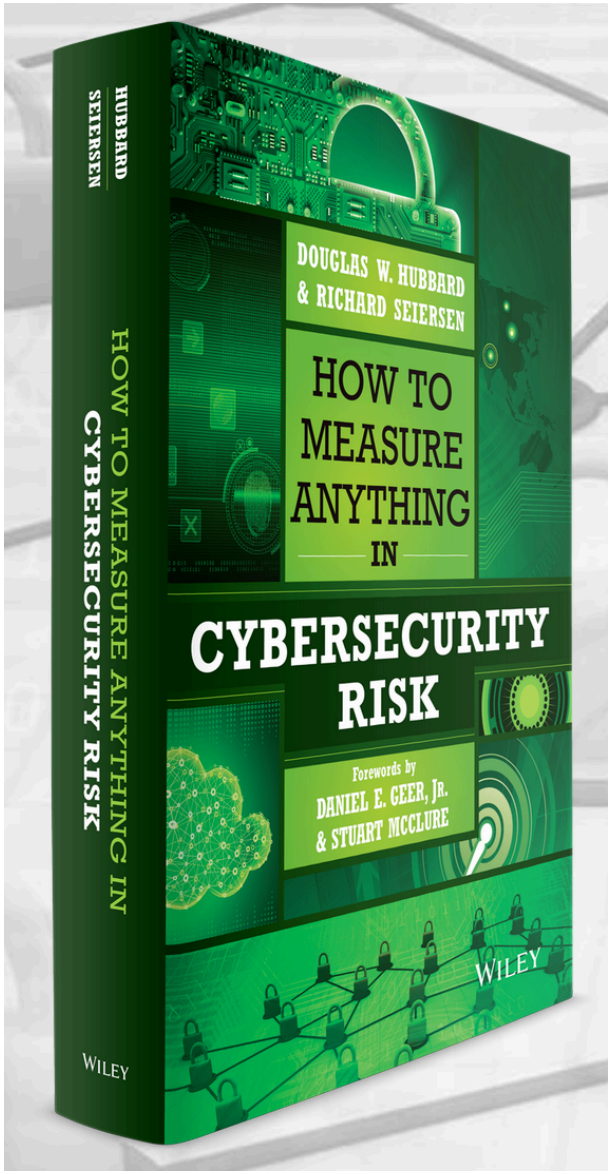
Section 4

Cybersecurity Risk Management

"Cyber security is about risk management at the end of the day, network technology will never be completely secure." @TomBossert45

—Thomas Bossert[1], Cyber Week 2017, Tel Aviv

[1] Assistant to the President for Homeland Security and Counterterrorism



Reasoning about uncertainty in security

"The book in particular offers an alternative to a set of deeply rooted risk assessment methods now widely used in cybersecurity but that have no basis in the mathematics of risk or scientific method. We argue that these methods impede decisions about a subject of growing criticality. We also argue that methods based on real evidence of improving decisions are not only practical but already have been applied to a wide variety of equally difficult problems, including cybersecurity itself."

Topics include

- security
- uncertainty
- measurement
- risk assessment
- misunderstandings of the above terms
- a better approach to measuring cybersecurity risk
- measuring the performance of cybersecurity risk analysis
- a simple quantitative measuring method
- how to improve models with minimal data

Cyber Threat Perspective

"Nation-states, organized crime, hacktivist entities, and insider threats want our secretes, our money, and our intellectual property, and some want our complete demise. Sound dramatic? If we understand the FBI correctly, they expect to ***spend as much or more*** *on protecting us from cyber threats than from those who would turn airplanes, cars, pressure cookers, and even people into bombs."* — Hubbard and Seiersen, 2016

FireEye Labs: Cyber Threat Map

The rules of engagement in today's threat landscape are changing rapidly and as cyber-crime evolves, there is a security gap that can be exploited. As our dependency on technology further permeates our daily habits, the threats that exploit the security gap will have graver consequences.
— FireEye, Inc., 2018

Cyber Threat Perspective

Nation-states, organized crime, hacktivist entities, and insider threats want our secretes, our money, and our intellectual property, and some want our complete demise. Sound dramatic? If we understand the FBI correctly, they expect to ***spend as much or more*** *on protecting us from cyber threats than from those who would turn airplanes, cars, pressure cookers, and even people into bombs.* — Hubbard and Seiersen, 2016

But there is also good news:

> "We're as close as possible to our unemployment rate being zero," says Sam Olyaei, senior research analyst, at Gartner Security & Risk Assessment Summit in National Harbor, MD. "If you're a cybersecurity professional with any kind of skill set, you already have a job and multiple offers on the table." … There are currently more than 348,000 open security positions, according to CyberSeek. By 2022, there will be **1.8 million unfilled positions**, according to the Center for Cyber Safety and Education.[2]

[2] Kasey Panetta, Confront the Cybersecurity Talent Shortage. Gartner, June 2017 [Online]. https://www.gartner.com/smarterwithgartner/solve-the-cybersecurity-talent-shortage/ [Accessed: October 2018]

Global Attack Surface

"Perhaps the total attack surface that concerns all citizens, consumers, and governments is a kind of `**global attack surface**': the total set of cybersecurity exposures—across all systems, networks, and organizations—we all face just by shopping with a credit card, browsing online, receiving medical benefits, or even just being employed. This global attack surface is a macro-level phenomenon driven by at least four macro-level causes of growth: …

"Perhaps the total attack surface that concerns all citizens, consumers, and governments is a kind of `**global attack surface**': the total set of cybersecurity exposures—across all systems, networks, and organizations—we all face just by shopping with a credit card, browsing online, receiving medical benefits, or even just being employed. This global attack surface is a macro-level phenomenon driven by at least four macro-level causes of growth:

1. ***The increasing number of persons on the Internet.***
   - Internet users worldwide grew by a factor of 6 from 2001 to 2014 (from half a billion to 3 billion).
   - It may not be obvious that the number of users is a dimension in some attack surfaces, but some measures of attack surface also include the value of a target, which would be partly **a function of number of users** (e.g., gaining access to more personal records).
   - Also, on a global scale, it acts as an important multiplier on the following dimensions.

2. ***The number of uses per person for online resources.***
   - The varied uses of the Internet, total time spent on the Internet, use of credit cards, and various services that require the storage of personal data-automated transactions are growing. Per person. Worldwide.
   - For example, since 2001 the number of websites alone has grown at a rate five times faster that the number of users—a billion total by 2014.
   - Connected devices constitute another potential way for an individual to use the Internet even without their active involvement. One forecast regarding the 'Internet of Things' (IoT) was made by Gartner, Inc: "4.9 billion connected things will be in use in 2015, up 30 percent from 2014, and will reach 25 billion by 2020." …

3. ***Vulnerabilities increase.***
   - A natural consequence of the previous two factors is the number of ways such uses can be exploited increases. This is due to systems and devices with potential vulnerabilities, even if vulnerabilities per system or device do not increase.
   - At least the number of discovered vulnerabilities will increase partly because the number of people actively seeking and exploiting vulnerabilities increases. …

4. ***The possibility of a major breach "cascade".***
   - More large organizations are finding efficiencies from being more connected. The fact that Target was breached through a vendor raises the possibility of the same attack affecting multiple organizations. Organizations like Target have many vendors, several of which in turn have multiple large corporate and government clients.
   - Mapping this **cyber-ecosystem of connections** would be almost impossible, since it would certainly require all these organizations to divulge sensitive information.
   - So the kind of publicly available metrics we have for the previous three factors in this list do not exist for this one.

