Acta Univ. Sapientiae, Mathematica, **10**, 2 (2018) 203–217

DOI: 10.2478/ausm-2018-0017

An improvement for a mathematical model for distributed vulnerability assessment

László Bognár
University of Dunaújváros, Hungary
email: bognarl@uniduna.hu

Antal Joós
University of Dunaújváros, Hungary
email: joosanti@gmail.com

Bálint Nagy
University of Dunaújváros, Hungary
email: nagyb@uniduna.hu

Abstract. Hadarics et. al. gave a Mathematical Model for Distributed Vulnerability Assessment. In this model the extent of vulnerability of a specific company IT infrastructure is measured by the probability of at least one successful malware attack when the users behaviour is also incorporated into the model. The different attacks are taken as independent random experiments and the probability is calculated accordingly. The model uses some input probabilities related to the characteristics of the different threats, protections and user behaviours which are estimated by the corresponding relative frequencies. In this paper this model is further detailed, improved and a numerical example is also presented.

1 Introduction

In recent decades information and infocommunication devices have become widely used. Besides their advantages previously unknown threats and malicious codes [8], [9] appeared. Traditionally measuring cyber risk usually consist of testing malicious activity [3] and penetration testing [10], [1]. Information can be obtained from the traffic of the network hence interactive metrics can

2010 Mathematics Subject Classification: 60A99, 94C99
Key words and phrases: vulnerability, probability, relative frequency

be evolved [5],[2], [7]. The behaviour of the users is usually regarded as a factor of secondary importance which can result in a model not adequately representing real life situations.

In an adequate model for assessing vulnerability of a specific business all three factors should be considered:

1. Malicious activity from the outher world threatening the IT network of the business.

2. Not properly protected elements of the IT network at the business.

3. Dangerous behaviours of users inside the business.

2 The model

Most of the notation of [4] will be used. For completeness these notations are to be reviewed.

Let $L\{l_1, \dots, l_\tau\}$ be the set of all available threat landscapes. In what follows a specific landscape will be used denoted by l. Let T_{all} be the set of all possible malware. Let $T = \{t_1, \dots, t_k\}$ be the set of all possible malware inside l. Let $U = \{u_1, \dots, u_r\}$ be the set of all users. Let $D = \{d_1, \dots, d_m\}$ be the set of all possible devices inside l. Let $P = \{p_1, \dots, p_n\}$ be the set of all available protections inside l. Let $UT = \{ut_1, \dots, ut_i\}$ be the set of all possible user tricks used by any malware inside l.

An integrated measure of vulnerability accounting for all three sources (attacker ingenuity, infrastructure weakness and adverse user behaviour) can be constructed.

For any given threat or class of threats for which the requisite IT infrastructure vulnerability and user facilitation is known, we can obtain a best estimate of:

1. The probability that an attacker will use a particular threat or class of threats against the enterprise (p_{prev}).
The probability p_{prev} is estimated by

$$p_{prev}(t, l) = \frac{\text{number of computers infected by } t \text{ inside } l}{\text{number of all computers inside } l}$$

for $t \in T$. Note, that p_{prev} can be based on a measurement or estimation and must be related to a time interval. Let

$$P_{prev} = \begin{array}{c|cccc} & t_1 & t_2 & \dots & t_k \\ \hline p & p_{prev}(t_1) & p_{prev}(t_2) & \dots & p_{prev}(t_k) \end{array}$$

be a vector. This means if we examine a particular attack, then the probability that this attack is in the form of the threat t_1 is $p_{prev}(t_1)$, etc.

2. The probability that the enterprise's IT infrastructure will allow the attack to be carried out successfully (p_{device}).

To elaborate the estimation of p_{device} first some auxiliary probabilities are defined and estimated.

