Hindawi
Security and Communication Networks
Volume 2019, Article ID 6716918, 8 pages
https://doi.org/10.1155/2019/6716918

Research Article

A Data-Driven Approach to Cyber Risk Assessment

Paolo Santini,[1] Giuseppe Gottardi,[2] Marco Baldi,[1] and Franco Chiaraluce[1]

[1]*Università Politecnica delle Marche, Ancona, Italy*
[2]*Fondazione F3RM1, Milan, Italy*

Correspondence should be addressed to Marco Baldi; m.baldi@univpm.it

Received 5 April 2019; Revised 2 July 2019; Accepted 11 August 2019; Published 9 September 2019

Guest Editor: Hyoungshick Kim

Cyber risk assessment requires defined and objective methodologies; otherwise, its results cannot be considered reliable. The lack of quantitative data can be dangerous: if the assessment is entirely qualitative, subjectivity will loom large in the process. Too much subjectivity in the risk assessment process can weaken the credibility of the assessment results and compromise risk management programs. On the other hand, obtaining a sufficiently large amount of quantitative data allowing reliable extrapolations and previsions is often hard or even unfeasible. In this paper, we propose and study a quantitative methodology to assess a potential annualized economic loss risk of a company. In particular, our approach only relies on aggregated empirical data, which can be obtained from several sources. We also describe how the method can be applied to real companies, in order to customize the initial data and obtain reliable and specific risk assessments.

1. Introduction

The process of risk assessment and treatment is fundamental to the implementation of an effective cyber security program and plays a crucial role for the national and international regulations in the field of data protection. A complete understanding of cyber risks is necessary, in order to ensure that the security controls an organization has in place are sufficient to provide an appropriate level of protection against cyber threats.

However, defining reliable models for the cyber risk exposure is still an open problem. Existing models [1–4] suffer from some important concerns that, for example, prevent the insurability market development [5]. First of all, cyber risk evaluation and the study of its related impact are performed mostly in qualitative ways, which are usually affected by errors and misrepresentations of the risk. They also exhibit several disadvantages, such as the approximate nature of the achieved results and the difficulty of performing a cost-benefits analysis [6]. Quantitative approaches, in their turn, are usually based on scoring systems that associate a certain score to a technological/organizational context. The idea is commendable but, as it will be clarified afterward, the way in which it is commonly implemented does not give a realistic measure of the cyber risk and the related impact. Reliable models for the measure of the cyber risk are not available or have significant limitations, like the lack of generalization and the fact that most works consider only the analysis of past data to derive probabilistic models, while it is not clear how to obtain reliable estimates about future events [7, 8]. Moreover, some quantitative approaches, like the well-known HTMA (how to measure anything) [9] and the FAIR [10] methods rely on a subjective evaluation of the likelihood of an event (in particular, of the probability of a successful attack due to a certain threat) given by a team of experts [11, 12]. These kinds of probabilities usually show some level of inaccuracy and should be replaced by more objective models. The impact of the set of considered threats is then measured in terms of economic loss, which is also subjectively estimated. Based on these premises, the need to improve the quantitative evaluation of the cyber risk of an organization, through a dynamic monitoring of the attacks and vulnerabilities the organization is experiencing, clearly emerges.

A recent study, for the case of data breaches in IT and information security, has been developed in [13]. However, it is clear that the whole panorama involves other sectors (for

instance, business and finance companies) and other cyber threats. This means that in order to obtain reliable representations of this wide scenario, more general analyses are required. A step towards the generalization of this work has been done in [14], where the authors consider the whole range of cyber risks and associated costs, by analyzing sufficiently large and available datasets. This kind of approach is not new in the literature (see, for instance, [8, 15] for a study based on data breaches): through the analysis of a large amount of historical data, probabilistic models about the registered cyber crime events can be derived. However, as the authors in these papers clearly state, such an approach has some important limitations. Indeed, cyber crime is a dynamic and continuously evolving phenomenon: considering only past data is clearly not enough to obtain reliable estimates about future events.

For instance, let us suppose that a company wants to determine its risk exposure and eventually define possible strategies to increase its security against cyber crime. In such a case, relying on probabilistic models is very likely to be not enough, since specific countermeasures and consequent attack strategies should be taken into account. One example of this process is given in [12], where the authors consider a dynamic model that describes possible interactions between a defender and an attacker. In such a model, each countermeasure is considered as a software update of an existing cyber security system and is characterized by an effectiveness score; the optimal strategy is thus obtained by taking into account the interplay between the defender and the attacker.

