Quantitative Cyber-Security

Colorado State University
Yashwant K Malaiya

CS559

L23 Breach Probability 2



CSU Cybersecurity Center
Computer Science Dept

Topics

Recall that Risk = Likelihood x Impact

- Probability of a breach: this data is hard to obtain
- Gordon-Loeb Model: impact of additional security in investments
- Breach cost components, models
- A Case study

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.

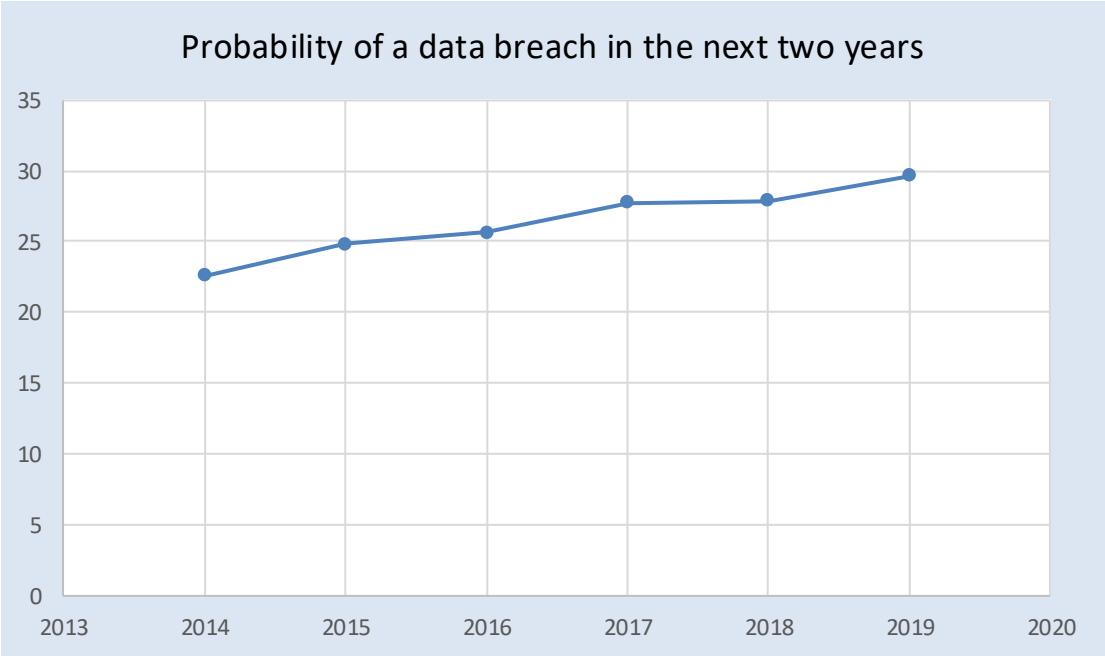
Updating Submodels

- The tables for the factor submodels should be updated time to time to account for recent variations.
- Cost of a Data Breach report is published annually.
- A website A Guide to Calculating the Cost of Data Breaches in Netherlands gives python code to implement our model!

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
- 2015-2019 (4 years): up by 5, i.e. 1.25%/year