4. ***The possibility of a major breach "cascade".***
   - More large organizations are finding efficiencies from being more connected. The fact that Target was breached through a vendor raises the possibility of the same attack affecting multiple organizations. Organizations like Target have many vendors, several of which in turn have multiple large corporate and government clients.
   - Mapping this **cyber-ecosystem of connections** would be almost impossible, since it would certainly require all these organizations to divulge sensitive information.
   - So the kind of publicly available metrics we have for the previous three factors in this list do not exist for this one.

It seems reasonable that of these four trends the earlier trends magnify the latter trends. If so, the risk of the major breach "cascade" event could grow faster than the growth rate of the first couple of trends."

Source: Hubbard and Seiersen, 2016

Hypothesis

Attack surface and breach are correlated. If this holds true, we haven't seen anything yet.

Example: Intelligent transportation systems (ITS)

Automated and autonomous vehicles

Connected and automated vehicle (CAV) technologies are among the most heavily researched automotive technologies.

- The vehicle technologies currently available are only a fraction of what is being developed for the future.

- The technologies for autonomous cars, connected cars, and advanced driver assistance systems overlap:

  Fully automated, autonomous, or "self-driving" vehicles are defined by the U.S. Department of Transportation's National Highway Traffic Safety Administration (NHTSA) as "those in which **operation of the vehicle** occurs without direct driver input to control the steering, acceleration, and braking and are designed so that the driver is not expected to constantly monitor the roadway while operating in self-driving mode."

Cyber Attack Surface

- CAV technology offer enhanced safety, reduced congestion, improved emissions and advanced road design through connectivity between vehicles and the wider CAV infrastructure within urban areas.
- Wireless networks will allow vehicles to **communicate with smart traffic management systems** in real time, sharing information such as live traffic conditions and emergency situations and enable automatic controls to respond by adjusting signal phases to optimize routing and minimize overall congestion or navigate emergency vehicles.
- Given the vital role of **urban transport networks for smart cities** and their increasing reliance on automated services, cyber security is a key aspect in designing and rolling out smart infrastructure on which we will depend: "The current attack surface for cities is huge and wide open to attack." (Perlroth2015).
- A cyber attack on **CAV smart infrastructure** at peak traffic hours could have catastrophic consequences; for instance, producing massive gridlock by taking over traffic controls for major intersections could cripple emergency services for the entire city region over extended time periods.

Cybersecurity needs better risk measurements

Risk management is problematic

- the size of the attack surface and the volume of vulnerabilities, attacks, and compromises means organizations must make tough choices
- not everything can be fixed, stopped, recovered, etc.
- there will thus always be some *acceptable* (or *tolerable)* loses
- **what risks are acceptable** is often not documented or not clearly stated in quantifiable terms
- making calculations to determine if a given **expenditure is justified or not** infeasible

Cybersecurity needs better risk measurements

Risk management is problematic

- the size of the attack surface and the volume of vulnerabilities, attacks, and compromises means organizations must make tough choices
- not everything can be fixed, stopped, recovered, etc.
- there will thus always be some *acceptable*, or *tolerable*, loses
- **what risks are acceptable** is often not documented or not clearly stated in quantifiable terms
- making calculations to determine if a given **expenditure is justified or not** infeasible

"Vulnerability management" and "threat management" can be generalized to **security management**

How do organizations conduct security management?

How can one

- prioritize the allocation of limited resources to address an ever growing list of vulnerabilities?
- reason about cybersecurity decisions in a fight against uncertain and growing risks?

What was risk again?

Risk is the intersection of assets, threats, and vulnerabilities.

Risk is a function of threats exploiting vulnerabilities.

… quantifying the potential for loss, damage or destruction of an asset.

$$\text{Risk} = \left(\frac{\text{Vulnerability } x \text{ Threat}}{\text{Counter-Measure Score}}\right) x \text{ Valuation}$$

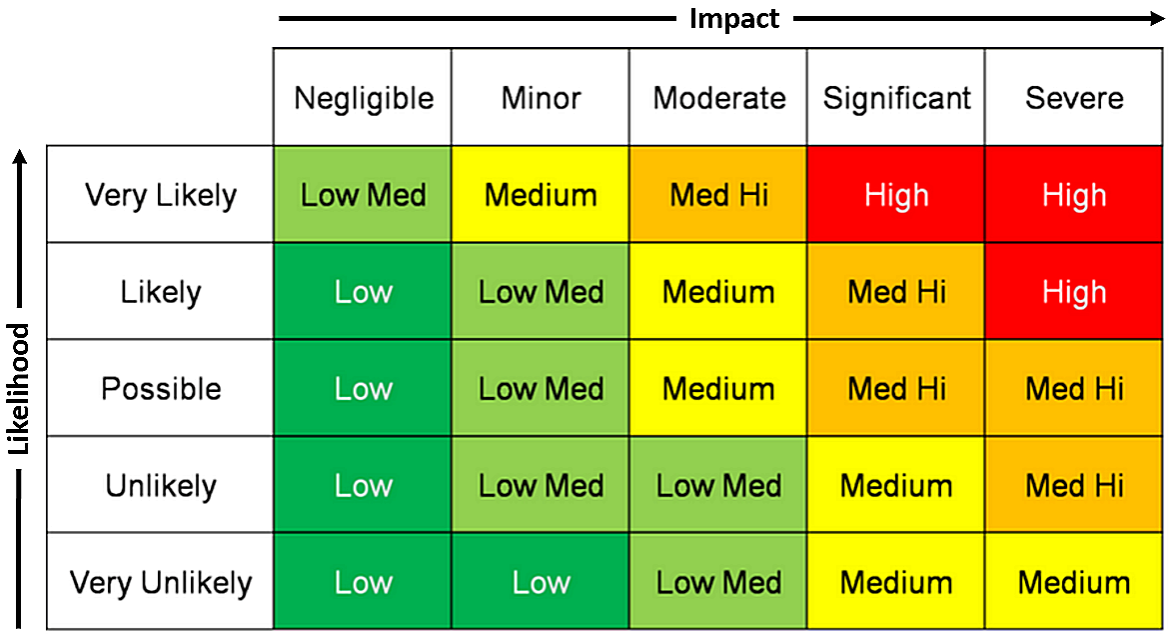
The valuation is the *estimated value* to the organization of the asset that is at risk.

Countermeasures mitigate threats (e.g., a firewall to stop unauthorized access to servers).

Common ~~best~~ practices

Expert intuition generally helps with management in many situations

For more systematic approaches, the vast majority of organizations resorts to some sort of "scoring" method to rank risks: the higher the score, the sooner one needs to address the risk.