Previous works highlight the fact that the whole scenario of cyber risk, with the heterogeneous ensemble of all possible players and events, looks like a phenomenon that difficultly allows deriving reliable probabilistic models. In particular, applications on some real case studies might strongly depend on specific aspects of the involved subjects. For instance, conducting a risk assessment with a qualitative approach may result in a good level of controls for malware defense, but this does not provide evidence on whether these security measures are effective in counteracting malware attacks. As a consequence, inaccurate prioritization could result in valuable resources being spent on risk areas that may not be very important and which may not deserve such resources and vice versa.

As mentioned above, one way to reduce the uncertainty in this scenario is the one of relying on opinions of experts [11, 12]. This kind of approach is followed in HTMA [9], where a set of threats is characterized by a likelihood value and the corresponding impact. Basically, the likelihood is the probability of successful attack due to each threat, while the impact expresses the subsequent economic loss. In particular, in the HTMA method, the impacts are estimated through interviews to experts that, for each threat, are asked with the 90% confidence range for possible economic losses. Their answers are then used to define the random variable associated to the impact; a log-normal distribution is assumed for each one of the considered threats.

In this paper, we use the HTMA approach but assuming experts' opinions only as a starting point to be progressively and continuously improved through the acquisition of new and updated information on the organization's behavior against cyber threats. More precisely, we propose to exploit a combination of probabilistic techniques and objective data as an input to HTMA. Our goal is to define a methodology for fitting a probabilistic model into a real case study. The procedure we propose is substantially based on checking the effectiveness of the applied measures through a data-driven approach. As previous studies by SANS Institute already outline [16], one of the key focuses of an effective cyber risk assessment is the measurement of the security controls implementation effectiveness. Indeed, as seen in most risk assessment methods, a risk matrix only representing impact and likelihood is a commonly used tool to assess cyber risk. According to previous approaches, these values are obtained through the analysis of a sufficiently large amount of historical data. However, these values hardly reflect the actual state of a particular entity under analysis. Indeed, a reliable assessment of the risk exposure without taking into consideration the effectiveness of the applied measures is a hard task. For this purpose, our analysis considers the CIS 20 Critical Security Controls (Center for Internet Security, https://www.cisecurity.org/), but the same procedure can be applied to any other suitable set of security issues.

The paper is organized as follows. In Section 2, we briefly remind the probabilistic model we consider, based on the HTMA approach. In Section 3, we describe how empirical data can be used to fit the model into an actual entity (for instance, a company). In Section 4, we show an application of this methodology to some real case scenarios. Finally, in Section 5, we draw some conclusive remarks.

2. Probabilistic Model

In this section, we shortly describe the method we use for evaluating the cyber risk exposure of a company. We first present a generic model that fits into our study case and give the basic notions that are fundamental for our analysis. We then briefly introduce the HTMA methodology and describe how it can be used to derive a quantitative measure for the risk exposure.

2.1. General Model. We consider a set of events $\mathbf{I}_E = \{E_1, \ldots, E_n\}$, the i-th one with probability of occurrence p_i, for $i = 1, \ldots, n$; in this paper, we assume that the events are independent, which means that the occurrence of E_i does not influence the occurrence of all the other events E_j, with $j \neq i$. More complex scenarios, in the presence of correlation between different events, will be analyzed in future work. Each event E_i is also linked to a random variable c_i, which is associated to the impact and is described through a probability distribution $f^{(i)}(c_i)$: each time the event E_i occurs, it has an impact c_i, whose particular value depends on $f^{(i)}(c_i)$. Then, we denote as $C(\mathbf{I}_E)$ the random variable defined as the sum of the impacts of the considered events. Our goal is to characterize the statistical properties of $C(\mathbf{I}_E)$.

Actually, unless specific choices for the distributions of c_i are made, providing a closed form for the distribution of

$C(\mathbf{I}_E)$ is unfeasible. However, we can proceed with numerical simulations and estimate its cumulative distribution function (CDF). More precisely, we can simulate N different scenarios: in each scenario, we simulate the occurrences of the events E_i, and for each occurred event, we randomly extract the corresponding impact, according to $f^{(i)}(c_i)$. Let $C^{(j)}$ be the resulting total impact obtained in the j-th simulated scenario; then, the CDF of $C(\mathbf{I}_E)$ can be estimated as

$$F_{C(\mathbf{I}_E)}(a) = P\{C(\mathbf{I}_E) \leq a\} = \frac{\#_j\{C^{(j)} \leq a\}}{N}. \quad (1)$$

It is clear that, in order to obtain reliable estimates via numerical simulations, the value of N must be sufficiently large.

