Dynamic Risk Management Response System to Handle Cyber Threats

G. Gonzalez-Granadillo[a], S. Dubus[b], A. Motzek[c], E. Alvarez[a], M. Merialdo[d], S. Papillon[b], H. Debar[a], J. Garcia-Alfaro[a]

[a] *Institut Mines-Telecom, Telecom Sudparis, CNRS SAMOVAR UMR 5157 9 rue Charles Fourier, 91011 EVRY, France*
[b] *Bell Labs, NOKIA Route de Villejust, 91625 NOZAY, France*
[c] *Universität zu Lübeck, Institute of Information Systems, Ratzeburger Allee 160, 23562 Lübeck, Germany*
[d] *RHEA Group, Avenue Pasteur 23, 1300 Wavre, Belgium*

Abstract

Responding adequately to possible and ongoing cyber attacks must have the aim to reduce risk on a mission, while not sacrificing a mission for security. Existing approaches often solely consider impacts without considering negative-side effects of missions, or are manually based on traditional risk assessments leaving aside technical difficulties. In this paper we propose a dynamic risk management response system (DRMRS) consisting of a proactive and reactive management software that aims at evaluating threat scenarios in an automated manner, as well as to anticipate the occurrence of potential attacks. We adopt a quantitative risk-aware approach that provides a comprehensive view of the threats, by considering the likelihood of success of the considered threats, the induced impact, the cost of the possible responses, as well as negative side-effects of a response. The proposed DRMRS dynamically evaluates threat scenarios. Responses are selected and proposed to operators based on financial, operational and threat assessments. The DRMRS is applied to a real world use case study of a SCADA environment, to validate the approach in scenarios with multiple threats.

Keywords: Dynamic System, Automated Response, Risk Assessment, Graph Attack, Security Assurance, Cybersecurity

1. Introduction

Efficient information security relies on an exhaustive identification of the scenarios that cause undesired incidents or events on monitored systems. Security standards relying on risk assessment (e.g. ISO 27000 series [1, 2], ETSI TVRA [3], NIST 800-30 [4], EBIOS v2 [5]) often define threat as the possibility for an identified agent or source, to perform adverse actions that exercise a

Email addresses: gustavo.gonzalez_granadillo@telecom-sudparis.eu (G. Gonzalez-Granadillo), samuel.dubus@nokia-bell-labs.com (S. Dubus), motzek@ifis.uni-luebeck.de (A. Motzek), ender.alvarez@telecom-sudparis.eu (E. Alvarez), m.merialdo@rheagroup.com (M. Merialdo), serge.papillon@nokia-bell-labs.com (S. Papillon), herve.debar@telecom-sudparis.eu (H. Debar), joaquin.garcia_alfaro@telecom-sudparis.eu (J. Garcia-Alfaro)

vulnerability on identified assets of an organization. Assets themselves rely on technical parts of interconnected infrastructures, usually referred to as supporting assets. In order to achieve a comprehensive yet accurate mitigation of possible and ongoing attacks on the managed ICT system, we develop a security management system that relies on a response framework that continuously quantifies risks and decides how to respond to cyber-threats that target on the monitored system. This framework is called the *dynamic risk management response system* (DRMRS). This latter is a software implementation of a real world critical infrastructure system that has been replicated in order to have a total control of all the assets of the organization. The DRMRS consists of over 13,000 nodes categorized as entry points (e.g., remote terminal units or devices that are the first target point of entrance for an attack), intermediate nodes (remote terminal units or devices that are used to reach business devices), and business devices (e.g., organizational critical devices).

The aim of our proposal is to provide improved protection of critical infrastructures against cyber-physical attacks. We assume upgraded infrastructures with novel Information and Communications Technology (ICT) capabilities, e.g., in terms of computing, communication and interconnection capabilities. This adoption of upgraded features comes at the cost of introducing new threats that are required to be holistically handled both in terms of safety and security (in the traditional ICT sense). This interaction requires to conciliate several protection disciplines from both safety and security areas (e.g., from service continuity and recovery requirements to threat analysis and defense in depth techniques). Without loss of generality, we report in this paper a first attempt that combines well established techniques from both areas: risk assessment and automated ICT response management to complement traditional techniques in terms of safety analysis and reliability engineering in Industrial Control Systems (ICS). Without underestimating the safety and reliability literature (with well established techniques and solutions world-wide deployed over many decades) we focus in this paper on the ICT security counterpart of the proposal.

Traditional risk assessments are rather organizational (business-aware) than technical, and enable security officers to manage risks on the long run. However, both ICT systems and threat landscape do not cease to evolve, and dynamic cyber security management becomes paramount to address potential breaches. In the process of risk assessment, physical-critical systems consider the trade-off between cost, likelihood of occurrence, and potential consequences. This argument applies equally well for cyber-critical systems, where failure is replaced by attack, and safety is replaced by security. The operational security management is based on technical processes, executed by administrators who are not necessarily aware of organization's business and strategic aspects. This gap between technical and organizational levels renders traditional risks assessment methods cumbersome and obsolete. Our DRMRS leverages a novel concept of elementary risk (ER) that represents a quantum of risk for an organization. Composite risks (CRs) enable dynamic calculation of risk posture while considering the system's state [6].

In cyber security domains, attack scenarios are frequently represented by the use of attack graphs. Various kinds of attack graphs have been proposed in the scientific literature in order to represent at an abstract level (i.e., not a specific occurrence of an attack scenario, but rather a template of a possible multi-steps attack) scenarios composed of several elementary attack steps [7, 8, 9, 10]. However, in order to compute the most exhaustive list of possible attack scenarios, attack graphs must rely on algorithms that base their processes on an up-to-date knowledge of the network connectivity between each piece of equipment of the monitored system. In addition, the algorithm must base its process on the knowledge of an up-to-date vulnerability inventory (i.e., a list of all pieces of equipment in the infrastructure with the exhaustive list of its current vulnerabilities).

Our framework is designed to be schematically decomposed in three stages (i) situation awareness (exploiting attack modeling), (ii) risk assessment (exploiting risk modeling), and (iii) response assistance (exploiting response modeling). It is important to recall that in a proactive chain, the DRMRS evaluates possible attack scenarios before they originate in the system. As such, vulnerabilities and mitigation actions are modeled and analyzed through simulation or projection of data, making it possible to identify the best response set of actions that would be triggered after an attack is detected and the reactive mode is activated. Eliminating a vulnerability by either applying a patch, or implementing corrective actions is not required in the proactive chain. Instead, we provide an exhaustive evaluation of all possible mitigation actions (and their combinations) that can be implemented as a response to a specific attack scenario. The optimal response plan is selected and will only be implemented if the threat is realized in the system.

The contributions on this paper can be summarized as follows:

- A dynamic and automated response framework that integrates operational, financial, and threat impact models to evaluate and select response plans for a given threat;
- A prototype integrating the various components of the dynamic risk management response system;
- A quantitative model for the assessment and selection of mitigation actions;
- A mission impact assessment pondering the potential negative side-effects of responses.
- A fully integrated system to generate enriched response plans based on particular threat scenarios;
- Testing and experimentation of the integrated DRMRS prototype that enable verification of its interfaces regarding both behavior and correctness of the processed data;
- The deployment of the DRMRS over a SCADA environment.

This article extends previous work [11]. The original work has been significantly extended by an in-depth description of the complete system, and, most importantly, a complete extension towards an automated incident handling of multiple threat scenarios. Moreover, we evaluate proactive and reactive behaviors of the system in one consistent use case involving real world experts on real world data.

Unlike early research works that consider only the impact, our response system aims at minimizing the risks of both possible and ongoing attacks, as well as a potential negative side-effect of our response itself. We adopt a quantitative risk-aware approach that provides a comprehensive view of the threats, by considering, (i) the likelihood of success of the considered attacks, (ii) their induced impact, and (iii) the cost and impact of the possible responses. Moreover, the proposed response system aims at dynamically managing the security of a monitored ICT system proactively (i.e., before an attack is detected) and reactively (i.e., after an attack is detected).

It is important to highlight that in critical infrastructure systems, even though safety engineers have a better ability to identify failure modes through stress testing than security engineers may have, safety engineers do not know the complete set of system and component-level failure modes. Our research is therefore limited to identified threats and known vulnerabilities that, if exploited, may lead to undesirable consequences on the system (e.g., the execution of an attack). We exclude

from our research unknown threats and/or unknown vulnerabilities that lead to the execution of known/unknown attacks.

The remaining of the paper is structured as follows: Section 2 describes our proposed dynamic risk management response system (DRMRS) and details the input and output data. Section 3 details the functional DRMRS modules. Section 4 discusses the integration of the DRMRS into a SCADA environment. Section 5 presents tool testing and experimentation. Section 6 presents related work. Conclusions and future work are presented in Section 7.

2. Dynamic Risk Management Response System (DRMRS)

The dynamic risk management response system is composed of two complementary chains of treatment: proactive and reactive risk management, which are further refined in three stages: (i) attack modeling; (ii) risk modeling; and (iii) response modeling. A simplified diagram of the functional architecture of a dynamic risk management response system is presented in Figure 1.

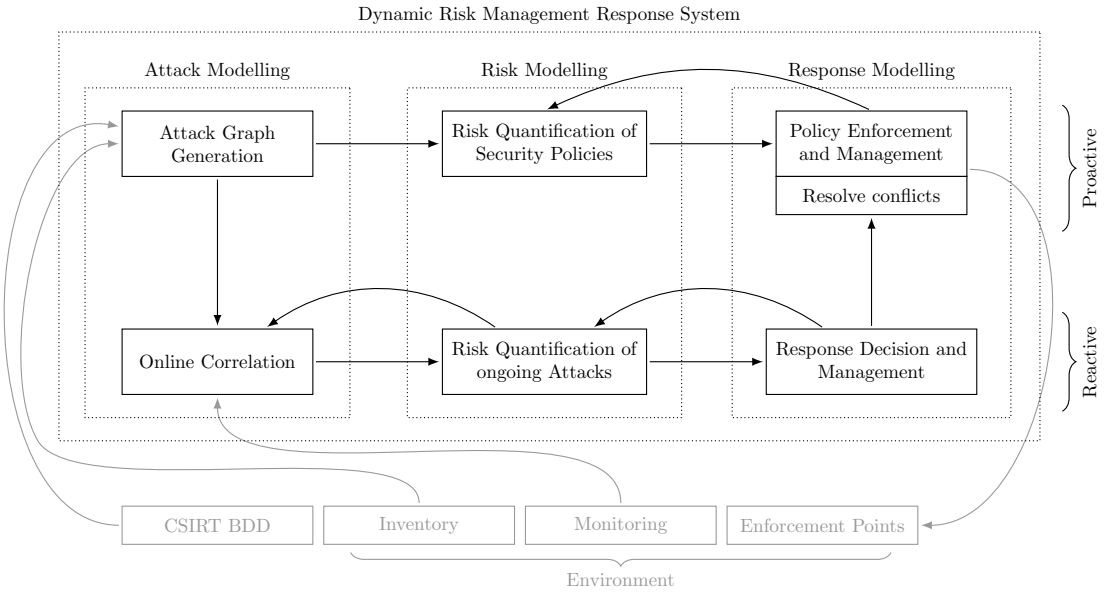

Figure 1: Diagram of the dynamic risk management response system (DRMRS) of ICT Systems. Note that the complete DRMRS is based on local and global feedback loops.

The generic high-level functions that enable the three stages of a dynamic risk management are the following:

- Attack Modeling: a situation awareness function that is defined as the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future [12]. In our perspective, this is interpreted as the regrouped features enabling a manager of monitored or protected ICT systems to grasp the global threats and current ongoing attacks weighing on its systems/businesses/missions at any time, and to project that status into the near future using attack graphs and online correlation tools.

- Risk Modeling: using a risk assessment approach, it identifies the features that enable a manager to qualify and quantify the risks of any identified threat or ongoing attack (risk quantification of ongoing attacks); as well as, the possible impacts and costs on the systems/businesses/missions that should result from the activation and deployment of possible response mechanisms that mitigate these risks (risk quantification of security policies).

- Response Modeling: the use of response modeling encompasses the features that propose response possibilities to mitigate the identified risks, enable choice of the most suitable response possibilities to reduce the identified risks below an admissible level (policy enforcement and conflict management), and then compute the mitigation actions to be deployed on monitored or protected ICT systems (response decision and management).

The proactive risk management chain of treatment consists of a policy-driven management of potential risks during the life cycle of the system. It assumes that the existence of vulnerabilities in the system can lead to events specified in a security policy as prohibitions. Such events are interpreted as potential attack scenarios that must be handled before they become active instances of policy violations. As soon as a new update during the risk evaluation process is reported, this chain triggers an automated process to guide security officers in the tasks of updating and deploying new security measures. The goal is to anticipate the occurrence of potential attacks, faiures, or any kind of security issues, and enforce the appropriate response plans.

Active monitoring of vulnerabilities and exchange of diagnosis reports are provided to this chain. These vulnerabilities and reports are mapped to previously defined strategic policies, specified by the security officers of the organization. Policy updates are achieved by enabling conditions already specified in the strategic policies. The conditions are specified in terms of logical predicates that can either be enabled or disabled. The resulting procedure establishes the link between preventive rules and risk evaluation, so that a potential vulnerability exploitation raised as likely activates the enforcement of system reconfiguration before policy violations occur.