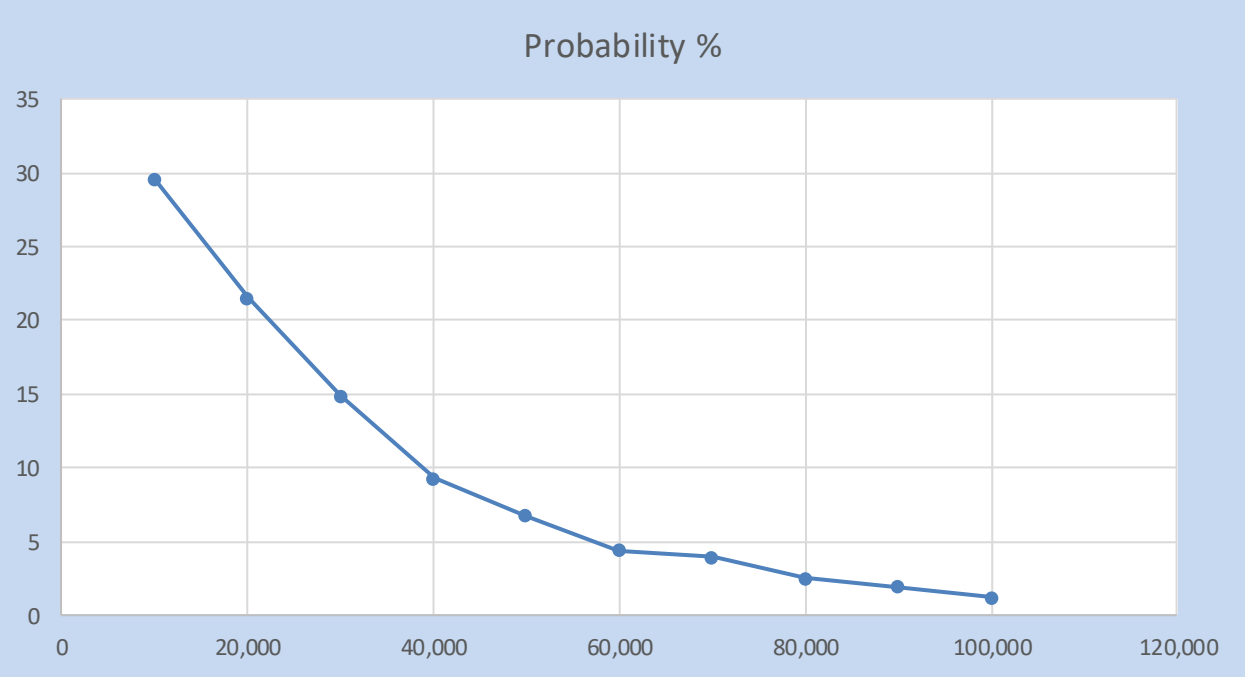


Colorado State University

Probability of a data breach by number of records lost

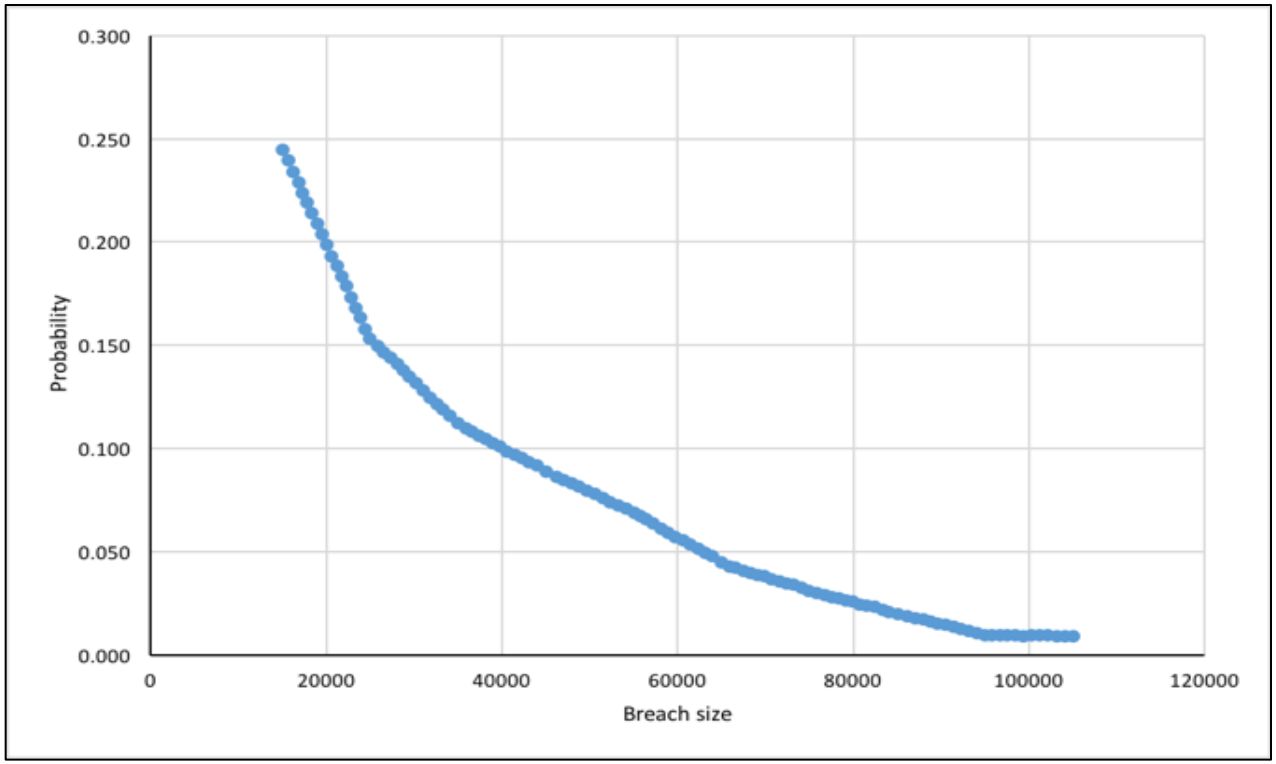
Over the next two years, involving minimum of 10,000 and maximum of 100,000 records.

