UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

Cyber Security Risk Assessment
Fall 2016

Lecture 14

Quantitative Risk Analysis

Uncertainty

UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

Two Types of Uncertainty

- ***Epistemic***
  - The world is deterministic but we don't know it
    - What is the value of an already tossed coin hidden in my hand?
- ***Stochastic***
  - The world is not deterministic
    - What will be the value of tossing coin?
- ***In security both aspects are present***
  - Some attacks depends on the random layout of the memory → may not work all the time
  - Some attacks took place but we don't know it yet
- ***Mostly we use a Bayesian Interpretation***
  - Probability is a subjective degree of belief that we update given some information

UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

What to believe?

- ***Linda is***
  - 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student she was deepy concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
- ***Which is more likely?***
  - Linda is a bank teller
  - Linda is a feminist bank teller

UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

Where is the fallacy?

- ***What description implies***
  - Pr(L feminist) = High
  - Pr(L feminist bank teller | L is bank teller) = Very High
- ***What the question was about***
  - Pr(L is feminist bank teller) =

    Pr(L feminist bank teller | L is bank teller) *

    Pr(L is bank teller)
- ***Conditional vs Absolute Probability***

UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

How to protect?

- ***Bomber pilots can***
  - carry either a flak jacket or a parachute because of weight limitations. The probability of being strafed by enemy guns is ¾ (requiring flak jacket to survive) the probability of plane being shot down is ¼ (requiring parachute to survive)
- ***What is best?***
  - Flak jacket at all times
  - Parachute at all times
  - Flak jackets 3 times out of 4 and parachute 4th time
  - Flak jackets 1st time and parachute 3 out of 4 times

UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

The fallacy

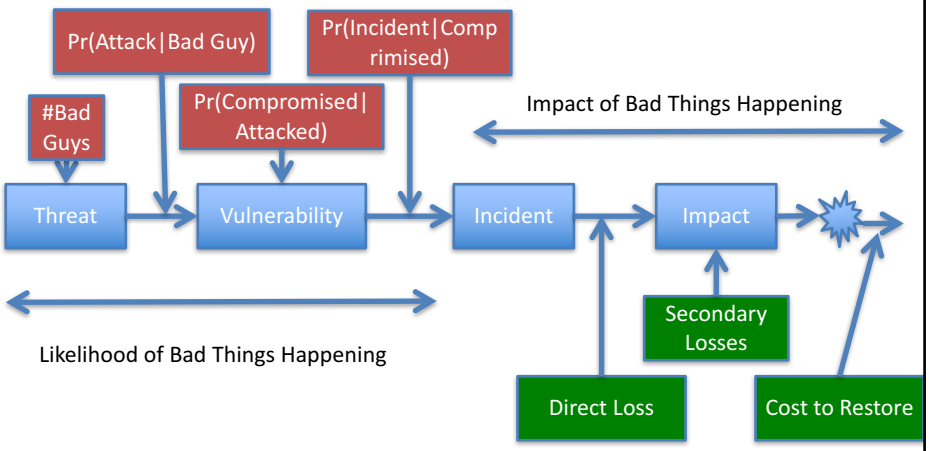
- ***Fallacy is chances don't repeat themselves***
  - The law of large numbers is actually working on large numbers…
  - UNLESS you really know this is a process that has memory (but then probability should be described differently)
- ***Pilot taking flak jacket first three times***
  - Clearly has not been shot down on the fourth one
  - So he has seen the series "strafed;strafed;strafed" → next time it is going to be "shotDown"
  - By taking the parachute the fourth time he has ¾ chance to die, ¼ of survival
    - Strafing and shooting down are independent on the previous event