The probability $p_{prot}(t, p)$ is introduced

$$p_{prot}(t, p) = \frac{\text{number of successful attempts of } t \text{ through the protection } p}{\text{number of all attempts of } t \text{ through the protection } p}$$

for any $t \in T$ and $p \in P$. Let

$$P_{prot} = \begin{array}{c|cccc} & p_1 & p_2 & \dots & p_n \\ \hline t_1 & p_{prot}(t_1, p_1) & p_{prot}(t_1, p_2) & \dots & p_{prot}(t_1, p_n) \\ t_2 & p_{prot}(t_2, p_1) & p_{prot}(t_2, p_2) & \dots & p_{prot}(t_2, p_n) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ t_k & p_{prot}(t_k, p_1) & p_{prot}(t_k, p_2) & \dots & p_{prot}(t_k, p_n) \end{array}$$

be a $k \times n$ matrix. This means that the probability of a successful attempt of t_1 through the protection p_1 is $p_{prot}(t_1, p_1)$, etc.

The value $z_{device-elements}(d, t)$ is introduced

$$z_{device-elements}(d, t) = \begin{cases} 1 & \text{if } t \text{ can work on } d \\ 0 & \text{if } t \text{ can not work on } d \end{cases}$$

(or shortly $z_{dev-elem}(d, t)$) for any $t \in T$ and $d \in D$. Let

$$Z_{device-elements}$$

$$= \begin{array}{c|cccc} & t_1 & t_2 & \dots & t_k \\ \hline d_1 & z_{dev-elem}(d_1, t_1) & z_{dev-elem}(d_1, t_2) & \dots & z_{dev-elem}(d_1, t_k) \\ d_2 & z_{dev-elem}(d_2, t_1) & z_{dev-elem}(d_2, t_2) & \dots & z_{dev-elem}(d_2, t_k) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ d_m & z_{dev-elem}(d_m, t_1) & z_{dev-elem}(d_m, t_2) & \dots & z_{dev-elem}(d_m, t_k) \end{array}$$

be an $m \times k$ matrix.

The value $z_{device-prot-install}(d, p)$ is introduced

$$z_{device-prot-install}(d, p) = \begin{cases} 1 & \text{if } d \text{ does not have the protection } p \\ 0 & \text{if } d \text{ has the protection } p \end{cases}$$

(or shortly z_{d-p-i}) for any $d \in D$ and $p \in P$. Let

$$Z_{device-prot-install}$$

$$= \begin{array}{c|cccc} & p_1 & p_2 & \dots & p_n \\ \hline d_1 & z_{d-p-i}(d_1,p_1) & z_{d-p-i}(d_1,p_2) & \dots & z_{d-p-i}(d_1,p_n) \\ d_2 & z_{d-p-i}(d_2,p_1) & z_{d-p-i}(d_2,p_2) & \dots & z_{d-p-i}(d_2,p_n) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ d_m & z_{d-p-i}(d_m,p_1) & z_{d-p-i}(d_m,p_2) & \dots & z_{d-p-i}(d_m,p_n) \end{array}$$

be an $m \times n$ matrix. Let

$$P_{device-prot-install-d_j}$$

$$= \begin{array}{c|cccc} & p_1 & p_2 & \dots & p_n \\ \hline t_1 & p_{d-p-i-d_j}(t_1,p_1) & p_{d-p-i-d_j}(t_1,p_2) & \dots & p_{d-p-i-d_j}(t_1,p_n) \\ t_2 & p_{d-p-i-d_j}(t_2,p_1) & p_{d-p-i-d_j}(t_2,p_2) & \dots & p_{d-p-i-d_j}(t_2,p_n) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ t_k & p_{d-p-i-d_j}(t_k,p_1) & p_{d-p-i-d_j}(t_k,p_2) & \dots & p_{d-p-i-d_j}(t_k,p_n) \end{array}$$

be a $k \times n$ matrix where

$$p_{d-p-i-d_j}(t_x,p_y) = \max\{p_{prot}(t_x,p_y), z_{d-p-i}(d_j,p_y)\}$$

for any $j \in \{1,\dots,m\}$, $x \in \{1,\dots,k\}$ and $y \in \{1,\dots,n\}$. This means that if the threat t_1 can work on d_j, then the probability of a successful attempts of the threat t_1 through the protection p_1 on the device d_j is $p_{d-p-i-d_j}(t_1,p_1)$, etc. The probability $p_{device-prot-d_j}(t)$ is introduced

$$p_{device-prot-d_j}(t) = \min_{\text{for all } p \text{ protecting } d_j} p_{prot}(t,p)$$

