Front.Comput.Sci.
DOI

RESEARCH ARTICLE

New Multi-objective Approach for Dynamic Risk-Driven Intrusion Responses

Chaker Katar(✉)[1], **Ahmed Badreddine** [1]

1 LARODEC, Institut Supérieur de Gestion de Tunis, 41 Avenue de la liberté, 2000 Le Bardo, Tunisie

Abstract Despite the abundant literature reporting on intrusion detection and response systems (IDRSs), the cost models adopted in planning security responses have not paid significant attention to security risks. Even though the ISO 27005 and NIST (National Institute of Standard and Technology) and FIPS (Federal Information Processing Standard) guidelines provide sufficient specifications for planning security responses, there still have not been many studies in which the IDRS systems were designed accordingly. In this context, we propose a new risk-driven response mechanism that conforms to ISO 27005 and NIST and FIPS guidelines. This mechanism follows established specifications for collecting risk information, assessing basic risks, mitigating risks, and generating risk-driven optimal security strategies. In order to ensure optimality in computing the security strategies, we adopt the multi-objective influence diagram technique to search for the optimal security strategy that minimizes risks and security investment cost under strict constraints. In order to demonstrate the working of our proposed risk-driven intrusion detection response mechanism, we provide a detailed numerical example using simulated test environment and network traffic.

Keywords Intrusion response. Risk management. Security strategy. ROSI criterion. Multi-objective optimization. Influence diagrams

Received May 05, 2018; accepted December 19, 2018

E-mail: *chaker.katar@isg.rnu.tns*

1 Introduction

The intrusion detection literature has increasingly paid attention to intrusion response mechanisms. These are commonly classified as passive or active [1, 2]. Passive response mechanisms simply alert the site security officer (SSO) to the detected attacks. Active mechanisms, in contrast, propose multiple response actions that range from simple rule revision to the deployment of a security control. They assist the SSO in selecting and implementing response actions against detected threats depending upon their automation levels, manual or automatic [3]. Automatic response mechanisms have been also categorized based on response time criteria into proactive and reactive classes [3]. Proactive response mechanisms preempt attacker action sequences and react before attacks take place. Reactive response mechanisms, in contrast, implement their corrective actions after the detection of mounted attacks. Automated responses have been further structured into the categories static, dynamic, and cost-sensitive according to how response actions are selected [4]. Additional response categorizations have been identified involving criteria such as adjustment and cooperation capabilities of automated components [1, 2].

Several active intrusion response mechanisms have been designed and prototyped, including EMERALD and AAIRS [2, 5]. They adapt well to specificities of the monitored system, though they lack explicit and clear processes to select new security controls or evaluate existing ones. In the last decade, cost-effective responses have gained increased insight. The cost-sensitive response mechanisms involve var-

ious cost factors such as damage and response costs, which assess attack harm and security control impacts, respectively, on the target system. They rely mainly on trade-offs between cost factors to select appropriate security controls. However, they have never considered risk models as recommended by ISO 27002 and ISO 27005 to assess and mitigate risks of mounted attacks [6, 7]. Additionally, they have several drawbacks in cost factor assessment as well as in countermeasure selection. On one hand, these mechanisms involve arbitrarily and statically assigned cost factors. Some mechanisms use constant response cost, damage cost, and response financial impact [8, 9], which are useless and require stringent revision to be realistic for target computing environments. Additionally, other mechanisms lack appropriate processes to assess cost factors. For instance, intrusion damage cost in [8, 10] is approximated, similarly to the response cost, using attack implementation cost and impact on the target asset. However, the implementation cost of an attack involves difficult-to-estimate parameters, including the attacker's skill. Existing processes for determining the damage cost also ignore required information about attack likelihood, deployed security controls, and the vulnerability state of the target asset. On the other hand, control selection processes of existing cost-sensitive mechanisms are commonly based on trade-offs [2, 11]. For every detected attack, they select the security control subset associated with the highest expected value or expected utility [8, 11]. However, selected controls are useless if the target asset is not critical to the organization business functions and incurred damage is under the tolerance level. These processes also involve techniques such as Bayesian networks, system maps, and dependency trees [11, 12]. They are seldom formulated as optimization tasks. Although several goals such as minimization of attack damage and response cost should be satisfied to reach the optimal security control subset, few response mechanisms have integrated multi-objective optimization techniques to control selection processes.

To address these issues, we propose in this paper a new risk-driven intrusion response mechanism. The latter relies on a new risk model that complies with the ISO 27005 and FIPS 65 standards and abides by the recommendations of guidelines such as NIST's [13–15]. This model is divided into two parts:

1. **The risk assessment part** concerns the determination and the estimation of inflicted damage on target assets due to detected attacks.
2. **The risk treatment part** focuses on selection of the optimal security control subset to mitigate risks of detected attacks.

The remainder of this paper is organized as follows. Section 2 reviews main cost-sensitive response mechanisms and highlights their shortcomings. Section 3 discusses our research motivations and states treated problem. Section 4 introduces the proposed response approach and presents in detail the designed risk management model with its two parts. Sections 5 and 6 illustrate the processing steps of the approach and discuss results achieved, respectively. Finally, section 7 concludes and presents perspectives on this work.

2 Related work

Although several classes of active response mechanisms have been identified, the cost-sensitive class has gained more attention in the last decade. The studies concerning cost-sensitive intrusion response can be divided into two categories:

1. Mechanisms without system model: This category includes mechanisms such as those proposed by Stakhanova et al. [4, 8] and Iafarov et al. [10]. These mechanisms are commonly based on trade-offs between the damage cost and the response cost of corrective actions. Additionally, various criteria are used to select the most appropriate response action, including expected value of the response [8], the return on response investment (RORI) metric [9], the expected utilities of candidate response plans [11]. In a recent work, Nejat and Kabiri [16] have proposed an adaptive and cost-based response framework that involves cost factors such as response cost, attack cost, and cost balance. Moreover, this mechanism supports multiple processing units including priority resolver, response selection, and launcher units that respectively determine costs of detected attacks, collateral cost and merit value of a response. Response launcher then implements response action associated with higher merit value.
2. Mechanisms with system model: This category groups intrusion response mechanisms based on system and cost models. Cost model includes factors such as penalty cost [12], intrusion and response costs [17, 18], damage cost and response impact [19] (see Table 1). System model instead reproduces relationships between assets of the computing environment. It corresponds to a dependency tree in [12]; whereas in [17–19] a graph

structure is used to capture dependencies between system assets. In recent response framework of Shameli-Sendi et al. [20], the system model is composed of a service dependency graph (SDG) and an attack and defense tree (ADT). The SDG reproduces dependencies between different services of the system, their importance, and the severities of propagated impacts. The ADT illustrates both steps followed by an attacker and security controls in place to defend against them. The framework of Shameli-Sendi et al. takes account also of several cost factors including damage and deployment costs.

The response mechanisms discussed above are capable of implementing reactive and proactive reactions such as those described in [11, 20]. Additionally, they allow dynamic revision of current selected responses on the basis of past ones by including factors such as response success and response goodness [8, 20]. Mechanisms of both groups commonly select appropriate responses using trade-offs between damage and response costs or may also consider the costs and benefits of responses. However, in spite of the selection criterion involved, such as RORI [9] or response merit [16], they use unbounded costs and benefit factors, which lead to confusing situations. Therefore, they select response actions to defend against every detected attack, even those whose incurred damage is below the tolerated damage cost. Additionally, they implement response actions that are costly with respect to the maximum accepted response cost in order to reinforce the security of assets that are not critical to business functions of the organization. Even mechanisms based on selection criteria such as RORI or response merit are lacking lower bounds on acceptable RORI or merit; otherwise, response actions with financial impact exceeding their security impacts would be selected [9, 16].

3 Research motivations and problem statement

The response mechanisms reviewed above are based on cost models that include various factors such as penalty, response, and damage costs, as shown in Table 1. Therefore, they need a stringent initialization step to better fit the security requirements of the monitored computing environment. For instance, in [17] (see Table 1), several cost factors are determined based on statically assigned parameters. Moreover, a determination process is defined specifically for response benefit but not for other cost factors. In this mechanism, optimization methods are not involved in designing and selecting response actions to be deployed. However candidate responses are determined based on their evaluated benefits and costs. A few mechanisms instead use dynamically estimated parameters that depend on the target environment and the detected attack [20]. They do not require extensive historical databases of assets, detected attacks, and deployed security controls to determine cost factors, and thus design candidate response strategies. Indeed, in mechanisms summarized in Table 1, a candidate consists of single or multiple security controls identified using criteria such as low penalty cost and "damage cost exceeds response cost". The most appropriate strategy to be implemented is then determined from the candidates based on selection criteria including the lowest penalty cost and the highest expected value. However, selection criteria for these mechanisms could produce inappropriate choices, for two main reasons. First, they lack appropriate parameters to assess different cost factors and select suitable response actions; such parameters include asset value, probability of exploiting a vulnerability, and attack likelihood. Indeed, these mechanisms usually take account of the damage cost factor, but its estimation does not involve parameters such as asset value, vulnerability impact, or attack likelihood. Additionally, they have never considered appropriate risk models to conduct a thorough assessment of damage inflicted on the target asset by detected attacks. These mechanisms may also select unacceptable response actions, such as the common case in which a response cost exceeds the value of the target asset. Second, these mechanisms use unbounded cost factors [4, 10], and consequently they react against every detected attack even if the incurred damage is below a given acceptance level. Additionally, they could select a wrong response, one that meets criteria such as the highest benefit or utility but not the acceptance levels [8, 11]. Furthermore, response selection involves simultaneous and conflicting objectives including maximization of response benefit and minimization of damage cost, and therefore, it is regarded as a typical multi-objective optimization problem (MOP). However, few mechanisms have formulated this problem as a multi-criteria optimization task and used appropriate methods to solve it. Moreover, though information security standards such as ISO 27002 recommend the involvement of risk management in designing or revising security strategies, it was commonly ignored by existing response mechanisms. These issues motivate us to propose a **new risk-driven intrusion response mechanism based on a compliant risk management model**, with the ISO 27005 and NIST standards, as depicted in figure 1, **and a multi-objective approach**. As such, our proposed risk model integrates two interdependent

