INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 15, Issue: 3, Month: June, Year: 2020
Article Number: 3865, https://doi.org/10.15837/ijccc.2020.3.3865

CCC Publications

AGORA UNIVERSITY PRESS

Dynamic Expert System-Based Geographically Adapted Malware Risk Evaluation Method

D. Vitkus, J. Jezukeviciute, N. Goranin

Donatas Vitkus*
Vilnius Gediminas Technical University,
Vilnius, Lithuania
*Corresponding author: d.vitkus@vgtu.lt

Justina Jezukeviciute
Vilnius Gediminas Technical University,
Vilnius, Lithuania
justina.jezukeviciute@stud.vgtu.lt

Nikolaj Goranin
Vilnius Gediminas Technical University,
Vilnius, Lithuania
nikolaj.goranin@vgtu.lt

Abstract

Fast development of information systems and technologies while providing new opportunities for people and organizations also make them more vulnerable at the same time. Information security risk assessment helps to identify weak points and preparing mitigation actions. The analysis of expert systems has shown that rule-based expert systems are universal, and because of that can be considered as a proper solution for the task of risk assessment automation. But to assess information security risks quickly and accurately, it is necessary to process a large amount of data about newly discovered vulnerabilities or threats, to reflect regional and industry specific information, making the traditional approach of knowledge base formation for expert system problematic. This work presents a novel method for an automated expert systems knowledge base formation based on the integration of data on regional malware distribution from Cyberthreat real-time map providing current information on newly discovered threats. In our work we collect the necessary information from the web sites in an automated way, that can be later used in a relevant risk calculation. This paper presents method implementation, which includes not only knowledge base formation but also the development of the prototype of an expert system. It was created using the JESS expert system shell. Information security risk evaluation was performed according to OWASP risk assessment methodology, taking into account the location of the organization and prevalent malware in that area.

Keywords: information security risk analysis, expert systems, knowledge base formation, JESS, information acquisition.

1 Introduction

Information security risk assessment is a good way to manage the information security of the enterprise [1]. It helps not only in identifying weak points of enterprise information security management system but also in reducing costs of implementing information security controls because there is no necessity to implement all possible controls but only those, which are identified as necessary during the information security risk analysis.

Today's fast development of information systems and technologies provides new opportunities for people and organizations [5]. While it enables business to be faster and more competitive, it also makes them more vulnerable at the same time. Ontime conducted information security risk assessment allows to prepare a mitigation plan and reduce risks to an acceptable level [10]. On the other hand, companies need to employ information security risk manager or buy information security risk assessment service. Sometimes it becomes a problem, especially for small-medium enterprises, that is why automation of risk analysis process by employing expert systems is considered as a perspective approach.

This paper proposes a novel approach of knowledge base formation for expert systems compared to the traditional one, that requires real experts, knowledge base formation process is long, sometimes expensive [7]. It is suitable in cases where long-lasting data can be used, but not in information security risk assessment. Knowledge base used in information security risk assessment should reflect constantly changing threat landscape, evaluate the risk probability depending on the region and other factors.

The method proposed is based on the acquisition of data from trusted and open websites providing current information on newly discovered threats. The proposed method was implemented by data acquisition from the Cyberthreat real-time map website by Kaspersky Lab [19].

Also, a prototype of an expert system using a knowledge base with automatically acquired data was developed. The prototype was designed to perform the information security risk assessment according to the OWASP methodology [6]. The formed knowledge base was implemented on the Java Expert System Shell (JESS) [21]. The acquired data was automatically imported into JESS using XML format. It is necessary to state, that automatically acquired data has formed only part of a knowledge base, i.e. was integrated into the existing knowledge base and was used to estimate the malware probability, used in risk calculation in the specific region.

The paper is organized as follows. Section 1, current section, is an introduction. Section 2 presents related works in information security risk analysis and the development of expert systems knowledge base. Section 3 presents an approach of the integration of data on regional malware distribution from Cyber-threat real-time map into the knowledge base of risk analysis expert system for geographically adapted risk evaluation. Section 4 is results and discussions and finally, section 5 concludes the paper.

2 Related work

The analysis of expert systems has shown that rule-based expert systems are universal, and because of that, can be considered as a proper solution for the task of risk assessment automation. Expert systems can be a proper solution for enterprises because it allows to reduce costs for the risk management [2]. But to assess information security risks quickly and accurately, it is necessary to process a large amount of data about newly discovered vulnerabilities or threats, to reflect regional and industry specific information [9].

One of the most difficult points in creating expert systems is the creation of a knowledge base. In this part of the expert system are stored system rules, patterns, facts, data specifications and other structures [6]. It allows the inference engine to give the right answer to the task [3]. There are several ways of knowledge base formation methods [14]:

Manual – the knowledge is created by the real expert and knowledge engineer.

Semi-automatic – the knowledge is created partly by humans and partly in an automated way.

Fully-automatic – the knowledge is created without the participation of human experts.

In this paper we are analyzing automated knowledge base formation methods. One of such methods utilizes data import from specific domain ontologies [15]. It is a good way of knowledge base formation to use specific domain ontologies. This method is good because it allows us to reuse specific data, especially, where exact concepts are needed [18]. But the main disadvantage of this method is that there are several different ontologies languages, that means it is difficult to format knowledge base in universal way [8]. The other disadvantage is, that ontologies are not created every day, that means, we cannot get up to date data from ontologies format [17]. In the case of risk analysis, it is a must, since many threats, such as malware and vulnerabilities are dynamic in nature.