Likelihood \ Impact	Negligible	Minor	Moderate	Significant	Severe
Very Likely	Low Med	Medium	Med Hi	High	High
Likely	Low	Low Med	Medium	Med Hi	High
Possible	Low	Low Med	Medium	Med Hi	Med Hi
Unlikely	Low	Low Med	Low Med	Medium	Med Hi
Very Unlikely	Low	Low	Low Med	Medium	Medium

RISK ASSESSMENT MATRIX

RISK RATING KEY	LOW	MEDIUM	HIGH	EXTREME
	0 – ACCEPTABLE	1 – ALARP (as low as reasonably practicable)	2 – GENERALLY UNACCEPTABLE	3 – INTOLERABLE
	OK TO PROCEED	TAKE MITIGATION EFFORTS	SEEK SUPPORT	PLACE EVENT ON HOLD

		SEVERITY			
		ACCEPTABLE LITTLE TO NO EFFECT ON EVENT	TOLERABLE EFFECTS ARE FELT, BUT NOT CRITICAL TO OUTCOME	UNDESIRABLE SERIOUS IMPACT TO THE COURSE OF ACTION AND OUTCOME	INTOLERABLE COULD RESULT IN DISASTER
LIKELIHOOD	IMPROBABLE RISK IS UNLIKELY TO OCCUR	LOW – 1 –	MEDIUM – 4 –	MEDIUM – 6 –	HIGH – 10 –
	POSSIBLE RISK WILL LIKELY OCCUR	LOW – 2 –	MEDIUM – 5 –	HIGH – 8 –	EXTREME – 11 –
	PROBABLE RISK WILL OCCUR	MEDIUM – 3 –	HIGH – 7 –	HIGH – 9 –	EXTREME – 12 –

Dilemma

"You can't control what you can't measure." — Tom DeMarco[1]

"Literally hundreds of security vendors and even standards bodies have come to adopt some form of **scoring system**. Indeed, scoring approaches and risk matrices are at the core of the security industry's risk management approaches. In all cases they are based on the idea that such methods are of some sufficient benefit. That is, they are assumed to be at least an improvement over not using such a method.

… let's be clear about our position on current methods: They are a failure. They do not work."

"A thorough investigation of the research on these methods and decision-methods in general indicates the following:

- There is **no evidence** that the types of scoring and risk matrix methods widely used in cybersecurity improve judgement.
- …

[1]Controlling Software Projects, Management Measurement & Estimation, (1982), p. 3.

- On the contrary, there **is evidence** these methods *add noice and error* to the judgment process … Any appearance of "working" is probably a type of "analysis placebo." That is, a method may make you feel better even though the activity provides no measurable improvement in estimating risks (or even adds error).

- There is **overwhelming evidence** in published research that quantitative, probabilistic methods are effective.

- Fortunately, most cybersecurity experts seem willing and able to adopt better quantitative solutions. But common misconceptions held by some—including misconceptions about basic statistics—create some obstacles for adopting better methods."

Hubbard and Seiersen, 2016, pp. 14-15

Risk Assessment

Video materials

- Introduction to risk assessment
  https://www.youtube.com/watch?v=EWdfovZ1g2g

- Why most risk assessments are wrong and dangerous
  https://www.youtube.com/watch?v=PA9rqNBZWIw

Cybersecurity Measurement Primer

Three reasons why anyone ever thought something was immeasurable, cybersecurity included, and all three are rooted in misconceptions:

1. Concept of measurement. A lot more things become measurable, once one understands the meaning of "measurement".

2. …

Cybersecurity Measurement Primer

Three reasons why anyone ever thought something was immeasurable, cybersecurity included, and all three are rooted in misconceptions:

1. *Concept of measurement.* A lot more things become measurable, once one understands the meaning of "measurement".

2. *Object of measurement.* What is being measured is often not well defined, using sloppy and ambiguous language.

3. …

Yahoo! data breaches

- Yahoo! reported two major data breaches of user account data to hackers during the second half of 2016. The first announced breach, reported in September 2016, had occurred sometime in late 2014, and affected over 500 million Yahoo! user accounts.
- A separate data breach, occurring earlier around August 2013, was reported in December 2016. Initially believed to have affected over 1 billion user accounts, Yahoo! later affirmed in October 2017 that all **3 billion of its user accounts** were impacted.
- Both breaches are considered the largest discovered in the history of the Internet. Specific details of material taken include: names, email addresses, telephone numbers, encrypted or unencrypted security questions and answers, dates of birth, and hashed passwords.
- Yahoo! has been criticized for their late disclosure of the breaches and their security measures, and has been facing several lawsuits as well as investigation by members of the United States Congress.
- The breaches have impacted Verizon Communications's July 2016 plans to acquire Yahoo! for about $4.8 billion, which resulted in **a decrease of $350 million** in the final price on the deal closed in June 2017.

Fallout: severe damage to the company's reputation

Cybersecurity Measurement Primer

Three reasons why anyone ever thought something was immeasurable, cybersecurity included, and all three are rooted in misconceptions:

1. *Concept of measurement.* A lot more things become measurable, once one understands the meaning of "measurement".

2. *Object of measurement.* What is being measured is often not well defined, using sloppy and ambiguous language.

3. *Methods of measurement.* Many procedures of empirical observation are not well known; otherwise, it would become apparent that many things thought to be immeasurable have been measured already.

1. Concept of measurement

The *meaning of measurement* is often not well understood by (cyber)security experts and described as

- "to quantify something"
- "to compute an exact value"
- "to reduce to a single number"
- "to choose a representative amount"

implying that measurement is a single, exact number with no room for error.

1. Concept of measurement

The *meaning of measurement* is often not well understood by (cyber)security experts and described as

- "to quantify something"
- "to compute an exact value"
- "to reduce to a single number"
- "to choose a representative amount"

implying that measurement is a single, exact number with no room for error.

Typical statements include something like:

> "We can't measure the true impact of a data breach because some of the consequences can't be known exactly."

> "There is no way we can put a probability value on being the target of a massive DDoS attack because there is too much uncertainty."

These and similar statements indicate a presumed definition of measurement that is **unscientific** and **unrelated to real decision making**.

Definition of measurement

Practical decision-making is based on an understanding of measurement as ***observations that quantitatively reduce uncertainty***. Indeed, a mere reduction—not necessarily an elimination—of uncertainty is what one can expect from measurement.

Measurement is only a probabilistic exercise.

The fact that some (amount of) error is **unavoidable** but the result can still be an improvement on prior knowledge is central to how experiments, surveys, and other scientific measurement are performed.

Definition of Measurement

Practical decision-making is based on an understanding of measurement as ***observations that quantitatively reduce uncertainty***. Indeed, a mere reduction—not necessarily an elimination—of uncertainty is what one can expect from measurement.

Measurement is only a probabilistic exercise.

The fact that some (amount of) error is **unavoidable** but the result can still be an improvement on prior knowledge is central to how experiments, surveys, and other scientific measurement are performed.

> ***Measurement***: A quantitatively expressed reduction of uncertainty based on one or more observations.

Measurements that meet basic standards of scientific validity report results with some specified degree of uncertainty like,

> "There is a 90% chance that an attack would cause this system to be down somewhere between 1 and 8 hours."

Ultimately, a measurement is *just information*, and there is a rigorous theoretical foundation for **information theory**, developed by Claude Shannon in the 1940s: "A Mathematical Theory of Communication". (Source: Shannon, 1948)

Shannon proposed a mathematical definition of *information* as the amount of **uncertainty reduction** in a signal measured in terms of the *entropy*[3] removed by a signal.

[3] Generally, lack of order or predictability; in information theory, a logarithmic measure of the rate of transfer of information in a particular signal or message.

