Acta Univ. Sapientiae, Mathematica, **12**, 1 (2020) 14–29

DOI: 10.2478/ausm-2020-0002

Assessing the effect size of users' consciousness for computer networks vulnerability

László Bognár
University of Dunaújváros,
Táncsics Mihály utca 1/A,
Dunaújváros, Hungary 2400
email: bognarl@uniduna.hu

Antal Joós
University of Dunaújváros,
Táncsics Mihály utca 1/A,
Dunaújváros, Hungary 2400
email: joosa@uniduna.hu

Bálint Nagy
University of Dunaújváros,
Táncsics Mihály utca 1/A,
Dunaújváros, Hungary 2400
email: nagyb@uniduna.hu

Abstract. In this paper the conditions and the findings of a simulation study is presented for assessing the effect size of users' consciousness to the computer network vulnerability in risky cyber attack situations at a certain business. First a simple model is set up to classify the groups of users according to their skills and awareness then probabilities are assigned to each class describing the likelihood of committing dangerous reactions in case of a cyber attack. To quantify the level of network vulnerability a metric developed in a former work is used. This metric shows the approximate probability of an infection at a given business with well specified parameters according to its location, the type of the attack, the protections used at the business etc. The findings mirror back the expected tendencies namely if the number of conscious user is on the

2010 Mathematics Subject Classification: 60A99, 94C99
Key words and phrases: vulnerability, users' awareness, computer networks, simulation study, cyber security

rise the "relative improvement of the cyber security" is increasing. The tendencies in the change of this relative improvement are established, different graphs and curves are constructed to give an overall view for the influence of the different parameters. In addition to these general conclusions assessments are made for the magnitude and for the range of the relative cyber security improvement. An interesting findings that even in the case of small differences in skills making the users more conscious in their reactions can significantly enhance the level of cyber security at a business.

1 Introduction

Assessing the extent of vulnerability of a net of computers against outer cyber threats is of prime interest for both the IT experts sector and the businesses using computer networks. Most of the IT solutions concentrate on different hardware and software protections against the threats and little attention is paid on the effect of the users' behavior during their daily routine handling potentially risky situations. Opening potentially dangerous websites, clicking on links in emails from unknown source, downloading files to the computers are typical "user tricks" which may raise the level of risk of infections.

Making the users more conscious in their computer usage is an evidential tool for enhancing the cyber security of a business. However to give some measure for the effectiveness of these kind of efforts (trainings for employees, incentives, penalties, etc.) is essential for those offering these services and for the managements of the businesses as well.

In this paper the findings of computer simulation studies are presented where the users at a business are categorized according to their everyday computer usage. The three categories (Naive, Typical, and Conscious) encompasses the different types of user groups mainly reflecting their attitudes and behaviors in risky cyber threat situations.

To make the influence of the different groups sensible the $\mathtt{p}_\mathtt{s}$ metric for measuring cyber vulnerability developed in [4], [3], [2] and [1] is used. This $\mathtt{p}_\mathtt{s}$ metric is the probability that at least one cyber malware (virus) can successfully go through the IT protections and infects a certain net of computers.

The features of the given net and the prevailing threats at the location investigated are precisely characterized by different matrices describing the presence or absence of the different types of protections, the relative frequencies and the level of danger of different viruses and the level of danger of the different user tricks (Simple, Moderate, Complex).

This p_s metric holds the intrinsic characteristic of being monotonously increasing as the number of viruses, devices or users increases. In this study to ensure getting such a p_s metric which reflects the real differences of vulnerability attributable exclusively to the different categories of users' consciousness (awareness) and excludes the effect of the business size the number of users, devices, and threats are kept constant while the proportions of the different types of users are changed.

Besides these kept-constant parameters other input values like probabilities of occurrences of certain threats or probabilities of applying certain user tricks, etc. were also temporarily fixed in the analysis (at the values which are typical for the majority of businesses) but later the sensitivity of the p_s value against these varying input values is demonstrated.

The final goal was to illustrate the extent of changes in the value of p_s when the level of consciousness at a business increases due to some measures introduced by the management. The increase of consciousness is embodied in the increase of the proportion of Conscious users reevaluating some users' status from Typical or Naive to Conscious or in the increase of the proportion of Typical users reevaluating some users' status from Naive to Typical.

The direct relationship between the p_s value and some specific business financial indicator is not investigated here.