for any $t \in T$. Let

$$P_{device-prot-d_j} = \begin{array}{c|c} & p \\ \hline t_1 & p_{device-prot-d_j}(t_1) \\ t_2 & p_{device-prot-d_j}(t_2) \\ \vdots & \vdots \\ t_k & p_{device-prot-d_j}(t_k) \end{array}$$

be the column vector where

$$p_{device-prot-d_j}(t_x)$$

$$= \min\{p_{d-p-i-d_j}(t_x, p_1), p_{d-p-i-d_j}(t_x, p_2), \dots, p_{d-p-i-d_j}(t_x, p_n)\}$$

for any $j \in \{1, \dots, m\}$ and $x \in \{1, \dots, k\}$. This means that if the threat t_1 can work on d_j, then the probability of a successful attempts of the threat t_1 through any protection protecting the device d_j is $p_{device-prot-d_j}(t_1)$, etc.
The probability $p_{device-prot}(d, t)$ is introduced

$$p_{device-prot}(d, t) = \min_{\text{for all } p \text{ protecting } d} p_{prot}(t, p)$$

for any $t \in T$ and $d \in D$. Let

$$P_{device-prot} = \begin{array}{c|cccc} & t_1 & t_2 & \dots & t_k \\ \hline d_1 & p_{device-prot}(d_1, t_1) & p_{device-prot}(d_1, t_2) & \dots & p_{device-prot}(d_1, t_k) \\ d_2 & p_{device-prot}(d_2, t_1) & p_{device-prot}(d_2, t_2) & \dots & p_{device-prot}(d_2, t_k) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ d_m & p_{device-prot}(d_m, t_1) & p_{device-prot}(d_m, t_2) & \dots & p_{device-prot}(d_m, t_k) \end{array}$$

be an $m \times k$ matrix where

$$p_{device-prot}(d_x, t_y) = p_{device-prot-d_x}(t_y)$$

for any $x \in \{1, \dots, m\}$ and $y \in \{1, \dots, k\}$.
The probability $p_{device}(d, t)$ is introduced

$$p_{device}(d, t) = z_{decive-elements}(d, t) \cdot p_{device-prot}(d, t)$$

for any $t \in T$ and $d \in D$. Let

$$P_{device} = \begin{array}{c|cccc} & t_1 & t_2 & \dots & t_k \\ \hline d_1 & p_{device}(d_1, t_1) & p_{device}(d_1, t_2) & \dots & p_{device}(d_1, t_k) \\ d_2 & p_{device}(d_2, t_1) & p_{device}(d_2, t_2) & \dots & p_{device}(d_2, t_k) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ d_m & p_{device}(d_m, t_1) & p_{device}(d_m, t_2) & \dots & p_{device}(d_m, t_k) \end{array}$$

be an $m \times k$ matrix where

$$p_{device}(d_x, t_y) = z_{dev-elem}(d_x, t_y) \cdot p_{device-prot}(d_x, t_y)$$

for any $x \in \{1, \dots, m\}$ and $y \in \{1, \dots, k\}$. This means that the probability of a successful attempts of the threat t_1 through any protection protecting the device d_1 is $p_{device}(d_1, t_1)$, etc.

3. The probability that users of the enterprise's IT infrastructure will provide sufficient facilitation for the attack to succeed (p_{user}).
The $p_{usertrick}(t, ut)$ probability is introduced

$$p_{usertrick}(t, ut) = \frac{\text{number of attempts of } t \text{ where } t \text{ used } ut}{\text{number of all attempts of } t}$$

for any $t \in T$ and $ut \in UT$. Let

$$P_{usertrick}$$

$$= \begin{array}{c|cccc} & ut_1 & ut_2 & \dots & ut_i \\ \hline t_1 & p_{usertrick}(t_1, ut_1) & p_{usertrick}(t_1, ut_2) & \dots & p_{usertrick}(t_1, ut_i) \\ t_2 & p_{usertrick}(t_2, ut_1) & p_{usertrick}(t_2, ut_2) & \dots & p_{usertrick}(t_2, ut_i) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ t_k & p_{usertrick}(t_k, ut_1) & p_{usertrick}(t_k, ut_2) & \dots & p_{usertrick}(t_k, ut_i) \end{array}$$

be a $k \times i$ matrix. This means that the probability that the threat t_1 uses usertrick ut_1 is $p_{usertrick}(t_1, ut_1)$, etc.
The $p_{user-usertrick}(u, ut)$ probability is introduced

$$p_{user-usertrick}(u, ut) = \frac{\text{number of successful attempts of } ut \text{ on } u}{\text{number of all attempts of } ut \text{ on } u}$$

