VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Justinas JANULEVIČIUS

METHOD OF INFORMATION SECURITY RISK ANALYSIS FOR VIRTUALIZED SYSTEMS

DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES,
INFORMATICS ENGINEERING (07T)

VGTU
LEIDYKLA
TECHNIKA
Vilnius 2016

Doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2012–2016.

Supervisor

Prof. Dr Habil. Antanas ČENYS (Vilnius Gediminas Technical University, Informatics Engineering – 07T).

The Dissertation Defense Council of Scientific Field of Informatics Engineering of Vilnius Gediminas Technical University:

Chairman

Prof. Dr Olegas VASILECAS (Vilnius Gediminas Technical University, Informatics Engineering – 07T).

Members:

Prof. Dr Arnas KAČENIAUSKAS (Vilnius Gediminas Technical University, Informatics Engineering – 07T),
Dr Tomas KRILAVIČIUS (Vytautas Magnus University, Informatics – 09P),
Assoc. Prof. Dr Raimundas MATULEVIČIUS (University of Tartu, Estonia, Informatics Engineering – 07T),
Prof. Dr Habil. Rimantas ŠEINAUSKAS (Kaunas University of Technology, Informatics Engineering – 07T).

The dissertation will be defended at the public meeting of the Dissertation Defense Council of Informatics Engineering in the Senate Hall of Vilnius Gediminas Technical University at **10 a. m. on 19 December 2016**.

Address: Saulėtekio al. 11, LT-10223 Vilnius, Lithuania.
Tel.: +370 5 274 4956; fax +370 5 270 0112; e-mail: doktor@vgtu.lt

A notification on the intend defending of the dissertation was sent on 18 November 2016. A copy of the doctoral dissertation is available for review at the VGTU repository http://dspace.vgtu.lt/ and at the Library of Vilnius Gediminas Technical University (Saulėtekio al. 14, LT-10223 Vilnius, Lithuania).

VGTU leidyklos TECHNIKA 2389-M mokslo literatūros knyga

ISBN 978-609-457-967-7

justinas.janulevicius@vgtu.lt

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Justinas JANULEVIČIUS

VIRTUALIZUOTŲ SISTEMŲ INFORMACIJOS SAUGOS RIZIKOS ANALIZĖS METODO KŪRIMAS IR TAIKYMAS

DAKTARO DISERTACIJA

TECHNOLOGIJOS MOKSLAI,
INFORMATIKOS INŽINERIJA (07T)

VGTU
LEIDYKLA
Vilnius TECHNIKA 2016

Disertacija rengta 2012–2016 metais Vilniaus Gedimino technikos universitete.

Vadovas

prof. habil. dr. Antanas ČENYS (Vilniaus Gedimino technikos universitetas, informatikos inžinerija – 07T).

Vilniaus Gedimino technikos universiteto Informatikos inžinerijos mokslo krypties disertacijos gynimo taryba:

Pirmininkas

prof. habil. dr. Olegas VASILECAS (Vilniaus Gedimino technikos universitetas, informatikos inžinerija – 07T).

Nariai:

prof. dr. Arnas KAČENIAUSKAS (Vilniaus Gedimino technikos universitetas, informatikos inžinerija – 07T),
dr. Tomas KRILAVIČIUS (Vytauto Didžiojo universitetas, informatika – 09P),
doc. dr. Raimundas MATULEVIČIUS (Tartu universitetas, Estija, informatikos inžinerija – 07T),
prof. habil. dr. Rimantas ŠEINAUSKAS (Kauno technologijos universitetas, informatikos inžinerija – 07T).

Disertacija bus ginama viešame Informatikos inžinerijos mokslo krypties disertacijos gynimo tarybos posėdyje **2016 m. gruodžio 19 d. 10 val.** Vilniaus Gedimino technikos universiteto senato posėdžių salėje.

Adresas: Saulėtekio al. 11, LT-10223 Vilnius, Lietuva.
Tel.: (8 5) 274 4956; faksas (8 5) 270 0112; el. paštas doktor@vgtu.lt

Pranešimai apie numatomą ginti disertaciją išsiųsti 2016 m. lapkričio 18 d.

Disertaciją galima peržiūrėti VGTU talpykloje http://dspace.vgtu.lt/ ir Vilniaus Gedimino technikos universiteto bibliotekoje (Saulėtekio al. 14, LT-10223 Vilnius, Lietuva).

Abstract

The growth of usage of Information Technology (IT) in daily operations of enterprises causes the value and the vulnerability of information to be at the peak of interest. Moreover, distributed computing revolutionized the outsourcing of computing functions, thus allowing flexible IT solutions. Since the concept of information goes beyond the traditional text documents, reaching manufacturing, machine control, and, to a certain extent – reasoning – it is a great responsibility to maintain appropriate information security.

Information Security (IS) risk analysis and maintenance require extensive knowledge about the possessed assets as well as the technologies behind them, to recognize the threats and vulnerabilities the infrastructure is facing. A way of formal description of the infrastructure – the Enterprise Architecture (EA) – offers a multiperspective view of the whole enterprise, linking together business processes as well as the infrastructure. Several IS risk analysis solutions based on the EA exist. However, lack of methods of IS risk analysis for virtualization technologies complicates the procedure, thus leading to reduced availability of such analysis.

The dissertation consists of an introduction, three main chapters and general conclusions. The first chapter introduces the problem of information security risk analysis and its' automation. Moreover, state-of-the-art methodologies and their implementations for automated information security risk analysis are discussed.

The second chapter proposes a novel method for risk analysis of virtualization components based on the most recent data, including threat classification and specification, control means and metrics of the impact.

The third chapter presents an experimental evaluation of the proposed method, implementing it to the Cyber Security Modeling Language (CySeMoL) and comparing the analysis results to well-calibrated expert knowledge.

It was concluded that the automation of virtualization solution risk analysis provides sufficient data for adjustment and implementation of security controls to maintain optimum security level.

Reziumė

Informacinių technologijų (IT) taikymo augimas kasdienėje organizacijų veikloje pasižymi svarbios informacijos talpinimo informacinėse sistemose padidėjimu. Dėl šios priežasties, informacijos vertė ir jos pažeidžiamumas yra itin kritiški organizacijų veiklų tęstinumui užtikrinti. Šiuolaikinės kompiuterizacijos paradigmos, pavyzdžiui, išskaidytosios kompiuterinės architektūros, tapo pagrindu kuriant lanksčius IT sprendimus paslaugų teikimą perleidžiant trečiosioms šalims. Informacijos sąvoka yra itin plati ir apima tekstinę informaciją, gamybą, įrenginių valdymą ar net išvadų generavimą. Todėl informacijos apsauga yra labai svarbi organizacijos veiklos dalis.

Informacijos apsaugos (IS) rizikos analizė ir valdymas reikalauja nuodugnių žinių apie turimas informacines vertybes bei technologijas. Tai reikalinga tam, kad būtų atpažįstamos grėsmės ir pažeidžiamumai, kylantys turimai infrastruktūrai. Formalus infrastruktūros aprašas – verslo architektūra (*angl. Enterprise Architecture*) leidžia sudaryti daugialypį organizacijos vaizdą, jungiantį verslo procesus bei infrastruktūrą. Yra sukurta nemažai IS rizikos analizės technologijų, veikiančių verslo architektūros pagrindu, tačiau trūksta šių technologijų taikymo virtualizacijos technologijoms pavyzdžių. Jų nebuvimas reikalauja papildomų veiksmų aprašant ir vertinant virtualizacijos saugumą. Taigi, tai komplikuoja šių technologijų panaudojimą infrastruktūroje. Šioje disertacijoje nagrinėjami virtualizacijos technologijų informacinės apsaugos rizikos vertinimo verslo architektūroje ypatumai ir šio vertinimo metodo sudarymas ir pritaikymas.

Disertaciją sudaro įvadas, trys pagrindiniai skyriai ir bendrosios išvados. Pirmajame skyriuje apibrėžiama informacijos apsaugos rizikos analizės automatizavimo problema. Šiame skyriuje taip pat apžvelgiami informacijos apsaugos rizikos analizės pažangiausi metodai ir jų taikymas.

Antrajame skyriuje siūlomas naujas informacijos apsaugos rizikos analizės metodas virtualizacijos technologijoms. Siūlomas metodas apima virtualizacijos grėsmių klasifikavimą ir specifikavimą, kontrolės priemonių parinkimą bei kiekybinį rizikos analizės vertinimo metodą.

Trečiajame skyriuje pristatomas eksperimentinis pasiūlyto metodo vertinimas įgyvendinant jį kibernetinės apsaugos modeliavimo kalbos (CySeMoL) aplinkoje ir palyginant vertinimo rezultatus su kalibruotomis ekspertinėmis šios srities žiniomis.

Eksperimentinio tyrimo metu nustatyta, kad siūlomas virtualizacijos technologijų rizikos analizės automatizavimo metodas suteikia pakankamai duomenų apie IS lygį, kad būtų galima modeliuoti virtualizacijos komponentų saugumo priemonių taikymą.

Notations

Symbols

AC – Access vulnerability variable (page 40);
I_e – The information score (page 47);
$C(e)$ – The calibration score (page 47);
CF – Configuration vulnerability variable (page 40);
H_e – The average response entropy score (page 47);
s – Standard deviation (page 44);
SV – Software Vulnerability (page 40);
TI – Traffic Isolation vulnerability variable (page 40);
VF – Virtual Firewall vulnerability variable (page 40).

Abbreviations

ADtree – An Attack-Defense tree concept (Kordy *et al.* 2012);
CCM – The Cloud Control Matrix issued by the Cloud Security Alliance (Cloud Security Alliance 2014);
CERT – Computer Emergency Response Team;
CVE – Common Vulnerabilities and Exposures code identifier managed by NIST;
CVSS – Common Vulnerability Scoring System;
CySeMoL – the Cyber Security Modeling Language;

DHS – The United States Department of Homeland Security;
DoS – Denial of Service attack;
EA – Enterprise Architecture;
EDB – Exploit Database;
FIRST – Forum of Incident Response and Security Teams;
IS – Information Security;
ISRA – Information security risk analysis;
IT – Information Technology;
NCCIC – National Cybersecurity & Communications Integration Center;
NIST – National Institute of Standards and Technology;
NVD – National Vulnerability database managed by NIST;
OS – Operating System;
SANS – System administration, Audit, Networking and Security Institute;
VM – Virtual Machine.

Domain Specific Definitions

Asset – A major application, general support system, high impact program, physical plant, mission critical system, personnel, equipment, or a logically related group of systems (Kissel 2013);

Attack – attempt to destroy, expose, alter, disable, steal or gain unauthorized access to or make unauthorized use of an asset (ISO 27000: 2014);

Attack Scenario – algorithm or calculation combining one or more measures with associated decision criteria (ISO 27000: 2014);

Control – measure that is modifying risk. Controls include any process, policy, device, practice, or other actions which modify risk (ISO 27000: 2014);

Cybersecurity – The ability to protect or defend the use of cyberspace from cyber attacks (Kissel 2013);

Data – A subset of information in an electronic format that allows it to be retrieved or transmitted (Kissel 2013);

Domain – A set of subjects, their information objects, and a common security policy (Kissel 2013);

Enterprise Architecture (EA) – The description of an enterprise's entire set of information systems: how they are configured, how they are integrated, how they interface to the external environment at the enterprise's boundary, how they are operated to support the enterprise mission, and how they contribute to the enterprise's overall security posture (Kissel 2013);

Expertise – expert opinion claiming the most expertise for a given item (Cooke 1991);

Exploitability – metrics that reflect the characteristics of the thing that is vulnerable, which we refer to formally as the vulnerable component (FIRST 2014);

Information Security – The protection of information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide confidentiality, integrity, and availability (Kissel 2013);

Information Security Risk – The risk to organizational operations (including mission, functions, image, reputation), organizational assets, individuals, other organizations, and

the Nation due to the potential for unauthorized access, use, disclosure, disruption, modification, or destruction of information and/or information systems (Kissel 2013);

Risk – effect of uncertainty on objectives. An effect is a deviation from the expected – positive or negative. Uncertainty is the state, even partial, of deficiency of information related to, understanding or knowledge of, an event, its consequence, or likelihood. Risk is characterized by reference to potential events and consequences, or a combination of these and is expressed in terms of a combination of the consequences of an event and the associated likelihood of occurrence (ISO 27000: 2014);

Risk Analysis – process to comprehend the nature of risk and to determine the level of risk. It provides the basis for risk evaluation and decisions about risk treatment (ISO 27000: 2014);

Risk Management – coordinated activities to direct and control an organization with regard to risk (ISO 27000: 2014);

Security Requirement – need or expectation that is stated, generally implied or obligatory (ISO 27000: 2014);

Taxonomy – a controlled vocabulary consisting of preferred items, all of which are connected in a hierarchy or polyhierarchy (ANSI 2010);

Threat – potential cause of an unwanted incident, which may result in harm to a system or organization (ISO 27000: 2014);

Vulnerability – weakness of an asset or control that can be exploited by one or more threats (ISO 27000: 2014).

Contents

[1] The annexes are supplied in the enclosed compact disk.

Introduction

Problem Formulation

Information Security (IS) is one of the major concerns of the modern organization. As the usage of Information Technology (IT) is constantly growing, evolving and reaching most of the activities within and outside the organization, the IS plays the key role in ensuring the fluent process execution. Due to the expanding interest for the IS in the last decades, there have been numerous studies worldwide on how to control the IS up to a chosen level. This has resulted in a good understanding in what is required given the situation. To simplify the process of ensuring the IS, specific automation tools have been developed, implementing the results of the scientific studies.

Defining the IS situation requires knowledge of the analyzed infrastructure, the threats and vulnerabilities it is facing and a method to represent the magnitude of risk for each component that leads to a thorough picture of the IS situation. Based on the situation, a plan for risk mitigation can be made. Factors influencing the IS situation are described and analyzed in the study. They vary depending on the types of components used in the infrastructure and the relations between them.

One of the modern ways of defining the IS situation is by performing a risk analysis of (a part of) the IT infrastructure. A formal way of representing the IT

infrastructure – the Enterprise Architecture (EA) – allows a formal definition of the existing hardware and software components, legal agreements and business processes. Therefore, it gives an ability to represent the role of the IT infrastructure within the organizational environment.

Virtualization, being a relatively new paradigm, has revolutionized the architecture of the IT infrastructure, as it is no longer required to use dedicated hardware for specific processes. Moreover, it motivates the outsourcing of IT infrastructure services to third-party vendors, as the costs can be cut down dramatically.

Virtualization technology risk analysis is a very important topic when dealing with modern IT architecture IS. It helps to correctly architect the IT infrastructure, while minimizing the IS risks of the organization.

However, lack of methods and implementations for virtualization information security risk analysis leaves a gap in this field. A method for information security automation for virtualization environments is the cornerstone for the virtualization IS, as automated analysis enables unbiased, real-time evaluation of the situation with instant reaction to changes.

Relevance of the Thesis

The use of virtualization technologies exist in most of the organizations and is currently growing. The most typical form of virtualization - Cloud Computing - reaches the organization's IT infrastructure by various services, including e-mail, file sharing and others. Understanding the virtualization risks and their magnitude is crucial in ensuring the IS of these components.

To fully understand the risks of virtualization, a method, evaluating these risks within the context of the organization's infrastructure and process architecture is required. It can be achieved by analyzing a formally described model. Modeling the overall organization's architecture, including business processes, hardware and software can be achieved using the formal Enterprise Architecture model. However, lack of special meta-concepts for virtualization components in strictly formed Enterprise Architecture models is a drawback, as such components cannot be described properly, using dedicated elements.

Due to the nature of virtualization components of sharing physical resources between the virtual machines – additional threat vectors exist compared to the regular single machine computing. Specifying these vectors and acquiring metrics is required for successful risk analysis.

Due to the complex nature of the possible architectures and their components in the Enterprise Architecture model as well as the probabilistic output of Enterprise Architecture analysis tools, the metrics for the attack vectors are re-

quired to be quantitative, i.e. provide numerical scores for the exploitability of the existing vulnerabilities.

While there have been numerous approaches on IS risk analysis, virtualization is still vaguely covered. Moreover, a method for information security automation for virtualization environments to be used in an Enterprise Architecture model is not present.

This research is dedicated to develop a method for quantitative IS risk analysis for virtualization components that can be implemented to an automated EA analysis solution.

The Object of Research

The object of this research is information security risk analysis of virtualized computer systems.

The Aim of the Thesis

The main goal of this research is to provide a method for information security risk analysis for virtualized computer systems.

The Tasks of the Thesis

The following tasks are executed to achieve the goal:

1. Review of the existing methods for the evaluation of information security risk in virtualized systems.
2. Develop a virtualization threat landscape, defining the scope of the virtualization threats.
3. Propose a method for quantitative risk analysis by adapting existing IS threat database data to eliminate and/or support the results of expert evaluation.
4. Perform comparative statistical analysis to evaluate the accuracy of calculated threat scorings in relation to expertise knowledge and define its uncertainty.
5. Evaluate the method improvements in a working Enterprise Architecture analysis system to provide scenario-analysis.

Research Methodology

Methods of comparative and literature analysis were used to analyze the research object. The literature is chosen based on the impact factor, reputation of the author and relevance to the topic. Methods of information security risk analysis, scoring and threat identification are used to develop a method for virtualization technology information security risk analysis. The methods of experimental research have been used to validate the correctness of the proposed method and its implementation.

Scientific Novelty of the Thesis

The scientific novelty of this study is specified as follows:

1. A novel method of virtualization information security risk analysis is proposed. It includes an extensive definition of the threat landscape of virtualization technologies, association of the threats with covering control means, definition of requirements for secure virtualization component usage, attack scenarios, and variables influencing the security of the components.
2. A new numerical procedure, combining the recent exploitability scoring to define the probability values of the variables, influencing the virtualization security is introduced and analyzed.
3. A new variable scoring correctness validation method using adapted statistical procedures for expertise elicitation is proposed. The results of this procedure provide calibration and entropy scores of uncertain information, therefore novelty uncertainty evaluation of the scoring to define the correctness of the scoring.

Practical Value of the Research Findings

A method for virtualized system information security risk analysis using expert knowledge and validated vulnerability and exploit databases has been proposed. This method can be used as a stand-alone tool, or implemented into existing tools for information security analysis automation. An implementation of the method to the Cyber Security Modeling Language is proposed in this thesis.

The Defended Statements

The following statements based on the results of present investigation are the hypotheses to be defended:

1. The proposed method allows for the information security risk probabilities to be evaluated from processed vulnerability exploitation scoring of Common Vulnerability and Exploit database.
2. The proposed method substitutes the expert knowledge on virtualization information security risk.

Approval of the Research Findings

The author has published 4 publications in 4 reviewed scientific journals. The author has made three presentations at three scientific conferences:

- 12th IMEKO TC10 Workshop on Technical Diagnostics: New Perspectives in Measurements, Tools and Techniques for Industrial Applications. June 6–7, 2013, Florence, Italy.
- The 2014 International Conference on Information and Network Security (ICINS 2014). April 11–12, 2014, Jeju, South Korea.
- The XXI IMEKO World Congress. August 29–September 5, Prague, Czech Republic.

Structure of the Dissertation

The dissertations consists of 112 pages, includes 20 figures and 18 tables.

1

Virtualization Information Security Assurance and Methods

This chapter provides an overview of existing research in the field virtualization information security, reviews the theoretical aspects of virtualization technologies, information security and its extent. Moreover, methods for information security risk analysis and the components for risk modeling are analyzed in this chapter as well as various applications of the techniques in real life solutions are presented.

Recent growth of information technology (IT) application fields across various industries has led the industry to dramatic optimization results, thus saving the resources. One of the revolutionary conceptions – virtualization – is not only efficient in the monetary perspective, but also if handled correctly – security wise. The threat landscape of virtualization technologies is broader compared to traditional computing technologies due to the fact that physical resources are unavoidably shared, thus leaving more opportunities of exploits in this field. This leads to the fact that for virtualization to be beneficial, security must be in the top priorities.

However, virtualization being a relatively new concept, the methodologies for security and risk management of such environments are mostly under development (Chandramouli 2014). Moreover, proper security assessment, including risk analysis of the virtualization components is important for identification of the security strategy aspects.

The results of experiments, presented in this chapter are published in two papers (Janulevičius, Goranin 2013, Janulevičius, Čenys 2014).

1.1. Virtualization of Information Technology Systems

Virtualization is a process of simulating a physical computer on the existing hardware platform. It is achieved by using a virtualization layer that decouples the computer software from the existing hardware resources. The virtualization layer distributes the present hardware resources to the instances of virtualization, called virtual machines (SP 800-125: 2011). Virtual machine is a portion of the physical resources, dedicated to run as simulation of a physical machine (Portnoy 2012). Although the formal requirements for virtualization have been defined over fourty years ago (Popek, Goldberg 1974) such kind of technology came to prominence only very recently.

Decoupling of the physical hardware from the software can be beneficial when a powerful hardware setup needs to be divided into weaker-resource machines. On the contrary, when having two machines with less than required power, virtualization can serve in merging the resources (Portnoy 2012).

In real-life applications, virtualization has shown the benefits of resource elasticity, saving of energy and isolation of applications. Beyond this, virtualization also provides disaster recovery capabilities, fault tolerance ant the ability of cloning and migration of virtual machines. Therefore the efficiency of the IT increases as most of these processes can be automated using advanced virtualization systems (Ottenheimer, Wallace 2012).

One of the most successful virtualization applications is the Cloud Computing. It facilitates the beneficial features of virtualization by providing it as a service. This means that the customer of the cloud does not have to worry about the physical hardware. Once the service is acquired, the resources can be easily increased by simply adding as many resources from as many physical machines as necessary.

The hypervisor is a layer of software placed between the hardware and one or more virtual machines that it supports. Therefore, it controls and manages the interactions between virtual machines and the hardware shared by the virtual machines. The main hypervisor characteristics are (Popek, Goldberg 1974):

- provision of an environment identical to the physical environment;
- providing it with minimal performance cost;
- retaining complete control of the system resources.

Two types of hypervisors exist – Type 1 (Fig. 1.1 a) and Type 2 (Fig. 1.1 b) differing by the deployment method.

Type 1 hypervisors run directly on the physical hardware. They communicate directly with the hardware resources. This leads to better performance compared to the Type 2 hypervisors, as less processing overhead is required to run the hypervisor. Moreover, they are considered to perform better security-wise, as there are less points for security breaches.

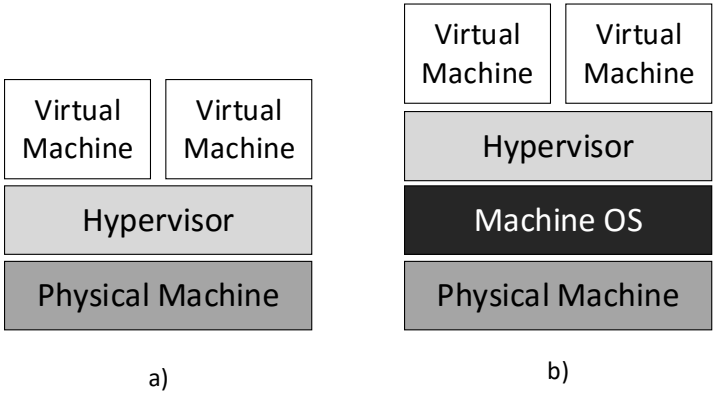


Fig. 1.1. Comparison of: a) Type 1 and b) Type 2 hypervisor architecture (source: author)

A Type 2 hypervisor runs as an application on an operating system. Support for hardware components therefore is inherited from the underlying operating system, therefore making it much easier to deploy. However, the processes of the underlying systems reduce the performance of such hypervisors. Moreover, compatibility issues reduce the reliability of performance (Portnoy 2012).

Virtual machines serve as an environment for traditional operating systems and applications, running on top of a hypervisor or a physical server. Within a virtual machine the processes are similar to the ones in a physical machine, however there is no dedicated physical hardware. Instead there is a share of a resource pool. Virtual machines allow the operating systems to access virtual devices, which, from an operating system point of view are considered to be exactly the same as physical ones.

The most popular realization of such virtualization technologies is Cloud Computing. It is a type of service based on the abilities of virtualization, bringing the benefits of resource pooling, scalability and advanced recovery capabilities. National Institute of Standards and Technology defines cloud computing as a "*model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models* (SP 800-145: 2011)".

The features, characterizing the cloud are (SP 800-145: 2011):

- on-demand self-service – the user of the cloud computing service can adjust the services automatically, without the need of supporting staff;
- broad network access – all the services can be accessed through the network, without any need of physical access;

- resource pooling – the resources of the provider are pooled, using a multi-tenant model, so there is no association of services with physical devices, thus leading to abstract location (e.g. country, datacenter);
- rapid elasticity – capabilities can be rapidly provisioned by automatically adjusting the share of pooled resources;
- measured Service – the services are leveraged and monitored. The costs are based on the usage of resources rather than ownership of assets over time.

Cloud Computing is also classified by the service models that they provide, defining the levels of responsibility by the user and the service provider (Kavis 2014). These service models are (as presented in Fig. 1.2):

- Software as a Service (SaaS). The user can facilitate the provider's applications running on a cloud infrastructure. They are accessible from client devices through a thin client interface.
- Platform as a Service (PaaS). The user can deploy consumer-created or acquired applications supported by the provider.
- Infrastructure as a Service (IaaS). The user can provision processing, storage, networks, and other computing resources, deploy and run arbitrary software including operating systems and applications.

