Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

OSINT in the Slovak Republic

Bachelor Thesis

2020
Michal Sládeček

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

OSINT in the Slovak Republic

Bachelor Thesis

Study Programme:	Computer Science
Field of Study:	Computer Science
Department:	FMFI.KI Department of Computer Science
Supervisor:	doc. RNDr. Martin Stanek PhD.

Bratislava, 2020
Michal Sládeček

22995305

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta:	Michal Sládeček
Študijný program:	informatika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor:	informatika
Typ záverečnej práce:	bakalárska
Jazyk záverečnej práce:	anglický
Sekundárny jazyk:	slovenský

Názov: OSINT in the Slovak Republic
OSINT v prostredí Slovenskej republiky

Anotácia: Preskúmať možnosti OSINT (Open-source intelligence) v prostredí Slovenskej republiky. S využitím znalostí slovenských reálií identifikovať špecifické zdroje voľne prístupných informácií o fyzických osobách, spoločnostiach a IT infraštruktúre. Porovnať s klasickými "globálnymi" zdrojmi informácií. Navrhnúť vhodné postupy pri získavaní informácií a zhodnotiť možnosti automatizácie.

Vedúci:	doc. RNDr. Martin Stanek, PhD.
Katedra:	FMFI.KI - Katedra informatiky
Vedúci katedry:	prof. RNDr. Martin Škoviera, PhD.
Dátum zadania:	03.10.2019

Dátum schválenia: 24.10.2019

doc. RNDr. Daniel Olejár, PhD.
garant študijného programu

.... študent

.... vedúci práce

22995305

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Michal Sládeček
Study programme:	Computer Science (Single degree study, bachelor I. deg., full time form)
Field of Study:	Computer Science
Type of Thesis:	Bachelor´s thesis
Language of Thesis:	English
Secondary language:	Slovak

Title:	OSINT in the Slovak Republic
Annotation:	Explore the possibilities of OSINT (Open-source intelligence) in the Slovak Republic. Identify specific Slovak sources of freely available information about individuals, companies and IT infrastructure. Compare the sources with classic "global" sources of information. Propose suitable information retrieval procedures and evaluate the possibilities of automation.

Supervisor:	doc. RNDr. Martin Stanek, PhD.	
Department:	FMFI.KI - Department of Computer Science	
Head of department:	prof. RNDr. Martin Škoviera, PhD.	
Assigned:	03.10.2019	
Approved:	24.10.2019	doc. RNDr. Daniel Olejár, PhD. Guarantor of Study Programme

....

Student Supervisor

Poďakovanie: Chcem sa poďakovať svojmu školiteľovi doc. RNDr. Martinovi Stanekovi, PhD., za jeho rady a veľkú snahu o to, aby táto práca vyzerala čo najlepšie. Tiež ďakujem doc. Mgr. Bronislave Brejovej, PhD., za jej užitočné materiály ktoré mi pomohli s technickými časťami písania práce.

Abstrakt

V tejto práci sumarizujeme otvorené zdroje informácií v Slovenskej Republike. Venujeme sa zdrojom o osobných údajoch, organizáciach a internete. Pre jednotlivé zdroje uvádzame príklady ich využitia. Práca hodnotí využiteľnosť zdrojov predovšetkým z pohľadu informačnej bezpečnosti, no mnohé poznatky sú aplikovateľné aj v iných oblastiach (napr. v žurnalistike). V práci sme tiež navrhli automatický nástroj na hľadanie spriaznených webstránok na slovenskom internete.

Kľúčové slová: OSINT, otvorené zdroje, penetračné testovanie, red teaming

Abstract

In this thesis, we summarize open sources of information that are available in the Slovak Republic. We discuss sources about personal information, organizations and the internet. The thesis evaluates these sources from the perspective of information security, but many sources are applicable in other areas (for example, in journalism). We also created an automated tool to search correlated webpages on Slovak internet.

Keywords: OSINT, Open sources, penetration testing, red teaming

Contents

List of links

All the links were accessed on 27.5.2020.

Introduction

Open Source Intelligence (referred to as OSINT) is the process of gathering, processing, and analyzing information from publicly available information [18].

Before the internet, OSINT consisted mostly of analyzing press and public transmissions like radio and television. It is believed that OSINT accounted for the biggest part of military intelligence on the Soviet Union. For example, various intelligence services used information from the Soviet periodical released by Soviet Ministry of Defence *Krasnaya Zvezda* [16].

In recent years, the internet contributed to huge changes in how OSINT is gathered and processed [15]:

- **Collection got easier.** With the arrival of search engines, it is easier to search for an information than it was during the Cold War. The collection is so easy that performing OSINT is no longer a privilege of Intelligence Agencies. Some journalists investigate almost purely using public sources, for example, the Bellingcat[1].
- **Volume of data increased.** This is both an advantage and disadvantage because the amount of data nowadays is too big to be processed. A common technique to overcome this problem is automation, for example web scraping. Not all investigations can be automated though. Another way to resolve the issue of data volume is to crowdsource the investigation. For example, Europol crowdsourced investigation of child abuse by sharing photos of places or objects [9].
- **Source validation became more important.** The internet allows publishing anything anonymously. This creates a problem for all professionals using OSINT, as they can not usually rely on a single source and need to verify most of the information on the internet.

Because of these changes, OSINT became useful outside of intelligence agencies. Common performers of OSINT are:

- **Penetration testers** use OSINT when searching for weak spots in the target's infrastructure [14].

[1] `https://www.bellingcat.com/about/`

- **Investigative Journalists** use OSINT during their investigations, for example when researching public procurements [23].
- **Business analytics** use OSINT when analyzing competition, markets, or customers. For example, analysis of competitor's offers is used to find market holes [20].
- **Recruiters** use OSINT to find people who match the employers's requirements.
- **Common people** uses OSINT daily. Reading reviews of an e-shop before buying something or looking through someone's Facebook profile to find where the person moved are examples of OSINT.

About this thesis

The procedure of OSINT is very specific to each country. By using only the world-wide sources, it is often not possible to dig into the necessary depth when investigating a subject.

The goals of this thesis are to:

- Analyze sources of OSINT available in Slovakia.
- Suggest approaches that utilize these sources.
- Evaluate the approaches on real targets and demonstrate how the Slovak specific sources helped in investigations.

For the scope of this thesis, we limit the scope of OSINT from any publicly available information only to the information available on the internet. This excludes OSINT sources like nespapers, broadcasts or information requests to public organizations.

We do not use any paid tools. When we mention paid tools in this thesis, we always just inform about what these tools claim to be able to do. This might not reflect the real abilities of these tools.

For better clarity, we usually use the word *investigator* to refer to the person that performs OSINT. Sometimes, when we believe that the technique is highly useful in a specific area, we specify the person's profession.

The thesis is divided into the following chapters:

- The chapter **Personal information** shows how to collect information about people. This includes information like are address, phone number, or date of birth.
- The chapter **Organizations** shows how to collect information about both public and private organizations. This includes public procurements.
- The chapter **Webpages** shows how Slovak sources of domain data can be utilized in various investigations.

- The chapter **Procedure** contains an OSINT procedure that can be used by penetration testers before an red teaming assignment. The procedure and the results are shown on an example target.

The thesis also has scripts and tools that we used attached.

Current state of research

Most resources we found about OSINT are usable either worldwide or only in the USA. The book *Open Source Intelligence Techniques* by Michael Bazzell is a very thorough reference of sources. Many of the detailed techniques can be used globally. There are also many guides on the internet, for example, the *Bellingcat's Online Investigation Toolkit*[2]. We used some techniques from these sources and compare the results with specific Slovak sources.

In Slovakia, we did not find much research about OSINT. All the found articles were targeted at investigative journalists. These articles helped us with the section 2.2 about public procurements.

- *Práca novinára s otvorenými zdrojmi a dátami* by Mgr. Ján Hacek, PhD. [10]
- *Breaking the Public Procurement Act and open sources* by Patrícia Voľanská [23]

[2] `https://docs.google.com/document/d/1BfLPJpRtyq4RFtHJoNpvWQjmGnyVkfE2HYoICKOGguA/edit`

Chapter 1

Personal information

This chapter is about investigating individuals. In first section we discuss techniques that are useful in following scenarios:

- Find information about a given individual.
- Find an unknown individual from a piece of information about the person.

In the second part of this chapter we show a real-life investigation and analyze, how useful were the techniques from the first part. We also show some interesting points that the reader can take from these investigations.

An investigator needs to be careful with publishing or acting on any personal information. The results of the investigation should always be taken only as leads in further investigations as they do not always show the whole picture. For example, we heard of a case in which the owner of a company did not change the registered address of his company after selling the property. An investigator almost published an article about the connection of the unknowing new owner to illegal activities conducted by the company. In this case, the investigator asked the new property owner before publishing the article, so no harm was done. It is still a good example of how an open-source investigation can reach the wrong conclusions and cause potential harm to people.