The reactive risk management system acts complementary to the proactive chain, in which some of the risks might be accepted without being eliminated. However, during operations, some of the detected attacks may occasionally see their risk arise to an unacceptable level (e.g. when the likelihood of an attack increases drastically towards its success, for instance, when the attacker exploits a zero day vulnerability). It may also happen that new opportunities appear for some attackers which open new means to threaten critical assets of the monitored ICT system (e.g. if the filtering rules of a compromised firewall are modified by an attacker, or the measurements of the system are tampered by the attacker). This kind of situation stresses the need for a reactive chain used to continually and dynamically manage instantiated risks.

In contrast to the strategic scope of the proactive risk management system, the reactive chain refers to risk management after the detection of an ongoing attack. Consequently, an insightful decision to deploy an efficient *tactical* response has to be determined in order to eliminate, or mitigate, the ongoing attack. Possible conflicts must be managed to produce a security policy instantiation that complies with both strategic and tactical response requirements before enforcement.

Being an extremely broad field of research, we limit the scope of this article in the following to details on the components and behaviors of the proactive chain. Nevertheless, fundamental building blocks remain directly equivalent and adaptable to the reactive part, as discussed in future work.

2.1. DRMRS Architecture

Our proposed dynamic risk management response system (DRMRS) has a modular architecture composed of several integrated elements that interact to evaluate, assess and mitigate the impact of identified threats. Components are grouped according to their functions in three main blocks: (i) input data, that provide the information required to analyze the impact of threats and mitigation actions on the target system; (ii) the processing modules, composed of the strategic response decider (SRD), the attack graph generator (AGG), the response operational impact assessment (ROIA), and the threat risk quantifier (TRQ), whose mission is to evaluate individual and combined mitigation actions in financial and operational perspectives in order to generate the corresponding response plans; and (iii) the output data, composed of the primitive and/or enriched response plans, which indicate the security actions to be deployed in order to mitigate the current threats. The remaining of this section details each DRMRS component.

The DRMRS is a dynamic system that involves information coming from different sources of the system environment (e.g., abstract security policies, network inventory, authorized mitigation actions), as shown in Figure 2.

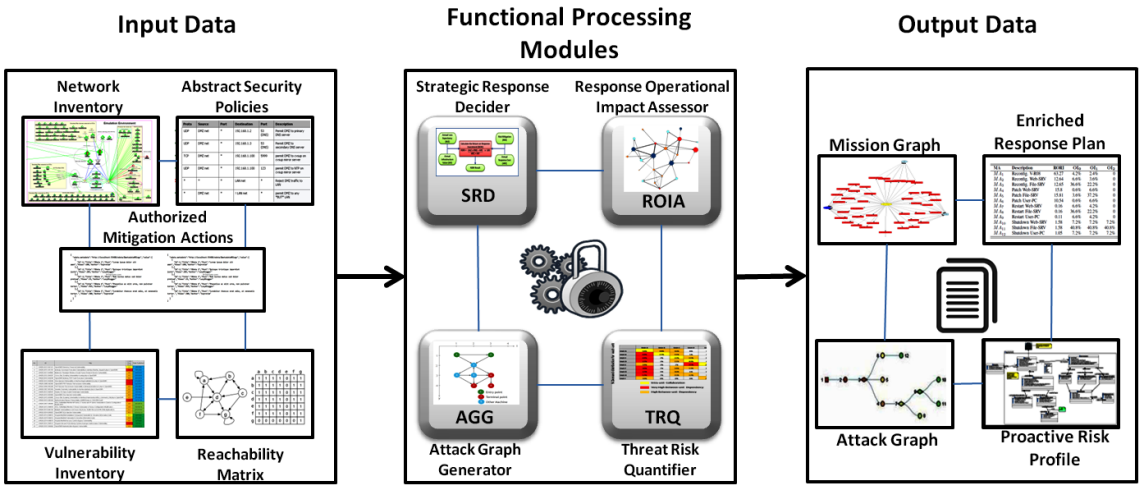

Figure 2: Dynamic risk management response system (DRMRDS) conceptual workflow on acquiring, processing, and generating data.

The DRMRS automates the process of activating/deactivating policies. Responses are activated when new threats are reported and the cost of deploying an optimal response (provided by the financial impact assessment as contextual data, then validated by human officers guided by the operational impact assessment) is acceptable. The system can also deactivate previous strategic or tactical actions (e.g., when threats have vanished, or the response costs, in terms of financial or operational costs, is higher than executing no action at all). The remaining of this section details each module of the DRMRS.

2.2. DRMRS Input Data

The DRMRS accepts input data using JSON encoded files [13] that are required for the system to generate the response plans against specific threat scenarios. An extraction of a JSON file is presented in Listing 1.

Listing 1: JSON File Example

```
{"monitored_System_Ident": "",
 "snapshot_Ident": "03/14/2016 15:15:54",
 "riskProfile_Ident": "11bd2fa4-c390-4fc2-819d-5c776880f01a",
 "attackPathElementaryRisks": [
  {"riskprofileattackpath":
     {"ident": "1",
     "likelihood": "0.193"},
   "elementaryrisk": [
    {"relatedDetrimentalEvent": "DE-BP-M-C",
          "impact": "0.1",
          "risk": "High"},
         {"relatedDetrimentalEvent": "DE-BP-H-C",
            "impact": "0.55",
            "risk": "Extreme"}]]
 },}
```

Examples of input data are: Network Inventory (as a database that stores all information about the devices of the system); Abstract Security Policies (as a configuration file that includes detailed information about threat scenarios); Authorized Mitigation Actions (as a configuration file that details information about mitigation actions); Reachability Matrix (as a configuration file that contains information about the communication among devices in a network topology); and Vulnerability Inventory (as a database that stores the information about the vulnerabilities associated to each devices on the system). The remaining of this section details each input file.

2.2.1. Network Inventory

The network inventory is an application that collects and stores topology information in a database at high speed (due to its internal parallel computation). It can be deployed from small to large size networks, maintaining interesting speed results, as well as in multiple instances on different IP domains, centralizing the results on a common database. The Network Inventory contains information of all active devices of the monitored system. For each device we have a unique identifier (ID), the type of device (PEP_Type), the annual equipment cost (AEC), the associated vulnerabilities (CVE), protocol information, etc. The sum of the AEC for all devices result into the annual infrastructure value (AIV). This latter is a dynamic value used to compute the RORI index.

2.2.2. Abstract Security Policies

Abstract security policies include information required for the instantiation of security policies (e.g., information of threats, policy enforcement points (PEP), and mitigation actions of a given organization). In this sense, a threat scenario is described by an attack vector, which establishes the target business devices (and the possible entry points and middle targets that could lead to them) that, in the event of an attack, are used to determine and differentiate the possible threat

scenarios occurring in the system. An attack vector is composed of the possible steps towards one or several business devices; each step has one or more PEP_Type under which security devices are grouped.

Each threat defined in an abstract default security policy is also described by a threatID, severity, frequency, and a likelihood threshold. Such threshold allows the instantiation process to determine that detrimental events with a likelihood value above the threshold are considered as involved in a Threat Scenario.

In addition, a policy includes details of the policy enforcement points, e.g., name, annual equipment value, PEP_Type, quantity, and weighting factor. This latter is a numerical value (generally from zero to ten) that represents the contribution of a given PEP in the execution of an attack. Such a value is used to compute the coverage of the mitigation action over a particular threat.

Finally, the file includes information associated to the mitigation actions (e.g., ID, annual response cost, nodeID, restrictions, and effectiveness) that is required to evaluate the Return On Response Investment (RORI) index in the security policy instantiation process.

2.2.3. Authorized Mitigation Actions

It contains a list of mitigation actions that are authorized to be executed as a response to a given threat. We specified a format that aims at offering a language to structure the knowledge that defines the characteristics of mitigation actions, that are *a priori* authorized on devices of the monitored system, in order to enable modules of the DRMRS (i.e., strategic and tactical response deciders) to compute response plans.

Each mitigation actions is defined by (i) an identifier (e.g., "FirewallRule_on_dorete_filtering_ftp"); (ii) a scope of enforcement, indicating whether it is applicable for *strategic* or *tactical* response; (iii) the enforcement points on which it is permitted to be deployed; (iv) the annual equipment value (associated to the PEP); (v) the annual response cost (associated to the mitigation enforcement of the action).

Authorized mitigation actions allow modeling precise actions to be executed. For instance, patching a specific vulnerability (corresponding to a specific vulnerability identifier in the vulnerability inventory) on a given device (corresponding to its node identifier in the network inventory and reachability matrix). In addition, the format allows representing abstract mitigation action kinds. For instance,it is possible to specify a mitigation action representing a firewalling rule which prohibits a range of source and destination ports between one or more IP addresses, and whose implementation involve multiple changes on several firewall devices of the monitored system.

2.2.4. Reachability Matrix

The aim of the reachability matrix formalism is to provide the attack graph generator (AGG) functional module of the DRMRS with the knowledge of the communication possibilities (i.e., at a logical network level) between devices of the monitored system for which the possible attack graphs must be generated. This format provides an exhaustive view of the communication among devices in a network topology. The knowledge not only provides a view of the logical routing possibilities, but also encompasses the knowledge of, e.g., a filtering policy currently deployed in the topology.

For a TCP/IP network, the reachability matrix carries the knowledge that could be represented by a two dimensions matrix, with each column and row being devices' IP addresses. In this matrix, each cell carries the information of the IP ports and protocols authorized between the couple of IP addresses.

In order to save room during storage and in the memory of our prototypes while processing the knowledge, a dedicated structured format, following a connectivity graph approach, has been defined to represent the information of the Reachability Matrix in a more compact way compared to a matrix. Even if more efficient than a matrix, the representation has not been designed for highest compactness, but rather to ease the further process of computing an attack graph with graph algorithms in the attack graph generation (AGG) functional module. Nevertheless, this connectivity graph representation enables grouping several nodes of a network infrastructure sharing the exact same connectivity characteristics, to reduce the size of data representing the Reachability Matrix.

Moreover, the proposed structure models each node of the monitored system by considering: (i) its communicating nodes, expressing, for each protocol or port, through which routing paths other nodes can reach it; and, (ii) a list of vulnerability existing on the node, using their unique identifier in the vulnerability inventory.

2.2.5. Vulnerability Inventory

A vulnerability inventory provides, in a structured format, the attack graph generation (AGG) functional module of the DRMRS with the knowledge representing possible elementary attack actions on devices of the monitored system.

It describes, at an abstract level, the characteristics of elementary attack actions corresponding to the exploitation possibilities of each vulnerability that could exist on devices of a monitored system, using a pre-condition/post-condition formalism. For each described elementary attack action, the pre-condition expresses the requirements that would allow the exploitation of a vulnerability by an attacker. Symmetrically, the post-condition explicitly expresses consequences produced after the successful exploitation of the vulnerability.

Several attack action models using this formalism, enabling reasoning with logical inference using first or second-order logic, has been proposed in the literature [14, 8, 15]. Also syntactically different, our modeling approach is very similar to the Lambda language [16], but it uses a restrictive set of possible predicates for post-condition and pre-condition terms: (i) vulnerability(Host, Reference), indicating a given Host is vulnerable to the vulnerability identified in the Reference set; (ii) reachability(Host1, Host2, Protocols, Ports), indicates Host2 can be accessed or exploited from Host1 with the protocols and ports given in Protocols and Ports sets; and, (iii) privilege(Host, Level), indicates the current privilege Level, between *User* and *Root*, of an hypothetical attacker on a Host.

Also unlike the other approaches which explicitly model attack actions, the vulnerability inventory rather models each known vulnerability of the monitored system. It gives first, identification information for the considered vulnerability: specifically, a unique identifier; and, *references* to complementary identification on the considered vulnerability (e.g. other identifiers, links on information page of public vulnerability databases). Then, *requirements* and *consequences* are defined.

The elementary attack actions are inferred by crossing the vulnerability models with the list of identifier of vulnerabilities found for each *node* of the monitored system in the reachability matrix.

2.3. DRMRS Output Data

The DRMRS produces output data as JSON files that are generated by the functional processing modules of the DRMRS. Examples of these files are the attack graph, the mission graph, the proactive risk profile and the enriched response plan. The remaining of this section details each proactive output file.

2.3.1. Mission Graph

A mission graph is a JSON representation of information provided by and obtained from the mission dependency model and resource dependency model. Therefore, the mission graph describes the business model of an organization, i.e., it identifies each business function, which are the assets with a value for the business of an organization. Further, it associates to each of these assets one or several business processes, which are crucial for the economical and operational success of the organization. Any failure of these processes will lead to an impact onto the company, mission or organization. Further, each business process is associated with natures of Feared Events are defined using the classical security axes: confidentiality, integrity and availability. If a business resource linked through the business model is compromised with an attack that induces this kind of impact, it means that it could cause inability to accomplish the mission of the related business process. Attacks towards these business resources are likely to origin from prone devices, i.e., entry points, which are identified for each modeled mission. Further, every business process is associated with a consequence, representing the economical, operational or environmental consequence of a failure of said business process.