(or shortly $p_{u-utrick}(u, ut)$) for any $u \in U$ and $ut \in UT$. Let

$$P_{user-usertrick}$$

$$= \begin{array}{c|cccc} & ut_1 & ut_2 & \dots & ut_i \\ \hline u_1 & p_{u-utrick}(u_1, ut_1) & p_{u-utrick}(u_1, ut_2) & \dots & p_{u-utrick}(u_1, ut_i) \\ u_2 & p_{u-utrick}(u_2, ut_1) & p_{u-utrick}(u_2, ut_2) & \dots & p_{u-utrick}(u_2, ut_i) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ u_r & p_{u-utrick}(u_r, ut_1) & p_{u-utrick}(u_r, ut_2) & \dots & p_{u-utrick}(u_r, ut_i) \end{array}$$

be an $r \times i$ matrix. This means that the probability that the user u_1 uses usertrick ut_1 is $p_{u-utrick}(u_1, ut_1)$, etc.
From the probabilities $p_{usertrick}$ and $p_{user-usertrick}$ we can calculate the probability $p_{user}(u, t)$ which is the probability that the threat t infects using at least one usertrick through the user u. This is

$$p_{user}(u, t)$$

$$= 1 - \prod_{\text{for all } ut \text{ used by } t} (1 - p_{usertrick}(t, ut) \cdot p_{user-usertrick}(u, ut))$$

for any $u \in U$, $t \in T$ and $ut \in UT$. Let

$$P_{user} = \begin{array}{c|cccc} & t_1 & t_2 & \dots & t_k \\ \hline u_1 & p_{user}(u_1, t_1) & p_{user}(u_1, t_2) & \dots & p_{user}(u_1, t_k) \\ u_2 & p_{user}(u_2, t_1) & p_{user}(u_2, t_2) & \dots & p_{user}(u_2, t_k) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ u_r & p_{user}(u_r, t_1) & p_{user}(u_r, t_2) & \dots & p_{user}(u_r, t_k) \end{array}$$

be an $r \times k$ matrix where

$$\begin{gathered} p_{user}(u_1, t_1) \\ = 1 - (1 - p_{usertrick}(t_1, ut_1) \cdot p_{u-utrick}(u_1, ut_1)) \\ \cdot (1 - p_{usertrick}(t_1, ut_2) \cdot p_{u-utrick}(u_1, ut_2)) \cdot \dots \\ \cdot (1 - p_{usertrick}(t_1, ut_i) \cdot p_{u-utrick}(u_1, ut_i)), \end{gathered}$$

etc. This means that the probability that the threat t_1 infects using at least one usertrick through the user u_1 is $p_{user}(u_1, t_1)$, etc.

2.1 The probability of infection

These three probabilities (p_{prev}, p_{device}, p_{user}) can be combined to obtain an overall probability of malicious success, (provided each relevant combination of attack, user, and component of IT infrastructure is accounted for) [6]. The (p_{prev}, p_{device}, p_{user}) values are related to a given threat, a given user and a given device. The aggregated vulnerability would be an index of the whole organization related to all of the users, all of the devices and all of the possible threats. The probability of the infection is p_s which is the probability that the investigated landscape will be infected by at least one malware. This can be calculated in the following form

$$p_s = 1 - \prod_{\text{for all } t,u \text{ and } d} (1 - p_{user}(t, u) \cdot p_{device}(t, d) \cdot p_{prev}(t, l))$$

for any $u \in U$, $t \in T$ and $d \in D$.

