MODELING CYBER INCIDENT OCCURRENCE FOR PROBABILISTIC CYBER RISK ASSESSMENT

Giulia RAFAIANI, Massimo BATTAGLIONI, Marco BALDI and Franco CHIARALUCE
Department of Information Engineering, Polytechnic University of Marche, Ancona, Italy

---

ABSTRACT

Cyber security challenges have been the national security in today's world, organizations ranging from small to large enterprises, government and private universities, hospitals, all prone to cyber-attacks from across the globe. National and international data protection regulations dedicate particular attention to cyber risk assessment and management. In the literature, a great deal of effort has been devoted to the development of tools and methods for cyber risk assessment. However, existing methodologies often lack straightforwardness, and their implementation may result burdensome in real scenarios. An intuitive, but quantitative model to estimate the likelihood of occurrence of a cyber incident in a certain period is here provided. Then, multiplying such a quantity by the impact of the corresponding threat, a cyber risk index is obtained. Our model combines three indexes (maturity, complexity and attractiveness) characterizing the considered organization and exploits a generalized logistic function and the properties of conditional probabilities to compute the desired likelihood. We validate the effectiveness and practicality of our method with numerical examples.
***Keywords*:** Cyber incident, cyber risk, FAIR, HTMA, logistic function, probabilistic risk assessment

1. INTRODUCTION

In recent years, different privacy and data protection regulations have been introduced to protect the increasing amount of personal data that is daily processed. Processing of personal data is as expected subject to different risks in terms of confidentiality, integrity, availability, authenticity, and reliability of data and services (Gritzalis *et al.*, 2018). This leads to the need for efficient tools for assessing and managing such risks. Risk is defined as the likelihood

of occurrence of a threat multiplied by the potential adverse impact of its occurrence (Taubenberger *et al.*, 2011; NIST SP 800-30, 2012). The use of risk assessment methods is essential for identifying and categorizing risks. However, a unique and globally recognized risk assessment method does not exist. Yet, the International Organization for Standardization (ISO) defines some steps for risk assessment methods (ISO 31000, 2018; ISO 27005, 2018): after the context establishment, the main steps are risk identification, analysis, and evaluation. Risk identification aims to identify critical assets and the associated threats and vulnerabilities, while risk analysis is needed to determine the likelihood of occurrence and the impact of threats (Shamala *et al.*, 2013; ISO 27005, 2018). Eventually, risk evaluation is useful to compare the results with previously defined risk acceptance criteria (ISO 31000, 2018; ISO 27005, 2018). Risk assessment methods could be classified into quantitative and qualitative approaches, mainly according to the risk analysis step. Quantitative methods rely on numerical categories; their results are robust, comparable and reproducible. However, these methods are costly in terms of time and resources (NIST SP 800-30, 2012). In fact, they are often difficult to implement in real contexts and require some experts to perform them. Qualitative methods, instead, are simple to interpret and fast to implement, since they use nonnumerical categories. However, the results are very subjective, making it difficult to reproduce and compare them (NIST SP 800-30, 2012). In the literature, there is a large number of approaches to cyber risk assessment. Some of them have been proposed by national and international organizations, or by public and private organizations (a complete list can be found in ENISA (2022)). Gritzalis *et al.* (2018) and Giuca *et al.* (2021) provided an exhaustive analysis of these methods. Furthermore, lots of efforts have been done in the literature in order to find an optimal solution to the likelihood and impact estimation problem (Freund and Jones 2015; De Gusmão *et al.*, 2016; Hubbard and Seiersen 2016; Aksu *et al.*, 2017; Handa *et al.*, 2019; Khosravi-Farmad *et al.*, 2020; Schmitz *et al.*, 2020; Kim and Weber, 2022; Kure *et al.*, 2022). These approaches are often difficult to interpret and complex to manage, especially for small and medium organizations. In fact, they usually need a list of possible threats, of all the vulnerabilities and/or of the relations among all components of the considered system. It follows that providing an exhaustive and complete list may be unfeasible. Furthermore, while information security is a continuously evolving subject, those lists heavilyrely on the knowledge of past events, that are often hard to obtain (Patel *et al.*, 2008).

