EXAMINATION OF PERSONAL DATA AVAILABILITY ON THE INTERNET

Bitinaitė R., Liuiza A.
Kaunas University of Applied Sciences

Abstract

Availability of sensitive personal data online is one of the leading causes for identity theft and hacking by social engineering. An experiment was carried out to examine and evaluate how much of personal information is shared online by Lithuanian youth. Results show that, on average 4-5 pieces (out of 10 usual pieces looked for in this experiment) of personal information can be found about a person in 15 minutes of searching online.

Keywords: *personal data, privacy, online security, cybersecurity, Internet privacy;*

Introduction

Every day, millions of people use the Internet – browse, search for information, send email, participate in discussions, connect via social networks. Most of them have no idea about the dangers to their privacy that arise while being online. Privacy and personal data protection on the Internet is an area, now undergoing mayor shifts. United States Government Accountability Office. Information security in 2014 has issued a report, that over the period from 2009 to 2013, number of personal data breaches online has almost doubled, with 3/4 of them happened online. The phenomenon of social networks has significantly encouraged the spread of personal data online, and shortcomings of privacy policies of said networks is one of the lessons for such prolific number of breaches.

Social networks were created to encourage communication and sharing of information. People who are connected to social networks often publish a lot of personal data there, without ever thinking how this data will be handled. EU General Data Protection Regulation (GDPR), that is coming into effect this year, aims to force parties that are handling personal data to be more transparent and careful. Lithuanian law states the same data handling principles – personal data has to be collected and handled for defined and lawful purposes, and not to be handle in ways that were not defined at the time of data collection.

In practice, people, especially youth, rarely thinks about what types of personal information they share online and accessible to all kinds of people and organizations, with unknown intentions. Such information can be used in identity theft, social engineering, when cyber criminals try to pretend to be someone else online with either the goal of getting access with some more senistive systems, like online banking, or to cover their own identity in other cyber crimes.

European Parliament is seeking meaningful reform to achieve better privacy on the Internet. Currently, information that is published online, stays online – it is very hard to remove it, it can be accessed and used several time by a lot of different people. There is a hope that GDPR rules will strengthen persons' rights to privacy online, and will make it easier to prosecute cyber criminals who make use of such information in their crimes.

Research aim – to investigate and evaluate, how much personal data can be found online about particular set of young Lithuanian people.

Research object – personal information of participants of this experiment, that is available publicly or with minimal permissions in social network.

Research problems:

1. Collect publicly or with minimal permissions available personal data of participants of this experiment.

2. Analyze and evaluate the collected data.

Methodology

On January 17th, 2018 (Kaunas) and February 8th-10th, 2018 (Vilnius) a national fair of institutions of higher education „Studijos 2018" took place. During this fair at the booth of Kaunas University of applied science an experiment has been carried out. Visitors were encouraged to take place at a prize draw – they could win a branded thermo-cup or some other simple prizes. All they gad to do was to scan a prominently displayed QR code and register as a participant using their Facebook profile. It is not an unusual request – a lot of sites on the Internet use Facebook profiles to identify their users. After that, participants were asked to come back in 20 minutes to claim their prize.

After users completed the registration, a custom computer program would read all data that is available via Facebook API with that simple permission – first and last name, email address, in some cases – phone number, information about participants' family, etc. After that a team of students from Kaunas University of Applied Sciences Technology faculty Multimedia technology study program would manually check that data and amend it using information, found by doing simple Google search queries.

10 particular pieces of information were being looked for:
- First and last name
- E-mail address,
- Phone number,
- Date of birth,
- Place of residence (city, address),
- Family members,
- Relationship status,
- School / Employer,
- Name of a pet,
- Mother's maiden name.

Students had 15 minutes, counting from person's registration to complete this task. After that, all collected data would be put into a report and printed out at the fair booth. This report was presented to the participants, with explanation of the experiment t hey have taken part in. Depending on amount of information that was found, different prizes were given to every participant.

Results

One of the things that were monitor, were bounce rates – how many people actually completed the registration process. Out of 688 unique visitors who scanned the QR code, 58.9 percent completed the registration and took part in the experiment.