One crucial quantity in our analysis is the *loss exceedance curve* defined as

$$L_{C(\mathbf{I}_E)}(a) = P\{C(\mathbf{I}_E) \geq a\} = 1 - F_{C(\mathbf{I}_E)}(a) + P\{C(\mathbf{I}_E) = a\}. \quad (2)$$

By definition, the value of $L_{C(\mathbf{I}_E)}(a)$ corresponds to the probability that the total impact of the considered events is equal to or exceeds a threshold value a. If, for instance, we have $L_{C(\mathbf{I}_E)}(a) = 0.2$, then this means that the probability that the set of considered events results in a total impact $\geq a$ is equal to 0.2. Also by definition, we have $L_{C(\mathbf{I}_E)}(0) = 1$ and $L_{C(\mathbf{I}_E)}(\infty) = 0$.

2.2. Cyber Risk Assessment Based on HTMA. In this section, we briefly describe the use of the HTMA method for assessing cyber risk. We consider a set of n different cyber threats and, with reference to the notation introduced in the previous section, define E_i as the event that the i-th threat has resulted in an economic loss for the company. Thus, p_i is the probability of such an occurrence, for instance, in the time interval of one year.

Following the HTMA methodology, each economic impact is obtained on the basis of interviews: in particular, for each threat, experts are asked with the 90% confidence interval of economic losses that their company might sustain, in case of occurrence of the considered threat. The impact of each cyber threat is then associated to a range in the form $[c_i^{(\min)}; c_i^{(\max)}]$, corresponding to the 90% confidence interval. In particular, as mentioned, the HTMA method assumes that each economic impact follows a log-normal distribution, with mean μ_i and standard deviation σ_i that are obtained as

$$\mu_i = \frac{\log\left(c_i^{(\max)}\right) + \log\left(c_i^{(\min)}\right)}{2}, \quad (3)$$

$$\sigma_i = \frac{\log\left(c_i^{(\max)}\right) - \log\left(c_i^{(\min)}\right)}{3.29}, \quad (4)$$

respectively. Thus, starting from likelihood values and the corresponding economic impacts distribution, Monte Carlo simulations can be performed to obtain the loss exceedance curve $L_{C(\mathbf{I}_E)}(a)$.

In each simulated scenario, we consider all possible (i.e., identified) threats. For each threat, we randomly sample a variable with continuous uniform distribution in $[0; 1]$. If such value exceeds p_i, then the corresponding impact is sampled (according to the corresponding log-normal distribution); otherwise, it is set as 0. Then, by computing the sum of all impacts, we obtain the value of the total impact. If we consider a sufficiently large number of scenarios, then we can obtain a reliable estimate of the loss exceedance curve.

3. Cyber Risk Assessment Framework

The loss exceedance curve introduced in the previous section strongly depends on the company assets and characteristics. In other words, in order to obtain a reliable estimate of the loss exceedance curve, we need to customize the values p_i and c_i to the actual organization we are considering. This operation is usually performed through surveys submitted to the company, in order to obtain a reliable overview of what the company is currently doing to prevent cyber threats. However, it is clear that the use of such answers is likely not enough to obtain a complete and accurate picture of the actual state of the company. As an example, the implementation of a particular strategy (e.g., having good anti-malware tools) does not mean that the strategy is indeed effective, since its effectiveness is influenced by many other factors.

In other words, if the stakeholders' answers are evaluated with a more objective process, this will result in a more objective profiling of the company. This is our aim, which we pursue by introducing the use of some data-driven key risk indicators (KRIs) in the HTMA approach. The model we use is described next, while its application to some practical case studies is reported in Section 4.

3.1. Data-Driven KRIs. Suppose that we can dispose of a tool that monitors the company and returns a sufficient amount of quantitative evidences such as

(1) Malicious code/software activity (i.e., malware, ransomware, botnet evidences)

(2) Insecure/unencrypted/vulnerable protocols usage (i.e., P2P, vulnerable SSL, etc.)

(3) Deep web exposure (company targeted by criminals)

(4) Data breaches due to human errors, third parties, or hacking activity

(5) Software/infrastructure vulnerabilities

Clearly, we could use such information to establish a well-defined set of KRIs that negatively influence the effectiveness of the existing controls. For instance, the detection of malware incidents originating within the network of the organization is a clear indicator that some of the employees are not aware about phishing attacks and that the anti-malware tools used by the company are not enough, even if all due controls are implemented. In the same way, vulnerabilities or unnecessary/insecure services exposed on the Internet perimeter of the company are a clear indicator that the implemented controls are not effective. In both cases, it is clear that the level of security is not as high as claimed. So, the answers of the survey should be mapped

with a list of evidences pertaining to cyber incidents and technical vulnerabilities obtained, for example, with a set of cyber intelligence sources, in order to associate an effectiveness score to each one of the given answers. This way, one can get a more realistic picture of the company cyber profile: in the case of a "bad" score, the controls implemented by the company and declared by the answers are not effective. In other words, the likelihood of the associated threats should be increased, since the probability of incurring in some related incidents is higher than expected. On the contrary, in the case of a "good" score, the stakeholders' answers can be positively weighted consequently.