The followings were assumed:

1. the attacker usage of the given threat, the IT infrastructure allowance and the user acceptance are different from each other,

2. all of the attack attempts are independent from each other,

3. the computer usage behaviours of all users are the same and equal to the average usage in the organization.

Observe the calculated p_s value is related to the same time interval as the original p_{prev} was related to.

3 A numerical example

Let $T = \{t_1, \ldots, t_4\}$ be the set of malware. Let $U = \{u_1, \ldots, u_7\}$ be the set of all users. Let $D = \{d_1, d_2, d_3\}$ be the set of all devices. Let $P = \{p_1, \ldots, p_5\}$ be the set of all protections. Let $UT = \{ut_1, \ldots, ut_6\}$ be the set of all user tricks used by any malware in T. Let

$P_{prev} =$

t_1	t_2	t_3	t_4
0.25	0.25	0.25	0.25

and

$P_{prot} =$

	p_1	p_2	p_3	p_4	p_5
t_1	0.01	0.02	0.03	0.04	0.02
t_2	0.11	0.12	0.13	0.14	0.15
t_3	0.21	0.22	0.23	0.24	0.25
t_4	0.31	0.32	0.33	0.34	0.35

.

This means that the probability of a successful attempt of t_1 through the protection p_1 is 0.01, etc.
Let

$Z_{device\_elements} =$

	t_1	t_2	t_3	t_4
d_1	1	0	0	0
d_2	0	1	1	0
d_3	0	1	0	1

.

This means that t_1 can work on d_1, t_2 can not work on d_1, etc.
Let

$Z_{device\_prot\_install} =$

	p_1	p_2	p_3	p_4	p_5
d_1	1	0	1	0	1
d_2	0	1	1	0	1
d_3	1	0	0	1	1

.

This means that d_1 does not have the protection p_1, d_1 has the protection p_2, etc.

Thus

$$P_{\text{prot\_install\_d}_1} = \begin{array}{c|ccccc} & p_1 & p_2 & p_3 & p_4 & p_5 \\ \hline t_1 & 1 & 0.02 & 1 & 0.04 & 1 \\ t_2 & 1 & 0.12 & 1 & 0.14 & 1 \\ t_3 & 1 & 0.22 & 1 & 0.24 & 1 \\ t_4 & 1 & 0.32 & 1 & 0.34 & 1 \end{array}.$$

Observe

$$p_{d-p-i-d_1}(t_1, p_1) = \max\{p_{prot}(t_1, p_1), z_{d-p-i}(d_1, p_1)\} = \max\{0.01, 1\} = 1,$$
$$p_{d-p-i-d_1}(t_1, p_2) = \max\{p_{prot}(t_1, p_2), z_{d-p-i}(d_1, p_2)\} = \max\{0.02, 0\} = 0.02,$$

etc. This means that the probability of a successful attempts of the threat t_1 through the protection p_1 on the device d_1 is $p_{d-p-i-d_1}(t_1, p_1)$, etc. Similarly

$$P_{\text{prot\_intall\_d}_2} = \begin{array}{c|ccccc} & p_1 & p_2 & p_3 & p_4 & p_5 \\ \hline t_1 & 0.01 & 1 & 1 & 0.04 & 1 \\ t_2 & 0.11 & 1 & 1 & 0.14 & 1 \\ t_3 & 0.21 & 1 & 1 & 0.24 & 1 \\ t_4 & 0.31 & 1 & 1 & 0.34 & 1 \end{array},$$

$$P_{\text{prot\_intall\_d}_3} = \begin{array}{c|ccccc} & p_1 & p_2 & p_3 & p_4 & p_5 \\ \hline t_1 & 1 & 0.02 & 0.03 & 1 & 1 \\ t_2 & 1 & 0.12 & 0.13 & 1 & 1 \\ t_3 & 1 & 0.22 & 0.23 & 1 & 1 \\ t_4 & 1 & 0.32 & 0.33 & 1 & 1 \end{array}.$$

Furthermore

$$P_{\text{device\_prot\_D}_1} = \begin{array}{c|c} & P \\ \hline t_1 & 0.02 \\ t_2 & 0.12 \\ t_3 & 0.22 \\ t_4 & 0.32 \end{array}.$$