Most importantly, a mission graph carries information about the probability of global operational impact on the mission originating from widespread effects causing local impacts. The DRMRS utilizes this information to assess the global operational mission impact caused by proposed response plans, i.e., to assess the self-inflicted probability of negative side-effects and collateral damage due to executed actions. For example, it prohibits the critical consequences of shutting down all central control servers to mitigate a potential attack. Inside the DRMRS this information is produced by the response operational impact assessor (ROIA, Section 3.3) based on the underlying mathematical models, i.e., mission dependency model and resource dependency model.

2.3.2. Attack Graph

Attack graphs are a classical way used in security to represent the exposure of a system. Many proposals have been made in the literature to formalize attack graph models and propose generation algorithms for the IT domain [7, 8, 9, 10]. It is a convenient and compact way to aggregate the various attack paths depicting possible attack scenarios in a system.

In our DRMRS, the attack graph generation functional module is responsible for generating attack graphs starting from predefined Entry Points, and reaching business devices (i.e., critical devices) of the monitored system. Details on the generation of proactive attack graphs are given in Section 3.1.

2.3.3. Proactive Risk Profile

A proactive risk profile contains structured information representing the risk posture of the monitored system on the mid-long term. It is produced at the output of the Threat Risk Quantification functional module. The proactive risk profile includes information about attack scenarios, detrimental events, the risk associated to each detrimental events and the contribution of each attack scenario to the risk of each detrimental event (a detrimental event being defined as the fact of harming the accomplishment of an organization's objective or mission, i.e., a business process). Details on the various processes to construct a proactive risk profile are given in Sections 3.2, 3.2.1, 3.2.2 and 3.2.3.

2.3.4. Enriched Response Plan

An enriched response plan contains detailed information about the actions that best mitigate the threat scenarios. A primitive response plan (based on the financial assessment) is generated by the RFIA module. An enriched response plan (based on the financial and operational assessment) is generated by the ROIA module. Details on the response plan generation is given in Sections 3.4.2 and 3.4.3.

3. DRMRS Functional Modules

The Dynamic Risk Management Response System integrates software components that map detrimental events and attack graph evidences to potential attack scenarios and proactive conditions that were previously defined as strategic policies by the security officers of the organization. The goal is to anticipate the occurrence of potential attacks. The functional modules responsible for the generation of the response plans are (i) the attack graph generation, whose purpose is to calculate the exposure of the monitored system to threats; (ii) the threat risk quantification (TRQ), that aims at assessing the risk of each detrimental events feared by the organization, that uses the monitored system to support its business processes; (iii) the response operational impact assessor (ROIA), whose objective is to assess potentials of collateral damage onto a company and their associated business processes; and (iv) the strategic response decider (SRD), that analyses response plans composed of actions that mitigate the current situation.

3.1. Attack Graph Generation (AGG)

The formalism used by the AGG functional module to represent the exposure to threats is based on attack graphs, as introduced in Section 2.3.2. According to characteristics commonly used in the literature to classify attack graphs [17], those generated by our prototype are topologically directed graphs. This means each directed edge represents the attack action from the tail node towards the head node. Each node corresponds to the compromising status of a specific device of the monitored system. Two possible levels of compromising of a node are considered in our attack graph model: (i) $User$, when attack actions result in compromising a node with normal user's privilege; (ii) *Root*, when attack actions result in gaining the highest administrative privilege on a node.

On the proactive perspective, AGG implements a feature that computes attack graphs corresponding to all potential attack scenarios for the monitored system, regardless of the fact that there could be ongoing attack attempts in the monitored system. To generate an attack graph, the AGG bases its computation on three kinds of inputs: (i) the reachability matrix of the monitored system; (ii) the vulnerability inventory; and (iii) the mission graph.

The AGG approach follows a semi-explicit correlation technique [7, 8, 9, 10, 18], to model elementary attack actions and the logical inference among them as pre-conditions and post-conditions. The AGG approach does not rely on the explicit description of attack scenarios. Instead, it enables generating complex attack scenarios composed of several elementary attacks exploited in sequence, which represents, nowadays, the most threatening Advanced Persistent Threat (APT) in the ICT domain. However, unlike other semi-explicit approaches (e.g. [18]), that enable any kind of predicates to describe pre and post conditions, we restrict the predicates to three kinds: (i) predicates indicating the connectivity between nodes, (ii) predicates about the existence of a vulnerability on a node, and (iii) predicates indicating the level of privilege of a fictitious attacker on a compromised node. The elementary attack actions and their predicates are extracted and derived by the AGG functional module from the above mentioned structured input data.

As the main purpose of the AGG functional module is to transform a logical representation of possible attack actions in the monitored system in a directed graph (DG), it enables generating exhaustively the attack paths based on the attack graph, using efficient path search algorithms in DG (e.g., bread-first or depth-first search). Attack paths composed of complex chains of compromised devices are considered as possible attack scenarios.

Moreover, in order to restrict the predicates dictionary to three kinds, and thanks to the proposed attack graph model that decomposes the compromising status of each devices with the gained privilege (i.e., *Root* or $User$), the AGG functional module produces non-monotonic attack paths (i.e., attack paths with backtracking on already compromised nodes to escalate the compromising status from an unprivileged $User$ level to a privilege *Root* level) with the cost of the most efficient path search algorithm (i.e., $O(E + V)$, where E is the number of edges and V the number of node in the attack graph). These characteristics are significant differentiation of our approach over the wide majority of proposed attack path search solutions in the literature.

In our DRMRS prototype, AGG is a part of a unified software component (i.e., AGG-TRQ), which addresses both attack graph generation (AGG) and threat risk quantification (TRQ) functional modules on the two management perspectives. On the proactive perspective, AGG is responsible for (i) producing proactive attack graphs as soon as modifications occurs in the monitored system; (ii) sharing this information to the TRQ part; and (iii) communicating proactive attack graphs to the strategic response decider (SRD) software components.

3.2. Threat Risk Quantification (TRQ)

The exhaustive generation of all possible attack paths from the proactive attack graph is useful for the threat risk quantification (TRQ) functional module to compute the proactive risk profile. The TRQ functional module implements a feature that assesses a risk weighing on an Organization. To achieve this, the component exploits the business model of the organization expressed in the mission graph together with the list of all possible attack paths generated from the AGG.

The TRQ is responsible for (i) producing a proactive risk profile as soon as modifications occur in a attack graph, and (ii) communicating the proactive risk profile to the SRD module so that (together with the attack graph) the best response plan is derived.

The mission graph describes the business model of an organization including the consequences of potential impacts on business processes. For each business process, a consequence expresses the impact that it would incur in case it is unable to accomplish its mission, on a qualitative impact scale. The impact scale used by TRQ is a totally ordered set composed of named levels, as presented in Equation 1.

$$Critical \prec Serious \prec Moderate \prec Minor \prec None \tag{1}$$

It is then possible to derive a detrimental event as the inability for a business process to accomplish its mission, in case a business device is compromised on one of the security axis defined by the Feared Event related to this business process, with the impact defined by the consequence of this business process.

TRQ exploits the detrimental events, and their link with the business resource through the business model of the mission graph, to compute elementary risks (ER) based on computed attack paths. An elementary risk is defined by Kanoun et al. [6] as *"the quantum of risk inflicted by a single detrimental event to an asset through the exercise of a single attack scenario on one single supporting asset contributing to that asset"*.

The purpose of elementary risks is to bridge the gap between the technical exposure of a monitored system and the business perspective of the organization.

Unlike the classical definition of risk, an ER relates to the contribution of the risk on one occurrence of a specific event, caused by one (and only one) of the different possibilities that would cause this specific event. Although classical approaches calculate a scalar value using an evaluation function f that only combines two dimensions of the risk (i.e., *Likelihood* and *Impact*), we define an elementary risk ER_i as a three-tuple metric, as shown in Equation 2.

$$ER_i = [Likelihood_i, Impact_i, f(Likelihood_i, Impact_i)] \tag{2}$$

Equation 2 keeps track on the whole likelihood and impact information, which allows a distinguished treatment of the two risk dimensions in further management processes (i.e., prioritize treatments addressing the highest likelihood or impact rather than only reasoning on the risk). The composition function f used in TRQ is a two dimensional Risk Matrix based on qualitative scales, as recommended by many traditional Risk Management methodologies [5]. An example of such Risk Matrix can be found in the Annex E.2 of ISO/IEC 27005:2011 standard [1].

3.2.1. Likelihood of an ER Calculation

As we rely on attack paths to represent attack scenarios, the likelihood associated to an elementary risk ER_i depends on the technical characteristics of one and only one attack path that enables compromising its terminal node (i.e., with a nature of compromising compatible with the nature of the Detrimental Event's impact) such as: (i) the number of its intermediate steps; (ii) intrinsic characteristics of the vulnerabilities exploited in each intermediate steps.

Based on these parameters, we propose to calculate $Likelihood_i$ by considering an attack path, which is basically a sequence of exploited vulnerabilities, as an equivalent Markov chain in which the k^{th} state T_k of the Markov chain corresponds to the k^{th} step in the attack path. In the equivalent Markov chain, we set the exit rate λ_k of the sojourn time of the k^{th} state T_k to a value that is homogeneous to the difficulty score at the k^{th} step in the attack path.

The rational behind our choice to model an attack path as an equivalent Markov chain and to fix the exit rate of each node in the chain, is driven by two assumptions: (i) compromising a node with the exploitation of a vulnerability does not depend on the compromising of previous nodes in the chain (i.e. the process is memoryless or stateless); and, (ii) the time spent at each elementary attack step by the best fictitious attacker that follows the attack path is proportional to the difficulty of exploiting the vulnerability at this step.

We then derive λ_k from parameters relating to the difficulty to exploit the vulnerability exploited at the k^{th} step in the attack path. This difficulty score is computed based on parameters of metrics of the Common Vulnerability Scoring System version 2 (CVSSv2) [19], extracted from the National Vulnerability Database (NVD) [20], as presented in Equation 3:

$$\lambda_k = \text{RoundUp1}(AV \cdot AC \cdot Auth \cdot Expl \cdot RC) \tag{3}$$

From Equation 3, the function suggests to round up to one the product of the access vector (AV), access complexity (AC), Authentication (Auth), Exploitability (Expl) and Report Confidence (RC) CVSSv2 metrics of the related exploited vulnerability at step k of the corresponding attack path [19].

As proposed by Kanoun et al. [21], the Mean Time to Attack Object (MTAO) is first computed, comparably to the Mean Time To Failure (MTTF) in the dependable theory [22], as the summation

of the expectation of the mean sojourn time of each state in the equivalent Markov chain of the considered attack path (see Equation 4).

$$MTAO = \sum_{k} E\{T_k\} = \sum_{k} \frac{1}{\lambda_k} \tag{4}$$

It is worth noting that the notion of likelihood is homogeneous to a root-power quantity of a progression perceived as non linear, the likelihood of the attack path associated to ER_i is calculated in dB, as proposed in Equation 5.

$$Likelihood_i = -20 \cdot \log_{10}\left(\frac{MTAO - MTAO_{min}}{MTAO}\right) \tag{5}$$

From Equation 5, $MTAO_{min}$ is the lowest possible value for an MTAO. Such value corresponds to the simplest attack path, composed of one attack step, and for which the compromised vulnerability has the easiest possible CVSS. All CVSS metrics are values contained in the [0;1] interval. Moreover, the easier the exploitation of the vulnerability, the closer the parameters of its CVSS metrics to the value of one. The MTAO for the simplest attack path, is therefore, $MTAO_{min} = \lambda_{max}^{-1} = 1$.

The defined likelihood metric provides the following characteristics:

- the core part of the metric (i.e. the part inside the log function) is normalized between zero and one, as it is homogeneous to the probability of occurrence. This is achieved by dividing the numerator of the core part in Equation 5 by MTAO;
- the metric increases as a fictitious attacker progresses on the attack path. This is achieved since the core part of the metric decreases when considering shorter paths of an attack path. Assuming the likelihood is expressed in dB with a negative factor, the metric evolves from zero towards infinite values as an attacker approaches the target;
- the likelihood progression evolves as a non linear metric. More precisely, the metric increases its value as an attacker approaches the target in a given attack path;
- for two attack paths with the same number of steps, the lowest likelihood value is assigned to the attack path with the easiest vulnerability exploitation.

3.2.2. Impact of an ER Assessment

While the attack scenario (i.e., attack path) is relevant to calculate the likelihood, the impact of each ER_i (i.e., the Elementary Impact denoted EI_i) is decided by what follows the successful execution of the terminal step of an attack scenario. Both, the nature and magnitude of the impact, depend on the consequences of the exploitation of the last vulnerability V that resides in the business device (i.e., terminal node of the attack scenario). In particular, the consequences of V can be expressed with the boolean variables that indicate whether the successful exploitation of V leads to a violation of confidentiality (Imp_C), integrity (Imp_I), or availability (Imp_A). Note that Imp_C, Imp_I, and Imp_A are true if the associated CVSS metric of vulnerability V is not null.