1.1 Information gathering

Each part of this section assumes a scenario in which the investigator has a piece of information. We organized this chapter in such a way because it is consistent with how most OSINT investigations are performed. For example, a fictional investigation of possible connection between political party and a fake-news website can look like this:

1. Investigator finds a contact phone number on an extremist website.
2. Investigator finds a listing that shows the owner of the phone number.

3. By searching for the name, the investigator can find the permanent residence of the person.
4. Investigator then finds a company that is registered on the permanent residence of the owner.
5. Transparent account of the political party shows, that the company is getting paid for advertisement from them.

Following a procedure will sometimes not be enough to fulfill the investigation requirements. The investigator often needs to be very creative when analyzing the found information and it is impossible to create a procedure for every possible investigation scenario. We provide procedures for scenarios where investigator already has one of the following pieces of information:

- Name
- Address
- Phone number
- Online handles and emails
- Evidence number of vehicle

Use of search engines is a common technique that is useful in all of the following scenarios. The information found by search engines often comes from very unique sources that can not be all manually checked. As an example, once we were able to find information about a person including name, birth number, phone number, and address just by googling the target's email address. The information was located on a small webpage inside a scan of a contract between the target and an insurance company.

Name

The first step for the investigator is usually looking through social networks. From our experience, the most used ones are Facebook, LinkedIn, and Instagram. It is always worth to look for the target there. Some Slovaks also use the Russian social network VKontakte. The techniques for searching these social networks are not Slovak specific, therefore we will not discuss them here. From Slovak specific social networks, the most used is *pokec*[1.1] . We looked into this network and from viewing the public chatrooms, it seems that the users are mostly looking for dating or prostitution. It depends on the focus of investigation whether this network is worth looking into.

Website *kdeje*[1.2] provides a search for the name across multiple sources in the Slovak Republic. This might seem like an easy shortcut but we do not recommend the investigators to use this website. The website shows a list of people who were searched, and this could compromise the operation security.

Very powerful tool to learn information about a person is the real estate registry[1.3] . It is not possible to search real estate registry only by name, the investigator needs to know at least approximate location where the person owns a property. Once the investigators find a property that the person owns, they can find following interesting information in the real estate registry:

- Date of birth
- Maiden name (in case the person is female)
- The floor and flat number
- Permanent residence of the person
- Date when the person obtained the property
- Husband or wife of the person can be found if the person owns a property together with a partner.

There is no public registry that contains all people. We identified that information about person can be found in any of the following conditions:

- **The person owns a company or sits in its statutory body.** The Business Registry of Slovak Republic shows people along with their addresses of permanent residence[1.4] .
- **The person is self-employed.** All self-employed people have their addresses listed in the Trade Registry[1.5] .
- **The person has a lien.** All liens are listed in the Registry of Liens along with addresses and dates of birth[1.6] .
- **The person has debt.** People who did not pay social insurance, health insurance, or taxes, are listed in their registry of debtors. There is a webpage that collects and aggregates data from all these registries[1.7] . If the person's debt is against a private subject, the subject can add him to a different webpage[1.8] that shows him until he pays the debt. If the debts caused the person to go bankrupt, the person is listed in the insolvency registry[1.9] .
- **The investigator knows in which municipality the person owns a property.** In this case, it is possible to use the real estate search tool[1.10] .
- **The person is a public functionary.** Many public functionaries are required to publish their property declarations. However, they often find ways to bypass this requirement[1]. The declaration can be found either on the website of the institution or the website of the National Council of the Slovak Republic[1.11] .

[1] For example by publishing the declaration on private intranet (`https://domov.sme.sk/c/6897153/niektore-vysoke-skoly-taja-majetok-rektorov.html`) or by locking it in safe(`https://bratislava.sme.sk/c/20921536/majetkove-priznanie-primatora-je-v-trezore.html`)

- **The person is listed in an online phone directory.** Phone directory gets data directly from cellular service providers. It contains people who agreed to have their data shared by these providers. The data there contains phone number and address of permanent residence[1.12] [1.13] .
- **The person owns a property.** If the person owns a property, this information is in the real estate registry. Even if the investigator does not know the city where the person owns a property, it might be possible to find the property owned by the person. The investigator would need to make multiple guesses, for example, the city where the person or the city where most of the person's friends live.
- **The person had a contract with the government.** Public institutions publish almost all contracts on the Internet. The information that can be found there depends on the organization that publishes the contract. Some organizations redact the contracts to a degree where only the name remains from the personal information of the person. Nevertheless, sometimes investigators can find a contract that is not redacted or poorly redacted. More information about contracts is available in chapter 2. The contracts might expose even phone numbers or emails.

There are other sources we recommend to check. If the investigator knows both name and date of birth, it is possible to search for distraints[1.14] . Some of the registries of the Ministry of Internal Affairs[1.15] show a person's date of birth and address, but most of these registries are not searchable by name[2]. The only way to search them is to use Google query `inurl:"ives.minv.sk/rez/registre/" NAME`. During our investigations, we found that Google does not have these registries indexed well though, so this query does not always find the information even if it is there.

Address

If the investigator has an address, it is usually easy to find who owns the property. The best resource we found for this purpose is the map of real estate registry[1.17] . The only difficulty arises when the person that owns the property has permanent residence elsewhere, as this might mean the person lends the property. We did not find a way to find the person that lives there in such a case.

The investigator should also look at whether there are some companies or non-governmental organizations registered on the address. The source that we use for this purpose is the map on *verejne.digital*[1.18] .

[2] An exception is registry of non profit organizations, which has a search by natural person[1.16]

Telephone number

It is possible to search for number on some listings websites[1.19] [1.20] . Finding a listing with the phone number can be a breakthrough in the investigation, as it might show (depending on what the user provided) email, address, and name. However, the search on these websites performs exact strict matching, so the investigator needs to perform multiple searches. If the searched number is `09123456789`, we recommend to search at least for `+4219123456789`, `4219123456789`, `09123456789` and `9123456789`.

There are many websites that offer search for phone numbers. These websites contain phone numbers that people reported and their purpose is to warn people from fraudulent or annoying calls, so the chance of finding our target is small[3].

The investigators can also try to search the online phone directories for numbers, even though they don't support the functionality. The phone directories are sometimes indexed with google and information can be found with this google query: `site:zlatestranky.sk site:telefonny.zoznam.sk "+421XXXXXXXXX"`.

Usernames and emails

Online handles are very useful for investigators even if the person's name is known. There might be thousands of John Smith's online, but only one `jsmith777`. When investigating online crime, the objective is to find the real-life personality of a known username or email. If the person takes precautions like using a different handle, email, and password on each website, it might be very difficult to find the identity. The investigations usually exploit some mistakes that the person did.

One common mistake that people do is reusing the same handle or its variations across many websites. While on some webpages they might be anonymous, they might use the same handle on a social network where they use their real identity. By finding the same handle, investigator can connect the two identities. There are multiple tools for searching usernames across various platforms and websites [4], but none of them have good coverage of Slovak websites. We decided to add this functionality to Sherlock. Sherlock is a tool that searches for usernames on more than 300 different websites [8].

[3] For example following websites:

- https://www.vyhladavaniecisla.sk/
- https://www.neznamecislo.sk/
- https://www.ktomivolal.eu/
- https://www.09xy.sk/
- http://zmeskanyhovor.com/
- https://www.090.sk/
- http://zmeskanyhovor.com/

We added 23 Slovak sites on which it is possible to test for username existence to Sherlock. Adding new webpages to Sherlock is fairly easy, all that needs to be done is to extend the *data.json* file. We used the following procedure to find websites on which the existence of username can be verified:

1. We searched google for a username we thought would be commonly used, for example `michal123`[4].
2. On each of the result websites, we tried to find whether it was possible to verify that an username exists. This consisted of trying an username that exists for sure and an username that certainly does not exist and seeing whether there were some noticable differences in responses.