Another well known method for knowledge base formation is machine learning. Ethem Alpaydin (2010) suggested a simple method of machine learning for knowledge base formation [12]. There are several machine learning algorithms based on data classification (Bayes, Nearest neighbor method, regression, genetic algorithms, etc.) [16]. All of them are useful for prediction tasks where human can't make the prediction obviously. Machine learning algorithms don't need human interaction, they are self-learning systems and can work with various types of data. But the biggest disadvantage of such systems is that they are susceptible to errors [11]. It can take a lot time to discover the bug. Another limitation is related to the fact that the self-learning system need a lot of high-quality data [13]. It is not suitable for the information security analysis expert system, since, for example, the probability of a specific threat is a specific and dynamic value, and, as mention above, there is a need for actual up to date data.

The analysis of automated knowledge base formation methods for expert systems has shown that there are only a few such methods. The most popular are based on ontology integration and machine learning. The main disadvantage of ontology-based methods for risk analysis area is that data in ontology is typically static, outdated or suitable only for one risk assessment process part (e.g. for control selection), while machine learning is suitable when good datasets are available. It can be stated, that despite the fact that there are some automated methods for knowledge base formation, they are not suitable for creation of knowledge base of information security risk analysis expert system, responsible for the identification of threats in dynamic areas, such as malware and vulnerability threats.

3 Integration of data on regional malware distribution from Cyberthreat real-time map into the knowledge base

This work presents an automated expert systems knowledge base formation method based on the acquisition of data from trusted and open websites providing current information on newly discovered threats, by reading their HTML code and extracting information between tags. HTML was chosen because it is widely used [4] and structured, so it can be easily transformed into another format.

The method proposed is shown on in Figure 1.

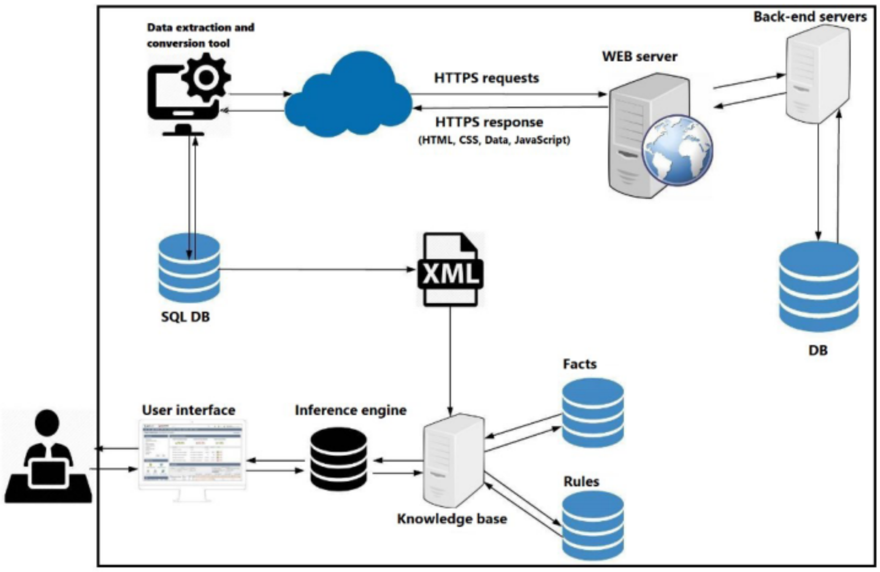
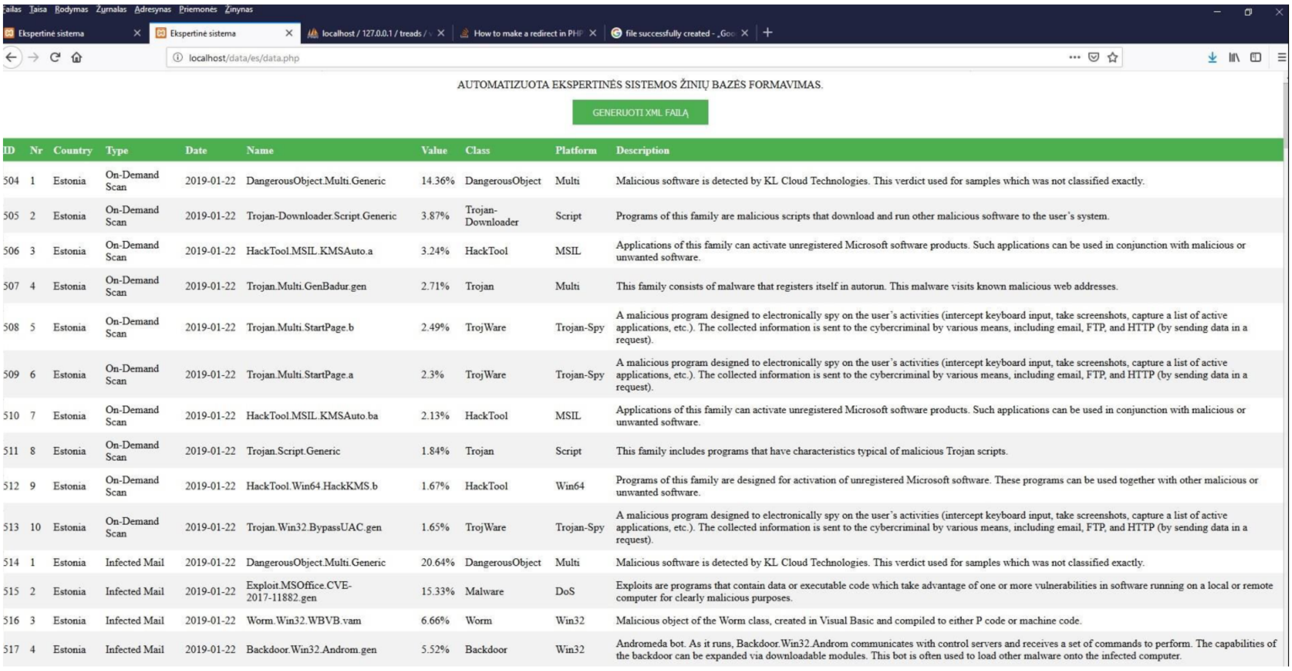


Figure 1: The architectural view of method implementation.


ailas Taisa Rodymas Žurnalas Adresynas Priemonės Žinynas
Ekspertinė sistema
Ekspertinė sistema
localhost / 127.0.0.1 / treads /
How to make a redirect in PHP
file successfully created - „Goo
localhost/data/es/data.php

AUTOMATIZUOTA EKSPERTINĖS SISTEMOS ŽINIŲ BAZĖS FORMAVIMAS.