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



Colorado State University

Breach probability -Breach size



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

Colorado State University

Breach probability -Breach size model

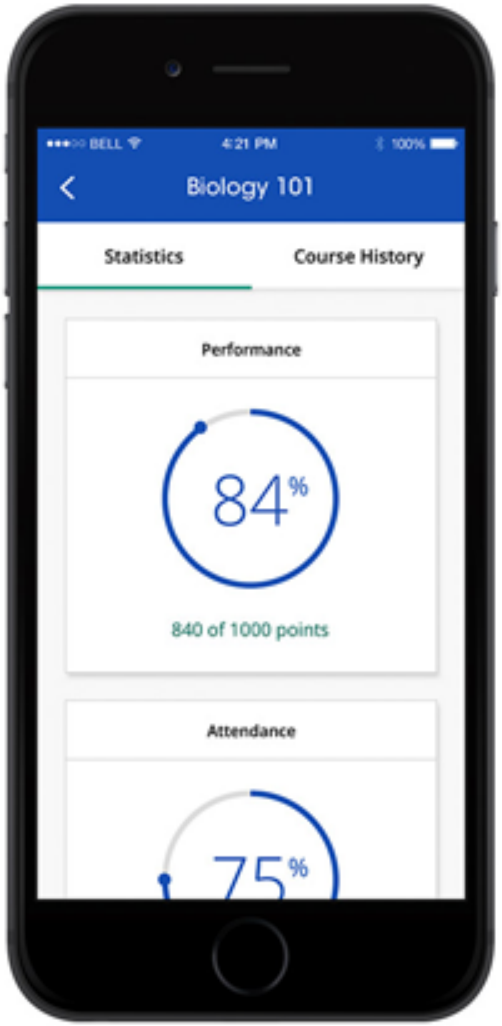
Probabilty of a breach in a year:

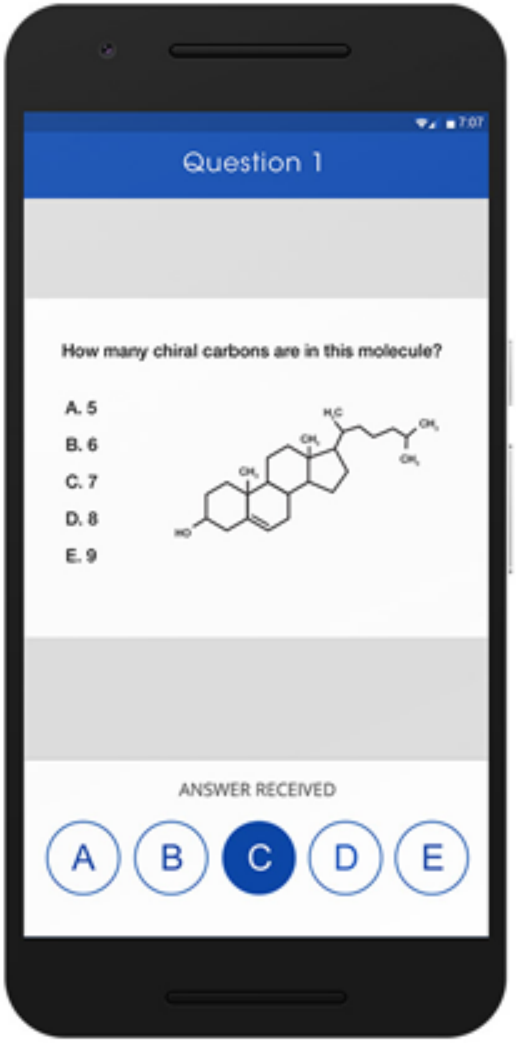
$F_{Breachsize} = \alpha \, exp\,(-\beta x)$

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

Based on 2015 Ponemon data. α may have gone by by 1.5% a year.

iClicker





Colorado State University

In class question

- What is the probably of a data breach in a 2-year period involving 20,000 records?
- A: 100%
- B: 21.5%
- C: 0.23%
- D: 53.5%