Table 1 Comparison of cost-based response mechanisms

Work	Parameters involved	Parameter evaluation		Decision criteria		Optimization	Output
		Static or Dynamic	Process	Response design	Response selection		
[12]	Penalty cost	Static	None	Minimum penalty	Lowest overall penalty cost	None	Single response action
[4, 8]	Response cost and Response value	Static	Response cost and Response value	Response benefit exceeds response cost	Highest response value	None	Set of response actions
[17]	Response cost, Damage cost and Response benefit	Static	Response benefit	Maximum benefit	Highest benefit Lowest cost	None	Set of response actions
[18]	Effectiveness index and Disruptiveness index	Static	None	Response index	Highest response index	None	Single response action
[9]	Technical and financial impacts	Static	None	Security impact and Financial impact	RORI	None	Set of response actions
[11]	Expected utility, Implementation cost, and Alert risk	Static	Propagated probability of an alert (BDN)	Expected utility and acceptable Implementation cost	Highest expected utility	None	Set of response actions
[10]	Attack cost and Response cost	Static	None	None	Attack cost exceeds response cost	None	Set of response actions
[20]	Damage cost, Response performance, Security impact, and Security cost	Dynamic	Available	Acceptable Response performance, Security impact, and Security cost	Attack damage, cost severity, and highest score	Multi-objective	Single response action
[16]	Attack cost , Response cost , and response effectiveness	static	Available	Response merit	Highest response merit	None	Single response action
Our work	Exposure, Vulnerability severity, and Control effectiveness	Dynamic	Available	Acceptable residual risk and Response cost	ROSI	Multi-objective	Set of response actions

parts:

1. **The risk assessment part** which involves the identification of several parameters, related to detected attacks, target assets, supported vulnerabilities, and deployed security controls. Moreover, it includes detailed steps to estimate different parameters and evaluate basic risks incurred by a target asset due to detected attacks. It ensures a quantitative and realistic assessment of inflicted damages, and thus eliminates statically attributed costs.
2. **The risk treatment part** which focuses on selection of the optimal security control subset to thwart a detected attack. Therefore, it is formulated as a MOP. This aims to determine the optimal security strategy that minimizes both attack risks and response cost with regard to the tolerated risk level and the allocated security budget, respectively. The implementation of the risk mitigation uses the multi-objective influence diagram (MID) technique. It consists of three steps, namely, MID structure learning, parameter learning, and evaluation.

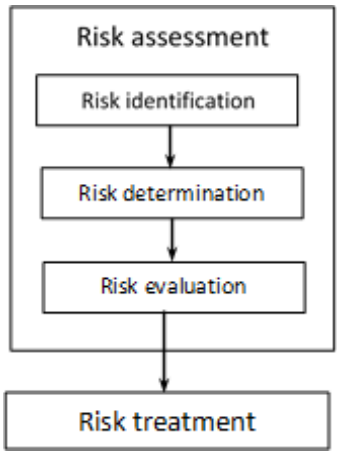


Fig. 1 Information security risk management according ISO 27005 [7]

4 Proposed framework for intrusion response

The detection mechanism of an IDRS identifies intrusive actions and eventually determines the probability of their true type. Then, the detection decision is forwarded to the response mechanism so that a suitable defense strategy can be designed and corrective actions against the detected attack can be implemented.

The risk-driven response mechanism proposed in this paper is depicted in figure 2 It relies on a new risk model. For every intrusion alert, this model starts by assessing damages incurred by the target assets. Next, it treats risks by reducing them to an acceptable level using an optimal combination of countermeasures. The latter is selected from those recommended with regard to security policy rules and strategic security objectives of the target organization. Risk assessment and treatment in the proposed risk model and the implemen-

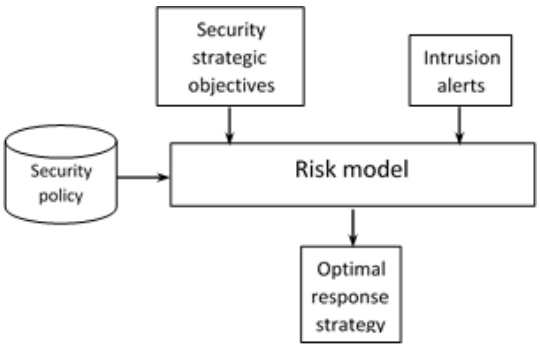


Fig. 2 Proposed risk-driven response framework

tations of the designed response mechanism are discussed in subsequent sections. Indeed our risk-driven response mechanism is dynamic. It reassesses damages incurred by target assets for every detected attack. Moreover, it revises and enhances deployed security solution depending upon inflicted damages, target asset, and detected attack. To illustrate risk assessment and mitigation in the frame of the proposed response mechanism (see section 5), only truly simulated and detected attacks (True Positive: TP) have been considered [8, 10, 17, 18, 20]. Additionally, this mechanism implements no reaction against detected normal behavior (True Negative: TN).

4.1 Proposed risk model

The proposed quantitative risk model complies with the ISO 27005 standard [7]. Furthermore, it is consistent with NIST recommendations [15]. It consists of two interdependent parts as illustrated in figure 1. On one hand, the assessment part aims at designing the graphical component (RM_G) and computing the numerical component (RM_N) of our risk model. Therefore, it respectively ensures the identification and the determination of risk parameters, and the evaluation of damage inflicted by detected threats. On the other hand, the treatment part proposes a multi-objective optimization formulation for modeling security control selection. The risk assessment part is concerned with the analysis and the evaluation of threat risks, whereas the risk treatment part concentrates on the post-assessment step, specifically, the risk mitigation.

4.2 Risk assessment

In our model, the risk assessment is performed in two steps. Initially, risks are analyzed; in this step, different risk parameters are identified and their determination processes are designed. Then, the basic risk to the target asset, a_i, is evaluated by considering the parameters determined in the previous step.

4.2.1 Risk identification and determination

The proposed risk management model relies on four main components, namely, assets, threats, vulnerabilities, and security controls, as depicted in figure 3. Threats and assets are critical components and are commonly considered by risk management methodologies. The threat component encompasses deliberate actions of insider or outsider entities who attempt to inflict damage on target assets. The asset component includes any computing resource valuable to the organization. The assets of the computing environment carry several vulnerabilities or weaknesses that might serve as the main source of harm to them. To defend against potential exploitation of supported flaws, appropriate security controls are selected and implemented with reference to the security policy. Moreover, they can reduce or neutralize negative effects arising from exploitation of weaknesses.

Different risk parameters are identified based on the diagram of figure 3. They are initially determined and then used in the evaluation of incurred losses caused by detected attacks. They include parameters commonly considered by normalized risk assessment methodologies, namely, asset value (AV_i), impact (VIP_i), likelihood ($L_{i,q}$), and exposure ($EL_{i,q}$). Moreover, our risk model explicitly takes account of other risk parameters such as severity of supported vulnerabilities (SV_i) and effectiveness of deployed countermeasures ($ES_{0,i}$). This improvement incorporated in our proposed risk model emphasizes the roles of environment-dependent parameters that were previously neglected or included only implicitly, through key elements, in risk assessment. Furthermore, it allows more precise and objective estimations of risk levels and a more thorough risk management than the existing

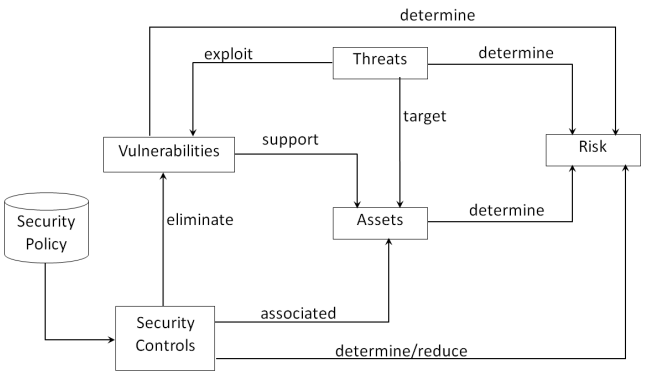


Fig. 3 Dependence diagram of risk components

ALE- (Annual Loss Expectancy) or NIST-inspired methodologies [13, 15]

- Asset value (AV_i) : Asset identification and valuation is a critical step in risk analysis methodologies. It is required for the enumeration and the determination of worth of potential target assets of the computing environment. In our model, a four-class taxonomy inspired by the asset groups in [21] and [22] is adopted. Identified assets of the computing environment are assigned to the hardware, software, or data group, or a combination of these assigned to the system group. Assets of different classes can be valuated using several qualitative or quantitative methodologies as discussed in [22–24]. The valuation process in our model, relies on an economical approach. AV_i is estimated by the difference between the income and the operation cost of an asset, as discussed in [25].
- Impact (VIP_i) : impact of an asset corresponds to the fraction of it potential losses due to the supported vulnerabilities. It determines the magnitude of harm that could result from potential exploitation of target asset vulnerabilities. The global impact of an asset is expressed as a weighted sum of relative impact factors in terms of CIA (Confidentiality, Integrity and Availability) and their importance coefficients, as illustrated by equation 1.

$$VIP_i = (IFC_i \times \alpha_1) + (IFI_i \times \beta_1) + (IFA_i \times \gamma_1) \quad (1)$$

Where:

  - IFC_i, IFI_i, IFA_i : relative impact factors respectively of confidentiality (IFC), integrity (IFI), and availability (IFA) of a_i due to its supported vulnerability set.
  - α_1, β_1, γ_1: the coefficients of importance respectively to CIA requirements of asset a_i,

Relative and global impact factors are determined using scores, given by the National Vulnerabilities Database (NVDB) and the Common Vulnerability Scoring System (CVSS), for all supported flaws of the target asset a_i, $V_i = \{v_j, j = 1, \ldots, n\}$ [26, 27]. Relative impact factors of confidentiality (IFC), integrity (IFI), and availability (IFA) are similarly determined using respectively confidentiality, integrity, and availability impact coefficients in the CVSS vector of every vulnerability $v_j \in V_i$. For instance, the IFC is determined as illustrated by equation 2.

$$IFC_i = \sum_{j}^{n} cif_j \quad (2)$$

Where:

  - cif_j: the confidentiality impact coefficient in the CVSS vector of the vulnerability $v_j \in V_i$.