GENERUOTI XML FAILĄ

ID	Nr	Country	Type	Date	Name	Value	Class	Platform	Description
504	1	Estonia	On-Demand Scan	2019-01-22	DangerousObject.Multi.Generic	14.36%	DangerousObject	Multi	Malicious software is detected by KL Cloud Technologies. This verdict used for samples which was not classified exactly.
505	2	Estonia	On-Demand Scan	2019-01-22	Trojan-Downloader.Script.Generic	3.87%	Trojan-Downloader	Script	Programs of this family are malicious scripts that download and run other malicious software to the user's system.
506	3	Estonia	On-Demand Scan	2019-01-22	HackTool.MSIL.KMSAuto.a	3.24%	HackTool	MSIL	Applications of this family can activate unregistered Microsoft software products. Such applications can be used in conjunction with malicious or unwanted software.
507	4	Estonia	On-Demand Scan	2019-01-22	Trojan.Multi.GenBadur.gen	2.71%	Trojan	Multi	This family consists of malware that registers itself in autorun. This malware visits known malicious web addresses.
508	5	Estonia	On-Demand Scan	2019-01-22	Trojan.Multi.StartPage.b	2.49%	TrojWare	Trojan-Spy	A malicious program designed to electronically spy on the user's activities (intercept keyboard input, take screenshots, capture a list of active applications, etc.). The collected information is sent to the cybercriminal by various means, including email, FTP, and HTTP (by sending data in a request).
509	6	Estonia	On-Demand Scan	2019-01-22	Trojan.Multi.StartPage.a	2.3%	TrojWare	Trojan-Spy	A malicious program designed to electronically spy on the user's activities (intercept keyboard input, take screenshots, capture a list of active applications, etc.). The collected information is sent to the cybercriminal by various means, including email, FTP, and HTTP (by sending data in a request).
510	7	Estonia	On-Demand Scan	2019-01-22	HackTool.MSIL.KMSAuto.ba	2.13%	HackTool	MSIL	Applications of this family can activate unregistered Microsoft software products. Such applications can be used in conjunction with malicious or unwanted software.
511	8	Estonia	On-Demand Scan	2019-01-22	Trojan.Script.Generic	1.84%	Trojan	Script	This family includes programs that have characteristics typical of malicious Trojan scripts.
512	9	Estonia	On-Demand Scan	2019-01-22	HackTool.Win64.HackKMS.b	1.67%	HackTool	Win64	Programs of this family are designed for activation of unregistered Microsoft software. These programs can be used together with other malicious or unwanted software.
513	10	Estonia	On-Demand Scan	2019-01-22	Trojan.Win32.BypassUAC.gen	1.65%	TrojWare	Trojan-Spy	A malicious program designed to electronically spy on the user's activities (intercept keyboard input, take screenshots, capture a list of active applications, etc.). The collected information is sent to the cybercriminal by various means, including email, FTP, and HTTP (by sending data in a request).
514	1	Estonia	Infected Mail	2019-01-22	DangerousObject.Multi.Generic	20.64%	DangerousObject	Multi	Malicious software is detected by KL Cloud Technologies. This verdict used for samples which was not classified exactly.
515	2	Estonia	Infected Mail	2019-01-22	Exploit.MSOffice.CVE-2017-11882.gen	15.33%	Malware	DoS	Exploits are programs that contain data or executable code which take advantage of one or more vulnerabilities in software running on a local or remote computer for clearly malicious purposes.
516	3	Estonia	Infected Mail	2019-01-22	Worm.Win32.WBVB.vam	6.66%	Worm	Win32	Malicious object of the Worm class, created in Visual Basic and compiled to either P code or machine code.
517	4	Estonia	Infected Mail	2019-01-22	Backdoor.Win32.Androm.gen	5.52%	Backdoor	Win32	Andromeda bot. As it runs, Backdoor.Win32.Androm communicates with control servers and receives a set of commands to perform. The capabilities of the backdoor can be expanded via downloadable modules. This bot is often used to load other malware onto the infected computer.

Figure 2: Example of automatically extracted data.

The proposed method has several steps:
1. Data extraction:
1.1. HTTPS requests to specified WEB page are sent.
1.2. HTTPS response is obtained from the WEB server.
1.3. Useful data about malware is extracted from HTML tags and stored to SQL DB.
2. Data conversion: all data in the SQL DB are converted to XML format.
3. Knowledge base formation: created XML file is imported to the existing expert system.