Answer

Colorado State University

In class question

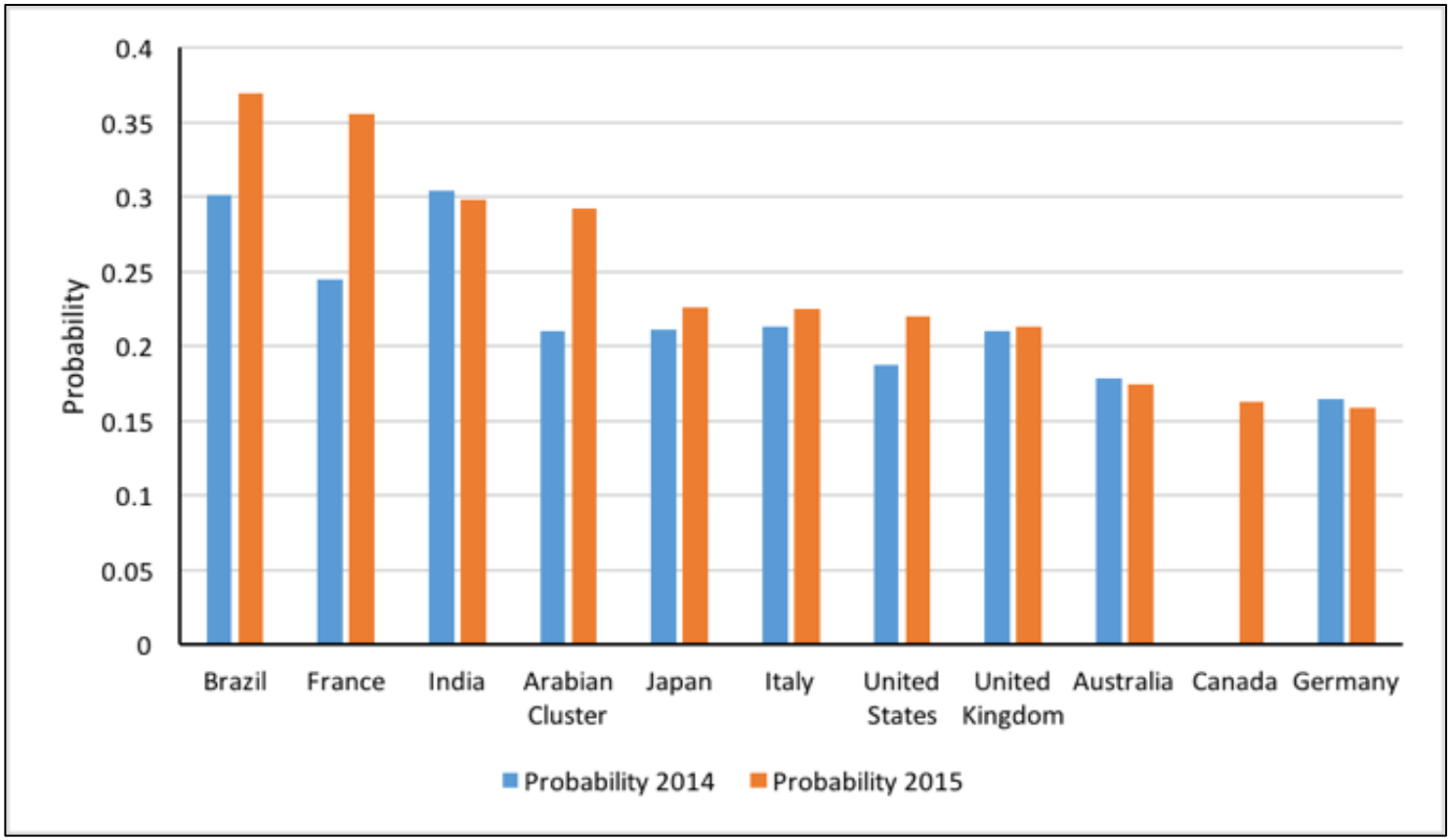
- What is the probably of a data breach in a 2-year period involving 20,000 records?
- A: 100%
- B: 21.5%
- C: 0.23%
- D: 53.5%

Back from iClicker



Colorado State University

Data breach probability by country



Data breach probability by country (Ponemon data 2015)

A minimum of 10,000 compromised records

Colorado State University

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)

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: $F_{encryption}$

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:

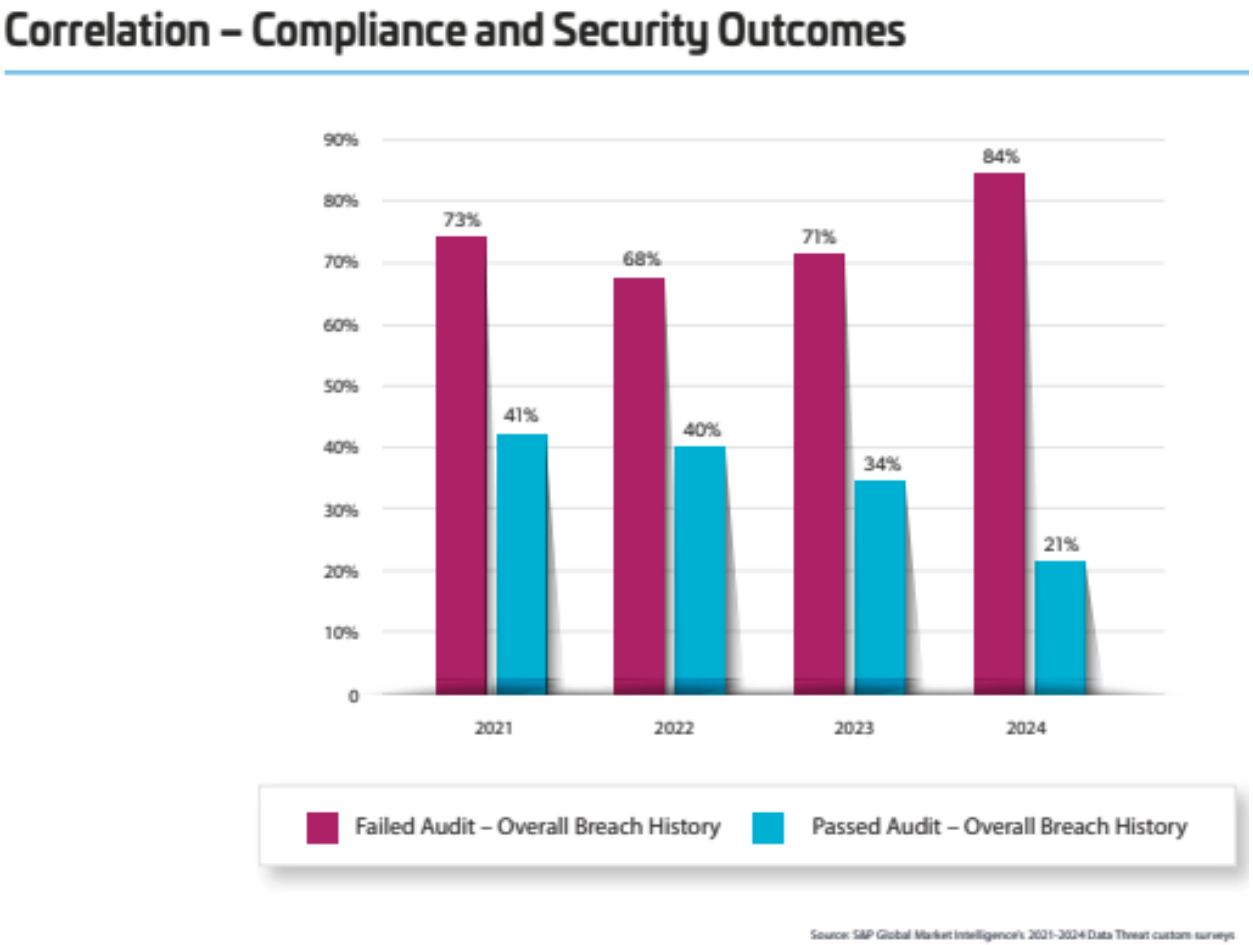
Data breach causes	Multiplier
Malicious or criminal attack	1.32
Negligence or mistakes (Human error)	0.82
System glitch	0.75
don't know	1 (default)

Updating Submodels

- The tables for the factor submodels should be updated time to time to account for recent variations.
- Cost of a Data Breach report is published annually; however the probability data is often not included.
- A website A Guide to Calculating the Cost of Data Breaches by a new cybersecurity company in Denmark gives python code to implement Algarni and Malaiya's model!

A. M. Algarni and Y.K. Malaiya, "A Consolidated Approach for Estimation of Data Security Breach Costs", 2nd Int.Conf. on Information Management (ICIM), London, 2016, pp. 26-39.
A. M. Algarni, V. Thayananthan and Y. K. Malaiya "Quantitative Assessment of Cybersecurity Risks for Mitigating Data Breaches in Business Systems" Appl. Sci, 19 April 2021, 11, 3678, pp. 1-24.

Compliance and Security Outcomes



In 2024, of the respondents whose organizations failed a compliance audit, 84% reported having some breach in their history, with **31% saying they experienced a breach in the last 12 months**. In contrast, for those that passed compliance audits, only 21% have a breach history and only **3% suffered a breach in the last 12 months**.
https://intexus.la/wp-content/uploads/2024/05/2024-thales-data-threat-report-global_PDF.pdf

Colorado State University

Quantitative Cyber-Security

Colorado State University

Yashwant K Malaiya

CS559

Gordon-Loeb Model impact of a security investment on the probability of a breach



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 Model

- 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*.
- I have kept the same notation which several following papers have used.
- Further work needed.

Comments on approach

Security breach probability function

Security breach probability function: Probability of a breach after an investment z is given by $S(z, v)$

- 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 $S(z, v)$ function :

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, $S'(z, v) < 0$ and $S''(z, v)>0$, where S' denotes the partial derivative with respect to z and S'' denotes the partial derivative of S' with respect to z.

- A3 implies that as the investment in security increases, the prob of a breach declines, but at a decreasing rate. *Law of diminishing returns.*
- 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.

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.