The present paper proposes a method for assessing the likelihood of occurrence of a cyber threat that combines the advantages of both quantitative and qualitative methods. In fact, the model we proposeprovides a quantitative approach that, at the same time, drastically reduces the costs in terms of

required time and resources. Since this approach is very simple, organizations can exploit it to perform self-assessments, without renouncing objectivity of the results. In the model we propose, we exploit and combine three main parameters: the maturity of the target organization, the complexity of its technological infrastructure, and its attractiveness. Therefore, with respect to the traditional gap analysis (Conceptivity 2018), we consider two additional parameters: complexity and attractiveness. Moreover, opposed to traditional gap analysis methods, we use a mathematical model for quantitatively evaluating the likelihood of occurrence of a cyber threat.

2. COMPONENTS AND MODELIZATION OF CYBER RISK

In the proposed model, the three key parameters as maturity, complexity and attractiveness of the target organization were used to assess the cyber risk. In fact, the risk of suffering one successful attack not only depends on the protection measures adopted by the organization (maturity), but also on the complexity of the organization itself and on the number and type of attack attempts experienced in a given period (attractiveness).

2.1. MODEL COMPONENTS

In this section, we describe the proposed model and discuss its individual components.

Maturity index

We define the maturity index as the level of adherence of the organization to the controls addressed by one or more cyber security frameworks. The choice of the framework determines the area of application of the model; for example, CIS controls (2021) can be used for evaluating cybersecurity compliance, the controls proposed ENISA (2017) can be used for assessing data protection compliance, while the controls of the Cybersecurity Framework (NIST, 2018) can help the organization to evaluating both cybersecurity and data protection compliance. This allows considering the actual cyber posture of the target organization, instead of relying on a general list of known threats. Furthermore, since the frameworks are constantly updated, their use leads to the creation of a dynamic model. In the model we propose, the evaluation is performed by determining the level of implementation of each control of the chosen framework. This can be done through a binary response or using a scale to evaluate at which degree every control is implemented. In any case, an N/A option should be included. Once the implementation of each control has been evaluated, a score is associated to every answer and, finally, a maturity index is obtained as a weighted mean of the scores; controls with N/A are not considered in the average. The weights

must be chosen according to the considered framework and to the type of organization. In fact, if some controls are considered more crucial (or less relevant) for the assets under exam, a higher (or lower) weight can be assigned to them. The final output of the evaluation is a maturity index, expressed as a real-valued variable ranging from 0 to 10. Notice that, in most existing frameworks, there are also some controls dedicated to the awareness of the employees. These controls are directly related to non-malicious threats, which therefore are intrinsically considered in the computation of the maturity index.

Complexity Index

We define the complexity index as the level of the intrinsic complexity of the organization technological infrastructure. We consider such a complexity index as a key component for cyber risk assessment because we assume that the cyber posture of an organization is strictly related to the complexity of its infrastructure. Indeed, this concept was introduced in (CIS, 2021), where the security controls to be implemented depend onthe dimension of the considered organization. We evaluate the complexity index through a set of punctual controls, grouped in 5 categories. The chosen controls consider not only the dimension of the organization, but also the characteristics of the components and their interconnections, as well as the number of services and their interconnections, and the IT system management.

In the model we propose, the evaluation is done by determining the level of implementation of each control. For each control, five possible answers are given: minimal, low, moderate, significant and high complexity. Every answer is supported by a description of all the five levels to simplify the answering process for the assessor, and making it as objective as possible. Once the implementation of each control has been evaluated, a score is associated to every answer, and the complexity index is obtained through a weighted mean of the scores for each of the five categories. The weight associated to the controlsin each category is simply obtained as the total number of controls in that category divided by the overall number of controls. The final output is a real-valued complexity index, ranging from 0 to 10.