More precisely, when considering the stakeholders' answers, we propose to use a mapping between all the questions and the set of considered threats: this way, we can obtain a *coverage score* that in the end is used to calibrate the initial likelihood values according to the company profile. For this purpose, we first define the *control coverage vector* $\widetilde{t} = [\widetilde{t}_1, \ldots, \widetilde{t}_n]$ as the ensemble of scores computed on the basis of the stakeholders' answers. Basically, each entry $\widetilde{t}_i$ is defined as the state of coverage against the i-th considered threat. In particular, we define a *control* as a series of atomic actions to protect the organization against internal and external threats. An example of control is "Continuous Vulnerability Management," that, as per CIS v7, could be realized carrying out operations like performing authenticated vulnerability scanning, deploying automated operating system patch management tools, etc. The more actions are performed, the higher the level of control's implementation will be. For instance, an effective measurement of this control can be achieved by taking into consideration the exposure to software and infrastructure vulnerabilities; moreover, an effective measurement of the malware defenses (CSC 8) can be achieved verifying the infections events originated within the network. Then, the stakeholders' answers can be mapped into the set of controls; in particular, the answers can be collected in a vector $\mathbf{g}$, while the controls in a vector $\mathbf{d}$. Then, let $\mathbf{W}$ be a matrix with number of rows and columns equal to, respectively, the number of answers and controls. We can write

$$\mathbf{d} = \mathbf{gW}, \tag{5}$$

where the entries of $\mathbf{W}$ are all ≤ 1. With this choice, the values in $\mathbf{d}$ correspond to weighted sums of elements of $\mathbf{g}$. In particular, the entry in position (i, j) in matrix $\mathbf{W}$, which we denote as $w_{i,j}$, expresses the impact of the i-th answer on the j-th control. These coefficients need to be normalized, in the sense that the sum of each column in $\mathbf{W}$ equals 1. If $w_{i,j} = 0$, this means that the action referred to the i-th answer has no impact on the j-th control. The entries of $\mathbf{W}$ are determined on the basis of statistical analyses and evaluations on empirical data. We point out that the whole procedure is independent of both the number of collected answers and the number of analyzed controls. Indeed, choosing a different set of answers or different controls (or both of these possibilities) will just result in different dimensions and different entries for the matrix $\mathbf{W}$ (and, obviously, different lengths for the vectors $\mathbf{g}$ and $\mathbf{d}$), while the whole procedure will remain unchanged.

For each threat E_i, we then combine the elements of $\mathbf{d}$, in order to obtain the elements of $\widetilde{t}$. Indeed, each control is associated to one or more threats: the formula we have used in our simulations is

$$\widetilde{t} = \mathbf{dK}, \tag{6}$$

where $\mathbf{K}$ is another matrix in which the entries of each column sum to 1, whose entries, analogously to those of $\mathbf{W}$, are evaluated after a statistical analysis performed on empirical data. The entry in position (i, j) in $\mathbf{K}$, which we denote as $k_{i,j}$, expresses the impact of control i on the j-th threat.

By combining (5) and (6), we obtain

$$\widetilde{t} = \mathbf{gWK}, \tag{7}$$

which shows how the entries of the control coverage vector are in linear dependence with the entries of the answer vector. It must be noticed that this matrix approach is rather classic and widely used in the literature [17, 18]. The main novelty of our analysis is in the fact that the vector $\widetilde{t}$ is not used directly but it is combined with another vector resulting from empirical data.

Actually, analogously to what was done for the control coverage vector, a set of measurements can be provided in order to obtain another length-n vector that we call *effectiveness vector* and denoted as $\widehat{t} = [\widehat{t}_1, \ldots, \widehat{t}_n]$. The vectors $\widetilde{t}$ and $\widehat{t}$ are then combined, in order to obtain the *effective coverage score*, which is a length-n vector $\mathbf{t}$ whose i-th entry is obtained as $t_i = \widetilde{t}(1 - (1/5)\widehat{t}_i)_i$. It is clear that this operation corresponds to scaling the entries of the control coverage vector on the basis of the evidences found. Finally, the customized likelihoods are obtained as $p_i' = (x p_i)^{2/3}$, where $x = \max\{1 - t_i, 0.06\}$. We point out that these expressions have been derived after evaluation of available empirical data and extrapolations of the values from [19].