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

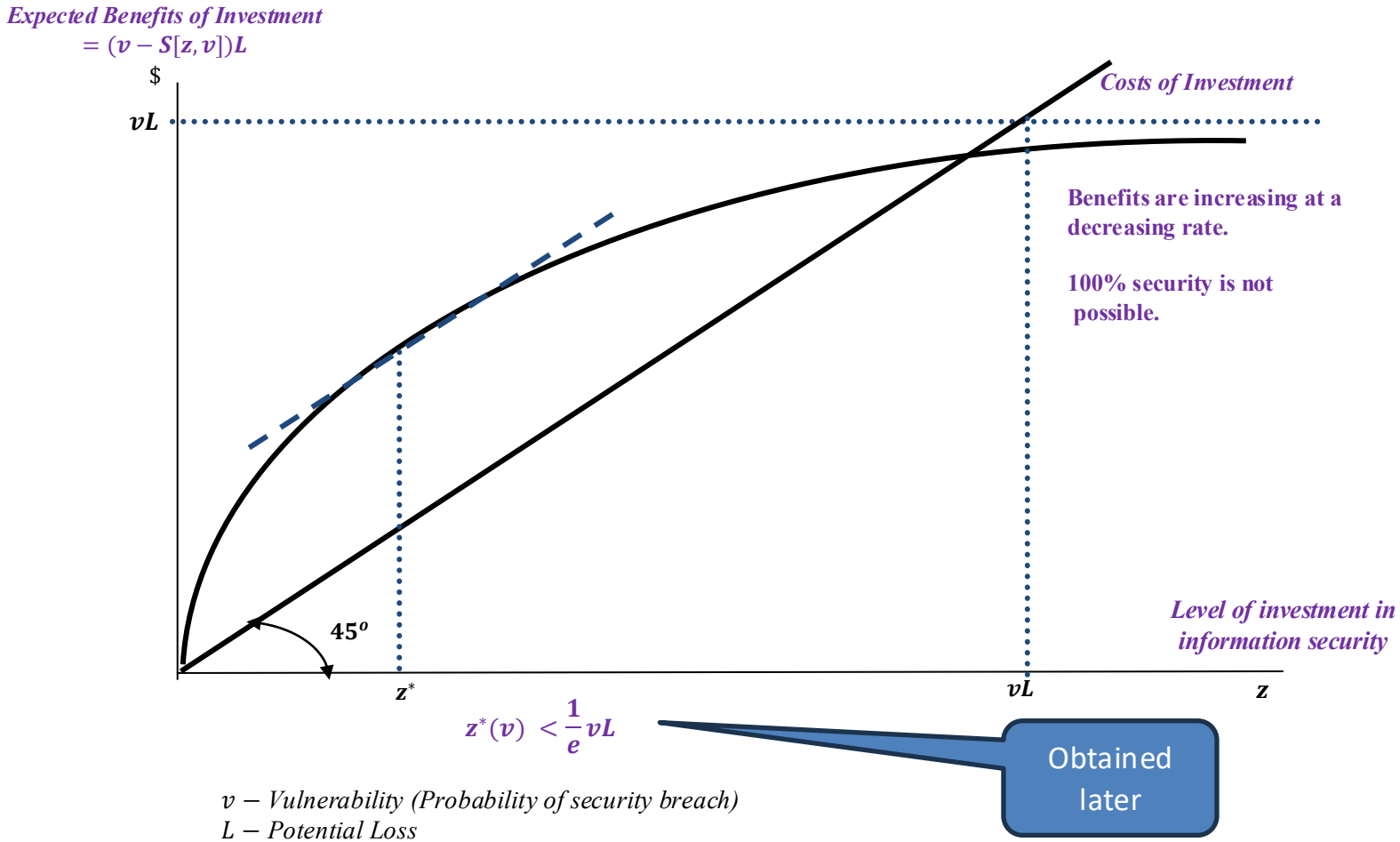
A risk measure like ALE

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*

Expected Benefits ($) vs Costs of an Investment (z) in 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)*

Optimal: The point the tangent to EBIS(z∗), has a slope one.

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 optimal investment is given by $z^*(v) < (1/e)vL$.

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

Optimal: The point the tangent to EBIS(z*), has a slope one.

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 $S^{I}(z,v)=v/(\alpha z+1)^{\beta}$ for some $\alpha>0$, $\beta\geq 1$) or to class II (i.e., it can be expressed as $S^{II}(z, v) = v^{\alpha z+1}$ for some $\alpha > 0$), 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 absence of any investment in security

Weakly increasing: may be occasionally flat

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 (risk matrix).

4. Focus spending where it should reap the largest net benefits based on productivity of investments.

Recent Developments

- Widely cited 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, $\alpha = 4x10^{-5}$, $\beta = 1$
    - Optimal about 0.2 v
- Table based investment distribution: based on risk values of each component.

Expected loss from information security breach, Cost saving from a 2 M investment

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: $vL \geq 70$

Value of Information Sets (in $ Million)

Vulnerability/Threat	Low	Medium	High		
	20	40	60	80	100
Low 20%	0.67	1.33	2.00	2.67	3.33
Medium 40%	1.11	2.22	3.33	4.44	5.56
60%	1.67	3.33	5.00	6.67	8.33
High 80%	1.41	2.81	4.22	5.63	7.04

Figure 5. Cost savings from 2nd million dollar investment.

Expected loss from information security breach
(Risk Matrix: each entry is risk in M$)

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

Colorado State University

Research Questions

I would like to get answers to these questions.