After multiple iterations of the process we found these websites:

- `azet.sk`
- `birdz.sk`
- `brainquest.sk`
- `ceknito.sk`
- `debata.pravda.sk`
- `df.sk`
- `jaspravim.sk`
- `mileneckyvztah.sk`
- `mimiaukcie.sk`
- `mimibazar.sk`
- `modrykonik.sk`
- `modrastrecha.sk`
- `mojevideo.sk`
- `motoride.sk`
- `mtbiker.sk`
- `novalaska.sk`
- `pokec.sk`
- `predajdielo.sk`
- `psickar.sk`
- `rande.sk`
- `taktospravim.sk`
- `tortyodmamy.sme.sk`
- `varecha.pravda.sk`
- `zariadim.sk`

Finding handles can be often easier than finding emails because very few websites are indexed by email. Websites usually hide emails of users for privacy reasons. We still found some widely used webpages that offer search for an email:

- `https://www.bazos.sk/hodnotenie.php`
- `http://sbazar.sk/sbazar/mojeinzeraty.php`
- `https://inzercia.predaj.sk/moje-inzeraty`

A second technique that we use during searches for an email is to look through data leaks. Websites are often compromised and the data from them, for example emails and passwords, is leaked[5]. These leaks can be found on hacker forums[6]. We

[4] Google query used: `site:.sk "michal123"`

[5] Sometimes, but not always, the passwords are hashed. The leaked database of *SKTorrent.eu* contains cleartext passwords and many Slovak emails.

[6] A big collection of leaks can be found on `raidforums.com`

recommend investigators to maintain their collection of leaks. This allows them to not depend on a service that can be taken down at any moment. A useful trick is to search the leaks not only for emails but also for passwords. When investigators find a password belonging to the searched email, they can search for other places where this password was leaked. If the password is not among the common ones, its usage across multiple emails could mean that these emails are used by the same person [3]. More information about leaks can be found in section 4.2. The last technique that is used is trying to reset the target's password on various services [22]. This technique involves starting the password reset process and getting to the point where the website gives the user multiple password reset options to choose from. These options sometimes contain additional information about the target. For example, when trying to reset the password on our account on *azet.sk*, we were presented with the following information:

- The phone number area code
- Last three digits of the phone number
- First two characters of email
- Domain of email

Evidence number of car

If the investigator has an evidence number of the car, he can try to follow this process to find the owner:

1. Find the VIN of the vehicle. We provide two free ways to find VIN:
   - Website *overeniestk.sk* shows the VIN after searching for evidence number[1.21] .
   - Search for the insurance contract of the vehicle[1.22] . When the investigators find the number of the insurance contract, they can search for the contract[1.23] . The contract contains a VIN.
2. Searching the VIN in the registry of liens can reveal the owner when the vehicle has a lease[1.24] .

It is usually not possible to find the owner of the car with this process, and we did not find a better process to do it. The website *stkonline* claims to be able to find owner of the vehicle after sending a SMS[1.25] .

1.2 Case study

Preparations of the experiment To demonstrate some of the techniques from this chapter, we performed OSINT on 10 random people. We decided to not show any personally identifiable information in this thesis, but we describe the process by which we selected our random targets so that anyone can follow and replicate our findings. We chose the people by the following process:

1. Open advanced search on website `https://absolventi.uniba.sk/index.do`.
2. Choose *Faculty of mathematics, physics and informatics* in faculty list, year *2000* from the year list, *Bc.* from the title list and search.
3. Take first 5 men and first 5 women from the resulting list.
4. To preserve anonymity, we sorted these people randomly.

The objectives of the experiment are to find the following information about these people:

- Find the person in source that contains some recent information. This could be any source that shows some information about the person from the last two years. Examples are social networks, employer's websites, or news articles.
- Find the permanent residence of the person.
- Find the date of birth.
- Find phone number.

We also set three degrees of accomplishment for the objectives:

- **Y** when there is a direct link between the information and the target
- **P** when there is a strong connection between the information and the target, but no definite proof (for example the name is the same and the date of birth is similar to what we expect)
- **N** when we did not find any connection to the target.

Investigations process Our initial analysis found the following starting points for investigation:

- Earned degree enables us to filter from multiple people with the same name. We can find the highest achieved title in the graduates' list of Comenius University.
- Because the people earned Bc. degree in the year 2000, we can estimate that their date of birth probably lies between 1975 and 1980.
- Study specialization enables us to guess probable jobs that the person can do. This might be used to filter people with the same name.
- The targets know their classmates from the faculty, and they might have some of them befriended on Facebook.
- The targets will probably have Comenius University listed in their biography.

- Because many absolvents of the faculty find work abroad, we expect that some of our targets do not currently live in Slovakia. Those that stayed are probably located near Bratislava or Kosice city because these cities have the most opportunities for people with this background.

For women, we had to first find the name that the woman uses after being married. We were able to find marital names of 4 out of 5 women. For some of them, we found the names in multiple different ways. The different successful approaches were these:

- We already knew the marital name of one woman because she got married before getting the degree.
- We found that 3 out of 5 women were using both marital and maiden names on Facebook. They all are friends on Facebook, therefore we can be sure that they all are former classmates.
- Google indexed a building permit that contained both woman's maiden and marital names.
- Business registry of Slovak Republic contained historical data and showed the change of name.

Once we found the new name of the woman, the investigation process was the same as for men:

1. Google the person's name
2. Search LinkedIn
3. Search Facebook
4. Go through all the sources from Section 1.1

In each investigation, there were multiple people with the same name. This created a problem and we had to rely on the chance that the two people are the same. For example, there is even a small probability that two people with the same name were born on the same day and both got Mgr. title. The investigator needs to have this in mind when evaluating each new piece of information. This is the reason why we are sure only about two found addresses. Interesting point is that these two addresses both belong to women, and we are sure because the real estate registry shows both marital and maiden names.

Second expected problem was that we did not know, whether the people are still living in Slovakia. It turned out that three out of ten targets are currently living abroad. They were found with help of the following sources:

- A Wikipedia page
- Facebook list of friends of another target

- LinkedIn profile

The third problem was the very low internet presence of two targets. We were able to find records of people with the same name and date of birth in the expected range, but because these people do not have an internet presence, we were unable to verify that these records belong to them.

The most interesting point taken from the investigation is how much information can be mined from the Business Registry. A woman from our targets owned a company and therefore was listed in the Trade Registry. By following historical changes, we could trace her life through the last 10 years:

- We found a date when she disappeared from the company's owners. The next day, a woman with the same first name, but a different address and surname, entered the company. This showed us the approximate date of her marriage. We also know she married to a co-owner of the company, because of the new surname and address.
- We could see the approximate time when she and her husband moved out of a flat into a house because they both changed address.

		Objectives			
		Found	Address	Phone	Birth
Person	1	Y	P	P	P
	2	Y	Y	N	Y
	3	N	N	N	N
	4	Y	P	P	P
	5	Y	N	N	N
	6	Y	Y	N	Y
	7	P	P	P	P

Table 1.1: Fullfilled objectives

Results The results are in Table 1.1. We do not show people who moved from Slovakia in these results. Following list shows how much the sources were helpful:

- Facebook was helpful in 4 investigations.
- LinkedIn was helpful in 4 investigations.
- The Trade Registry of Slovak Republic did not help us with these people.
- The Business Registry of Slovak Republic was helpful in 2 investigations.
- Registry of Liens was helpful in 2 investigations.
- Online phone directory was helpful in 3 investigations.

Chapter 2

Organizations

In this chapter, we show gathering of information about both public and private organizations. There are many different types of organizations, for example

- Companies
- Institutions
- Non-profit organizations
- Political parties
- Foundations

2.1 Sources

Searching for organizations is more straightforward than searching for people because organizations have a public unique identifier, the Organisation Identification number (ICO). This number can be used to search almost any source.

Following sources contain information that can help investigator to understand the organization [10]:

- **Basic information** like associated names, addresses, date of entry or former mergers can be found in various sources, depending on the type of organization. The Business Registry[2.1] contains data on corporations. There are various small registries on the website of the Ministry of Internal Affairs that contain other types of organizations, for example, civil associations[2.2] .
- **Financial data** can be found in the Registry of Financial Statements[2.3] . This data contains financial indicators, for example, the total value of assets that the organization owns, revenue, profit, and expenses. It also sometimes contains a contact phone number and an email.
- **Complex ownership schemes** can be uncovered with help of Registry of Partners of Public Sector[2.4] . This registry contains companies that conduct business

with the government. Some companies have very complex ownership schemes that would take a long time to uncover. This registry always contains the beneficial owners of the company, along with an analysis of the company's ownership. An example of a company with very difficult to understand ownership scheme is *VÁHOSTAV - SK, a.s.*. The proof of ownership references 34 different documents that were needed to prove who owns the company[1]. Some of the documents used for the proof are not public. An investigator would not be able to fully prove who owns this company without the registry.

- **Judicial decisions** can be found in the Registry of Judicial Decisions[2.5] . We recommend to search this source by the name of the company because the ICO numbers are often indexed with spaces in them.
- **Regulated information** can be found in Central Register of Regulated Information[2.6] . The issuers of securities publicly traded on markets regulated by Central Bank of Slovak Republic are required to publish regulated information, like the Annual Financial Report.
- **Information about EU funding** can show whether the company applied for some grants or participated in some projects funded by European Union[2.7] .

