x^{0.4971}$	0.54

Note: R^2 of 0.5 suggests moderate correlation. There are other factors that impact cost.

Annual Cost Models

- $\boldsymbol{Expected\ Annual\ Security\ Cost} =$
  Annual expected costs due to breaches +
  Costs regardless of any breaches
- $\boldsymbol{Annual\ Expected\ Cost\ due\ to\ Breach} =$
  Σ Probability of a breach of data type i $\times$
  Total cost per breach for type i

Overall risk evaluation model



Colorado State University

Models for Partial costs

- Details in Abdullah Algarni's dissertation: Quantitative economics of security : software vulnerabilities and data breaches, CSU

- $\boldsymbol{Investigation\ cost\ per\ record}$
$$= [a * (size)^{b-1}\ for\ factors\ 4,5,6] * Fbreach\_cause * Fencryption * Fprivacy$$
- $\boldsymbol{Crisis\ Management\ Cost\ per\ Record}$
$$= [a * (size)\ ^\wedge\ (b-1)\ for\ factor\ 11] * FBCM$$
- $\boldsymbol{Sanctions\ cost\ per\ record}$
$$= a * (size)\ ^\wedge\ (b-1)\ \ for\ factor\ 14$$
- $\boldsymbol{Class\ Action\ Lawsuit\ Cost\ per\ record}$
$$= a * (size)\ ^\wedge\ (b-1)\ for\ factor\ 15\ and\ 16$$
- ***Opportunity cost: considered separately***

2020 Data

Ponemon Global Cost of Data Breach Study 2020

- 3,400-99,730 records
- Excludes mega-breaches, considered separately