- What is the actual security breach probability function based on real data or current state of technology/economics/threats?
- Find information on approaches for security enhancements and their effectiveness.

Costs for outsourced cybersecurity services

- Minimum costs for outsourced cybersecurity services start around $2,000 - $3,500 per month
  - On a per-user basis between $195 and $350 per user, including support and maintenance.
- Organization already having support (in-house or outsourced),
  - the cybersecurity portion on its own is typically $35 - $65 per user.

VC3: How Much Do Managed Cybersecurity Services Cost?

Managed Cybersecurity Services

- **Advisory services**: regular reports and analyses of security events, incidents, and security posture. compliance with regulatory frameworks.
- **Cybersecurity software tools**: see next slide
- **Incident response**: a remediation plan
- **Policy writing**: guidelines for how you allow people and systems to access your network
- **Regulatory compliance**: specific regulations such as HIPAA, CMMC, GDPR, SEC, etc.
- **Software patching**software, applications, and systems are kept up to date
- **Vulnerability management**: identifying and addressing weaknesses.
- **Security awareness training**: Employee training programs about phishing, social engineering, and best practices for maintaining security.
- **Identity and access management**: managing and verifying the identities of individuals and devices.

Cybersecurity Services: Tools

- **Cybersecurity software tools**:
  - **Endpoint detection and response (EDR) coupled with 24/7/365 human monitoring:** endpoints such as computers, servers, and mobile devices are secure
  - **Multi-factor authentication (MFA)**
  - **Microsoft 365 protection and backups**
  - **Credential monitoring**: dark web monitoring
  - **Email protection:** encryption, filtering, and scanning
  - **Web and content protection**: to block employees from accessing potentially malicious or risky websites.
  - Security Incident Event Management (SIEM) / Security Orchestration, Automation, and Response (SOAR): real-time analysis of network hardware and application-generated security alerts.
  - Advanced malware protection: to detect, prevent, and respond to advanced malware threats (including zero-day attacks)
  - **Advanced cloud security:**
  - **Next-generation firewall:** firewall capabilities with additional features and functionalities to provide enhanced protection against modern cyber threats

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 -2023)
- **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.
- **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

Some major breaches

Organization	Records	Breach Date	Type	Source	Industry	Estimated Cost
Anthem insurance	78 million	Jan-15	Identify theft	Malicious outsider	Healthcare	$100 million
Yahoo	500 million	Dec-14	Account access	Russia?	Business	$350 million
Home depot	109 million	Sep-14	Financial access	Malicious outsider	Business	$28 million
JPMorgan chase	83 million	Aug-14	Identify theft	Malicious outsider	Financial	$13 billion
Benesse	49 million	Jul-14	Identify theft	Malicious insider	Education	$138 million
Korea credit bureau	104 million	Jan-14	Identify theft	Malicious insider	Financial	$100 million
Target	110 million	Nov-13	Financial access	Malicious outsider	Business	$252 million
Adobe System	152 Million	Sep-13	Financial access	Malicious outsider	Business	$714 Million
Equifax	147 Million	May-17	Credit ratings	China PLA?	Credit Rating	$1700 million

Long Cheng, Fang Liu, Danfeng (Daphne) Yao, Enterprise data breach: causes, challenges, prevention, and future directions, Wires, 09 June 2017.

Equifax to Pay $575 Million as Part of Settlement with FTC, CFPB, and States Related to 2017 Data Breach, 2019

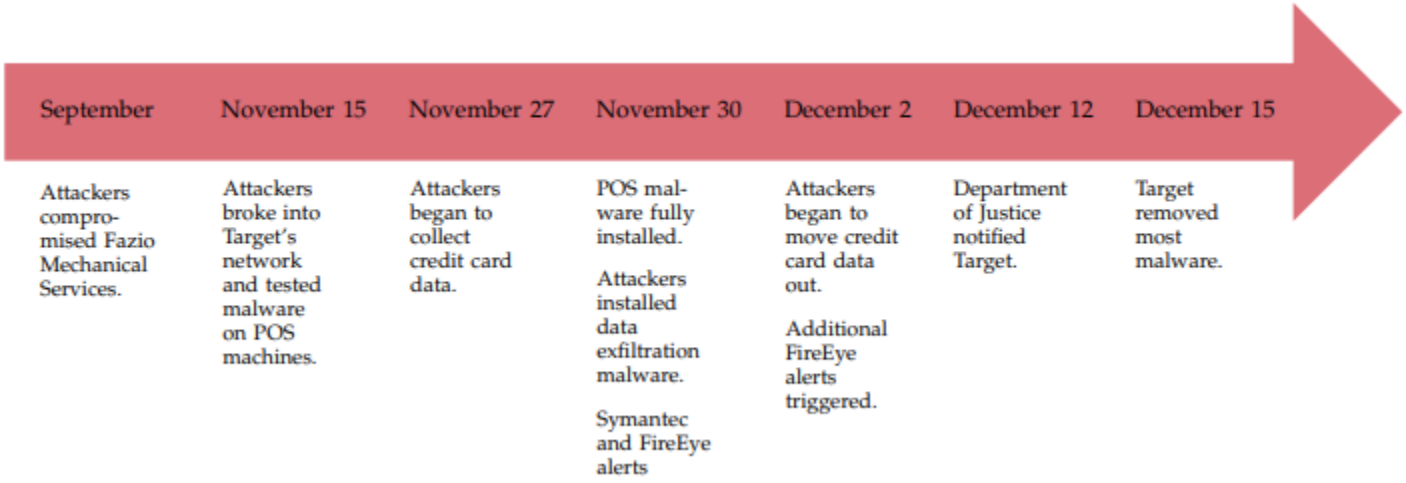
Note that the costs are *estimated*.