On the other hand, from a practical point of view, it is extremely useful to define a global indicator that measures the general cyber security posture of a company. We call this value *Security Control Score* (SCS) and we compute it as the average value of the entries of $\mathbf{t}$. Basically, a high SCS indicates substantial and effective investments in people and technology to protect the digital assets and a low exposure to costs following from cyber threats.

4. Application to Real Case Scenarios

In this section, we describe the application of our model to some real case scenarios. Consistently with the existing literature, we first consider historical data to obtain reliable values for the likelihood and impact values of a set of cyber threats. Then, we show a real case study, by applying the procedure described in Section 3, that is, introducing the evidences resulting from the intelligence tool into the estimate.

4.1. Obtaining Initial Likelihood Values. In order to provide all the required inputs for the HTMA model, we need to define a set of cyber threats for which the likelihood and

impact values can be reliably estimated. For this purpose, we rely on data reported in [19]; this choice is motivated by the fact that this report contains a large amount of information about events in 2017, so it offers a significant and recent picture of the current cyber crime panorama. According to [19], we consider nine different threats and the corresponding likelihood values; such values are listed in Table 1. An ID number is assigned to each threat, in order to simplify the notation.

For each one of the considered threats, starting from [19], we have determined the 90% confidence ranges for the impacts, on the basis of the sector in which a company operates; such ranges are listed in Table 2. For instance, for a company operating in the industrial sector, the range of losses due to malware events goes from a minimum of 1.95 M$ to a maximum of 2.19 M$.

The distinction between operating sectors might not be enough to obtain reliable estimates of the economic losses: as [19] clearly shows, the company size is another aspect that must be taken into account. In particular, coherently with [19], we can define the company size as a function of the number of seats (i.e., number of employees). This dependence can be heuristically modelled through a coefficient α defined as follows:

$$\alpha = \begin{cases} \dfrac{as^3 + bs^2 + cs}{11.7}, & \text{if } s \leq 40000, \\ \dfrac{de^{ks} + le^{ms}}{10^6}, & \text{if } s > 40000, \end{cases} \tag{8}$$

where s is the number of seats and the values of the coefficients in (8) have been obtained on the basis of extrapolations from data contained in the report ($a = 1.04 \cdot 10^{-12}$, $b = -6.54 \cdot 10^{-8}$, $c = 1.41 \cdot 10^{-3}$, $d = 1.815 \cdot 10^{7}$, $k = 2.125 \cdot 10^{-7}$, $l = 0.5838$, and $m = 6.398 \cdot 10^{-5}$).

The coefficient α can then be used to adjust the ranges listed in Table 2. In order to clarify this aspect, let us suppose that a company operating in the industrial sector has $\alpha = 0.1$: then, all the impacts reported in Table 2 must be multiplied by a coefficient equal to 0.1. For instance, the range for malware attacks becomes $[0.195; 0.219]$: these values correspond to $c_1^{(\min)}$ and $c_1^{(\max)}$ that are used in (3) and (4).

The values computed this way can then be used as reliable values for the likelihoods and impacts that are exploited in the HTMA methodology.

4.2. Case Studies. In order to validate our approach, we have run some simulations considering three different organizations O_1, O_2, and O_3, with the following common properties:

(i) There are 2000 workstations

(ii) Their business sector is industrial/manufacturing

(iii) Their annual revenue is about 350,000,000 $

It must be said, however, that the proposed approach is quite general and can be equally applied to different sectors. What usually changes passing from one scenario to another is obviously the numerical values of the quantities involved, while the approach and the set of formal relationships at the basis of the model remain substantially unchanged.

Table 1: Considered threats with the corresponding likelihood values.

ID	Threat	Likelihood
1	Malware	0.98
2	Web-based attacks	0.67
3	Denial of services	0.53
4	Malicious insiders	0.40
5	Phishing and social eng.	0.69
6	Malicious code	0.58
7	Stolen devices	0.43
8	Ransomware	0.27
9	Botnets	0.63

An intelligence tool is supposed to be used to detect cyber evidences on the cyber perimeters of O_1, O_2, and O_3, to verify the effectiveness of the security controls implemented by them. Table 3 reports the number of evidences, of the type listed in Section 3.1, monitored for the three organizations in a precise period of time (e.g., 1 year), distinguishing them on the basis of the impact: trivial, middle, and critical. Eight cyber intelligence sources S_i, with $i = 1, \ldots, 8$, have been adopted. Suitably processed, the values in Table 3 permit to determine the effectiveness vector $\widehat{t}$. By combining it with $\widetilde{t}$, we can compute the effective coverage score, $\mathbf{t}$, and finally, the security control score, SCS. All these values are reported in Table 4. Finally, the corresponding ranges of the economic losses (90% confidence intervals) are listed in Table 5.