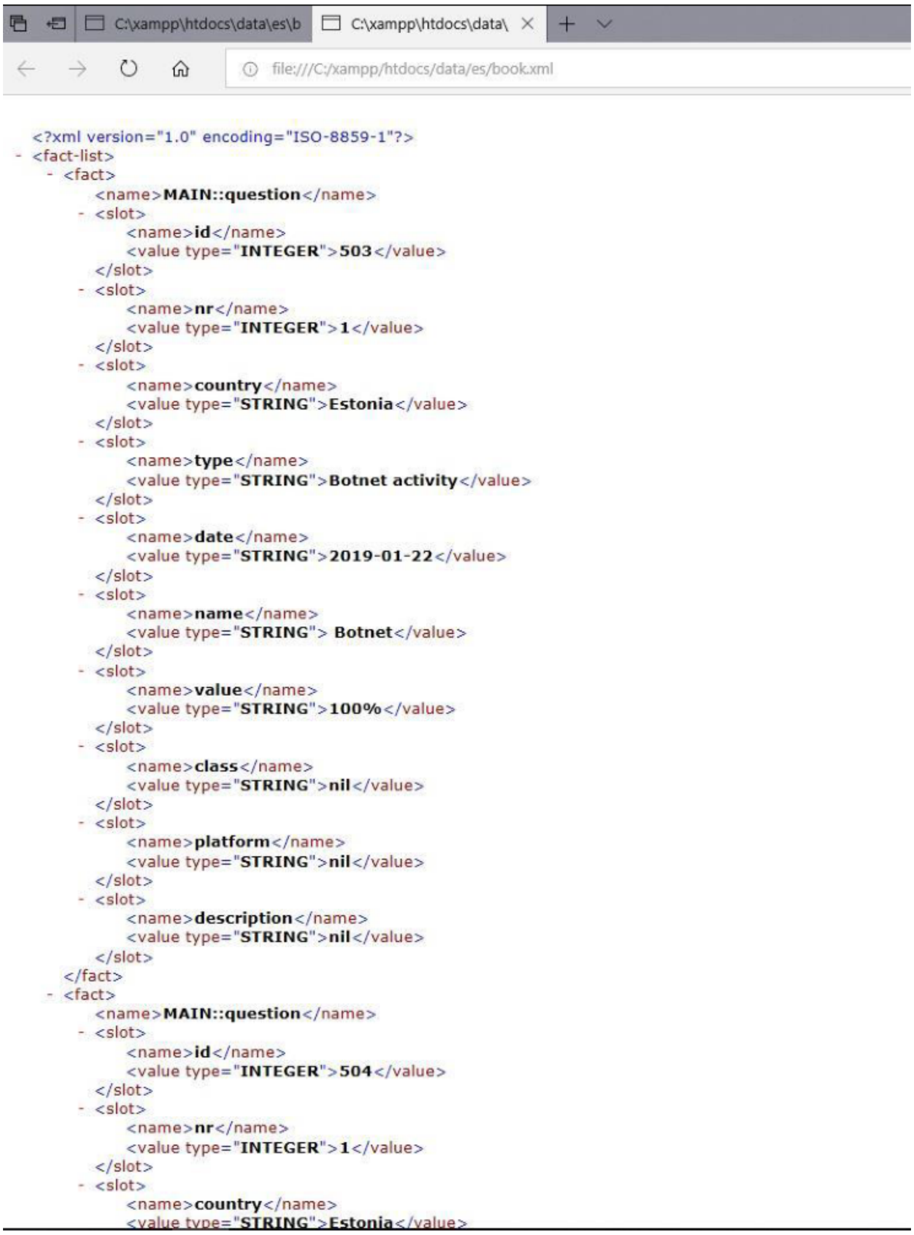
The method is implemented with a tool created in PHP language, which addresses the specified website by sending the HTTPS request. When a web server responds to the query, the tool reads the HTML code and processes the information between the specified tags. The collected data is stored in the SQL database. In this way, we get the needed knowledge from the web sites in an automated way, that can be later used in relevant risk calculations. Example of the automatically extracted data is shown in Figure 2.

The knowledge base formation is done by converting extracted data from SQL DB to XML format and importing it to JESS. Example of data conversion is shown in Figure 3.

The formed knowledge base was implemented on the Java expert system shell (JESS). All collected data was automatically imported into JESS using XML format. On Figure 4 an example of knowledge base import into JESS expert system prototype is demonstrated.

It is necessary to state, that automatically acquired data has formed only part of a knowledge base, i.e. was integrated into the existing knowledge base and was used only for estimation of malware probability used in risk calculation in the specific region. The prototype interacts with the user via a dialogue session. Risk identification is done by matching virus distribution platforms and platforms used in the organization. Also, the location of the organization and the prevalence of viruses in that area is evaluated. If an organization is operating in a country where the prevalence of the virus is recorded, the potential risk is calculated.