The detrimental event associated to EI_i has two main properties that are obtained during the process of a traditional risk assessment: (i) $Viol_C$, $Viol_I$, and $Viol_A$ are Boolean variables that indicate whether the detrimental event arises following a violation of confidentiality, integrity and

availability respectively; and (ii) Magnitude is a quantitative or qualitative variable that indicates the impact's magnitude on the organization.

Hence, EI_i (i.e., the impact of ER_i) is calculated by crossing the technical consequences of the attack scenario (i.e., Imp_C, Imp_I and Imp_A) with the nature of the Detrimental Event ($Viol_C$, $Viol_I$ and $Viol_A$) as presented in Equation 6.

$$EI_i = [(Imp_C \wedge Viol_C) \vee (Imp_I \wedge Viol_I) \vee (Imp_A \wedge Viol_A)] \cdot Magnitude \tag{6}$$

3.2.3. Proactive Risk Profile Construction

Each possible attack path computed on the proactive perspective may lead to zero, one or several detrimental events DE_j. To a given possible attack path related to DE_j, denoted P, correspond a set of proactive elementary impacts $\{PEI_{i_j}^{p_j=P}\}$ (i.e., elementary impacts on DE_j caused by P), and ultimately a set of proactive elementary risks $\{PER_{i_j}^{p_j=P}\}$ (i.e., elementary risks relative to DE_j caused by P) where each $PER_{i_j}^{p_j=P}$ is composed as described by Equation 7.

$$PER_{i_j}^{p_j=P} = [Likelihood_{p_j=P}, PEI_{i_j}^{p_j=P}, f(Likelihood_{p_j=P}, PEI_{i_j}^{p_j=P})] \tag{7}$$

For each detrimental event DE_j, we have a set $\{PER_{i_j}^{p_j}\}$ composed of all the proactive elementary risks derived from the set of possible attack paths $\{p_j\}$ relevant to DE_j. The proactive risk profile is composed by aggregating the various computed elementary risks, using Equations 8 and 9.

A set of all proactive elementary risks, with its relating detrimental events is defined as follows:

$$\left\{[PER_{i_j}^{p_j}, DE_j]\right\} \tag{8}$$

For each detrimental event DE_j, the risk level which corresponds to the elementary risk relating to DE_j with the maximum risk level is computed using Equation 9.

$$Risk_{DE_j} = max\left(PER_{i_j}^{p_j}\right) \tag{9}$$

3.3. Response Operational Impact Assessor (ROIA)

Assessing how local, widespread events globally affect a higher goal, such as a mission or a company's business goals, is a process frequently called *missionimpactassessment*. The response operational impact assessor (ROIA) aims at using a mission impact assessment to ponder about negative side-effects of response plans and individual mitigation actions. Any executed action to prevent a proactive infiltration of a network and individual nodes, such as patching nodes or strategically enforced firewall rules, or any reactive measure to mitigate an ongoing attack, such as ad-hoc connection drop-downs or deactivation of individual nodes, inherently represents a threat to the operational capability on involved nodes. For instance, shutting down a node in a network will inevitably reduce the operational capability of this node with a probability of *one*. Further, employing a patch on a node might lead to an immediate conflict with, say, a probability of 10%, requires a reboot in some intermediate time, i.e., a probability of *one* and will eventually fully resume its operational capability, i.e., it will be reduced with a null probability. We call this reduction of operational capability, an *impact on a node*. Moreover, what needs to be considered is that any local impact may spread throughout a network: if a node is highly dependent on receiving

information from a node which has been shutdown, it definitely will be impacted as well as it is not able to operate as intended anymore. In the end it may be worse—from an operational perspective assuring mission success—to defend or eliminate an attack surface by an action, i.e., one sacrifices mission success for a false sense of security by a too narrow perspective on the problem. The ROIA is used to overcome these issues and broaden the point of view on defending attacks with the ultimate goal to assure mission success.

The spreads of local impacts are often considered in existing approaches to mission impact assessments. However, existing approaches frequently suffer from a fourfold problem, which we identify in our previous work [23] and [24] and characterize in the following four categories: (i) Self-crafted propagation algorithms are used, which are not based on any mathematical problem and require deep training of experts to parametrize a framework, (ii) require large sets of reference results to interpret obtained results, (iii) force experts out of their expertise, and (iv) require validation of end-results against ground truth. Extremely often experts to a framework are not available, which leads to a *guessing* of parameters at an initial setup, which may lead to spurious results. In the end, operator will become dull to obtained results, as $itisalwaysasevereredererroofcategoryfive$ and may first react at a stage of six. Moreover, such holistic system are extremely hard to verify against ground truth, as such, large sets of ground truth with exact information about spreads, local impacts and inflicted global impacts are frankly not available or are confidential.

While for ongoing and past attacks, such ground truth may be available at a very little extent, we intend to utilize these assessments for assessing the global consequences of mitigation actions, i.e., self executed responses. This is an extremely novel research field and definitely no large sets of ground truth are available. However, which is available are different experts from different expertise, where some deeply understand a company and its business goals (i.e., a mission), some deeply understand an ICT/ICS infrastructure, and some deeply understand the local effects of attacks and responses, but are not deeply aware on the global consequences of them. In our approach outlined previously [23, 24], we introduce a probabilistic graphical model combining all three (possibly disagreeing views) into one sound and well-defined model to assess the impact of locally executed actions onto a mission. By the reduction of mission impact assessment onto a known mathematical problem—probabilistic inference in probabilistic graphical models—every defined parameter is understandable directly, locally and without deep training of experts. For example, the introduced local probabilities in the first paragraph of section are directly understandable and do not require to overlook the complete infrastructure, i.e., do neither need to consider each and every dependency of each node nor the transitive dependencies of the company on these nodes. This is highly beneficial for our application, as each defined model, i.e., a mission dependency model from a business perspective, a resource dependency model from an infrastructure perspective, and a local impact model from a security and operational perspective, is validatable locally at parameter level. This means that the defined models can be validated independently, can be validated against expert knowledge and can be validated against very small locally executed ground truth sets.

In addition, based on a validated and well-defined probabilistic graphical model, each inference results must be seen as validated as well, i.e., obtained mission impact assessment are correct and validated by definition and do not require validation against holistic ground truth anymore. The ROIA module is composed of a learned resource dependency model, and a mission dependency model. The former has been validated by internal IT consultants, whereas the latter has been created by business experts. The remaining of this sections details each model (a visualized demonstration is depicted in Figure 4).

3.3.1. Resource Dependency Model

A resource dependency model is a mathematical model representing dependencies of involved resources supporting a company in the form of a probabilistic graphical model (PGM). The PGM represents every involved resource as a random variable and represents every dependency among them as a conditional probability of failure. A resource dependency model is constantly and automatically learned from network traffic analyses. By doing so, it supports a mission dependency model in identifying mission critical devices and dynamically captures changing environments.

3.3.2. Mission Dependency Model

A mission dependency model is a mathematical model representing dependencies among a company or mission in the form of a probabilistic graphical model (PGM). The PGM captures dependencies of a *company* or mission on its *business processes* that need to be accomplished. Business processes are supported by (i.e., are depended on) *business functions*, which are provided by *business resources*, i.e., mission critical devices, and in an ICT use case *business devices*. Each node of a mission dependency model represents a random variable, and every dependency represents a conditional probability of impact. A mission dependency model directly reflects expertise of business experts and is locally and directly modelable by experts. A mission dependency model is sees as static over time, as changes in a company may reflect themselves at a lower resource level.

A detailed identification of involved business resources may be disputable by different experts coming from different expertise. For example, a business function *provideaccesstocustomerdata* may depend on various business resources such as a central database server, a computational cluster analyzing customer behavior and a web-frontend. The mission dependency model accepts disagreeing information on the correct identification of those business resources by considering transitive dependencies on other resources by the use of a resource dependency model.

For technical details, an in-depth evaluation of the complexity for obtaining mission impact assessment, details on the process of obtaining and validating models, and multiple use case studies we refer to our previous work [23, 24, 25].

3.4. Strategic Response Decider (SRD)

The strategic response decider (SRD) handles identified threats, authorized mitigation actions and strategic policies (i.e., default and contextual policy rules, as well as contextual definitions). It extracts concrete entities from reported threats, and infers concrete policy instances to eventually guide the system into new updates and reconfigurations. These are provided as concrete response plans on a long-term proactive perspective. Response plans are validated by human operators, prior final enforcement. The goal of the SRD is the automated administration of policy-related activities, including addition of new rules, removal of unnecessary conditions, and activation of strategic responses (i.e., activation of new mitigation and response plans). The SRD interacts with the attack graph generator and threat risk quantifier (AGG-TRQ) and the response operational impact assessor (ROIA) components in order to evaluate and select the best proactive response.

Strategic responses are specified using response policies. Such policies are activated, deployed and enforced onto the system as soon as threats likely to lead to detrimental events are notified. Expert (manual) human knowledge is expected to define the strategic policies. This includes the definition in terms of default rules, contextual rules and context types. The SRD relies on the response financial impact assessor (RFIA) component to quantify the level of benefit perceived per response plan on a financial basis. The RFIA provides an assessment concerning the potential financial impact that a given response plan may cause to an organization. Response plans represent

proposed mitigation of the assessed risks and are assumed to be composed of one or more mitigation actions.

3.4.1. Response Financial Impact Assessor (RFIA)

The RFIA performs the calculation of the return-on-response-investment (RORI) index associated to the mitigation actions composing a response plan. The RORI index is used to evaluate optimal plans, by ranking them as a trade-off between their efficiency in stopping potential attacks, and their ability to preserve, at the same time, the best service to legitimate users. The RORI index is calculated for each mitigation action, according to Equation 10.

$$\frac{(ALE \cdot RM) - ARC}{ARC + AIV} \cdot 100 \tag{10}$$

Where ALE (annual loss expectancy) refers to the financial cost expected from the threat, in the absence of applying mitigation; RM (risk mitigation) estimates the effectiveness and coverage of an action in mitigating the threat; ARC (annual response cost) expresses the expected cost of applying the mitigation action; and AIV (annual infrastructure value) is a fixed cost associated to the system infrastructure (e.g., cost of equipment, services, etc.), regardless of applying or not mitigation. More information about the computation of each parameters of the RORI index can be found in our previous research [26].

The RFIA component evaluates and selects mitigation actions from a pool of candidates, by ranking them in terms of RORI values. The higher the RORI value associated to a mitigation action, or to a combination of mitigation actions, the higher the associated ranking. The purpose of this process is to pre-select sets of combined mitigation actions that are identified as optimal from a financial perspective and propose them to reduce the risk of threats against the monitored system. Pre-selected sets are sent to the Response Operational Impact Assessor (ROIA) and to the Visualization Environment, prior their eventual deployment over the monitored system.

With regard to Equation 10, a first improvement has been proposed to enhance the risk mitigation (RM) function of the RORI expression. The work, reported by Gonzalez-Granadillo et al. [27, 28] extends the concept of attack surface used in previous versions of the RORI metric. It identifies authorization and contextual dimensions that may directly contribute to the exposition of system vulnerabilities. New properties associated to the vulnerabilities, such as temporal conditions (e.g., granted privileges only during working hours), spatial conditions (e.g., granted privileges when connected within the company premises), and historical conditions (e.g., granted privileges only if previous instances of the same equivalent events were already conducted) can now be included and combined with the RORI cost-sensitive metric.

The process undertaken by the strategic response decider (SRD) extends initial work reported by Gonzalez-Granadillo et al. [26]. The approach proposes the combination of authorization models and quantitative metrics, for the selection of mitigation actions. The actions, modeled in terms of contextual rules, are prioritized based on a cost-sensitive metric that extends the return on investment (ROI) concept. The goal is finding an appropriate balance between the financial damages associated to a given threat, and the benefits of applying some mitigation actions to handle the threat, with respect to the loss reduction. The RORI metric addresses such a goal.

The adaptation of the selection process, based on financial and operational assessment functions, has been presented in a previous research [25], which reports the combination of both assessment approaches, over a representative set of mitigation actions. The combination, based on a multi-dimensional minimization proposal, proposes the choice of semi-optimal responses that, on the one

hand, bear the highest financial attractiveness on return on investment; and, on the other hand, bear the lowest probability of conflicting with the organization's missions. This is seen as beneficial for its application in the scenarios of the project, where highly critical missions and resources must be protected, without sacrificing missions in favor of security.

3.4.2. Security Policy Instantiation for Primitive Response Plans construction

A response plan (RP) is a vector of mitigation actions, representing individual actions to be performed as a response to an adversary or threat opposed to an organization. The generation of a response plan considers the information of the threat scenarios coming from the abstract security policies (ASP) file and the information of detrimental events (DEs) coming from the proactive risk profile (PRP). We compare predefined conditions in both input files (e.g., is the likelihood of the threat scenario greater than or equal to the likelihood of the detrimental event). In such a case, we collect all attack path IDs that will be used in the attack graph parsing process. If the condition is not met, the process generates an empty response plan.

Having the most updated information of the network inventory and using the attack vector from the ASP, we generate a concrete attack vector file. A determination is made on whether there is a partial concrete attack vector (i.e., for each path of the attack vector, we search all active nodes from the network inventory). If at least one concrete attack vector is found, the process searches for a match of entry points and business devices from the obtained attack vector and the mission graph. However, if there is no concrete attack vector, the process generates an empty response plan.

