Behind passwords: An analysis of preliminary results in order to understand how users protect their privacy
by Tibor Roskó and Gergő József Szőllősi

Abstract
Nowadays, the Internet is the most common source of information, it is the means of almost any activity, such as shopping, online banking, getting informed or keeping contacts. Users need to know privacy regulations, read privacy and cookie policies to protect themselves against illegal or unwanted data use. The answers obtained to our questionnaire have been analyzed and they prove our hypothesis that users’ privacy awareness is low, and highlight that users need a deeper education in privacy protection. In our questionnaire people were asked general questions, such as “Do you always read cookie policies?” or “Do you have deep knowledge of privacy regulations?”, and more specific questions, such as “Have you ever used the Hungarian National eHealth Infrastructure (EESZT)?” or “Do you know that Facebook plugins follow all your online activities?”.

Contents

1. Introduction
2. Problem statement
3. Proposal
4. Methods
5. Results
6. Discussion
7. Conclusion

1. Introduction

In the last few years, online presence has significantly increased, which EUROSTAT and NTIA statistics confirm (Roskó, 2020a). Figures 1 and 2 highlight the growing tendency to use online social networks and financial services. Despite these facts, users usually do not pay attention to carefully read privacy policies for sites that they use, such as online shopping applications, news sources or Facebook. Moreover, users do not read policies, and most of the time, unnecessarily share a huge amount of personal information about themselves, as highlighted by Nyoni and Velemíni (2018). A large percentage of Facebook users (75 percent) have shared or often share their geo-location with friends in posts (Figure 3), while 33 percent provide complete personal data and 67 percent share partial personal details (Figure 4). Many posts are primarily status updates, picture sharing and liking. Most respondents think that these actions are safe and do not cause problems, but the truth is that cybercriminals can use this information to build profiles about specific users, for example, daily routines, frequent activities or studies. All of these might also contribute a steady increase in cybercrimes. Identity theft, phishing and information leakage are the most common cybercrimes (European Union Agency for Cybersecurity, 2019). Watters, et al., 2019 emphasized that since 76.5 percent of services apply SMS-based two-factor authentication, the identity of a given sender can be easily guessed or named in a message. In other cases, unnecessarily shared personal information can be the biggest source of successful identity theft attacks,
Behind passwords: An analysis of preliminary results in order to understand how users protect their privacy

embarrassment or stalking (boyd and Hargittai, 2010).

Figure 1: The proportion of online social network use in U.S., EU and Hungary between 2015–2019 (Source: EUROSTAT and NTIA statistics).
Figure 2: The proportion of online financial services use in U.S., EU and Hungary between 2015–2019 (Source: EUROSTAT and NTIA statistics).
Figure 3: Geo-location sharing by users (Source: Nyoni and Velempini, 2018).
Behind passwords: An analysis of preliminary results in order to understand how users protect their privacy

Analyzing recent papers, we found that most of them only inspect the awareness of Facebook users’ over privacy, and only some of them cover a wider part of the online market, such as Android users, platform-independent users or browser extensions. In the terms of privacy awareness, we paid special attention to covering a wide range of privacy aspects. We tried to reach not only Facebook or other social media users in our survey but also identify those focus areas where users need to be educated more intensively in order to raise their awareness of cybersecurity issues.

Having analyzed the preliminary results of our surveys on privacy awareness, we provide a preliminary hypothesis validation since significant measures must appear in small samples as well. This paper presents these results and discusses the commonalities and differences compared to other recent research. Based on these results, this paper proposes educational methods to develop privacy awareness as a final outcome.

2. Problem statement

People do almost everything online, but they do not read carefully — or not at all — privacy policies of the sites that they use. Additionally, they share unnecessarily too much personal information on social media sites, such as images of certificates or documents. All of information can be an intense resource for cybercriminals to successfully perpetrate cybercrimes, such as identity theft and extortion.

3. Proposal
The primary goal of this paper is to validate the hypothesis that privacy awareness is low based on scores from users on privacy and the following criteria:

a. Users know that there are privacy regulations, but they do not know their details.
b. A huge part of users do not read privacy policies on sites.
c. Despite national propagation and support, the Hungarian National eHealth Infrastructure (Elektronikus Egészségügyi Szolgáltatási Tér or EESZT) is not well-known among Hungarians, which significantly highlights that most might have not made privacy settings for their medical records.

4. Methods

In this study, we applied mixed research methods, such as qualitative and quantitative methods as discussed in Addo and Eboh (2014). As a first step, we analyzed our survey results to prove or discard our hypothesis that privacy awareness is low.