- Expected potential loss (EPL_i) : the potential loss of an asset a_i due to its supported flaws is determined by the product between its value (AV_i) and impact (VIP_i) of the target asset using equation 3.

$$EPL_i = VIP_i \times AV_i \quad (3)$$

Where:

  - AV_i: the estimated value of target asset a_i
  - VIP_i : see equation 1.

- Threat likelihoods ($L_{i,q}$) : One of the challenging steps in security risk assessment methodologies is the determination of likelihoods of malicious events experienced by different assets of the computing environment. The likelihood of an attack expresses how likely the target asset is to be compromised when an attack (AT_q) takes place. It is approximated by the probability that AT_q *occurs* and *affects the target asset* (a_i). Based on Bayes' theorem of conditional probability, it is determined by the product of the probability of the occurrence of AT_q and the probability that AT_q affects a_i when it occurs. It can be expressed as follows:

$$L_{i,q} = P(AT_q) \times P(AT_q \text{ affects } a_i \mid AT_q \text{ occurs}) \quad (4)$$

$P(AT_q)$ is the probability that AT_q takes place. It is always determined by considering the probabilistic output of the detection mechanism and its certainty coefficient. The probability of occurrence of an attack AT_q is estimated using equation 5

$$P(AT_q) = P(\text{detected attack is of type} AT_q) \times cc_{IDS} \quad (5)$$

  - $P(\text{detected attack is of type } AT_q)$ corresponds to the detection probability of an IDS.
  - cc_{IDS}: the certainty coefficient of an IDS corresponds to its accuracy determined as discussed in [28, 29]

The conditional probability in equation 4 can be estimated by the failure of deployed security controls, $SS_{0,i}$, to protect the target asset a_i against exploitation of supported flaws that accomplish the objective of attack type AT_q (see equation 6)

$$F_{i,q} = \prod_{j}^{|SS_q|} 1 - e_{j,q} \quad (6)$$

Where:

- $SS_{0,i} = \{dc_j, j = 1,\ldots,m\}$: the set of deployed security controls to protect asset a_i against different exploit types.
- AT_q: the detected attack type on the asset a_i
- $e_{j,q}$: the expected effectiveness of security control dc_j against exploits leading to attacks of type AT_q
- SS_q: the subset of applicable controls against attack type AT_q in $SS_{0,i}$,

• Exposure ($EL_{i,q}$) : It expresses the expected damage incurred by the target asset a_i due to the detected attack AT_q. It takes account of potential loss of the target asset owing to supported vulnerabilities as well as the probability of success of a detected attack. It is estimated by the product between the expected potential loss (EPL_i) and the likelihood of the detected attack ($L_{i,q}$) on the target asset (see equation 7).

$$EL_{i,q} = EPL_i \times L_{i,q} \tag{7}$$

• Vulnerability severity (SV_i) : In our risk model, this parameter gives an insight into the potential exploitation of a subset of supported flaws. On one hand, it expresses an attacker's opportunity to expand the attack surface by exploiting a vulnerability subset. On the other hand, it focuses on the extent of incurred damage due to mounted multi-step attacks. It takes account of severity scores of supported flaws grouped into patched, unpatched, and unresolved (SVP, SVNP, and SVNR, respectively). First, flaws of different groups and their gravity coefficients are determined with the assistance of security experts and information available worldwide [30]. Second, relative severity factors of the considered groups are computed on the basis of normalized CVSS scores, as illustrated by equation 8 for SVP group.

$$SVP_i = \frac{\sum_j^{n_p} cvss_j}{n_p} \tag{8}$$

$$SV_i = (SVP_i \times \alpha_2) + (SVNP_i \times \beta_2) + (SVNR_i \times \gamma_2), \tag{9}$$

Where:

- $V_i = \{v_j, j = 1,\ldots,n\}$: the set of supported vulnerabilities by asset a_i,
- $VP = \{v_j, j = 1,\ldots,n_p\}$: the subset of patched vulnerabilities of asset a_i
- $cvss_j$: the normalized CVSS severity score of vulnerability v_j,
- α_2, β_2, γ_2: the weight coefficients that reflect gravities of respectively patched, unpatched and unresolved flaws of asset a_i.

Finally, the severity of supported flaws is expressed as a weighted sum using computed relative factors and their gravity coefficients (see equation 9).

• Security control effectiveness ($ES_{0,i}$) : Damages inflicted on assets of the computing environment due to mounted threats are always mitigated using security controls of multiple categories. For federal organizations, NIST has imposed a minimum required security control collection [14,23,31] and presented also a guideline and a tool, ASSET, for assessing deployed security countermeasures [24, 32]. Moreover, security audit reports provide detailed information about several features that concern security safeguards and their implementation and deployment environments including correctness and strength features [33]. In the proposed risk model, we assume that the assessment of control effectiveness is conducted by security experts using these features and datasets collected from audit reports and control vendors. Each element in the resulting effectiveness matrix is a numerical value ranging between 0 and 1 that expresses the expected efficacy of the concerned security control against the exploitation of flaws of a given group. The effectiveness of deployed security controls ($SS_{0,i}$) is determined based on this matrix using equation 10.

$$ES_{0,i} = \frac{\sum_q^Q 1 - F_{i,q}}{Q}, \tag{10}$$

Where Q attack types can be mounted on the target asset a_i.

4.2.2 Basic risk estimation (BR)

In our risk model, an asset a_i may be a target of a single attack AT_q at a given time point t. The basic risk $BR_{i,q}$ to a_i due to AT_q is determined as illustrated in figure 4, which corresponds to RM_G, the graphical component of the proposed risk model. It is evaluated by combining the exposure, $EL_{i,q}$, the severity of supported vulnerabilities, SV_i, and the effectiveness of the deployed security strategy, $ES_{0,i}$, using equation 11.

$$BR_{i,q} = EL_{i,q} * SV_i * (1 - ES_{0,i}) \tag{11}$$

Furthermore, the basic risk of the computing environment at a time point t is expressed in terms of damages inflicted on its target assets, $A_t = \{a_i, i = 1,\ldots,p\}$, for detected attack AT_q using equation 12.

$$BR_q = \sum_{i}^{p} BR_{i,q} \tag{12}$$

All above discussed equations have been encoded into several algorithms in order to determine the numerical component RM_N of the proposed risk model. Both RM_G and RM_N are involved in designing appropriate security strategies to the computing environment with regard to mounted attack, as discussed in next section.

4.3 Risk treatment

The risk treatment focuses on post-assessment steps, namely, the selection and the implementation of a treatment option. Although various options exist, including avoidance and transference, our response framework takes account of a risk reduction option, as discussed above. The latter involves the design and the selection of optimal response strategies, SS^*, that mitigate incurred risks and reduce security investment cost to within imposed constraints. For such, we propose the following optimization problem: Let be:

- $RR_{k,i}$: the residual risk to target asset a_i due to the deployment of security strategy SS_k against detected attack AT_q, such that

$$RR_{k,i} = BR_{i,q} * (1 - ES_{k,i}) \tag{13}$$

- $ES_{k,i}$: the estimated effectiveness of security strategy SS_k to defend against detected attack AT_q on target asset a_i,
- W_k : the deployment cost of security strategy SS_k, as follows:

$$W_k = \sum_{j}^{l} cost_j, \forall sc_j \in SS_k, \tag{14}$$

  Where $cost_j$ is the cost of security control sc_j
- RR_k : the residual risk to the computing environment due to the deployment of security strategy SS_k against detected attack AT_q, such that

$$RR_k = \sum_{i}^{p} RR_{k,i}, \forall a_i \in A_t, \tag{15}$$

The optimal security strategy SS^* that minimizes both residual risk and security investment cost, as discussed above, is identified by solving the following MOP.

$$\begin{aligned} \underset{SS_k}{\text{minimize}} \quad & \Psi(BR, SS_k) = RR_k + W_k \\ \text{subject to} \quad & RR_k \leq \tau \\ & W_k \leq \beta \\ & SS_k = \{sc_j, j = 1, \ldots, H_k\} \subseteq SC_q \end{aligned} \tag{16}$$

Where τ is the tolerated risk to the target computing environment, β is the allocated security budget, and SC_q is the set of security controls recommended by security experts for defending against an attack of type AT_q, with reference to the security policy and the controls given in ISO 27001 Annex A [34].

To solve this problem, our idea is to transform the previously constructed risk management model into a graphical decision model for reasoning under uncertainty. To approach this, we propose three main phases (see figure 5):

1. *Structure learning phase*: in this phase, we propose to transform the proposed risk model structure RM_G into the graphical component of a multi-objective influence diagram MID_G
2. *Parameter learning phase*: in this phase, we propose to carry out the whole multi-objective influence diagram MID by adding a numerical component MID_N to MID_G
3. *Evaluation phase*: in this phase, we propose to evaluate the obtained MID in order to obtain the appropriate set of security controls SS^*

Before detailing our approach, we provide a review of multi-objective influence diagrams [35].