Reprinted with corrections from *The Bell System Technical Journal,*
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

Source: http://math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf

Ultimately, a measurement is *just information*, and there is a rigorous theoretical foundation for **information theory**, developed by Claude Shannon in the 1940s: "A Mathematical Theory of Communication".

Shannon proposed a mathematical definition of *information* as the amount of **uncertainty reduction** in a signal measured in terms of the *entropy* removed by a signal.

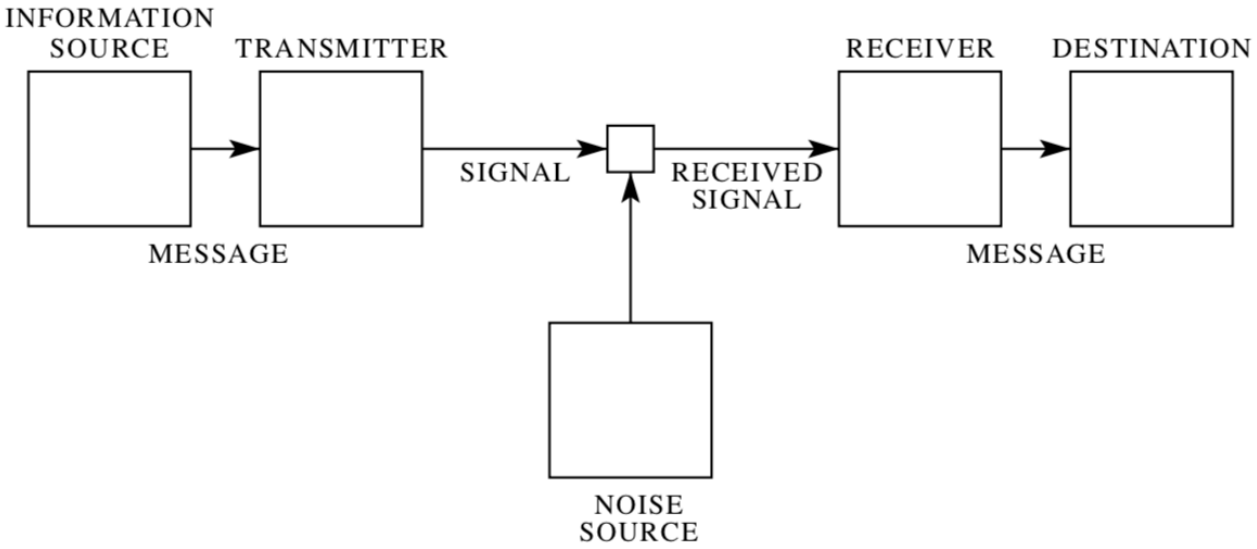


Fig. 1 — Schematic diagram of a general communication system.

(Source: Shannon, 1948)

Shannon, 1948:

> *The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.*
>
> *Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These* ***semantic aspects of communication are irrelevant*** *to the engineering problem.*
>
> *The significant aspect is that the* ***actual message is one selected from a set of possible messages****. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.*
>
> *If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as* ***a measure of the information produced*** *when one message is chosen from the set, all choices being equally likely. …the most natural choice is the logarithmic function.*

Assuming that the receiver of information has some **prior state of knowledg**e but has also some uncertainty—meaning, the knowledge is incomplete—, the new information reduces the uncertainty (not necessarily completely) in a quantifiable way.

> This *uncertain reduction point of view* is what is critical to business. Major decisions made under a state of uncertainty, such as whether to approve large IT projects or new security controls, can be made better—even if just slightly—by reducing uncertainty. Sometimes even small uncertainty reductions can be worth millions of dollars.
>
> (Hubbard and Seiersen, 2016)

Measurement Scales Taxonomy

Measuring cybersecurity is like any other type of measurement in the sense that it does **not require certainty**.

Any notion of measurement that presumes measurements are exact quantities ignores the usefulness of simply reducing uncertainty, whenever eliminating uncertainty is not possible or not economical.

When making big (risky) decisions under uncertainty, e.g. $$$$ business investments, uncertainty reduction has considerable value.

Uncertainty reduction implies a prior state of uncertainty to be reduced:

> "Probability" refers to the state of uncertainty (or *degree of believe*) of an observer.
>
> > E.g., there is an *X*% probability that we will file a $*Y, or more,* insurance claim within the next 18 months.

This view of probabilities is called the *subjectivist* or *Bayesian interpretation*.

In probability theory and statistics, Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event (a.k.a. ***prior***):

> Given a ***hypothesis A*** and ***evidence B***, Bayes' theorem states the relationship between the probability of the hypothesis *P*(*A*) prior to getting the evidence and the probability *P*(*A* | *B*) of the hypothesis after getting the evidence:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}, \text{ and } P(B) \neq 0.$$

- *P*(*A*), the *prior*, is the initial degree of belief in *A*.
- *P*(*A* | *B*), the *posterior* is the degree of belief having accounted for *B*.
- the quotient $\frac{P(B|A)}{P(B)}$ represents the support *B* provides for *A*.

Probability Theory

Does probability measure the real, physical tendency of something to occur or is it a measure of how strongly one believes it will occur, or does it draw on both these elements?

- In answering such questions, mathematicians **interpret the probability values** of probability theory.
- There are two broad categories of probability interpretations which can be called *physical probabilities* and *evidential probabilities*.

Common interpretations:

- Classical (Principle of indifference)
- Frequentist (Frequency of occurrence)
- Subjective (Degree of belief)
- Propensity (Degree of causal connection)

Physical probabilities

This category, also called objective or frequency probabilities, is associated with random physical systems such as roulette wheels, rolling dice and radioactive atoms. In such systems, a given type of event tends to occur at a persistent rate, or "relative frequency", in a long run of trials.

Classical interpretation

Developed from studies of games of chance (such as rolling dice) it states that probability is shared equally between all the possible outcomes, provided these outcomes can be deemed equally likely.

- If a random experiment can result in N mutually exclusive and equally likely outcomes and if N_A of these outcomes result in the occurrence of event A, the probability of A is defined by

$$P(A) = \frac{N_A}{N}.$$

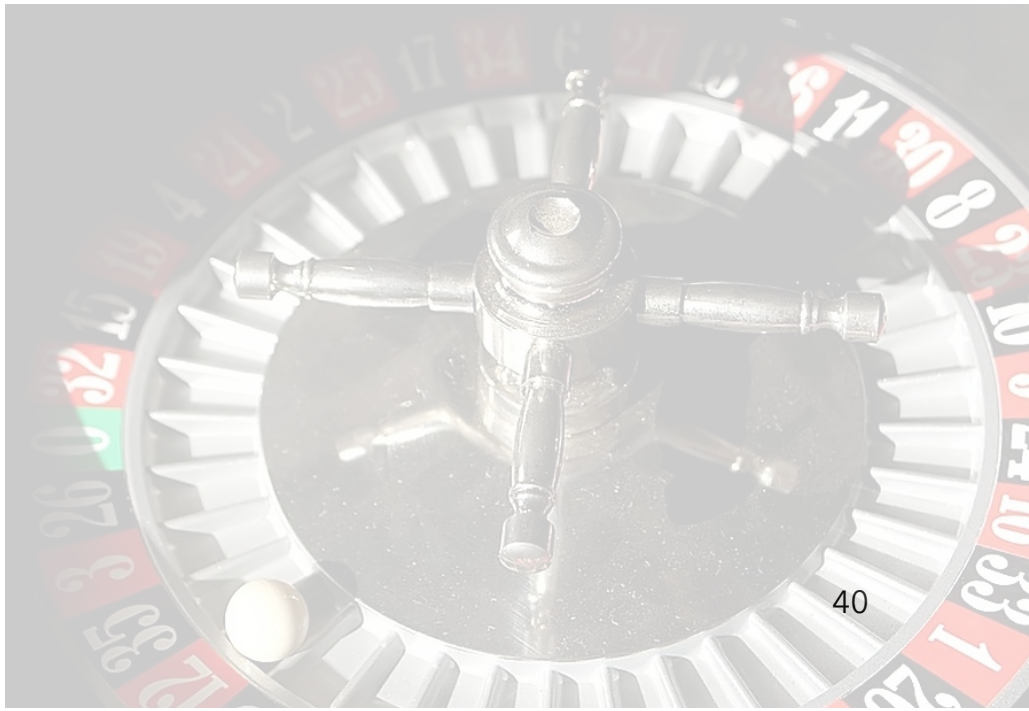
- Based on hypothetical symmetry
- Applicable only to situations in which there is only a **finite number** of possible outcomes