This is all we need to run numerical simulations, following the HTMA approach, for the three considered companies. The resulting loss of exceedance curves is shown in Figure 1. These curves have been obtained by applying the theoretical approach discussed in Section 2.2. So, according to (2), each curve represents, for the specific organization it refers to, the probability that the economic loss is equal to or greater than the values of a reported in abscissa.

As it clearly results from the figure, this company profiling might have some serious consequences in the cyber risk assessment. For instance, suppose that the maximum loss that the three companies can sustain is equal to 1 M$. We see that, for the three companies, the probabilities of exceeding this value are significantly different and go from a minimum of approximately 0.05 for the company O_3 to maximum of approximately 0.75 for the company O_1. These values might be compliant or not with the profile and expectations of the organization, and in the latter case, they should suggest the adoption of correcting actions to reduce the risk. On the other hand, the picture so obtained is provisional, since it is expected to change, getting better or worse in subsequent assessment campaigns.

5. Conclusion

In this paper, we have described a data-driven approach to assess cyber risk and associate a score to the cyber exposure of a company. Our model relies on the well-known HTMA

Table 2: Impact ranges of the considered threats, expressed in M$, for different industrial sectors.

ID	Financial	Utilities and energy	Aerospace and defense	Technology and software	Health care
1	[3.51; 3.93]	[3.29; 3.69]	[2.78; 3.11]	[2.51; 2.81]	[2.4; 2.69]
2	[2.99; 3.35]	[2.8; 3.14]	[2.36; 2.65]	[2.14; 2.4]	[2.04; 2.29]
3	[2.32; 2.61]	[2.18; 2.44]	[1.84; 2.06]	[1.66; 1.86]	[1.59; 1.78]
4	[2.1; 2.35]	[1.97; 2.21]	[1.66; 1.86]	[1.5; 1.68]	[1.44; 1.61]
5	[1.93; 2.16]	[1.81; 2.03]	[1.52; 1.71]	[1.38; 1.55]	[1.32; 1.48]
6	[1.91; 2.13]	[1.79; 2]	[1.51; 1.69]	[1.37; 1.53]	[1.31; 1.46]
7	[1.28; 1.44]	[1.2; 1.35]	[1.01; 1.14]	[0.915; 1.03]	[0.875; 0.983]
8	[0.79; 0.885]	[0.74; 0.829]	[0.625; 0.7]	[0.565; 0.633]	[0.54; 0.605]
9	[0.521; 0.585]	[0.488; 0.548]	[0.413; 0.463]	[0.373; 0.418]	[0.356; 0.4]
ID	Services	Industrial	Retail	Public sector	Transportation
1	[2.11; 2.37]	[1.95; 2.19]	[1.78; 1.99]	[1.58; 1.77]	[1.4; 1.57]
2	[1.8; 2.01]	[1.66; 1.87]	[1.51; 1.7]	[1.34; 1.51]	[1.19; 1.34]
3	[1.4; 1.57]	[1.29; 1.45]	[1.18; 1.32]	[1.04; 1.17]	[0.926; 1.04]
4	[1.26; 1.42]	[1.17; 1.31]	[1.06; 1.19]	[0.944; 1.06]	[0.838; 0.939]
5	[1.16; 1.3]	[1.07; 1.21]	[0.976; 1.1]	[0.866; 0.973]	[0.769; 0.863]
6	[1.15; 1.28]	[1.06; 1.19]	[0.968; 1.08]	[0.859; 0.959]	[0.762; 0.851]
7	[0.769; 0.864]	[0.713; 0.801]	[0.648; 0.728]	[0.575; 0.646]	[0.51; 0.573]
8	[0.475; 0.532]	[0.44; 0.493]	[0.4; 0.448]	[0.355; 0.398]	[0.315; 0.353]
9	[0.314; 0.351]	[0.29; 0.326]	[0.264; 0.296]	[0.234; 0.263]	[0.208; 0.233]
ID	Consumer products	Communications	Life science	Education	Hospitality
1	[1.4; 1.57]	[1.35; 1.52]	[1.24; 1.39]	[0.955; 1.07]	[0.955; 1.07]
2	[1.19; 1.34]	[1.15; 1.29]	[1.06; 1.19]	[0.813; 0.912]	[0.813; 0.912]
3	[0.926; 1.04]	[0.897; 1.01]	[0.823; 0.924]	[0.632; 0.709]	[0.632; 0.709]
4	[0.838; 0.939]	[0.811; 0.909]	[0.745; 0.834]	[0.572; 0.641]	[0.572; 0.641]
5	[0.769; 0.863]	[0.744; 0.836]	[0.683; 0.767]	[0.525; 0.589]	[0.525; 0.589]
6	[0.762; 0.851]	[0.738; 0.824]	[0.678; 0.756]	[0.52; 0.581]	[0.52; 0.581]
7	[0.51; 0.573]	[0.494; 0.555]	[0.454; 0.51]	[0.348; 0.391]	[0.348; 0.391]
8	[0.315; 0.353]	[0.305; 0.342]	[0.28; 0.314]	[0.215; 0.241]	[0.215; 0.241]
9	[0.208; 0.233]	[0.201; 0.226]	[0.185; 0.207]	[0.142; 0.159]	[0.142; 0.159]

Table 3: Evidences of intelligence.