4.3.1 Multi-objective influence diagrams (MIDs)

MIDs [35] are an extension of classical influence diagrams [36] (IDs) to solve decision problems having multiple objectives; they are among the most commonly used graphical decision models for reasoning under uncertainty. Their success is due to their clarity and their simplicity since their topology (chance node, value node, and decision node) is easily comprehensible to decision makers. Moreover, their evaluation provides the optimal solutions while maximizing the decision makers' utilities. Formally, an influence diagram has two components:

1. The **graphical component** (or qualitative component) is a directed acyclic graph (DAG), denoted by $G = (N, A)$, where A is the set of arcs in the graph and N its node set. The node set N is partitioned into subsets CN, DN, and VN such that

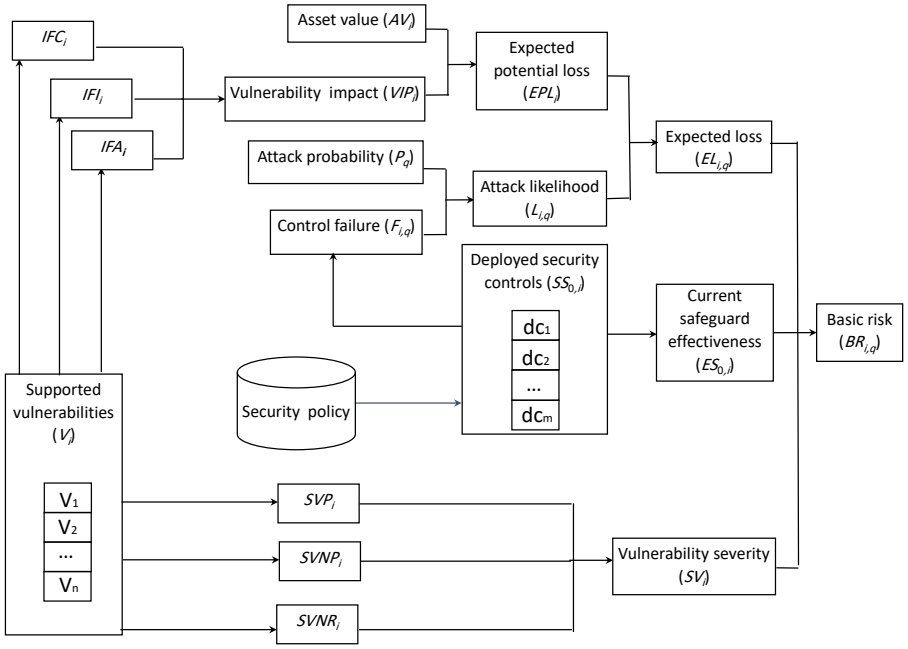


Fig. 4 Risk model: graphical component (RM_G)

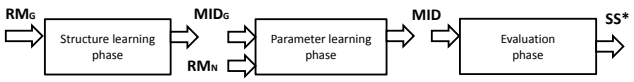


Fig. 5 Risk treatment phases

- $CN = \{CN_1, \ldots, CN_N\}$ is a set of chance nodes, which represent uncertain factors relevant to the decision problem. Chance nodes are represented by circles.
- $DN = \{DN_1, \ldots, DN_H\}$ is a set of decision nodes, which depict decision options. These nodes should respect a temporal order. Decision nodes are represented by rectangles.
- $VN = \{VN_1, \ldots, VN_K\}$ is a set of value nodes, which represent utilities to be optimized; these are represented by lozenges.

Arcs in A have different meanings according to their targets. We can distinguish *conditional arcs* (those directed into chance and value nodes), which are those that have as their target chance nodes representing probabilistic dependencies, and *informational arcs* (those directed into decision nodes), which imply time precedence.
The graphical component of MID encodes different conditional dependencies between chance nodes. Influence diagrams are required to satisfy some constraints to be *regular*; in particular, value nodes cannot have children, and there is a directed path that contains all of the decision nodes. As a result of this last constraint, influence diagrams will satisfy the *no-forgetting* property in the sense that a decision node and its parents should be parents to all subsequent decision nodes.

2. The **numerical component** (or quantitative component) consists in evaluating different links in the graph. Namely, each conditional arc having as its target a chance node CN_j is quantified by a conditional probability distribution of CN_j in the context of its parents. Such conditional probabilities should respect the probabilistic normalization constraints. Thus,
   - If $Pa(CN_j) = \emptyset$ (CN_j is a root), then the a priori probability relative to CN_j should satisfy

$$\sum_{c_{j,k} \in \omega_{c_j}} P(c_{j,k}) = 1 \tag{17}$$

   - If $Pa(CN_j) \neq \emptyset$, then the relative conditional probability relative to CN_j in the context of its parents $Pa(CN_j)$ should satisfy

$$\sum_{c_{j,k} \in \omega_{c_j}} P(c_{j,k} \mid Pa(CN_j)) = 1 \tag{18}$$

Chance nodes represent uncertain variables characterizing the decision problem. Each decision alternative

may have several consequences according to uncertain variables. The set of consequences is characterized by a utility function. In *MID*, consequences are represented by different combinations of parents of a value node. Hence, each value node is quantified by a utility function, denoted by U, in the context of its parents. In general, the definition of the numerical component is constructed by experts and decision makers.

Once the graphical and numerical components of the MID are defined, it can be used to generate the optimal decision, that yielding the highest expected utility, via an evaluation algorithm [35].

4.3.2 Structure learning phase

In this phase, we propose to construct the graphical component of *MID* (i.e., MID_G) from

- RM_G: the graphical component relative to the risk model (RM) with respect to the target computing environment (see section 3)
- SC_q: the set of recommended security controls to thwart an attack of type AT_q
- O_1, O_2: the objective relative to residual risk and that relative to the security investment cost, respectively

Algorithm 1 illustrates the structure learning phase.

Algorithm 1 Structure learning algorithm

Require: RM_G,SC_q, O_1, and O_2
Ensure: MID_G

1. Create a value node VN and add O_1 and O_2.
2. Create several chance nodes relative to: $BR_{i,q}$, $EL_{i,q}$, $ES_{0,i}$, SV_i, EPL_i, $L_{i,q}$, AV_i, The VIP_i, AT_q, AV_i, $F_{i,q}$, $SS_{0,i} = \{dc_1, \ldots, dc_m\}$, IFC_i, IFI_i, IFA_i, $V_i = \{v_1, \ldots, v_n\}$, SVP_i, $SVNP_i$, $SVNR_i$
3. Create a set of decision nodes, each of which is relative to a security control in $SC_q = \{sc_1, \ldots, sc_H\}$.
4. Connect
   - $BR_{i,q}$ to VN
   - $EL_{i,q}$, $ES_{0,i}$, and SV_i to $BR_{i,q}$
   - EPL_i and $L_{i,q}$ to $EL_{i,q}$
   - AV_i and VIP_i to EPL_i
   - AT_q and $F_{i,q}$ to $L_{i,q}$
   - Each dc_j, $j = 1,\ldots, m$, to $ES_{0,i}$ and to $L_{i,q}$
   - Each v_k, $k = 1,\ldots, n$, to IFC_i IFI_i, IFA_i, SVP_i, $SVNP_i$, and $SVNR_i$
5. Connect the decision nodes SC_q by respecting the precedence order.
6. Connect the chance node AT_q to the decision nodes SC_q.
7. Connect each decision node sc_h, $h = 1,\ldots, H$, to the value node VN.

4.3.3 Parameter learning phase

The quantification of different nodes of the designed MID is based on the probability theory and uses Bayes' conditional probability theorem. The evaluation of the conditional probability tables (CPTs) of chance nodes in the MID is conducted using several developed algorithms. Because of the lack of space, these latter can not be illustrated in this paper. Table 3 details how initializing different types of nodes in the designed MID. At this level, the numerical component of the risk model, RM_N, is also involved.

4.3.4 Evaluation phase

In this phase, the generated MID of the proposed intrusion response mechanism is evaluated in order to design and determine the optimal response strategy against detected attack. Various steps of this phase are summarized in [35]

5 Illustrative example

5.1 Experimental environment

The designed response component relies on two main processes, namely, risk assessment and treatment, as commonly recommended by security standards and guidelines [7, 15]. Both processes are performed to design and select the most appropriate security strategy to defend against the detected attack on the target asset. Basic risk assessment steps and different phases of risk mitigation based on our multi-objective optimization approach are illustrated in the example detailed herein. In previous studies [11,20], illustrative examples have simulated their proper test environments. In our example, we have simulated a test environment similar to the most commonly adopted intrusion detection and response testbed of the DARPA evaluation project [39, 40]. The simulated test environment using GNS3 [41] is similar to the DARPA 1999 test bed for intrusion detection and evaluation. Our simulation includes two network segments (see figure 6). The inside segment includes two domain broadcast that correspond to users and victim PCs. The victim domain includes several platforms such as Windows and Linux as well as an inside sniffer. The outside segment supports three domain broadcast namely user PCs, web servers and outside sniffer. Indeed, each domain broadcast has its own IP address. Moreover, inside and outside sniffers serve to collect network traffic. The latter is then preprocessed to be analyzed by the detection mechanism. All domains are involved in setting several parameters