UNIVERSITY OF TRENTO - Italy

Digital MASTER SCHOOL

Quantitative Risk Analysis - I

- ***Risk = Likelihood * Impact (negative)***



UNIVERSITY OF TRENTO - Italy

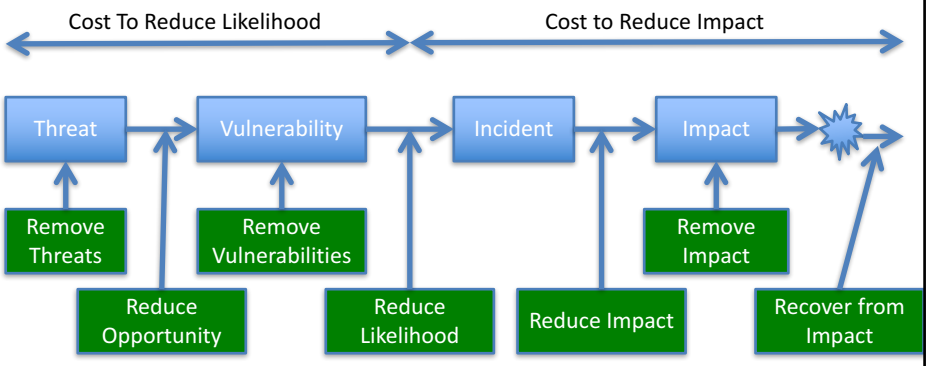
Digital MASTER SCHOOL

Quantitative Risk Analysis - II

- ***Fix an interval of observation (say N years)***

Benefit = + Likelihood*Impact – NewLikelihood*NewImpact

Value = + Benefit - Cost for NewLikelihood - Cost for NewImpact



UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

What we need to estimate

- ***#Threats***
  - Intentions to attack by cyber-terrorist, financially motivated criminals, hackivists, disgruntled employees, etc.
- ***Pr(Attack|Threat)***
  - If a given threat is active how many attacks are we going to get?
- ***Pr(Compromise|Attack)***
  - Once we are attacked would this generate an actual compromise of the machine (so the exploit would actually work)
- ***Pr(Incident|Compromise)***
  - Once we have been exploited, has this exploit been transformed into an incident that has a specific cost?
- ***May not be possible to estimate everything individually***
  - E.g. some probability might be difficult to disentangle from actual data

UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

Example Verizon DBiR

- ***Verizon Reports***
  - #Number of Incidents x Victim Type
  - #Number of Data Breaches x Victim Type
  - #Typology of attacks
- ***Example in 2015***
  - Retail 370 incidents, 182 breaches
  - Professional services 916 incidents, 53 breaches
  - Cyberspies 247 incidents, web app attack 5334
- ***What can be calculated?***

UNIVERSITY OF TRENTO - Italy

Digital MASTER SCHOOL

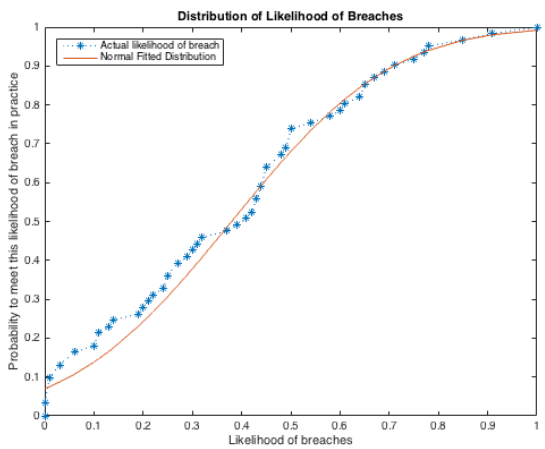
What we have

- ***#Threats → don't know***
- ***Pr(Attacks|Threats) * Threats → "incidents" in Verizon terminology***
- ***Exploited Vulnerabilities → don't know***
  - From the data I report here, the information of exploited attack is actually there in the DB, they only tell us the gross totals in this report
- ***Pr(Incidents|Compromise) * Pr(Compromise|Attack) * Pr(Attack|Threats) * Threats → "data losses"***
- ***Can reconstruct***
  - Pr(Incidents|Threats) = "Verizon data losses"/"Verizon's incidents"