2 The model for users classification

According to their consciousness there are three distinguished class of users

- *Naive* users,
- *Typical* users,
- *Conscious* users.

As the names of the categories suggest Naive users are assumed to commit dangerous actions even in cases requiring very simple user tricks while Conscious users are victims only of threats requiring more sophisticated user tricks.

Let ν be

$$\begin{bmatrix} c \\ t \\ n \end{bmatrix}$$

the vector where c (t, n reps.) is the number of the Conscious (Typical, Naive resp.) users. Let r be the number of users. Observe $c + t + n = r$.

There are three distinguished type of user tricks

- *Simple* user trick,
- *Moderate* user trick,
- *Complex* user trick.

The Simple (Complex resp.) user trick is the easiest (complicated resp.) trick for the threat to attack a given device.

Let P_{skills} be the

	Simple	Moderate	Complex
Conscious	$\beta_{C,1}$	$\beta_{C,2}$	$\beta_{C,3}$
Typical	$\beta_{T,1}$	$\beta_{T,2}$	$\beta_{T,3}$
Naive	$\beta_{N,1}$	$\beta_{N,2}$	$\beta_{N,3}$

3×3 matrix. The real number $\beta_{C,1}$ ($\beta_{C,2}$, $\beta_{C,3}$ resp.) is the probability that a Conscious user uses a Simple (Moderate, Complex resp.) user trick. The real number $\beta_{T,1}$ ($\beta_{T,2}$, $\beta_{T,3}$ resp.) is the probability that a Typical user uses a Simple (Moderate, Complex resp.) user trick. The real number $\beta_{N,1}$ ($\beta_{N,2}$, $\beta_{N,3}$ resp.) is the probability that a Naive user uses a Simple (Moderate, Complex resp.) user trick. It is assumed that $\beta_{i,1} < \beta_{i,2} < \beta_{i,3}$ and $\beta_{C,j} < \beta_{T,j} < \beta_{N,j}$ for $i = C, T, N$ and $j = 1, 2, 3$.

Combining ν and P_{skills} let $P_{user-usertrick}$ be the

	Simple	Moderate	Comlex
u_1	$\beta_{C,1}$	$\beta_{C,2}$	$\beta_{C,3}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_c	$\beta_{C,1}$	$\beta_{C,2}$	$\beta_{C,3}$
u_{c+1}	$\beta_{T,1}$	$\beta_{T,2}$	$\beta_{T,3}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_{c+t}	$\beta_{T,1}$	$\beta_{T,2}$	$\beta_{T,3}$
u_{c+t+1}	$\beta_{N,1}$	$\beta_{N,2}$	$\beta_{N,3}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
u_{c+t+n}	$\beta_{N,1}$	$\beta_{N,2}$	$\beta_{N,3}$

$r \times 3$ matrix. Here $u_1, \ldots, u_c$ denote the users belonging to the group of Conscious users, $u_{c+1}, \ldots, u_{c+t}$ denote the users in the Typical group and $u_{c+t+1}, \ldots, u_{c+t+n}$ stand for the users in the Naive group.

In this paper this $P_{user-usertrick}$ matrix is the main tool to study the effect of users' behaviour at a business against different cyber threats. Changing the values of c, t, n within the fixed value of r ($r = c + t + n$) or changing the proportion (distribution) of probabilities in the rows/columns of this matrix enables us to study different situations and give an overview about the magnitude of the effect size of users' consciousness to the cyber vulnerability.

3 The p_s probability

In [2] the probability of infection p_s was introduced which is the probability that the investigated landscape will be infected by at least one malware. This can be calculated in the following form

$$p_s = 1 - \prod_{t=1,\dots,k;u=1,\dots,r;d=1,\dots,m} (1 - p_{user}(t,u) \cdot p_{device}(t,d) \cdot p_{prev}(t)) \qquad (1)$$

for any $u \in U$, $t \in T$ and $d \in D$, where U (T and D resp.) symbolizes the set of users (threats and devices resp.) at the landscape investigated. In (1) $p_{user}(t,u)$ is the probability that the threat t infects the landscape using at least one usertrick through the user u. In (1) $p_{device}(t,d)$ is the probability of a successful attempts of the threat t through any protection protecting the device d. In (1) $p_{prev}(t)$ is the probability that an attack is in the form of the threat t.