Table 2 Quantification of MID nodes

Node	Type	Quantification
Vulnerability, $v_j, j = 1,\ldots,n$	Barren node	The vulnerability set V_i of the target assets a_i are initially determined using vulnerability scanners such as Nessus or OpenVAS. Their severity scores and impact coefficients are determined by their correspondent vectors in NVDB and OSVDB [27, 37]
Deployed security control, $dc_k, k = 1,\ldots,m$	Barren node	The deployed security strategy, $SS_{0,i}$, to protect the target asset includes different controls. Every control is characterized by a vector of effectiveness coefficients, each of which specifies its approximated efficacy against a flaw class, as enumerated by the CWE and DARPA taxonomies [38, 39].
Asset value, AV_i	Barren node	The values of target assets in the computing environment are determined as discussed in section 1 by the difference between their incomes and operational costs.
Attack type, AT_q	Barren node	The types and the probabilities of detected attacks are determined by the detection mechanism of an IDRS. Most commonly, this mechanism provides a probabilistic output interpreted as the probability of occurrence of the detected attack type. Attack types depend on the adopted attack taxonomy including the DARPA reduced classification [39].
Relative impact factor of vulnerabilities in terms of CIA requirements, IFC_i, IFI_i, and IFA_i	Chance nodes	They are respectively determined based on CIA impact coefficients of every supported flaw with regard to NVDB and equation 1. The CPTs of these nodes in the designed MID are similarly quantified using samples from discovered flaws in V_i.
Vulnerability impact, VIP_i	Chance node	The vulnerability impact factor regarding CIA requirements involves IFC_i, IFI_i, and IFA_i nodes. It is estimated using algorithm equation 2. The CPT of the correspondent chance node is approximated using several samples of supported vulnerabilities V_i.
Expected potential loss, EPL_i	Chance node	It corresponds to potential loss of the target asset due to supported flaws. The CPT of this node is estimated using samples of supported vulnerabilities.
Failure of deployed controls against given attack type, $f_{i,q}$	Chance node	It expresses the bypassing rate of deployed security controls due to mounted attack of given type AT_q (see equation 6). The CPT of the $f_{i,q}$ node is determined using samples of deployed controls, $SS_{0,i}$.
Threat likelihood, $L_{i,q}$	Chance node	The likelihood of an attack AT_q is estimated in our risk model using equation 4. Its CPT for the designed MID is determined using samples of deployed controls and detected attacks
Expected loss, $EL_{i,q}$	Chance node	It expresses damage incurred by the asset a_i due to the detected attack AT_q. The CPT of node $EL_{i,q}$ is determined using samples of supported flaws and types of mounted attacks.
Relative severity factors of patched, unpatched and unresolved vulnerabilities, SVP_i, $SVNP_i$, and $SVNR_i$	Chance nodes	They are determined based on CVSS severity scores, in NVDB, of flaws in patched, unpatched and unresolved groups, respectively. The CPTs of these nodes are similarly quantified using samples of patched, unpatched and unresolved flaws respectively.
Severity of supported vulnerabilities, SV_i	Chance node	It estimates the severity of supported flaws and the opportunity given to the attacker to extend his attack scenario. The CPT of this node is determined based on samples of patched, unpatched, and unresolved flaws in V_i.
Effectiveness of deployed controls, $ES_{0,i}$	Chance node	Its determination takes account of the estimated efficacy of every security control dc_k in the deployed strategy $SS_{0,i}$ (see equation 10). The CPT for this node is estimated using samples of deployed controls in $SS_{0,i}$.
Basic risk of an asset, $BR_{i,q}$	Chance node	It involves three chance nodes, namely, expected loss $EL_{i,q}$ due to the detected attack AT_q, severity of supported weaknesses SV_i, and the effectiveness of the deployed security strategy, $ES_{0,i}$. It is approximated in our risk model using equation 11. The CPT of the $BR_{i,q}$ node in the designed MID is determined using samples of supported flaws, deployed controls and detected attack types.
Recommended security controls, sc_j	Decision nodes	For every detected attack, security experts recommend applicable security controls with respect to the security policy and control of Annex A of ISO 27001. Every decision node in the designed MID is associated with a recommended control and its effectiveness vector against possible exploit classes (see table 5).
Residual risk, $RR_{k,i}$	Value node (O_1)	It corresponds to the remaining unmitigated risk to target asset a_i after the selection of the security strategy to be deployed, SS_k. The residual risk of the target asset due to deployment of the security strategy SS_k is determined by equation 13.
Security investment cost, W_k	Value node (O_2)	It corresponds to the cumulated cost of security controls in the selected security strategy SS_k. It is determined using equation 14.

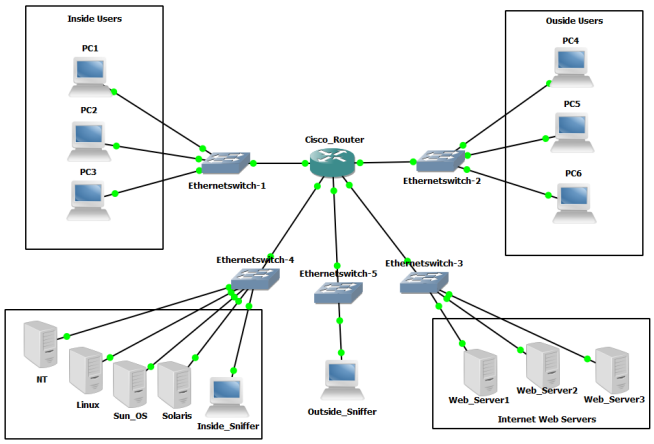


Fig. 6 Simulated test environment

in our response framework, namely supported flaws and deployed security controls, as well as in designing and selecting the optimal response strategy. Furthermore, we assume that they are protected using a basic control subset [23] (see Table 3).In this environment, we have simulated also different attacks with respect to the DARPA reduced taxonomy. In fact, the simulated attacks include DOS, U2R, R2L and probe classes. For instance, the SYN flooding, ps, imap and ipscan are respectively examples of these attacks that were simulated in our experimentation. The simulated traffic was processed by a simple machine-learning-based detection mechanism in order to evaluate certain parameters in our risk model, including the detected attack type and the probability of occurrence [39,40]. The prototyped detection mechanism is based on decision trees that were recommended among the most suitable learning technique for intrusion detection [28].

5.2 Assumptions and example settings

In the example presented, the following assumptions and others are imposed:

- Assets: The insider network segment of the simulated test environment is the single system asset to be monitored and protected. Additionally, the single system asset, a_1 has a value AV_1, of \$2,000,000.
- Attack taxonomy: The DARPA taxonomy with four classes namely denial of service (DOS), user to root (U2R), remote to local (R2L), and probing (Prb) respectively , $\{AT_0, AT_1, AT_2, AT_3\}$, is adopted in our experiment.
- The illustrative example focuses on a truly recognized DOS attack instance by our detection mechanism with a probability $P(AT_0)$ (or P_0) of 0.9.
- Vulnerabilities: Supported vulnerabilities of the monitored asset are determined based on the banners of included software and operating systems. A sample of supported vulnerabilities V_1 is shown in table 4. They are classified based on the following criteria:
  - Exploit objective: This identifies service degradation (0), privilege elevation (1), remote exploitation (3), and information disclosure (4) vulnerability classes.
  - Remediation state: This categorizes vulnerabilities as patched (P), unpatched (UP), and unresolved (UR).
- Security controls: Deployed and recommended security controls to defend against different attack types are determined with respect to rules of the security policy and the Annex A of ISO 27001.
  - Initial security strategy, $SS_{0,1}$, deployed on the target computing environment is reduced and includes controls given by table 3. For instance, the deployed Kerberos authentication system, dc_1 of $SS_{0,1}$, costs \$7500. Its effectiveness coefficients against exploitation of service degradation, privilege elevation, remote exploitation, and information disclosure vulnerabilities are assumed to be 0.1, 0.4, 0.4, and 0.6, respectively.
  - The recommended security control set, SC_0, against the detected DOS attack AT_0 is shown in table 5. Each control's mean deployment cost and effectiveness coefficients are also estimated by security experts.

  In tables 3 and 5, effectiveness of a security control against exploit of vulnerabilities leading to resource privation, privilege elevation, remote exploit or information disclosure are estimated by security experts. Expert methodology to estimate control effectiveness involves several features such as correctness and strength [33]. Information about these features are collected using several reports available to security expert including security audit, benchmarking and vendor reports.
- Importance coefficients of security objectives: For our simulated test environment that corresponds to a military base, security experts recommend confidentiality, integrity and availability objectives. The confidentiality is the most important objective in such context, thus we have associated to it the highest coefficient (.9). Additionally, the importance coefficients of integrity and

Table 3 Deployed security strategy, $SS_{0,1}$ on target asset a_1

Control id	Control title	Mean cost ($)	Estimated control efficacy			
			Resource privation	Privilege escalation	Remote exploit	Information disclosure
dc_1	Kerberos authentication system	7500	0.1	0.4	0.4	0.6
dc_2	Ip filtering	9000	0.4	0.1	0.5	0.5
dc_3	Content filtering	11000	0.2	0.8	0.6	0.5
dc_4	Anti-virus	7500	0.5	0.6	0	0.2
dc_5	Anti-rootkit	12500	0	0.6	0.1	0

Table 4 Supported vulnerabilities V_1 of target asset a_1

Flaw id	CVE	Exploit objective	Remediation state	CVSS Score	Confidentiality impact	Integrity impact	Availability impact
v_1	CVE-1999-1035	0	UP	5.0	0	0	0.275
v_2	CVE-1999-0153	0	UP	5	0	0	0.275
v_3	CVE-1999-0667	0	UR	10	0.66	0.66	0.66
v_4	CVE-1999-1199	0	UR	10	0.66	0.66	0.66
v_5	CVE-1999-0107	0	UR	5	0	0	0.275
v_6	CVE-1999-0016	0	UR	5	0	0	0.275
v_7	CVE-1999-0250	0	UR	10	0.66	0.66	0.66
v_8	CVE-1999-1504	0	UP	5	0	0	0.275
v_9	CVE-1999-0116	0	UR	5	0	0	0.275
v_{10}	CVE-1999-0128	0	UR	5	0	0	0.275
v_{11}	CVE-1999-0377	0	UR	5	0	0	0.275
v_{12}	CVE-1999-1423	0	UR	2.1	0	0	0.275
v_{13}	CVE-1999-0513	0	UR	5	0	0	0.275
v_{14}	CVE-2002-1024	0	UP	7.1	0	0	0.66
...	...	...	...	...	...	...	...
v_{57}	CVE-1999-0151	3	UP	7.6	0.66	0.66	0.66

availability are .7 because they have been considered as important in such environment.

- Gravity coefficients of vulnerability remediation states: To determine severity coefficients of supported flaws, we have identified three main groups of vulnerabilities based on their remediation states. Determined groups include patched, unpatched, and unresolved flaws. The most searched vulnerabilities are those resolved but unpatched in the target environment [30]. Therefore, the corresponding group is associated with the highest gravity coefficient (.6). Patched and unresolved flaws are in general exploited by expert attackers. The latter rarely identify misconfigured patchs to exploit their corresponding flaws, thus the proposed gravity coefficient of this group is .1. However, they are capable of implementing their own attack strategies using unresolved flaws, which are assigned a gravity coefficient of .4. the fixed gravity coefficients for patched, unpatched, and unresolved vulnerability groups are 0.1, 0.6, and 0.4, respectively [30, 42].
- Acceptance levels: The tolerated risk and allocated security budget are initialized to $5000 and $15,000, respectively. In fact, expert recommendations to determine tolerated risk level and security budget rely on standards and guide lines such as FIPS199 and NIST SP800-60 and security best practices [6, 23, 43]. FIPS 199 concerns categorization of information and information systems, whereas, NIST SP800-60 focuses on mapping information and information systems to security categories. They serve to assess impact levels of identified business processes of the target environment. Security expert use then determined impact level to estimate tolerated risk of the target environment. Common methodology adopted by security experts to estimate security spending includes several steps. It starts by identifying security problems associated to physical and technical security. Then, it determines possible threats leading to these problems. For every threat, security ex-