Some sources that we already discussed in Section 1.1 can be used to search for companies:

- Insolvency Registry
- Registry of Liens
- Real Estate Registry
- Lists of debtors

Private service Finstat[2.8] aggregates many of these sources. This service offers some data for free, for example, financial data and owner information. Second aggregator is FOAF[2.9] . These aggregators can be very handy if the investigator needs to quickly get an overview a company.

2.2 Public procurements and contracts

In Slovakia, the Freedom of Information Act[2] dictates that almost all contracts by government organizations need to be published online. There are few exceptions from this law, for example:

- Contracts by Slovak Information Service or Military Intelligence

[1] https://rpvs.gov.sk/rpvs/Partner/Partner/Dokument/111291

[2] https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2000/211/

- Employment or service contracts
- Security contracts for prisons
- Contracts by embassies

The contracts can be found in three places:

- Central register of contracts (CRZ)[2.10] contains most contracts by government.
- Some institutions, for example municipalities or the National Bank of Slovakia, must publish contracts on their webpages. These websites sometimes do not have implemented search function, so the investigator needs to search them manually or rely on the contracts being indexed by Google.
- Institutions that do not have their webpage, for example, small villages, publish the contracts in the Commercial journal[2.11]

The most common use of these contracts is to find overpriced purchases. Investigative journalists often do this process. The prices in these contracts might also be interesting for business competitors.

However, the contracts are also a source of other useful information about the owner and the supplier. For example, we were able to find contracts that contained the following pieces of information:

- Emails
- Used software
- Phone numbers
- Names of employees
- Unproperly redacted sensitive information

Public institutions need to perform procurements before ordering some service. Sometimes, private companies also have this obligation [23]. The information about procurements can be found in the Journal of public procurements[2.12] . Some procurements are published in the Electronic Contractation System(EKS)[2.13] system. Application *verejne.digital*[2.14] tries to analyze new procurements and find suspicious ones automatically.

Search using CPV codes

Public procurements contain technical information that is very interesting for security professionals. Penetration testers can use Common Procurement Vocabulary(CPV) codes for both physical and network attacks. There is an easy way to search for technical information regarding information security. Each procurement has assigned at least one CPV code. The CPV code consists of eight main digits and one check digit.

Each digit of the code further specifies the procurement object from broad categories to very specific [5]. To show an example:

1. **35000000-4** Security, fire-fighting, police and defence equipment
2. **35100000-5** Emergency and security equipment
3. **35120000-1** Surveillance and security systems and devices
4. **35120000-1** Security equipment
5. **35125300-2** Security cameras

CPV codes are very useful during security assesments. For example, if burglars are interested in all security devices, they can first search using code **35000000-4** (Security, fire-fighting, police and defence equipment). If they find more results than they can process, they can specify the search only for the most interesting items, for example, cameras.

The sources where investigators can use CPV codes are EKS and the Journal of public procurements. The investigator often needs to cross-reference these sources with CRZ and look at the contract, as we show in following investigation:

1. We found a procurement for cameras at the Comenius University (`https://www.uvo.gov.sk/vyhladavanie-zakaziek/detail/dokumenty/142612`)
2. The procurement contains link to a contract.
3. We searched the CRZ for all contracts between the supplier and the University (`https://www.crz.gov.sk/index.php?ID=2171273&art_zs2=EMM+International%2C+spol.+s+r.o.&art_predmet=&art_ico=&art_suma_zmluva_od=&art_suma_zmluva_do=&art_datum_zverejnene_od=&art_datum_zverejnene_do=&art_rezort=0&art_zs1=Univerzita+Komensk%C3%A9ho+v+Bratislave&nazov=&art_ico1=&odoslat=Vyh%C4%BEada%C5%A5`).
4. We found more contracts that were not linked from the webpage of Office for Public Procurement (UVO). Each contract described locations of cameras in a different building.

verejne.digital

The free toolset **verejne.digital** aggregates data from multiple sources shown in this chapter. It contains the following tools:

- A map that shows places where companies are registered or places where politicians live.
- Profiles of politicians and their assets. These profiles are created from published property declarations that we mentioned in Section 1.1. The tool also cross-references this data with the real estate registry.

- A tool that can find the shortest link between companies and people. The tool also provides graph visualizations of the links.
- Procurements list that shows current public procurements and rates whether they are suspicious.

2.3 Case studies

2.3.1 Example of OSINT on organizations

We decided to perform OSINT on 7 organizations to demonstrate how much information can be found using the sources outlined in the previous section. We decided to use a list of organizations that were suppliers in the most followed contracts on 25.4.2020[3]:

1. LACORP, s.r.o
2. REMAPLAST spol. s.r.o
3. BONUL, s.r.o
4. Naša Planéta o.z.
5. A-testing, s.r.o
6. NEWTON Media, spol. s r. o.
7. Plastonic s.r.o.

We went through the previously mentioned sources, procurements, and contracts. We counted each piece of information. The results are in the table 2.1.

Some contracts and financial statements have the names and emails redacted while some do not. We were also able to find some contracts that had sensitive information poorly redacted. It was possible to copy the blacked-out text and paste it elsewhere.

All the companies registered on the EKS had their email published there. This might be interesting when the investigator wants to contact the company.

2.3.2 Finding antivirus software used on ministries

There are currently 13 ministries in Slovakia. Our objective in this scenario is to determine what antivirus programs they use.

There are multiple codes for antivirus software that we can use.

We searched EKS for the CPV code 48760000-3 (Virus protection software package). We found procurements for antivirus software in the following ministries:

- Ministry of Agriculture and Rural Development[4]

[3] The most followed contracts are those that had the most downloads during last 30 days. The archived version of the list can be found on `https://archive.is/Cziew`

[4] `https://portal.eks.sk/SpravaZakaziek/Zakazky/Detail/32956`

		Company						
		1	2	3	4	5	6	7
Type of information	Judicial decisions	0	12	93	0	0	18	0
	Liens	0	2	2	0	0	0	0
	Contracts in crz	4	4	172	1	2	42	6
	Emails	0	1	3	0	1	3	1
	Names	3	5	8	3	1	5	2
	Phone numbers	0	1	1	0	0	3	0
	Entries in EKS	0	1	5	0	0	2	0

Table 2.1: Number of found pieces of information per company

- Ministry of Education, Science, Research and Sport[5]
- Ministry of Health[6]
- Ministry of Environment[7]
- Ministry of Transport, Construction and Regional Development[8]
- Ministry of Foreign and European Affairs[9]
- Ministry of Justice[10]

We were not able to find antivirus information about other ministries in EKS. We tried to find information about antivirus programs in CRZ, but it is complicated because CRZ does not support full-text search. The process is very time consuming. We spent around a hour looking through various interesting contracts for the other ministries. We tried the Ministry of Culture, Ministry of Internal Affairs and Ministry of Defence.

We found an antivirus contract for the Ministry of Internal affairs (`https://www.crz.gov.sk/index.php?ID=2541923`). However, the Ministry withdrawed from this contract.

We did not find Antivirus used in the Ministry of Culture. However, we found that many organizations under the Ministry publish annual reports that contain antivirus names (for example, the contract `http://www.culture.gov.sk/extdoc/7938/NOC_SPRAVA_2018_FINAL` contains information about buying ESET NOD32 on page 57).

We used Google dork `ministerstvo obrany site:crz.gov.sk "ESET" licencie` to find a contract from the Ministry of Defence. We found that Google had multiple contracts indexed. The contracts were between the *Ministry of Defence* and *Aliter*

[5] https://portal.eks.sk/SpravaZakaziek/Zakazky/Detail/246351

[6] https://portal.eks.sk/SpravaZakaziek/Zakazky/Detail/148108

[7] https://portal.eks.sk/SpravaZakaziek/Zakazky/Detail/239949

[8] https://portal.eks.sk/SpravaZakaziek/Zakazky/Detail/110558

[9] https://portal.eks.sk/SpravaZakaziek/Zakazky/Detail/55093

[10] https://portal.eks.sk/SpravaZakaziek/Zakazky/Detail/237456

Technologies, a.s.. The company was renewing the licences for ESET each year, and last year we found was 2017 (`https://www.crz.gov.sk/index.php?ID=3263562&l=sk`). We did not find whether the Ministry decided to change the antivirus or whether it found new supplier for the same product.

Chapter 3

Websites

Investigators might need to research a specific domain or a Facebook page. The objective of the investigation is often an attribution to a specific actor. This is not always possible. Anyone can buy a domain and server completely anonymously. Nevertheless, malicious actors do not always do it. The investigators can assume that the creator of the website probably made some mistakes. We do not go into details of what mistakes the creators do, because they are not specific to Slovakia. These are methods like analysing whois information or analysing website's content [12].