Number of participants

● Kaunas ● Vilnius
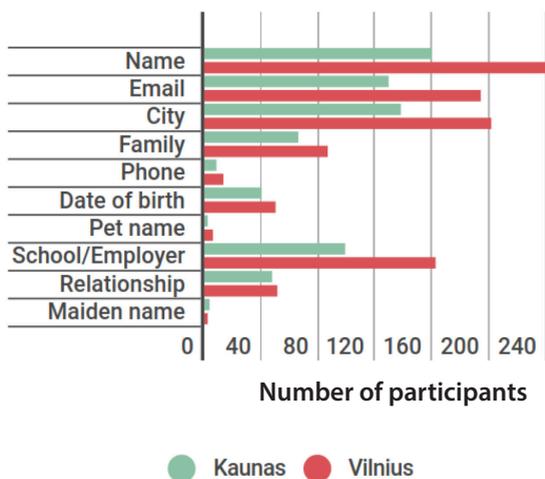
*Picture 1. Collected data by information groups*

Statistical data (Picture 1) shows, that name, city and e. mail were the easiest to find, with 100, 83.7 and 79.8 percent found, respectively. Hardest to find were phone number (5.2 percent), pet's names (2.0 percent) and mother's maiden name (1.7 percent). Every piece of information was found in at least 7 cases, so neither of the fields was impossible to find.

Picture 2 displays how many fields of information were found about particular participant. Most often, 4 (28.7 percent of cases) and 5 (26.7 percent) fields of information were found. Most rare result was 1 field (1.1 percemnt) and 8 fields (0.7 percent) were collected. 2 fields were found in 9.8 percent of the cases, 3 fields – 17.8 percent. 6 and 7 fields were colected for 11.7 and 3.5 percent of participants respectively. No more than 8 pieces of information were collected for any specific participant. This shows that while people are not sharing everything, a lot of information can still be found online even with a very short and cursory search.
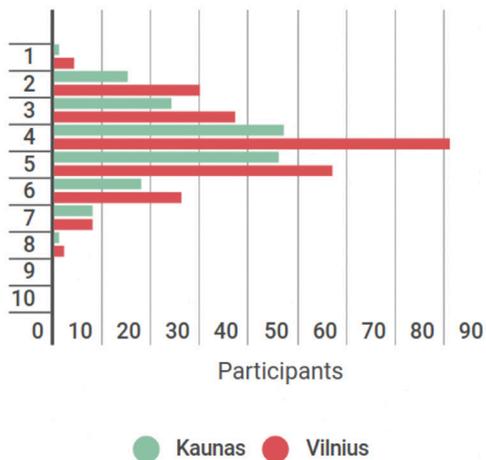
Picture 2. Number of information fields found, per participant

Prize statistics (Picture 3) show, that the most common was the intermediate prize – a notebook. It was presented to participants who had from 3 to 5 pieces of information available online. If participant shared more that 5 items of information, they would get a pen. About 60 of this smallest prize were distributed. And only person who scored well in this experiment (less than 3 pieces of information available) got the thermo-cup that was advertised as the main prize. In total, 16 cups were given away.
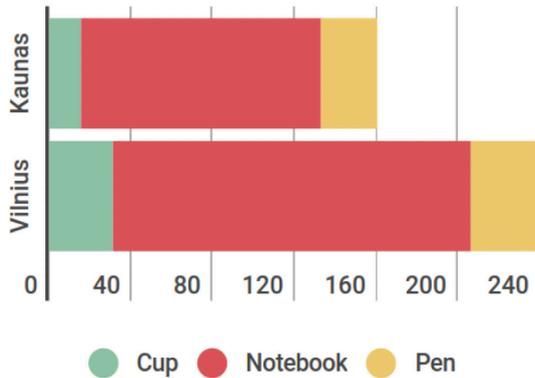
Picture 3. Prize statistics

Conclusions

1. Results of this experiment show, that 90 percent of participants share more that 3 pieces of information available online publicly or with minimal permissions. And they were willing to give access to it for a chance of a fairly cheap and simple reward.

2. Least sensitive data – name, email, city – were shared the most (more than 70 percent of the cases. Most sensitive data, like popular security questions (pet's name, mother's maiden name) were amongst the best guarded, but students still managed to find those in at least 7 cases.

3. The results show that Lihuanian youth has a quite average approach to guarding personal data.

References

1. United States Government Accountability Office. Information security: Federal Agencies Need to Enhance Responses to Data Breaches (GAO-14- 487T), 2014, p. 2–4 [interactive, accesed 2018-03-20]. <https://www.gao.gov/assets/670/662227.pdf>.

2. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [interactive, accesed 2018-03-20] <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679>