Table 5 Recommended security controls SC_0 against DOS attack AT_0

Control id	Control title	Mean cost ($)	Estimated control efficacy			
			Resource privation	Privilege escalation	Remote exploit	Information disclosure
sc_1	Review of the information security policy	12500.0	0.5	0.4	0.4	0.4
sc_2	Controls against malicious code	14000.0	0.6	0.8	0.8	0.8
sc_3	Controls against mobile code	9500.0	0.7	0.7	0.7	0.7
sc_4	Security of network services	12500.0	0.3	0.7	0.8	0.6
sc_5	Monitoring system use	7500.0	0.5	0.5	0.5	0.0
sc_6	Protection of log information	8500.0	0.2	0.6	0.5	0.6
sc_7	Clock synchronization	3250.0	0.1	0.5	0.5	0.3
sc_8	Review of user access rights	3750.0	0.2	0.8	0.7	0.3
sc_9	User authentication for external connections	7500.0	0.5	0.5	0.9	0.3
sc_{10}	Equipment identification in networks	4000.0	0.3	0.6	0.4	0.3
sc_{11}	Segregation in networks	9500.0	0.4	0.5	0.7	0.2
sc_{12}	Network connection control	10500.0	0.7	0.3	0.7	0.5
sc_{13}	Network routing control	6500.0	0.4	0.2	0.4	0.2
sc_{14}	Secure log-on procedures	12500.0	0.5	0.8	0.7	0.5
sc_{15}	Session time-out	1750.0	0.2	0.8	0.8	0.0
sc_{16}	Limitation of connection time	3000.0	0.5	0.6	0.8	0.0
sc_{17}	Information access restriction	8000.0	0.5	0.7	0.7	0.6
sc_{18}	Sensitive system isolation	8500.0	0.5	0.5	0.5	0.3
sc_{19}	Control of operational software	8000.0	0.4	0.5	0.5	0.5
sc_{20}	Control of technical vulnerabilities	8500.0	0.5	0.7	0.7	0.7

perts estimate its impact on the target asset. Additionally, they determine security spending to restore normal function of the target if this threat takes place. Security experts use also several metrics such as ROSI and collected information from security posture and audit reports to estimate global impact of potential threats and consequently, allocated security budget. Security budget is revised by the senior management depending upon strategic goals of security plan and recommendations of security reports [44, 45].

5.3 Risk assessment

The assessment part of the proposed response mechanism allows two main components, namely, graphical and numerical components. These components are required in order to fulfill the next part, risk mitigation.

- Graphical component (RM_G): With respect to the example settings, the basic risk $BR_{i,q}$ to the target asset a_1 due to the detected DOS attack, AT_0, is determined as depicted by figure 4.
- Numerical component (RM_N): In our example, risk parameters and basic risk $BR_{i,q}$ to the target are estimated as shown in table **??**. Detailed lists of supported vulnerabilities and deployed security controls in our experimental test environment are given in tables 3 and 4, respectively.

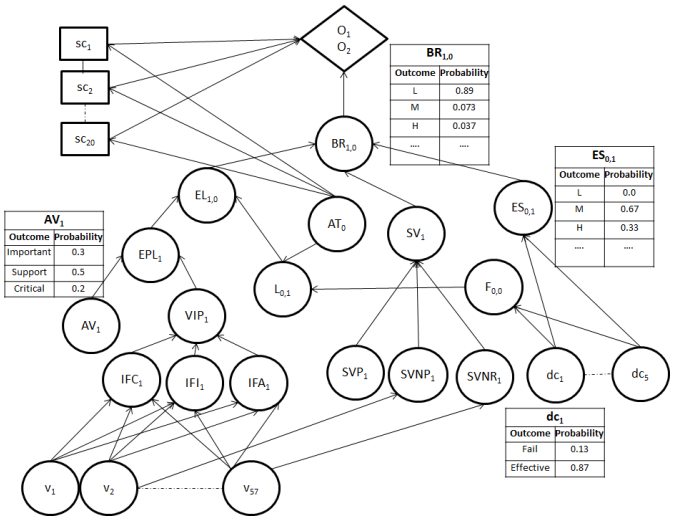


Fig. 7 Learned MID structure and parameters

5.4 Risk treatment

5.4.1 Structure learning phase

As shown in figure 7, **chance, decision, and value nodes are represented respectively by circles, rectangles, and lozenges.** The building algorithm (see Algorithm 1)) adds the objectives (i.e O_1, O_2) in the same value node V. Then, each element in RM_G is created as chance node (i.e $BR_{1,0}$, $EL_{1,0}$, AT_0, SV_1, $ES_{0,1}$, EPL_1, $L_{0,1}$, $F_{0,0}$, AV_1, VIP_1, IFC_1, IFI_1, IFA_1, SVP_1, $SVNP_1$, $SVNR_1$, $\{dc_1 \dots dc_5\}$ and $\{v_1 \dots v_{57}\}$). Then each recommended security controls (see table 5 are created as decision node (i.e ,sc_1, $sc_2 \dots sc_{20}$).

Algorithm 1 connects

- $BR_{1,0}$ to V
- ($EL_{1,0}$, SV_1, $ES_{0,1}$) to $BR_{1,0}$,
- (EPL_1, $L_{0,1}$) to $EL_{1,0}$,
- (AV_1, VIP_1) to EPL_1,
- (AT_0, $F_{0,0}$) to $L_{0,1}$,
- (IFC_1, IFI_1, IFA_1) to VIP_1,
- (SVP_1, $SVNP_1$, $SVNR_1$) to SV_1,
- $\{dc_1 \dots dc_5\}$ to ($ES_{0,1}$, $F_{0,0}$),
- $\{v_1 \dots v_{57}\}$ to (IFC_1, IFI_1, IFA_1, SVP_1, $SVNP_1$, $SVNR_1$),
- AT_0 to $\{sc_1 \dots, sc_{20}\}$
- $\{sc_1 \dots, sc_{20}\}$ to V.

Then, nodes in the RM_G are quantified as discussed in table 3. In fact, the probability table of every chance node is estimated using a sampling technique and identified outcomes [35]. It expresses the probability distribution of the node over its outcomes conditioned on its parents. For instance, the CPT of the chance node $ES_{0,1}$ is determined with respect to its parents namely, $\{dc_1 \dots, dc_5\}$. As example, the probability that the effectiveness of $SS_{0,1}$ is medium (M) known the set of deployed controls $\{dc_1 \dots, dc_5\}$ is .67.

5.4.2 Parameter learning phase

In this phase, datasets concerning vulnerabilities, V_1; deployed security controls, $SS_{0,1}$; the detected attack, AT_0; and RM_N (see table **??**), are involved estimating CPTs of chance nodes of the MID as illustrated in figure 7.

5.4.3 Evaluation phase

The evaluation of the MID allows to determine candidate control combinations regarding the objectives of residual risk, O_1, and of security investment cost, O_2, to defend against a detected DOS attack (see Table 7).

The most appropriate control combination is then identified based on the return on security investment (ROSI) criterion [9]. Indeed, ROSI metric is widely adopted in information security domain to compare several investment strategies. Additional other metrics are also applicable including net present value (NPV) and expected benefit of information security (EBIS) [45, 46]. To determine ROSI of a designed security strategy, we consider both its cost and benefit as illustrated by equation 19. ROSI metric does not consider time value of investment and the NPV metric is commonly applied to this aim. However, in our response framework, detection and response decisions are near real-time, therefor ROSI is suitable to make appropriate decision about designed security strategies. Moreover, the ROSI metric is determined only if

Table 6 Risk model: numerical component (RM_N)

Supported vulnerabilities of asset a_1		$V_1 = \{v_1, \ldots, v_{57}\}$	
Relative impact factors in terms of CIA	$IFC_1 = 0.473$	$IFI_1 = 0.478$	$IFA_1 = 0.419$
Importance coefficients of CIA requirements for asset a_1	$\alpha_1 = 0.9$	$\beta_1 = 0.7$	$\gamma_1 = 0.7$
Impact factor for supported flaws of a_1		$VIP_1 = 0.458$	
Value of asset a_1		$AV_1 = \$2000.000$	
Expected potential loss of a_1 due to supported flaws		$EPL_1 = \$916,287.864$	
Security controls deployed on a_1		$SS_{0,1} = \{dc_1, \ldots, dc_5\}$	
Probability of detected DOS attack on a_1		$P_0 = 0.9$	
Failure of $SS_{0,1}$ against detected DOS attack		$F_{1,0} = 0.216$	
Likelihood of detected DOS attack		$L_{1,0} = 0.1944$	
Expected loss of a_1 due to detected DOS attack AT_0		$EL_{1,0} = \$178,126.361$	
Relative severity factors of patched, unpatched, and unresolved flaws in V_1	$SVP_1 = 0.0$	$SVNP_1 = 0.7255$	$SVNR_1 = 0.6543$
Gravity coefficients for flaw categories	$\alpha_2 = 0.6$	$\beta_2 = 0.4$	$\gamma_2 = 0.1$
Severity of supported flaws V_1		$SV_1 = 0.632$	
Effectiveness of deployed controls $SS_{0,1}$		$ES_{0,1} = 0.8936$	
Basic risk to the asset a_1 due to detected DOS attack		$BR_{1,0} = \$11,970.439$	

incurred risk exceeds the tolerated level.

In our illustrative example, to determine the most appropriate control combination, we first evaluate both objectives for every candidate combination, SS_k, using equations 13 and 14, respectively. Then, the trade-off between O_1 and O_2 is estimated using the $ROSI_k$ ratio:

$$ROSI_k = (BR - RR_k) - W_k/W_k \tag{19}$$

The ROSI ratio, $ROSI_k$, assesses the global effectiveness of a candidate strategy according to its cost. It guides the selection of a cost-effective security strategy for mitigating the risk level currently reached. The most appropriate security strategy, SS^*, is associated with the highest ROSI that exceeds 1. This ensures that returns of selected security strategies always exceed their costs.

In our example, the most appropriate response strategy corresponds to the combination number (id) 49152, associated with an $ROSI_{49152}$ value of 1.51 (see figure 8(c)). As shown in table 5, it consists of the combination $\{ sc_{15} = T, sc_{16} = T \}$. Additionally, it ensures the minimum security investment cost and a residual risk below the tolerated level as illustrated by figures 8(a) and 8(b).