The first section of this chapter shows our research of options available to researchers of Slovak internet. All the research in first section is based on the fact that the list of all *.sk* domains is publicly available. The most important part of this section is about finding domains that are associated to each other.

The second part of this chapter is about investigating and monitoring disinformation domains and Facebook pages. We show a project created to combat disinformation, along with two example case studies of the project's usage.

3.1 Research

The difference between Slovakia and other countries that makes all the research in this section possible is that the *.sk* domain owner, the SK–NIC shares a list of all registered domains[3.1] . We refer to this list as *domains.txt* in this chapter. This list is generated daily.

We noticed that it is uncommon for other country-code zone registrars to disclose a list of all registered domains. We were not able to find a full domain list for any other country TLD. There are some lists of domains available online. These lists are maintained by crawling the internet[3.2] . Because the lists are maintained using crawls, they only contain websites with incoming links from other websites. This is a huge limitation – for example, newly registered domains are not on this list.

We came up with multiple uses for this list and show them in the following subsections.

Phishing attack prevention

A company can set up a service that downloads the list every day and performs fuzzy string matching of the company's domains through all newly registered domains. With fuzzy string matching, they can find phishing or typosquatting attempts as soon as the domain is registered.

For example, let's say that a bank *examplebank* with the webpage *examplebank.sk* wants to protect its clients. The company monitors newly registered domains, so when a malicious actor registers domain *examplebamk.sk*, the bank is notified and can react appropriately.

To demonstrate this, we performed an experiment with a few common domain names. We used the tool *agrep* to perform fuzzy string matching [24].

We found one person that registers many domains with names that sound similar to some known webpages. Examples of domains that the person owns are:

- `kafland.sk`
- `jeureka.sk`
- `ministerstvodopravy.sk`
- `modykonik.sk`
- `okresnysud.sk`
- `tatrabaka.sk`
- `ylavomat.sk`
- `zlavy-sme.sk`

A reader with knowledge of the whole Slovak internet can easily spot that these domains are mimicking some real Slovak domains. Many of the changes rely on the user missing a key or hitting the wrong key while typing the domain name. After closer examination, we determined that this is not a phishing network but just a marketing scheme. Almost all of the pages redirect to the webpage `zlavy.odpadnes.sk`. This shows that monitoring registered domain names can be applied outside of security to protect the intellectual property and business interests of the company.

Slovak internet research

A list of all domains can help security researchers to scan the whole Slovak internet. Such a scan might answer many research questions, for example:

- Where are Slovak domains hosted
- How many websites are using encryption
- What is the most common content management systems on the Slovak internet.

A scan of the whole Slovak internet can also help greatly in domain investigations. For example, let's say that we do not know who is the author on a webpage but we know what handle he or she uses on Wordpress. One way to find his identity would be to perform a scan of all Slovak domains, determining if the website runs WordPress and if it does, then enumerating the authors. It might be the case that the author has more Wordpress blogs and uses the real name on some of them.

In the next subsection, we also show multiple techniques for domain attribution that are only possible because the investigators can scan the whole Slovak internet.

Domain correlation

Domain correlation is the process of enumerating domains related to one we already know. There are two different types of domain enumeration [11]:

- **Vertical domain correlation**, also called subdomain enumeration, is the process of finding subdomains under a known second-level domain.
- **Horizontal domain correlation** is the process of finding domains owned by the same company that has different second-level domain names.

Because the *domains.txt* file contains only second-level domains, it can not be used for vertical domain correlation. We did not find any sources specific to Slovakia that would help in vertical domain correlation, so investigator needs to use classic subdomain enumeration techniques (more in Section 4.1).

In the next paragraphs, we demonstrate different ways to horizontally correlate domains. We demonstrate the techniques on the example domain `uniba.sk`. Some techniques may return thousands of related domains, for example, if the domain is hosted on a shared web hosting. The investigator needs to decide which methods return false positives.

Owner The first simple thing to do is to look for the domains owned by the same owner. The owner is the third column in the *domains.txt* file. The owner for the University is `UNIV-0027` We find the domains sharing the same owner by using the `cut` and `grep` commands:

```
cat domains.txt | cut -d';' -f1,3 | grep 'UNIV-0027'
```

We found 8 domains that had the string `UNIV-0027` in their owner column:

- `comuniba.sk`
- `fwlslovakia.sk`
- `geonet.sk`
- `radavs.sk`
- `uniba.sk`
- `fallingwallslabslovakia.`
- `comeniusuniversity.sk`
- `konfuciovinstitut.sk`

There are two interesting domains. The domain `geonet.sk` is owned by `UNIV-0027-4059` and the domain `radavs.sk` is owned by `UNIV-0027-1227`. Both of these seem to belong also to the University, but for some reason they have numbers added to the end.

Registrator Some organizations (including Comenius University) have their own domain registrars. This is the second column in the *domains.txt* file, and for the University it is again `UNIV-0027`. We again searched for this with bash command:

```
cat domains.txt | cut -d’;’ -f1,2 | grep ’UNIV-0027’
```

This time, we found more domains. This is because many of the newfound domains are owned by the faculties, for example, `FMFI-0001` is the owner of two of these domains. There are 36 domains registered with the `UNIV-0027` registrar:

- `biomedmartin.sk`
- `centa.sk`
- `ceved.sk`
- `cezap.sk`
- `comuniba.sk`
- `fakultamanazmentu.sk`
- `fmtest.sk`
- `fsev.sk`
- `fwlslovakia.sk`
- `informatickyden.sk`
- `kamako.sk`
- `kmn.sk`
- `konfuciovinstitut.sk`
- `lefa.sk`
- `martinet.sk`
- `matika.sk`
- `mba.sk`
- `muzeologia.sk`
- `nasauniverzita.sk`
- `oamt.sk`
- `ose.sk`
- `radavs.sk`
- `radioaktiv.sk`
- `sosmt.sk`
- `ssmt.sk`
- `studujnafifuk.sk`
- `studujnauk.sk`
- `uniba.sk`
- `webjournal.sk`
- `webzurnal.sk`
- `zdravyspanok.sk`
- `zssmt.sk`
- `comeniusuniversity.sk`
- `fakultamanagementu.sk`
- `questionsofjournalism.sk`
- `fallingwallslabslovakia.sk`

Nameservers Two domains might be correlated if they share the same nameserver. The domain `uniba.sk` has two nameservers listed, `dns1.uniba.sk` and `dns3.uniba.sk`. We did not find any new subdomains this way, but there is some possibility that two domains are correlated only by nameserver so we include this method.

IP Address The domains might be correlated if they are hosted on the same IP address. We used the list of resolved hosts from Redlab SK [1]. Because the file is a json, we used `jq` to parse it:

```
cat sk-www-domains-resolved.json |
jq ".[][\"$(dig +short "uniba.sk")\"] | select (.!=null) | .[]" |
cut -d '"' -f 2
```

We found 3 domains cohosted with `uniba.sk`:

- `comeniusuniversity.sk`
- `ose.sk`
- `studujnauk.sk`

Mailserver Correlated domains might also share a mail server. Because we have a list of all domains, we can create a list of pairs of domains and mail servers by simply performing an `MX` DNS request to each domain. We found that domain `uniba.sk` uses mail server `uniba-sk.mail.protection.outlook.com`. By searching this list for other hosts that use the same mail server, we did not find any.

However, we were able to find mail servers in the network of Comenius University like `mail.dcs.fmph.uniba.sk`. This mail server is used by 4 domains that we did not encounter before because they are not registered or owned by Comenius University:

- `ecdl.sk`
- `informatika.sk`
- `mfcs.sk`
- `sofsem.sk`

Content of websites Even if the domains are registered anonymously, there is often a way to prove a horizontal correlation. There might be distinct marks in the content of a website that can be used, for example:

- Web designers sometimes leave signatures in the source code of websites they made. This can be in the form of an HTML comment or a visible text like *Created by the Example studio* at the bottom of the webpage.
- The website might contain some IDs in its source code. Examples of this are tracking IDs. Because the IDs are unique per owner, they can be used to prove the existence of a connection between two websites if it is used on both of them [19].

We decided to find domains similar to `uniba.sk` by looking for its tracking IDs. There are many online tools to find these correlations[1]. We tried these tools and found

[1] For example `https://analyzeid.com` or `https://dnslytics.com/reverse-analytics`

that they do not perform their searches across all Slovak websites, but find only the most known websites. Because we have a full list of all Slovak websites, we can create a better database than these tools have.

The script is attached to this thesis. We used GNU parallel for this script [21].