Having the attack graph file and the attack vector, we search for paths that match both input files. A determination is made on whether there is a final concrete attack vector (i.e., for each path of the attack vector, there is a node that matches with the attack graph). If at least one matching node is found, the process collects the set of nodes from the attack vector involved in the attack graph, otherwise, the process generates an empty response plan.

Having the list of authorized mitigation actions, a determination is made on whether or not there are involved nodes in the process. If it is the case, the process extracts all mitigation actions associated to the PEP type of the nodes obtained from the attack graph, otherwise, the process generates an empty response plan. The RORI evaluation is performed on the extracted mitigation actions and response plans are generated accordingly.

The output of this module is a set of response plans representing individual actions to be performed as a response to an adversary or threat opposed to an organization. The response plan contains an ID, the mitigation action ID an type, the policy enforcement point and the RORI index. Response plans are of two types: individual, when only one mitigation action is proposed; and combined, when two or more mitigation actions are proposed to be implemented. In such a case, a new parameter called 'probability of conflict' is included in order to manage restrictions among the proposed actions.

3.4.3. Enriched Response Plan as the Best Primitive Response Plan

This module obtains the primitive response plans and performs an operational evaluation in order to select the best response plan in financial and operational terms. We consider that response plans, while highly effective, could lead to operational negative side-effects inside the network and therefore onto a mission. Response plans are therefore evaluated based on local impact and assessments of dependencies inside an organization's business. As a result, response plans are enriched with operational information that indicates the impact over the organizational mission(s).

Based on a multi-dimensional minimization proposal, the number of response plans is reduced to only one that best satisfies both: the financial and the operational impact modules [25]. The method searches for the best semi-optimal response plan with the lowest operational impact assessment and the highest RORI index.

A response plan is said to be semi-optimal since it might not be the best solution neither in financial nor in operational terms, but it proposes a set of mitigation actions that on the one hand, bears the highest financial attractiveness on return on investment, and, on the other hand, bears the lowest probability of conflicting with a company's mission. This is beneficial for applications, where highly critical missions and resources must be protected, without sacrificing missions in favor of security.

In order to find a semi-optimal response plan, we consider the financial impact given by the RORI index, and the operational impact of the business company at time T0 (i.e., OI_0 = short-term), T1 (i.e., OI_1 = mid-term), and T=2 (i.e., OI_2 = long-term). The approach focus on keeping RORI high, short-term impact low, mid-term impact down, and long-term impact low [23].

The approach searches for a boundary of acceptable elements (acceptable as a compromise). This boundary is a numerical value representing a deviation (ϵ) of the optimum. For instance, with ϵ=0.1, we accept 10% deviation of the optimum in each dimension. The acceptance criteria for the financial and operational impact are different. For the financial impact, we keep response plans whose RORI index are greater or equal to 90% of the highest RORI value. For the operational impact, we keep response plans whose 'OI' are up to 10% of the lowest 'OI' value. Then, we check if there is a match in all evaluated response plans. If there is a match, we stop the process; otherwise, we increase the ϵ value until we find a tuple that matches. In particular, we search the ϵ where we obtain the smallest set of values.

Once a semi-optimal response plan is found, the information is sent to the visualization module, which depicts such results to the security operator.

4. Integration into a SCADA environment

We study the infrastructure environment of an energy distribution organization called ACEA[1]. For testing purposes, some real control systems of ACEA in cold standby mode (e.g., at their disaster recovery sites) have been replicated into a cloned SCADA environment for scalability emulation. The replicated system consists of a distributed network of Remote Terminal Units (RTU) in energy stations of medium voltage (MV = 20,000 Volts) and high voltage (HV = 150,000 Volts), that acquire data from electrical devices (e.g., PLC, sensors, etc), and send them to the supervisor terminal unit (STU) of the headquarters. Some video and screencasts of the replicated environment are available at `http://j.mp/DRMRSVis`.

The aforementioned environment is a controlled system that replicates several sections of the organization, making it possible to test and validate the framework and its components. The system uses industrial SCADA protocols for energy distribution. The replicated network used for our evaluation is composed of over 13,000 energy stations, 6,000 of which are controlled by the central system. All data used in the computation of the operational and financial impact assessment modules have been obtained and validated by use case providers. Using statistical

[1]Critical Infrastructure organization based in Italy whose main activities focus on energy distribution and water supply, see `http://www.acea.it`

data, expert knowledge, simulation and risk assessment tools, we are able to quantify the different parameters composing our proposed prototype.

Several techniques for the quantification of cost and impact on security-critical systems (in the ICT sense) exist in the literature [27, 28, 29, 30, 31]. Quantification of costs and impacts on security-critical systems are built upon existing mathematical and numeric simulation models to assess cyber security events (e.g., threats, errors, countermeasures). We have used such techniques during the interviews with use case providers aiming at estimating each of the parameters composing our models for the use case of this article. Contrary to most security approaches that treat the risk analysis as a binary search (true or false), indicating that either the system has been compromised or not; we provide a proactive approach that explores multiple parameters (e.g., criticality, vulnerability, effect, etc.) associated to every asset of the system and evaluates the consequences and all possible mitigation actions that could be deployed if a threat is realized.

The remaining of this section discusses the integration and main results of our DRMRS in a scenario with multiple threats.

4.1. Input Information

This sections provides the input data used by the DRMRS prototype to analyze and evaluate threat scenarios and to generate response plans accordingly.

4.1.1. Network Inventory

The current snapshot (i.e., 05/20/2016 12:26:42) of the system shows that there are 17 active PEPs with an AIV equivalent to 6,925,555.00 €, as shown in Table 1. Please note that AIV corresponds to the value obtained out of the sum of all PEP's cost (i.e., annual equipment cost, AEC) that are active at the time of the snapshot. The AIV parameter is a variable value that depends on the time of the evaluation and the PEP that are detected by the system.

4.1.2. Abstract Security Policies

Use case providers have identified five threat scenarios that could lead to severe consequences on the target system. The abstract security policies show the information associated to all possible threats, PEPs and attack vectors. Table 2 summarizes this information.

The annual loss expectancy (ALE) is computed as the product of the severity times the frequency of the threat (both are input values that could be obtained qualitatively and transformed into quantitative values). The attack vector shows the attack path which considers the involved policy enforcement points (PEP) as entry points (EP), intermediate targets (T1, T2), and business devices (BD). The likelihood is associated to detrimental events. Those threats whose likelihood is above a specific threshold (e.g., in this case threshold = 0.5) are therefore analyzed. Threats AS01MV, AS03 and AS04 are discarded since their likelihood of occurrence is lower than the proposed threshold.

4.1.3. Authorized Mitigation Actions

Each threat has PEP that are directly or indirectly affected. Each PEP_Type is responsible of applying one or more mitigation actions. The following three types of mitigation actions can be applied in our system:

- Patching, refers to a piece of software designated to update a computer program or its supporting data, to fix or improve it. This includes fixing or removing security vulnerabilities and other bugs with such patches and improving the usability or performance;

Table 1: Network Inventory Input Data

PEP	PEP_Type	Description	AEC (€)
PEP16	FWCEDET	Logical Firewall and IPS working in CEDET	105,000
PEP2	SRVMSCADA	Medium Voltage Server	355,000
PEP1	SRVXSCADA	High Voltage Server	355,000
PEP3	FEXSCADA	High Voltage Front End	1,320,000
PEP13	FTPSRV	FTP Server	3,000
PEP11	HMISCADA	Human-Machine Interface	80,000
PEP15	NTPSRV	NTP Server	2,000
PEP20	VRTX	Edge Router on Remote Sites	206,796
PEP14	USERPC	User PC	1,000
PEP5	GWMSCADA	Medium Voltage Gateway	410,532
PEP6	GWXSCADA	High Voltage Gateway	615,800
PEP17	FWDR	Firewall IPS/DR	105,000
PEP10	WEBCADA	Web Server	45,000
PEP18	MGMSRV	Management Server	3,000
PEP9	RTUSCADA	Remote Terminal Unit	2,621,927
PEP4	FEMSCADA	Medium Voltage Front End	660,000
PEP7	VGROUTER	Virtual Router	36,500
Annual Infrastructure Value (AIV)			6,925,555

Table 2: Abstract Security Policies

Threat	Description	ALE (€)	Attack Vector	Likelihood
AS01HV	DoS to High Voltage nodes	20,000,000	EP=VGROUTER; T1=WEBSCADA; T2=FTPSRV; BD=FEXSCADA	0.75
AS01MV	DoS to Medium Voltage nodes	2,000,000	EP=RTUSCADA; T1=GWSCADA; T2=FEXSCADA; BD=SRVSCADA	0.25
AS02	Data corruption or leakage	12,000,000	EP=VGROUTER; T1=WEBSCADA; T2=USERPC; BD=FTPSRV	0.60
AS03	DoS against SCADA electrical devices	100,000	EP=RTUSCADA; T1=GWSCADA; T2=FEXSCADA; BD=SRVSCADA	0.20
AS04	DoS against business Services	2,000,000	EP=VGROUTER; T1=FTPSRV; T2=USERPC; BD=WEBSCADA;	0.25

- Reboot, refers to the process of restarting a device or a computer program;
- Shutdown, refers to completely remove any possibility to access a device by powering off a device.

Each type of mitigation action has an associated effectiveness (EF) and cost (ARC). The EF value is assigned by the system (e.g., Reboot=1%, Shutdown=15%, Patching=100%), whereas the

ARC value is assigned by expert knowledge and statistical data. Table 3 provides an example of the list of mitigation actions authorized to threat AS01HV.

Table 3: Authorized Mitigation Action Information

PEP_Type	**WF**	**Affected Node**	**Q**	**COV**	**MA_Type**	**EF**	**ARC (€)**
WEBSCADA	3	STWEB	1	0.09	Shutdown	0.15	15.00
					Reboot	0.01	15.00
					Patching	1.00	25.00
FEXSCADA	4	mferp1, mferp2	4	0.50	Shutdown	0.15	200.00
		xfep1, xferp2			Reboot	0.01	200.00
MGMSRV	1	LANGUARD, KALI	2	0.00	No action	0.00	0.00
RTUSCADA	5	TP2000-T2	1	0.16	Shutdown	0.15	15.00
					Reboot	0.01	15.00
FTPSRV	2	ARCHIVESRV,	3	0.19	Shutdown	0.15	15.00
		FTPSRV01,			Reboot	0.01	15.00
		dorete			Patching	1.00	25.00

As shown in Table 3, each PEP_Type has an associated weighting factor (WF) that indicates the level of priority or criticality inherent to the type of PEP in the execution of a mission to the organization. The COV value is computed using Equation 11. Both 'EF' and 'COV' parameters are required to compute the risk mitigation (RM) value of individual and combined mitigation actions needed for the computation of the RORI index as described in Section 3.4.1.

$$COV = \frac{Q_i \cdot WF_i}{\sum_{j=0}^{n} QT_j \cdot WF_j} \tag{11}$$

Where:
Q_i = number of nodes from a PEP_Type that are affected by a given mitigation action
WF_i = weighting factor associated to the affected PEP_Type
QT_j = Total number of node types that appears in the attack graph
WF_j = Weighting factor associated to each node type

4.1.4. Vulnerability Inventory

Two vulnerabilities (i.e., CVE-2008-4250, and CVE-2006-3439) are associated to the attack scenarios AS01HV and AS02. Such vulnerabilities have been found in the nodes STWEB, dorete, and user-PC. The list of protocols and ports used to compromise a node or that can be open after exploiting the aforementioned vulnerabilities are shown in Table 4. The privilege level required before and after exploitation is 'root'.

Table 4 represents the list of protocols and ports to which an attacker must have access on the attacked device containing the considered vulnerability, to be able to exploit successfully the considered vulnerability on the device. The instances are linked to a list of protocol and port classes, which are defined in the reachability matrix.

Table 4: List of Protocols and Ports associated to the threats AS01HV and AS03

Port	TCP	Port	UDP
21	FTP control (Command)	67	Bootstrap Protocol Server and DHCP
22	Secure Shell (SSH)	69	Trivial File Transfer Protoco
23	Telnet (Encrypted telecommunications)	123	Network Time Protocol
111	Sun Remote Procedure Call	111	Sun Remote Procedure Call
512	Rexec Remote Process Execution	161	Simple Network Management Protocols
514	Shell	10000	XHX
		631	Internet Printing Protocol
		33434	Traceroute

4.1.5. Rechability Matrix

The reachability matrix provides an exhaustive view of the communication among devices and the network topology up to the ISO/OSI level 4. It is composed of a reachanility matrix correlator (RMC) that performs the reachability computation across the monitored ICT network to determine whether two nodes are reachable from each other in the network, and this for all pairs of nodes representing ICT devices. RMC utilizes a novel, ontology based approach to combine information from various sources to obtain a reachability matrix, which is required by various components, such as the attack graph generator.