Data were collected from a randomly selected high school in Debrecen via a paper-based questionnaire and the University of Debrecen via an electronic questionnaire in 2019. To filter out contradictory answers, records which meet the following conditions were deleted:

1. Validation A: Q2 = A1 AND Q3 = A1 or A2,
2. Validation B: (Q6 = A1 AND Q9 = A2) OR (Q6 = A1 AND Q7 = A2).

We analyzed the dataset in order to generate descriptive statistics, summarized in Table 1.

<table>
<thead>
<tr>
<th>Age group</th>
<th>SUM before erasure</th>
<th>Validation A</th>
<th>Validation B</th>
<th>A+B</th>
<th>SUM after erasure</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school</td>
<td>74</td>
<td>8</td>
<td>15</td>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>University</td>
<td>199</td>
<td>8</td>
<td>11</td>
<td>0</td>
<td>180</td>
</tr>
</tbody>
</table>

After cleaning the dataset of contradictory answers, chi-square tests were used to indicate if there were significant differences between answers and to examine if there were relationships between age groups (high school or university students) and answers. All statistical analyses were performed with SPSS, version 26, and we considered our results significant if the \( p \) value was less than 0.05. The abbreviation Q stands for the question and abbreviation A for answer; for example, Q1 = A1: Question 1 = Answer 1. In order to analyze sharp differences, we grouped answers, the method of grouping detailed in the case of the specific question. By this step, all the variables, except score, are dichotomous because they only consist of yes and no. All the datasets and questionnaires are available in Roskó (2020b).

5. Results

This section presents the results of our survey analyses on privacy awareness as detected by a knowledge about privacy regulations as well as the Hungarian National eHealth Infrastructure.
5.1. Scores on privacy awareness

The privacy awareness scores were counted with a specific set of criteria detailed in Roskó (2020b); Table 2 summarizes the scores grouped by age group (high school and university students), and Figure 5 shows the proportion of good (9–12 points) and bad (0–8 points) classes.

<table>
<thead>
<tr>
<th>Score</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school</td>
<td>1.85</td>
<td>0.00</td>
<td>1.85</td>
<td>1.85</td>
<td>5.56</td>
<td>18.52</td>
<td>12.96</td>
<td>1.85</td>
<td>22.22</td>
<td>11.11</td>
<td>14.81</td>
<td>5.56</td>
<td>1.85</td>
</tr>
<tr>
<td>University</td>
<td>2.22</td>
<td>1.11</td>
<td>12.22</td>
<td>3.33</td>
<td>18.33</td>
<td>12.78</td>
<td>13.89</td>
<td>6.11</td>
<td>12.22</td>
<td>3.89</td>
<td>5.00</td>
<td>7.22</td>
<td>1.67</td>
</tr>
</tbody>
</table>

Table 2: Scores on privacy awareness (yellow: the most commonly achieved score, bold: the highest score achieved in class ‘bad’ and green: the highest score achieved in class ‘good’; all values in percent)

Figure 5: Proportion of privacy awareness scores whether user scored 9 points or not.

To precisely determine if privacy awareness was low, we summarized scores in Table 2. For high school students, a
score of 8 was the most commonly achieved value, which is also the most commonly achieved value in class ‘bad’. In the case of university students, the score of 4 was the most commonly achieved value, also the most commonly achieved value of class ‘bad’. Figure 5 shows the proportions of classes ‘bad’ and ‘good’, highlighting that a large number of both groups of respondents could only score less than 9 points. The threshold value of privacy awareness was about 75–80 percent (9 points) of a total of 12 points, where it could be said that a user was conscious of privacy protections. A large number of respondents did not have enough knowledge to protect their privacy. University students differed significantly; χ²(12)=24.795, p=0.016, ϕ_c=0.326. To prove that the results were accurate, we analyzed the datasets in more detail.

5.2. Only a low percentage of users consistently read privacy policies

We analyzed how many users always read privacy policies, another key test to prove or discard our hypothesis that privacy awareness was low.