1. Visit each webpage from *domains.txt*.
2. On each website, look for following regular expressions:
   - `'UA-\d*-\d*'` - Google Analytics (GA) IDs
   - `'ca-pub-\d*'` - Google AdSense IDs
   - `'src="[^\""]*analytics[^\"]*\.js\"'` - sometimes, even though the website uses GA it does not have ID in its source code, because the ID is loaded from a javascript file. We try to match javascript files with the word *analytics* inside them and resolve these script files.
3. Filter the file for duplicates.

We found that `uniba.sk` uses Analytics ID `UA-11431529-28`. Similar IDs are used by following websites:

- `data-science.sk`
- `datova-veda.sk`
- `datovaveda.sk`
- `poistna-matematika.sk`
- `poistnamatematika.sk`
- `pravdepodobnost.sk`
- `studujnauk.sk`
- `uniba.sk`

We did not find six of these domains before because they are registered with domain provider Webhouse[2].

Automation of domain finding

We automated the process of finding horizontaly correlated domains by creating a bash script that gets a single domain as input and finds correlated domains. We demonstrate the tool by finding domains correlated to domain `uniba.sk`. Investigators should always first run the tool with the flag `-info`. This way they get an overview of how many domains are correlated by each method.

```
./similar_site.sh uniba.sk -info
#registrar:UNIV-0027
    Count: 37, filter out with -fr
#owner:UNIV-0027
    Count: 37, filter out with -fo
```

[2] `https://www.webhouse.sk/`

```
#ns1:dns1.uniba.sk
    Count: 16, filter out with -fns1
#ip:158.195.6.138
    Count: 4, filter out with -fip
#ns2:dns3.uniba.sk
    Count: 16, filter out with -fns2
#ns3:
    Count: 396567, filter out with -fns3
#ns4:
    Count: 396567, filter out with -fns4
#mailserver:uniba-sk.mail.protection.outlook.com.
    Count: 1, filter out with -fmx
Found Analytics ID: UA-11431529-17
    Count: 6, filter out with -fa=UA-11431529-17
Found Analytics ID: UA-11431529-28
    Count: 2, filter out with -fa=UA-11431529-28


To show only resulting domains, add -unique parameter

To show list sorted by domains, use -bydomain parameter
Recommended usage: ./similar_site.sh uniba.sk -fns3 -fns4
```

The fields for nameservers 3 and 4 are shared with many other domains, so they are certainly false positives. The tool automatically recommends filtering these methods. We try the recommended usage and get the following output:

```
./similar_site.sh uniba.sk -fns3 -fns4

owner:biomedmartin.sk
...
ns2:uniba.sk
mailserver:uniba.sk
analytics:data-science.sk UA-11431529-17
analytics:datova-veda.sk UA-11431529-17
analytics:datovaveda.sk UA-11431529-17
analytics:poistna-matematika.sk UA-11431529-17
analytics:poistnamatematika.sk UA-11431529-17
analytics:pravdepodobnost.sk UA-11431529-17
analytics:studujnauk.sk UA-11431529-28
```

```
analytics:uniba.sk UA-11431529-28
```

If investigators are not interested in the methods by which a domain was found, they can use switch `-unique` that produces just a list of unique found domains.

```
./similar_site.sh uniba.sk -fns3 -fns4 -unique

biomedmartin.sk
centa.sk
ceved.sk
...
webzurnal.sk
zdravyspanok.sk
zssmt.sk
```

3.2 Disinformation websites

Investigation of disinformation and hoaxes is a usual task for some investigators. To name a few examples, the investigator might be interested in the following tasks:

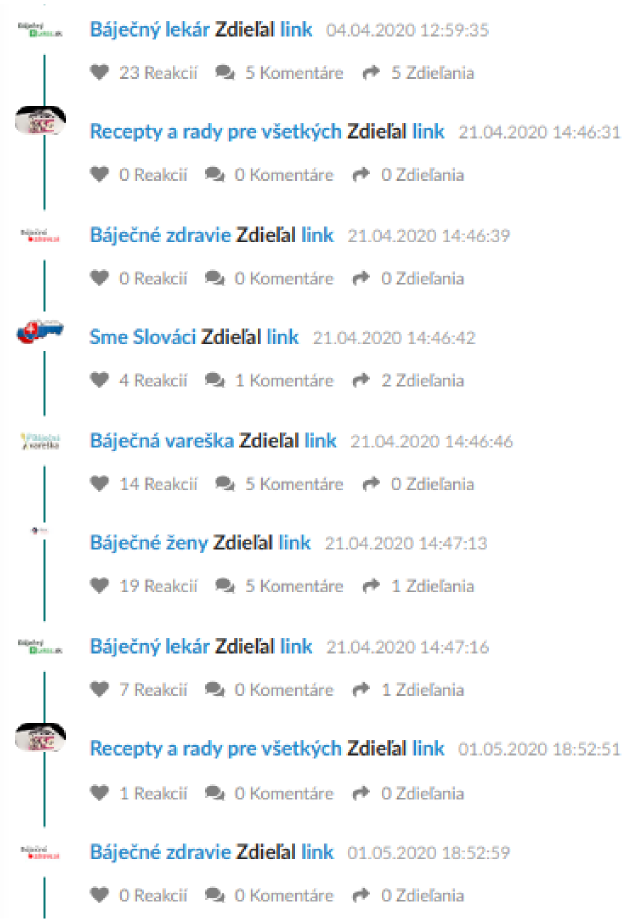
- People who create the content
- Political interests behind the hoax.
- Monitoring current trends in the disinformation scene.

A very good resource for analyzing disinformation websites is *blbec.online*[3.3] . The website follows many pages on Facebook and gathers their posts. The tool analyzes these posts and serves very detailed information. This data has many possible uses, we show only two that we find the most useful.

Because *blbec.online* shows the time of post shares, researchers can analyze times when a piece of content was shared. If two pages often happen to share the same content in matters of minutes, it means that the webpages have the same administrator. The tool provides investigators with an interface that enables them to spot these anomalies. The figure 3.1 shows that different Facebook pages shared the same link in one minute window[3]. By looking at different posts of these pages the investigator can verify that this happens very often. The investigator can thus conclude that these websites have the same administrator. In this scenario, it shows a connection between the pages sharing medical hoaxes and the nationalist Facebook page[4].

[3]

[4] The connection between these pages was originally found by Infosecurity.sk by analyzing Google AdSense IDs – `https://infosecurity.sk/dezinfo/mr-bajecny`.



Báječný lekár Zdieľal link 04.04.2020 12:59:35
23 Reakcií 5 Komentáre 5 Zdieľania

Recepty a rady pre všetkých Zdieľal link 21.04.2020 14:46:31
0 Reakcií 0 Komentáre 0 Zdieľania

Báječné zdravie Zdieľal link 21.04.2020 14:46:39
0 Reakcií 0 Komentáre 0 Zdieľania

Sme Slováci Zdieľal link 21.04.2020 14:46:42
4 Reakcií 1 Komentáre 2 Zdieľania

Báječná vareška Zdieľal link 21.04.2020 14:46:46
14 Reakcií 5 Komentáre 0 Zdieľania

Báječné ženy Zdieľal link 21.04.2020 14:47:13
19 Reakcií 5 Komentáre 1 Zdieľania

Báječný lekár Zdieľal link 21.04.2020 14:47:16
7 Reakcií 0 Komentáre 1 Zdieľania

Recepty a rady pre všetkých Zdieľal link 01.05.2020 18:52:51
1 Reakcií 0 Komentáre 0 Zdieľania

Báječné zdravie Zdieľal link 01.05.2020 18:52:59
0 Reakcií 0 Komentáre 0 Zdieľania

Figure 3.1: Pages that probably have the same admin

The second feature that is interesting is data about how many times were webpages linked on specific Facebook pages. For example, by looking at the Facebook site *World of politics*[5], we see that by far the most shared website is `magazin1.sk` and that the webpage `vitaminovo.sk` also was shared by this page.

Both of these leads are interesting. While `magazin1.sk` is a known disinformation website, it would be strange for the page to share exclusive content from this webpage unless there is some connection. This alone might be an indication that the Facebook page is a front to this webpage.

The second website, `vitaminovo.sk` is not very known and it is not online anymore. The Facebook page of this website has only 640 fans. Interestingly, the page shared content from this website, because it means that the administrator of the Facebook page at least follows the webpage and because not many people know about this webpage it might mean a potential connection.

The two other shared domains do not give researchers much information, because both are pretty well-known portals so even the knowledge that the administrator of *World of politics* follows them does not give us much information.

An investigative journalist found that both `magazin1.sk` and `vitaminovo.sk` are linked to Mario Vidak[6]. Because we suspect that the *World of politics* is linked to both `magazin1.sk` and `vitaminovo.sk`, we can suspect that Mario Vidak is also administrator of the *World of politics*.