Attractiveness

The last key component of the model we propose is the attractiveness of the organization. The attractiveness is strictly related to the organization business, to the type and the amount of data the organization processes, and so on. We assume that an unattractive organization will be subjected, in a given period, to less attacks with respect to an attractive organization. Moreover, each attack will be composed of a relatively few attempts in case of an unattractive organization, and vice versa. The organization dimension does not (necessarily) affect its attractiveness. However, the attractiveness affects the

maturity of the attacks and of the attackers. In the model we propose, to evaluate the attractiveness of an organization we have examined data from cyber security reports as CLUSIT (2021). We have analyzed the number of attacks received by different types of businesses in relation to the total number of attacks documented in a year. This way, we have defined 5 different levels of attractiveness and we have classified the types of organizations receiving more than 10% of the total number of attacks as “Very high” attractive, those receiving more than 5% as “High”" attractive, those receiving more than 2.5% as “Medium” attractive, those receiving more than 1.25% as “Low” attractive, and, finally, those receiving less than 1.25% as “Very low” attractive.

Relations among the components

The parameters defined above are combined as in Figure 1 depicted. The attractiveness of the organization influences both the number of attacks and the maturity of attackers. The maturity and the complexity of the organization together influence, along with the maturity of the attackers, the probability of success of an attack. Maturity and probability are inversely related. Complexity and probability, instead, are directly related. The number of attacks and the probability of success of an attack influence the likelihood of occurrence of a successful attack. Finally, likelihood and impact determine the risk.

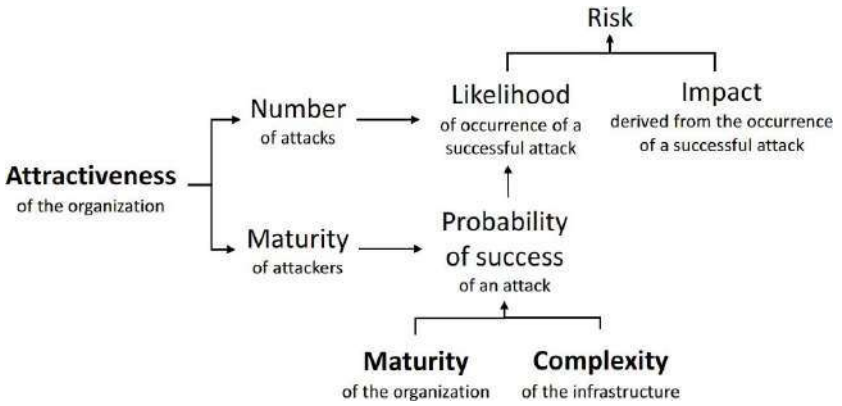

Fig. 1: Relations among the variables considered in the model.

2.2. CYBER RISK ASSESSMENT MODEL

We define an attack as an attempt to partially or totally disclose, expose and/or compromise data by an attacker. Attractive organizations are more likely to face more and more structured attacks, and vice versa. To be as close as possible to real scenarios, in our model each attack is associated to a certain probability to be successful and, therefore, more than one attempt may be needed to breach the organization. We assume that in the time slot Δt only one attack can be performed. We also assume that different attacks are not correlated. We note that in scenarios where such hypothesis is too optimistic, our model still provides a lower bound on the success probability. To quantitatively estimate the likelihood of occurrence of a successful attack, we exploit a function that, taking the complexity and the maturity indexes as arguments, returns the probability of success of a single attack. As the Figure 1 depicts, this probability, when combined with the attractiveness, could be used to estimate the likelihood of having a successful attack in a given period.