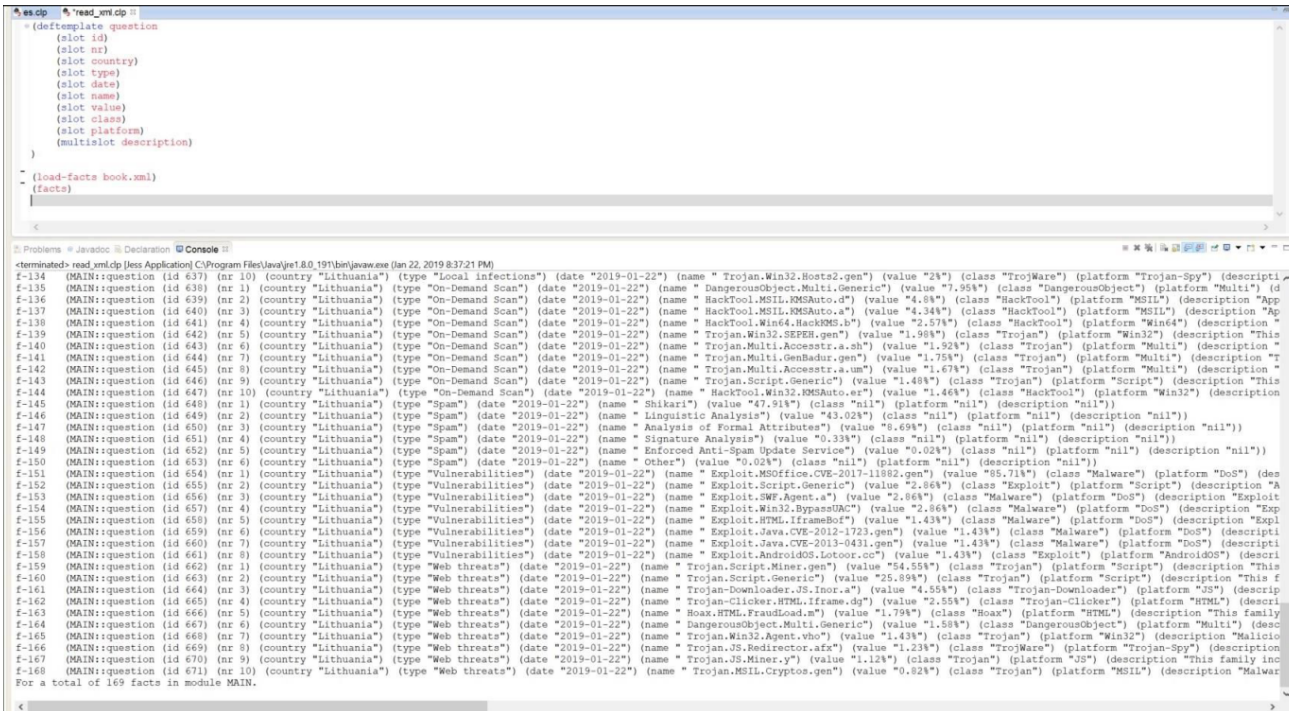
The proposed method was tested on data acquisition from the Cyberthreat real-time map website by Kaspersky Lab [19]. This site was chosen because it contains up to date information about known malicious code and its distribution location, in approximately 200 locations in total. In this work, information about 3 Baltic region countries was used: Lithuania, Estonia, Latvia. The extracted information contained the following information about malware: name, type, prevalence in the selected country, spreading platform, class, description, system, how it works, etc. Use case diagram of the method is shown in Figure 5.


C:\xampp\htdocs\data\es\b
C:\xampp\htdocs\data\
file:///C:/xampp/htdocs/data/es/book.xml

```
  <?xml version="1.0" encoding="ISO-8859-1"?>
- <fact-list>
  - <fact>
      <name>MAIN::question</name>
    - <slot>
        <name>id</name>
        <value type="INTEGER">503</value>
      </slot>
    - <slot>
        <name>nr</name>
        <value type="INTEGER">1</value>
      </slot>
    - <slot>
        <name>country</name>
        <value type="STRING">Estonia</value>
      </slot>
    - <slot>
        <name>type</name>
        <value type="STRING">Botnet activity</value>
      </slot>
    - <slot>
        <name>date</name>
        <value type="STRING">2019-01-22</value>
      </slot>
    - <slot>
        <name>name</name>
        <value type="STRING"> Botnet</value>
      </slot>
    - <slot>
        <name>value</name>
        <value type="STRING">100%</value>
      </slot>
    - <slot>
        <name>class</name>
        <value type="STRING">nil</value>
      </slot>
    - <slot>
        <name>platform</name>
        <value type="STRING">nil</value>
      </slot>
    - <slot>
        <name>description</name>
        <value type="STRING">nil</value>
      </slot>
    </fact>
  - <fact>
      <name>MAIN::question</name>
    - <slot>
        <name>id</name>
        <value type="INTEGER">504</value>
      </slot>
    - <slot>
        <name>nr</name>
        <value type="INTEGER">1</value>
      </slot>
    - <slot>
        <name>country</name>
        <value type="STRING">Estonia</value>
```

Figure 3: Example of data conversion to XML format.