In [2] and [1] it was shown how these probabilities can be computed using several parameters describing the present state of the investigated landscape, the prevailing threats, the devices, the state of protections etc.

Since the number of these parameters influencing the value of p_s is numerous it is not easy to see general tendencies due to some selected specific parameters without keeping the others at constant values.

In this study where the aim was to get information about the effect of users' consciousness the parameter values referring to the other features of the investigated business have been fixed. Hence two types of parameters have been distinguished:

- kept-constant parameters and
- study parameters.

The kept-constant parameters are those which are not varied in the study at all. They can simply be regarded as those specifying the basic unit of

comparison. For example the total number of users is fixed to 100 since this parameter can be regarded as the size of the users' sample taken from the population of all users at a given business.

The study parameters are being varied in the study. Since they are still rather numerous it was practical to temporarily select and fix them at some typical real world values (called "typical study values") and investigate the sensitivity of p_s to the deviation from these typical study values later.

4 The kept-constant parameters and their values in the study

The values of the kept-constant parameters:

- the number of malwares: $k = 10$,
- the number of users: $r = 100$,
- the number of devices: $m = 10$,
- the number of protections: $n = 10$,
- the number of user tricks: $i = 3$,
- the number of groups of user skills: $s = 3$,
- the probability that an attacker will use a particular threat or class of threats against the enterprise: $P_{prev} = [\ 1/k,\ \ 1/k,\ \ \ldots\ \ ,1/k\ ]$,
- the $Z_{device-elements}$ matrix which describes that in this situation each virus can work on each device:

$$Z_{device-elements} = \begin{array}{c|ccc} & 1 & \ldots & k \\ \hline 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ldots & \vdots \\ m & 1 & \ldots & 1 \end{array},$$

- the $Z_{device-prot-install}$ matrix which describes that in this situation each protection is installed on each device:

$$Z_{device-prot-install} = \begin{array}{c|ccc} & 1 & \ldots & n \\ \hline 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ m & 0 & \ldots & 0 \end{array},$$

- the P_{prot} matrix which describes that in this situation the probability of a successful attempt of a given threat through at a given protection:

$$P_{prot} = \begin{array}{c|ccc} & 1 & \dots & n \\ \hline 1 & 1/(nk) & \dots & 1/(nk) \\ \vdots & \vdots & \dots & \vdots \\ k & 1/(nk) & \dots & 1/(nk) \end{array}.$$

5 The study parameters and their "typical study values"

The "typical study values":

- The distribution of probabilities in the P_{skills} matrix:

$$P_{skills} = \begin{array}{l|ccc} & \text{Simple} & \text{Moderate} & \text{Complex} \\ \hline \text{Conscious} & p_0 & 3p_0 & 6p_0 \\ \text{Typical} & 3p_0 & 9p_0 & 18p_0 \\ \text{Naive} & 6p_0 & 18p_0 & 36p_0 \end{array}, \text{ where } p_0 = 0.0001.$$

- The distribution of probabilities in the $P_{usertrick}$ matrix:

$$P_{usertrick} = \begin{array}{c|ccc} & \text{Simple} & \text{Moderate} & \text{Complex} \\ \hline t_1 & \alpha_1 & \alpha_2 & \alpha_3 \\ \vdots & \vdots & \vdots & \vdots \\ t_k & \alpha_1 & \alpha_2 & \alpha_3 \end{array}, \text{ where } \begin{array}{l} \alpha_1 = 0.6, \\ \alpha_2 = 0.3, \\ \alpha_3 = 0.1. \end{array}$$

Here $P_{usertrick}$ is a $k \times 3$ matrix. The real number α_1 (α_2, α_3 resp.) is the probability that a threat uses a Simple (Moderate, Complex resp.) user trick. The integer number k denotes the total number of threats involved in the study.

The fact that each row of the $P_{usertrick}$ matrix has the same values of α_1, α_2 and α_3 tacitly assumes that each virus behaves similarly that is all viruses involved in the analysis belong to the same group of viruses with respect to their user tricks required to activate them. Obviously it does not hold for all groups of viruses so later the effect of differently distributed α probabilities will be investigated.