In our model, the probability of success of an attack, noted by $P(S)$, is related to the maturity index x through the following generalized logistic function:

$$\mathrm{P(S)} = A + \frac{K-A}{1+e^{-B(x-x_0)}}, \text{ with } 0 \leq x \leq 10, B<0 \qquad (1)$$

The reasons behind this choice are explained in (Rafaiani 2021).

The value of x is limited by the range of the maturity index. Moreover, we assume that the probability of success cannot reach 1 and 0, which are unrealistic values. Then, the maximum and the minimum values of the function are set at U for x = 0 and at L for x = 10. K and A depend on $x0$ and can be easily obtained by setting $f(0)$ = U and $f(10)$ = L. Finally, $x0$ is set equal to the complexity index, that depends on the infrastructure.

To take the attractiveness into account, we first weight $P(S)$ according to the organization attractiveness by computing

$$p = w\, x\, P(S), \qquad (2)$$

where w is the attractiveness weight. Some possible choices for w are given in Table 1.

The weighted probability of success p, computed according to (2), is then used to estimate the likelihood that the organization under exam will face, in the considered time interval (a year in our evaluation), one successful attack, after a certain number of failed attempts, whose probability of occurrence is

also analyzed in probabilistic terms. In fact, it is not useful to determine the likelihood to have at least one successful attack since, once an attack is successful, the organization will likely improve its initial conditions.

Table 1. Possible values for w as a function of the attractiveness level.

Attractiveness	Very Low	Low	Medium	High	Very High
w	0.6	0.7	0.8	0.9	1

2.3. PROBABILISTIC APPROACH

Let us assume that in a certain time slot Δt, the organization either faces an attack or not. So, in a period containing t time slots, at most t attacks can be experienced. Let us define the following events:

- A: the organization experiences an attack in the time slot Δt;
- T_N: the organization experiences exactly N attacks in a period of t time slots;
- T_N^*: the organization experiences at most N attacks in a period of t time slots.

Based on these premises, we can assume that the probability of experiencing exactly N attacks in a period of t time slots follows a binomial distribution, i.e.,

$$P(T_N) = \binom{t}{N} P(A)^N [1 - P(A)]^{(t-N)} \quad (3)$$

where $\mathrm{P(A)} = \frac{N_{avg}}{t}$ and N_{avg} is the average number of attack attempts experienced in a period of t time slots.

Similarly, when t is chosen relatively large, we can use a Poisson distribution, which is indeed the limit of a

binomial distribution when the number of Bernoulli trials, t, goes to infinity. Accordingly, for large t, we have

$$P(T_N) = \frac{\lambda^N}{N!} e^{-\lambda} \quad (4)$$

where $\lambda = \mathrm{N_{avg}}$. It is easy to conclude that, whatever distribution is used for $P(TN)$, the probability of experiencing at most N attack attempts in a period containing t time slots results in

$$P(T_N^*) = \sum_{i=0}^{N} P(T_i) \quad (5)$$

Single Attack

Under the assumptions of Section 2.2, the probability that the organization will suffer an attack in a certain time slot Δt, and this attack will be successful, is

$$L = pP(A), \tag{6}$$

having already considered the need to weight the probability of success, according to (2). L has therefore the meaning of likelihood that the event will happen.

To provide a single output for the probability that the organization faces a single attack in a given time slot Δt, we can randomly sample the chosen (binomial or Poisson, depending on the values of Navg and t) distribution. Finally, by definition, the risk associated to a certain event can be computed as

$$R = L\,x\,I \tag{7}$$

where I is the impact of the threat occurrence.