Observe

$$\begin{aligned} &p_{device-prot-d_1}(t_1) \\ &= \min\{p_{d-p-i-d_1}(t_1, p_1), p_{d-p-i-d_1}(t_1, p_2), \ldots, p_{d-p-i-d_1}(t_1, p_n)\} \\ &\quad \min\{1, 0.02, 0.03, 1, 1\} = 0.02. \end{aligned}$$

This means that if the threat t_1 can work on d_1, then the probability of a successful attempts of the threat t_1 through any protection protecting the device d_1 is 0.02, etc. Similarly

$$P_{device-prot-d_2} = \begin{array}{c|c} & P \\ \hline t_1 & 0.01 \\ t_2 & 0.11 \\ t_3 & 0.21 \\ t_4 & 0.31 \end{array},$$

$$P_{device-prot-d_3} = \begin{array}{c|c} & P \\ \hline t_1 & 0.02 \\ t_2 & 0.12 \\ t_3 & 0.22 \\ t_4 & 0.32 \end{array}.$$

Thus

$$P_{device-prot} = \begin{array}{c|cccc} & t_1 & t_2 & t_3 & t_4 \\ \hline d_1 & 0.02 & 0.12 & 0.22 & 0.32 \\ d_2 & 0.01 & 0.11 & 0.21 & 0.31 \\ d_3 & 0.02 & 0.12 & 0.22 & 0.32 \end{array}.$$

Observe

$$p_{device-prot}(d_1, t_1) = p_{device-prot-d_1}(t_1),$$
$$p_{device-prot}(d_1, t_2) = p_{device-prot-d_1}(t_2),$$

etc. This means that if the threat t_1 can work on d_1, then the probability of a successful attempts of the threat t_1 through any protection protecting the device d_1 is 0.02, etc. Furthermore

$$P_{device} = \begin{array}{c|cccc} & t_1 & t_2 & t_3 & t_4 \\ \hline d_1 & 0.02 & 0 & 0 & 0 \\ d_2 & 0 & 0.11 & 0.21 & 0 \\ d_3 & 0 & 0.12 & 0 & 0.32 \end{array}.$$

Observe

$$p_{device}(d_1, t_1) = z_{dev-elem}(d_1, t_1) \cdot p_{device-prot}(d_1, t_1) = 0.02 \cdot 1 = 0.02,$$
$$p_{device}(d_1, t_2) = z_{dev-elem}(d_1, t_2) \cdot p_{device-prot}(d_1, t_2) = 0.12 \cdot 0 = 0,$$

etc. This means that the probability of a successful attempts of the threat t_1 through any protection protecting the device d_1 is 0.02. Since t_2 can not work on d_1, the probability of a successful attempts of the threat t_2 through any protection protecting the device d_1 is 0, etc. Let

$P_{usertrick} =$

	ut_1	ut_2	ut_3	ut_4	ut_5	ut_6
t_1	0.141	0.142	0.143	0.144	0.145	0.146
t_2	0.151	0.152	0.153	0.154	0.155	0.156
t_3	0.161	0.162	0.163	0.164	0.165	0.166
t_4	0.171	0.172	0.173	0.174	0.175	0.176

This means that the probability that the threat t_1 uses usertrick ut_1 is 0.141, etc. Observe the sum of the probabilities in any row is not greater than 1. Let

$P_{user\_usertrick} =$

	ut_1	ut_2	ut_3	ut_4	ut_5	ut_6
u_1	0.031	0.032	0.033	0.034	0.035	0.036
u_2	0.041	0.042	0.043	0.044	0.045	0.046
u_3	0.051	0.052	0.053	0.054	0.055	0.056
u_4	0.061	0.062	0.063	0.064	0.065	0.066
u_5	0.071	0.072	0.073	0.074	0.075	0.076
u_6	0.081	0.082	0.083	0.084	0.085	0.086
u_7	0.091	0.092	0.093	0.094	0.095	0.096

This means that the probability that the user u_1 uses usertrick ut_1 is 0.031, etc. Thus