6 The effect of users' consciousness for the p_s value in the case of "typical study values"

To illustrate the magnitude of the effect size of users' consciousness extreme situations have been analyzed where first all users were assumed to belong to one specific class of consciousness ("Original class") and gradually each user is trained to step up into a "higher" class of consciousness ("Improved class"). Three extreme versions of this "class change" are detailed:

- from Naive to Typical,
- from Naive to Conscious,
- from Typical to Conscious.

Various graphs have been constructed to visualize the effect size.

On the first kind of graphs the change of the absolute value of p_s probability is depicted as the function of the number of "reevaluated" users. This probability is denoted by $p(x)$ where x refers to the number of "reevaluated" users. (Sometimes they are called to "reeducated" users.) Accordingly $r - x$ refers to the number of users still in the "Original class". If $r = 100$, $p(0)$ refers to the case when all users are in their "Original class" while $p(100)$ indicates the situation when all users have been "reevaluated".

On the second kind of graphs the $\Delta(x)$ function, the relative change of the p_s probability is illustrated

$$\Delta(x) = \frac{p(0) - p(x)}{p(0)}.$$

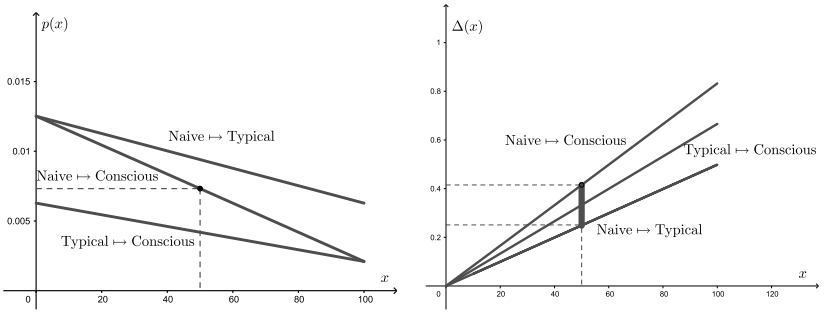
Since the absolute value of p_s is very much dependent on the different "class change" situations more practical to calculate the ratio of the probability change to the p_s value in the "Original class". This change can be interpreted as the "relative improvement of the defence".

From business point of view this relative improvement can be the basis of any management measures for the sake of improving cyber security through the enhancement of users' consciousness.

In Fig. 1(a) the $p(x)$ curves are shown for the three extreme "class change" situations. The changes of $p(x)$ are almost linear in all three cases, naturally the slope of the Naive $\mapsto$ Consious line is the steepest one.

On the Naive $\mapsto$ Consious curve in Fig. 1(a) it can be seen that the initial probability of the infection is $p(0) = 0.01252$ and if 50 Naive users

are reeducated to Conscious users, then the probability of the infection is $p(50) = 0.00732$.



(a) The $p(x)$ curve (b) The $\Delta(x)$ curve

Figure 1: The "typical study values".

The $\Delta(x)$ curves can be constructed from the corresponding $p(x)$ curves hence for all three "class change" situations these curves also show almost perfect linearity. If 50 Naive users are reeducated to Conscious users, then the change of the defence is $p(0)-p(50) = 0.01252-0.00732 = 0.0052$ and the normalized change of defence with respect to the initial probability of the infection is $\Delta(50) = (p(0) - p(50))/p(0) = 0.41511$ which can be seen in Fig. 1(b).

These $\Delta(x)$ curves can be regarded as the most important findings of the simulation studies. Even for those who are not very familiar with the issues of cyber security the magnitude of the improvement can be convincing. Seeing the different extreme situations of "class change" one can find that significant improvement can be reached through the enhancement of the users' consciousness.

The range of this relative improvement for a specific x value can be assessed as the difference of the $\Delta(x)$ values for the Naive $\mapsto$ Conscious and the Naive $\mapsto$ Typical "class change" situations. Both the magnitude and the range of the improvement is monotonously increasing as the number of the "reevaluated" users is increasing. For example the relative improvement of the defence can vary from about 25 to 40 percent when half of the users' status has been changed. In Fig. 1(b) this range is indicated with a thick solid line.