Multiple Attacks

If we consider the more realistic scenario in which multiple attacks can occur, we also need to define the following event σ: the first successful attack is experienced by the organization after at most N attack attempts. We can describe the random variable corresponding to σ with a cumulative (due to the "at most" part, having assumed that the attempts are independent) geometric distribution, obtaining

$$P(\sigma|T_N^t) = \sum_{i=1}^{N} p(1-p)^{i-1} \tag{8}$$

By definition of conditional probability, we have that

$$P(T_N^t \cap \sigma) = P(T_N^t)P(\sigma|T_N^t) \tag{9}$$

where $\cap$ defines the intersection of two events and $P(T_N^t)$ represents the probability of the conditioning event. Moreover, by Bayes' theorem, we have that

$$P(T_N^t \cap \sigma) = P(T_N^t)P(\sigma|T_N^t) = P(\sigma)P(T_N^t|\sigma). \tag{10}$$

We can assume that $P(T_N^+|\sigma) = P(T_N^+)$, therefore obtaining $P(\sigma) = P(\sigma|T_N^+)$. Finally, we obtain

$$P(T_N^+ \cap \sigma) = P(T_N^+)\, P(\sigma) \qquad (11)$$

which is the probability that the organization will face at most N attack attempts and one of them will be successful.

3. APPLICATION OF THE MODEL TO REAL DATA

To validate our approach, we have compared the results obtained through it with some real data found in the literature. We have considered a recent cybersecurity report by Accenture (Accenture 2020). In such a report, organizations are classified into Leaders and Non-leaders. The former is identified as those organizations that, among the sample, have been able to better identify and manage data breaches (Accenture, 2020). We can associate this classification to different maturity indexes in our model. As an example, we have associated a maturity index of 9 to Leaders, and a maturity index of 6 to Non-leaders.

Based on (Accenture 2020), only 1 out of 27 cyber-attacks (3.7%) actually resulted in a security breach for the leader organizations, while the non-leader organizations suffered 1 security breach out of 8 cyber-attacks (12.5%). Unfortunately, the report does not contain any information about the complexity of the organizations in the sample and, therefore, we have arbitrarily set $x0$ to its mid-value, i.e, $x0 = 5$ for both types of organizations.

Starting from the aforementioned (average) probabilities of success of a single attack, we obtain the following parameters to be used in our model: $B = -2$, $U = 0.97$, and $L = 0.03$. The corresponding logistic function is in Figure 2 depicted.

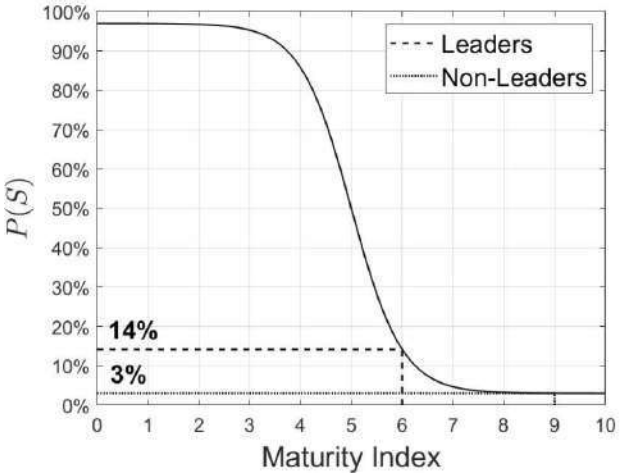

Fig. 2: Probability of success of a single attack for leader and non-leader organizations.

In order to take the attractiveness of the organization into account, we can fix $w = 1$ for leader organizations and $w = 0.7$ for non-leader organizations.

All these data can be given as input to the considered probabilistic model, to estimate the probability that the organization will face a certain number of attacks, and one of them will be successful. By considering t =365 (and therefore a period of one year with daily intervals), $N_{avg} = 239$, and $p = 0.037$ for leader organizations, and $t = 365$, $N_{avg} = 166$, and $p = 0.088$ for non-leader organizations, and by performing Monte Carlo simulations, we obtain the binomial distributions depicted in the Figure 3. We remark that we consider the binomial distribution since N_{avg} and t have the same order of magnitude. When $t \gg N_{avg}$, instead, we can use the Poisson distribution.