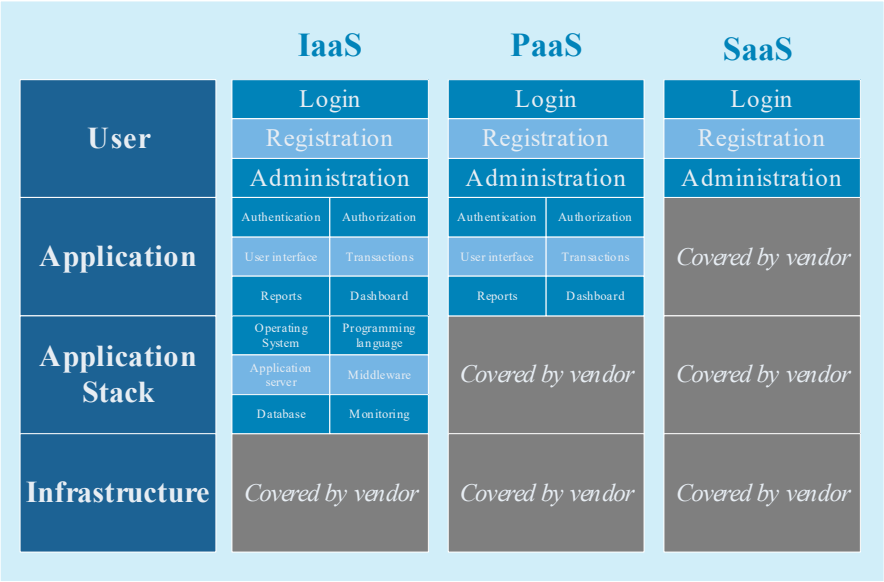


Fig. 1.2. Responsibility areas based on the cloud computing service model (Kavis 2014)

Cloud Computing services can also be classified by the deployment models (SP 800-145: 2011):

- private cloud is operated internally by an organization;
- community cloud is shared by several organizations. It supports a specific community with shared concerns;
- public cloud is available to the general public or a large industry group and is owned by an organization selling cloud services;
- hybrid cloud is a composition of two or more clouds deployment models.

Most of the service aspects inside the architecture of the cloud, including the security issues, are managed by the provider. Therefore the usage of the outsourced infrastructure is based on trust between the client and the supplier. Level of trust and responsibility from technical aspects is managed by setting up a contractual relationship between the parties. It is typically achieved by applying the service level agreement (SLA). An international standard covering this domain (ISO/IEC CD 19086-1: 2015) is being prepared, although guidelines for cloud service level agreement standardization (European Commission 2014) have already been published and fully cover the SLA aspect of cloud computing security. Based on the resource management architecture provided in (Marinescu 2013), the SLA is an essential component that must be taken into account. The Service Level Agreement (SLA) is described by the relevant Service Level Objectives (SLOs) based on (European Commission 2014).

The Service Level Agreement is the cornerstone of contractual agreements (Kyriazis 2013) that provides fundamental grounds for (European Commission 2014):

1. Quality of Service (QoS) – ensuring that the infrastructure would ensure proper quality of service.
2. Quality of Protection (QoP) – that denotes the means to ensure the information security.

The material, presented in this chapter presents the benefits of virtualization technologies and the services of Cloud Computing. However, besides these benefits, certain issues of usage exist. One of the most important of these issues is ensuring that the information is secure.

1.2. Information Security

Information security deals with the protection of information from unauthorized use. It is described by three main attributes – Confidentiality, Integrity and Availability. They serve for the following security features (Kissel 2013):

- Confidentiality – Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information;
- Integrity – Guarding against improper information modification or destruction, and includes ensuring information non-repudiation and authenticity;
- Availability – Ensuring timely and reliable access to and use of information.

As any other phenomenon, information security requires a metric to define the present situation. For this purpose, information security risk analysis is used. The objective of a risk analysis is the description of risk. Therefore understanding of the risk concept and how it works is essential to perform risk analysis (Aven 2008).

There is no unanimous agreement upon the meaning of the term "risk". However, commonly "risk" is focused on negative deviation from an expected target state (Ackermann 2012). Expectation of an organization can have significant deviation from the real situation. Therefore the risk is focused on the uncertainties in relation to the market average values. It serves as the most important information security metric. Depending on the type of the risk analysis method, the risk either be expressed in a qualitative or quantitative way (Table 1.1).

Risk Analysis (RA) is a process for estimation of the frequency and impact of risk scenarios (ISACA 2013). It serves as the basis for optimizing the processes in the planning phase, setting requirements for solutions and measures, drawing conclusions on whether these solutions and measures meet the requirements and documenting an acceptable risk level (Aven 2008).

Mathematical risk expression evaluates Risk Exposure (RE) for the probability (P) of an Unsatisfactory Outcome (UO) multiplied by the loss of this outcome (Ackermann 2012):

$$RE = P(UO) \cdot L(UO). \tag{1.1}$$

Quantitative risk scoring can be based on the probability-based models. The probability-based risk evaluation models provide a description of risk exposure. The descriptive scoring is a model function called the Key Risk Indicator (KRI), representing the degree to which the analyzed object is subject to the particular risk. The particular measure of Value-at-risk (VaR) is the standard risk measure used to evaluate the exposure to risk (Panjer 2006). This measure shows the amount of effort required to ensure proper functioning of an architecture component.

VaR estimation can be performed by three competing approaches. The historical data approach determines loss probabilities in a statistically nonparametric way. However, the main drawback of this approach is that historical data

may not adequately represent present conditions. The stress testing approach is based on calculating losses under various scenarios of unlikely but plausible conditions. The Extreme Value Theory (EVT) approach characterises the lower tail behaviour of the distribution of returns without tying the analysis down to a single parametric family (Smith 2002).

It is worth noticing that the quantification of risk is dealing with future events, which leads to information imperfectness. Moreover, technological developments tend to modify the risk as the adoption of new technologies bring additional threat vectors with them.

Risk analysis methods are organized into three main categories as presented in Table 1.1.

Table 1.1. Categories of risk analysis methods

Category	Type of Analysis	Description
Simplified RA	Qualitative	Informal procedure, no requirements for input or output
Standard RA	Qualitative or Quantitative	Formalized procedure, using recognized RA methods. Risk matrices are the typical output.
Model-based RA	Quantitative	Event-tree and scenario based analysis providing full traceability of conclusions.

The results of risk analysis are relevant if the quality of such analysis is high. The quality of risk analysis can be determined depending on the criteria, defining the soundness of it. A sound risk analysis must satisfy the following requirements (Haimes 2004):

- Comprehensiveness – the scope of risk analysis should include all the aspects of the domain, with a required level of knowledge on each element;
- Adherence to evidence – every portion of risk assessment should be based on hard evidence, including documentation of procedures and policies as well as evidence of certain functions in operation;
- Logical soundness – regardless of the proof provided by the evidence, the process of risk analysis should follow the principles of logic. Items, included in the analysis must reflect to the investigated problem;
- Practicality – the output of the risk analysis should be practical. This means that based on the risk analysis report, actions and inferences should be possible to draw to improve the situation;

- Openness to evaluation – every step of the process should be well documented and clear, keeping in mind the idea, that anyone interested in the process should be able to understand it solely from the description itself;
- Explicit assumptions and premises as a basis – only assumptions and premises that are clear and are well-based should be used for the risk analysis;
- Compatibility with institutions – aligning the risks to the institutionary requirements is essential to have a regulatory compliant output;
- Conduciveness to learning – the process of risk analysis should be the basis for learning and should be adapted to the things learned;
- Attuning to risk communication – if possible, risk analysis should be performed in a way that the output follows a certain protocol, so that a standardized way of communicating the risk situation can be established;
- Innovativeness – as risks are always developing and changing, the innovativeness of the process has to be built in and follow the trend of the newly arising changes.

If the risk analysis satisfies the following criteria, it can be used as a measure for defining the information security situation. Information security within the organization is considered as a directing and supporting concept in the protection of the information assets from intentional or unintentional disclosure, modification, destruction, or denial. It is achieved by implementing appropriate information security and organizational planning policies, procedures, and guidelines.

Information security risk analysis is a technique to identify and assess factors that may have negative influence on the success of achieving a goal from the perspective of information technology. Risk analysis also involves definition of preventive measures to reduce the probability of these factors from occurring. Also, it includes identification of the required countermeasures to successfully deal with such constraints when they develop. The identified risks can be used to support the development of system requirements, including security needs (Peltier 2010).

The process of information security risk analysis consists of the components presented in Fig. 1.3.

Fig. 1.3. Main components of the information security risk analysis process (Peltier 2010)

Information security risk analysis is a part of the information security life cycle, where it provides the initial information required for the successful information security management. The information security life cycle is presented in Fig. 1.4.

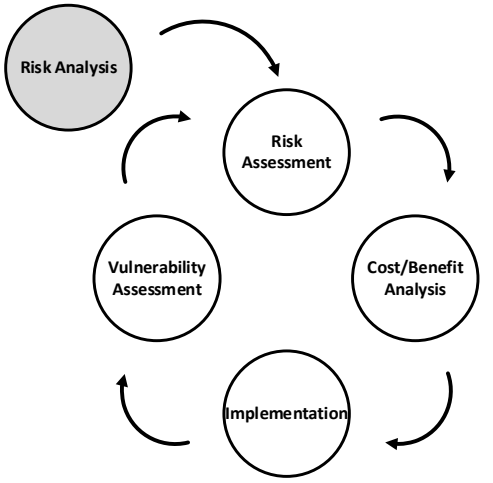


Fig. 1.4. Information security life cycle (Peltier 2010)

Risk analysis consists of the following steps (ISO 27005: 2011):

- Risk Identification:
  - Identification of Assets – producing a finite list of IT related assets that are taken into account for the analysis;
  - Identification of Threats – specifying the threats based on the assets;
  - Identification of Existing Controls – specifying countermeasures for the existing threats;
  - Identification of Vulnerabilities – specifying the uncovered areas of the security aspects;
  - Identification of Consequences – specifying probable scenarios.
- Risk Estimation:
  - Assessment of Consequences;
  - Assessment of Incident Likelihood;
  - Level of Risk Estimation.

Several methodologies for information security risk analysis exist. They differ by the field of application, nature of the output and the purpose of the analysis. Moreover, they differ by the terminology used for the process. This study focuses on the process applicability for risk analysis rather than the specific term used to describe it, therefore “method” and “methodology” are

considered equal, disregarding the differences of the semantic meaning of the two terms.

The research of the available information security risk analysis methods is based on the European Union Agency for Network and Information Security Inventory of Risk Management and Risk Assessment Methods has resulted in sixteen methods including the Austrian IT Security Handbook (Bundeskanzleramt 2013), CCTA Risk Analysis and Management Method (CRAMM) (Yazar 2002), Coras (Braber *et al.* 2006), Dutch A&K Analysis (Dutch Ministry of Internal Affairs 1996), EBIOS (French National Information Systems Security Agency 2011), FAIR (Jones 2011), FRAP (Peltier 2000), IRAM2 (Information Security Forum 2014), OCTAVE (Alberts, Dorofee 2002), MAGERIT (Lopez *et al*. 2005), Marion (CLUSIF 1998), MEHARI (CLUSIF 2011), MIGRA, RiskSafe Assessment, IT-Grundschutz (BSI 2008) and ISAMM (Fenz *et al.* 2014) methods. The detailed description of the methodologies is presented in Annex A.

The comparison of the existing risk analysis methodologies provides a systematic approach of representing the sixteen of the most widely used methodologies and their features as well as compliance to National and International standards. Based on the analysis of the existing methodologies, presented in Annex A it is observed that the majority (13/16) of the methodologies are designed for the qualitative analysis of information security risk. Two of them (Annex A: M04 and M11) are obsolete and no longer used. 13 methodologies declare compliance with existing international or national standards, most popular of which is the ISO/IEC 27000 family of standards (9/16). Four of them are commercial (Annex A: M08, M13–M14, M16). The analysis has also revealed that the methodologies typically consist of several steps ("phases", "steps" and "stages" are considered as synonyms of different method descriptions in this case), defining the objectives of the risk analysis, threats and controls, and performing the risk analysis.

Methodologies, analysed in this chapter are used as reference material for further research and the development of the proposed method.

Information security risk modeling involves the detailed description of the prerequisites, important for risk analysis. The categories of aspects to be described in such model are presented in Fig. 1.3.

Due to the fact that risk is dealing with future events, a concept of uncertainty is introduced to risk modeling, thus defining the risk modeling to be of probabilistic (stochastic) nature. Probabilistic modeling deals with uncertainty of variables and parameters describing the structure of the elements and constrains.

Uncertainty is defined as a phenomenon where "*potential outcomes cannot be described in terms of objectively known probability distributions, nor can they be estimated by subjective probabilities* (Haimes 2004)".

In the process of risk modeling likelihood of possible undesired deviations from the plan through objective or subjective probabilities ar assessed. Afterwards, causal relationships between the sources of risk and their impacts are modeled. Quantification of the random and decision variables and the relations of these variables to the state variables, objective functions, and constraints is considered to be the crucial step in the risk analysis process (Haimes 2004).

When dealing with complex situations, advanced tools and methods for risk analysis are required. Such tools and methods are presented in the next chapter.

1.3. Information Security Modeling and Analysis

Management of complex information technology (IT) infrastructure requires specific solutions for automation because such infrastructure typically faces constant dynamic changes. Since the introduction of Enterprise Architecture (EA) concept, there have been numerous approaches in systemizing and automating the management of the IT infrastructure. One of the sensitive domains of the EA – the IT infrastructure security assessment – requires in-depth analysis and expertise as well as a design, responsive to changes to ensure proper operation of the IT applications. Comprehensive and detailed EA models serve as reference models for such analysis. Moreover, modern EA modelling language based tools have built-in analysis and assessment capabilities.

Providing sufficient resources for cyber security risk analysis is challenging resource-wise, therefore usually found almost exclusively only in security oriented and large enterprises. Meanwhile small and medium enterprises (SMEs) typically tend to have limited resources, thus requiring combined solutions, capable to serve both for security as well as business process modelling. Having a cybersecurity oriented EA model, designed in a security-assessment-specific EA modelling language, enables automatic cyber security assessment of the architecture. However, enterprise management-related data, including the security-oriented architecture model information is organized in an enterprise-friendly form. Extraction of such information and transformation of it to a form suitable for EA modelling language with analytic features enables automation of cyber security assessment of an EA.

Architectural modeling languages is a very important component in management and development of information systems. Modeling languages such as SySML (Friedenthal *et al.* 2014), Business Process Modeling Notation (BPMN) (Chinosi, Trombetta 2012) enable creation of information system architecture and system environment through diagrams that can be used for various forms of analysis, one of which is security. Moreover, they offer extensions for Industrial Control System Security Analysis (Lemaire, Lapon 2014), and on top layer,

Cloud Security (Machida *et al.* 2011). Moreover, there are various tools available for this purpose including CORAS and its' extension for ISO standard compliance (Beckers *et al.* 2014), MulVAL (Ou *et al.* Appel 2005) or NetSPA (Artz 2002).

However, the modeling languages do not offer the required reasoning, while the tools either suffer from being subjective or vague, or even obsolete. Thus, there are some solutions that offer modeling capabilities along with the reasoning based on the systemized expert knowledge base. One of them is OpenMADS (Andrade *et al.* 2013), the other is Cyber Security Modeling Language (CySeMoL) (Sommestad *et al.* 2013). A comparison of the existing tools and methods is provided in Table 1.2.

Table 1.2. Comparison of the existing Architectural modeling language for Cyber Security risk analysis

No.	Name	Tool Provided?	Automatic Analysis?	Technical Issue Assessment?	Socio-technical Issue Assessment?	Multi-domain?	Multi-perspective Attack Analysis?
1.	CySeMoL	YES	YES	YES	YES	YES	YES
2.	CORAS	YES	YES	YES	NO	YES	NO
3.	Common Criteria	NO	NO	YES	YES	YES	NO
4.	OpenMADS	YES	YES	YES	NO	YES	NO
5.	MulVAL	NO	NO	YES	NO	NO	NO
6.	NetSPA	YES	YES	YES	NO	NO	NO

Using Enterprise Architecture Security Assessment Tools, such as Cyber Security Modeling Language (CySeMoL) for implementation of information security related models provides the end user with improved accessibility and the ability to assess critical areas of the design prior to its' deployment, preventing possible damages. Therefore, implementation of cloud security assessment model comes as a natural solution to this problem. Moreover, since CySeMoL already covers a broad range of domains of enterprise architecture, cloud security assessment enables users of the system to cover their complete IT infrastructure that, in case of an enterprise, is a combination of diverse technology (Sommestad *et al.* 2013). On top of it, the decision makers of the enterprise require solutions for cyber security estimation that are easy to understand (Holm *et al.* 2013). And the CySeMoL provided easy-to-understand graphical output of

the situation suits this requirement well. Therefore, this study uses the CySeMoL as a basis.

CySeMoL is defined as a modeling framework and calculation engine for estimation of the cyber security of enterprise-level system architectures. Within the CySeMoL the knowledge of the possible attacks is included, therefore the framework operates as an expert system. Such solution only requires the system architecture as an input. Modeling and calculation of the vulnerabilities is achieved using the Predictive, Probabilistic Architecture Modeling Framework (P2AMF) which extends the Object Constraint Language (OCL) to be used for probabilistic assessment and prediction of the properties. This framework enables the expression of uncertainties, relations and attributes in Unified Modeling Language (UML) and perform assessments using these uncertainties. The P2AMF creates a prediction model, representing the predicted situation.

1.4. Information Security Risk Analysis in Virtualizated Systems

Virtualization came to prominence rather recently but the exponential speed of growth of usage of it has drawn a major interest from many perspectives. They range from resource optimization to the security concerns and their solutions. As the latter has the main focus – the overview of research is targeted to the existing scientific research on virtualization information security.

Virtualization security was first publicly noticed as an issue over ten years ago by W. Wong (2005). The author notices the raising usage of virtualization for more powerful processes and notices the need to evaluate and ensure the security of virtualized systems. Since then, rapid growth of interest has been observed for this topic and various approaches have been analyzed for this matter. One of the early approaches by S. Vaarala contribute to the field of virtualization security by providing a threat model, including available security mechanisms with recommendations for security improvements through virtualization (Vaarala 2006). Garfinkel and Warfield analyzed the virtualization impact as a security measure and pointed out the benefits of data isolation and snapshot abilities as the arguments against the traditional computing (Garfinkel, Warfield 2007). Finally in 2008 Vaughan-Nichols published an introductory article about the virtualization security concerns (Vaughan-Nichols 2008).

But not only scientific approaches are observed. From the technical point of view, the first virtualization risk and risk management considerations were raised rather recently as well, when McAfee's Foundstone released a whitepaper on virtualization and risk (Hau, Araujo 2007), introducing the need of risk assessment for virtualized systems. Later developments build on this topic, analyz-

ing secure virtualization configuration (Hietala 2009) and server virtualization security and risk evaluation by L. Kai (2012).

Most of the virtualization security research, however, focuses on cloud computing. Technical reports on cloud computing security assessment (Context Information Security 2011) deal with this topic based on technical documentation, issued by trusted sources. One of the sources, the ENISA report on Cloud Computing (Catteddu, Hogben 2009) provides a generous amount of possible vulnerabilities, covering the most of the domain. Cloud computing specific security issues in various researches are classified according to specific service models by which different types of services are delivered to the end user (Subashini, Kavitha 2011; Modi *et al.* 2013). Specific cases of security and privacy have been discussed in number of researches and surveys with the most up to date and comprehensive being and pointing out the latest tendencies of scientific researches (Alia *et al.* 2015). Gruschka and Jensen (2010) are of the first to introduce a cloud computing scenario security model using attack vectors based on three classes of participants: users, services and providers. In this model, every cloud computing scenario interaction can be addressed to two entities. Therefore, every attack vector here is detailed as a set of three-class bi-directional model interactions. These studies serve as the basis for the baseline cloud computing security.

However, no study to provide quantitative characteristics for describing risk in the field of cloud computing have been observed in this research. This is partially solved in the field of doctoral thesis – where one of the main reference works – a very thorough dissertation, published as a study on IT Security risk management dealing with perceived IT Security risks in the context of Cloud Computing (Ackermann 2012) – provides a quantitative empirical survey to examine how potential users perceive IT security risks and how these risk estimations affect the adoption of Cloud Computing. The research takes an even deeper approach by investigating the scenario parameter effect to the distribution of potential losses. Regardless of the sophistication of this work, however, it can be referred to, but does not cover the method for the numerical procedures to obtain the risk evaluation scores.

Thesis by M. Carroll (2012) focuses on the governance of cloud computing and virtualization, providing a set of legislative requirements to deal with virtualization and cloud computing. This framework contributes to the contractual requirement portion of the research.

Other worth noticing doctoral dissertations include T. Garfinkel (2010). In this dissertation several paradigms for enhancing host security leveraging are introduced. They include virtualization based approach to trusted computing allowing multiple virtual hosts with different assurance levels to run concurrently on the same platform using a novel “open box” and “closed box” model, vir-

tual machine introspection an approach to enhancing the attack resistance, intrusion detection and prevention systems and overshadowing data protection approach for providing a last line of defense for application data even if the guest OS running an application has been compromised.

From the technical perspective, Z. Wang (2012) has conducted a research for hypervisor integrity assurance as the basis for the doctoral thesis. A new system – HyperSafe, enabling the self-protection for Type I hypervisors has been proposed. It combines two techniques, non-bypassable memory lockdown and restricted pointer indexing. Another system for Type II hypervisor integrity – the HyperLock for secure isolation of a vulnerable or compromised type-II hypervisor has been proposed, including hypervisor isolation runtime and hypervisor shadowing. Finally an anti-rootkit solution called HookSafe has been proposed in this thesis. This, combined with another technical solution by J. Szefer (2013) proposes a method to leverage hardware to help provide protection for data execution inside virtual machines on the remote cloud servers. Szefer also introduces a new threat, associated with virtualization – extraction of sensitive or confidential code or data from virtual machines using an attacker with virtualization layer privileges. These works have contributed to the forming of the overall picture of the cloud computing security situation, and therefore – the threat landscape of the field.

1.5. Virtualization and Cloud Computing Security Research in Lithuania

The field of information security is within the focus areas of Lithuanian researchers. Therefore, plenty of approaches towards the development or improvement of portions of information security exist. Special interests for researchers include cryptography, with applications in protocol level (Sakalauskas *et al.* 2007) as well as as its applications (Sakalauskas *et al.* 2010), infrastructure analysis (Kajackas, Rainys 2011) and even legislative issues (Stitilis *et al.* 2011).

Research of virtualization and cloud computing has not been in the main focus in Lithuania. However, there are some notable works in the field. A rootkit detection experiment within a virtual environment has been carried out to present a framework for investigation of kernel-level rootkit behaviour within the virtual environments (Toldinas *et al.* 2015). From the forensic point of view, virtualization security issues are dealt with by Goranin and Mažeika, providing a methodology for forensics in virtualization (Goranin, Mažeika 2011).

1.6. Conclusions of the First Chapter and Formulation of the Dissertation Tasks

The first chapter of this thesis provides an overview and definition of the domain specific concepts. It includes aspects of their implementation and operation as well as comparison between the various modifications. The following conclusions have been drawn:

1. The overview of the risk analysis methods has revealed the lack of ability of such methods to evaluate virtualization related risks.
2. Sixteen most widely used risk analysis methodologies and their features have been analysed. The analysis of the existing methodologies, presented in Annex A shows that the majority (13/16) of the methodologies are designed for the qualitative analysis of information security risk. Two of them (Annex A: M04 and M11) are obsolete and no longer used. 13 methodologies declare compliance with existing international or national standards, most popular of which is the ISO/IEC 27000 family of standards (9/16). Four of them are commercial (Annex A: M08, M13–M14, M16).
3. The reviewed scientific research literature provides a strong foundation for virtualization and cloud computing security assurance, spanning from governance and management issues to the very technical details for the assurance of security. However, lack of research in the field of virtualization security risk analysis has directed this research towards the development of a new structured way to analyze the information security risks that affect virtualization technologies.

Based on the conclusions, the following tasks are formulated to achieve the goal:

1. Review of the existing methods for the evaluation of information security risk in virtualized systems.
2. Develop a virtualization threat landscape, defining the scope of the virtualization threats.
3. Propose a method for quantitative risk analysis by adapting existing IS threat database data to eliminate and/or support the results of expert evaluation.
4. Perform comparative statistical analysis to evaluate the accuracy of calculated threat scorings in relation to expertise knowledge and define its uncertainty.
5. Evaluate the method improvements in a working Enterprise Architecture analysis system to provide scenario-analysis.

2

Risk Analysis Modeling and Application for Virtualization Technologies

This chapter provides the knowledge and links between the methods to allow building a thorough model of virtualization information security risks. It includes the existing infrastructure, contractual agreements, up-to-date threat sources and means to control them and expert evaluation into account. The outcome of this chapter is a newly proposed method for virtualization information security risk analysis based on the newest scientific research in the field.

2.1. Method for Quantitative Information Security Risk Analysis for Virtualization Technologies

The components of this method are based on a combination of well-established and validated methods (Beckers 2015), (ISO 27005:2011). Together, these methods provide a framework to organize the present data, and based on it – provide a scientific approach for a quantified evaluation of risks. The processes of this method are required to:

1. Define the coverage of the threats for the researched area.

2. Link the threats with the control means that cover them.
3. Define and specify the attack scenarios associated with the research area.
4. Define the most important variables influencing the risk magnitude of each analyzed attack scenario.
5. Acquire numerical values for each variable through an experimental research.
6. Validate the numerical risk values by a separate independent experimental research.

As the nature of the application of this method requires, these main steps are expanded by the following: the security controls are organized into larger security requirement groups for easier analysis and input minimization, attack scenarios are introduced to provide the traceability of the analysis and result validation is introduced to prove the correctness of the method performance.

The process, used to achieve the objectives of the method, is graphically presented in Fig. 2.1.

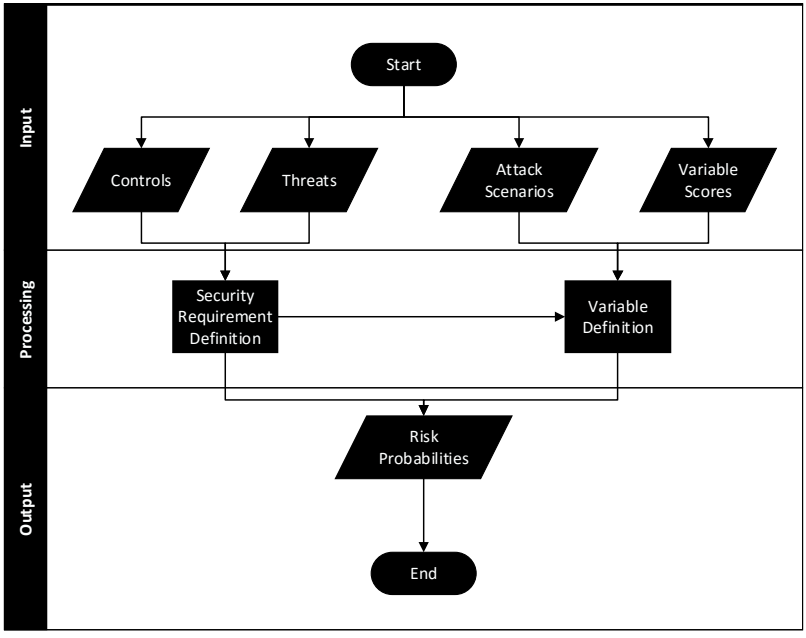


Fig. 2.1. The flowchart of processes of the proposed risk analysis method (Source: author)

As seen from Fig. 2.1 the method has four input channels. It generates one output – risk, expressed as a probability. The rest of this chapter is dedicated to describe the methods to achieve the required results, including the collection of input data, the processing and the output.