6 Discussion

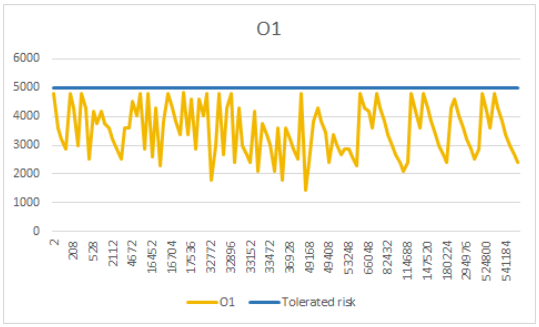
The proposed response mechanism is capable of dynamically taking account of target environment settings. Each time some of the supported vulnerabilities are patched or new security controls are deployed, the learned MID takes account of the new changes. Moreover, it generates several applicable security strategies to thwart detected attacks while respecting imposed constraints. As illustrated by figures 8(a) and 8(b), in our example multiple candidate response strategies that achieve both objectives O_1 and O_2 are generated. Additionally, in table 7 optimal response strategies in terms of O_1 and O_2 are given by the combinations 49152 (yellow row) and 49156 (green row), respectively. To select the most appropriate strategy and comply with the security principle that states that the cost of a security solution should not exceed its benefit, we apply the ROSI criterion. The latter was evaluated for all candidate strategies as shown in figure 8(c). According to this criterion, the optimal response strategy in terms of residual risk given by combination number 49156 is rejected because its ROSI is 0.74, and therefore it does not comply with the above-mentioned security principle. Moreover, by applying the ROSI criterion, we have identified additional acceptable candidates including the control combinations 32896 and 33280 (blue rows), which ensure returns of 1.06 and 1.11, respectively. However, the most appropriate response strategy is associated with the highest ROSI value of 1.51, as shown in table 7. It ensures the minimum security investment cost, \$4750, and reduces the residual risk to below the tolerated level, \$4788.176.

Our response mechanism is capable of reacting against different attack types including DOS, U2R, R2L and probe. However, for the lack of space, we have illustrated only a DOS attack case. Yet, the case of coordinated attack such as DDOS needs some improvements at detection and response levels [47, 48]. On one hand, the detection level requires distributed mechanisms to recognize DDOS attempts. On the other hand, the response level needs appropriate mechanisms able to deploy different security controls against DDOS agents. In this context, our response mechanism is suitable to thwart DDOS attacks if an appropriate implementation approach is adopted. Since DDOS attacks involve several agents, distributed architecture is the most suitable to im-

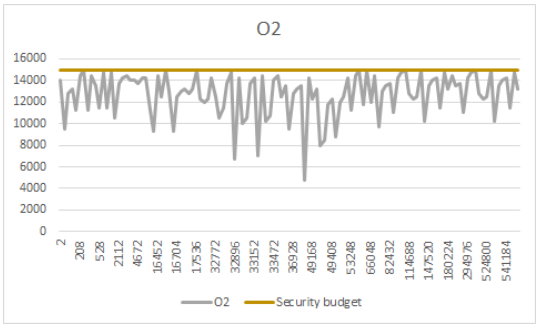
Table 7 Generated candidate response strategies at the evaluation phase of MID

id	O_1	O_2	sc_1	sc_2	sc_3	sc_4	sc_5	sc_6	sc_7	sc_8	sc_9	sc_{10}	sc_{11}	sc_{12}	sc_{13}	sc_{14}	sc_{15}	sc_{16}	sc_{17}	sc_{18}	sc_{19}	sc_{20}	$BR_{1,0}$	ROSI
2	4788.176	14000	F	T	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	H	0.51
...	...	...	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
32864	4309.358	14750	F	F	F	F	F	T	T	F	F	F	F	F	F	F	F	T	F	F	F	F	H	0.52
32896	4788.176	6750	F	F	F	F	F	F	F	T	F	F	F	F	F	F	F	T	F	F	F	F	H	1.06
32912	2394.088	14250	F	F	F	F	T	F	F	T	F	F	F	F	F	F	F	T	F	F	F	F	H	0.67
32960	4309.358	10000	F	F	F	F	F	F	T	T	F	F	F	F	F	F	F	T	F	F	F	F	H	0.77
33024	2992.610	10500	F	F	F	F	F	F	F	F	T	F	F	F	F	F	F	T	F	F	F	F	H	0.86
33088	2693.349	13750	F	F	F	F	F	F	T	F	T	F	F	F	F	F	F	T	F	F	F	F	H	0.67
33152	2394.088	14250	F	F	F	F	F	F	F	T	T	F	F	F	F	F	F	T	F	F	F	F	H	0.67
33280	4189.654	7000	F	F	F	F	F	F	F	F	F	T	F	F	F	F	F	T	F	F	F	F	H	1.11
33296	2094.827	14500	F	F	F	F	T	F	F	F	F	T	F	F	F	F	F	T	F	F	F	F	H	0.68
33344	3770.688	10250	F	F	F	F	F	F	T	F	F	T	F	F	F	F	F	T	F	F	F	F	H	0.80
33408	3351.723	10750	F	F	F	F	F	F	F	T	F	T	F	F	F	F	F	T	F	F	F	F	H	0.80
33472	3016.551	14000	F	F	F	F	F	F	T	T	F	T	F	F	F	F	F	T	F	F	F	F	H	0.64
33536	2094.827	14500	F	F	F	F	F	F	F	F	T	T	F	F	F	F	F	T	F	F	F	F	H	0.68
33792	3591.132	12500	F	F	F	F	F	F	F	F	F	F	T	F	F	F	F	T	F	F	F	F	H	0.67
34816	1795.566	13500	F	F	F	F	F	F	F	F	F	F	F	T	F	F	F	T	F	F	F	F	H	0.75
36864	3591.132	9500	F	F	F	F	F	F	F	F	F	F	F	F	T	F	F	T	F	F	F	F	H	0.88
36928	3232.019	12750	F	F	F	F	F	F	T	F	F	F	F	F	T	F	F	T	F	F	F	F	H	0.69
36992	2872.905	13250	F	F	F	F	F	F	F	T	F	F	F	F	T	F	F	T	F	F	F	F	H	0.69
37376	2513.792	13500	F	F	F	F	F	F	F	F	F	T	F	F	T	F	F	T	F	F	F	F	H	0.70
49152	4788.176	4750	F	F	F	F	F	F	F	F	F	F	F	F	F	F	T	T	F	F	F	F	H	1.51
49156	1436.453	14250	F	F	T	F	F	F	F	F	F	F	F	F	F	F	T	T	F	F	F	F	H	0.74
49168	2394.088	12250	F	F	F	F	T	F	F	F	F	F	F	F	F	F	T	T	F	F	F	F	H	0.78
49184	3830.540	13250	F	F	F	F	F	T	F	F	F	F	F	F	F	F	T	T	F	F	F	F	H	0.61
49216	4309.358	8000	F	F	F	F	F	F	T	F	F	F	F	F	F	F	T	T	F	F	F	F	H	0.96
49280	3830.540	8500	F	F	F	F	F	F	F	T	F	F	F	F	F	F	T	T	F	F	F	F	H	0.96
49344	3447.486	11750	F	F	F	F	F	F	T	T	F	F	F	F	F	F	T	T	F	F	F	F	H	0.73
49408	2394.088	12250	F	F	F	F	F	F	F	F	T	F	F	F	F	F	T	T	F	F	F	F	H	0.78
49664	3351.723	8750	F	F	F	F	F	F	F	F	F	T	F	F	F	F	T	T	F	F	F	F	H	0.98
...	...	...	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
573440	2394.088	13250	F	F	F	F	F	F	F	F	F	F	F	F	F	F	T	T	F	F	F	T	H	0.72

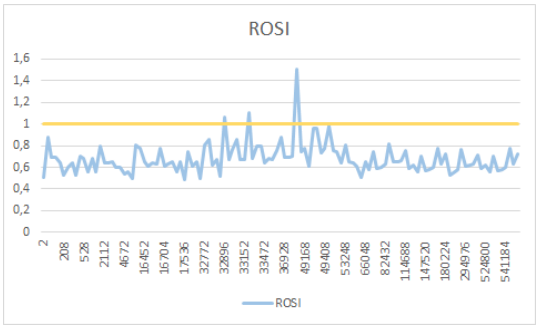
plement our response mechanism. The latter involves two types of mobile agents, namely responder and coordinator agents. A responder agent is deployed on every node of the monitored environment. Additionally, a response plan co-



(a)



(b)



(c)

Fig. 8 Reached results of treated example (O_1, O_2, and ROSI of candidate response strategies)

ordination agent is required at this level. It ensures a coordinated reaction against detected DDOS attack by consolidating all response strategies designed by involved response agents. Moreover, it cooperates with detection agents to send back up to date information about detected DDOS attack to response agents. It is important to note that, security experts are involved in initializing several parameters of our risk model including deployment costs, effectiveness coefficients, and recommended security controls for every node in the monitored network.

Although our dynamic response mechanism is capable of designing and selecting optimal response strategies against detected attacks, as discussed above, we lack a common test environment to compare this mechanism to the existing ones. Indeed, a widely adopted test environment, such as that proposed in the DARPA project framework for detection mechanisms, is also required for response mechanisms. On one hand, it would allow a detailed evaluation of response mechanism capabilities. On the other hand, it would ensure an objective comparison of results obtained to those of existing mechanisms.

7 Conclusion

The designed dynamic response mechanism relies solely on a compliant risk model. To the best of our knowledge, this is the first cost-sensitive mechanism based on a quantitative risk model. The latter takes account of risk parameters, namely the severity of supported flaws and effectiveness of deployed controls, implicitly involved in existing risk models. Additionally, the assessment process involves several parameters dependent on target assets and mounted attacks. It allows a thorough evaluation of risks of the monitored computing environment. Assessed basic risks reflect real security states of the target environment and its assets. Furthermore, they appropriately guide the mitigation process in designing effective security strategies. The mitigation part of the designed response mechanism is formulated as a MOP. Moreover, the MID techniques are adopted to solve the formulated problem. The ROSI criterion is involved as well to select the optimal strategy to mitigate inflicted risk by the target asset. The designed response mechanism takes account of various factors dependent on the computing environment, its assets, and security policy, as well as mounted attacks. Additionally, its mitigation process involves both the damages of mounted attacks and the consequences of the selected controls in designing effective security strategies. Furthermore, it is capable of determining the optimal security strategy to defend against the detected attack as illustrated by the detailed example.