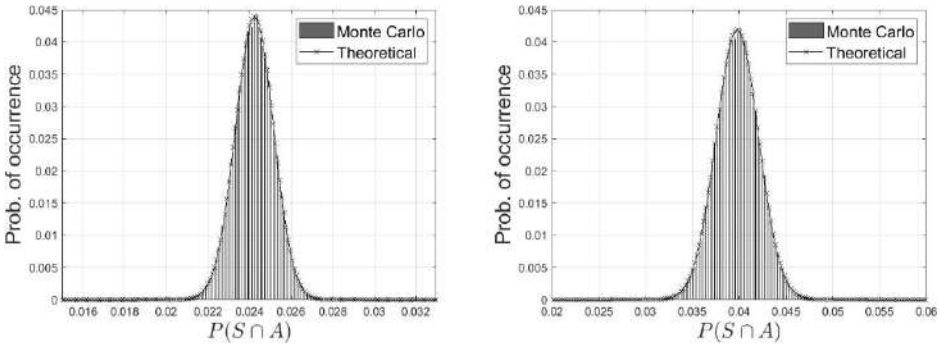

Fig. 3: Binomial distribution for $P(S \cap A)$ when t = 365, N_{avg} = 239, p = 3.7% (a) and N_{avg} = 166, p = 8.8% (b).

At this point, using (8) and (11), we are able to estimate the probability that the organization will face a successful attack, after a certain number of attempts and the probability that the organization will face a certain number of attempts and one of them will be successful. The results are in Figure 4 shown.

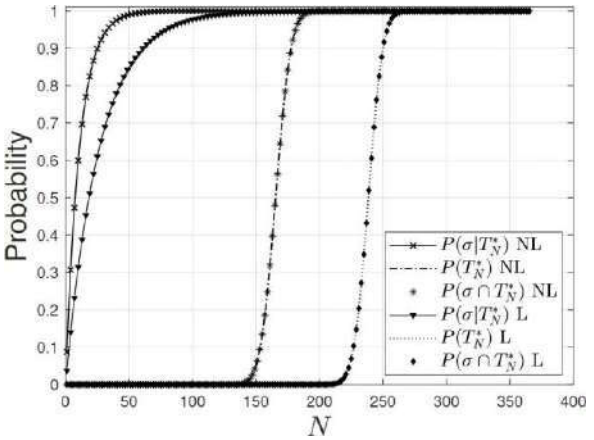

Fig. 4: Conditional and conditioning probabilities for p = 3.7% (L = Leader organizations) and p = 8.8% (NL = non-Leaderorganizations).

Notice that, since N_{avg} is rather high, it is very unlikely that the organization will face only a relatively small number of attacks. However, when N is slightly larger than N_{avg}, $P(\sigma \cap T_N^*)$ saturates the cumulative geometric curve, since after all those trials, it is very likely that the attacker will succeed. Clearly, a leader organization is more likely to face a larger number of attacks due to its higher attractiveness but, as the cumulative distribution curve suggests, leader organizations (that are characterized by a larger maturity index with respect to non-leader organizations) are more prepared to recognize and counter the attacks they receive.

4. CONCLUSIONS

The model here proposed provides a simple, quantitative, and cost-effective way for an organization to estimate the likelihood of receiving a successful cyber-attack in a specified period. The evaluation of the main components of the model is done through questionnaires; this enables also small and medium organizations to easily assess cyber risk, without relying on

external experts. The outputs of the model can be considered as a good starting point for cyber risk management. In fact, the model can be used to estimate the number of improvements the organization should perform for an acceptable likelihood of successful attacks, i.e., an acceptable risk. The validation of the model on real data proves that the proposed method is practical; the obtained results show how the model works for different organizations with different initial conditions and postures.

REFERENCES

Accenture. 2020. Third annual state of cyber resilience report.