The RMC exploits SWRL (Semantic Web Rule Language) rules. These rules describe all possible scenarios to be investigated in order to detect all nodes that are reachable from any given pair of node (with its specific routing instructions) and associated network interfaces (each being connected to one particular network). Each node of the monitored system can be selected in order to check all reachable nodes (highlighted nodes). Information shown for generating the diagram in Figure 3 is extracted from the computed reachability matrix.

4.2. Output Information

This section presents the output files generated by our DRMRS upon the reception of the input data described in Section 4.1. It is important to note that after comparing the likelihood values of detrimental events in the proactive risk profile against the threat scenario threshold values, two threats (i.e., AS01HV and AS02) present a likelihood of occurrence higher than the predefined threshold. The system therefore retrieves a concrete attack vector indicating the attack path for each threat scenario. The remaining of this sections describes each DRMRS output file for both threat scenarios.

4.2.1. Mission Graph

Based on the information from the mission graph, we have retrieved the nodes in paths pointing to business devices for threat 'AS01HV' and 'AS02'. Each node has a unique identifier, a host name that corresponds to an instantiated device, a PEP_Type which corresponds to the abstraction class of the PEP, and a Node_Type, which indicates whether the node is an entry point, an intermediate node, a target node or a business device. Please note that business devices are the most critical node types from the SCADA environment. They are required to accomplish a business process

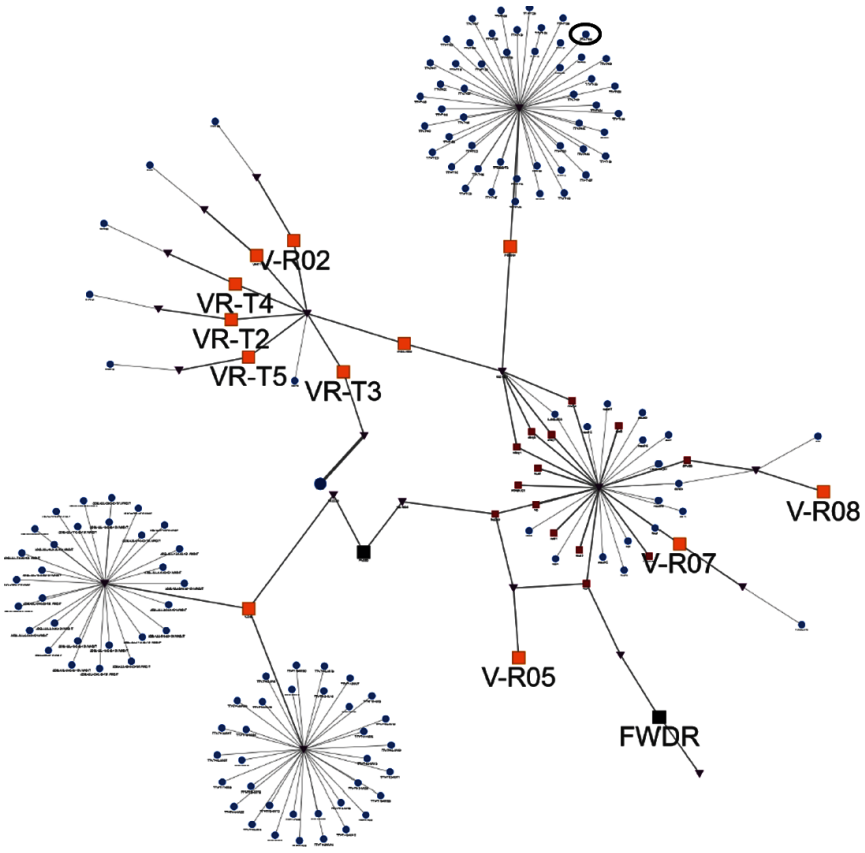

Figure 3: Computed reachability for each node starting from the node with the circle. The tested version of the environment is composed of 166 nodes (blue dots), divided into 19 broadcast domains (purple triangles). It is possible to identify a set of routers (yellow squares) which connect the broadcast domains. Routers and firewalls (black squares) imply firewalling rules on network traffic in order to protect the command and control network from intrusion.

within the organization. An example of the retrieved node information for the threat 'AS01HV' is given in Table 5.

4.2.2. Attack Graph

The Attack Graph Generation (AGG) function extracts the target nodes of the monitored system and the hypothetical entry points from the mission graph in order to construct an attack graph by inferring possible elementary attack actions with the knowledge of the reachability matrix and vulnerability inventory.

The reachability matrix data set represents a topology of 163 devices. From this set, we have identified 19 devices (i.e. 11.65% of the devices) with at least one vulnerability. However, most of these devices have more than one associated vulnerability (with an average of 11 vulnerabilities) for a total of 219 identified vulnerabilities. The produced attack graph is composed of 16 nodes and 207 edges. From the 28 devices identified in the mission graph as business devices, 8 of them are the target of at least one attack path (Table 6).

Table 5: Retrieved Node Information

Node Identifier	**Host Name**	**PEP_Type**	**Node_Type**
b992e600-0de2-496c-kkk0-ssjenqaa123h7	mferp1	FEXSCADA	Business device
718bc323-9d78-4ada-9629-8176f42a9703	dorete	FTPSRV	Target2
e06496d2-6120-4c9d-a310-cd93e9305b48	LANGUARD	MGMSRV	Intermediate Node
94d37c8d-bc68-47bf-ad60-7524a77e1464	ARCHIVESRV	FTPSRV	Target2
19b2bb1e-9f23-4fe8-902e-1a26feaf58e8	KALI	MGMSRV	Intermediate Node
e470baab-5d88-4b20-ac28-61ea42b37da3	FTPSRV01	FTPSRV	Target2
876hhezq-77tg-4897-665g-jjhsfaqzs665	xferp2	FEXSCADA	Business device
d3480ddc-fe4a-4b94-9dc5-5cf94e658291	mferp2	FEXSCADA	Business device
b54b235d-116a-49b4-9052-353a0d15caad	xferp1	FEXSCADA	Business device
c9fa4086-d979-4794-9b6e-cd0478040856	STWEB	WEBSCADA	Target1
c6dd8687-c791-4f91-bf58-497de6032088	TPT2000-T2	RTUSCADA	Intermediate Node

Table 6: Target business devices in the proactive attack graph

Node Identifier	**Host name**	**Type**
718bc323-9d78-4ada-9629-8176f42a9703	dorete	Terminal
876hhezq-77tg-4897-665g-jjhsfaqzs665	xferp2	Server
94d37c8d-bc68-47bf-ad60-7524a77e1464	ARCHIVESRV	Server
b54b235d-116a-49b4-9052-353a0d15caad	xferp1	Server
b992e600-0de2-496c-kkk0-ssjenqaa123h7	mferp1	Server
c9fa4086-d979-4794-9b6e-cd0478040856	STWEB	Server
d3480ddc-fe4a-4b94-9dc5-5cf94e658291	mferp2	Server
e470baab-5d88-4b20-ac28-61ea42b37da3	FTPSRV01	Server

With the two entry points identified in the mission graph (Table 7) as possible start for attacks, and although the attack graph itself is relatively small, when computing the attack paths of length 4 reaching one of the 8 targets of the graph AGG finds 42.182 different possible attack paths.

This number is close to the theoretical value of possible paths in a full-meshed graph of 16 nodes with 2 starting nodes and 8 target nodes (i.e. $2 \times 8 \times (15 \times 14 \times 13 + 15 \times 14 + 15 + 1) = 47.296$). It is due to the flat nature of the topology of the monitored system, composed of large number of similar SCADA devices on same sub-networks. It validates the usefulness and compactness of attack graphs to represent the possible attack scenarios on such kind of topology.

Computing the exhaustive number of attack paths in such a use case would probably results in a value close to the theoretical number of paths in a fully-meshed Directed Graph of 16 nodes (i.e. $(2 \times 8 \times \sum_{i=1}^{16} \frac{15!}{(15-i+1)!}) > 5.6 \times 10^{13}$). We then parametrize the AGG prototype to compute only the attack paths of length 4 and shorter.

Table 7: Devices of the Network Inventory that are entry points in the proactive attack graph

Node Identifier	Host name	Type
c6dd8687-c791-4f91-bf58-497de6032088	TPT2000-T2	Terminal
19b2bb1e-9f23-4fe8-902e-1a26feaf58e8	KALI	Terminal

4.2.3. Proactive Risk Profile

Detrimental events are extracted from the mission graph by deriving them from the defined business processes. From the ACEA use case, four different detrimental events are extracted as presented in Table 8.

Table 8: Detrimental Events extracted from mission graph

Identifier	Name	Description
DE-BP-M-D	Potential loss of power.	Unability to [Medium Voltage Distribution] can induce [Potential loss of power.] with a [1.0] magnitude (i.e. impact degree).
DE-BP-M-C	High manual work load. Manual Control of Medium Voltage Distribution is possible. Is redundant, not everything must work.	Unability to [Medium Voltage Remote Control] can induce [High manual work load. Manual Control of Medium Voltage Distribution is possible. Is redundant, not everything must work.] with a [0.1] magnitude (i.e. impact degree).
DE-BP-H-C	Extremly High manual work load. Manual Control of High Voltage Distribution is possible. All substations must work constantly.	Unability to [High Voltage Remote Control] can induce [Extremly High manual work load. Manual Control of High Voltage Distribution is possible. All substations must work constantly.] with a [0.55] magnitude (i.e. impact degree).
DE-BP-H-D	Complete Rome without power. Extremly catastrophic.	Unability to [High Voltage Distribution] can induce [Complete Rome without power. Extremly catastrophic.] with a [0.55] magnitude (i.e. impact degree).

Using the attack graph computed by AGG, the TRQ computes Elementary Risks based on the Impact extracted from the consequence associated to each business process in the mission graph. Although the attack graph correspond to 42.182 attack paths, the TRQ produces 63.208 Elementary Risks based on the computation presented in Section 3.2. And, the risk values computed for each Detrimental Events on the proactive perspective are, definitely, those presented in Table 9.

4.2.4. Enriched Response Plan

For the generation of the enriched response plan, we compare the RORI values of all individual and combined mitigation actions with the short-term, med-term, and long-term operational impact (i.e., OI_0, OI_1, OI_2). As such, we are able to find a semi-optimal response plan that matches the

Table 9: Risk values of Detrimental Events in the Risk Profile on the proactive perspective

DE Identifier	Likelihood	Impact magnitude	Risk nature	Risk level
DE-BP-M-D	0.0	1.0	CIA	Null
DE-BP-M-C	1.0	0.1	CIA	Extreme
DE-BP-H-C	1.0	0.55	CIA	Extreme
DE-BP-H-D	1.0	0.55	CIA	Extreme

conditions of having the highest RORI value and the lowest OI values. There is a total of 224 response plans to be evaluated for threat 'AS01HV', 797 response plans to be evaluated for threat 'AS02'. Details on the evaluation process are found in the work performed by Gonzalez-Granadillo et al. in [11, 25].

The enriched response plan for threat 'AS01HV' has a RORI index equivalent to 71.34%, and the following operational impacts: $\langle OI_0 : 0.2724, OI_1 = 0.2161, OI_2 = 0.1781 \rangle$. The selected response plan against threat 'AS01HV' proposes the following concrete actions:

- Install patches to the node STWEB against CVE-2008-4250, and CVE-2006-3439;
- Shutdown the node TPT2000-T2;
- Reboot nodes ARCHIVESRV and FTPSRV01;
- Install patches to the node dorete against CVE-2008-4250, and CVE-2006-3439.

The enriched response for threat 'AS02' RORI index of 82.8514 and the following operational impacts $\langle OI_0 : 0.108, OI_1 : 0.09, OI_2 : 0.0 \rangle$ and a RORI index of 82.8514. The selected response plan against threat 'AS02' proposes the following concrete actions:

- Install patches to the node STWEB against CVE-2008-4250, and CVE-2006-3439;
- Install patches to the node user-PC against CVE-2008-4250, and CVE-2006-3439;
- Reboot the node STWEB;
- Reboot User-PC.

4.3. DRMRS Visualization and Enforcement

The DRMRS allows visualization of many features e.g., Configuration, which allows editing input files in a standard format (e.g., XML, JSON) containing relevant data that is used to initialize the system; Vulnerability, which allows the user to inspect the actual system vulnerabilities, arranging them according to different analysis objectives; Topology, which presents the user with visual information about the network topology and its geographic location; mission graph, response plans; Attack and Mission Graph, which allows to inspect the current attack graphs and depicts information about goals and services that represent the core business of the organization. In addition, the prototype proposes a view of response plans, which provides information about the financial and

operational impact evaluation of attacks and mitigation actions. Some visualization examples are available at `http://j.mp/DRMRSVis`.

It is important to know that although the visualization process is of automated nature, the system operator has the possibility to manually click on different nodes to explore more deeply the vulnerabilities associated to it and to extract more information about the possible issue(s). Furthermore, the visualization tool allows for verification of the impact level i.e., low (green), medium (yellow), and high (red) by simulating the enforcement of a given mitigation action which eliminates or rearranges vulnerabilities on the system.

In order to avoid conflicts, every time a response plan is generated, the system verifies the current implemented actions and compare them with the actions proposed by the new generated response plan. After discarding duplicates actions (e.g., shutting down a device that is currently off, installing a patch that has been already installed, etc), the system enforces those actions that do not require manual intervention (e.g., reboot, shutdown). However, for those actions that require the installation or removal of hardware/software, the system generates a ticket indicating the action to be performed by a SCADA operator.