$P_{user} =$

	t_1	t_2	t_3	t_4
u_1	0.028516	0.030477	0.032434	0.034388
u_2	0.036891	0.039418	0.041939	0.044455
u_3	0.045206	0.048290	0.051366	0.054434
u_4	0.053460	0.057094	0.060716	0.064326
u_5	0.061655	0.065830	0.069989	0.074132
u_6	0.069791	0.074498	0.079185	0.083852
u_7	0.077868	0.083099	0.088305	0.093487

Observe

$$\begin{aligned} p_{user}(u_1, t_1) = 1 - (&1 - p_{usertrick}(t_1, ut_1) \cdot p_{u-utrick}(u_1, ut_1)) \\ &\cdot (1 - p_{usertrick}(t_1, ut_2) \cdot p_{u-utrick}(u_1, ut_2)) \\ &\cdot \ldots \cdot (1 - p_{usertrick}(t_1, ut_i) \cdot p_{u-utrick}(u_1, ut_i)) \\ = 1 - (&1 - 0.141 \cdot 0.031) \cdot (1 - 0.142 \cdot 0.032) \cdot \ldots \cdot (1 - 0.146 \cdot 0.036) \\ = 0.&028516, \end{aligned}$$

etc. Therefore,

$$\begin{aligned} p_s &= 1 - (1 - p_{user}(t_1, u_1) \cdot p_{device}(t_1, d_1) \cdot p_{prev}(t_1)) \\ &\quad \cdot (1 - p_{user}(t_1, u_2) \cdot p_{device}(t_1, d_1) \cdot p_{prev}(t_1)) \\ &\quad \cdot \ldots \cdot (1 - p_{user}(t_4, u_7) \cdot p_{device}(t_4, d_3) \cdot p_{prev}(t_4)) \\ &= 1 - (1 - 0.028516 \cdot 0.02 \cdot 0.25) \cdot (1 - 0.036891 \cdot 0.02 \cdot 0.25) \\ &\quad \cdot \ldots \cdot (1 - 0.093487 \cdot 0.32 \cdot 0.25) = 0.079774. \end{aligned}$$

This means that the probability of the infection of the investigated company with users $u_1, \ldots, u_7$, devices d_1, d_2, d_3, protections $p_1, \ldots, p_5$ and matrices as above is 0.079774. Thus we get that the probability of an infection by at least one malware is 0.079774.

4 Simulations

In this section results of simulation studies are presented. Businesses with different sizes (different number of devices and users) are modelled and the p_s probabilities are calculated when certain number of threats are present. The results are summarized in Table 1 and Table 2.

The *Micro* (*Small*, *Medium*, *Big*, resp.) business is a company (or department) with 10 (50, 100, 1000, resp.) devices and 10 (50, 100, 1000, resp.) users. In real life the probabilities p_{prev}, p_{prot}, $p_{usertrick}$ and $p_{user-usertrick}$ can be estimated by relative frequencies but in the simulations these were estimated by random uniform probabilities. In the Table 1 the probabilities p_{prev} (p_{prot}, $p_{usertrick}$, $p_{user-usertrick}$, resp.) are in the interval $[0.9, 1]$ ($[0, 0.1]$, $[0, 0.1]$, $[0, 0.1]$, resp.). The results in the Table 1 correspond to the case when the number of protections is 5 and the number of usertrick is 5.

The probability 0.25 in the cell of the third row of the second column in Table 1 means that the approximate probability of p_s is 0.25 if there are 10 devices, 10 users in the company, the number of threats is 10, the number of protections is 5, the number of usertricks is 5, the random elements of the vector P_{prev} lie on the interval $[0.9, 1]$, the random elements of the matrix P_{prot} lie on the interval $[0, 0.1]$, the random elements of the matrix $P_{usertrick}$ lie on the interval $[0, 0.1]$ and the random elements of the matrix $P_{user-usertrick}$ lie on the interval $[0, 0.1]$. Of course the matrices $Z_{device-elements}$ and $Z_{device-prot-install}$ are random matrices with elements 0 or 1.

Observe that if the number of the devices (or users) or the number of the threats is large, then the probability is close to 1.