Aksu M, Dilek M, Tatli E, Bicakci K, Dirik H, Demirezen M, Aykir T. 2017. A quantitative CVSS-based cyber security risk assessment methodology for IT systems. In: Proceedings 2017 International Carnahan Conference on Security Technology (ICCST). Madrid, Spain.

Associazione Italiana Per La Sicurezza Informatica. 2021. Rapporto Clusit 2021 Sulla Sicurezza ICT In Italia.

Center of Internet Security (CIS). 2021. CIS Controls v.8.

Conceptivity. 2018. Cybersecurity standard gap analysis, White paper.

De Gusmão APH, Silva, LC, Mendonça Silva M, Poleto T, Seixas Costa APC. 2016. Information security risk analysis model using fuzzy decision theory. International Journal of Information Management 36(1), 25–34.

ENISA (European Union Agency for Network, Information Security). 2017. Handbook on security of personal data processing.

ENISA (European Union Agency for Network, Information Security). 2022. Compendium of risk management frameworks with potential interoperability - Supplement to the Interoperable EU Risk Management Framework Report.

Freund J, Jones J. 2014. Measuring and Managing Information Risk: A FAIR Approach. Butterworth- Heinemann.

Giuca O, Popescu T, Popescu A, Prostean G, Popescu D. 2021. A survey of cybersecurity risk management frameworks. In: V. Balas, L. Jain, M. Balas, S. Shahbazova (eds.) Soft Computing Applications. SOFA 2018. *Advances in Intelligent Systems and Computing*, vol. **1221**. Springer, Cham.

Gritzalis D, Iseppi G, Mylonas A, Stavrou V. 2018. Exiting the risk assessment maze: A meta-survey.

ACM Comput. Surv. 51(1), 1–30.

Handa A, Mukhopadhyay S, Mallick S, Kumar N, Shukla S, Minz R, Nagarmat S, Rakesh R. 2019. Cyber risk assessment of networked cyber assets using probabilistic model checking. In: Proceedings 2019 IEEE

Conference on Information and Communication Technology, Allahabad, India.

Hubbard D, Seiersen R. 2016. How to Measure Anything in Cybersecurity Risk. John Wiley & Sons Inc.

International Organization for Standardization (ISO). 2018. ISO 31000:2018 - risk management – guidelines.

International Organization for Standardization (ISO). 2018. ISO/IEC 27005:2018 information technology - security techniques - information security risk management.

Khosravi-Farmad M, Ghaemi-Bafghi A. 2020. Bayesian decision network-based security risk management framework. *Journal of Network and Systems Management*, **28(4):** 1794–1819.

Kim S, Weber S. 2022. Simulation methods for robust risk assessment and the distorted mix approach.

European Journal of Operational Research, **298(1):** 2022, Pages 380-398.

Kure HI, Islam S, Mouratidis H, 2022. An integrated cyber security risk management framework and risk predication for the critical infrastructure protection. Neural Comput & Applic.

NIST (National Institute of Standards, Technology). 2012. Special publication 800-30 revision 1 - information security: Guide for conducting risk assessments.

NIST (National Institute of Standards, Technology). 2018. Framework for improving critical infrastructure cybersecurity - version 1.1.

Patel S, Graham J, Ralston, P. 2008. Quantitatively assessing the vulnerability of critical information systems: A new method for evaluating security enhancements. International Journal of Information Management 28(6), pages 483–491.

Rafaiani G, Battaglioni M, Chiaraluce F, Libertini G, Spalazzi L, Cancellieri G. 2021. A functional approach to cyber risk assessment. Proc. AEIT 2021 International Annual Conference, Second Virtual Edition.

Schmitz C, Pape S. 2020. LiSRA: Lightweight security risk assessment for decision support in information security. Computers & Security 90, 1–20.

Shamala P, Ahmad R, Yusoff M. 2013. A conceptual framework of info structure for information security risk assessment (ISRA). *Journal of Information Security and Applications* ,**18(1):** 45– 52.