[5] https://blbec.online/blb/1232-svet-politiky

[6] https://infosecurity.sk/dezinfo/mr-bajecny/

Chapter 4

OSINT in red teaming

OSINT always begins with a piece of known information and objectives. For example:

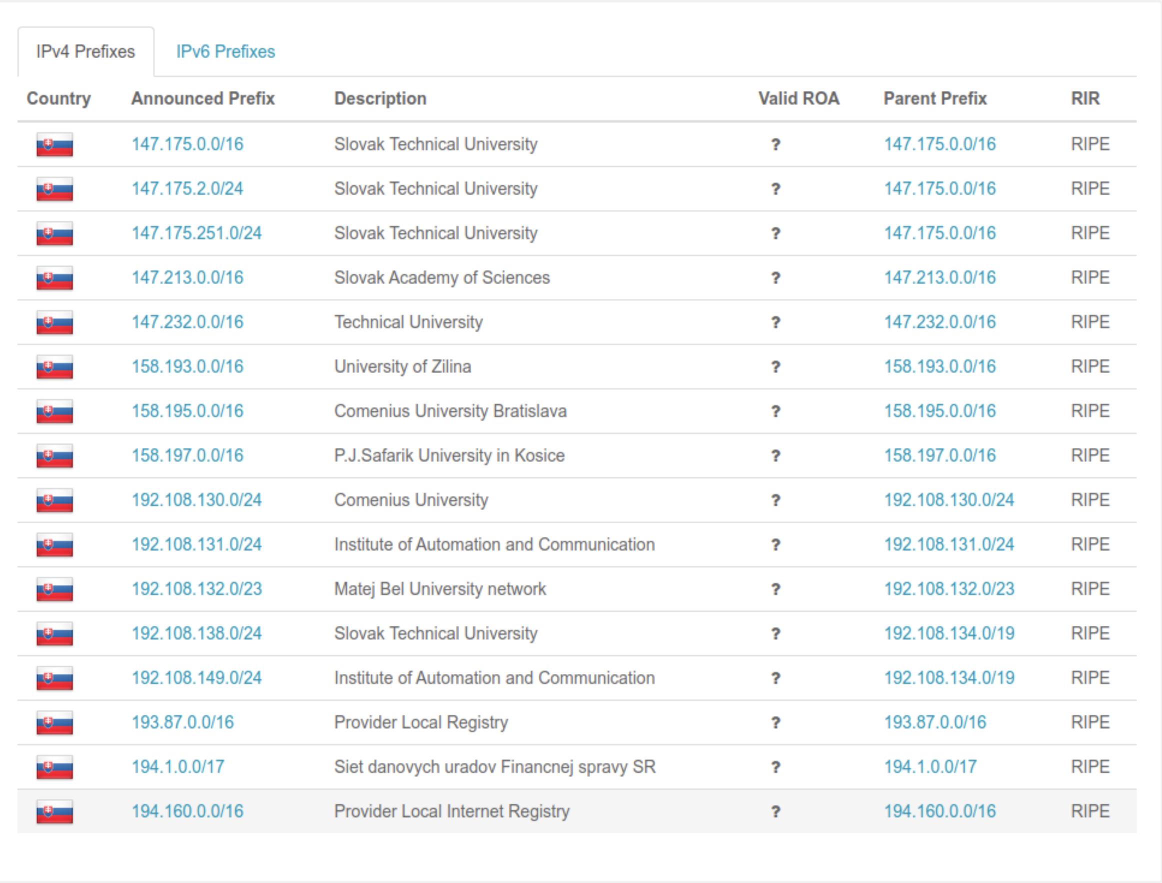
- **Known information:** A possibly overpriced contract between a company and a public institution. The contract contains the name of the company and the public institution.
- **Objectives:**
  1. Find who owns the company.
  2. Find whether owners have some relations to the public institution.
  3. Find whether the contract was overpriced.
  4. Find how the procurement for this contract was done.

Red teaming exercise is a simulated attack on a company. The exercise should help the company to see how they would react to a real attack. Red teaming exercises often have less restricted scope than penetration tests. OSINT can often help red teamers to find additional assets of the company that could help them in their attack [7].

In this chapter, we show an example OSINT before a red teaming exercise. We start by defining the objectives:

- **Known information:** Name of the organization.
- **Objectives:**
  1. Find information about an organization's online infrastructure.
  2. Find information that could aid in a phishing attack.
  3. Find whether company leaks some data.
  4. Find information that could aid in a physical attack.

We show methods that can help the red team to achieve these objectives. Most methods we found are from the MITRE PRE-ATT&CK framework [17]. We demonstrate all methods on the Comenius University. For other organizations, the results would be different. The University has a huge infrastructure, and it is a very open



Country	Announced Prefix	Description	Valid ROA	Parent Prefix	RIR
🇸🇰	147.175.0.0/16	Slovak Technical University	?	147.175.0.0/16	RIPE
🇸🇰	147.175.2.0/24	Slovak Technical University	?	147.175.0.0/16	RIPE
🇸🇰	147.175.251.0/24	Slovak Technical University	?	147.175.0.0/16	RIPE
🇸🇰	147.213.0.0/16	Slovak Academy of Sciences	?	147.213.0.0/16	RIPE
🇸🇰	147.232.0.0/16	Technical University	?	147.232.0.0/16	RIPE
🇸🇰	158.193.0.0/16	University of Zilina	?	158.193.0.0/16	RIPE
🇸🇰	158.195.0.0/16	Comenius University Bratislava	?	158.195.0.0/16	RIPE
🇸🇰	158.197.0.0/16	P.J.Safarik University in Kosice	?	158.197.0.0/16	RIPE
🇸🇰	192.108.130.0/24	Comenius University	?	192.108.130.0/24	RIPE
🇸🇰	192.108.131.0/24	Institute of Automation and Communication	?	192.108.131.0/24	RIPE
🇸🇰	192.108.132.0/23	Matej Bel University network	?	192.108.132.0/23	RIPE
🇸🇰	192.108.138.0/24	Slovak Technical University	?	192.108.134.0/19	RIPE
🇸🇰	192.108.149.0/24	Institute of Automation and Communication	?	192.108.134.0/19	RIPE
🇸🇰	193.87.0.0/16	Provider Local Registry	?	193.87.0.0/16	RIPE
🇸🇰	194.1.0.0/17	Siet danovych uradov Financnej spravy SR	?	194.1.0.0/17	RIPE
🇸🇰	194.160.0.0/16	Provider Local Internet Registry	?	194.160.0.0/16	RIPE

Figure 4.1: IPv4 Prefixes of Slovak Academic Network

institution. If the red team would perform OSINT on a small institution, they might not find any useful information. However, from our experience we believe that it is possible to find something interesting in any bigger institution.

For each technique, we show few examples of results. We did not perform full OSINT nor did we verify that the found vulnerabilities really exist or that the credentials really work.

4.1 Infrastructure

IP ranges

Larger organizations often own IP ranges. Knowledge of these ranges is important for the red teamers because it enables them to look for vulnerable services in these ranges. We used the tool *bgpview*[4.1] to search for University's IP ranges.

We found Slovak Academic Network, and its IPv4 prefixes. All IP ranges are shown on Figure 4.1. We can see that the Comenius University has two prefixes there. Later in this chapter we show that it also hosts parts of infrastructure in the IP ranges *193.87.0.0/16* and *194.160.0.0/16*.

Domain enumeration

There are many reasons why the red team needs to enumerate domains:

- Many organizations do not have their IP ranges, and they host all their services in the cloud.
- Even if an organization has its IP ranges, it is possible that it also has some domains hosted elsewhere.
- If a web server is running on an IP and it uses virtual hosts, it is necessary to find what domain names point to this IP. Each virtual host is a new server with potentially different attack surface.

There are many tools and methodologies for this purpose, and each of them can yield different results. We decided to use the following tools and techniques for horizontal domain enumeration:

- Our tool that enumerates Slovak domains (see Chapter 3). This tool found 44 second-level domain names.
- Open-source tool *Amass*[4.2] . We ran the tool on University's CIDR (`'amass intel -cidr 158.195.0.0/16`) and found 20 domains. Some of these domains belonged to some teachers or other organizations associated with University (for example Trojsten). This is interesting, because our tool can not find such domains.
- Website *Built With*[4.3] . The webpage can show domains that share Google Analytics ID. It found the webpage `skvprifuk.info`. This webpage was not found by our tool because the tool checks only Slovak domains. However, *Built With* did not find the *.sk* domains our tool found.