The proposed virtualization risk analysis and management method is designed to be quantitative and easily applicable to probabilistic evaluation. To achieve this goal the following is performed:

- the amount of input information is sorted by grouping the controls into requirements;
- attack scenarios are defined for automated machine reasoning;
- variables are introduced for detailed output of risk probabilities;
- quantitative risk analysis is performed to generate a probabilistic output for the virtualization component information security risk.

The method consists of threat identification and selection of control means to cover the threats. The next step is to match countermeasures covering each control mean, link them and combine to provide the minimum amount of countermeasures for consistent risk analysis. Therefore the method consists of the following:

- definition of threats – taxonomical classification of the threats;
- definition of controls – threat association with the hypervisor specific control means;
- definition of security requirements – proposal of the requirements representing the controls;
- definition of attack scenarios – attack scenario association with the security requirements;
- definition of attack variables – assignment of the main attack variables;
- definition of attack variable scores – association of the exploitability scores with the variables;
- evaluation – risk evaluation based on the collected data;
- validation – risk evaluation validation based on the expert data.

The proposed method for virtualization information security risk analysis is of quantitative nature. The output of the analysis is the probability of risk that represents the likelihood of one or more professional penetration testers to successfully complete the attack step in the object model within the time designated for the attack. These probabilities are based on the estimates for the attack steps associated with it (Holm *et al.* 2013). Quantitative methods benefit over the qualitative ones by providing a more accurate image of risk and enabling quantitative cost-benefit analysis but at the same time requiring greater expertise of the field (Lee 2014).

Threat identification is performed using an own-developed reference virtualization security threat taxonomy (presented in subchapter 1.2.2) that provides a thorough landscape of the threats and threat sources that the virtualization components may face. Controls for these threats (presented in subchapter 1.2.3) are described and linked to provide a mitigation plan for the threats. The realization of regulatory controls for the threats in a practical approach consists of specific

solutions, serving as one or more controls that mitigate the risk. Data required for variable definition – the attack scenarios, that describe the nature of the variables and the variable scores that assign a numerical value to the variables are presented in subchapter 1.2.4.

To be able to solve the problem, it is necessary to understand the scope and the broadness of the problem in the first place. The first step is to define the limits, within which, the problem exists. Knowledge of the analyzed field requires a formal representation, which is typically organized to serve the purpose of the research in the most convenient way. Organizing the knowledge enables the definition of the research domain limits. Choosing taxonomy over an ontology in this approach is based on the requirement of the knowledge to be of an easily definable and classifiable domain for taxonomy (and as opposed – universal bodies of knowledge for ontologies) (Smiraglia 2014).

It includes construction of a taxonomy consisting of the component entities of the domain. A low-level "ad-hoc" taxonomy is required to define the coverage of the area as the definition requires a clear specification of knowledge. Hierarchical classification is a classification arranged according to the general-specific relations. Providing a hierarchical order in the classification process organizes the knowledge in a convenient, easy readable form.

A classification is considered appropriate if it has the following properties (Smiraglia 2014):

- Systematic – have rules for inclusion of component entities to a class, division or sub-division;
- Expansible – the design has the built-in flexibility to be expanded;
- Well described entities – all single and composite subject concepts should have designations assigned and have a location in the classification.

Building a taxonomy includes determination of the domain and scope, review of subject domain authorities, extraction of concepts, organization of concepts and validation. In the case of this thesis, a new taxonomy is needed, as the existing approaches do not cover the whole threat landscape (subchapter 1.2.2).

Risk evaluation process assigns a value to a certain risk factor. It is a multi-perspective measure, incorporating the evaluation of threat occurrence likelihood and impact analysis. Based on the nature of the risk analysis method (Table 1.1), the output of the risk evaluation can be qualitative or quantitative. As the proposed method deals with quantitative risk analysis, it requires incorporating a method to provide a quantitative risk evaluation.

In the case of this method, risk is evaluated using two sources of information. At first, to define the objectives of the risk analysis the variables representing the objectives are specified. To achieve the quantitative information about these variables statistically processed scores of widely recognized vulnerability databases are used. The validation of this quantification is then achieved by collecting expert knowledge and comparing it to the variable scores.

To gain quantitative measures of security risk model, based on a qualitative nature is a complicated task as risk measurement is typically of the qualitative nature (Creswell 2003). A study on security risk model parameter quantification (Ryan *et al.* 2012) suggests using a well-established method of expert judgement elicitation for the needs of information security. Moreover, this study suggests and validates the point that sound results can be achieved using a set of experts as small as twenty-one individuals.

An expert is considered to be a skilful, well trained person with extensive knowledge in a specific field. Such expert provides an opinion in the process of expert elicitation (Ayyub 2001). The methods of expert elicitation can be of direct and indirect nature.

The indirect methods are based on betting rates by experts to reach a point of indifference among presented options related to an issue. Such methods, however, have a major disadvantage – the utility value of money is not necessarily linear with the options presented to an expert, and the utility value of money is independent of the answer to an issue, such as failure rate. However, indirect techniques are useful to elicit probabilities from probability-illiterate experts as real-life values are used to estimate the probabilities.

Direct methods elicit a direct estimate of the degree of belief of an expert on some issue. Direct methods include the Delphi method designed for technological forecasting and policy analysis and risk studies as well as the nominal group technique which allows iterative evaluation where a structured discussion is conducted after the experts have provided initial opinions and the final judgment is made individually on a second cycle of opinion elicitation and aggregated mathematically similar to the Delphi method (Ayyub 2001).

The main issue when dealing with expert data is the uncertainty of the metrics caused by bias of individual experts that arises due to their background. The Cooke classical model (Cooke 1991) offers a solution for synthesizing the expert knowledge of various individuals by assigning weights to their judgement based on their ability to estimate the true value of the seed questions, which have known answers prior to the study. Seed questions help identifying the most suitable individuals for the study by defining their rele-

vance to the background knowledge and the ability to quantify their knowledge (Ryan *et al.* 2012).

Virtualization security extends the attack surface of the host-based computing with additional attack surfaces caused by virtualization and networking issues. Therefore, it is important to understand the complete attack surface to avoid the inconsistency of the security strategy due to undocumented weak links. Multiple approaches of describing the virtualization attack surface exist, the most comprehensive of which are described below.

Pearce et al. have proposed a study on the security threats and solutions for virtualization. The authors suggest a virtualization threat classification into three categories, the causes being: strong virtualization properties, core virtualization implementation and control software and data flows (Pearce *et al*. 2013).

A special publication by the National Institute of Standards and Technology (NIST) offers security recommendation for hypervisor deployment. In this publication, the recommendations are based on a set of potential threats, organized into five categories as well as supplemental threat source description (Chandramouli 2014). The areas of execution isolation, devices emulation and access control, execution of privileged operations for guest systems by the hypervisor, virtual machine management as well as host and hypervisor software are covered in this publication.

From the commercial application point of view, SANS Institute has reviewed and organized the virtualization threats and hardening process of commercial virtualization solutions (Shackleford 2010). This study provides the controls based on the facing threats as well as deployment and configuration recommendations for commercial virtualization products.

Other studies covering the description of virtualization threats to a certain extent include Gupta and Kumar research on the virtualization threats, provided as a taxonomy of cloud security, organizing the threats into two main categories – the hypervisor and the virtual machine threats (Gupta, Kumar 2013). Another study worth mentioning is carried out by Hashemi and Ardakani (Hashemi, Ardakani 2012). The taxonomy of the security aspects of cloud computing systems partially covers the virtualization issues as well.

The study by Pearce *et al.* (2013) focuses on the technical threats, mostly arising from the hypervisor and virtual machine interaction, thus lacking threats arising from the virtualization management. The NIST special publication deals with the hypervisor deployment issues, while the SANS Institute whitepaper covers the aspects of real-life virtualization implementation scenarios.

2.2. The Proposed Virtualization Security Threat Taxonomy

Virtualization security assessment requires a base point, consisting of known security threats that the technology is facing. Analysis of the existing research in this field revealed that although studies proposing virtualization security categorization exist, they represent a specific field of application, thus limiting the threat analysis to a certain point of view – technical (Pearce *et al.* 2013), hypervisor deployment and governance (Chandramouli 2014) and commercial product deployment (Shackleford 2010). However, the research of this paper aims at providing a multi-perspective risk analysis for virtualization solutions. The comparison is conducted using the threat lists, provided in the three documents. The overlapping threat entries are merged. This merged threat list has resulted in thirty entries. It is assumed that such categorization approach provides a full virtualization threat landscape to ensure the completeness of the analysis.

As seen in Table 2.1. the results of the analysis of existing virtualization information security threat taxonomies has revealed that there is no approach covering the whole threat landscape.

As a result of this comparison, a taxonomy, combining all the threats found in the threat landscapes is proposed. As the three analyzed sources differ by the application field (technical) (Pearce *et al*. 2013), hypervisor deployment and governance (Chandramouli 2014) and commercial product deployment (Shackleford 2010) as stated above), the proposed taxonomy represents a complete virtualization threat landscape. Other categorization approaches are disregarded, as the threats presented in them fully overlap with the threats of the three main taxonomies. The proposed taxonomy is presented in Fig. 2.2.

The virtualization threats in the proposed taxonomy are organized into two categories and six subcategories. It ensures that this representation is clear and the information, required for analysis of the threat origin, is sufficient. The threats are organized according to (Steiner 2012). This document suggests that the threats of virtualization are of logical, physical and premise security nature. The presence of virtualization extends the threat landscape compared to traditional computing by adding additional threats, related to the hypervisor (a layer between the hardware resource pool and virtual machines), unknown geographical location of the stored data, data breaches between the virtual machines within the same hypervisor and management of virtual network security and monitoring. To emphasize the importance of virtualization in this taxonomy virtualization specific threats are placed under Logical Security category, while the physical and premise security threats are placed under category Other.

Table 2.1. Comparison of the existing virtualization security threat landscapes

No.	Threat	(Chandramouli 2014)	(Shackleford 2010)	(Pearce *et al.* 2013)
1.	Compromise of VN Traffic Confidentiality		+	
2.	MAC Address Spoofing		+	
3.	Abuse of Ports and Services		+	
4.	Management Network Abuse		+	
5.	Denial of Service Attacks		+	
6.	Hyperjacking (Rootkit)	+		+
7.	Hypervisor Compromise			+
8.	Data Leakage Through Shared Memory		+	
9.	Exploitation of Unnecessary Virtual Devices	+	+	
10.	Data Leakage Through External Devices		+	
11.	Hypervisor Trust Model Breaches			+
12.	Hypervisor Intervention			+
13.	Dataflow Breaches			+
14.	Virtualization Attacks			+
15.	Guest System Attacks and Compromise	+		
16.	Isolation Breaches	+		
17.	Unauthorized Access		+	
18.	Brute Force Access		+	
19.	Privilege Abuse and Escalation Attacks		+	+
20.	Over-Privilege Abuse	+	+	
21.	Man-in-the-middle Attacks		+	
22.	Exposure to Untrusted Networks		+	
23.	Attacks on Unhardened Host Systems	+	+	
24.	VM Cloning			+
25.	VM Nonlinearity Breaches			+
26.	Control Channel Breaches			+
27.	VM Migration	+		
28.	No Access Monitoring Requirements		+	
29.	No Forensics and Incident Response		+	
30.	No Regulatory Compliance		+	

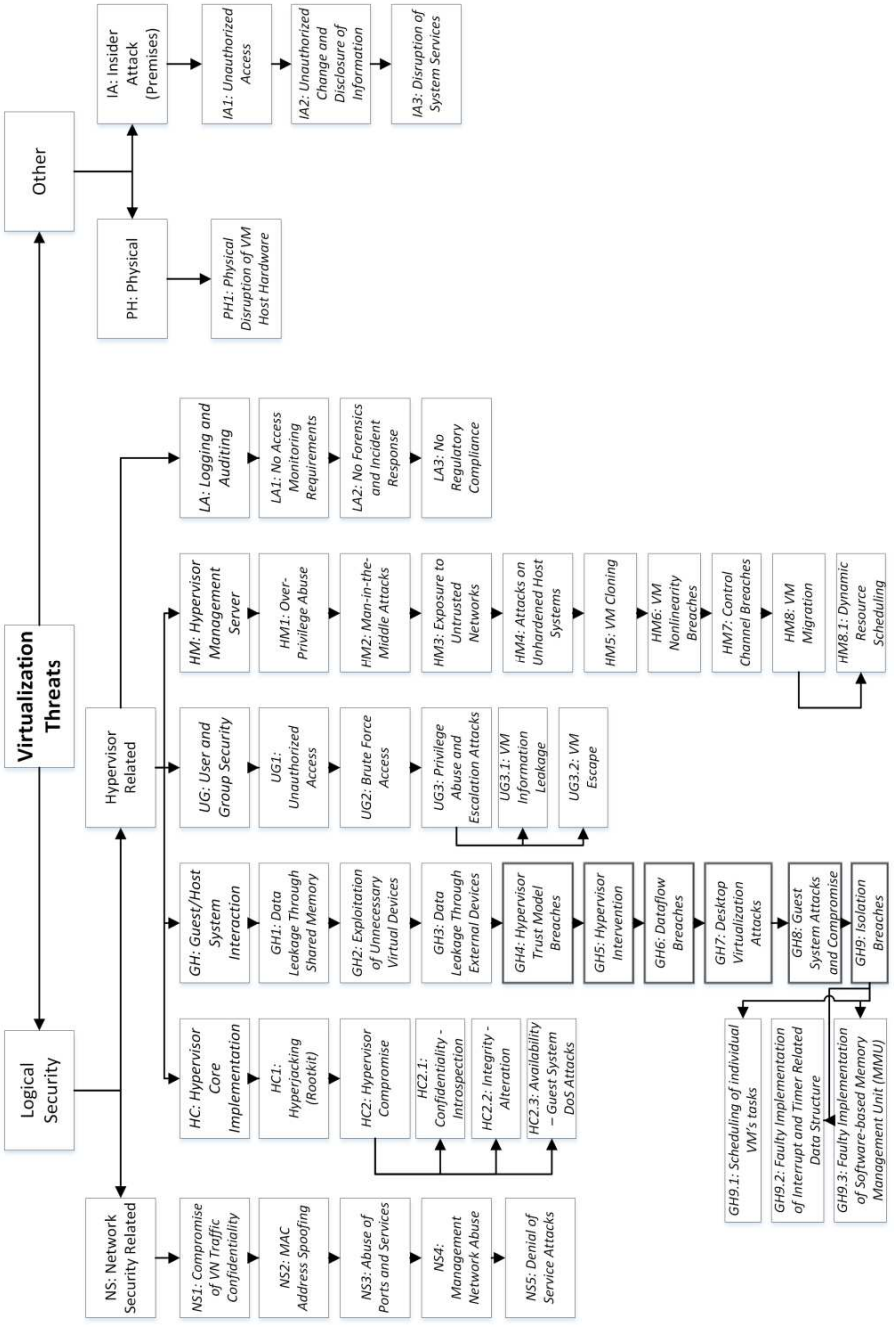


Fig. 2.2. The proposed virtualization security threat taxonomy

Virtualization Specific threats consist of Network Security Related and Hypervisor Related. While Network Security Related threats seem to be more generic then Virtualization Specific, it is worth mentioning that the selected threats of this category deal with network virtualization issues.

The Hypervisor Related threats are in the main focus of this study. They are organized into five sections:

- *HC: Hypervisor Core Implementation* – Consists of security implications due to unmet requirements and leads to transparency and resource control breaches.
- *GH: Guest/Host System Interaction* – Security implications caused by interactions of VMs with the host fall in this category.
- *UG: User and Group Security* – Security implications due to improper implementation of authorization and authentication controls.
- *HM: Hypervisor Management Server* – Misuse of management controls and protocols leads to gaining unauthorized privileges than may be exploited.
- *LA: Logging and Auditing* – Security implications due to unmet compliance mandates and disability to perform forensics in case of an incident fall in this category.

The Hypervisor Core Implementation (HC) specifies the threats of Hyperjacking and Hypervisor Compromise that are dependent of the way the hypervisor is implemented and configured.

Hyperjacking is considered to be one of the upcoming virtualization threats. It works by placing malicious control over the hypervisor, thus infecting virtual machines and having certain advantages over the users of the operating systems within the VM. While the automated threat detection solutions do not search for malicious activities at the VM level, stealthy activities become possible. Hyperjacking can be realized in four different types depending on the way it acts (Rutkowska 2006).

Hypervisor Compromise is expanded into three sub-concepts of Introspection, Alteration and Guest System DoS Attacks. Introspection is caused when the hypervisor has full low-level visibility of operation and can intervene it. Hypervisor alteration is caused by a compromised hypervisor and can affect all the running virtual machines. Moreover, the hardware supporting the hypervisor can be affected as well. Guest system DoS attacks exploit the hypervisor bugs as well as resource starvation to compromise the availability of a guest system (Pearce *et al.* 2013).

Guest/Host System Interaction (GH) specifies the threats caused by the virtualization environment and the virtual instances. Most typically they include various types of data leakage and other isolation breaches.

Threats falling to this category include Data Leakage through Shared Memory and External Devices. Since virtual machines share the same physical

hardware, there is a possibility of data leakage through shared hardware components. Compromising the shared memory and altering the identifiers of ownership of information portions leads to arbitrary acquisition of information. Moreover, inconsistent hypervisor trust model and disregarding of the least privilege principle, allowing unnecessary ports and services may lead to exploitation of security issues, unknown to the owner (Schneider 2004), such as dataflow breaches.

The guest system, running over the hypervisor also faces threats, caused by compromised hypervisor (through hypervisor intervention), as it controls the processes and hardware management for the guest system. Thus, threats like desktop Virtualization Attacks, Guest System Attack and Compromise and other Isolation Breaches exist.

User and Group Security (UG) specifies the threats caused by misuse of access control management. They include unauthorized access, brute force access, as well as privilege abuse and escalation related attacks.

Hypervisor Management Server (HM) specifies the threats arising of misusage of management layer of the hypervisor that tends to have higher privileges. Apart of the over-privilege abuse, threats of control channel breaches, exposure to untrusted networks, VM nonlinearity, man-in-the-middle attacks and VM migration fall in this category.

Logging and Auditing (LA) is the regulatory and forensic threat group, where compliance to certain security management documentation and nonexistent forensic capabilities are the main threat sources.

The proposed taxonomy is evaluated based on the National Information Standards Organization standard ANSI/NISO Z39.19 (ANSI 2010) and the main requirements of:

- *Structure*: hierarchical classification of information – this is represented by the number of (sub)categories. The proposed taxonomy that has 8 exceeds the nearest approach by two (sub)categories additionally providing three levels of subcategorization for easier implementation;
- *Comprehensiveness*: the taxonomy should cover the objective domain – the number of nodes in the proposed taxonomy – 34 exceeds the nearest approach by fifteen entries;
- *Soundness*: the taxonomy should meet the standard requirements – the number of narrower terms for a broader term falls into the recommended range of [3; 20], the identifiers for each node for easier navigation are assigned;

Based on these criteria the newly proposed taxonomy is well built, meets the requirements of structure, comprehensiveness and soundess and exceeds the previous approaches by informativeness and organization (ANSI 2010).

Table 2.2. Comparison of categorization approaches to the proposed taxonomy

Characteristic	Categorization by					
	Pearce *et al.* (2013)	NIST SP800-125a (Chandramouli 2014)	SANS Institute (Shackleford 2010)	Gupta and Kumar (2013)	Hashemi and Ardakani (2012)	Proposed Taxonomy
Number of threats	12	13	19	8	11	34
Number of (sub)categories	3	6	6	2	1	8
Depth of subcategorization	1	1	1	1	1	3
Average concepts per lowest level (sub)category	4.0	2.2	3.2	4.0	11.0	4.3

2.3. Security Controls and Attack Scenarios for the Hypervisor Specific Threats

To ensure proper level of hypervisor security, threats, specified in Fig. 2.2 under sub-category "Hypervisor Related" must be covered by at least one security control. The security controls in this case are considered as means and methods to prevent a threat from occurrence, thus meeting a set of predefined security requirements (800-53A: 2014).

Cloud Security Alliance's Cloud Control Matrix version 3.0.1 (Cloud Security Alliance 2014) is used as a source security controls. The controls, relevant to the proposed threat taxonomy are selected and linked to the threats of the proposed taxonomy. The relationships between the threats and security controls are provided in Table 2.3. For the convenience of the layout, only the IDs of both sources are provided, and the color marking matches the one of the CCM to be used as a reference.

Based on the selection of hypervisor specific security controls, tools and methods for realization of each control is selected. It was noticed that multiple threats that virtualization faces can be covered by one control mean. Since simple input without the loss of initial information is an important feature, groups of control means, serving as security requirements have been proposed. The *requirements for secure hypervisor deployment and management* are as follows:

Table 2.3. Relationships between CCM security controls (Cloud Security Alliance 2014) and proposed taxonomy threats

CCM V3.0.1 Control ID	Proposed Taxonomy Threat ID
AIS-02	UG1, UG2, UG3
AIS-03	GH1, GH2, GH3, GH6, HM2, HC2
AIS-04	HC2
AAC-03	LA3, GH6, HM7
DCS-01	HM5, HM6, GH4, HC1, HC2
DCS-03	GH3
DCS-04	HC2
EKM-01	UG1, UG3, HM1
EKM-03	HC2, UG1
EKM-04	UG1, UG2, UG3
GRM-01	LA3
GRM-06	LA3
IAM-01	LA1
IAM-03	LA2, HM1

CCM V3.0.1 Control ID	Proposed Taxonomy Threat ID
IAM-04	LA3
IAM-07	UG1
IAM-08	GH4, HM3
IAM-09	UG1
IVS-01	LA2, HM7
IVS-02	LA2, HM5
IVS-03	GH9
IVS-07	GH2, GH5, GH7, GH8, HM2, HM4
IVS-09	GH6, GH9, HC2
IVS-10	GH6, HM7, HM8
SEF-02	LA2
SEF-04	LA3
TVM-01	GH7
TVM-02	GH7

- *REQ-1: A Security Policy compliant to existing Information Security management standards exists and is operative.* This is due to the requirements of established baseline security requirements (GRM-01) and an information security policy (GRM-06), addressing of access policies (IAM-04, IAM-07, IAM-09) and customer access (AIS-02), as well as policies, ensuring protection of confidentiality, integrity and availability of data exchanged between the two entities (AIS-04). Moreover, regulatory compliance has to be maintained by defined roles and responsibilities (AAC-03) as well as a trust model established (IAM-08). The policy also requires legal preparation for incident response (SEF-04). It also requires asset classification (DCS-01), equipment identification and inventory (DCS-03) as well as authorization prior to relocation or transfer of hardware, software or data (DCS-04).
- *REQ-2: Fully established hypervisor technologies are used and properly configured.* Hypervisors must ensure their competence se-

curity-wise. Therefore, they must ensure proper encryption and key management (EKM-01, EKM-04) as well as the protection of sensitive data (EKM-03). Moreover, established virtualization solutions include audit (IAM-01) and diagnostic (IAM-03) capabilities. Hypervisors and the operating systems running on virtual machines inside the hypervisor must be hardened, leaving only necessary ports, protocols and services open (IVS-07). A hypervisor also must ensure that there is no information leakage between the VMs running on it (IVS-09) as well as safety and proper level of encryption during the migration of VMs due to resource distribution (IVS-10).

- *REQ-3: Fully established provisioning and security solutions are used and properly configured.* Integrity of data (AIS-03) and changes are monitored (IVS-02) as well intrusion detection systems are running (IVS-01). The clock is synchronized to facilitate activity timelines (IVS-03). Incident response and forensic solutions are available (SEF-02) as well as solutions for malicious software prevention (TVM-01) and patch management (TVM-02).

Information security risk analysis procedure is dependent not only on the vulnerable objects facing the network, but the nature of the attack as well. The nature of the attack depends on the goal of the attacker. Defining the attack leads to modeling of various scenarios that an asset may face security-wise.

A typical way of representing the attack steps is by using attack trees (Schneier 1999). While attack tree is a comprehensive tool for complex system security evaluation, the inability to represent the interaction between attacks and defenses limits the depth of the defense strategy analysis. This leads to inability of system security evolution representation using attack trees, as they do not take defender's actions into account. An extension to attack trees – the attack-defense trees (ADTrees), proposed by Kordy *et al.* (2012), adds the ability to represent interaction between an attacker and defender thus allowing a comprehensive model to be built.

An attack-defense tree has two types of nodes: attack and defense, which represent goals of an attacker and a defender respectively. Such tree also has a capability of refinement and countermeasure representation. Each node can have one or more children, specifying the sub-goals of the main goal. Each node can have one opposite that represents a countermeasure. The refinement can be of conjunctive or disjunctive nature, where disjunctively refined node is achieved if at least one of its children's goals is achieved. Conjunctively refined node is achieved if all of the children's goals are achieved. This method allows formal representation of attack-defense scenarios that, with proper node weights, can be simulated to acquire predictions for real-life situations. The formalization of this type of trees requires that attack nodes are represented as circles and defense nodes as rectangles. Refinement relations are indicated by solid lines.