Propensity interpretation

An interpretation as a physical propensity, or disposition, or tendency of a given type of physical situation to yield an outcome of a certain kind, or to yield a long run relative frequency of such an outcome.

- Propensities are invoked to explain why repeating a certain kind of experiment will generate a given outcome type at a persistent rate.
- A central aspect of this explanation is the law of large numbers (LLN), which is a consequence of the axioms of probability.
- *Example:* Consider a coin is tossed repeatedly many times, in such a way that its probability of landing heads is the same on each toss, and the outcomes are probabilistically independent, then the relative frequency of heads will be close to the probability of heads on each single toss.

- LLN: a theorem that describes the result of performing the same experiment a large number of times;
- the average of the results obtained from a large number of trials should be close to the expected value, and tends to become closer as more trials are performed.



Bayesian (or subjective[4]) interpretation

An interpretation of the concept of probability, in which (instead of frequency or propensity of some phenomenon) probability is interpreted as reasonable expectation representing a state of knowledge or as quantification of a personal belief

- This interpretation can be seen as an extension of **propositional logic** that enables reasoning with hypotheses, that is propositions whose truth or falsity is uncertain.
  In the Bayesian view, a probability is assigned to a hypothesis.
- Bayesian probability belongs to the category of *evidential probabilities;* to evaluate the probability of a hypothesis, one specifies some prior probability, which is then updated to a posterior probability in the light of new evidence obtained.

Bayesian methods are characterized by concepts and procedures as follows:

- The use of random variables, or more generally *unknown quantities*, to model all sources of uncertainty in statistical models including uncertainty resulting from lack of information.
- The need to determine the prior probability distribution taking into account the available (prior) information.
- The sequential use of Bayes' formula: when more data become available, calculate the posterior distribution using Bayes' formula; subsequently, the posterior distribution becomes the next prior.

---

[4] a measure of the **degree of belief** of an individual assessing the uncertainty of a particular situation

Stating **priors** allows to compute the *value of additional information* since this value (at least partly) depends on the state of uncertainty before obtaining this information.

A fundamental irony is when someone in cybersecurity says they lack the data to assign probabilities. After all, we use probability ***because*** we lack perfect information, not *in spite* of it.

Important

- While the state of uncertainty—or degree of believe—may reflect a **subjective estimate**, this estimate needs to be based on some "process" that is mathematically coherent and consistent.[5]
- This condition implies that we can reject subjective uncertainties on objective grounds if someone's predictions are constantly wrong (or highly unreliable).

[5] For a hypothesis to be a **scientific hypothesis**, the scientific method requires that one can test it.

2. Object of Measurement

Even when the more useful concept of measurement—*as uncertainty reducing observations*—is adopted, some things seem immeasurable because we simply don't know what exactly (unambiguously) we mean when we first consider the ***object*** of measurement.

Identifying the objects of "measurement" really is the beginning of almost any scientific inquiry.

"Cybersecurity experts and executives need to realize that some things seemed intangible only because they have been poorly defined. Avoidably vague terms like

- *damage to reputation*
- *customer confidence*
- *threat capability*
- *cyber threat*

seem immeasurable at first, perhaps, only because what they mean is not well understood.

These terms may actually represent a list of distinct and observable phenomena that need to be identified in order to be understood." (Hubbard and Seiersen, 2016)

Once we figure out **what someone means and why it matters**, the issue in question starts to look a lot more measurable—usually this is the first level of analysis.

Basic Definitions

Uncertainty and Risk, and their Measurements

- **Uncertainty:** Lack of complete certainty, that is the existence of more than one possibility—the true outcome is not known.

  Measurement of Uncertainty: a set of probabilities assigned to a set of possibilities.

  E.g., there is a 35% chance we will have a serious data breach within the next five years.

- **Risk:** A state of uncertainty where some of the possibilities involve a loss, catastrophe, or other undesirable outcome.

  Measurement of Risk: a set of possibilities, each with quantified probabilities and quantified losses.

  E.g., there is an 8% chance that a data breach will result in a legal liability exceeding $10M.

3. The Methods of Measurement

It's not what you don't know that will hurt you,
it's what you know that ain't so.

—Mark Twain

Direct vs. indirect forms of measurement, depending on the available access to the entire object of measurement:

- Measuring downtime of a system or dollar amounts spent on security upgrades of IT assets is straightforward; there is no 'obscured' portion of a population to assess.

- Many situations are more challenging and involve populations that are too large or dynamic to see all at once, calling for measurement methods that involve indirect deductions and inferences.

This case definitely applies to cybersecurity, where we often need to **infer something unseen from something we can see**.

- "Cybersecurity is not some exceptional area outside the domain of statistics but rather exactly the kind of problems statistics was made for. …

- Cybersecurity experts may believe they correctly recall and understand enough about statistics and probability so that they can make confident declarations about what inferences can be made from some data without attempting any math.

Unfortunately, their "mental math" is often not at all close to correct. There are misconceptions about the methods of measurement that get in the way of assessing risk in many fields, including cybersecurity." (Hubbard and Seiersen, 2016)

> "I don't have any tolerance for risk at all because I never take risks."
> —The words of a midlevel manager at an insurance company.*

*Douglas Hubbard, How to Measure Anything (Third Edition), Wiley, 2014.

Small samples tell you more than you think

Someone saying *"We don't have enough data to measure this,"* makes a very specific mathematical claim—often without being aware of it and without providing actual math to support it.

Example

Statistics actually helps us make some informative inferences from surprisingly small samples:

> Consider a random sample of just *five of anything*—assuming some population on which a linear order is defined (e.g., numbers). What is the chance that the **median** of the entire population is between the largest and smallest of that sample of five?

Small samples tell you more than you think (continued)

Someone saying *"We don't have enough data to measure this,"* makes a very specific mathematical claim—often without being aware of it and without providing actual math to support it.

Example

Statistics actually helps us make some informative inferences from surprisingly small samples:

> Consider a random sample of just *five of anything*—assuming some population on which a linear order is defined (like numbers). What is the chance that the **median** of the entire population is between the largest and smallest of that sample of five?
>
> With a sample this small, the range might be very wide, but if it is any narrower than your previous range, then it **counts as a measurement** according to our previous definition.