Detailed Examples

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

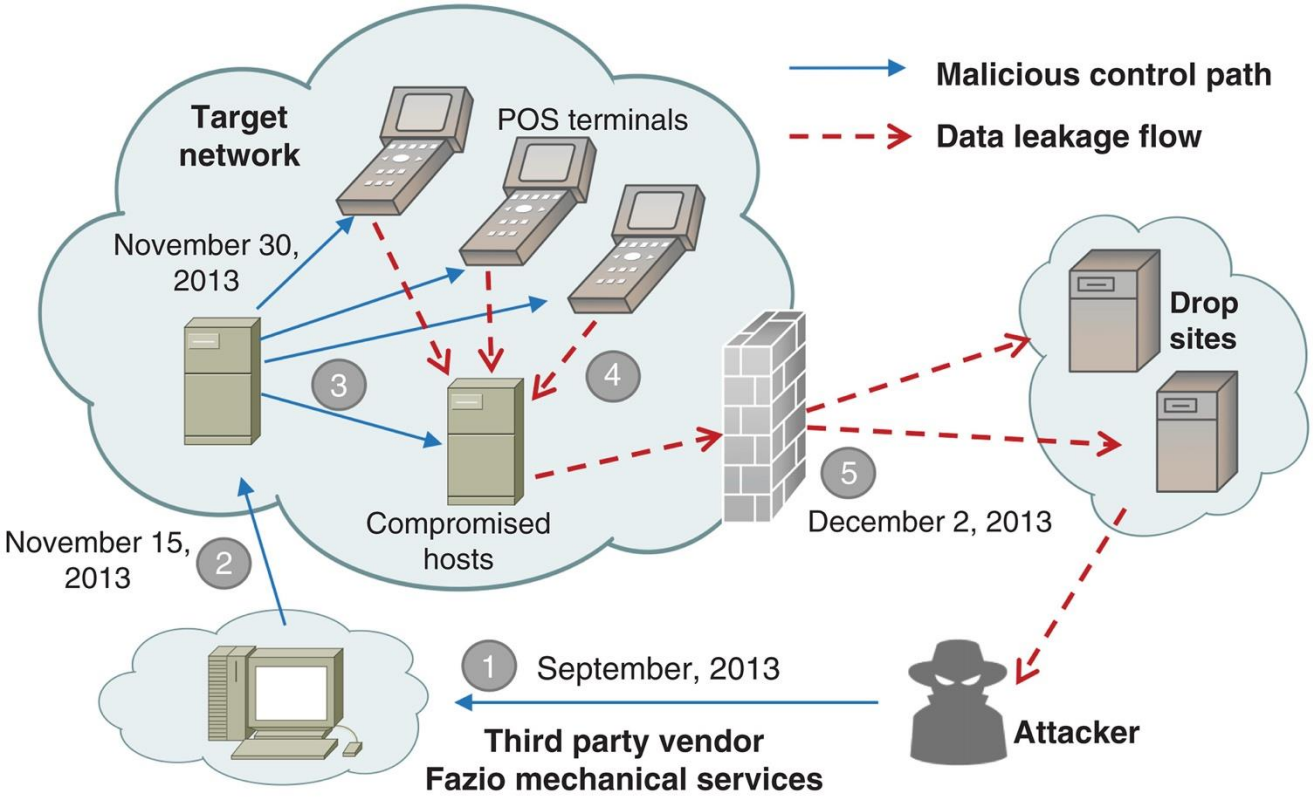
Target data breach (2013)

- 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

Breakdown and analysis of the Target data breach



WIREs Data Min & Knowl, Volume: 7, Issue: 5, First published: 09 June 2017, DOI: (10.1002/widm.1211)

Colorado State University

Target data breach (2013)



- TGT Price chart (Yahoo Finance)

Note: Announced Dec 19, 2013 to media.

Question: what was the impact on the enterprise value?

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)				

Observation: Insurance covered 35.7%. Govt covered 27.0%. Who ultimately paid for these?

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

Colorado State University

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, which had access to the Home Depot network.

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