It is possible to improve and conduct detailed tests of proposed risk-driven response mechanism. In fact, the mitigation process of the mechanism could involve additional factors in designing security strategies. Impacts of designed strategies on security policy rules and the normal function of computing environment assets, as well as locations of and conflicts between security controls, are useful factors for revising or designing new strategies. Multiple criteria such as those involved in [1, 11, 20] are useful. In addition, vulnerability graphs are very interesting for designed response mechanism. The assessment process could be based on identified

sequences of exploited flaws, using these graphs to ensure more precise and realistic estimations of incurred damages. Additionally, detailed and well-designed tests are necessary to validate the implemented prototype of the response component. A thorough complexity study of designed processes may also be required for future improvements.

As future work, the false negative (FN), false positive (FP) and misclassified hit (Mis) can be considered in order to improve capabilities of the risk-driven response mechanism. Indeed, they require thorough treatment processes at the detection and response stages. On one hand, a rule-based system that focuses on validation experiments of an IDS and concerns FN, FP, and Mis is obviously needed (extremely useful to our response framework) [29]. It reduces uncertainty about detection decisions of an IDS, and consequently, enhances efficiency of intrusion response systems. On the other hand, FN, FP and Mis costs [29, 49, 50] can be considered in our MOP to improve designed response strategies and decision process. Additional perspectives of this work concerns collateral damage of designed response strategies. In fact the indirect impact of selected security controls on target assets can be considered in our risk model to enhance response cost estimation [8, 17] as well as propose an improved ROSI metric adapted to our response framework) [9].

References

1. S. Anwar, J. Mohamad-Zain, M. Zolkipli, Z. Inayat, B. Khan, S.and Anthony, and V. Chang. From intrusion detection to an intrusion response system: Fundamentals, requirements, and future directions. *MDPI Algorithms*, 10(2):1–24, 2017.
2. N. Stakhanova, S. Basu, and J. Wong. A taxonomy of intrusion response systems. *Int. J. Inf. Comput. Secur.*, 1(1/2):169–184, 2007.
3. N. Anuar, M. Papadaki, S. Furnell, and N. Clarke. An investigation and survey of response options for intrusion response systems. In *Information Security South Africa Conference 2010, Sandton Convention Centre, Sandton, South Africa, August 2-4, 2010.*, 2010.
4. N. Stakhanova, S. Basu, and J. Wong. A cost-sensitive model for preemptive intrusion response systems. In *21st Int. Conf. on Adv. Inf. Net. and App. (AINA 2007), N.F., CA, May 21-23, 2007.*, 2007.
5. P. Porras and P. Neumann. Emerald: Event monitoring enabling responses to anomalous live disturbances. Technical report, SRI International, 1997.
6. ISO. Iso 27002: Code of practice for information security controls. Iso, International Organization for Standardization, 2013.
7. ISO. Iso 27005: Information security risk management. Iso, International Organization for Standardization, 2011.
8. N. Stakhanova, C. Strasburg, S. Basu, and J. Wong. Towards cost-sensitive assessment of intrusion response selection. *Computer Security*, 20(2-3):169–198, 2012.
9. G. Gonzalez-Granadillo, C. Ponchel, G. Gregory Blanc, and H. Debar. Combining technical and financial impacts for countermeasure selection. *CoRR*, abs/1411.0654, 2014.
10. R. Iafarov, R. Gad, and M. Kappes. Improving attack mitigation with a cost-sensitive and adaptive intrusion response system. In *Proceedings of The Fourteenth International Conference on Networks, ICN 2015, Barcelona, Spain,April 19 - 24, 2015*, 2015.
11. M. Khosravi-Farmad, A. Ramaki, and A. Bafghi. Risk-based intrusion response management in ids using bayesian decision networks. In *5th International Conference on Computer and Knowledge Engineering (ICCKE), 2015.*, 2015.
12. T. Toth and C. Kruegel. Evaluating the impact of automated intrusion response mechanisms. In *18th Annual Computer Security Applications Conference, (ACSAC 2002), 9-13 December, 2002.*, 2002.
13. Computer Security Division. Fips 65: Guideline for automatic data processing risk analysis. Technical report, FIPS, 1979.
14. Joint Task Force Transformation Initiative. Sp 800-39: Managing information security risk: Organization, mission, and information system view. Technical report, NIST, 2011.
15. G. Stoneburner, A. Goguen, and A. Feringa. Sp 800-30: Risk management guide for information technology systems. Technical report, NIST, 2002.
16. S.K. Nejat and P. Kabiri. An adaptive and cost-based intrusion response system. *Cybernetics and Systems*, 2017.
17. I. Balepin, S. Maltsev, J. Rowe, and K. Levitt. Using specification-based intrusion detection for automated response. In *Recent Advances in Intrusion Detection, 6th International Symposium,RAID 2003, Pittsburgh, PA, USA, September 8-10, 2003.*, 2003.
18. Y. Wu, B. Foo, B. Blake Matheny, T. Olsen, and S. Bagchi. Adepts : Adaptive intrusion containment and response using attack graphs in an e-commerce environment. In *2005 International Conference on Dependable Systems and Networks (DSN'05), Yokohama, Japan, 28 June-1 July 2005*, 2005.
19. M. GhasemiGol, H. Takabi, and A. Bafghi. A foresight model for intrusion response management. *Computers & Security*, 62:73–94, 2016.
20. A. Shameli-Sendi, H. Louafi, W. He, and M. Cheriet. Dynamic optimal countermeasure selection for intrusion response system. *IEEE Trans. Depen. Sec. Comp.*, PP(99):1–1, 2016.
21. C. Pfleeger and S. Pfleeger. *Security in Computing (4th Edition)*. Prentice Hall PTR, 2006.
22. F. Farahmand. *Developing a Risk Management System for Information Systems Security Incidents*. 2004.
23. Computer Security Division. Fips 199: Standards for security categorization of federal information and information systems. Technical report, FIPS, 2004.
24. Computer Security Division. Fips 191: Guideline for the analysis of local area network security. Technical report, FIPS, 2015.
25. C. Bellefeuille. Quantifying and managing the risk of information security braches to the supply chain. Master's thesis, MIT, 2005.

26. P. Mell, K. Scarfone, and S. Romanosky. *A Complete Guide to the Common Vulnerability Scoring System (CVSS) Version 2.0*. NIST, 1 edition, 2007.
27. National vulnerability database (nvdb). http://nvd.nist.gov/nvd.cfm, 2017. Accessed: 2017-03-30.
28. Sandhya Peddabachigari, Ajith Abraham, and Johnson Thomas. Intrusion detection systems using decision trees and support vector machines. *International Journal of Applied Science and Computations, USA*, 11(3):118–134, 2004 .
29. L. Wenke, F. Wei, M. Matthew, and S. Salvatore. Toward cost-sensitive modeling for intrusion detection and response. *Computer Security*, 10(1-2):5–22, 2002.
30. A. Arora, A. Nandkumar, and R. Telang. Does information security attack frequency increase with vulnerability disclosure? an empirical analysis. *Information Systems Frontiers*, 8(5):350–362, 2006.
31. Joint Task Force Transformation Initiative. Sp 800-53 rev. 3: Recommended security controls for federal information systems and organizations. Technical report, NIST, 2009.
32. Joint Task Force Transformation Initiative. Sp 800-53a rev. 1: Guide for assessing the security controls in federal information systems and organizations: Building effective security assessment plans. Technical report, NIST, 2010.
33. G. Beauchemin and Dansereau G. Harmonized threat and risk assessment methodology. Technical report, 2007.
34. ISO. Iso 27001: Information security management systems requirements. Iso, International Organization for Standardization, 2013.
35. M. Diehl and Y. Haimes. Influence diagrams with multiple objectives and tradeoff analysis. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 34(3):293–304, 2004.
36. RD. Shachter. Evaluating influence diagrams. *Operations research*, 34(6):871–882, 1986.
37. Open source vulnerability database (osvdb). http://osvdb.org, 2017. Accessed: 2017-03-30.
38. MITRE. Common weakness enumeration, 2017. Accessed: 2017-05-10.
39. The 1999 darpa intrusion detection evaluation program. http://www.ll.mit.edu/IST/ideval, 2017. Accessed: 2017-03-30.
40. J. Haines, R. Lippmann, D. Fried, M. Zissman, E. Tran, and S. Boswell. 1999 darpa intrusion detection evaluation: Design and procedures. Technical report, 2001.
41. EPITECH Innovative Project. Gns3.
42. M. Ahmed, E. Al-Shaer, and L. Khan. A novel quantitative approach for measuring network security. In *INFOCOM 2008. 27th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies,13-18 April 2008.*, 2008.
43. K. Stine, R. Kissel, W. Barker, A. Lee, and J. Fahlsing, J.and Gulick. Sp 800-60 rev. 1: Guide for mapping types of information and information systems to security categories. Technical report, NIST, 2008.
44. Barbara Filkins and GM Hardy. It security spending trends. *A SANS Survey. Swansea, UK: SANS Institute*, page 23, 2016.
45. Lawrence A Gordon and Martin P Loeb. The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)*, 5(4):438–457, 2002.
46. Wes Sonnenreich, Jason Albanese, Bruce Stout, et al. Return on security investment (rosi)-a practical quantitative model. *Journal of Research and practice in Information Technology*, 38(1):45, 2006.
47. Huaqiang Wei, Deb Frinke, Olivia Carter, and Chris Ritter. Cost-benefit analysis for network intrusion detection systems. 2001.
48. Saman Taghavi Zargar, James Joshi, and David Tipper. A survey of defense mechanisms against distributed denial of service (ddos) flooding attacks. *IEEE communications surveys & tutorials*, 15(4):2046–2069, 2013.
49. Jalal Baayer, Boubker Regragui, and Aziz Baayer. False positive responses optimization for intrusion detection system. *Journal of Information Security*, 5(02):19, 2014.
50. Sapon Tanachaiwiwat, Kai Hwang, and Yue Chen. Adaptive intrusion response to minimize risk over multiple network attacks. *ACM Trans on Information and System Security*, 19(1-30):95–96, 2002.