Small samples tell you more than you think (continued)

Solution

- Assume you randomly pick five values that were all above or below the median. Then the median would be outside of the range of the samples considered here.
- What is the chance this happens in a random experiment?

Small samples tell you more than you think (continued)

Solution

- Assume you randomly pick five values that were all above or below the median. Then the median would be outside of the range of the samples considered here.

- What is the chance this happens in a random experiment?

  The chance for randomly picking a value above the median is 50%.

  Now, doing this five times in a row has a chance of 1 in 32, or 3.125%.

  Thus, the chance that five randomly picked values are either **all above** or **all below** the median is calculated as follows: 100% – 3.125% x 2, or 93.75%

Small samples tell you more than you think (continued)

Solution

- Assume you randomly pick five values that were all above or below the median. Then the median would be outside of the range of the samples considered here.

- What is the chance this happens in a random experiment?

  The chance for randomly picking a value above the median is 50%.

  Now, doing this five times in a row has a chance of 1 in 32, or 3.125%.

  Thus, the chance that five randomly picked values are either **all above** or **all below** the median is calculated as follows: 100% – 3.125% x 2, or 93.75%

Rule of Five

There is a 93.75% chance that the median of a population is between the smallest and the largest values in any random sample of five from that population.

Small samples tell you more than you think (continued)

Solution

- Assume you randomly pick five values that were all above or below the median. Then the median would be outside of the range of the samples considered here.

- What is the chance this happens in a random experiment?

The chance for randomly picking a value above the median is 50%.

Now, doing this five times in a row has a chance of 1 in 32, or 3.125%

Thus, the chance that five randomly picked values are either **all a**
the median is calculated as follows: 100% – 3.125% x 2, or 93.75%

Cybersecurity is now such a critical endeavour that even small reductions in uncertainty can be extremely valuable.

Rule of Five

There is a 93.75% chance that the median of a population is between the small
and the largest values in any random sample of five from that population.

A Practical Quantitative Method for Cybersecurity

We start with a simple model to replace the common risk matrix by providing a way to capture *subjective estimates* of likelihood and impact probabilistically.

- You may think of this model as a simple one-for-one substitution—meaning, we still **depend on** the subjective judgement of cybersecurity experts about likelihood and impact.

- Thus, the proposed method is really just another way of expressing a *perceived current state of uncertainty*; and does **not reflect** a proper "measurement" that would have further reduced any prior uncertainty based on additional observations.

So what's the point then?

A Practical Quantitative Method for Cybersecurity

We start with a simple model to replace the common risk matrix by providing a way to capture *subjective estimates* of likelihood and impact probabilistically.

- You may think of this model as a simple one-for-one substitution—meaning, we still **depend on** the subjective judgement of cybersecurity experts about likelihood and impact.

- Thus, the proposed method is really just another way of expressing a *perceived current state of uncertainty*; and does **not reflect** a proper "measurement" that would have further reduced any prior uncertainty based on additional observations.

So what's the point then?

- We express uncertainty in a way that allows us to unambiguously communicate risk and systematically update *prior uncertainty* with new information.
- Capturing our current state of uncertainty is an important starting point in any measurement problem.

Modelling framework

The basic methodical framework of this method comprises the following steps (Hubbard and Seiersen, 2016):

1. Define a LIST OF RISKS to be analyzed.
2. Define a specific **time period** over which a risk event could materialize.
3. For each risk, assign a **probability** that the stated event will occur within the specified time period.
4. For each risk, assign a **range for a monetary loss** if such an event occurs as a *90% confidence interval (CI)*–i.e., a range wide enough that one is 90% certain that the actual loss is within the stated range.
5. Get estimates from **multiple experts**, where possible. In case of deviating estimates (not uncommon), simply average their responses.

Measure and improve your skills at assessing probabilities: *calibrated probability assessment*.

Managing Uncertainties

Relying on value ranges to represent uncertainties rather than on unrealistically precise point values clearly has advantages. By using ranges and probabilities one avoids to assume anything *one really doesn't know for a fact*. But it also means one has to use probabilistic modelling methods to "do the math."

> How to perform arithmetic operations (say, in a spreadsheet) if one has no exact values but only ranges?

- Use the ***Monte Carlo simulation method*** (introduced in the 1940s for the Manhattan Project).

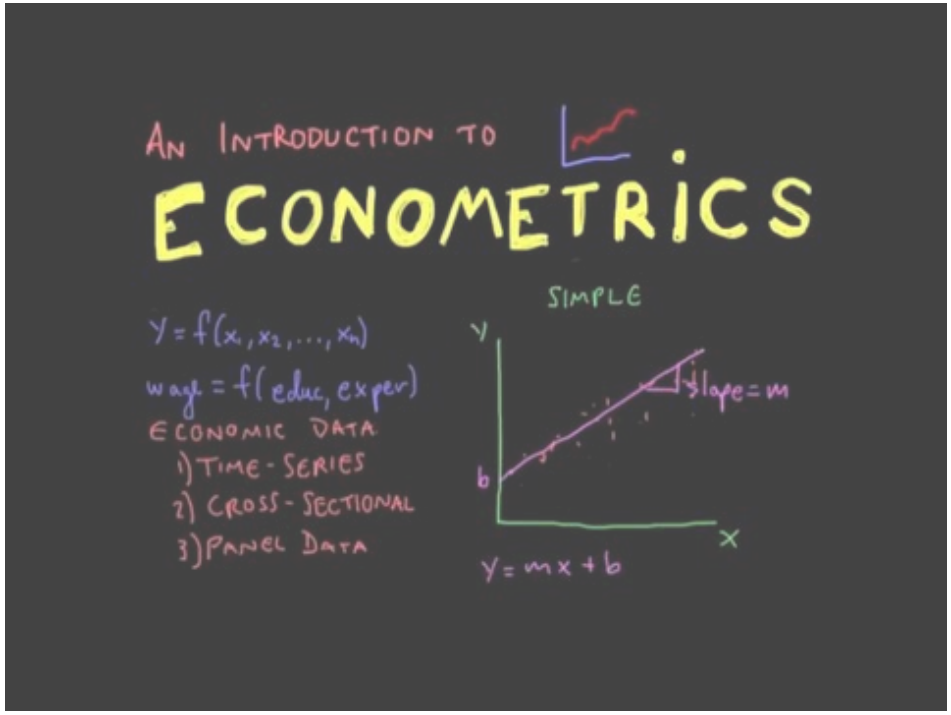
- This method involves the use of random sampling techniques and normally computer simulation to obtain approximate solutions to mathematical or physical problems, especially in terms of **a range of values**, each of which has a calculated probability.

- Monte Carlo simulations have been used to explore risk models for power plants, supply chains, insurance, …, and cybersecurity.

- A Monte Carlo simulation generates a large number of scenarios based on probabilities for input variables. By simulating (easily thousands) of scenarios for every risk, one can determine the chances that an event occurs for each type of risk and, if so, what the impact will be.

  Intro to Monte Carlo simulation

  https://www.youtube.com/watch?v=5nM5e2_1QQ0



... the branch of economics concerned with the use of mathematical methods (especially statistics) in describing economic systems.

Population Process

A *population process* is a Markov chain in which the state of the chain is analogous to the **number of individuals** in a population (0, 1, 2, etc.), and changes to the state are analogous to the addition or removal of individuals from the population.