While implementing the proposed method for the experiments, it was created a data extraction and conversion tool to get data from HTML, format the knowledge base and convert it to XML format. Also, a prototype of an expert system using a knowledge base with automatically acquired data was developed. The prototype was designed to perform the SME's information security risk assessment.



```
(deftemplate question
    (slot id)
    (slot nr)
    (slot country)
    (slot type)
    (slot date)
    (slot name)
    (slot value)
    (slot class)
    (slot platform)
    (multislot description)
)

(load-facts book.xml)
(facts)
```

```
<terminated> read_xml.clp [Jess Application] C:\Program Files\Java\jre1.8.0_191\bin\javaw.exe (Jan 22, 2019 8:37:21 PM)
f-134   (MAIN::question (id 637) (nr 10) (country "Lithuania") (type "Local infections") (date "2019-01-22") (name " Trojan.Win32.Hosts2.gen") (value "2%") (class "TrojWare") (platform "Trojan-Spy") (descripti
f-135   (MAIN::question (id 638) (nr 1) (country "Lithuania") (type "On-Demand Scan") (date "2019-01-22") (name " DangerousObject.Multi.Generic") (value "7.95%") (class "DangerousObject") (platform "Multi") (d
f-136   (MAIN::question (id 639) (nr 2) (country "Lithuania") (type "On-Demand Scan") (date "2019-01-22") (name " HackTool.MSIL.KMSAuto.d") (value "4.8%") (class "HackTool") (platform "MSIL") (description "App
f-137   (MAIN::question (id 640) (nr 3) (country "Lithuania") (type "On-Demand Scan") (date "2019-01-22") (name " HackTool.MSIL.KMSAuto.a") (value "4.34%") (class "HackTool") (platform "MSIL") (description "Ap
f-138   (MAIN::question (id 641) (nr 4) (country "Lithuania") (type "On-Demand Scan") (date "2019-01-22") (name " HackTool.Win64.HackKMS.b") (value "2.57%") (class "HackTool") (platform "Win64") (description "
f-139   (MAIN::question (id 642) (nr 5) (country "Lithuania") (type "On-Demand Scan") (date "2019-01-22") (name " Trojan.Win32.SEPEH.gen") (value "1.98%") (class "Trojan") (platform "Win32") (description "This
f-140   (MAIN::question (id 643) (nr 6) (country "Lithuania") (type "On-Demand Scan") (date "2019-01-22") (name " Trojan.Multi.Accesstr.a.sh") (value "1.92%") (class "Trojan") (platform "Multi") (description "
f-141   (MAIN::question (id 644) (nr 7) (country "Lithuania") (type "On-Demand Scan") (date "2019-01-22") (name " Trojan.Multi.GenBadur.gen") (value "1.75%") (class "Trojan") (platform "Multi") (description "T
f-142   (MAIN::question (id 645) (nr 8) (country "Lithuania") (type "On-Demand Scan") (date "2019-01-22") (name " Trojan.Multi.Accesstr.a.um") (value "1.67%") (class "Trojan") (platform "Multi") (description "
f-143   (MAIN::question (id 646) (nr 9) (country "Lithuania") (type "On-Demand Scan") (date "2019-01-22") (name " Trojan.Script.Generic") (value "1.48%") (class "Trojan") (platform "Script") (description "This
f-144   (MAIN::question (id 647) (nr 10) (country "Lithuania") (type "On-Demand Scan") (date "2019-01-22") (name " HackTool.Win32.KMSAuto.er") (value "1.46%") (class "HackTool") (platform "Win32") (description
f-145   (MAIN::question (id 648) (nr 1) (country "Lithuania") (type "Spam") (date "2019-01-22") (name " Shikari") (value "47.91%") (class "nil") (platform "nil") (description "nil"))
f-146   (MAIN::question (id 649) (nr 2) (country "Lithuania") (type "Spam") (date "2019-01-22") (name " Linguistic Analysis") (value "43.02%") (class "nil") (platform "nil") (description "nil"))
f-147   (MAIN::question (id 650) (nr 3) (country "Lithuania") (type "Spam") (date "2019-01-22") (name " Analysis of Formal Attributes") (value "8.69%") (class "nil") (platform "nil") (description "nil"))
f-148   (MAIN::question (id 651) (nr 4) (country "Lithuania") (type "Spam") (date "2019-01-22") (name " Signature Analysis") (value "0.33%") (class "nil") (platform "nil") (description "nil"))
f-149   (MAIN::question (id 652) (nr 5) (country "Lithuania") (type "Spam") (date "2019-01-22") (name " Enforced Anti-Spam Update Service") (value "0.02%") (class "nil") (platform "nil") (description "nil"))
f-150   (MAIN::question (id 653) (nr 6) (country "Lithuania") (type "Spam") (date "2019-01-22") (name " Other") (value "0.02%") (class "nil") (platform "nil") (description "nil"))
f-151   (MAIN::question (id 654) (nr 1) (country "Lithuania") (type "Vulnerabilities") (date "2019-01-22") (name " Exploit.MSOffice.CVE-2017-11882.gen") (value "85.71%") (class "Malware") (platform "DoS") (des
f-152   (MAIN::question (id 655) (nr 2) (country "Lithuania") (type "Vulnerabilities") (date "2019-01-22") (name " Exploit.Script.Generic") (value "2.86%") (class "Exploit") (platform "Script") (description "A
f-153   (MAIN::question (id 656) (nr 3) (country "Lithuania") (type "Vulnerabilities") (date "2019-01-22") (name " Exploit.SWF.Agent.a") (value "2.86%") (class "Malware") (platform "DoS") (description "Exploit
f-154   (MAIN::question (id 657) (nr 4) (country "Lithuania") (type "Vulnerabilities") (date "2019-01-22") (name " Exploit.Win32.BypassUAC") (value "2.86%") (class "Malware") (platform "DoS") (description "Exp
f-155   (MAIN::question (id 658) (nr 5) (country "Lithuania") (type "Vulnerabilities") (date "2019-01-22") (name " Exploit.HTML.IframeBof") (value "1.43%") (class "Malware") (platform "DoS") (description "Expl
f-156   (MAIN::question (id 659) (nr 6) (country "Lithuania") (type "Vulnerabilities") (date "2019-01-22") (name " Exploit.Java.CVE-2012-1723.gen") (value "1.43%") (class "Malware") (platform "DoS") (descripti
f-157   (MAIN::question (id 660) (nr 7) (country "Lithuania") (type "Vulnerabilities") (date "2019-01-22") (name " Exploit.Java.CVE-2013-0431.gen") (value "1.43%") (class "Malware") (platform "DoS") (descripti
f-158   (MAIN::question (id 661) (nr 8) (country "Lithuania") (type "Vulnerabilities") (date "2019-01-22") (name " Exploit.AndroidOS.Lotoor.cc") (value "1.43%") (class "Exploit") (platform "AndroidOS") (descri
f-159   (MAIN::question (id 662) (nr 1) (country "Lithuania") (type "Web threats") (date "2019-01-22") (name " Trojan.Script.Miner.gen") (value "54.55%") (class "Trojan") (platform "Script") (description "This
f-160   (MAIN::question (id 663) (nr 2) (country "Lithuania") (type "Web threats") (date "2019-01-22") (name " Trojan.Script.Generic") (value "25.89%") (class "Trojan") (platform "Script") (description "This f
f-161   (MAIN::question (id 664) (nr 3) (country "Lithuania") (type "Web threats") (date "2019-01-22") (name " Trojan-Downloader.JS.Inor.a") (value "4.55%") (class "Trojan-Downloader") (platform "JS") (descrip
f-162   (MAIN::question (id 665) (nr 4) (country "Lithuania") (type "Web threats") (date "2019-01-22") (name " Trojan-Clicker.HTML.Iframe.dg") (value "2.55%") (class "Trojan-Clicker") (platform "HTML") (descri
f-163   (MAIN::question (id 666) (nr 5) (country "Lithuania") (type "Web threats") (date "2019-01-22") (name " Hoax.HTML.FraudLoad.m") (value "1.79%") (class "Hoax") (platform "HTML") (description "This family
f-164   (MAIN::question (id 667) (nr 6) (country "Lithuania") (type "Web threats") (date "2019-01-22") (name " DangerousObject.Multi.Generic") (value "1.58%") (class "DangerousObject") (platform "Multi") (desc
f-165   (MAIN::question (id 668) (nr 7) (country "Lithuania") (type "Web threats") (date "2019-01-22") (name " Trojan.Win32.Agent.vho") (value "1.43%") (class "Trojan") (platform "Win32") (description "Malicio
f-166   (MAIN::question (id 669) (nr 8) (country "Lithuania") (type "Web threats") (date "2019-01-22") (name " Trojan.JS.Redirector.afx") (value "1.23%") (class "TrojWare") (platform "Trojan-Spy") (description
f-167   (MAIN::question (id 670) (nr 9) (country "Lithuania") (type "Web threats") (date "2019-01-22") (name " Trojan.JS.Miner.y") (value "1.12%") (class "Trojan") (platform "JS") (description "This family inc
f-168   (MAIN::question (id 671) (nr 10) (country "Lithuania") (type "Web threats") (date "2019-01-22") (name " Trojan.MSIL.Cryptos.gen") (value "0.82%") (class "Trojan") (platform "MSIL") (description "Malwar
For a total of 169 facts in module MAIN.
```