Formal ADTree analysis is performed by defining the ADTree with an abstract syntax term (ADTerm) that is linked to a signature (AD-signature). An AD-signature is defined by an unranked function F with domain D and a range R that denotes a family of functions $(F_k)_{k\in\mathbb{N}}$, where $F_k: D^k \rightarrow R$, for $k > 0$. An ADTerm can be of the proponent's (denoted as $\mathbb{T}_{\Sigma}^{p}$) and opponent's (denoted as $\mathbb{T}_{\Sigma}^{o}$) type, where the proponent's type constitute a formal representation of AD-Trees. An ADTree is considered to be a finite ordered tree T over the set of labels $L_T = \mathbb{B}^p \cup \mathbb{B}^o \cup \{\vee^p, \wedge^p, \vee^o, \wedge^o\}$ and function $\lambda: Pos(T) \rightarrow \{\circ, \square\}$ with two conditions – condition 1 ensures that each node p of an ADTree is either refined in a conjunctive or disjunctive way. Condition 2 specifies the requirement that each node p may only have one child of the opposite type. And it must be depicted as the rightmost child node of p (Kordy *et al.* 2012).

The function λ distinguishes between attack and defense nodes. Value $\lambda(\varepsilon)$ determines whether the attacker or the defender is the proponent of the tree. Comparison of the λ values of the parent node with the values of λ applied to its children refined and non-refined nodes are specified. A node p is considered to be refined with the condition that it has at least one child p_i. A non-refined node can have at most one child p_1, and this child needs to satisfy $\lambda(p) \neq \lambda(p1)$.

For this case, the attack and defense surface is modeled as an attack-defense tree where the root represents the goal of the attack, while the rest of the nodes represent attack steps and defenses to mitigate the risk. As the main goal is considered to be the control of the hypervisor. The attack-defense tree is designed to represent the requirements for secure hypervisor deployment and management, specified in subchapter 2.3.

From the attacker point of view hypervisor security is evaluated by the possibility to apply certain attack scenarios to gain the desired attack goal. Defining attack scenarios enable the definition of the attack goals and modeling the means and methods to acquire them. Common hypervisor attack goals, specified by (EC-Council 2011) include:

1. *Denial of service (DoS)* (ATT_1) – the main aim is to shut down a hypervisor or plant a backdoor to access the VMs. The attack for this goal is modeled based on the classification of DoS attacks provided by (Prasad *et al.* 2014), extending them with the countermeasures provided by (Mirkovic, Reiher 2004). Analysis of this attack scenario for the purpose of risk assessment is provided in (Sommestad *et al.* 2011). The attack detection defense mechanism in this case covers the following requirements presented in subchapter 2.3:
    a) Pattern, anomaly detection – REQ-3, as it involves technical provisioning and security solutions, including change monitoring and intrusion detection systems;
    b) 3rd party detection – REQ-1, as it involves passing the responsibilities to a 3rd party supplier that requires legislative and governance documentation;

2. *Jump Virtual machines (VM)* (ATT_2) – using a security hole in the hypervisor, a user logged into one VM can jump to another VM. VM jumping can be achieved due to non-secure (old, obsolete and vulnerable) OS running on the VMs and/or no separation between traffic between VMs and external network (Reuben 2007). Although this attack scenario may seem to be virtual machine specific, the concerns of hypervisor weaknesses make it a hypervisor attack scenario as well. The countermeasures in this case cover the following requirements presented in subchapter 2.3:
    a) Patching – REQ-3, as it involves malicious software prevention and patching itself;
    b) Network segregation – REQ-2, as it is a technical operational requirement, including configuration hardening;
3. *Intercept host traffic* (ATT_3) – exploiting a vulnerability in the hypervisor to track system calls, paging files, memory, and disk activity. Caused by non-hardened configuration which leads to promiscuous virtual network adapter mode and disabled switch traffic examination. It allows for placing of frames with a forged MAC address to the network, thus changing the destination of the packet from the legal recipient virtual machine to the malicious one (Borza *et al.* 2004). The resulting attack-defense model includes configuration hardening requirement (REQ-2) and traffic monitoring and examination (REQ-3).

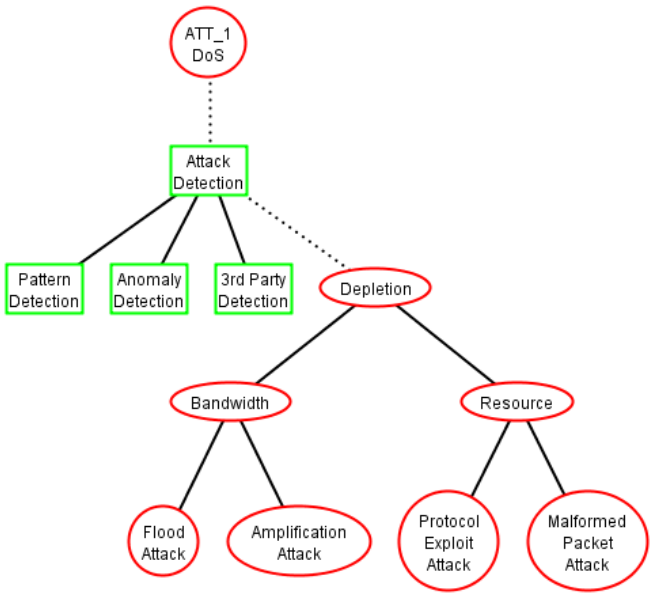


Fig. 2.4. Attack-defense tree for Denial of Service (source: author)

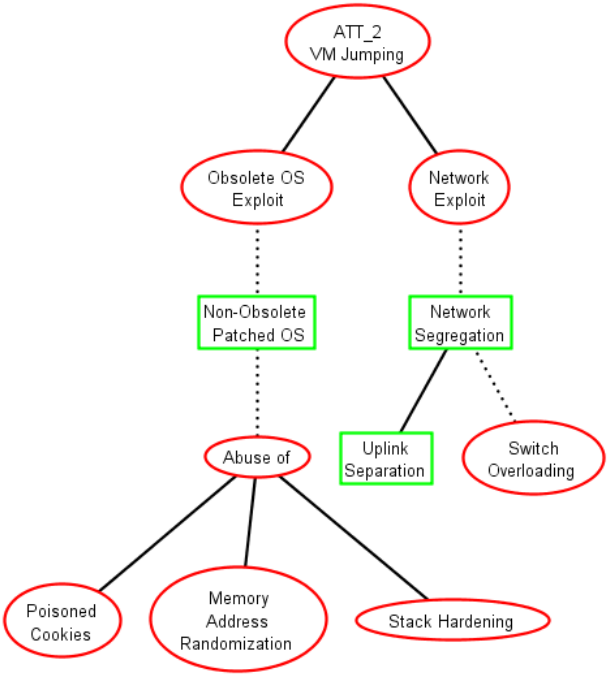


Fig. 2.5. Attack-defense tree for Virtual Machine Jumping (source: author)

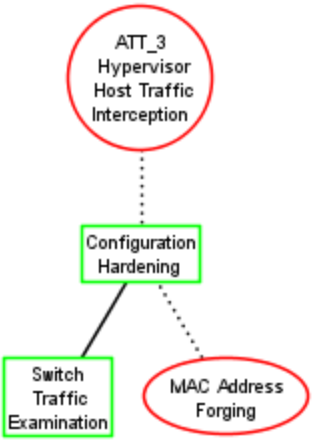


Fig. 2.6. Attack-defense tree for Hypervisor Host Traffic Interception (source: author)

Expanding them with the relevant ones used by the Operating System concept of Cyber Security Modeling Language (Sommestad *et al.* 2013) adds the following attack goals with predefined attack-defense trees (Holm *et al.* 2013):

- *Find unknown service* (ATT_4) – if the hypervisor is not managed based on the least privilege principle, unknown services can be exploited to gain control over the hypervisor;
- *Find exploit* (ATT_5) – involves looking for unknown or unmanaged vulnerabilities to gain control over the hypervisor;
- *Execute arbitrary code* (ATT_6) – access and privilege management issues as well as security holes lead to possibility of injecting and executing arbitrary code. Analyzed in (Sommestad *et al.* 2012);
- *Deploy exploit* (ATT_7) – find a security hole and deploy an exploit;
- *Compromise* (ATT_8) – intentional or unintentional impact on confidentiality, integrity or availability by an untrusted source.

2.4. Definition of the Virtualization-security related Variables

The variables, analyzed for this study are based on the attack scenarios of the hypervisor. While most of these attack scenarios are analyzed in the previous research (Sommestad *et al.* 2011, 2012; Sommestad 2012) the two new attack scenarios dealing with *virtual machine jumping* (ATT_2) and *host traffic interception* (ATT_3) require expert information acquisition. The variables of these scenarios are presented in Table 2.4.

Table 2.4. Variables for hypervisor risk analysis

Scenario	Variable	Description and validation
ATT_2	SV	*Software vulnerability,* allowing the VM jumping (e.g. exploiting memory corruption vulnerabilities (National Vulnerability Database 2014), faulty port and virtual device management (Kortchinsky 2015)).
	AC	*Access vulnerability* enabling the attacker to access the administrative environment and make changes to the system (e.g. privilege escalation (National Vulnerability Database 2012)).
ATT_3	CF	*Secure configuration* plays an essential role in the overall hypervisor security, as improper configuration of the hypervisor can expose it to misusage (Shackleford 2010).
	TI	*Traffic isolation* is a very important aspect of hypervisor security, as there is a tendency of host traffic interception by guest systems on the hypervisor (Oracle 2012).
	VF	*Virtual firewalls* prevent any undesired activity within the virtualization environment (Garber 2012).

Numerical values for the variables presented in Table 2.4 are required for the risk analysis process. These numerical values have to be acquired from a highly validated and trusted source.

Vulnerabilities and their exploitation are on the main focus for the past decade, with initial research approaches analysing the vulnerabilities on a large scale (Frei *et al.* 2006). As the topic evolved, associations between the exploitability scores and real-life situations have been developed to predict future incidents (Bozorgi *et al.* 2010), along with a deeper approach, specifically analysing CVSS exploitability evaluation (Allodi, Massacci 2014). The overview of the existing information security vulnerability datasets is presented in Table 2.5.

Table 2.5. Existing vulnerability and exploit databases

No.	Name	Size	Description
1.	NVD	74 100 entries	National Vulnerability Database. Contains all disclosed vulnerabilities (CVE – Common Vulnerabilities and Exploits (International Telecommunication Union 2012)) and respective CVSS assessment. Maintained by US Governmental institutions: NIST, DHS, NCCIC and US-CERT. High validity and reliability.
2.	EDB	35 220 entries	The Exploit Database is a CVE compliant archive of public exploits and corresponding vulnerable software. Contains reports of the vulnerabilities for which a "proof-of-concept" exploit exists. Maintaned by private company – Offensive Security.
3.	EKITS	216 entries	Specifies exploit kits and market service prices in the black market (Allodi, Massacci 2012).
4.	Symantec Security Response threat writeups	1 125 entries	Contains the exploits for the computer operating systems that run Symantec security products. Types of exploits are: virus, worm, macro, misleading application, potentially unwanted app, parental control, adware trojan, dialer adware, adware, removal information, hoax, spyware and security assessment tool trojan. Source validated and reliable.
5.	0day.today	24 707 entries	Collects exploits from submittals and various mailing lists and concentrates them in a database. Source not validated.
6.	Rapid7 Vulnerability and Exploit Database	69 092 entries	Replicates the CVEs from the NVD.

The Table 2.5 entries No. 5 and 6 are excluded from the analysis due to the lack of validity of No. 5 and the replicative nature of No. 6. The mapping of the rest of the provided dataset coverage, created by (Allodi, Massacci 2012) is presented in Fig. 2.7, where the numbers match the "No." column of Table 2.5. The colors red, orange and cyan represent high, medium and low score vulnerabilities respectively. It is worth noticing that the amount vulnerabilities in the categories of low and medium in the NVD dataset is disproportionally high with respect to the others.

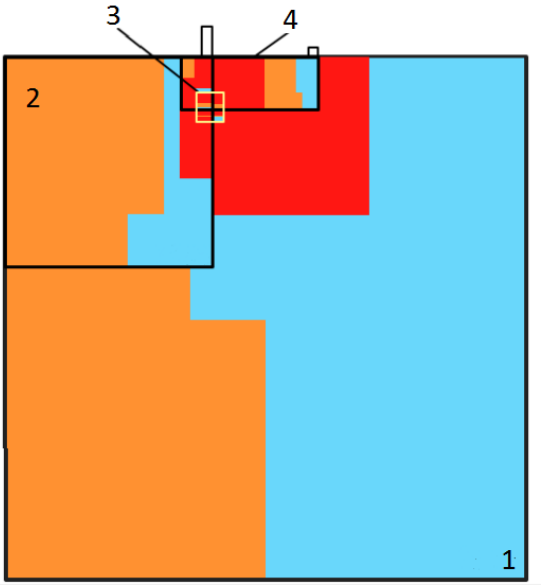


Fig. 2.7. Relative map of vulnerabilities per dataset: 1 – NVD; 2 – EDB; 3 – EKITS; 4 – Symantec; 5 – 0day.today; 6 – Rapid7 (Allodi, Massacci 2012)

The analysis of the informativeness of datasets has shown, that the NVD dataset includes the rest of the analyzed datasets (Allodi, Massacci 2012), therefore only the NVD is used as the primary dataset for the scoring.

The Common Vulnerability Scoring System (CVSS) managed by National Institute of Standards and Technology and FIRST provides a robust scoring system for IT vulnerabilities with input from representatives of a broad range of industry sectors, from banking and finance to technology and academia. It runs incorporated to the National Vulnerability Database which is the U.S. government repository of standards based vulnerability management data. The CVSS scores are considered to be the de facto standard for risk measurement (Allodi, Massacci 2012). Usage of CVSS scores for definition and assessment of quanti-

tative security risk measures has been introduced by Joh and Malaiya (Joh, Malaiya 2011).

The system contains vulnerability scores of three categories provided in decimal scoring scale: *Base scores, Temporal* and *Environmental*. The interest of this study is for the *Exploitability* score found under the *Base scores* category. The exploitability score is an empirical metric, evaluating: *attack complexity (C_AC), privileges required (C_PR), attack vector (C_AV)* and *user interaction (C_UI)* parameters. This metric reflect the ease and technical means by which the vulnerability can be exploited. The attack complexity metric describes the conditions beyond the attacker's control that must exist in order to exploit the vulnerability. These conditions may require the collection of more information about the target, the presence of certain system configuration settings, or computational exceptions. The privileges required metric describes the level of privileges an attacker must acquire before the exploitation of the vulnerability. This metric if greatest if no privileges are required. The attack vector metric reflects the context by which vulnerability exploitation is possible. The user interaction metric represents the requirement for a user, other than the attacker, to participate in the successful compromise of the vulnerable component. This metric determines if the vulnerability can be exploited by only an attacker, or a separate user must participate in some manner (FIRST 2014). The method (FIRST 2014) suggests that the *Exploitability* score is calculated as follows and presented in (2.1):

$$E_{CVSS} = 8.22 \cdot \prod_{k=1}^{4} C_k, \tag{2.1}$$

where the parameter values C and their associations to the indexes k are presented in Table 2.6.

Table 2.6. CVSS v. 3.0 Variable metric value association to numerical values (FIRST 2014)

Index	k parameter	Metric Value	Numerical Value
AV	1	Network Adjacent Network Local Physical	0.85 0.62 0.55 0.20
PR	2	None Low High	0.85 0.62 0.27
AC	3	Low High	0.77 0.44
UI	4	None Required	0.85 0.62

The score of the hypervisor risk analysis variables is defined by extracting the related vulnerability *Exploitability* scores from the CVSS. The sample set for each of the variables is set based on the sample size requirements of (Bartlett *et al.* 2001). The requirements specify that appropriate sample size is one of four features of a study design that have influence on detection of significant differences, relations or interactions. This means that appropriate sample size minimizes the *alpha error* (finding a difference that actually does not exist in the population) and *beta error* (failing to find an actual difference in the population). Therefore the magnitude of the alpha error is opposite to confidence level. The sample size is calculated using Cochran's sample size formula for continuous data (Bartlett *et al*. 2001):

$$n_{(o)} = \frac{t^2 \cdot s^2}{d^2}, \tag{2.2}$$

where t is value for selected alpha level in each tail of the curve, s is the estimate of standard deviation in the population and d is acceptable margin of error. The standard deviation s calculated based on (Bartlett *et al.* 2001) is:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{2.3}$$

where x_i is value of each member of the set and $\bar{x}$ is the mean of the set.

Given the confidence level is 0.95, and therefore the alpha error 0.05, the value of t is set to 1.96 based on (Bartlett *et al.* 2001), continuous data margin error is .03 multiplied by the number of points on primary scale, which, based on the CVSS method (FIRST 2014) is 10. The calculations result in standard deviation $s = 2.128365$ and the sample size $n_{(o)} = 19.33579 \approx 20$ for each variable, specified in Table 2.4.

The overall population of the set in this study is within the CVSS entry *exploitability* scores of the three most recent years associated with virtualization technologies.

Additional study of *exploitability* score distribution over time from 2013 to 2015 has shown that the average *exploitability* score of discovered vulnerabilities has a trend to increase over time (Fig. 2.8).

To determine the procedures of usage of the CVSS data, the normality of the sample data is determined by performing two tests – D'Agostino skewness and Anscombe-Glynn kurtosis interpretation (Ghasemi, Zahediasl 2012), and statistical normality testing using Kolmogorov-Smirnov test for datasets with the size of 50 and more and Shapiro-Wilk test if datasets are smaller (Devore 2012).

The D'Agostino skewness test (1970) and kurtosis (Anscombe, Glynn 1983) are measures of fit of departure from normality defining whether or not

the given sample comes from a normally distributed population. A perfectly normal distribution should return a score of 0. Positive value indicates positive skew or kurtosis and negative value indicates the negative. The higher the absolute value, the greater the skew or kurtosis.

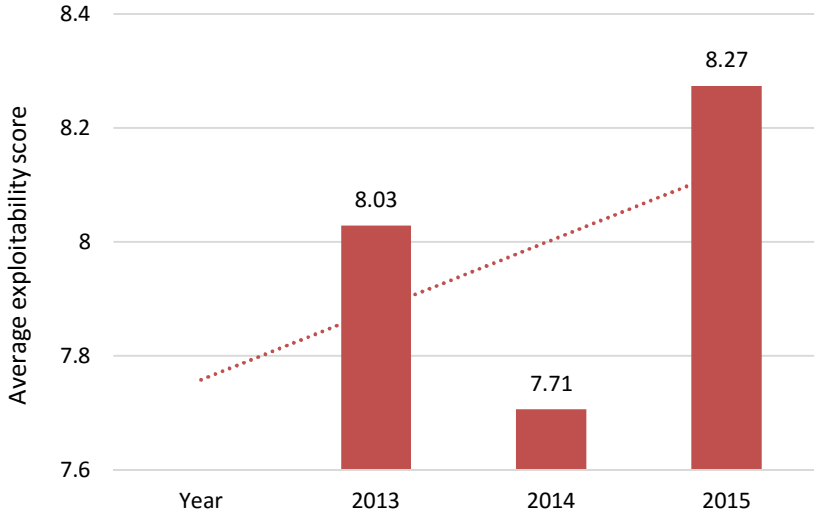


Fig. 2.8. Average CVSS exploitability score distribution over 2013–2015 (source: author)

The Kolmogorov-Smirnov test (Razali, Wah 2011) statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. However, for datasets with less than 50 elements the Shapiro-Wilk test is used. It utilizes the null hypothesis principle to check whether a sample came from a normally distributed population (Devore 2012).

Mutually exclusive and exhaustive events are described using the Bayes' Law of Total Probability is applied to define the variable scores, expressed as probabilities of occurrence. The Law of Total Probability is expressed as (Devore 2012):

$$P(B) = \sum_{i=1}^{k} P(B|A_i) \cdot P(A_i), \tag{2.4}$$

where the Total probability *P(B)* is the sum of conditional probabilities of the occurrence of individual events $P(B|A_i)$ multiplied by the prior probabilities $P(A_i)$ where the prior probabilities of all events are considered to be equal.

2.5. Expertise acquisition and processing

To prove that information security risk occurrence probabilities can be evaluated from processed vulnerability exploitation scoring of validated sources is true, expert data is used for comparison. The expert data, required for the risk analysis is acquired based on the attack scenarios). This approach is of an empirical nature, when evaluations of multiple human-experts are synthesized to deliver a unanimous metric. Each of the attack scenario has a different attack goal, and depending on the goals specific data, expressed as probabilities is collected. Some of the attack goals, as Denial of Service (Sommestad *et al.* 2011) and arbitrary code execution (Sommestad *et al.* 2012) are thoroughly covered and described in detail, while the newly introduced ones – such as VM Jumping and host traffic interception require an additional study. The rest of the attack scenarios are implemented to CySeMoL, therefore they are reused for this study (Sommestad *et al.* 2013).

Collection of the expert data is performed based on the method introduced in (Sommestad *et al.* 2011) and extensively described in (Holm *et al.* 2014). The method suggests that at first the variables that have an impact on the situation have to be defined for the study. These variables are factors that play an important role in the performance and safety of the assessed component. They are used as the basis for modeling the questions to acquire expertise.

A study on risk assessment probability combinations provides a method for combining varying risk assessment results as well as supports the need of multiple experts by concluding that uncertain quantities are better assessed by groups rather than a single expert (Clemen, Winkler 1999). This research uses the Cooke's classical model (Cooke 1991) to combine the risk assessment data, as it significantly outperforms the other methods (Cooke 2008).

The Cooke's classical model (Cooke 1991) builds a weighted expert probability assessment combination based on the proper scoring rule theory, where good calibration and low entropy are the main factors. This model elicits quantiles from the experts' distributions.

These scores are calculated based on the experts' answers to the specific seed questions. A seed question is a specific type of question, for which the correct answer is known at the time of analysis and is used for evaluation of the experts' knowledge.

The calibration score shows the deviation of the respondent's evaluation scores from the true values of the known seed questions. These questions require the respondents to specify a probability distribution to describe an uncertain continuous variable that is divided into a number of ranges. For this calibration it is divided into four ranges with the dividers being 5^{th}, 50^{th} and 95^{th} quantile values

based on (Cooke 2008). Let $s = s_1, \ldots, s_n$ be a probability distribution and assuming $p_i > 0$, $i = [1; 4]$; then the relative information of s with respect to p is (Cooke 1991):

$$I(s,p) = \sum_{i=1}^{4} ln \frac{s_i}{p_i}. \tag{2.5}$$

$I(s, p)$ is an index of the information learned if it was believed that p was correct, but subsequently learnt that s is correct.

A set of experts $e = 1, \ldots, E$ assesses probabilities of each uncertain event. They assign the corresponding indicator functions to one of B probability bins that are associated with a distribution over the possible outcomes. These bins are described by the probability p_b of occurrence, in the range of [0; 1], $b = 1, \ldots, B$.

Based on the observed values and the assignments, weights w_e are determined for each expert. The weights have to satisfy the following: $w_e \geq 0$ and $\sum w_e = 1$. The weight w_e is defined for each expert individually.

Let n_b be a number of variables assigned to b, s_b – the sample distribution of variables in bin b and N – sum of the variables n_b and $H(p_b)$ – the probability vector. Then the average response entropy is:

$$H_e(n) = \frac{1}{N} \sum n_b \, H(p_b). \tag{2.6}$$

The calibration score is:

$$C(e) = 1 - \chi_B^2 [\sum 2 n_b I(s_b, p_b]. \tag{2.7}$$

If the expert sample distribution realizations are drawn independently from a distribution with quantiles as stated by the expert, then the likelihood ratio statistic $2NI(s, p)$ is asymptotically distributed as a chi-square variable with 3 degrees of freedom (Cooke 2008). Then (2.7) becomes:

$$C(e) = 1 - \chi_3^2 \big(2NI(s,p)\big). \tag{2.8}$$

As opposed to the entropy, the information score is the second variable for the scoring. In a distribution, the information is the distribution concentration degree. Concentration or dis-concentration is measured relative to some other distribution. The information is expressed as:

$$I_e(n) = \frac{1}{N} \sum_{i=1}^{N} I_i. \tag{2.9}$$

Expert assessment combination is called a decision maker. A "good expertise" is considered to have good calibration and good information. Weights are associated to reward the "good expertise" in the process of decision making.

If the expert calibration score is above the set threshold, the weight of the expert e is the multiplication of calibration and information scores:

$$w_\alpha(e) = C(e) \cdot I_e(n). \tag{2.10}$$

Otherwise, the weight is set to zero. The threshold is set at the optimal position that is described as the highest possible weight of a virtual expert (Sommestad *et al.* 2011).

The seed questions are required to be very well validated and fall for the same domain of which the unknown variables are. These seed questions serve as an expert performance evaluation mechanism leading to weighting the importance of individual expert's opinion to the whole dataset. It is important to introduce these questions to the whole evaluation as seamless as possible for the expert to be able to identify them (Cooke 1991). The robustness of the weighting is highly dependable on the number of seeds used. Based on (Cooke 1991) eleven questions is enough to recognize substantial difference in calibration. These questions include both – the vulnerabilities to the hypervisor with known characteristics as well as the ones that the values are to be determined. The known answer values of known questions are taken from the National Vulnerability Database (NIST 2015) managed by the US Department of Commerce (questions 1–5), ENISA Threat Landscape (Marinos 2014) (question 6), Alert Logic Cloud Security Report (Coty *et al.* 2014) (questions 7–8), Bitglass Cloud Security Spotlight Report (Bitglass 2015) (questions 9–10) and 2nd Watch AWS Scorecard January – March 2015 (2nd Watch 2015) (question 11).

To retrieve the values of unknown variables, a survey is built for the interaction with the experts. Based on the requirements, specified in this chapter and the variables which vales need to be retrieved, specified in subchapter 2.4 a survey containing seed questions is built. The seed question of the survey are presented in Table 2.7. The questions for variable evaluation are presented in Table 2.8.

The unknown probabilities of the risk variables are acquired from the expert questions a.–e. of Table 2.8. The quality of expert knowledge is to be expressed by the weights that are to be assigned to the experts depending on their knowledge. The quality of knowledge is assessed using the Cooke's expert elicitation method and the expert input correlation to the known facts in Table 2.7.

To collect the required information – a set of experts of the domain is selected and their expertise acquired.