Most of our respondents confessed to be aware of privacy protections, but high school students appeared to be significantly more aware than university students (Figure 6); χ²(1)=10.466, p=0.001, ϕ_c=0.211. Unfortunately, these results were refuted by answers to Questions 4 and 6. In the case of reading cookie policies, a significant proportion of the respondents said that they did not read them because they were not interested; χ²(3)=1.161, p=0.762, ϕ_c=0.070. Only 3.70 percent of high school students and only 1.67 percent of university students always read cookie policies (Figure 7). Figure 8 shows a summarized result (yes=A1, A2 and no=A3, A4) of answers to Question 4 if the respondents read cookie policies. These results highlighted that the proportion of answers was almost equal for both groups of respondents; a huge part of them (about 80 percent) never read cookie policies; χ²(1)=0.001, p=0.976, ϕ_c=0.002. The summarized results (yes=A1, A2 and no=A3) of reading Facebook registration terms also carried the same negative results because about 60 percent of the respondents had not read the terms (Figure 9). Significant differences between the high school and the university students could not be demonstrated; χ²(1)=0.878, p=0.349, ϕ_c=0.062. Results in Figure 9 came from Facebook users only (Q5 = A1), but we also analyzed non-Facebook users. Only one secondary school student did not have a Facebook account, but the student had read registration terms (Q6 = A2), and the student was aware of privacy protection (Q10 = A1), while university students were somehow different, i.e., three students did not have a Facebook account and none had read registration terms (Q6 = A3), but one was aware of privacy protection (Q10 = A1).
Figure 6: Proportion of Question 10: Are you aware of what personal information is being processed by online systems that you use and for what purposes?
Figure 7: Proportion of Question 4: Do you usually read cookie policies?
Figure 8: Proportion of Question 4, grouped into “yes” and “no”: Do you usually read cookie policies?
Answers to Questions 4 and 6 refuted the hopeful result that a significant proportion of users were aware of privacy protection, such as reading registration terms, policies and having some knowledge about safe information sharing. Most respondents confessed to being aware of privacy protections, but the sad reality was that they did not care to read policies or terms.

5.3. The proportion of knowledge about privacy regulations

On average, more than 85 percent of both groups of respondents knew that data protection regulations existed (Figure 10): $\chi^2(1)=0.382$, $p=0.536$, $\phi_c=0.040$. 

Figure 9: Proportion of Question 6, grouped into “yes” and “no”: Did you read Facebook’s privacy policies?
We only analyzed those summarized answers (yes=A1 and no=A2, A3) to Question 3 where the answer to Question 2 was “yes”. We found that a large number of respondents did not have deep knowledge about privacy regulations. A significant difference between high school and university students was demonstrated in a question about these regulations. This deficiency was extremely high only for university students with 74.69 percent answering “no”; \( \chi^2(1)=9.573, p=0.002, \phi_c=0.214. \)

Analyzing answers to Question 15, we found that a low percentage of respondents participated in some form of privacy protection education, 18.52 percent of high school students and only 9.44 percent of the university students (Figure 11). Significant differences between high school and the university students could not be demonstrated; \( \chi^2(1)=3.351, p=0.067, \phi_c=0.120. \)
Behind passwords: An analysis of preliminary results in order to understand how users protect their privacy

Figure 11: Proportion of Question 15: Have you attended classes on privacy protection?

5.4. Most of the respondents were unaware about EESZT

EESZT has been in use since 2017, allowing users to precisely set privacy settings to control access to medical records. There are three options for restrictions: the first is a total restriction, with no access to medical records, the second is full permission where everyone has access to medical records, and the third is when specific rules are set for specific medical staff to access records. The default setting allows medical staff to access information, except psychiatric, addiction and STD medical records which can only be accessed by doctors of the same specific medical specialty.

Analyzing the summarized (yes=A1, no=A2, A3) answers to Question 13, we found that most respondents (98%) had never used EESZT, so most had not made privacy settings in EESZT (Figure 12). Significant differences between high school and university students could not be demonstrated, both equally had never used EESZT to set privacy settings; $\chi^2(1)=0.008$, $p=0.927$, $\phi_c=0.006$. 

\[\text{Figure 12: Proportion of Question 13: Have you made privacy settings in EESZT?}\]
Analyzing unsummarized answers to Question 13, we found that most respondents (more than 70 percent) were aware about EESZT; $\chi^2(2)=0.215$, $p=0.898$, $\phi_c=0.030$ (Figure 13).
Behind passwords: An analysis of preliminary results in order to understand how users protect their privacy

6. Discussion

Comparing the results of prior research on privacy awareness to the findings of this study, several significant commonalities could be observed. This section contrasts these commonalities to each other to confirm hypotheses based on our data.

We started our hypothesis validation with an analysis of privacy awareness scores, where we found that both high school and university students, in most cases, could only achieve less than 9 points, meaning that their privacy awareness was low. To precisely investigate the results, we analyzed each related question step-by-step. Having knowledge on privacy regulations exist could be most informative in determining a level of awareness.