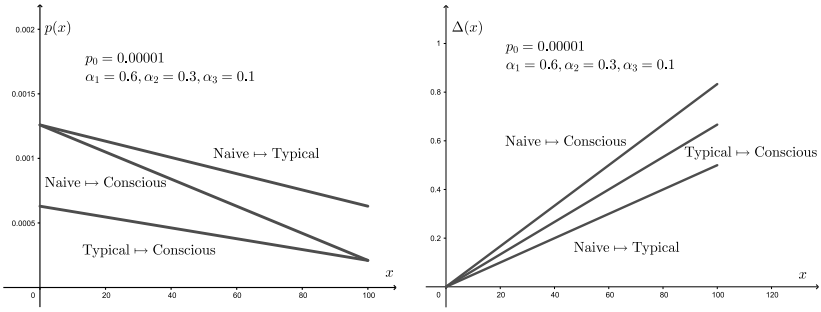
7 The sensitivity of the relative cyber security improvement to the deviations from the "typical study values"

In the remaining sections the sensitivity of the relative security improvement is investigated. As it was stated earlier the simulation studies were elaborated for those "typical study values" of the study parameters which are believed to be characteristic for real world average size businesses in every day cyber risk situations.

However it is worth to check whether slight or moderate deviations from these study values results in basically different conclusions or the findings are rather insensitive to these deviations.

7.1 Varying p_0

In the Fig. 2(a) - Fig. 4(b) the influence of the deviation from the $p_0 = 0.0001$ study value is shown. The range of the p_0 values goes from 0.00001 to 0.01.



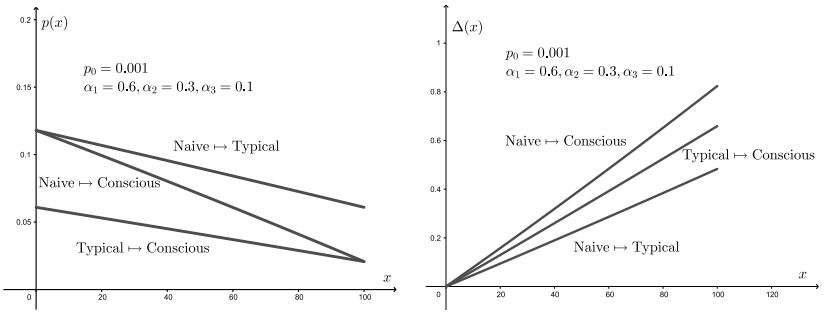
(a) The p(x) curve

(b) The Δ(x) curve

Figure 2: The value $p_0 = 0.00001$.

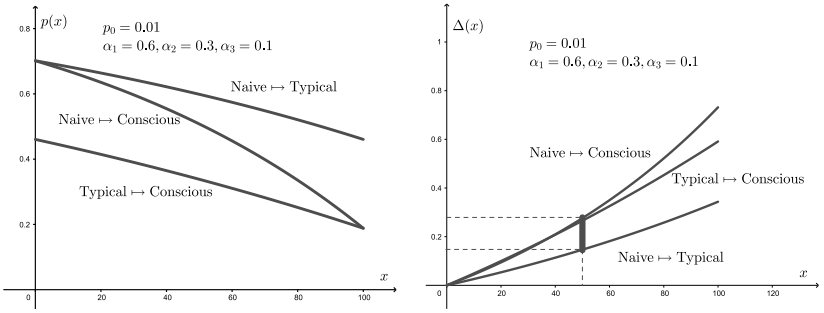
The magnitude of the $p(x)$ probabilities varies and the shape of the $p(x)$ and $\Delta(x)$ curves are slightly different from the almost linear "typical study values" curves.

With the increase of the p_0 probability the $\Delta(x)$ curves deviates from the linear relationships and tend to be more "exponential-like". At the same time the $\Delta(x)$ values for a specific x value are typically less then those for the typical $p_0 = 0.0001$ study value.



(a) The p(x) curve (b) The Δ(x) curve

Figure 3: The value $p_0 = 0.001$.



(a) The p(x) curve (b) The Δ(x) curve

Figure 4: The value $p_0 = 0.01$.

Besides these slight differences one can see that the range of the relative cyber security improvement still remain similar to the previously investigated cases. For example for the $p_0 = 0.01$ value the range of $\Delta(x)$ is somewhere between 15 and 30 percent if half of the users' status has been "reevaluated". In Fig. 4(b) this range is indicated with a thick solid line.