Beyond studying biological populations in ***population dynamics***[6], population processes also apply to a much wider range of application fields, including (but not limited to):

- telecommunications
- queueing theory
- chemical kinetics
- financial mathematics

Thus, 'population' could comprise packets in a computer network, molecules in a chemical reaction, or even units in a financial index.

Biological populations are typically characterized by the basic demographic processes and broad environmental effects to which a population is subject.

---

[6] … the branch of life sciences that studies the size and age composition of populations as **dynamical systems**, and the biological and environmental processes driving them (such as birth and death rates, and by immigration and emigration). Example scenarios are ageing populations, population growth, or population decline.

Estimator

In statistics, an *estimator* is a rule for calculating an estimate of a given quantity **based on observed data**: thus the rule (the estimator), the quantity of interest (the estimand) and its result (the estimate) are distinguished.

An estimator is used to infer the value of an *unknown parameter* in a statistical model.

Suppose there is a fixed parameter θ that needs to be estimated. Then an estimator is a function that maps the sample space to a set of sample estimates. An estimator of θ is usually denoted by the symbol $\hat{\theta}$.

Estimator

In statistics, an *estimator* is a rule for calculating an estimate of a given quantity **based on observed data**: thus the rule (the estimator), the quantity of interest (the estimand) and its result (the estimate) are distinguished.

An estimator is used to infer the value of an *unknown parameter* in a statistical model.

Suppose there is a fixed parameter θ that needs to be estimated. Then an estimator is a function that maps the sample space to a set of sample estimates. An estimator of θ is usually denoted by the symbol $\hat{\theta}$.

Mean Squared Error

> The *mean squared error* of $\hat{\theta}$ is defined as the expected value (probability-weighted average, over all samples) of the squared errors; that is,
>
> $$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}[(\hat{\theta}(X) - \theta)^2].$$

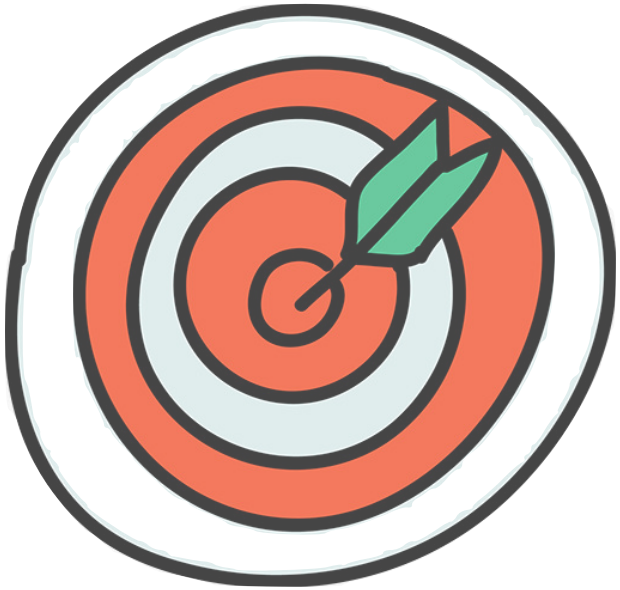
The mean square error indicates how far, on average, the collection of estimates are from the single parameter being estimated.

Example

Suppose we are shooting arrows at a target:

- The parameter is the bull's-eye of a target.
- The estimator is the process of shooting arrows at the target.
- The individual arrows are estimates (samples).

Then high MSE means the average distance of the arrows from the bull's-eye is high, and low MSE means the average distance from the bull's-eye is low.



Visualizing Risk

Compared to the risk matrix approach, the proposed method replaces the likelihood scale with an explicit probability, and the expected impact with a 90% CI representing a **range of potential losses**.

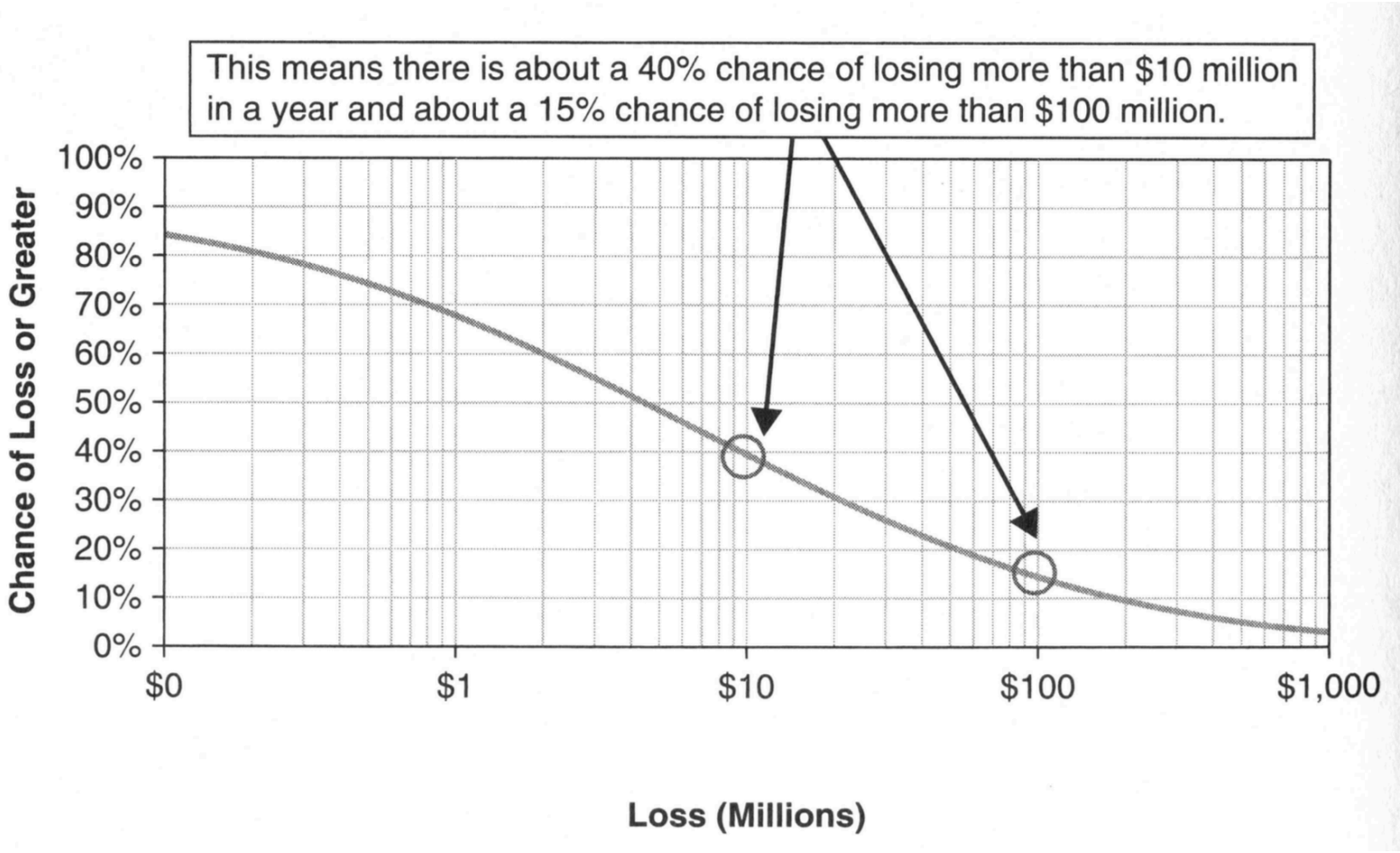
Thus, the impact is shown by more than a single point in a graphical representation.