Figure 4: Example of the knowledge base import to the expert system.

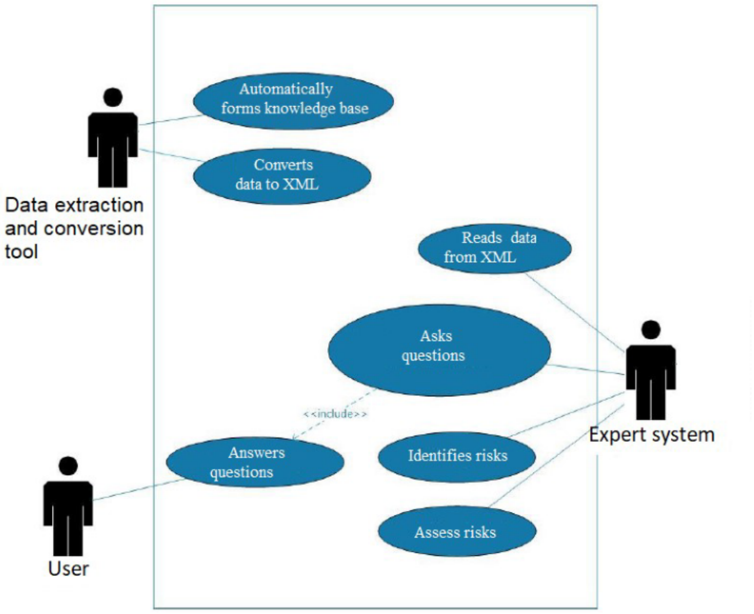
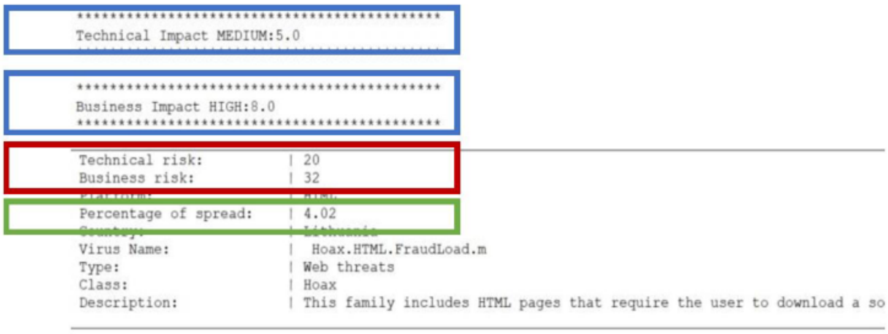


Figure 5: Method use case diagram.



```
**********************************************
Technical Impact MEDIUM:5.0
**********************************************

**********************************************
Business Impact HIGH:8.0
**********************************************

Technical risk:              | 20
Business risk:               | 32
Platform:                    | HTML
Percentage of spread:        | 4.02
Country:                     | Lithuania
Virus Name:                  |  Hoax.HTML.FraudLoad.m
Type:                        | Web threats
Class:                       | Hoax
Description:                 | This family includes HTML pages that require the user to download a so
```

Figure 6: Example of the risk calculation.

4 Results and discussion

During the experiment 225 web tags were scanned in total and 201 entries of knowledge base were formed automatically.

The proposed method has demonstrated a principal possibility of automated data extraction and import to the expert system knowledge base. Risk assessment, using automatically formed knowledge base, was conducted according to the OWASP methodology. According to the user answers, technical and business impacts were calculated in values from 0-9. The likelihood was evaluated according to the comparison of enterprise information systems platforms and malware platform.

Finally, the risk is calculated by formula: RISK = LIKELIHOOD * IMPACT. Example of the risk calculation is given in Figure 6.

Blue rectangles in the figure show impact, green – likelihood, red – the risk rate. The impact is evaluated by the business criteria, likelihood is a percentage of the malware spread in a specific region.

The main advantages of the method proposed are possibility to keep the knowledge base up to date with relevant information on current threats, to integrate information from multiple sources of information and to minimize workload while creating the knowledge base of an expert system. Limitations of the method are related to the changing structure of web sites providing data and the need to update the tool code manually because of that the further research is to be concentrated on tool adaptability for website structure.