5. Testing and Experimentation

We conducted a thorough experimentation using data retrieved from scans in the SCADA environment. The remaining of this section details some of the tests performed on each component of the DRMRS platform.

In order to test, verify and validate the attack graph component, we performed a twofold approach: (i) Case-study based experimentation, and (ii) replicated data based experimentation. The first one focuses on the core functions of the attack graph, and verifies that output results are produced correctly and accurately considering well mastered input. The second one focuses on the integration of the attack graph with the rest of the modules, and verifies if all information exchange is performed correctly and within the expected time.

Experiments on the attack graph component use data retrieved from scans in the SCADA environment. A data-set is composed of several files (e.g., reachability matrix XML data file, vulnerability inventory XML data file, mission graph JSON file, scored vulnerability inventory JSON file). Based on that, we assessed the order of magnitude of the attack paths that would be generated by the generation algorithm. Some vulnerability exploitation graph visualizations are available at `http://j.mp/DRMRSVis`.

From a theoretical point of view, this graph is formed of 16 nodes fully meshed (at the center of the graph). If we consider the sources (i.e., Entry Points) and the target (i.e., critical devices) that exist among these nodes, we obtain a combination of $14! = 87,178,291,200$ possibilities (only for the longest paths). If the source is outside of this group of 16 fully meshed nodes and still connected to all of the 16 nodes (i.e., in one direction only, so that the source node is not fully meshed with the others), we obtain 15 times more possibilities. This latter indicates the number of all possible combinations for a connection.

In addition, we need to consider the fact that each connection has 2 possible vulnerabilities. In term of attack paths, meaning that the numbers should be multiplied by a factor 2^{14} for the first case, and 2^{15} for the second case. We note that the mathematical exhaustive calculation is not adapted to full meshed or even strongly connected network. A first optimization of the attack graph generation algorithm has been therefore proposed. Such an optimization considers the fact that an attacker with a root access has at least all the exploitation possibilities of an attacker with

a simple user access. As a result, the vulnerability exploitation graph is reduced to a full mesh of 8 machines, which corresponds to 109,600 connection paths (i.e., clustered paths).

In order to verify and validate the functionality of the ROIA component, we use a three step approach.

- Code inspection tests, identifying crucial regions of code providing and fulfilling specialized requirements. This is used to assure that mathematical principles are correctly embedded;
- Automated tests: Functional syntactic tests, testing correct syntactic behavior when given syntactically correct input data. Functional behavior tests, testing the intended behavior and reaction to specific kinds of input data;
- Semantic tests, testing the usefulness and correctness of obtained results by the Mission Impact Model (MIM) and the ROIA component. The MIM and ROIA are based on a probabilistic graphical model, which provides local semantics that allow validating individual parameters. Based on probabilistic inference we assure that obtained results are validated as well. We describe this approach deeply in the research work proposed by Motzek et al. [24].

In addition, various scalability, performance and accuracy tests have been executed on the ROIA evaluation. In particular, we show that an employed approximation algorithm is verified against exact inference and provides an expected convergence depending on specific parameters. We show the linear scalability of a ROIA, i.e., that the computation time required for one ROIA evaluation scales at most linearly with each parameter in various experiments. For these experiments we use artificially generated data to produce test sets way beyond the complexity of the target environment. In particular, we show the scalability beyond a network consisting of 400,000 dependencies, i.e., dependencies between individual resources, which is the most deciding parameter of complexity in our approach.

A visualized demonstration of the automated learning approach and utilized mathematical models is given in Figure 4, where the target company is represented in dark green, critical devices are highlighted in green, while business functions are in blue and business processes in orange. In fact, this model was validated by business-, operation- and security-experts to the company and contains all crucial dependencies, bears reasonable dependency degrees, and realistically represents the goals and aims of the company.

In terms of the SRD component, we conducted several integration tests, where it is stated that:

- The component is able to receive information from the system in a *push* mode at any time. In this mode the integration framework (via the SRD Proxy) transfers the most up-to-date information of the target system to the component using a SFTP server. The component is also able to receive the following data types: network inventory, proactive risk profile, reachability matrix, authorized mitigation action, abstract default security policy, abstract response policy context, TRD enriched response plan, SRD selected response plan and attack graph.
- The component retrieves information from the system in a *pull* mode using a set of web services client interfaces. Those interfaces request to the Integration Framework (via the SRD Proxy) the most up-to-date or previous information of the system. The module is able to request the following data types network inventory, proactive risk profile, reachability matrix, authorized

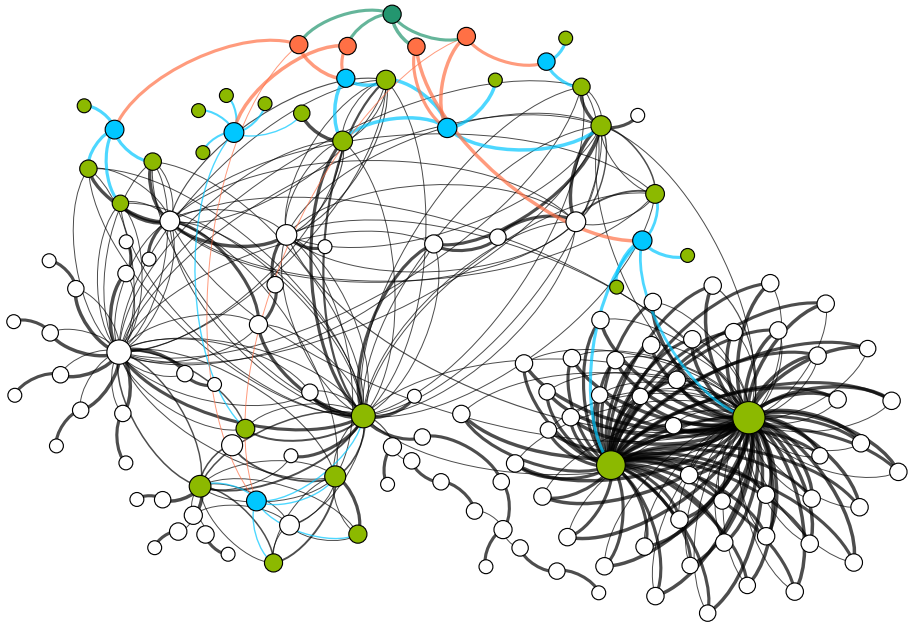

Figure 4: Visualized resource dependency model and mission dependency model, both learned in the backup environment for the disaster recovery site of the target company (dark green). Business resources display in green, business functions in blue, business processes in orange. Some screen videos are available online at: http://j.mp/DRMRSVis

mitigation action list, abstract default security policy, abstract response policy context and attack graph.

Regarding the RFIA component, several test cases have been executed in order to evaluate the computation speed in the combined evaluation of mitigation actions. The number of combination for a set of non-restrictive candidates is given by the expression $X = (2N) - (N + 1)$, therefore, in case $N = 4$, the number of combinations will be equivalent to 11, in case $N = 12$, the number of combinations is equivalent to 4083. Since the total number of combinations grows exponentially, we measured the time at which the system is able to perform the evaluation of multiple candidates. Results are plotted in a curve as shown in Figure 5.

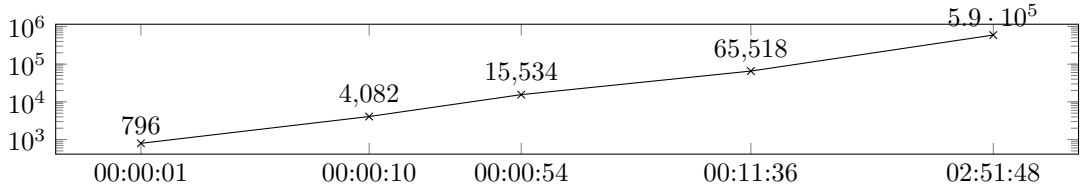

Figure 5: Computation time (abscissa) to evaluate all combinations of mitigation actions is linear in the number of combinations (ordinate) *(double logarithmic plot)*.

As depicted in Figure 5, a combination of 12 restrictive mitigation actions results into 796 combinations. Results are obtained in less than one second. For 12 non-restrictive mitigation actions, there are a total of 4 082 combinations, which are performed in less than 10 seconds. Similarly, for 16 restrictive mitigation actions, there are a total number of 15 534 combinations performed in less than one minute, whereas for 16 non-restrictive mitigation actions, there are a total of 65 518 combinations, which are performed in a total elapsed time of 00:11:36,26 (less than 12 minutes). Going up to 24 restrictive mitigation actions, results into 590 464 combinations. Such number of combinations is executed in almost three hours. As a result, in order to keep the evaluation process within a reasonable time (less than one minute), the system must process up to 14 non-restrictive mitigation actions (16 369 combinations). Beyond this threshold, the elapsed the performance of the evaluation is highly degraded.

In addition, we have tested the interaction with the Response Operational Impact Assessor (ROIA) module in order to select the best response plan in financial and operational terms. As such, we reduce the number of response plans to only one that best satisfies the RFIA and the ROIA modules. The method searches for the best semi-optimal response plan with the lowest operational impact assessment and the highest RORI index.

6. Related Work

Automated systems that evaluate and select actions to mitigate complex attack scenarios is an open research that represents a big challenge to critical infrastructures. Some research works have been conducted in the assessment of security measures. Kotenko et al. [32, 33], for instance, propose a framework for cyber attack modeling and impact assessment based on attack graph generation, real-time event analysis techniques, prognosis of future malefactor steps, attack impact assessment, and anytime approach for attack graph building and analysis. The framework evaluates security aspects in order to provide impact analysis for detection of malefactors and determining the countermeasures in near real time. We differ from this research in that we do not propose new algorithms or methods of attack graph construction, instead, we propose a novel framework that processes the input data to generate response plans for pre-defined threat scenarios.

Current researches focus on considering the impact of attacks by evaluating their severity and consequences, leaving aside the impact of security actions in mitigating the effects of such attacks. Dini and Tiloca [34], for instance, propose a simulation framework that evaluates the impact of cyber-physical attacks, discusses the attack ranking process, and analyses different mitigation actions. Kundur et al. [35], propose a paradigm for cyber attack impact analysis that employs a graph-theoretic structure and a dynamical systems framework to model the complex interactions amongst the various system components. Squoras et al. [36] present a qualitative assessment of the cyber attack impact on critical Smart Grid infrastructures. The approaches involve quantifying the effects of given classes of cyber attack, providing information on the degree of disruption that such class of attacks enable, and identifying sophisticated dependencies between the cyber and physical systems, but leave aside the impact of mitigation actions in the attack's impact calculation.

In contrast to previous research, Arpan et al. [37] focus on cost optimization for generation and evaluation of response plans using Attack Countermeasure Trees (ACTs). This approach overcomes the limitations of attack trees and defense trees approaches by taking into account attacks as well as countermeasures in the form of detection and mitigation events. Our proposed model differ from ACTs in the way that we evaluate ongoing and possible attack scenarios using reactive and

proactive strategies while considering dependencies to evaluate the operational and financial impact of security events.

In terms of operational impact assessment, probabilistic models have been researched as an adequate assessment of impacts or risks posed due to attacks or found vulnerabilities [38, 39, 40]. However, often imperfect knowledge is not considered [38] or dependency cycles pose a problem [39, 40]. Barreto et al. [41, 42], only consider direct impacts as approaches to mission modeling, neglecting transitive impacts and/or defining a manual description of all dependencies between individual devices inside one organization, which is, in most of the cases an unfeasible process. Other impact propagation approaches, able to handle such details, are not probabilistic based and degrade to a handcrafted propagation algorithm with arbitrary scores [43, 44]. Chung et al. [45] consider a probabilistic approach as well to determine the likelihoods of explicit attack paths. However, Chung et al.'s probability theory [45] is not sound and voids fundamental principles of probabilistic inference in multiply connected graphs.

7. Conclusion

In this paper we introduced a dynamic and automated risk management response system that generates response plans containing mitigation actions and their corresponding financial and operational assessments. The prototype provides several features (e.g., attack graph generation, threat risk quantification, operational and financial impact assessments) that focuses in three main aspects: (i) a dynamic evaluation of mitigation actions, (ii) an automated response plan generation of all possible combinations of mitigation actions, and (iii) an automated selection of the best response plan for a given threat scenario.

The prototype is said to be dynamic, since it performs a snapshot with a regular frequency (within minutes), at which, the current conditions of the system are assessed. Input data vary at each snapshot, indicating the exact number of active devices on the system (if a given equipment has been turned off, it will be discarded for the current snapshot). Since the Annual Infrastructure Value (AIV) and the coverage (COV) parameters depend directly on the number of the system's active devices, its computation dynamically changes at each snapshot, which in turn changes the risk mitigation (RM) and the RORI values for the set of evaluated mitigation actions. Similarly, the number of authorized mitigation actions may vary at each snapshot as a consequence of an update of a rule. In such a case, the combination process will either increase or reduce, making it possible to change the order of the best response plans for the same threat scenario.