Table 2.7. Seed questions of the survey for hypervisor security expertise acquisition

No.	Question	Value, %
1.	What is the share of known vulnerabilities with some impact on virtualization technologies?	3
2.	Of the known vulnerabilities with some impact on virtualization technologies, what is the portion that is hypervisor related?	73
3.	Of the known hypervisor related vulnerabilities what is the average severity?	52
4.	Of the known hypervisor related vulnerabilities what is the highest recorded severity?	93
5.	What is the share of known vulnerabilities with some impact on virtualization technologies that have affect VMware products?	52
6.	What is the average growth of bandwidth of DDoS attacks between years 2013 and 2014?	70
7.	What is current public IT Cloud Computing annual growth rate?	24
8.	What is the share of the overall incidents of the Cloud Hosting Providers to web application attacks?	44
9.	What is the share of increase of the security effectiveness that security training and awareness increases?	45
10.	What portion of the data, stored in the Cloud is email?	45
11.	What is the share of Linux-based cloud servers to overall number for Amazon Web Services?	24

Table 2.8. Questions for expertise variable evaluation

a.	What is the probability of hypervisor exploitation using software vulnerabilities?
b.	What is the probability of hypervisor exploitation using access vulnerabilities?
c.	What is the probability of hypervisor exploitation due to insecure configuration?
d.	What is the probability of hypervisor exploitation due to traffic isolation issues?
e.	What is the probability of hypervisor exploitation due to lack of virtual firewalls?

The experts for this study have been selected based on their publications in the field of virtualization security. Therefore, authors with publications within 2010 and 2015 that have keywords of virtualization, hypervisor and virtual machine monitor and are included in Thomson Web of Knowledge, SCOPUS

and INSPEC databases have been selected as experts. Of those, the ones with public contact information have been included to the list, which resulted in 638 entries. 117 of these entries turned out to be outdated, therefore they were discarded from the list. The rest of the experts have been invited to participate in the study. 31 of 521 responded to the call. Based on the results of (Ryan *et al.* 2012) and (Cooke 2008) this number of experts is sufficient to achieve consistent results.

2.6. Conclusions of Second Chapter

1. A taxonomy, combining all the threats found in the threat landscapes was proposed. Additionally, four threats are expanded in this taxonomy to provide a more detailed picture. As the three main analyzed sources differ by the application field (technical (Pearce *et al.* 2013), hypervisor deployment and governance (Chandramouli 2014) and commercial product deployment (Shackleford 2010)), the proposed taxonomy represents a complete virtualization threat landscape.
2. The comparison of the previous categorization approaches to the newly proposed taxonomy is evaluated based on the National Information Standards Organization standard ANSI/NISO Z39.19 and the main requirements of:
   a) structure: hierarchical classification of information – this is represented by the number of (sub)categories. The proposed taxonomy that has 8 exceeds the nearest approach by two (sub)categories additionally providing three levels of subcategorization for easier implementation;
   b) comprehensiveness: the taxonomy should cover the objective domain – the number of nodes in the proposed taxonomy – 34 exceeds the nearest approach by fifteen entries;
   c) soundness: the taxonomy should meet the standard requirements – the number of narrower terms for a broader term falls into the recommended range of [3; 20], the identifiers for each node for easier navigation are assigned.
3. To simplify the input of risk analysis without the loss of initial information, three groups of control means for secure hypervisor deployment and management, serving as security requirements have been proposed.
4. A survey is built to acquire the expert evaluation on the variables, required for this study. The unknown probabilities of the risk variables are acquired from the expert questions a.–e. of Table 2.7. The quality of expert knowledge is to be expressed by the weights that are to be assigned to the experts depending on their knowledge. The quality of knowledge is assessed using the Cooke's expert elicitation method and the expert input correlation to the known facts.

3

Application of the Virtualization Risk Analysis to CySeMoL

Combining the information, presented in chapter 1 with the existing CySeMoL capabilities leads to an introduction of a new – Hypervisor meta-class to the CySeMoL meta-model for the automated information security risk analysis. The implementation of information presented in chapter 2 is achieved by performing analysis of:

1. Risk of occurrence of hypervisor specific threats, provided in Fig. 2.2 serves as the cornerstone for this risk analysis as it defines the existing sources for the risks.
2. Specific controls for threat impact mitigation, presented in Table 2.3 provide detailed means and methods to cover the existing threats and therefore minimize the risk.
3. General requirements for hypervisor security assurance, provided in subchapter 2.3 is a generalized approach to represent the data as requirements to ensure the security of the hypervisor.
4. Hypervisor attack scenarios incorporate the data of the first three aspects, supported by external technical documentation to provide the attack steps and countermeasures, required for traceable analysis of component risk.

The results of experiments, presented in this chapter are published in two papers (Janulevicius *et al.* 2016a, Janulevicius *et al.* 2016b). Subchapter 3.4.1 summarizes the publication with co-authors: Extension of CYSEMOL for cloud computing information security assessment; subchapter 3.4.2 summarizes the publication with co-authors: Content based model transformations: solutions to existing issues with application in information security.

3.1. Experimental variable evaluation

The required variable evaluation is performed by using two data sources – the CVSS scores obtained from the CVE database and the expert evaluations. The results of the two approaches are compared to draw conclusions regarding the possibility of substitution between the two sources.

To provide the data of hypervisor risk analysis variable scores (Table 3.1), a sample of 20 CVSS *exploitability* scores, of the common vulnerabilities and exposures (CVE), related to each variable are analyzed as described in sub-chapter 2.4.

Table 3.1. Score analysis for hypervisor analysis variables

Var.	CVE code	Exploit. score
SV	CVE-2014-8891	10.0
	CVE-2015-3629	3.9
	CVE-2013-5878	10.0
	CVE-2014-0428	10.0
	CVE-2014-0422	10.0
	CVE-2014-0416	10.0
	CVE-2014-0373	10.0
	CVE-2014-0368	10.0
	CVE-2013-5893	8.6
	CVE-2013-3515	3.9
AC	CVE-2013-2212	5.5
	CVE-2014-4632	8.6
	CVE-2014-8750	8.0
	CVE-2013-1211	10.0
	CVE-2013-3079	8.0
	CVE-2012-1833	10.0
	CVE-2014-1211	8.6
	CVE-2013-3107	8.6
	CVE-2013-1405	10.0
	CVE-2015-0259	4.9

Var.	CVE code	Exploit. score
SV	CVE-2014-8867	3.9
	CVE-2014-8866	3.4
	CVE-2015-5166	3.9
	CVE-2015-4106	3.9
	CVE-2015-2152	3.4
	CVE-2015-2044	3.9
	CVE-2014-4947	10.0
	CVE-2013-4361	3.9
	CVE-2013-4355	2.7
	CVE-2015-3456	5.1
AC	CVE-2013-6470	10.0
	CVE-2013-4471	10.0
	CVE-2014-2828	10.0
	CVE-2014-1948	1.9
	CVE-2013-2157	8.6
	CVE-2013-2059	6.8
	CVE-2013-0282	10.0
	CVE-2015-1950	3.9
	CVE-2012-4116	8.6
	CVE-2013-1186	10.0

The end of the Table 3.1

Var.	CVE code	Exploitability score
CF	CVE-2014-8370	10.0
	CVE-2015-3259	3.1
	CVE-2015-2152	3.4
	CVE-2013-4369	3.4
	CVE-2015-3646	8.0
	CVE-2015-1852	8.6
	CVE-2014-3703	10.0
	CVE-2014-7821	8.0
	CVE-2014-3632	4.9
	CVE-2011-4347	1.9
TI	CVE-2014-1207	8.6
	CVE-2013-5970	8.6
	CVE-2014-0852	8.6
	CVE-2014-3477	3.9
	CVE-2014-0984	8.6
	CVE-2014-0076	8.6
	CVE-2014-0006	8.6
	CVE-2013-4576	3.9
	CVE-2013-5915	8.6
	CVE-2013-4242	3.9
VF	CVE-2013-6398	5.5
	CVE-2013-5566	10.0
	CVE-2015-2841	10.0
	CVE-2014-3703	10.0
	CVE-2014-3555	8.0
	CVE-2015-4769	6.8
	CVE-2015-4767	3.2
	CVE-2015-2639	6.8
	CVE-2015-0593	8.6
	CVE-2015-0592	10.0
CF	CVE-2014-7144	8.6
	CVE-2014-3621	8.0
	CVE-2014-5356	8.0
	CVE-2013-1068	10.0
	CVE-2013-6470	10.0
	CVE-2013-6433	4.9
	CVE-2013-0266	3.9
	CVE-2015-2682	10.0
	CVE-2014-0201	3.9
	CVE-2014-0200	3.9
TI	CVE-2013-4165	8.6
	CVE-2013-1624	4.9
	CVE-2013-1623	8.6
	CVE-2013-1620	8.6
	CVE-2013-1619	4.9
	CVE-2013-1618	4.9
	CVE-2013-0169	4.9
	CVE-2013-6398	5.5
	CVE-2013-5566	10.0
	CVE-2015-0695	10.0
VF	CVE-2014-6383	10.0
	CVE-2014-6242	8.0
	CVE-2014-5350	10.0
	CVE-2014-4746	10.0
	CVE-2014-2519	8.6
	CVE-2014-3857	8.0
	CVE-2014-3813	10.0
	CVE-2013-7182	8.6
	CVE-2013-5092	8.6
	CVE-2014-0655	8.6

Further investigation of the variables includes analysis and graphical representation of variable distribution for the defined variables, presented in Figs. 3.1–3.5.

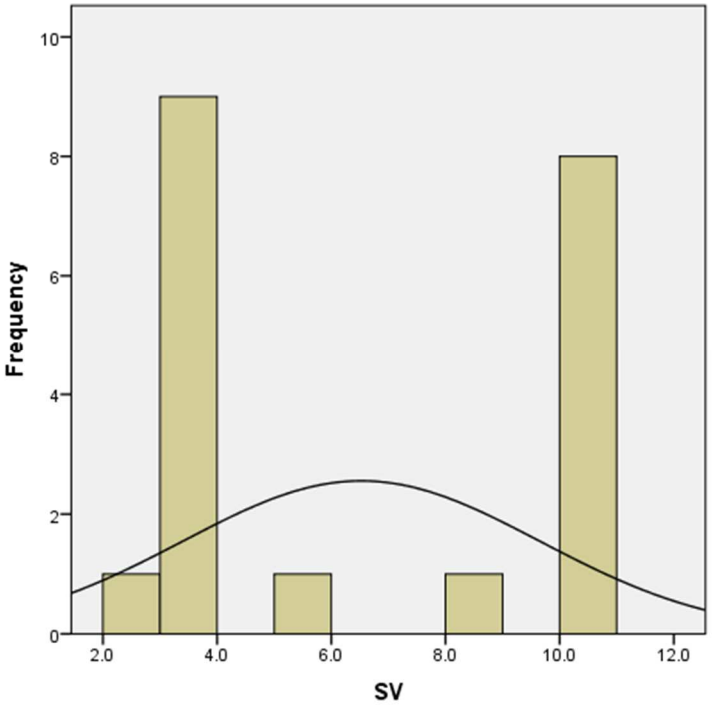


Fig. 3.1. Distribution histogram of the SV variable (source: author)

Variable SV has these characteristics: mean of $\mu_{SV} = 6.53$, standard deviation $\delta_{SV} = 3.12$, skewness $\gamma_{SV} = 0.19$, kurtosis $\kappa_{SV} = (–2.08)$. The Shapiro-Wilk test of normality returns the value of $W_{SV} = 0$.

From the graphical variable distribution inspection using Fig. 3.1 it is assumed that the normality of the distribution is false. Skewness and kurtosis characteristics, however, deny this assumption, as skewness and kurtosis of variable SV distribution falls in the range of [–1;1] which typically indicates whether the distribution can be considered normal (Devore 2012).

The null hypothesis (Shapiro-Wilk test) for the test of normality claims that the distribution of the variable is equal to the expected normal distribution, but due to the fact that the probability associated with the test of normality, less than 0, is less than or equal to the level of significance (0.01), the null hypothesis is rejected as non-descriptive. Therefore, it is concluded that the SV variable is not normally distributed.

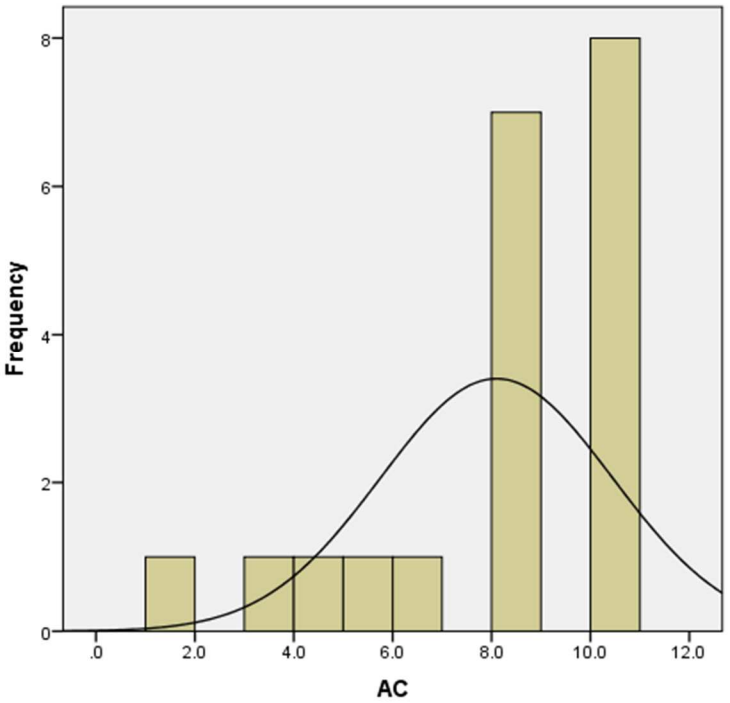


Fig. 3.2. Distribution histogram of the AC variable (source: author)

Variable AC has these characteristics: mean of $\mu_{AC} = 8.10$, standard deviation $\delta_{AC} = 2.35$, skewness $\gamma_{AC} = (-1.40)$, kurtosis $\kappa_{AC} = 1.31$. The Shapiro-Wilk test of normality returns the value of $W_{AC} = 0.001$.

From the graphical variable distribution inspection using Fig. 3.2 it is assumed that the normality of the distribution is false. Skewness and kurtosis characteristics support this assumption, as these characteristics are outside of the range of [–1;1];

The null hypothesis (Shapiro-Wilk test) for the test of normality claims that the distribution of the variable is equal to the expected normal distribution, but due to the fact that the probability associated with the test of normality, less than 0.001, is less than or equal to the level of significance (0.01), the null hypothesis is rejected as non-descriptive. Therefore, it is concluded that the AC variable is not normally distributed.

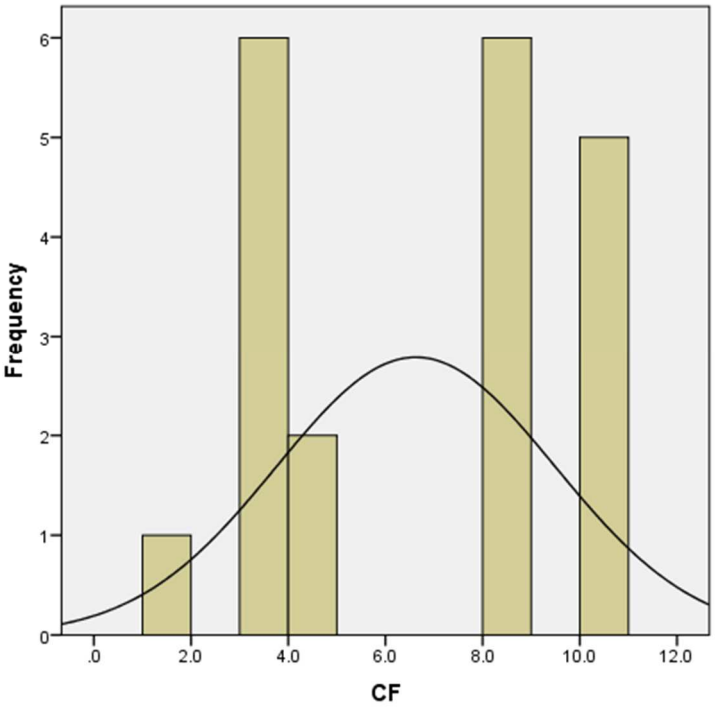


Fig. 3.3. Distribution histogram of the CF variable (source: author)

Variable CF has these characteristics: mean of $\mu_{CF} = 6.63$, standard deviation $\delta_{CF} = 2.87$, skewness $\gamma_{CF} = (–0.17)$, kurtosis $\kappa_{CF} = –(1.67)$. The Shapiro-Wilk test of normality returns the value of $W_{CF} = 0.001$.

From the graphical variable distribution inspection using Fig. 3.3 it is assumed that the normality of the distribution is false. Skewness opposes and kurtosis characteristic supports this assumption.

The null hypothesis (Shapiro-Wilk test) for the test of normality claims that the distribution of the variable is equal to the expected normal distribution, but due to the fact that the probability associated with the test of normality, less than 0.001, is less than or equal to the level of significance (0.01), the null hypothesis is rejected as non-descriptive. Therefore, it is concluded that the CF variable is not normally distributed.

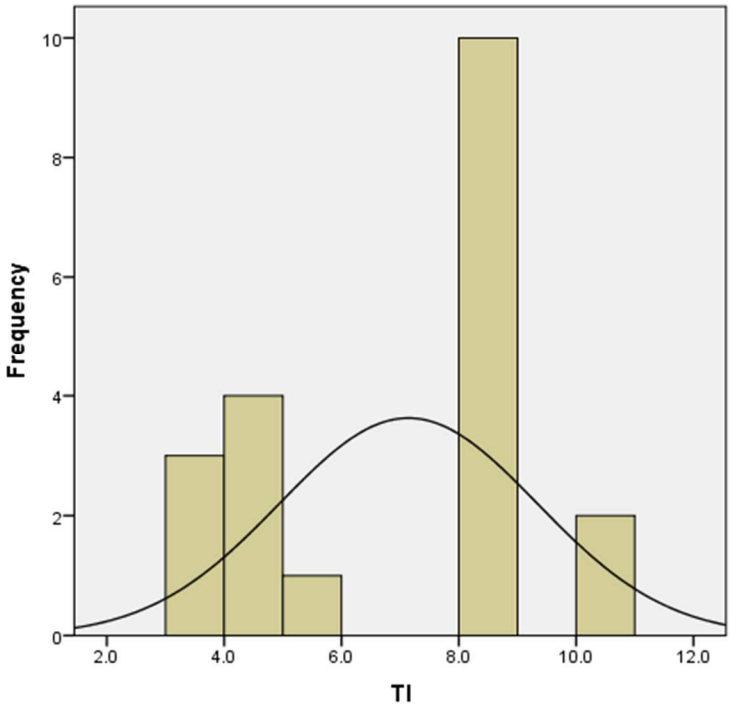


Fig. 3.4. Distribution histogram of the TI variable (source: author)

Variable TI has these characteristics: mean of $\mu_{TI} = 7.14$, standard deviation $\delta_{TI} = 2.20$, skewness $\gamma_{TI} = (–0.40)$, kurtosis $\kappa_{CF} = –(1.63)$. The Shapiro-Wilk test of normality returns the value of $W_{TI} = 0.001$.

From the graphical variable distribution inspection using Fig. 3.4 it is assumed that the normality of the distribution is false. Skewness and kurtosis characteristics support this assumption.

The null hypothesis (Shapiro-Wilk test) for the test of normality claims that the distribution of the variable is equal to the expected normal distribution, but due to the fact that the probability associated with the test of normality, less than 0.001, is less than or equal to the level of significance (0.01), the null hypothesis is rejected as non-descriptive. Therefore, it is concluded that the TI variable is not normally distributed.

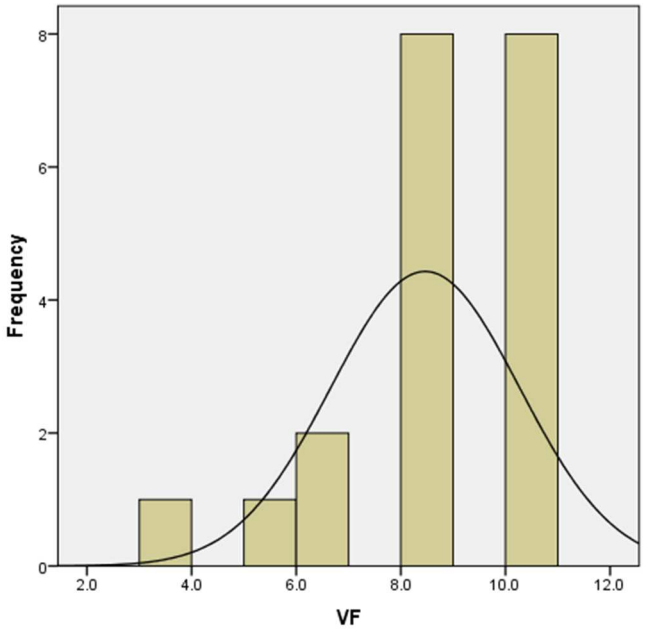


Fig. 3.5. Distribution histogram of the VF variable (source: author)

Variable VF has these characteristics: mean of μ_{VF} = 8.47, standard deviation δ_{VF0} = 1.80, skewness γ_{VF} = (–1.53), kurtosis κ_{CF} = 2.68. The Shapiro-Wilk test of normality returns the value of W_{TI} = 0.009.

From the graphical variable distribution inspection using Fig. 3.5 it is assumed that the normality of the distribution is false. Skewness and kurtosis characteristics support this assumption.

The null hypothesis (Shapiro-Wilk test) for the normality test claims that the variable distribution is equal to the expected normal distribution, but due to the fact that the probability associated with the test of normality, less than 0.001, is less than or equal to the level of significance (0.01), the null hypothesis is rejected as non-descriptive. Therefore, it is concluded that the VF variable is not normally distributed.

Since the normality of the score datasets is proven to be false, the Bayes' Law of Total Probability is not considered to be descriptive for such datasets. For such datasets a robust estimator that is less affected by the non-normality of the distribution is required. For this reason a statistical measure of central tendency – the winsorized mean – is used (Fuller 1991), with 10% of the extreme values winsorized. These values are used as the metrics of the variable. The observed variable scores are normalized by converting from decimal to centennial system to match with the output of other method's data for comparison. The scores are presented in Table 3.2.

Table 3.2. Variable scores (95% reliability)

Variable	SV	AC	CF	TI	VF
Upper_bound, %	80.47	93.00	80.81	82.09	95.18
Variable_score, % (Winsorized_mean_10%)	65.60	82.00	66.00	71.00	85.00
Lower_bound, %	51.23	71.00	53.99	61.51	78.32
Uncertainty, %	14.37	11.00	13.16	9.89	8.78

To acquire the expert assessment on the topic, a set of 31 experts, has been presented with the questions, provided in Tables 2.6 and 2.7. The assessment results of each expert are provided in Table 3.3. In this table, experts are identified by providing an identification code, consisting of the letter E and the sequence number. The cell values are the expert assigned probabilities for the occurrence of the previously expressed phenomena.

Table 3.3. Expert assessment of the seed and the hypervisor security variable probabilities (%)

Question	1	2	3	4	5	6	7	8	9	10	11	a	b	c	d	e
E1	10	65	75	100	60	55	40	50	40	50	30	60	70	70	80	70
E2	5	70	50	99	50	80	40	40	60	45	25	60	60	60	60	60
E3	8	70	20	90	35	40	20	30	30	60	50	50	70	60	65	50
E4	1	80	45	95	47	53	25	45	51	41	30	51	87	60	70	80
E5	10	60	36	99	55	65	30	42	55	33	20	70	80	65	69	55
E6	15	70	45	100	40	76	15	55	58	49	25	62	75	65	65	57
E7	4	75	50	90	48	70	25	45	50	45	25	67	80	68	70	80
E8	25	80	70	92	50	75	13	60	42	50	33	65	77	70	70	55
E9	56	40	80	70	20	25	60	50	20	70	50	90	50	80	30	40
E10	13	66	40	93	40	50	30	44	50	65	25	75	75	70	70	60
E11	9	50	45	99	65	35	22	40	40	60	25	70	75	60	70	50

The end of the Table 3.3

Question	1	2	3	4	5	6	7	8	9	10	11	a	b	c	d	e
E12	5	75	45	85	50	65	25	50	45	40	24	67	76	67	67	56
E13	40	40	80	85	10	10	45	70	30	50	45	50	50	60	60	50
E14	4	70	37	95	50	77	18	33	36	39	25	62	78	60	70	65
E15	3	72	55	95	50	70	20	40	45	40	20	65	85	65	72	80
E16	5	65	50	80	40	50	35	35	50	50	25	50	60	70	60	50
E17	10	75	45	97	45	55	30	40	50	50	24	60	70	60	70	60
E18	7	70	50	100	35	59	15	50	55	40	25	75	75	50	50	50
E19	1	72	50	99	50	50	20	49	50	33	25	45	50	45	60	80
E20	60	50	25	80	70	30	43	15	43	70	45	55	80	65	70	60
E21	10	76	57	90	50	65	21	50	50	45	24	60	65	50	70	60
E22	9	70	60	98	45	65	24	50	50	49	24	65	60	65	60	55
E23	5	75	50	90	50	70	25	45	40	40	25	65	80	65	70	85
E24	5	66	70	91	44	50	27	55	45	40	20	60	60	70	50	95
E25	2	67	67	95	51	47	30	47	50	45	24	40	65	75	70	50
E26	5	72	40	95	43	52	33	40	50	50	25	60	65	60	60	55
E27	10	70	45	95	48	70	25	40	55	55	24	60	80	70	70	60
E28	5	65	39	99	50	66	17	49	50	40	25	60	70	80	80	45
E29	2	70	40	98	49	70	21	35	50	45	25	50	70	65	65	60
E30	3	66	50	100	50	67	24	40	45	45	24	35	65	75	70	60
E31	2	71	50	97	50	56	20	45	40	35	25	50	80	70	65	55

The assessments of the experts are weighted to define the importance of each of their assessment using the Cooke's classical method. Based on the method three main metrics are acquired – the calibration score, mean relative total and the weight of an expert. The weights are then normalized, so that the sum of the weights equals 1. The results of these calculations are presented in Table 3.4.