Source	Evidences								
	Trivial			Middle			Critical		
	O_1	O_2	O_3	O_1	O_2	O_3	O_1	O_2	O_3
S_1	12	5	1	12	5	1	12	5	1
S_2	13	1	2	12	5	1	12	5	1
S_3	21	2	0	12	5	1	12	5	1
S_4	3	2	0	12	5	1	12	5	1
S_5	14	9	2	12	5	1	12	5	1
S_6	5	1	0	12	5	1	12	5	1
S_7	23	0	1	12	5	1	12	5	1
S_8	7	1	3	12	5	1	12	5	1

Table 4: Effective coverage score values for the three simulated companies.

Threat ID	O_1	O_2	O_3
1	0.5	0.61	0.85
2	0.23	0.79	0.78
3	0.64	0.74	0.95
4	0.12	0.34	0.67
5	0.21	0.87	0.90
6	0.16	0.56	0.75
7	0.15	0.23	0.57
8	0.68	0.75	0.91
9	0.91	0.87	0.82
SCS	0.4	0.64	0.8

Table 5: Ranges of economic losses for each considered organization (k$).

Threat ID	O_1 (0.4)	O_2 (0.64)	O_3 (0.8)
1	[397.53; 446.46]	[417.31; 468.68]	[209.13; 234.86]
2	[355.51; 400.48]	[175.87; 198.12]	[171.17; 192.82]
3	[136.53; 153.46]	[138.41; 155.58]	[49.15; 55.24]
4	[189.65; 212.34]	[192.48; 215.51]	[110.39; 123.60]
5	[229.01; 258.98]	[84.09; 95.10]	[67.10; 75.89]
6	[217.65; 244.34]	[171.48; 192.51]	[106.47; 119.52]
7	[120.56; 135.43]	[139.39; 156.60]	[87.31; 98.08]
8	[28.01; 31.38]	[30.55; 34.24]	[13.86; 15.53]
9	[13.93; 15.66]	[22.31; 25.0851]	[24.95; 28.04]

approach but reduces the margin for subjectivity by introducing the use of some quantitative key risk indicators. Its applicability in practice has been illustrated through some preliminary case studies taken from the industrial manufacturing world. Future works will concern application of the proposed method in a variety of different sectors, with the aim to catch practical evidences of the advantages it offers with respect to previous methods.

The proposed model should allow overcoming the intrinsic limits of the existing risk assessment approaches, which are based on the estimate of the threats' occurrence probability and on the observation of events that occurred in the past, thus not guaranteeing an adequate protection for the future. The proposed approach is also expected to

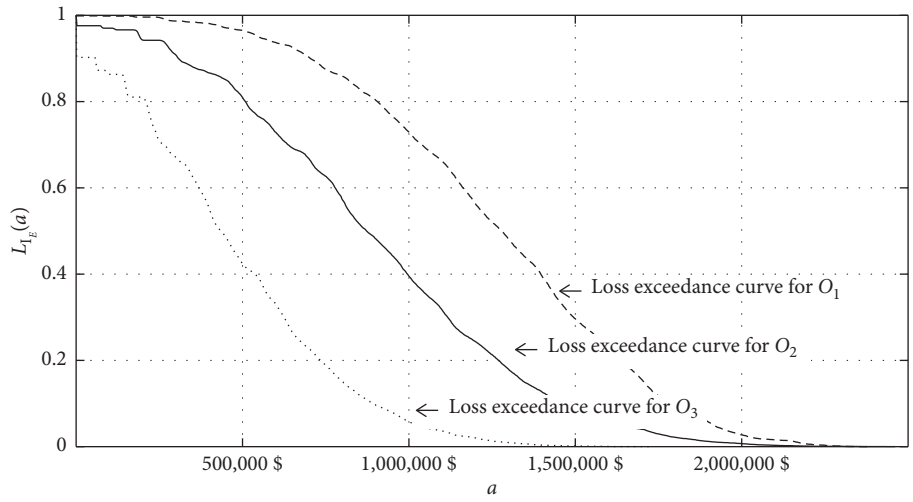

Figure 1: Loss exceedance curves for the three considered organizations.

provide companies and institutions with a new practical tool able to assess their cyber risk exposure and helping the definition of data protection policies for processes, systems, and infrastructures. The proposed solution is wide-ranging and applicable to different contexts, maintaining versatility and possibility to be used by companies and institutions of different size and working in different fields.