UNIVERSITY OF TRENTO - Italy

Digital MASTER SCHOOL

Global Data of Breach Likelihood



UNIVERSITY OF TRENTO - Italy

Digital MASTER SCHOOL

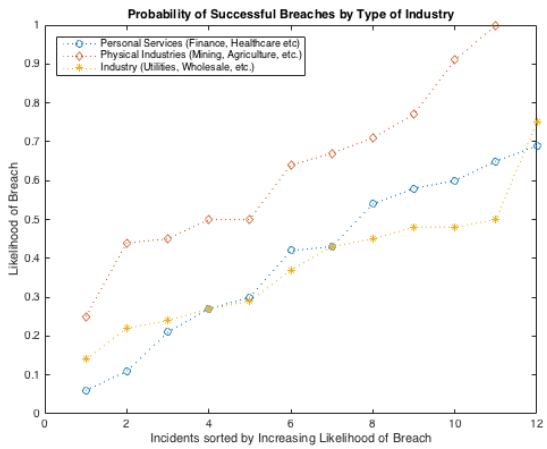
Data on Three Industries

- ***Average Attacks***
  - Personal Services (Finance, health) = 435.2
  - Physical Production (Agriculture etc) = 8.5
  - Industries (Utilities, Wholesale, etc.) = 112.1
- ***Average Probability of Success***
  - Personal Services (Finance, health) = 0.41
  - Physical Production (Agriculture etc) = 0.62
  - Industries (Utilities, Wholesale, etc.) =0.39

UNIVERSITY OF TRENTO - Italy

Digital MASTER SCHOOL

Also distribution and type matter



UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

Scale down to the company

- ***What are we missing?***
  - We don't know how many industries are in the sample by Verizon.
  - We must investigate that into the database
  - 200 Attacks over 100 Companies means 2 attacks against your company per year.
- ***Compute the final Likelihood***
  - Avg(Attacks) * Avg(Prob of Breaches) / Firms
- ***Multiply by impact → (average) risk***

UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

Refining the analysis

- ***Exploited Vulnerabilities***
  - If we have access to the data we can use this information to estimate the effect of countermeasures
- ***Pr(Compromise|Attack) → Pr(Compromise|Attack & CVSS=x)***
  - How many vulnerabilities with a given CVSS score have been attacked
  - How many of them has been the cause of a data breach?
  - If we remove the vulnerabilities with highest probability → reduce likelihood
- ***Approximate calculation also possible***
  - Assume that vuln with CVSS=10 yields a compromise with Pr=1
  - Conservative but may be an overkill

UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

What About Extreme Risks?

- ***So far we calculated averages of success***
  - Personal Services (Finance, health) = 0.41
  - Physical Production (Agriculture etc) = 0.62
  - Industries (Utilities, Wholesale, etc.) =0.39
- ***Which is the 90-percentile of success for breaches?***
  - Personal Services (Finance, health) = 0.58
  - Physical Production (Agriculture etc) = **0.95**
  - Industries (Utilities, Wholesale, etc.) = 0.66

UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

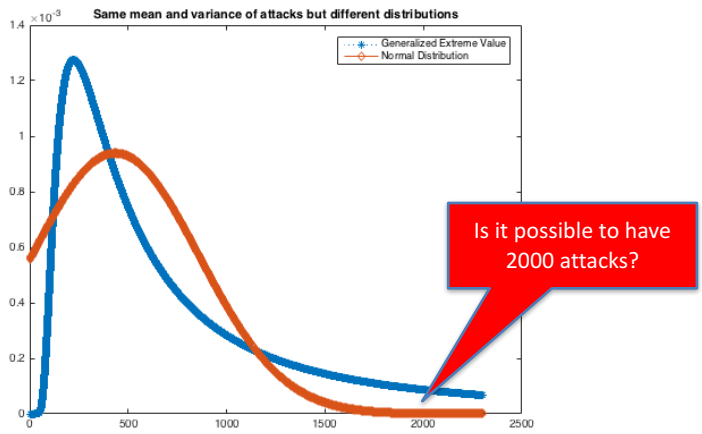
The Problem

- ***With Normal/Poisson distributions***
  - as we we go away from the average → there are very few dangerous cases
  - So we can reasonably use the average, at most moderate with the standard deviation
- ***Our data is not normal***
  - we have very fat tails of the distribution → dangerous cases may not be so few
  - Extreme risks may be less rare than we thought
- ***We need to estimate "worst cases"***