Table 3.4. Weights of expert assessments

Id	Calibration	Mean relative total	Unnormalized weight	Normalized weight w_e
E1	0.37000	0.120	0.046	0.150
E2	0.20000	0.240	0.048	0.150
E3	0.20000	0.130	0.026	0.082
E4	0.01200	0.290	0.000	0.000
E5	0.20000	0.100	0.021	0.067
E6	0.20000	0.130	0.027	0.086
E7	0.01200	0.300	0.000	0.000
E8	0.39000	0.120	0.046	0.150
E9	0.15000	0.170	0.000	0.000
E10	0.01200	0.190	0.000	0.000
E11	0.20000	0.100	0.020	0.065
E12	0.01200	0.210	0.000	0.000
E13	0.15000	0.200	0.000	0.000
E14	0.01200	0.230	0.000	0.000
E15	0.00054	0.250	0.000	0.000
E16	0.06800	0.190	0.000	0.000
E17	0.06800	0.110	0.000	0.000
E18	0.06800	0.220	0.000	0.000
E19	0.00054	0.280	0.000	0.000
E20	0.39000	0.140	0.053	0.174
E21	0.01200	0.120	0.000	0.000
E22	0.01200	0.140	0.000	0.000
E23	0.01200	0.210	0.000	0.000
E24	0.01200	0.210	0.000	0.000
E25	0.06800	0.280	0.000	0.000
E26	0.06800	0.210	0.000	0.000
E27	0.20000	0.120	0.023	0.076
E28	0.01200	0.220	0.000	0.000
E29	0.00054	0.260	0.000	0.000
E30	0.00054	0.250	0.000	0.000
E31	0.00054	0.230	0.000	0.000

Based on the normalized expert weights and their assessment of each variable, the value of the variables is evaluated as a sum of multiplications of the two values. Therefore, the evaluation is performed using:

$$\sum_{i=1}^{n} w_{Ei} \cdot V_{Ei}, \tag{3.1}$$

where w_{Ei} is the weight of expert i, and V_{Ei} is the value assigned to the variable by the same expert.

Using the previously described statistical procedures, individual variable scores have been obtained from the provided expertise. Moreover, the magnitude of uncertainty has been defined. Based on this magnitude, it is assumed that the variable falls within the range, defined by the upper and lower bounds with a confidence level of 95%. The detailed results of the variable evaluation and distribution using expert knowledge are presented in Table 3.5.

Table 3.5. Variable evaluation using expert knowledge

Variable	SVe	ACe	CFe	TIe	VFe
Upper_bound, %	63.83	74.12	67.91	68.88	65.6
Variable_score, %	59.81	70.42	65.00	65.42	60.90
Lower_bound, %	55.78	66.71	62.09	61.96	56.20
Uncertainty, %	4.02	3.70	2.91	3.46	9.40

3.2. Comparison of the Results

To check if the CVSS scores can be substituted with the expert evaluation and vice-versa, a statistical procedure to measure the equivalence between the two datasets is carried out.

As a primary characteristic, a correlation test of the two datasets is performed. The correlation plot is presented in Fig. 3.6 The returned correlation coefficient value of $r = 0.0849$ suggests that the correlations is weak. However, visual inspection of the plot suggests that only the variable VF is strongly deviated from the tendencies. Therefore, disregarding this variable in the correlation analysis returns $r = 0.8964$ correlation coefficient, which shows a very strong relation between the two datasets. However, correlation only shows the similarity of pattern, leaving other important parameters,

such as magnitude out of scope. Therefore, a more detailed investigation is carried out.

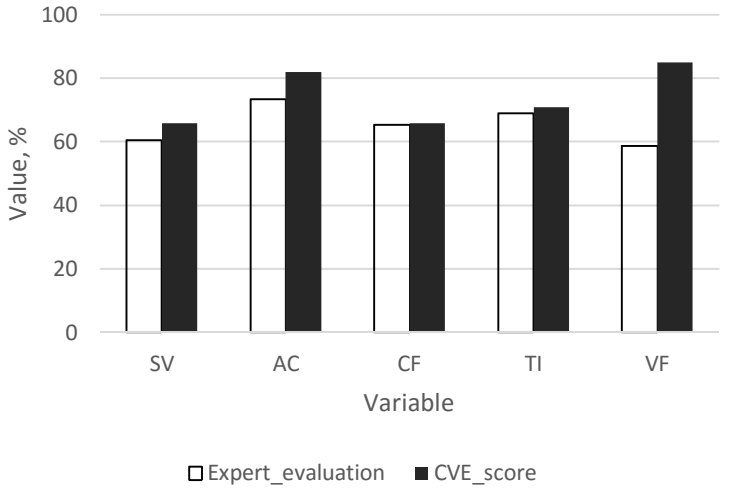


Fig. 3.6. Correlation of Expert evaluation and CVE scores (source: author)

Further investigation of the score equivalence is performed by using the Kruskal-Wallis H test, due to the variable data being non-normally distributed (Corder, Foreman 2009). It is a rank-based nonparametric test used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable. The test provides the χ^2 metric, which specifies the significance of the differences.

The Kruskal-Wallis H test is performed by the following:

$$\boldsymbol{H} = (\boldsymbol{N}-1)\frac{\sum_{i=1}^{g} \boldsymbol{n_i}\left(\left(\frac{\sum_{j=1}^{n_i} \boldsymbol{r_{ij}}}{\boldsymbol{n_i}}\right) - \frac{1}{2}(\boldsymbol{N}+1)\right)^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(\boldsymbol{r_{ij}} - \frac{1}{2}(\boldsymbol{N}+1)\right)^2}, \tag{3.2}$$

where n_i is the number of observations in group i; r_{ij} is the rank of observation j from group i; N is the total number of observations across all of the groups.

Based on this test, the calculated statistics are provided in Table 3.6.

Table 3.6. Assessment of the significance of the differences between the CVSS score and expert acquired data

Var.	df	p	χ^2	Sig. bound.	Conclusion
SV; SVe	5	0.05	4.500	11.07	non-significant
AC; ACe	4	0.05	6.523	9.49	non-significant
CF; CFe	2	0.05	1.436	5.99	non-significant
TI; TIe	4	0.05	5.000	9.49	non-significant
VF; VFe	4	0.05	1.768	9.49	non-significant

The comparison of the variables of Table 3.6 has concluded that there is no significant difference between the CVSS obtained scores and the expert evaluation. Based on the (NIST 2006) the significance boundaries (Sig. bound. of Table 3.6) are associated accordingly to the degrees of freedom (df).

3.3. Implementation of Research Data to CySeMoL

The Cyber Security Modeling Language (CySeMoL) is a sophisticated formalization for the assessment of cyber security on information system architectures. CySeMoL provides the ability of modeling in unified modeling language (UML) along with Bayesian attack graph realization through the object constraint language (OCL).

The output of the CySeMoL – a class diagram based heat-map specifying the difficulty for an attacker to reach and compromise different assets within the architecture from a specific entry point or multiple points. Since the attacker can be attached to any node within the infrastructure, the scenarios of attacks can be simulated according to the possible situations.

The CySeMoL is based on knowledge acquired from domain experts and observation studies. CySeMoL incorporates various components of the architecture from both – software and hardware sides, including computing and networking hardware, operating systems, web-servers, firewalls, network interfaces and many more (Sommestad *et al.* 2013).

The architecture here is represented as a graph where the nodes represent the components of the architecture with the edges representing their relationships by having specific conditional relationships between the nodes. The CySeMoL is based on the P2AMF framework (Johnson *et al.* 2013) that

provides advanced probabilistic reasoning about architecture models in the form of UML class and object diagrams. It works under the Object Constraint Language (OCL), adding a probabilistic inference mechanism.

The meta-class attributes in CySeMoL represent attack steps and countermeasures to these attack steps. The attack steps are parts of attack scenarios, which are executed in order to reach the attack goals. Countermeasures are described as counter-actions to prevent from the attack goals to be achieved. While the existing CySeMoL attack goals have the attack trees defined (Holm *et al.* 2013).

The hypervisor meta-class attribute list along with extended explanations is presented in Table 3.7. In this table an Identification number (*ID*) is given to every *Attribute* for future reference. The attributes are selected according to the requirements (*Req.*) of subchapter 2.3, representing the attack steps and countermeasures *Value source* describes the way data for the attribute value is acquired. *LOC* denotes that the value has to be specified manually and locally in the hypervisor class instance. *INH* denotes that the value of the attribute is inherited based on the *Condition* that requires a certain component to be present in the infrastructure. *Default Value* specifies whether the attribute is present (1) or absent (0) in the default state of the analysis.

The links between the variables and the proposed CySeMoL Hypervisor meta-class attributes are presented in Fig. 3.7.

Virtual Machine is an emulation of a physical computer system. From the software point of view the difference between the two types of machines is not visible. A VM performs the same tasks and procedures as a physical one, yet the only difference is that is not attached to a particular hardware setup, rather the resources are managed through the specific underlying layer of Hypervisor. Therefore, the only difference between the two is the Hypervisor being in between the OS and the hardware.

Since Virtual Machine is treated as a physical machine from the procedural side, to preserve the universal nature of the EA concepts, the Virtual Machine meta-class is rather expressed using the existing concepts of CySeMoL. More specifically – OperatingSystem and SoftwareProduct represent the Virtual Machine, on which the architecture of the machine can be built.

Table 3.7. Meta-class attributes and their values for the proposed CySeMoL hypervisor model

ID	Attribute	Req.	Value Source	Condition	Default Value
Countermeasures					
Policies and Standards					
C1.1	ProductSupported	REQ-1	LOC	–	1
C1.2	SecurityPolicyActive			–	1
Configuration Hardening					
C2.1	VNTrafficIsolated	REQ-2	LOC	–	1
C2.2	MACChangeRejected			–	1
C2.3	RootOverSSHOff			–	1
C2.4	HypervisorPatched			–	1
Provisioning					
C3.1	FirewallOn	REQ-3	INH	IF *Firewall* exists – 1, ELSE – 0	–
C3.2	IntrusionDetectionOn			IF *IDSSensors* exist – 1, ELSE – 0	–
C3.3	AntiMalwareOn				
Attack Steps					
A01	ExploitFloodAttack	–	INH	IF *IDSSensors* exist – 0, ELSE – 1	1
A02	ExploitAmplificationAttack		INH	IF *IDSSensors* exist – 0, ELSE – 1	1
A03	ExploitProtocol		INH	IF *IDSSensors* exist – 0, ELSE – 1	1
A04	ExploitMalformedPacket		INH	IF *IDSSensors* exist – 0, ELSE – 1	1
A05	ForgeMACAddress		INH	IF *IDSSensors* exist – 0, ELSE – 1	1
A06	AbuseMemoryAddress		INH	IF *IDSSensors* exist – 0, ELSE – 1	1
A07	AbuseStackHardening		INH	IF *IDSSensors* exist – 0, ELSE – 1	1
A08	AbusePoisonedCookies		INH	IF *IDSSensors* exist – 0, ELSE – 1	1
A09	ExploitSharedNetwork		INH	IF *IDSSensors* exist – 0, ELSE – 1	1

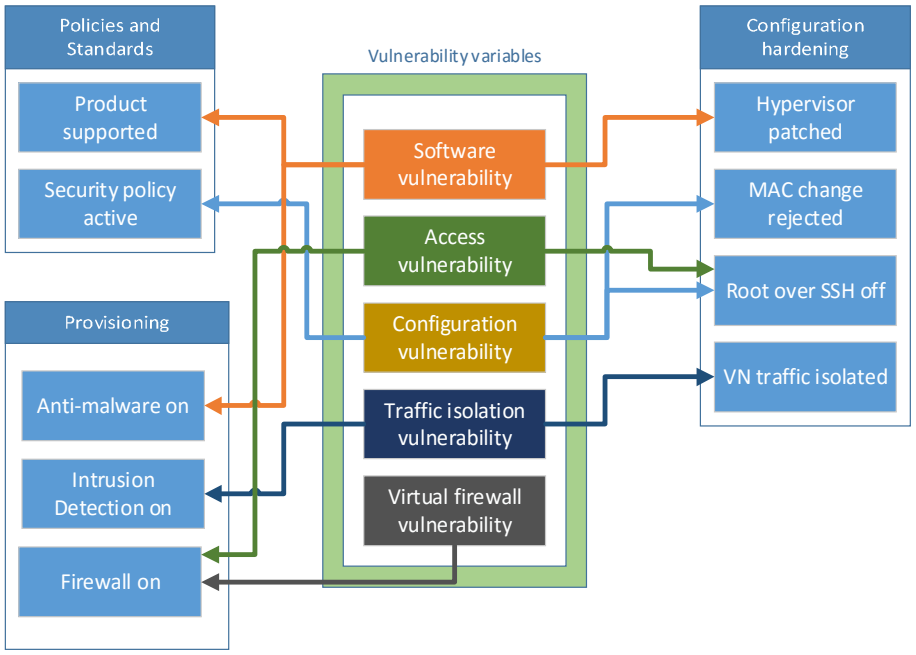


Fig. 3.7. Links between the analyzed risk assessment variables and Hypervisor meta-class attributes (source: author)

3.4. Improvements of the experimental setup

During this study, areas for additional improvement have been noticed. Such improvements would allow implementing the proposed method and its' practical usage easier and more fluent to current workflow of an organization. Such improvement, therefore, have been carried out and proposed in this thesis as well.

3.4.1. Model transformation solution for improvement of the accessibility

The process of this study has revealed the benefits of the Enterprise Architecture Modeling and the CySeMoL realization that is specifically designed to evaluate the information security risk. However, when it comes to using the CySeMoL for the information security related needs of a small or medium enterprise, the problem of accessibility arises.

While most of the sophisticated small and medium enterprises have their ICT infrastructure including the hardware and the software side well documented, the specific knowledge, required to represent the same layout in the form of CySeMoL model requires either additional training or assistance from the side. However, it was presumed, that most of the data, required to generate a CySeMoL model can be acquired from the attributes of network and service level diagrams of looser form. The presumption lead to an additional research and a newly proposed method for model transformation, as well as a tool, providing this functionality.

This method provides a model transformation from an OPC/XML format stored diagram to the CySeMoL model. OPC is a container file, storing a combination of files within organized into directories. This standardized format (ISO 29500-2: 2012) is a well-documented solution enabling easy implementation and versatility of usage fields. A format, based on OPC for storing graphical notation has an extension of .vsdx. The structure of this file is a hierarchy of separate files and directories, placed in a container to ensure that extracting the required information is a well-structured process.

The main tags of XML used in OPC/XML files include:

- Shapes – describes a shape array;
- Shape – describes a shape and its' identification number, name, type and master template;
- Cell – it is a versatile tag, containing information about name and value of many properties of cells under Shape and Section tags;
- Text – gives text output, most commonly an object of instance, visible graphically;
- Section – contains attribute information under it;
- Row – stores attribute information;
- Connects – describes array of connections;
- Connect – defines a connector between instances, specifying sheets, cells and parts connected.

The proposed approach is based on text analysis, structure comparison and relationship identification as presented in Fig. 3.8. It uses CySeMoL concept database to describe all possible components, its' attributes, values and relationships between concepts. While there is no specific metamodel of OPC/XML drawing file, only abstract elements described by textual data and connections between these elements exist. Structure of such metamodel does not allow direct transformation, as static relationships between two metamodels cannot be used.

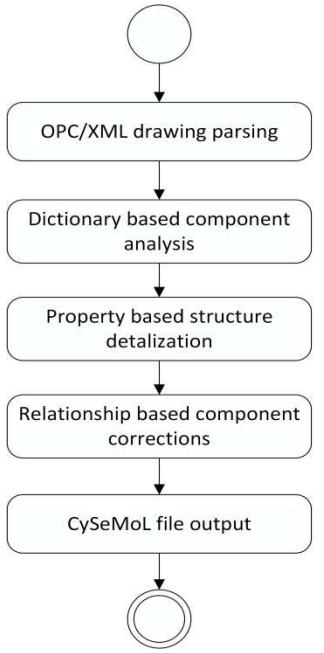


Fig. 3.8. Basic principle of the source model transformation to the target CySeMoL model (source: author)

The parsing of the source model extracts specific data, related to model visual presentation. The rest information, irrelevant for the transformation, can be disregarded. Therefore the extracted data only contains information about objects and relationships between them, disregarding the layout, theme and other supplemental information. Moreover, additional information describing an object is gathered by extracting meta-information including shape name, description and defined attributes and their values. This data is obtained from .vsdx file package *pages*. All additional attributes can be obtained from .vsdx file package *masters*. While the parsing does not require deep analysis, transformation of extracted data is more complex. OPC/XML drawing file often is more abstract comparing to CySeMoL model and additional information has to be appended automatically.

To make links between source and target models this approach proposes usage of CySeMoL concept dictionary with supplemented multiple synonyms for each concept. The dictionary for source model analysis includes all elements, specified as text based descriptions. This is very content dependent specification of elements with a lot of ambiguity, therefore a list of synonyms is used for each CySeMoL element to make sure that similar elements from source model are identified correctly. Additional context analysis is performed to define links between different level target model elements. This is done by using synonyms for

source model shape attributes (name and value). If the source model is well described by choosing appropriate diagrams and defining associated property values, the data can be used to detail each shape and its content.

To make the transformation even more detailed, example CySeMoL objects can be stored as templates. This enables creation of detailed CySeMoL objects rather than empty objects, requiring additional specification later. According to this data 3 possible results might occur in this transformation phase:

- Fully described CySeMoL object – appropriate CySeMoL object example is stored in the knowledge database and the name of the example has synonyms in source model or enough property values of the example object matches source model property values. The margins of matching property can be configured to get different transformation precision.
- Partially described CySeMoL object – appropriate CySeMoL object example is not stored in the knowledge database but some CySeMoL attribute values are matched to source model attribute values.
- Empty CySeMoL object – relevant CySeMoL object example is not stored in the knowledge database and no attribute values of source model match CySeMoL attribute values.

A specific software tool, based on this method has been developed and tested. The results of the performance of the tool are presented in Table 3.8.

Table 3.8. Summary of transformation accuracy between the OPC/XML and CySeMoL,%

Property	SMEs network			Web server			Total
	Components and links only	Component name added	Component properties added	Components and links only	Component name added	Component properties added	
Correctly identified element percentage	95	100	100	96	100	100	98
Fully described	87	93	97	73	93	92	88
Partially described	0	0	0	0	0	0	0
Empty	8	7	3	22	7	8	10
Correctly identified object chain percentage	81	98	100	64	98	97	87
Total	88	99	100	79	99	98	94

Text, dictionary based analysis is used for element identification, however further reasoning is required for definition of the element from source model relation to destination metamodel. The combination of dictionary association, structure comparison and relationship similarities showed a 94% accuracy in this model transformation. This requires a detailed list of attributes in order to increase the model transformation accuracy leading to a detailed situation description, useful for model transformation as well as EA formalization.

The proposed method simplifies SMEs cybersecurity assessment analysis and allows existing resource usage for acquiring new models as well as providing additional knowledge. Though the method was tested in security area for CySeMoL model design, the approach does not limit its' applicability, therefore it can be adjusted for a broader range of model transformations.

3.4.2. The Service Level Agreement Meta-Class

One of the main and most beneficial usage of virtualization technologies is the Cloud Computing service of many varieties. If the Cloud service is local, the regulatory basis of the service provider matches the one of the customer, as both belong to the same entity. However, when using Public Cloud services, the regulatory basis of the service provider may not only differ due to the organizational differences, but due to geographical, political and technical factors as well. The leveraging of these factors is a crucial factor when assessing the information security risk, thus it should be taken into account. However, the CySeMoL does not have such functionality implemented.

Implementing the Service Level Agreement class allows (but not obliges) disregarding the security measures of the service provider when designing the layout. This means that the SLA class overrides the service provider security measures by ensuring, that these measures are described and agreed upon in the contract, as presented in Fig. 3.9.

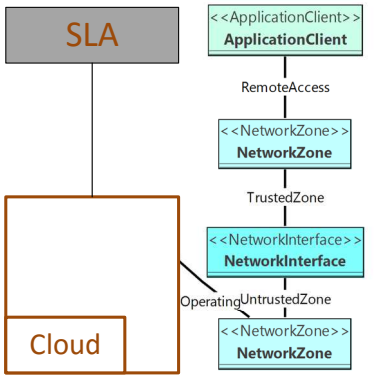


Fig. 3.9. Graphical representation of the newly proposed meta-concept of SLA and its' relation to the model (source: author)

Based on this new concept, the customer is not required to know the technical realization of the security measures of software provider, rather relying on the measures defined in the contract.

3.5. Conclusions of the Third Chapter

During the experimental phase of the research the following conclusions have been reached:

1. It was determined that the CVE database entry based score, used as a metric of the risk of exploiting the vulnerabilities is: *SV*=65.60% (±14.37), *AC*=82.00% (±11.00), *CF*=66.00% (±13.16), *TI*=71.00% (±9.89), *VF*=85.00% (±8.78).
2. It was determined that the expert evaluation score, used as a metric of the risk of exploiting the vulnerabilities is: *SVe*= 59.81% (±4.02), *ACe*=70.42% (±3.70), *CFe*=65.00% (±2.91), *TIe*=65.42% (±3.46), *VFe*=60.90% (±9.40).
3. Statistical analysis of expert evaluation using the classical Cooke's method has revealed that 7 of the 31 expert respondents showed significant knowledge in the field, therefore taken into account.
4. The comparison of the variables acquired from CVE and from expert knowledge has concluded that there is no significant difference between the two scores, with significance parameter χ^2 in the range of [1.436; 6.523].

General Conclusions

1. The overview of existing literature has revealed the need for a thorough security threat categorization for virtualized systems. Although many approaches exist, they do not provide a consistent landscape for security analysis reference. The analyzed sources allow for a taxonomical categorization to be developed that is required for the newly developed Cyber Security Modeling Language Hypervisor class.
2. The proposed generalized security threat categorization taxonomy for virtualized systems consists of 34 threats within 8 categories. It can serve as the basis for the categorization of control means.
3. The security of a virtualized system depends on the five parameters describing the overall health of the system: software vulnerabilities, access control, secure configuration, traffic isolation and virtual firewalls.
4. Statistical values of the parameters are defined based on the exploitability subscore of the CVSS scoring database.
5. Statistical values of the parameters are validated using the results of expert evaluation.
6. Implementation of the method to automated risk analysis solutions requires for the situation, including the software and hardware architec-

ture and configuration, to be described in a strictly described model. To improve the accessibility to the proposed method, a model transformation method has been developed.

References

2nd Watch. 2015. *AWS scorecard*. January – March 2015. 2nd Watch.

800-53A:2014. NIST Special Publication 800-53A Revision 4. National Institute of Standards and Technology, Gaithersburg, 2014.

Ackermann, T. 2012. *IT security risk management: perceived IT security risks in the context of cloud computing*. Darmstadt: Springer Gabler.

Ayyub, B. M. 2001. *Elicitation of expert opinions for uncertainty and risks*. Boca Raton: CRC Press. http://dx.doi.org/10.1201/9781420040906

Alberts, C. J.; Dorofee, A. 2002. *Managing information security risks: the OCTAVE approach*. Boston: Addison-Wesley Longman Publishing Co., Inc.

Alia, M.; Khan, S. U.; Vasilakos, A. V. 2015. Security in cloud computing: opportunities and challenges, *Information Sciences* 305(1): 357–383. http://dx.doi.org/10.1016/j.ins.2015.01.025

Allodi, L.; Massacci, F. 2012. A preliminary analysis of vulnerability scores for attacks in wild: the EKITS and SYM datasets, in *2012 ACM Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, 16–18 October 2012, Raleigh, NC, USA, 17–24. http://dx.doi.org/10.1145/2382416.2382427

Allodi, L.; Massacci, F. 2014. Comparing vulnerability severity and exploits using case-control studies, *ACM Transactions on Information and System Security (TISSEC)* 17(1): 1–20. http://dx.doi.org/10.1145/2630069

Andrade, E. C.; Alves, M.; Matos, R.; Silva, B.; Maciel, P. 2013. OpenMADS: an open source tool for modeling and analysis of distributed systems, in *Computer safety, reliability, and security*. Lecture notes in computer science, vol. 8153. Heidelberg: Springer, 277–284. http://dx.doi.org/10.1007/978-3-642-40793-2_25

Anscombe, F. J.; Glynn, W. J. 1983. Distribution of the kurtosis statistic b2 for normal statistics, *Biometrika* 70(1): 227–234.

ANSI. 2010. Guidelines for the construction, format, and management of monolingual controlled vocabluaries. National Information Standards Organization, Baltimore.

Artz, M. L. 2002. *NetSPA: a network security planning architecture*. Massachussetts Institute of Technology, Cambridge.

Aven, T. 2008. Risk analysis. Assessing uncertainties beyond expected values and probabilities. Chichester: John Wiley & Sons Ltd.

Bartlett, J. E.; Kotrlik, J. W.; Higgins, C. C. 2001. Organizational research: determining appropriate sample size in survey research, *Information Technology, Learning, and Performance Journal* 19(1): 43–50.

Beckers, K.; Heisel, M.; Solhaug, B.; Stølen, K. 2014. ISMS-CORAS: a structured method for establishing an ISO 27001 compliant information security management system, in *Engineering secure future internet services and systems*. Lecture notes in computer science, vol. 8431. Heidelberg: Springer, 315–344. http://dx.doi.org/10.1007/978-3-319-07452-8_13

Bitglass. 2015. Cloud security spotlight report. Bitglass.

Bond, J. 2015. The enterprise cloud: best practices for transforming legacy IT. Sebastopol: O'Reilly Media Inc.

Borza, A.; Duesterhaus, D.; Grabczynski, C.; Johnson, J.; Kelly, R.; Miller, T. 2004. *Cisco IOS switch security configuration guide.* Report Number: I33-010R-2004. National Security Agency, Fort Meade.

Bozorgi, M.; Saul, L. K.; Savage, S.; Voelker, G. M. 2010. Beyond heuristics: learning to classify vulnerabilities and predict exploits, in *16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 25–28 July 2010, Washington DC, USA, 105–114. http://dx.doi.org/10.1145/1835804.1835821

BSI. 2008. *Risk analysis based on IT-Grundchutz*. Standard. Bundesamt für Sicherheit in der Informationstechnik (BSI), Bonn.