5 Conclusions

The performed analysis has shown that increasing demand for automated information security risk assessment can be covered by the use of rules-based expert systems. However, the biggest problem in the development of this type of system is the formation and updating of the knowledge base especially in such a volatile area as cybersecurity. Manual methods of knowledge base formation are too slow, while existing semi-automatic and automatic knowledge base formation methods are concentrating on integration of static data, such as ontologies and can not depict the changing state in the cyberspace risk landscape. There is also a need to get risk assessment results fitted for the specific geographic location, since risk probabilities can be different in different regions due to variating malware population and other factors.

The new automated expert systems knowledge base formation method based on the acquisition of data from trusted and open websites providing current information on newly discovered malware and its geographic distribution was proposed. Method is based on scanning the HTML code, importing processed tag information into the database with further conversion into XML code and integration into the knowledge-base of a developed JESS-based expert system for geographically adapted malware risk assessment. Risk calculation is based on the widely accepted OWASP methodology. The method proposed differs from earlier methods by the ability to dynamically update the knowledge base rules and facts, thus reacting to the changing cyberthreat landscape, automatically present the emerging threats in a risk assessment with linking them to a geographic company location.

Kaspersky Lab's cyberthreat real-time map website was chosen as a data source for the method implementation tests. Data on geographical malware distribution was collected for 3 countries: Lithuania, Latvia and Estonia. The following data on cyber threats was collected in an automated way: malware description, related vulnerabilities, the prevalence of malware in the selected region, the malware spreading platform and typical impact on the system. The method tests on the developed JESS-based expert system has proved that automatically extracted data from the websites can improve the quality and expediency of risk assessment and provide geographically adapted results.

References

[1] Agrawal, V.A. (2017). *Comparative Study on Information Security Risk Analysis Methods*, JCP, 12.1, 57–67, 2017.

[2] Batista, L. O.; de Silva, G. A.; Araujo, V. S.; Araujo, V. J. S.; Rezende, T. S.; Guimaraes, A. J.; Souza, P. V. D. C. (2018). Fuzzy neural networks to create an expert system for detecting attacks by sql injection, *The International Journal of Forensic Computer Science*, 1, 8–21, 2018.

[3] Dheir, I.; Abu-Naser, S. S. (2019). Knowledge Based System for Diagnosing Guava Problems, *International Journal of Academic Information Systems Research (IJAISR)*, 3(3), 9–15, 2019.

[4] Dua, S.; Du, X. (2016). *Data mining and machine learning in cybersecurity*, CRC press, 2016.

[5] Dzitac, I.; Barbat, B. E. (2009). Artificial intelligence + distributed systems = agents, *International Journal of Computers Communications & Control*, 4(1), 17–26, 2009.

[6] Elsharif, A. A.; Abu-Naser, S. S. (2018). An Expert System for Diagnosing Sugarcane Diseases, *International Journal of Academic Engineering Research (IJAER)*, 3(3), 19–27, 2019.

[7] Kireeva, N.; Pozdnyak, I.; Gazizulina, A. (2019). Filling a Knowledge Base for Expert System in Information Security, *IOP Conference Series: Materials Science and Engineering*, 618(1), 2019.

[8] Kless, D.; Milton, S.; Kazmierczak, E.; Lindenthal, J. (2015). Thesaurus and ontology structure: Formal and pragmatic differences and similarities, *Journal of the Association for information science and technology*, 66(7), 1348–1366, 2015.

[9] Li, D.; Cai, Z.; Deng, L.; Yao, X.; Wang, H. H. (2018). Information security model of block chain based on intrusion sensing in the IoT environment, *Cluster Computing*, 22(1), 451–468, 2019.

[10] Losonczi, P.; Necas, P.; Nad, N. (2016). Risk management in information security, *Journal of Management*, 28, 2016.

[11] Mohamed E. (2017). Comparative study of four supervised machine learning techniques for classifications, *Information Journal of applied science and technology*, 7(2), 5-–18, 2017.

[12] Ramachandra, M. (2010). Information Mining, *Web-Based Supply Chain Management and Digital Signal Processing: Methods for Effective Information Administration and Transmission*, IGI Global, 223–231, 2010.

[13] Ristoski, P.; Paulheim, H. (2016). Web in data mining and knowledge discovery: A comprehensive survey, *Journal of Web Semantics*, 36, 1–22, 2016.

[14] Tandon, N.; Varde, A. S.; de Melo, G. (2017). Commonsense knowledge in machine intelligence, *SIGMOD Record*, 46(4), 2017.

[15] Vitkus, D.; Steckevicius, Z.; Goranin, N.; Kalibatiene, D.; Cenys, A. (2019). Automated Expert System Knowledge Base Development Method for Information Security Risk Analysis, *International Journal of Computers Communications & Control*, 14(6), 743–758, 2019.

[16] Xiao, H.; Rasul, K.; Vollgraf, R. (2017). Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms, *arXiv preprint*, 2017.

[17] Yadav, U.; Narula, G. S.; Duhan, N.; Jain, V.; Murthy, B. K. (2015). Development and visualization of domain specific ontology using protege, *Indian Journal of Science and Technology*, 9(16), 1–7, 2016.

[18] Zhong, W.; Liu, S.; Wan, F.; Li, Z. (2018). Equipment selection knowledge base system for industrial styrene process, *Chinese Journal of Chemical Engineering*, 26(8), 1707–1712, 2018.

[19] [Online]. Available: https://cybermap.kaspersky.com, Accesed on 27 March 2020.

[20] [Online]. Available: https://www.owasp.org/index.php/OWASP_Risk_Rating_Methodology, Accesed on 27 March 2020.

[21] [Online]. Available: https://jessrules.com/jess/download.shtml, Accesed on 27 March 2020.

Cite this paper as:
Vitkus, D.; Jezukeviciute, J.; Goranin, N. (2020). Dynamic Expert System-Based Geographically Adapted Malware Risk Evaluation Method, *International Journal of Computers Communications & Control*, 15(3), 3865, 2020.
https://doi.org/10.15837/ijccc.2020.3.3865