Table 1: The values of p_s probabilities in case of different business sizes

	Micro	Small	Medium	Big
threats	devices=10 users=10	devices=50 users=50	devices=100 users=100	devices=1000 users=1000
10	0.25	0.999935 91999547	1	1
50	0.75	0.999999 99996973	1	1
100	0.85	1	1	1
1000	0.999999 99715744	1	1	1

The probabilities in Table 1 can be regarded as overestimates of the real p_s probabilities since the sum of the elements in the random vector P_{prev} is greater than 1.

In the Table 2 the probabilities p_{prev} (p_{prot}, $p_{usertrick}$, $p_{user-usertrick}$, resp.) are in the interval $[0, 0.1]$ ($[0, 0.1]$, $[0, 0.1]$, $[0, 0.1]$, resp.). The results in the Table 2 correspond to the case when the number of protections is 5 and the number of usertrick is 5.

Table 2: The values of p_s probabilities in case of different business sizes

	Micro	Small	Medium	Big
threats	devices=10 users=10	devices=50 users=50	devices=100 users=100	devices=1000 users=1000
10	0.02	0.25	75	1
50	0.07	0.85	0.9986016 7849174	1
100	0.15	0.996973 10258718	0.999999 99963790	1
1000	0.7	0.999999 99998364	1	1

The difference between the Table 1 and Table 2 is the input random data p_{prev}.

5 Conclusions

From the simulation studies it can be seen that the model presented can be used for defining an index number reflecting the state of vulnerability of a certain company against cyber attacks. However these simulations also show that this model has constraints of applicability because if the size of the company is big enough, then the probability p_s is very close to 1 and no distinction can be made between the vulnerability of different companies. To overcome these constrains of the applicability it can be used either only to a smaller part of a large network or to a randomly selected smaller sample of users and devices.

This index can be a good measuring tool of comparing the vulnerability of different parts of a company or comparing the state of vulnerability of a company at different time instances.

Comparing different user behaviours can give valuable pieces of information for the company managements about the needs of improving employees awareness against cyber attacks.

References

[1] M. T. Chapman, Establishing metrics to manage the human layer, ISSA Security Education Awareness Special Interest Group, 2013.

[2] M. T. Chapman, *Advanced Persistent Testing: How to fight bad phishing with good*, PhishLine, 2015. http://www.phishline.com/advanced-persistent-testing-ebook

[3] S. E. Edwards, R. Ford, G. Szappanos, *Effectively testing APT defenses*, Virus Bulletin Conference, Prague, Czech Republic, 2015.

[4] K. Hadarics, K. Györffy, B. Nagy, L. Bognár, A. Arrott, F. Leitold, Mathematical Model of Distributed Vulnerability Assessment, In: Jaroslav Dočkal, Milan Jirsa, Josef Kaderka, Proceedings of Conference SPI 2017: Security and Protection of Information. Brno, 2017.07.01-2017.07.02. Brno: University of Defence, (2017), 45–57. (ISBN:978-80-7231-414-0)

[5] F. Lalonde Levesque, J. M. Fernandez, A. Somayaji, *Risk prediction of malware victimization based on user behavior*, Malicious and Unwanted Software: The Americas (MALWARE), 2014 9th International Conference on. IEEE, 2014.

[6] F. Leitold, A. Arrott, K. Hadarics, *Quantifying cyber-threat vulnerability by combining threat intelligence, IT infrastructure weakness, and user susceptibility*, 24th Annual EICAR Conference, Nuremberg, Germany, 2016.

[7] F. Leitold, K. Hadarics, Measuring security risk in the cloud-enabled enterprise, In: Dr Fernando C Colon Osorio, 7th International Conference on Malicious and Unwanted Software (MALWARE), Fajardo, Puerto Rico, 2012.10.16-2012.10.18. Piscataway (NJ): IEEE, (2012), 62–66. (ISBN:978-1-4673-4880-5)

[8] NIST SP 800-53r4 Security and Privacy Controls for Federal Information Systems and Organizations, 2013.

[9] NIST SP 800-83r1 Guide to Malware Incident Prevention and Handling for Desktops and Laptops, 2013.

[10] Pwnie Express, *Vulnerability assessment and penetration testing across the enterprise*, Whitepaper, 2014. http://www.pwnieexpress.com

Received: January 7, 2018