Bundeskanzleramt. 2013. *Österreichisches IT – Sicherheitshandbuch.* Austrian IT Security Handbook v4. Bundeskanzleramt (Austrian federal chancellery), Vienna.

Carroll, M. 2012. *A risk and control framework for cloud computing and virtualization*: Doctoral thesis. University of South Africa, Pretoria.

Catteddu, D.; Hogben, G. 2009. *Cloud computing: benefits, risks and recommendations for information security.* European Network and Information Security Agency, Heraklion.

Chandramouli, R. 2014. *NIST special publication 800-125 – A security recommendations for hypervisor deployment*. National Institute of Standards and Technology, Gaithersburg.

Chinosi, M.; Trombetta, A. 2012. BPMN: an introduction to the standard, *Computer Standards & Interfaces* 34: 124–134. http://dx.doi.org/10.1016/j.csi.2011.06.002

Cisco Systems, Inc. 2006. *Cisco unified CME solution reference network design guide*. Cisco Systems, Inc., San Jose.

Clemen, R. T.; Winkler, R. L. 1999. Combining probability distributions from experts in risk analysis, *Risk Analysis* 19(187): 187–204. http://dx.doi.org/10.1111/j.1539-6924.1999.tb00399.x

Cloud Security Alliance. 2014. *Cloud Control Matrix V3.0.1.* Cloud Security Alliance.
CLUSIF. 1998. *Marion Version 98*. Manual. CLUSIF, Paris.

CLUSIF. 2011. *Mehari 2010: reference manual of Mehari knowledge base*. Reference Manual. Clusif, PARIS.

Context Information Security, Ltd. 2011. *Assessing cloud node security*. Context Information Security, Ltd., London.

Cooke, R. 1991. *Experts in uncertainty: opinion and subjective probability in science*. Oxford: Oxford University Press.

Cooke, R. 2008. TU Delft expert judgement data base, *Reliability Engineering & System Safety* 93(5): 657–674. http://dx.doi.org/10.1016/j.ress.2007.03.005

Corder, G. W.; Foreman, D. I. 2009. *Nonparametric statistics for non-statisticians: a step-by-step approach*. New Jersey: Wiley. http://dx.doi.org/10.1002/9781118165881

Coty, S.; Snyder, P.; Stevens, K. 2014. *Alert Logic cloud security report spring 2014*. Houston: Alert Logic.

Creswell, J. W. 2003. Research design. Qualitative, quantitative and mixed methods approaches. 2nd ed. Lincoln: SAGE Publications.

FIRST. 2014. *CVSS v3.0 Specification (v1.7).* Technical Specification. FIRST, Morrisville, 21.

D'Agostino, R. B. 1970. Transformation to normality of the null distribution of g1, *Biometrika* 57(3): 679–681. http://dx.doi.org/10.1093/biomet/57.3.679

den Braber, F.; Braendeland, G.; Dahl, H. I.; Engan, I.; Hogganvik, I.; Lund, M. S.; Solhaug, B.; Stolen, K.; Vraalsen, F. 2006. *The CORAS model-based method for security risk analysis*. Handbook. SINTEF ICT, Oslo.

Devore, J. L. 2012. *Probability & statistics for engineering and the sciences*. 8th ed. Boston: Brooks/Cole.

Dutch Ministry of Internal Affairs. 1996. *Handleiding Afhankelijkheids- en Kwetsbaarheidsanalyse: stappenplan voor de uitvoering van een A&K-analyse v1.01.* Dutch Ministry of Internal Affairs, Hague.

EC-Council. 2011. *Virtualization security*. EC-Council Press.

European Commision. 2014. *Cloud service level agreement standardisation guidelines*. European Commission, The Cloud Select Industry Group, Subgroup on Service Level Agreements, Brussels.

Fenz, S.; Heurix, J.; Neubauer, T.; Pechstein, F. 2014. Current challenges in information security risk management, *Information Management & Computer Security* 22(5): 410–430. http://dx.doi.org/10.1108/IMCS-07-2013-0053

FIRST. 2014. *Common Vulnerability Scoring System v3.0*: specification document. FIRST.

Frei, S.; May, M.; Fiedler, U.; Plattner, B. 2006. Large-scale vulnerability analysis, in *SIGCOMM Workshop on Large-Scale Attack Defense*, 11–15 September 2006, Pisa, Italy, 131–138. http://dx.doi.org/10.1145/1162666.1162671

French National Information Systems Security Agency. 2011. *EBIOS detailed design documentation* (Logiciel EBIOS document de conception detaillee). Specification. French National Information Systems Security Agency, Paris.

Friedenthal, S.; Moore, A.; Steiner, R. 2014. *A practical guide to SysML*. Waltham: Elsevier.

Fuller, W. A. 1991. Simple estimators for the mean of skewed populations, *Statistica Sinica* 1(1): 137–158.

Garber, L. 2012. The challenges of securing the virtualized environment, *Computer* 1: 17–20. http://dx.doi.org/10.1109/mc.2012.27

Garfinkel, T. S. 2010. *Paradigms for virtualization based host security*: doctoral thesis. Stanford University, Stanford.

Garfinkel, T.; Warfield, A. 2007. What virtualization can do for security, *Login Journal* 32(6): 28–34.

Ghasemi, A.; Zahediasl, S. 2012. Normality tests for statistical analysis: a guide for non-statisticians, *International Journal of Endocrinology and Metabolism* 10(2): 486. http://dx.doi.org/10.5812/ijem.3505

Goranin, N.; Mažeika, D. 2011. *Nusikaltimai elektroninėje erdvėje ir jų tyrimo metodikos* [Cyber-crimes and their investigation methodologies]. Vilnius: TEV (in Lithuanian). http://dx.doi.org/10.5755/e01.9786094330544

Gruschka, N.; Jensen, M. 2010. Attack surfaces: a taxonomy for attacks on cloud services, in *IEEE 3rd International Conference on Cloud Computing*, 5–10 July 2010, Miami, FL, USA.

Gupta, S.; Kumar, P. 2013. Taxonomy of cloud security, International Journal of Computer Science, Engineering and Applications 3(5): 47–67.

Haimes, Y. Y. 2004. *Risk modeling, assessment and management*. Hoboken: John Wiley & Sons. http://dx.doi.org/10.1002/0471723908

Hashemi, S. M.; Ardakani, M. R. 2012. Taxonomy of the security aspects of cloud computing systems – a survey, *International Journal of Applied Information Systems* 4(1): 21–28. http://dx.doi.org/10.5120/ijais12-450611

Hau, W.; Araujo, R. 2007. Virtualization and risk – key security considerations for your enterprise architecture. Whitepaper. Foundstone, a division of McAfee, Santa Clara.

Hietala, J. 2009. *Top virtualization security mistakes*. Whitepaper. SANS Institute, Frederickburg.

Holm, H., M. Ekstedt, T. Sommestad, ir M. Korman. 2013. *A Manual for the Cyber Security Modeling Language*. Stockholm: Department of Industrial Information and Control Systems, Royal Institute of Technology.

Holm, H.; Sommestad, T.; Ekstedt, M.; Honeth, N. 2014. Indicators of expert judgement and their significance: an empirical investigation in the area of cyber security, *Expert Systems* 31(4): 299–318. http://dx.doi.org/10.1111/exsy.12039

Information Security Forum. 2014. *IRAM2: managing information risk is a business essential*. Specification. Information Security Forum, Walton-on-Thames.

International Telecommunication Union. 2012. *Series X: data networks, open system communications and security. Cybersecurity information exchange – Vulnerability/state exchange Common vulnerabilities and exposures*. Standard. Telecommunication Standartization Sector of ITU, Geneva, 2012.

ISACA. 2013. *Cobit 5 for risk*. ISACA, Rolling Meadows.

ISO 27000:2014. *Information technology – Security techniques – Information security management systems – Overview and vocabulary*. International Organization for Standartization, Geneva, 2014.

ISO 27005:2011. *Information technology – Security techniques – Information security risk management*. International Organization for Standartization, Geneva, 2011.

ISO 29500-2:2012. *Information technology – Document description and processing languages – Office open XML file formats – Part 2: Open packaging conventions*. 3rd ed. International Organization for Standartization, Geneva, 2012.

ISO/IEC CD 19086-1:2015. *Information technology. Cloud computing: Service level agreement (SLA) framework and Technology. Part 1: Overview and concepts.* International Organization for Standardization, Geneva, 2015.

Yazar, Z. 2002. *A qualitative risk analysis and management tool-CRAMM*. White paper. SANS Institute, Swansea.

Joh, H.; Malaiya, Y. K. 2011. Defining and assessing quantitative security risk measures using vulnerability lifecycle and CVSS metrics, in *The 2011 International Conference on Security and Management (SAM)*, 18–21 July 2011, Las Vegas, Nevada, USA. Worldcomp, 10–16.

Johnson, P.; Ullberg, J.; Buschle, M. 2013. P2AMF: predictive, probabilistic architecture modeling framework, in *Enterprise interoperability.* Lecture notes in business information processing, vol. 144. Berlin Heidelberg: Springer, 104–117. http://dx.doi.org/10.1007/978-3-642-36796-0_10

Jones, J. A. 2011. *An introduction to Factor Analysis of Information Risk (FAIR)*. Specification.: RMI (now RiskLens), Spokane.

Kai, L. 2012. Research on security and risk of server virtualization, *Journal of Tongren University* 4: 39.

Kajackas, A.; Rainys, R. 2011. Estimation of critical components of internet infrastructure, *Elektronika ir Elektrotechnika* 110(4): 35–28. http://dx.doi.org/10.5755/j01.eee.110.4.282

Kavis, M. J. 2014. *Architecting the cloud: design decisions for cloud computing service models (SaaS, PaaS, IaaS)*. Hoboken: John Wiley & Sons. http://dx.doi.org/10.1002/9781118691779

Kissel, R. 2013. *NISTIR 7298 revision 2. Glossary of key information security terms*. National Institute of Standards and Technology, Gaithersburg.

Kyriazis, D. 2013. *Cloud computing service level agreements. Exploitation of Research Results*. European Commission Directorate General Communications Networks, Content And Technology Unit E2 – Software And Services, Cloud. Brussels.

Kordy, B.; Mauw, S.; Radomirović, S.; Schweitzer, P. 2012. Attack–defense trees, *Journal of Logic and Computation*, 1–33. http://dx.doi.org/10.1093/logcom/exs029

Kortchinsky, K. 2015. *Escaping VMware Workstation through COM1*. Google Security Team Report.

Lee, M. C. 2014. Information security risk analysis methods and research trends: AHP and fuzzy comprehensive method, *International Journal of Computer Science & Information Technology* 6(1): 1–45.

Lemaire, L.; Lapon, J. 2014. A SysML extension for security analysis of industrial control systems, in *2nd International Symposium for ICS & SCADA Cyber Security Research 2014 (ICS-CSR 2014)*, 11–12 September 2014, St Pölten, Austria.

Lopez, F.; Amutio, M. A.; Candau, J.; Manas, J. A. 2005. *MAGERIT Version 2.* Methodology. Ministerio de Administraciones Publicas, Madrid.

Machida, F.; Andrade, E.; Kim, D. S.; Trivedi, K. S. 2011. Candy: component-based availability modeling framework for cloud service management using SysML, in *The 30th IEEE Symposium on Reliable Distributed Systems (SRDS 2011)*, 4–7 October 2011, Madrid, Spain.

Marinescu, D. C. 2013. *Cloud computing – theory and practice*. Amsterdam: Elsevier.

Marinos, L. 2014. *ENISA threat landscape 2014*. Overview of current and emerging cyberthreats. European Union Agency for Network and Information Security, Heraklion.

Mirkovic, J.; Reiher, P. 2004. A taxonomy of DDoS attack and DDoS defense mechanisms, *ACM SIGCOMM Computer Communication Review* 34(2): 39–53. http://dx.doi.org/10.1145/997150.997156

Modi, C.; Patel, D.; Borisaniya, B.; Patel, A.; Rajarajan, M. 2013. A survey on security issues and solutions at different layers of Cloud computing, *The Journal of Supercomputing* 63(2): 561–592. http://dx.doi.org/10.1007/s11227-012-0831-5

National Institute of Standards and Technology (NIST). 2006. *Engineering statistics handbook*. National Institute of Standards and Technology, Gaithersburg.

National Institute of Standards and Technology. (NIST). 2015. *National Vulnerability Database* [online]. US Department of Commerce, National Institute of Standards and Technology [cited 12 August 2015]. Available from Internet: https://nvd.nist.gov

SP 800-125:2011. Guide to security for full virtualization technologies. National Institute of Standards and Technology, Gaithersburg, 2011.

SP 800-145:2011. The NIST definition of cloud computing. National Institute of Standards and Technology, Gaithersburg, 2011.

National Vulnerability Database. 2012. *Vulnerability summary for CVE-2012-0217* [online], [cited 11 August 2015]. Available from Internet: https://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2012-0217

National Vulnerability Database. 2014. *Vulnerability summary for CVE-2014-0983* [online], [cited 11 August 2015]. Available from Internet: https://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2014-0983

Oracle. 2012. *Network isolation in private database clouds*. An Oracle white paper. Oracle.

Ottenheimer, D.; Wallace, M. 2012. *Securing the virtual environment. How to defend the enterprise against attack*. Indianapolis: John Wiley & Sons.

Ou, X.; Govindavajhala, S.; Appel, A. W. 2005. MulVAL: a logic-based network security analyser, in *14th USENIX Security Symposium*, 31 July – 5 August, 2005, Baltimore, USA.

Panjer, H. H. 2006. *Operational risk: modeling analytics*. Hoboken, New Jersey: John Wiley & Sons. http://dx.doi.org/10.1002/0470051310

Pearce, M.; Zeadally, S.; Hunt, R. 2013. Virtualization: issues, security threats, and solutions, *ACM Computing Surveys* 45(2): 1–39. http://dx.doi.org/10.1145/2431211.2431216

Peltier, T. R. 2000. *Facilitated Risk Analysis Process (FRAP)*. Specification. Boca Raton: CRC Press, 21.

Peltier, T. R. 2010. *Information security risk analysis*. 3rd ed. Boca Raton: CRC Press. http://dx.doi.org/10.1201/EBK1439839560

Popek, G. J.; Goldberg, R. P. 1974. Formal requirements for virtualizable third generation architectures, *Communications of the ACM* 17(7): 412–421. http://dx.doi.org/10.1145/361011.361073

Portnoy, M. 2012. *Virtualization essentials*. Indianapolis: John Wiley & Sons, Inc.

Prasad, K.; Munivara, A.; Reddy, R.; Rao, K. 2014. DoS and DDoS attacks: defense, detection and traceback mechanisms – a survey, *Global Journal of Computer Science and Technology* 14(7): 15–32.

Razali, N. M.; Wah, Y. B. 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests, *Journal of Statistical Modeling and Analytics* 2(1): 21–33.

Reuben, J. S. 2007. A survey on virtual machine security, in *T-110.5290 Seminar on Network Security*, 11–12 October 2007, Helsinki, Finland, 2(36).

Ryan, J. J.; Mazzuchi, T. A.; Ryan, D. J.; De La Cruz, J. L.; Cooke, R. 2012. Quantifying information security risks using expert judgment elicitation, *Computers & Operations Research* 39(4): 774–784. http://dx.doi.org/10.1016/j.cor.2010.11.013

Rutkowska, J. 2006. *Introducing Stealth Malware Taxonomy*. COSEINC Advanced Malware Labs, 9.

Sakalauskas, E.; Katvickis, A.; Dosinas, G. 2010. Key agreement protocol over the ring of multivariate polynomials, *Information Technology and Control* 39(1).

Sakalauskas, E.; Tvarijonas, P.; Raulynaitis, A. 2007. Key Agreement Protocol (KAP) using conjugacy and discrete logarithm problems in group representation level, *Informatica* 18(1): 115–124.

Schneider, F. B. 2004. Least privilege and more, in *Computer systems*. New York: Springer, 253–258. http://dx.doi.org/10.1007/0-387-21821-1_38

Schneier, B. 1999. Attack trees, *Dr. Dobb's Journal* 24(12): 21–29.

Shackleford, D. 2010. *A guide to virtualization hardening guides*. Swansea: SANS Institute.

Smiraglia, R. 2014. *The elements of knowledge organization*. Berlin: Springer. http://dx.doi.org/10.1007/978-3-319-09357-4

Smith, R. L. 2002. Measuring risk with extreme value theory, in M. AS. H. Dempster (Ed.). *Risk management: value at risk and beyond.* Cambridge University Press, 224–246.

Sommestad, T. 2012. *A framework and theory for cyber security assessments*: Dissertation. KTH, Royal Institute of Technology, Stockholm.

Sommestad, T.; Holm, H.; Ekstedt, M. 2011. Estimates of success rates of Denial-of-Service attacks, in *2011 IEEE 10th International Conference on Computing and Communications (TrustCom)*. 16–18 November 2011, IEEE, 21–28.

Sommestad, T.; Holm, H.; Ekstedt, M. 2012. Estimates of success rates of remote arbitrary code execution attacks, *Information Management & Computer Security* 20(2): 107–122. http://dx.doi.org/10.1108/09685221211235625

Sommestad, T.; Ekstedt, M.; Holm, H. 2013. The cyber security modeling language: a tool for assessing the vulnerability of enterprise system architectures, *IEEE Systems Journal* 7(3): 363–373. http://dx.doi.org/10.1109/JSYST.2012.2221853

Steiner, T. 2012. *An introduction to securing a cloud environment*. SANS Institute, Swansea, 24.

Stitilis, D.; Pakutinskas, P.; Daupariene, I.; Laurinaitis, M. 2011. Precondition for legal regulation of personal identification in cyberspace, *Jurisprudence* 2(18): 725–738.

Subashini, S.; Kavitha, V. 2011. A survey on security issues in service delivery models of cloud computing, *Journal of Network and Computer Applications* 34(1): 1–11. http://dx.doi.org/10.1016/j.jnca.2010.07.006

Szefer, J. M. 2013. *Architectures for secure cloud computing servers.* Princeton: Princeton University.

Toldinas, J.; Rudzika, D.; Stuikys, V.; Ziberkas, G. 2015. Rootkit detection experiment within a virtual environment, *Elektronika ir Elektrotechnika* 104(8): 63–68.

Vaarala, S. 2006. *Security considerations of commodity x86 virtualization*: licentiate thesis. Helsinki University of Technology, Helsinki.

Vaughan-Nichols, S. J. 2008. Virtualization sparks security concerns, *Computer* 41(8): 13–15. http://dx.doi.org/10.1109/MC.2008.276

Wang, Z. 2012. *Securing virtualization: techniques and applications*. Raleigh: North Carolina State University.

Wong, W. 2005. Platforms strive for virtual security – with partitioning and virtualization, platforms can run multiple environments securely, *Electronic Design* 53(17): 44–52.

List of Publications by the Author on the Topic of the Dissertation

Papers in the Reviewed Scientific Journals

Janulevičius, J.; Ramanauskaitė, S.; Goranin, N.; Čenys, A. 2016. Content based model transformations: solutions to existing issues with application in information security, *International Journal of Computers, Communications & Control (IJCCC)* 11(2): 233–247. Agora: Agora University Editing House. ISSN 1841-9836. Thomson ISI Web of Knowledge, citation Index: 0.746(F) (2014).

Janulevičius, J.; Čenys, A.; Goranin, N. 2016. Extension of CySeMoL for cloud computing information security assessment, *Proceedings of the Romanian Academy, Series A* 17(2): 186–192. Thomson ISI Web of Knowledge, citation Index: 0.746(F) (2014).

Janulevičius, J.; Čenys, A. 2014. Development of a risk assessment model for IT risk self-assessment expert system for SMEs, *International Journal of Computer and Communication Engineering (IJCCE)* 3(4): 306–309. San Bernardino: International Academy Publishing (IAP). ISSN 2010-3743.

Janulevičius, J.; Goranin, N. 2013. Ekspertinės sistemos smulkiojo ir vidutinio verslo rizikos valdymo problemoms spręsti, *Mokslas – Lietuvos ateitis = Science – Future of Lithuania: elektronika ir elektrotechnika = Electronics and electrical engineering* 5(2): 84–87. Vilnius: Technika. ISSN 2029-2341.

Summary in Lithuanian

Įvadas

Problemos formulavimas

Informacijos apsauga yra vienas svarbiausių šiuolaikinės organizacijos iššūkių. Nuolat augantis informacinių technologijų (IT) naudojimas šiuo metu aprėpia veiklas tiek organizacijos viduje, tiek išorėje. Dėl to informacijos apsauga yra itin aktuali užtikrinant sklandų procesų vykdymą. Dėl pastaruosius kelis dešimtmečius augančio IT naudojimo juntamas tyrimų, susijusių su šia tema, suaktyvėjimas. Juose gilinamasi kaip gali būti užtikrinama reikiamo lygio informacijos apsauga. Norint supaprastinti informacijos apsaugos užtikrinimo procesą taip pat yra kuriami automatizuoti įrankiai, diegiantys šių tyrimų rezultatus praktiškai. Šiame darbe terminai apsauga ir sauga laikomi tapačiais.

Informacijos apsaugos apibrėžimas reikalauja turimos infrastruktūros, grėsmių ir pažeidžiamumų išmanymo. Taip pat reikalingas ir rizikos dydžiui įvertinti skirtas metodas, aprėpiantis bendrą informacijos apsaugos situaciją. Remiantis šiais duomenimis sudaromas rizikos valdymo planas. Informacijos apsaugai įtakos turintys veiksniai, nagrinėjami šioje disertacijoje apima IT infrastruktūros komponentus ir ryšius tarp jų.

Informacijos apsaugos situacijos nustatymui įprasta atlikti rizikos analizę visai ar daliai turimos infrastruktūros. Infrastruktūros atvaizdavimo formalizavimui pasitelkiamas verslo architektūros (angl. Enterprise Architecture) modelis. Šis modelis leidžia formaliai aprašyti turimą aparatinę, programinę įrangą, teisinius įsipareigojimus bei verslo procesus. Taip užtikrinamas įvairiapusis organizacijos IT veiklos aprašymas.

Virtualizacija yra palyginti nauja paradigma, sukėlusi revoliuciją IT infrastruktūros architektūros srityje. Aparatinė įranga tapo nebūtina procesams įgyvendinti. Taip pat virtualizacija suteikia galimybę patogiau iškelti savo procesus į trečiųjų šalių teikiamas paslaugas, taip sumažinant kaštus IT eksploatacijai ir įsigijimui.

Virtualizuotų sistemų rizikos analizė yra labai svarbi šiuolaikinės IT informacijos apsaugos sritis. Ji padeda sudaryti tinkamą IT infrastruktūros architektūrą, sumažinant su informacijos apsauga susijusias rizikas. Tačiau dėl šios srities naujumo trūksta metodų ir jų taikymo pavyzdžių informacijos apsaugos rizikos analizei atlikti. Informacijos apsaugos automatizacijai pritaikytas metodas virtualizuotoms sistemoms vertinti yra pagrindinis komponentas užtikrinantis informacijos apsaugą šioje srityje. Toks metodas užtikrintų nešališką situacijos įvertinimą realiu laiku, nedelsiant vertinant pokyčius.

Darbo aktualumas

Virtualizacijos technologijos naudojamos beveik visose organizacijose ir jų naudojimas nuolatos auga. Tipiškiausi tokio naudojimo pavyzdžiai yra debesų kompiuterijos paslaugos – elektroninis paštas, dalinimosi bylomis paslauga ir kt. Dėl to kylančių rizikų ir grėsmių suvokimas yra labai svarbus užtikrinant organizacijos su IT susijusių veiklų sklandą.

Norint pilnai suprasti su virtualizacija susijusias rizikas reikalingas metodas, gebantis vertinti šias rizikas organizacijos infrastruktūros ir procesų kontekste. Tai galima atlikti analizuojant formaliai aprašytą modelį. Infrastruktūros architektūros modeliavimas, įskaitant verslo procesus bei aparatinę ir programinę įrangą gali būti atliekamas formalizuojant verslo architektūrą. Visgi specialių virtualizuotų sistemų konceptų trūkumas griežtai apibrėžtuose verslo architektūros modeliuose apsunkina jų naudojimą šioms sistemoms. Virtualizacija ir jos išteklių dalinimo savybė kuria naujus grėsmių vektorius, kurių tradicinė kompiuterija neturi. Šių vektorių apibrėžimas taip pat yra virtualizacijos informacijos apsaugos rizikos analizės dalis. Dėl galimų architektūrinių sprendimų sudėtingumo bei verslo architektūros sistemų tikimybinės išvesties reikalinga, kad šiais vektoriais apibrėžtos rizikos būtų išreiškiamos kiekybiškai, t. y. turėtų skaitines pažeidžiamumų išnaudojimo reikšmes.

Informacijos apsaugos rizikos analizės procesas yra nuodugniai aprašytas mokslininuose darbuose, tačiau jaučiamas šių darbų, skirtų virtualizacijos technologijoms, trūkumas. Todėl trūksta informacijos kaip tiksliai įvertinti naujas, su virtualizacijos taikymu susijusias grėsmes ir jų rizikas, susijusias su hipervizoriaus ir virtualios tinklo infrastruktūros apsauga (Catteddu, Hogben 2009).

Taip pat pastebėtina, kad metodų, skirtų naudoti informacijos apsaugos automatizavimui virtualizuotoms sistemoms vertinti šiuo metu nėra.

Šioje disertacijoje siekiama sukurti kiekybinį informacijos apsaugos rizikos analizės vertinimo metodą virtualizuotoms sistemoms, pritaikytą naudoti automatizuotuose verslo architektūros analizės įrankiuose.

Tyrimų objektas

Darbo tyrimų objektas – virtualizuotų sistemų informacijos apsaugos rizikos analizė.

Darbo tikslas

Disertacijos tikslas – pasiūlyti virtualizuotų sistemų informacijos apsaugos rizikos analizės metodą.