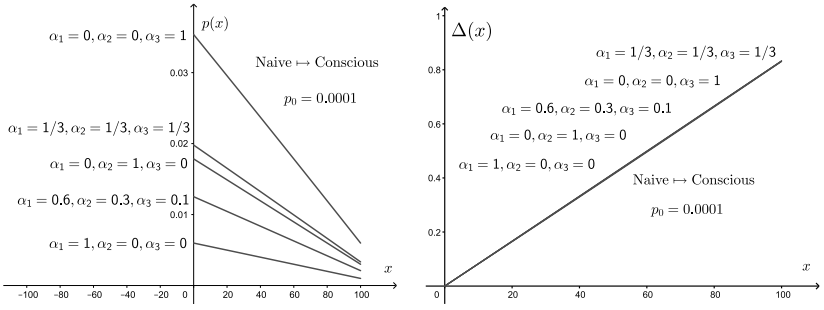
7.2 Varying α_1, α_2 and α_3

In the Fig. 5(a) - Fig. 5(b) the influence of the deviation from the $\alpha_1 = 0.6$, $\alpha_2 = 0.3$, $\alpha_3 = 0.1$ study values is shown.

The results are presented for the following sets of deviated α values:

- $\alpha_1 = 0,\ \alpha_2 = 0,\ \alpha_3 = 1,$
- $\alpha_1 = 0,\ \alpha_2 = 1,\ \alpha_3 = 0,$
- $\alpha_1 = 1,\ \alpha_2 = 0,\ \alpha_3 = 0,$
- $\alpha_1 = 1/3,\ \alpha_2 = 1/3,\ \alpha_3 = 1/3.$

These sets of α values can be regarded as the representations of different virus groups requiring different user tricks to activate them.



(a) The p(x) curve (b) The Δ(x) curve

Figure 5: Varying α_1, α_2 and α_3.

The Δ(x) curves practically coincide with each other convincingly demonstrating the insensitivity of the Δ(x) function to the deviations from the typical α_1, α_2, α_3 study values.

7.3 Varying the ratio of the columns of P_{skills}

To check the sensitivity of the relative improvement to the ratio of the values in the columns of the P_{skills} matrix it is more convenient to rewrite the matrix into the form:

$P_{skills} =$

	Simple	Moderate	Complex
Conscious	$\gamma_1 p_0$	$\gamma_2 p_0$	$\gamma_3 p_0$
Typical	$*$	$*$	$*$
Naive	$6\gamma_1 p_0$	$6\gamma_2 p_0$	$6\gamma_3 p_0$

Using these notations it is easy to express that the simulation studies covered the following sets of γ parameters:

- $\gamma_1 = 6$, $\gamma_2 = 3$, $\gamma_3 = 1$,
- $\gamma_1 = 1$, $\gamma_2 = 3$, $\gamma_3 = 6$,
- $\gamma_1 = 1$, $\gamma_2 = 1$, $\gamma_3 = 1$.

Having investigated all three cases one can establish that the $\Delta(x)$ function is entirely insensitive to the different sets of γ values. The details are not shown here.

7.4 Varying the ratio of the rows of P_{skills}

To check the sensitivity of the relative improvement to the ratio of the values in the rows of the P_{skills} matrix it is more convenient to rewrite the matrix into the form:

$$P_{skills} = \begin{array}{l|ccc} & \text{Simple} & \text{Moderate} & \text{Complex} \\ \hline \text{Conscious} & p_0 & 3p_0 & 6p_0 \\ \text{Typical} & * & * & * \\ \text{Naive} & \delta p_0 & 3\delta p_0 & 6\delta p_0 \end{array}$$

Using these notations it is easy to express that the simulation studies covered the following set of δ values: $\delta = 1, 2, 4, 6, 12$. In Fig. 6 the $p(x)$ curves are shown for the "Naive-Conscious" "class change" situation for this set of δ values.