- If an event that has a 5% chance of occurrence, we can't just say the impact will be exactly, say $10 million.
- There is really a 5% chance of loosing *something* at all, and perhaps a 2% chance of loosing more than $5 million, a 1% chance of loosing $10 million, et cetera.

We can represent the range of expected losses with a chart called a *loss exceedance curve, or LEC.*

- This concept is also used in financial portfolio risk assessment, actuarial science, and probabilistic risk assessment in nuclear power and other areas of engineering.

Example of a Loss Exceedance Curve - LEC



(Source: Hubbard and Seiersen, 2016)

Example of a Loss Exceedance Curve - LEC (continued)

A variation of the LEC is obtained by adding additional curves for *inherent risk*, *residual risk*, and *risk tolerance*:

- Inherent vs. residual risk is a common distinction in cybersecurity risk assessment to represent risks **prior to the application of proposed controls**—methods for mitigating risks—and risks **after the application of controls**.
- Inherent risks normally does not mean a complete lack of controls (not a viable alternative) but rather refers to basic and minimal controls that would be considered negligent to exclude. Typical examples include: password protection, firewalls, some frequency of applying security software updates (patches), et cetera.

Example of a Loss Exceedance Curve - LEC (continued)

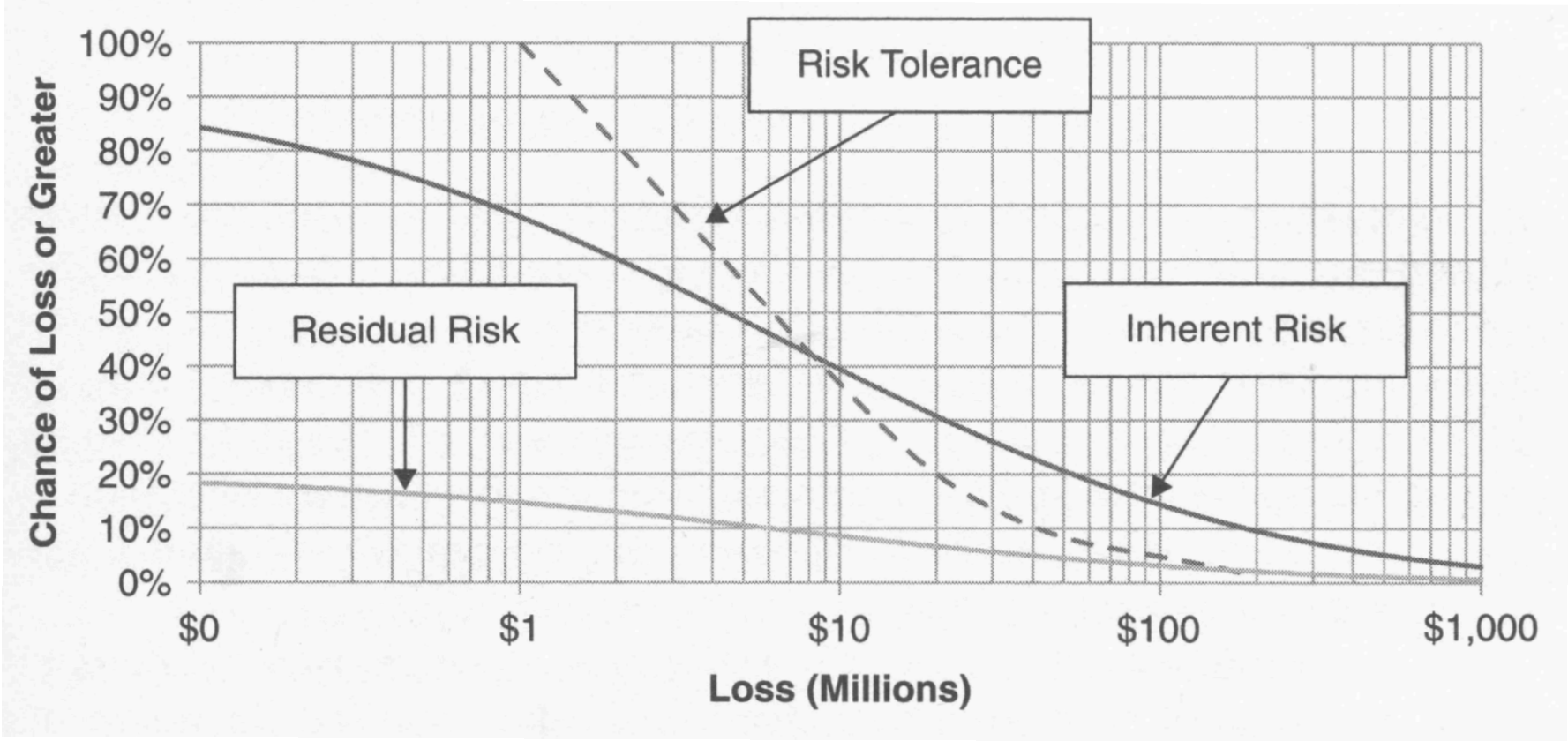
A variation of the LEC is obtained by adding additional curves for *inherent risk*, *residual risk*, and *risk tolerance*:

- Inherent vs. residual risk is a common distinction in cybersecurity risk assessment to represent risks **prior to the application of proposed controls**—methods for mitigating risks—and risks **after the application of controls**.
- Inherent risks normally does not mean a complete lack of controls (not a viable alternative) but rather refers to basic and minimal controls that would be considered negligent to exclude. Typical examples include: password protection, firewalls, some frequency of applying security software updates (patches), et cetera.
- The difference between inherent and residual risks are true discretionary controls that could be considered not absolutely necessary and may be excluded on reasonable grounds.

The LEC provides a simple and useful visual method for comparing a risk to a risk tolerance, expressed unambiguously and quantitatively.

In the example chart below, the part of the inherent risk curve **above** the risk tolerance curve *violates* or *breaks* the **risk tolerance**.

The residual risk curve is on or below the risk tolerance curve at all points; it therefore ***stochastically dominates*** the residual risk curve, meaning that the residual risks are acceptable.



(Source: Hubbard and Seiersen, 2016)

Supporting Decisions: Return on Risk Mitigation

Ultimately, the *point of risk analysis* is to support decisions on specific resource-allocation choices for specific controls.

- What is it worth to move one "high" risk to a "medium"?
- What if we have a budget of $8 million for cybersecurity investments to mitigate 80 lows, 30 mediums and 15 highs?
- What if we can mitigate more lows for the same investment as one medium?

What the CISO needs is a "Return on Control" calculation, i.e. monetized value of the reduction in expected losses divided by the cost of the control.

For any given year, we can show this as expressed by the following formula:

$$\text{Return on control} = \frac{\text{Reduction in expected losses}}{\text{Cost of control}} - 1$$

Sources

Hubbard and Seiersen, 2016

D. W. Hubbard and R. Seiersen. How to Measure Anything in Cybersecurity Risk. John Wiley & Sons, New Jersey, 2016.

Shannon, 1948

C. E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656, July, October, 1948.