Darbo uždaviniai

Darbo tikslui pasiekti ir mokslinei problemai spręsti darbe iškelti šie uždaviniai:

1. Atlikti esamų informacijos apsaugos rizikos analizės metodų, skirtų virtualizuotoms sistemoms, analizę.
2. Sudaryti su virtualizacija susijusių grėsmių aprašą, apibrėžiantį jų aprėptį.
3. Sukurti kiekybinį rizikos analizės metodą, pritaikant informacijos apsaugos duomenų bazių informaciją, siekiant pakeisti ir / ar pagrįsti ekspertinio vertinimo rezultatus.
4. Atlikti gautųjų įverčių lyginamąją statistinę analizę, įvertinant gautų rezultatų tikslumą ir nustatyti neapibrėžties dydį.
5. Įvertinti pasiūlyto metodo patobulinimus verslo architektūros analizės sistemoje.

Tyrimų metodika

Darbe taikomi lyginamosios analizės ir literatūros analizės metodai, naudoti siekiant išanalizuoti tyrimo objektą. Informacijos apsaugos rizikos analizės metodai taikomi siekiant sukurti virtualizacijos technologijų grėsmių identifikavimo ir įverčių suteikimo sistemą. Eksperimentinių tyrimų metodai naudojami pagrįsti gautų rezultatų tikslumui.

Darbo mokslinis naujumas

Darbo mokslinis naujumas pagrįstas šiais rezultatais:

1. Sukurtas naujas virtualizacijos informacijos apsaugos rizikos analizės metodas, leidžiantis įvertinti hipervizoriui kylančias grėsmes ir jų įvykimo tikimybes.
2. Informacija apie virtualiztuotų sistemų informacijos apsaugos rizikos analizę ir jos sudedamąsias dalis yra labai ribota. Šiai problemai spręsti siūlomas metodas, apibrėžiantis dalykinę sritį, grėsmių susiejimą su kontrolės priemonėmis, saugaus virtualizacijos komponentų naudojimo reikalavimus, atakų scenarijus ir kintamuosius, turinčius įtakos šių komponentų saugumui.
3. Pasiūlytas ir ištirtas naujas kiekybinis duomenų apdorojimo metodas, naudojantis aktualias išnaudojamumo įverčių reikšmes kintamųjų tikimybinėms reikšmėms aprašyti.
4. Pasiūlytas naujas įverčių tikslumo įvertinimo metodas, naudojantis pritaikytas statistines procedūras ekspertų išrankai. Šios procedūros rezultatai suteikia kalibravimo ir entropijos įverčius informacijai su neapibrėžtimi.

Darbo rezultatų praktinė reikšme

Darbe sukurto metodo praktinis įgyvendinimas leidžia analizuoti organizacijos infrastruktūros informacijos apsaugos riziką, esant virtualizuotoms sistemoms. Šio metodo taikymas verslo architektūros analizės įrankio (angl. *Enterprise Architecture Analysis Tool*) aplinkoje užtikrina rizikos analizės automatizavimą, pokyčių vertinimo efektyvumą bei architektūros apsaugos užtikrinimą projektavimo etape.

Ginamieji teiginiai

1. Siūlomas metodas leidžia informacijos apsaugos rizikos tikimybes įvertinti naudojant apdorotus pažeidžiamumo išnaudojamumo įverčius, gautus iš Bendrųjų pažeidžiamumų ir išnaudojamumų duomenų bazės.
2. Siūlomas metodas gali būti naudojamas ekspertinės informacijos apie virtualizuotų sistemų informacijos apsaugos riziką papildymui.

Darbo rezultatų aprobavimas

Disertacijos tema yra paskelbti keturi moksliniai straipsniai. Du iš jų yra publikuoti recenzuojamuose mokslo žurnaluose, kurie yra įtraukti į *Thomson Reuters ISI Web of Science* duomenų bazę ir turi citavimo indeksą.

Disertacijos rezultatai buvo aprobuoti 3 tarptautinėse konferencijose:

- 12th IMEKO TC10 Workshop on Technical Diagnostics: New Perspectives in Measurements, Tools and Techniques for Industrial Applications. June 6–7, 2013, Florence, Italy.
- The 2014 International Conference on Information and Network Security (ICINS 2014). April 11–12, 2014, Jeju, South Korea.
- The XXI IMEKO World Congress. August 29–September 5, Prague, Czech Republic.

Disertacijos struktūra

Disertaciją sudaro įvadas, trys pagrindiniai skyriai, bendrosios išvados, literatūros šaltinių sąrašas, autoriaus publikacijų disertacijos tema sąrašas, santrauka lietuvių kalba. Darbo apimtis – 112 puslapių neskaitant priedų, tekste yra 20 paveikslų ir 18 lentelių. Rašant disertaciją buvo panaudota 127 literatūros šaltiniai.

1. Informacijos apsaugos užtikrinimo virtualizuotoms sistemoms metodai

Skyriuje apžvelgiami esami informacijos apsaugos virtualizuotoms sistemoms tyrimai bei gilinamasi į teorinius virtualizacijos aspektus. Taip pat šiame skyriuje apibrėžiama informacijos apsaugos rizikos analizės sąvoka bei jos komponentai.

Pastaruoju metu stebimas spartus informacinių technologijų (IT) taikymo augimas įvairiose organizacijų veiklos srityse. Viena sparčiausiai populiarėjančių IT tai-

kymo formų – virtualizacija – yra ne tik finansiškai efektyvus sprendimas, bet, su sąlyga, kad įgyvendinta tinkamai, efektyvus ir apsaugos požiūriu. Svarbu paminėti, kad virtualizacijos apsaugos sritis yra platesnė už tradicinės kompiuterijos. Papildomi apsaugos aspektai šiuo atveju yra susiję su fizinių išteklių dalijimosi problema. Kadangi virtualizacijos apsauga pradėta domėtis palyginti neseniai – metodai apsaugos užtikrinimui šiuo metu tik vystomi.

Virtualizacija ir su ja susijusi debesų kompiuterijos paslauga – itin patogus IT resursų dalijimosi būdas, leidžiantis turėti tiek resursų, kiek yra reikalinga. Debesų kompiuterija pasižymi savitarnos galimybėmis, plačia tinklo aprėptimi, resursų apjungimu, elastingumu bei apmokėjimu tik už suteiktas pamatuotas paslaugas. Debesų kompiuterijos paslaugos, priklausomai nuo kontrolės lygmens, suteikiamo paslaugos gavėjui, skaidomos į: programinė įranga kaip paslauga; platforma kaip paslauga; bei infrastruktūra kaip paslauga. Pagal įdiegimo modelius debesų kompiuterija gali būti: privati; vieša; hibridinė; bendruomeninė.

Dėl virtualizacijos taikymo kyla papildomos informacijos apsaugos grėsmės, susijusios su papildomo hipervizoriaus komponento atsiradimu, tinklų virtualizacija ir atsiejimu nuo konkrečių fizinių įrenginių. Virtualiųjų tinklų ir jais perduodamų duomenų srautų apsauga yra sudėtingesnė nei fizinių tinklų. Konkrečių fizinių įrenginių svarbos sumažėjimas dėl orientavimosi į bendrą išteklių sąvadą sudėtingiau užtikrina šių resursų geografinį išsidėstymą. Visgi, virtualiųjų infrastruktūros komponentų naudojimas taip pat palengvina pasiekiamumo užtikrinimą, dėl galimybės nesudėtingai dubliuoti įrenginius ir srautus.

Informacijos apsaugą suvokiame kaip trijų kriterijų užtikrinimą. Šie kriterijai yra konfidencialumas, vientisumas ir prieinamumas (Kissel 2013). Rizikos analizė taip pat atliekama vertinant išteklių svarbą šiems kriterijams atitikti. Rizikos analizę apibrėžiame kaip procesą, skirtą įvertinti žalingų scenarijų įvykimą, jų dažnį bei poveikį. Rizikos analizė gali būti supaprastinta, standartinė arba grindžiama modeliavimu. Priklausomai nuo šios kategorijos skiriasi rizikos analizės išvesties rezultatas. Supaprastinta rizikos analizė neturi formalizuoto pavidalo ir atliekama kokybiniu principu. Standartinė rizikos analizė naudoja formalizuotą struktūrą, o jos rezultatas yra rizikų matricos. Modeliavimu grindžiamos analizės rezultatas pateikia sprendinių medį, todėl galimas įvertinimo atsekamumas. Rizikos analizės kokybei apibrėžti taikomi šie kriterijai (Haimes 2004):

- Aprėpiamumas – rizikos analizės metu tiriamų veiksnių aprėptis turi užtikrinti, kad įvertinami visi galimi poveikį turintys veiksniai;
- Įrodomumas – kiekvienas vertinamas veiksnys turi būti pagrįstas įrodymais;
- Logiškumas – turi būti atsižvelgiama (net jei įrodymai tam prieštarauja) į galimas logines pasekmes;
- Praktiškumas – rizikos analizės rezultatas turi būti praktiškai pritaikomas diegiant priemonių, skirtų rizikos mažinimui, planą;
- Atsekamumas – visi rizikos analizės procesai turi būti dokumentuojami, pagrįsti ir lengvai atsekami;
- Pagrįstumas – tik gerai pagrįsti kriterijai turi būti vertinami;

- Suderinamumas – rizikos analizė turi būti suderinama su galiojančiais potvarkiais, standartais ar teisės aktais;
- Mokymasis – rizikos analizė turi sudaryti pagrindą tobulėjimui ir žalingų veiksnių šalinimui;
- Formalumas – rizikos analizė turėtų laikytis protokolo, kad būtų suderinama su kitų rizikos analizių rezultatais;
- Inovatyvumas – rizikos analizė turi prisitaikyti prie situacijos pokyčių ir vertinti naujai atsirandančias grėsmes.

Disertacijoje apžvelgta šešiolika rizikos vertinimo metodų. Juos apibendrinus pastebėta, kad dauguma jų remiasi kokybiniu vertinimu, t. y. nepateikia skaitinių charakteristikų.

Modeliavimu grindžiamai informacinių technologijų rizikos analizei reikalingas formalus modelis ir metodas, kuriuo remiantis būtų atliekama rizikos analizė. Infrastruktūros ir procesų formalus modelis – verslo architektūra – tai modelis gebantis vaizduoti ir analizuoti objektus (komponentai, procesai) ir ryšius tarp jų. Architektūros modeliavimui naudojami specializuoti įrankiai leidžia ne tik atvaizduoti esamą situaciją, bet ir atlikti jos automatizuotą vertinimą. Atliktos lyginamosios šių modeliavimo aplinkų lyginamąją analizę nustatyta, kad plačiausias savybių rinkinys yra kibernetinio saugumo modeliavimo kalbos (CySeMoL) modelyje, kurio remiamasi tolimesniuose disertacijos skyriuose.

Verslo architektūros modelyje jos komponentai aprašomi klasėmis, atributuose nurodant jų parametrus. Vertinant duomenų apsaugos požiūriu, šie parametrai skirstomi į dvi grupes – atakos žingsnis (attack step) ir apsauga (defense). Pavyzdžiui operacinės sistemos klasė gali turėti tokius atakos žingsnius kaip patekimas, paslaugos sutrikdymas, kritinio pažeidžiamumo aptikimas, patekimas per vartotojo sąsają Apsaugos parametrai gali būti ugniasienės veikimas, saugos papildinių įdiegimas, apsauga nuo kenksmingo programinio kodo ir kt.

2. Rizikos analizės modeliavimas ir taikymas virtualizacijos technologijoms

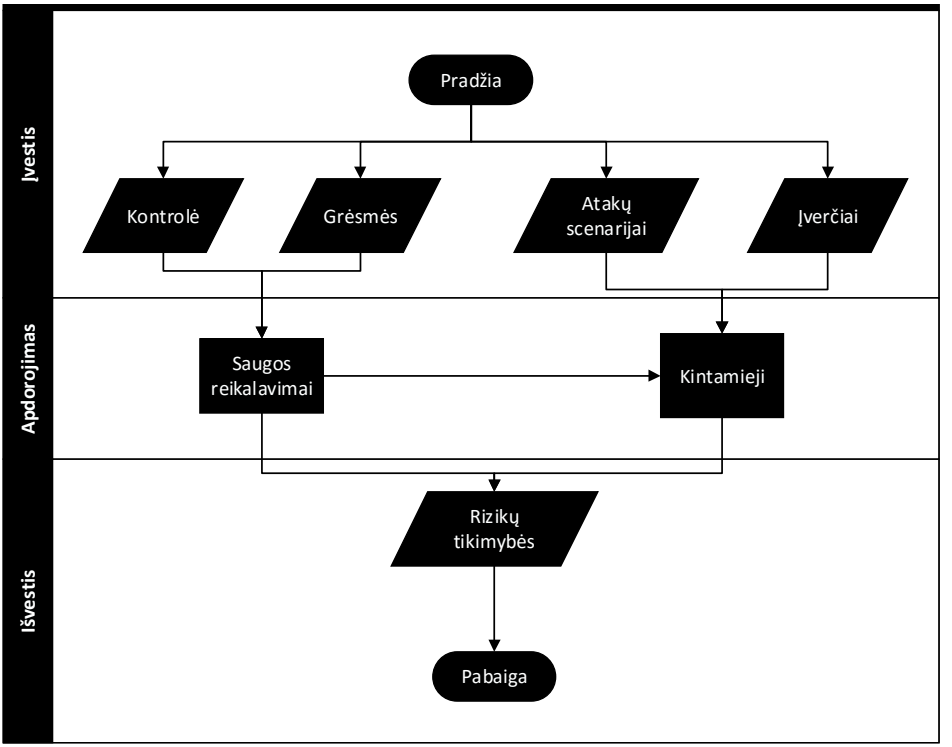
Šiame skyriuje detaliai aprašomas tyrimo metodas, leidžiantis jungti skirtingus metodus tarpusavyje, siekiant išanalizuoti rizikas kylančias virtualizuotoms sistemoms.

Šis metodas tarpusavyje apjungia įvairius metodus. Jis apima:

1. Grėsmių aprėpties nustatymą.
2. Grėsmių susiejimą su kontrolės priemonėmis.
3. Atakų scenarijų apibrėžimą.
4. Kintamųjų, apibrėžiančių rizikos dydį, nustatymą.
5. Skaitinių reikšmių suteikimą kintamiesiems.
6. Rezultatų korektiškumo įvertinimą.

Taip pat šio metodo pritaikomumui automatizuotam modeliavimui papildomai atlikami šie veiksmai: kontrolės priemonės grupuojamos į apsaugos reikalavimų grupes siekiant sumažinti įvesties apimtį; atakų scenarijai išplečiami atakos-gynybos scenarijais

siekiant įvertinti kontrolės priemonių efektyvumą. Metodo procesų aprašomoji diagrama pateikiama S2.1 paveiksle.



S2.1 pav. Aukšto lygio metodo procesų diagrama (blokinė schema)

Kaip matyti iš S2.1 paveikslo, šiame metode yra keturios įvestys ir viena išvestis. Išvestis yra rizikų tikimybių reikšmės. Kiekybinė charakteristika šiuo atveju yra reikalinga detaliam rizikos vertinimui verslo architektūros analizės aplinkoje.

Dalykinės srities apibrėžimui reikalinga sudaryti grėsmių, kylančių virtualizuotoms sistemoms, sąrašą ir jų sąsajas su atskiromis kategorijomis. Esamų kategorizavimo sprendimų analizė parodė, kad jie nėra pakankami, todėl buvo sudaryta nauja virtualizuotoms sistemoms kylančių grėsmių taksonomija, apimanti techninius, teisinius bei komercinius virtualizacijos įgyvendinimo aspektus. Įvertinus kokybinius šios taksonomijos aspektus gauta išvada, kad taksonomija sudaryta kokybiškai.

Gautos grėsmės susietos su kontrolės priemonių matricoje pateikiamomis kontrolės priemonėmis. Šios kontrolės priemonės sugrupuotos į reikalavimus saugiam virtualizacijos įgyvendinimui. Sudaryti trys pagrindiniai reikalavimai: Įdiegtos informacijos apsaugos valdymo politikos, suderinamos su standartais; naudojamas žinomas ir patikrintas hipervizorius; naudojamos tinkamos apsaugos priemonės.

Remiantis kitų autorių tyrimų duomenimis suformuotas penkių pagrindinių kintamųjų, turinčių įtakos rizikos dydžiui, sąrašas. Jis pateikiamas S 2.1 lentelėje.

S 2.1 lentelė. Kintamieji hipervizoriaus rizikos analizei

Kintamasis	Aprašas
SV	*Programinės įrangos pažeidžiamumas,* leidžiantis „peršokti“ iš vienos virtualiosios mašinos į kitą.
AC	*Prieigos pažeidžiamumas,* leidžiantis neteisėtai gauti administratoriaus privilegijų prieigą.
CF	*Saugios konfigūracijos pažeidžiamumas,* apimantis visą riziką bendrai.
TI	*Srautų izoliacijos nebuvimas,* leidžiantis informacijos nutekėjimą.
VF	*Virtualių ugniasienių nebuvimas,* neužtikrinantis apsaugos virtualizuotos aplinkos viduje.

Skaitinėms kintamųjų reikšmėms gauti pasirinkti duomenys iš JAV nacionalinės pažeidžiamumų duomenų bazės, kadangi jie turi plačiausią aprėptį ir didžiausią patikimumą. Duomenų korektiškumui patikrinti sudaryta srities ekspertų imtis, kurios atsakymais yra remiamasi šio tyrimo metu.

Ekspertų kalibravimui pasitelktas ekspertinės informacijos išrankos metodas (Cooke 1991), reikalaujantis sudaryti dviejų tipų klausimus: klausimus su žinomais atsakymais ekspertų kalibravimui; klausimus su nežinomais atsakymais ekspertiniam įvertinimui gauti. Disertacijos tyrimui sudaryti penki klausimai kalibravimui ir vienuolika klausimų kintamųjų įverčiams gauti.

3. Virtualizuotų sistemų rizikos analizės taikymas CySeMoL aplinkoje

Šiame skyriuje taikomi antrojo skyriaus tyrimuose gauti duomenys, siekiant įgyvendinti automatizuotą virtualizuotų sistemų rizikos analizės sistemą. Tam atlikti iš NVD duomenų bazės naudojami CVE tipo įrašai ir jų CVSS įverčiai. Konkrečiai naudojamas antrinis CVSS Išnaudojamumo įvertis. Kiekvienam kintamajam atrinktos naujausi su šio kintamojo grėsmėmis susiję CVE įrašai. Atlikus jų statistinę analizę nustatyti imčių aprašomieji parametrai. Remiantis šiais parametrais nustatyta, kad kintamųjų reikšmių pasiskirstymas neatitinka normaliojo skirstinio, todėl kintamiesiems aprašyti naudojamas centriškai orientuotas matavimo būdas – vindsorizuotas vidurkis, pašalinant 10% ribinių reikšmių (Fuller 1991). Kintamųjų reikšmės pateikiamos S3.1 lentelėje. Svarbu paminėti, kad šie rezultatai pateikiami dešimtbalėje sistemoje, o tikimybinis vertinimas atliekamas dešimtabalėje sistemoje, todėl įverčiai bus dauginami iš dešimties.

S3.1 lentelė. Kintamųjų įverčiai (CVSS)

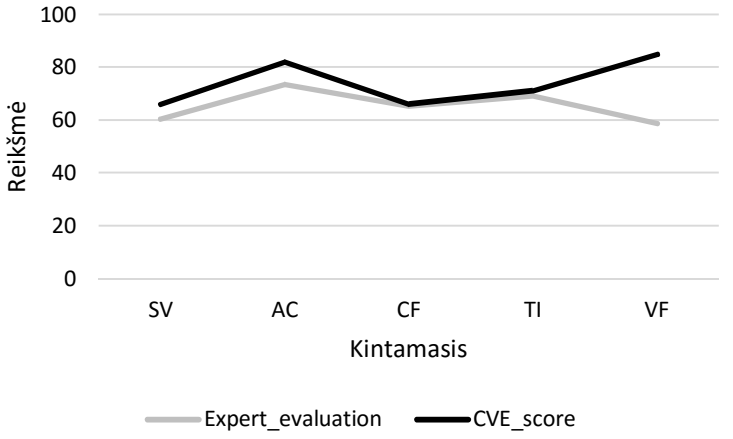
Kintamasis	SV	AC	CF	TI	VF
Viršutinė riba	8,047	9,300	8,081	8,209	9,518
Įvertis	6,560	8,200	6,600	7,100	8,500
Apatinė riba	5,123	7,100	5,399	6,151	7,832
Neapibrėžtis	1,437	1,100	1,316	0,989	0,878

Ekspertų kalibravimo analizė parodė, kad iš 31 apklausto eksperto 7 ekspertų svorinis koeficientas yra didesnis už 0, todėl jų žinios įvertintos kaip pakankamos dalyvauti šiame tyrime. Priklausomai nuo atsakymų koreliavimo su žinomomis teisingomis reikšmėmis šių ekspertų svoriai suteikti skirtingai. Ekspertų gautų rezultatų kintamųjų reikšmės pateikiamos S 3.2 lentelėje.

S3.2 lentelė. Kintamųjų įverčiai (ekspertinis vertinimas)

Kintamasis	SVe	ACe	CFe	TIe	VFe
Viršutinė riba	63,83	74,12	67,91	68,88	65,6
Įvertis	59,81	70,42	65,00	65,42	60,90
Apatinė riba	55,78	66,71	62,09	6196	56,20
Neapibrėžtis	4,02	3,70	2,91	3,46	9,40

Atliekant lyginamąją analizę, pirmoji charakteristika – koreliacinė analizė (S3.1 pav.). Ji parodė, kad koreliacija tarp dydžių yra silpna (r = 0,0849). Tačiau iš grafiko matyti, kad nesutampa tik vieno kintamojo – VF reikšmė. Atmetus šio kintamojo reikšmes koreliacija yra stipri (r = 0,8964).



S3.1 pav. Koreliacinė analizė tarp ekspertinio vertinimo ir duomenų bazių įverčių

Tolimesnei analizei naudojamas Kruskal-Wallis H testas. Tai yra neparametrinis testas naudojamas nustatyti statistiškai reikšmingus skirtumus tarp dviejų ar daugiau imčių, kuomet pasiskirstymas nėra normalus. Šio tyrimo rezultatai (S3.3 lentelė) parodė, kad skirtumai tarp šių imčių nėra reikšmingi, todėl galima teigti, kad galima keisti vieną duomenų šaltinį kitu.

S3.3 lentelė. Skirtumo tarp dviejų šaltinių duomenų reikšmingumo tyrimas

Kintamasis	df	p	χ^2	Reikšmingumo ribos	Išvada
SV; SVe	5	0,05	4,500	11,07	nereikšmingas
AC; ACe	4	0,05	6,523	9,49	nereikšmingas
CF; CFe	2	0,05	1,436	5,99	nereikšmingas
TI; TIe	4	0,05	5,000	9,49	nereikšmingas
VF; VFe	4	0,05	1,768	9,49	nereikšmingas

Remiantis šiais duomenimis pasiūlyta nauja hipervizoriaus meta-klasė CySeMoL aplinkoje, vertinanti virtualizuotų sistemų grėsmes automatiniu būdu.

Bendrosios išvados

1. Literatūros analizė atskleidė išsamaus vieningo virtualiztuotų sistemų apsaugos grėsmių skirstymo į kategorijas poreikį. Egzistuojantys sprendimai neužtikrina vientiso grėsmių apibrėžimo, reikalingo apsaugos analizei atlikti. Išnagrinėti šaltiniai tinkami būti virtualizuotų sistemų apsaugos grėsmių taksonomijos pagrindu, naudojamu naujoje kibernetinės apsaugos modeliavimo kalbos hipervizoriaus klasėje.
2. Sudaryta apibendrinta virtualizuotų sistemų apsaugos grėsmių aprašymo taksonomija. Apibendrinus grėsmes, aprašytas CVE duomenų bazėje ir esamus mokslinius tyrimus nustatyta, kad šiuo metu egzistuoja 34 grėsmės. Nutatyta, kad jos patenka į 8 kategorijas. Ši taksonomija naudotina kaip kontrolės priemonių skirstymo į kategorijas pagrindas.
3. Atlikus virtualizuotų sistemų apsaugai įtaką turinčių veiksnių tyrimą nustatyta, kad ji priklauso nuo penkių parametrų, aprašančių sistemos imunitetą: programinės įrangos pažeidžiamumai, prieigos valdymas, apsaugos konfigūracija, srautų atskirtis ir virtualios ugniasienės.
4. Nustatyta, kad statistiniai parametrų įverčiai gali būti sudaryti remiantis išnaudojamumo įverčiais, gautais iš CVE įverčių duomenų bazės.
5. Atlikus parametrų statistinių įverčių pagrindimą naudojant ekspertinio vertinimo rezultatus nustatyta, kad skirtumai tarp CVE įverčių duomenų bazės ir ekspertinio vertinimo nėra reikšmingi, todėl CVE įverčių duomenų bazės įverčius galima naudoti virtualizuotų sistemų informacijos apsaugos rizikos analizei.

6. Nustatyta, kad metodo praktiniam pritaikymui reikalingas esamos situacijos, įskaitant programinės ir aparatinės įrangos architektūrą ir konfigūraciją, aprašymas griežtai apibrėžto modelio pavidalu organizacijose yra sunkiai sudaromas. Siekiant pagerinti metodo pritaikomumą buvo sukurtas modelių transformacijos metodas skirtas šiai situacijai supaprastinti.

Annexes[1]

Annex A. Description of the existing risk analysis methodologies

Annex B. The coauthors‘ agreements to present publications for the dissertation defence

Annex C. Copies of scientific publications by the autor on the topic of the dissertation

[1] The annexes are supplied in the enclosed compact disk.

Justinas JANULEVIČIUS

METHOD OF INFORMATION SECURITY RISK ANALYSIS FOR VIRTUALIZED SYSTEMS

Doctoral Dissertation

Technological Sciences,
Informatics Engineering (07T)

Justinas JANULEVIČIUS

VIRTUALIZUOTŲ SISTEMŲ INFORMACIJOS SAUGOS RIZIKOS ANALIZĖS METODO KŪRIMAS IR TAIKYMAS

Daktaro disertacija

Technologijos mokslai,
informatikos inžinerija (07T)

2016 11 18. 9,50 sp. l. Tiražas 20 egz.
Vilniaus Gedimino technikos universiteto
leidykla „Technika“,
Saulėtekio al. 11, 10223 Vilnius,
http://leidykla.vgtu.lt
Spausdino UAB „BMK leidykla“,
J. Jasinskio g. 16, 01112 Vilnius