Results of the survey done in Nyoni and Velempini (2018) highlighted that a large percentage of Facebook users shared or often shares their geo-location with friends in posts (Figure 3), with 33 percent sharing personal data fully and 67 percent partially (Figure 4). The types of posts were primarily status updates, picture sharing and liking. Most of the respondents thought that these actions were safe, but cybercriminals could use this information to build profiles about specific users. This information indicated that users did not care to set stricter personal protection on Facebook, since 88 percent of Facebook users elected not to set specific or stricter privacy options (Figure 14). We found similarities related to this in the case of Question 13 on how few people had ever worked with EESZT to check medical records or make privacy settings, less than two percent of respondents. Johnson, et al. (2012) and Malik, et al. (2016) noted that photos shared on the Internet leak private and confidential information that users never intended to publicly share. In 2009, nine percent and in 2010, only two percent of all Facebook users made privacy settings. All of these

Figure 13: Distribution of EESZT use, Question 13.
results were significantly dependent on Internet skills because 12 percent of those who had low skills and five percent of those who had high skills did not make settings as noted by boyd and Hargittai (2010). The impact of skills in cybersafety was confirmed by Robinson, et al. (2020).

![Pie chart showing privacy settings usage on Facebook](Source: Nyoni and Velempini, 2018)

**Figure 14:** Privacy settings usage on Facebook (Source: Nyoni and Velempini, 2018).

In the case of reading privacy policies or terms of use, we also found commonalities. Alani (2017) measured awareness on permissions for applications with 13 questions, answered by 4,027 respondents. Only 35.71 percent of users always read these permissions (**Figure 15**). In the case of our results, despite the fact that a huge number of the respondents were aware of privacy protection (**Figure 6**), only 20 percent always read privacy policies (**Figure 8**). Only 37 percent used antivirus software (**Figure 16**), and a significant part of the users installed applications that required excessive permissions.
Figure 15: Frequency of permission reading (Source: Alani, 2017).
We might say that respondents in our survey were careless over privacy, especially, in the terms of privacy policies for those sites used on a daily basis.

7. Conclusion

It was obvious that privacy awareness was low among respondents to our survey and that was a need for education on privacy. This need for education was also noted in Nyoni and Velempini (2018) and Giannakas, et al. (2019).

A significant number of respondents confessed to being aware of privacy needs (Figure 6) and privacy regulations (Figure 10). However, a significant number almost never read privacy policies (Figures 7 and 8).

From another viewpoint, the most possible reason for a low proportion of stricter privacy settings in EESZT was that EESZT was nearly unknown among respondents (Figure 13).

Education is a key component, utilizing the Internet as a resource of information (Hopp and Sheehan, 2019). Privacy behavior of friends and colleagues is also an important factor (boyd and Hargittai, 2010). Educating users to develop their privacy awareness is a crucial task in preventing any future growth in personal cybercrime.

About the authors

Tibor Roskó, certified software engineer, obtained his diploma at the Faculty of Informatics of the University of
Debrecen, in 2017. He is currently a Ph.D. student at the Doctoral School of Informatics at the University of Debrecen in Hungary. His field of research is the development of the abstract model of global and centralized user authentication and examining the possibilities of its introduction. A subfield of his research is inspecting the possibilities of cooperation between separately operating systems, and supporting the practical implementation of data protection regulations. He took part in the DETEP Talent Program at the University of Debrecen and earned a national higher education scholarship. He is a member of the Information Security Section of John von Neumann Computer Science Society (Neumann János Számítógép-tudományi Társaság). Web: www.rtibor.hu

Gergő József Szollosi, a public health professional with a specialization in epidemiology, obtained his M.Sc. diploma at the Faculty of Public Health of the University of Debrecen in 2015. He is currently an assistant lecturer at the Family and Occupational Medicine in the Faculty of Public Health. His main field of interest is identifying protective and risk factors that might influence vaccination uptake in Hungary.
E-mail: szollosi [dot] gergo [at] sph [dot] unideb [dot] hu

Acknowledgements

We thank the support of Gyöngyi Bujdosó (University of Debrecen), Nyilas Istvánné (University of Debrecen) and Andrea Pákozdy (University of Debrecen). This research was partially supported by EFOP-3.6.3-VEKOP-16-2017-00002. The project was supported by the European Union, co-financed by the European Social Fund.

We wish to confirm that there are no known conflicts of interest associated with this study and there was no significant financial support for this work that could have influenced its outcome.

References


Behind passwords: An analysis of preliminary results in order to understand how users protect their privacy

364–382.


---

**Editorial history**

Received 19 April 2020; revised 13 October 2020; revised 1 November 2020; accepted 4 January 2021.

---

This paper is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Behind passwords: An analysis of preliminary results in order to understand how users protect their privacy by Tibor Roskó and Gergő József Szőlősi. First Monday, volume 26, number 8 (August 2021).
doi: http://dx.doi.org/10.5210/fm.v26i8.10616