The prototype is said to be automated, since it performs the process as a whole automated chain (i.e., from the detection of the threat, to the visualization of the selected response plan). The process is designed to help security administrators in the decision making process. The prototype generates the attack graph, quantifies the risk, assesses the operational and financial impacts, generates the response plans, and select the best semi-optimal response plan against the identified threat in an automated fashion. It does not enforce mitigation actions in an automated manner. Instead, it provides an assessment of the current system conditions in order to highlight the appropriate response strategies to be enforced by the administrator. For critical infrastructures, the selection of mitigation actions generally requires manual intervention from the operator, and in some cases, an approval of supervisors or a more advanced system operators. The DRMRS provides quantitative support to guide SCADA operators into the best financial and operational response action.

In terms of prototype verification and validation, a wide range of tests have been performed for each DRMRS component to evaluate that output results are produced correctly and accurately

considering well mastered input, and that all components are completed integrated to exchange information correctly and within the expected time. Results show that all components interact appropriately to generate the best response plan for a each threat scenario. Some DRMRS components (e.g., AGG, RFIA) have shown scalability issues due to the exponential growth of the total number of combinations for connections or mitigation actions. Such issues have been resolved with optimization algorithms that reduce the number of combinations by discarding those candidates that perform below a given threshold.

Conflicts among mitigation actions have been considered in the evaluation process. Every mitigation action is associated to a generic type (e.g., patching, restart, shutdown). Each mitigation action type has an associated restriction (e.g., mutually exclusive, totally restrictive, partially restrictive). For instance, an action that suggests to shutdown equipment E_1, is totally restrictive with any other action associated to E_1 but it can be perfectly combined with actions to be implemented on another equipment. Conflicts of restrictive mitigation actions are, therefore, avoided at the first stage of the evaluation process. In addition, mitigation actions for which some information is missing (e.g., cost, benefit, coverage), are discarded by the process before the combination. However, the prototype does not consider the financial impact of the following situations: (i) the selected response plan requires to implement a mitigation action that is already activated in the system, (ii) the selected response plan requires to deactivate an action that was previously active, (iii) mutually exclusive actions that are proposed as the best response plans for multiple threat scenarios affecting the same target system. Future work will concentrate on an evolution of the RORI metric that considers such situation in the evaluation of multiple mitigation actions.

Acknowledgments:

The research in this paper has received funding from the Panoptesec project, as part of the Seventh Framework Programme (FP7) of the European Commission (GA 610416). Authors would like to thank Panoptesec use case providers for their contribution of the case study, and partners from the university of Rome for their contribution of the visualization results.

[1] ISO/IEC International Standard, *Information technology - Security techniques - Information security risk management*, ISO/IEC 27005:2011(E), Second edition, Jun. 2011.

[2] ISO/IEC International Standard, *Information technology - Security techniques - Information security management systems - Requirements*, ISO/IEC 27001:2013, Second edition, Sep. 2013.

[3] ETSI, TISPAN, Methods and protocols, *Method and proforma for Threat, Risk, Vulnerability Analysis*, ETSI TS 102 165-1, v4.2.1, Dec. 2006. Available at: `http://portal.etsi.org/mbs/Referenced%20Documents/ts_10216501v040201p.pdf`, 2006.

[4] National Institute of Standards and Technology, *Information Security - Guide for Conducting Risk Assessments*, Computer Security Division, Information Technology Laboratory, NIST Special Publication 800-30, Revision 1. Available at: `http://csrc.nist.gov/publications/nistpubs/800-30-rev1/sp800_30_r1.pdf`, 2013.

[5] ANSSI, Agence Nationale de la Securite des Systemes D'Information, *EBIOS - Expression des Besoins et Identification des Objectifs de Securite*. Available at: `http://www.ssi.gouv.fr/IMG/pdf/EBIOS-1-GuideMethodologique-2010-01-25.pdf`, 2011.

[6] W. Kanoun, S. Papillon, and S. Dubus. *Elementary Risks: Bridging Operational and Strategic Security Realms*, In 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp. 278–286. IEEE, Bangkok, Thailand, 2015.

[7] S. Jajodia, S. Noel, and B. O'Berry, *Topological Analysis of Network Attack Vulnerability*, In Managing Cyber Threats: Issues, Approaches and Challenges, chapter 5. Kluwer Academic Publisher, pp. 247–266, 2005.

[8] O. Sheyner, J. Haines, S. Jha, R. Lippmann, and J. M. Wing, *Automated Generation and Analysis of Attack Graphs*, In 2002 IEEE Symposium on Security and Privacy, Oakland, California, USA, pp. 254–265, 2002.

[9] K. Ingols, R. Lippmann, and K. Piwowarski, *Practical Attack Graph Generation for Network Defense*, In ACSAC 2006: 22nd Annual Computer Security Applications Conference, Washington, D.C.: IEEE Computer Society, pp. 121–130, 2006.

[10] X. Ou, S. Govindavajhala, and A. W. Appel, *MulVAL: A Logic-based Network Security Analyzer*, In 14th USENIX Security Symposium, Baltimore, MD, USA, August 2005.

[11] G. Gonzalez-Granadillo, E. Alvarez, A. Motzek, M. Merialdo, J. Garcia-Alfaro, and H. Debar, *Towards an Automated and Dynamic Risk Management Response System*, 21st Nordic Conference on Secure IT Systems NordSec, 2016.

[12] Endsley, Mica R., *Towards a Theory of Situation Awareness in Dynamic Systems*, Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 37(1), pp. 32–64, March 1995.

[13] T. Bray, *The JavaScript Object Notation (JSON) Data Interchange Format*, Internet Engineering Task Force RFC7159, https://tools.ietf.org/html/rfc7159, 2014.

[14] Templeton, Steven J., and Karl Levitt, *A requires/provides model for computer attacks*, In Proceedings of the 2000 workshop on New security paradigms, pp. 31-38. ACM, 2001.

[15] Samarji, Layal, Frederic Cuppens, Nora Cuppens-Boulahia, Wael Kanoun, and Samuel Dubus, *Situation calculus and graph based defensive modeling of simultaneous attacks*, In Cyberspace Safety and Security, pp. 132-150. Springer International Publishing, 2013.

[16] Cuppens, Frederic, and Rodolphe Ortalo, *Lambda: A language to model a database for detection of attacks*, In International Workshop on Recent Advances in Intrusion Detection, pp. 197-216. Springer Berlin Heidelberg, 2000.

[17] S. Yi, P. Yong, X. Qi, T. Wang, Z. Dai, H. Gao, J. Xu, J. Wang, and L. Xu. *Overview on Attack Graph Generation and Visualization Technology*, In 2013 International Conference on Anti-Counterfeiting, Security and Identification (ASID), pp. 1-6. IEEE, 2013.

[18] Cuppens, Frederic, Alexandre Miege, *Alert correlation in a cooperative intrusion detection framework.*, In proceedings of 2002 IEEE Symposium on Security and privacy, Washington DC, pp. 202-215. IEEE Computer Society, 2002.

[19] P. Mell, K. Scarfone, and S. Romanosky. *CVSS: A Complete Guide to the Common Vulnerability Scoring System Version 2.0*, FIRST, Jun 2007, Available at: `http://www.first.org/cvss/cvss-guide.pdf`.

[20] U.S. National Institute of Standards and Technologies, *National Vulnerability Database: U.S. government repository of standards based vulnerability management data*, available at: `https://nvd.nist.gov/`

[21] W. Kanoun, S. Dubus, S. Papillon, N. Cuppens???Boulahia, and F. Cuppens. *Towards Dynamic Risk Management: Success Likelihood of Ongoing Attacks*, Bell Labs Technical Journal, vol. 17(3), pp. 61–78, 2012.

[22] G. Rubino, and B. Sericola, *Markov Chains and Dependability Theory*, Cambridge University Press, 2014.

[23] A. Motzek, R. Moller, M. Lange, and S. Dubus, *Probabilistic Mission Impact Assessment based on Widespread Local Events*, In NATO IST-128 Workshop on Cyber Attack Detection, Forensics and Attribution for Assessment of Mission Impact, 2015.

[24] A. Motzek, R. Moller, *Context- and Bias-Free Probabilistic Mission Impact Assessment*, Universität zu Lübeck, Institut für Informationssysteme, Tech. Rep., under review. Available at: `http://www.ifis.uni-luebeck.de/~motzek/techrep-miamim.pdf`, 2016.

[25] G. Gonzalez-Granadillo, A. Motzek, J. Garcia-Alfaro, and H. Debar, *Selection of Mitigation Actions Based on Financial and Operational Impact Assessments*, ARES 2016: 11th International Conference on Availability, Reliability and Security, Salzburg, Austria, August 2016.

[26] G. Gonzalez-Granadillo, M. Belhaouane, H. Debar, G. Jacob, *RORI-based Countermeasure Selection Using the OrBAC Formalism*, International Journal of Information Security, Springer, vol. 13(1), pp. 63–79, February, 2014.

[27] G. Gonzalez-Granadillo, J. Garcia-Alfaro, E. Alvarez, M. El-Barbori, and H. Debar,*Selecting Optimal Countermeasures for Attacks Against Critical Systems Using the Attack Volume Model and the RORI Index*, Computers and Electrical Engineering Journal vol. 47, pp. 13–34, 2015.

[28] G. Gonzalez-Granadillo, J. Garcia-Alfaro, and H. Debar, *A Polytope-based Approach to Measure the Impact of Events Against Critical Infrastructures*, Journal of Computer and System Sciences, Elsevier, vol. 83(1), pp. 3-21, 2016.

[29] O. Netkachov, P. Popov, K. Salako., *Quantification of the Impact of Cyber Attacks in Critical Infrastructures*, Computer Safety, Reliability, and Security, Springer vol 8696, pp. 316–327, 2014.

[30] U. Tatar, H. Bahsi, A. Gheorghe., *Impact assessment of cyber attacks: A quantification study on power generation systems*, 11th System of Systems Engineering Conference (SoSE'16), 2016.

[31] Y. Cherdantseva, P. Burnap, A. Blyth, P. Eden, K. Jones, H. Soulsby, K. Stoddart, *A review of cyber security risk assessment methods for SCADA systems*, Computers & Security, 56:1–27, 2016.

[32] I. Kotenko, A. Chechulin *A Cyber Attack Modeling and Impact Assessment Framework*, 5th International Conference on Cyber Conflict, 2013.

[33] I. Kotenko, E. Doynikova *Dynamical calculation of security metrics for countermeasure selection in computer networks*, 24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, 2016.

[34] G. Dini, M. Tiloca,*On Aimulative Analysis of Attack Impact in Wireless Sensor Networks*, 18th Conference on Emerging Technologies & Factory Automation (ETFA), 2013.

[35] D. Kundur, X. Feng, S. Liu, T. Zourntos, K.L. Butler-Purry, *Towards a Framework for Cyber Attack Impact Analysis of the Electric Smart Grid*, International Conference on Smart Grid Communications (SmartGridComm), pp. 244–249, 2010.

[36] K. I. Sgouras, A. D. Birda, D. P. Labridis, *Cyber Attack Impact on Critical Smart Grid Infrastructures*, Innovative Smart Grid Technologies Conference (ISGT), 2014.

[37] R. Arpan, K. Dong Seong, and S. T. Kishor, *ACT: Towards Unifying the Constructs of Attack and Defense Trees*, Security Communication Networks, pp. 1–15, 2012.

[38] L. Wang, T.a Islam, T. Long, A. Singhal, S. Jajodia *An Attack Graph-based Probabilistic Security Metric*, Data and Applications Security XXII. Springer Berlin Heidelberg, pp. 283–296, 2008.

[39] L. Yu, H. Man *Network Vulnerability Assessment Using Bayesian Networks*. International Society for Optics and Photonics, 2005.

[40] P. Xie, J. Li, X. Ou, P. Liu, R. Levy, *Using Bayesian networks for Cyber Security Analysis*, International Conference on Dependable Systems and Networks, pp. 211–220, 2010.

[41] A. Barreto, P. Costa, and E. Yano, *A Semantic Approach to Evaluate the Impact of Cyber Actions to the Physical Domain*, In Proceedings of the 7th International Conference on Semantic Technologies for Intelligence, Defense, and Security, pp. 64-71, USA, 2012.

[42] A. Barreto, P. Costa, and E. Yano, *Using a Semantic Approach to Cyber Impact Assessment*, In Proceedings of the 8th International Conference on Semantic Technologies for Intelligence, Defense, and Security, pp. 101–108, USA, 2013.

[43] N. Kheir, H. Debar, N. Cuppens-Boulahia, F. Cuppens, J. Viinikka *Cost Evaluation for Intrusion Response Using Dependency Graphs*, International Conference on Network and Service Security, 2009.

[44] J. Marko, C. Thul, P. Martini, *Graph-based Metrics for Intrusion Response Measures in Computer Networks*, 32nd IEEE Conference on Local Computer Networks, 2007.

[45] C. Chung, P. Khatkar, T. Xing, J. Lee, D. Huang, *NICE: Network Intrusion Detection and Countermeasure Selection in Virtual Network Systems*, IEEE Trans. Dependable Sec. Comput. vol. 10(4),pp. 198–211, 2013.