Data Availability

All data used in the paper come from cited references or are reported in the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. Ugur Aksu, M. Hadi Dilek, E. Islam Tatli et al., "A quantitative CVSS-based cyber security risk assessment methodology for IT systems," in *Proceedings of the 2017 International Carnahan Conference on Security Technology*, pp. 1–8, ICCST, Madrid, Spain, October 2017.

[2] K. Hartmann and C. Steup, "The vulnerability of UAVs to cyber attacks—an approach to the risk assessment," in *Proceedings of the 5th International Conference on Cyber Conflict*, pp. 1–23, CYCON, Tallinn, Estonia, June 2013.

[3] B. Karabacak and I. Sogukpinar, "ISRAM: information security risk analysis method," *Computers & Security*, vol. 24, no. 2, pp. 147–159, 2005.

[4] S. Naumov and I. Kabanov, "Dynamic framework for assessing cyber security risks in a changing environment," in *Proceedings of the 2016 International Conference on Information Science and Communications Technologies (ICISCT)*, pp. 1–4, Tashkent, Uzbekistan, November 2016.

[5] C. Biener, M. Eling, and J. H. Wirfs, "Insurability of cyber risk: an empirical analysis," *The Geneva Papers on Risk and Insurance—Issues and Practice*, vol. 40, no. 1, pp. 131–158, 2015.

[6] A. Rot, "IT risk assessment: quantitative and qualitative approach," in *Proceedings of the World Congress on Engineering and Computer Science 2008 (WCECS 2008)*, San Francisco, CA, USA, October 2008.

[7] R. Bojanc and B. Jerman-Blažič, "An economic modelling approach to information security risk management," *International Journal of Information Management*, vol. 28, no. 5, pp. 413–422, 2008.

[8] S. Romanosky, "Examining the costs and causes of cyber incidents," *Journal of Cybersecurity*, vol. 2, no. 2, pp. 121–135, 2016.

[9] D. W. Hubbard and R. Seiersen, *How to Measure Anything in Cybersecurity Risk*, Wiley, Hoboken, NJ, USA, 2016.

[10] J. Freund and J. Jones, *Measuring and Managing Information Risk—A FAIR Approach*, Butterworth-Heinemann, Oxford, UK, 2015.

[11] M. McNeil, T. Llansó, and D. Pearson, "Application of capability-based cyber risk assessment methodology to a space system," in *Proceedings of the 5th Annual Symposium and Bootcamp on Hot Topics in the Science of Security (HoTSoS'18)*, pp. 1–10, ACM, Raleigh, NC, USA, April 2018.

[12] M.-E. Paté-Cornell, M. Kuypers, M. Smith, and P. Keller, "Cyber risk management for critical infrastructure: a risk analysis model and three case studies," *Risk Analysis*, vol. 38, no. 2, pp. 226–241, 2018.

[13] M. Evans, L. A. Maglaras, Y. He, and H. Janicke, "Human behaviour as an aspect of cybersecurity assurance," *Security and Communication Networks*, vol. 9, no. 17, pp. 4667–4679, 2016.

[14] M. Eling and J. Wirfs, "What are the actual costs of cyber risk events?," *European Journal of Operational Research*, vol. 272, no. 3, pp. 1109–1119, 2019.

[15] B. Edwards, S. Hofmeyr, and S. Forrest, "Hype and heavy tails: a closer look at data breaches," *Journal of Cybersecurity*, vol. 2, no. 1, pp. 3–14, 2016.

[16] A. Baze, "Realistic risk management using the CIS 20 security controls," Technical Report, SANS Institute, Bethesda, MA, USA, 2016.

[17] S. Baddelpeli and G. Vert, "Adaptive security metrics for computer systems," in *Proceedings of the 2006 International Conference on Security & Management*, pp. 351–356, SAM, Las Vegas, NV, USA, June 2006.

[18] M. Jouini and L. B. A. Rabai, "Mean failure cost extension model towards security threats assessment: a cloud computing

case study," *Journal of Computers*, vol. 10, no. 3, pp. 184–194, 2015.

[19] Ponemon Institute LLC—Accenture, *Cost of Cyber Crime Study*, Technical Report, Ponemon Institute LLC—Accenture, North Traverse City, MI, USA, 2017.