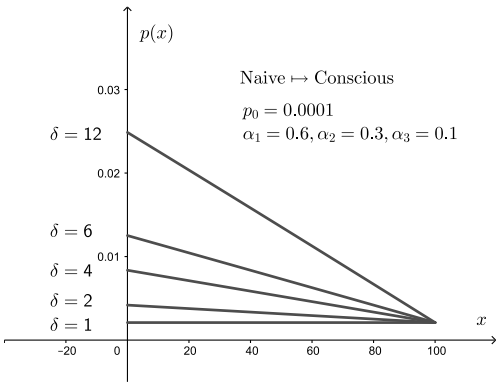
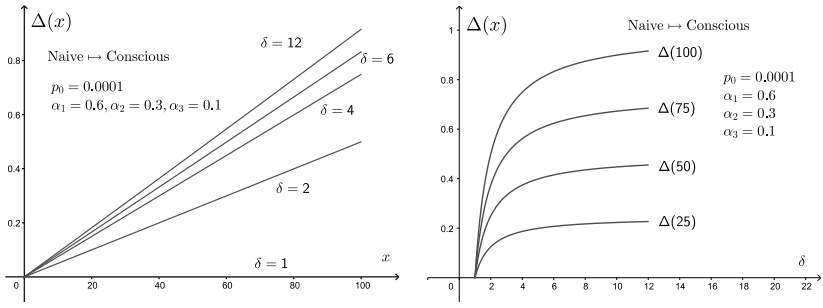


Figure 6: The $p(x)$ curve.

This sensitivity study when the effect of the varying δ value is being investigated is rather special. Since one specific δ value represents the differences between the Conscious and the Naive users in their skills (more precisely the ratio of the corresponding probabilities), it is straightforward that the effect of "reeducation" is larger if this difference is larger. If this difference is zero (the δ value is 1) the "reeducation" is obviously useless resulting in a $\Delta(x) = 0$ value.

Hence this sensitivity study is mainly for learning the form of the curves describing the relationship between the number of "reeducated" users and the relative cyber security improvement and also learning the relationship between the difference of skills and the relative cyber security improvement rather than simply establishing the fact of the existence of this sensitivity for the varying δ values.

In Fig. 7(a) the almost perfect linear association can be established between the number of "reeducated" users (x) and the relative cyber security improvement ($\Delta(x)$ values) for all δ values.



(a) The $\Delta(x)$ curve.

(b) The curves $\Delta(25)$, $\Delta(50)$, $\Delta(75)$, $\Delta(100)$

Figure 7: Varying δ.

In contrast the relationship between the difference in skills (δ) and the relative cyber security improvement ($\Delta(x)$) is far from linear. All curves for the different selected x values show steep increase in the relative improvement as the value of δ starts to depart from its initial value of 1 but later the slopes of the curves are becoming smaller and smaller seemingly tending to a limit value of $\Delta(x)$.

From practical point of view the steep starting phase of the curves can be of prime interest. It means that even in case of small differences between the user groups' skills it may be worth to take measures for enhancing the users awareness since significant increase may happen in the level of cyber security.

8 Conclusion

In this paper the conditions and the findings of a simulation study was presented for assessing the effect size of users' consciousness to the computer network vulnerability in risky cyber attack situations at a certain business.

First a simple model was set up to classify the groups of users according to their skills and awareness then probabilities were assigned to each class describing the likelihood of committing dangerous reactions in case of a cyber attack.

To quantify the level of danger a metric developed in a former work was used. This $\mathtt{p}_s$ metric shows the approximate probability of an infection at a given business with well specified parameters according to its location, the type of the attack, the protections used at the business etc.

To be able to see the tendencies in vulnerability exclusively attributable to the users consciousness the set of the numerous parameters were grouped to kept-constant and study parameters.

First the "typical study values" of the study parameters were used in the simulations then the sensitivity of the findings was investigated to the deviations from these typical values.

On one hand the findings mirrored back the straightforward and expected tendencies namely either the number of "reeducated" user is increasing or the surmounted difference of the users' groups in their skills is increasing the "relative improvement of the cyber security" is increasing.

On the other hand the tendencies in the change of this relative improvement have been established, different graphs and curves have been constructed to give an overall view for the influence of the different parameters.

In addition to these general conclusions assessments were made for the magnitude and for the range of the relative cyber security improvement. Slight sensitivity was experienced to the departures from the typical study values. It was shown that even in the case of small differences in skills making the users more conscious in their reactions can significantly enhance the level of cyber security at a business.

References

[1] Bognár L, Joós A, Nagy B, An Improvement for a Mathematical Model for Distributed Vulnerability Assessment, *Acta Universitatis Sapientiae, Mathematica*, **10** (2) (2018), 203–217.

[2] Hadarics K, Györffy K, Nagy B, Bognár L, Arrott A and Leitold F, 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. pp. 45–57 (ISBN:978-80-7231-414-0).

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

[4] Leitold F, Hadarics K, 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. pp. 62–66 (ISBN:978-1-4673-4880-5).

Received: September 11, 2019