UNIVERSITY OF TRENTO - Italy

Digital MASTER SCHOOL

Two distributions with same μ and σ



UNIVERSITY OF TRENTO - Italy

Digital MASTER SCHOOL

Estimating Extreme Risks

- ***Simplified version of Basel-II approach***
  - Used by banks to estimate rare but anyhow big risks
    - Banks do a double convolution: estimate likelihood and estimate losses → we take impact as given
- ***Approach***
  - Collect data of value of the variable of interests (attack, percentage of success, vulnerabilities, etc.)
  - Fit data into a distribution (try both slim and heavy tails)
  - Calculate at the alpha-percentile of the distribution
  - This is the number we use for the calculation of the final risk

UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

Poisson Distribution

- ***Key Idea***
  - Probabily that n incidents will happen in a give time interval → decreases linearly with the size
  - $Pr(k) = Pr(k-1) * \lambda/k$
- ***Typically very thin tail***
  - Large number of incidents are very very rare
- ***Cumulative Distribution***
  - $Pr(X < x) = e^{-\lambda} \Sigma^{x}_{i=0} \lambda^{x} /x!$
- ***Parameter estimation from data***
  - $\lambda = 1/n \, \Sigma^{n}_{j=0} x_j$

UNIVERSITY OF TRENTO - Italy

eit Digital MASTER SCHOOL

Pareto Distribution

- ***Key Idea***
  - Power Law for distribution of income
  - The people with a (large) pot of money m are progressively fewer and fewer i.e. they are only a/m^b
  - Used to model large losses (m) in property and liability insurance → the larger the b the more likely we have people with large losses
    - Typical values of b for earthquakes (1), fire industry (1.5), general liability (1.8), occupational injuries (2), motor liability (2.5)
- ***Cumulative Distribution***
  - $Pr(X<x) = 1 - (a/x)^b$
- ***Parameters estimation from data***
  - $a = \min(x_j)$
  - $b = n \, [\Sigma^{n}_{j=1} \log(x_j/a)]^{-1}$

UNIVERSITY OF TRENTO - Italy

Generalized Extreme Value Distribution

- ***Key Idea***
  - Try to captures the possible maxima (or minima) of a batch of random values
  - If the tail goes esponentially to zero → collapse to normal/Poisson distribution
  - If the tail goes polynomially to zero → Student's t distribution or Frechet's distribution
  - If the tail is bounded → Beta distribution
- ***E.g. Cumulative Distribution (Frechet)***
  - $Pr(X < x) = e^{-b/(x-a)^C}$

UNIVERSITY OF TRENTO - Italy

Estremal Values for Likelihood

- ***Data***
  - Use the "Incidents" in Verizon DBiR terminolgy
- ***Goal***
  - We want to now the worst possible number of attacks, at 95% percentile for different type of small firms
    - Banks have to calculate at the 99.9% (but we don't have enough data here)
- ***Process***
  - Compare Distributions
    - Actual (the empirical distribution), Poisson, Generalised Extreme Value, Pareto Tails
  - Find best distribution
    - We do this "visually", should be done with statistical tests → advanced courses
  - Return the inverse value of the 95% percentile

UNIVERSITY OF TRENTO - Italy

Digital MASTER SCHOOL

Extremal Number of Attacks

95%	Administr.	Consumers	Industry	Personal	Production
Empirical	26	179	18	50*	4
Normal (fit)	24	164*	13	50	3
Poisson	15	80	9	34	3**
GEV	30**	374*	16**	50*	1343
ParetoTails	24	169	17	49	4

Starred nodes correspond to the distributions that seem to fit best (from the plots)

UNIVERSITY OF TRENTO - Italy

Digital MASTER SCHOOL

Further reading

- ***Chapters 10, 11 on Textbook***
- ***Chapters 1-3, Claudio Franzetti, "Operational Risk Modelling and Management", CRC Press***
- *Ross Anderson's book*
- *L. Allodi, F. Massacci.* Comparing vulnerability severity and exploits using case-control studies. ACM Trans. on Information and System Security, 17(1):1 (2014).