Once the red team has created a list of second-level domains, the red team can start with vertical domain enumeration using the following methods :

- **Passive subdomain enumeration** is the process of finding subdomains with OSINT. Many tools provide a quick way to search through many sources. We decided to use the tool *Amass* for this.
- **Subdomain bruteforce** is the process of trying subdomain names from wordlist against the DNS resolver. The wordlist contains common words like *admin*, *jira*, *mail*. This method is very noisy and creates huge traffic for the organization's DNS server. We did not perform this method because of these properties.

We also performed reverse DNS Lookup on each host from the IP range. This method is for both subdomain and second-level domain enumeration. We found that many hosts in networks have reverse DNS pointers. We were able to find 2549 subdomains with this method. These need to be resolved again, as the PTR DNS queries and A queries do not have to be consistent.

We then performed the following steps:

1. We ran *Amass*'s passive subdomain enumeration on each of the found second-level domains. We found 5533 unique subdomains.
2. We combined this result with the 2549 domains found using reverse DNS queries. The resulting list has 5675 subdomains.
3. We tried to resolve each of these subdomains. We found that 4631 of them resolve to an IP address. Others are probably old domain names that are not used anymore.

Determine where are the IPs hosted

After resolving the domains to a list of IP addresses on which the organization hosts infrastructure, the red team can determine what providers or ranges the organization uses by looking at who owns these IP addresses. This is usually done by running whois query on the IP address or tool such as *bgpview*[4.4] . Some providers might come with a specific vulnerabilities, for example if red team knows that the organization uses Amazon AWS it can look for vulnerable S3 buckets.

We found following providers that host University's infrastructure (outside of University's own IP ranges):

- `Webhost.sk` – for example `www.data-science.sk`
- `193.87.0.0/16` and `194.160.0.0/16` – Provider Local Registry of Slovak Academic Network
- `Microsoft` – for example the MS Outlook or MS Office.

Open ports

The red team is interested in services that run in target infrastructure and are available from the internet. Each service provides an additional attack surface. Instead of port scanning, which is a slow and loud operation, we used passive service enumeration.

Passive enumeration utilizes the available port scans on the internet. Projects like Shodan[4.5] , Censys[4.6] or Project Sonar[4.7] are used for this purpose. The results are usually fast, and some of these services even contain historical data.

We performed search on University's IP ranges with Shodan. Because there are many results, we filtered out the most common ports like 80,443 or 25. We were able to find many services that would be interesting for red teamers, for example:

- Ethereum miner (`http://158.195.19.228:9090/`)

- CUPS print server (`http://158.195.31.99:631/printers/`)
- Webmin (`https://158.195.108.5:10000/`)
- Apache Tomcat administration (`http://158.195.68.49:8080/`)
- Remote desktop (`158.195.13.74`)

By using Shodan's command line interface, the investigator can also find when the ports were open and validate when the port was last found open: `shodan host -history IP`.

Directory and file finding

Websites often host unlinked content that is available only if one knows the specific URL or parameter. For example, an old administration interface can be exposed at the url `/old/administration.php`. There are two methods used to find such content:

- **Passive enumeration** There are many sources that aggregate URLs. Examples are Internet Archive[4.8] or Common Crawl[4.9] .
- **Bruteforce** This method uses a wordlist of possible directory or file names and tries to request them all.

We performed passive enumeration with tool *getallurls*[4.10] on all domains from our list and saved it. This tool uses the Internet Archive, Common Crawl, and AlienVault's Open Threat Exchange. We searched the urls for keywords indicating vulnerabilities or some interesting attack surface:

- `file=` – we found many results, all of them should be checked for potential file inclusion vulnerability.
- `.sql` – we found few SQL dumps, but they all turned out to be study materials.
- `admin` – we found some administration panels(`https://bpf2016.flaw.uniba.sk/administrator/` and `https://auth.ais2.uniba.sk/carbon/admin/login.jsp`). A red team could now proceed to try leaked credentials or create a phishing website based on them.

Google dorking

Red teamers often find leaks, useful information, or files with Google dorks. Real hackers also use Google dorks to find vulnerabilities[1].

We tried following dorks:

[1] This can be verified by searching for term Dork through hacking forums. There are often posts about collections of Google Dorks that can be used to hunt for specific vulnerability, like this one `https://web.archive.org/web/20200414054431/https://raidforums.com/Thread-Google-SQL-Injection-2020-4000-DORKS`

Leak	Year	Password format	Number of entries	Number of unique emails
SKTorrent.eu	2016	Cleartext	2	2
Adobe	2013	Encrypted	147	147
Dropbox	2012	Hashed	53	53
Myspace	2008	Hashed	19	19
000webhost.com	2015	Cleartext	3	3
Exploit.in	2016	Cleartext	437	356
Collection #1	2019	Mostly cleartext	542	294

Table 4.1: Leaks

- `site:uniba.sk intitle:"index of"` – this dork finds exposed open directories. We found 2590 results. We checked only a few of them and did not find any vulnerabilities; most of them were open because the teachers use them to share files with students.
- `allintext:password site:uniba.sk` – this dork looks for potentially leaking passwords. We found a file with passwords. However, closer examination showed that it is a config file for students[2].
- `site:uniba.sk "XSS" "SQL"` – this dork looks for mentions of vulnerability on the webpage. Previous reports or articles about vulnerabilities can help to direct the red team to an area where more bugs persist. We found an old security assessment of `blog.matfyz.sk`[3]. In this assessment, we can find that the blog contained many vulnerabilities and might be worth exploring. The document is ten years old, but the website seems to not have changed much in the meantime.

4.2 Leaks in organization

Leaked credential databases

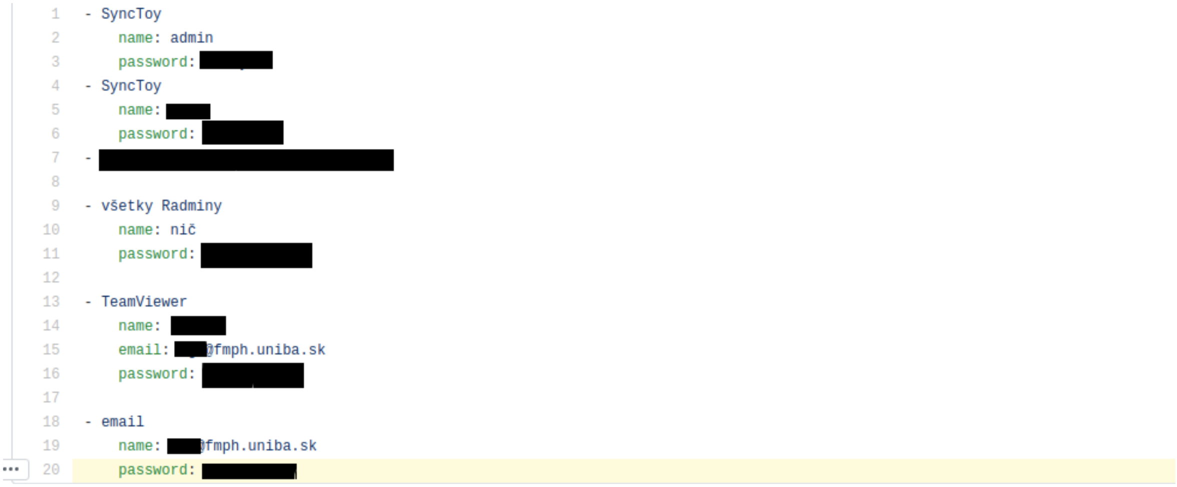
Employees often reuse their credentials in company network. If their credentials are leaked, the hackers can use them to attack the company. Copies of some (usually older) leaks are freely accessible on the internet.

We downloaded some of these accessible databases and searched them for *uniba.sk*. The results are in the Table 4.1.

Together, the leaks contained 636 unique email addresses. As we can see, the best results come from *Exploit.in* and *Collection #1*. This is because these are huge compilations of cracked hashes from many leaks.

[2] `http://danka.ii.fmph.uniba.sk/~vittek/unix/.netrc.html`

[3] `http://www.dcs.fmph.uniba.sk/bakalarky/obhajene/getfile.php/main.pdf?id=104&fid=199&type=application%2Fpdf`



```
- SyncToy
    name: admin
    password: [redacted]
- SyncToy
    name: [redacted]
    password: [redacted]
- [redacted]

- všetky Radminy
    name: nič
    password: [redacted]

- TeamViewer
    name: [redacted]
    email: [redacted]@fmph.uniba.sk
    password: [redacted]

- email
    name: [redacted]fmph.uniba.sk
    password: [redacted]
```

Figure 4.2: Passwords leaking on Github

Code repositories reconnaisance

Code repositories like Github, Bitbucket, or Gitlab are huge sources of information about the organization. Information that can be found in them contains [2]:

- Employee names and their positions
- Technologies used in the organization
- Domains of organization
- Leaked secrets or passwords

After searching for the University on the Github, we were quickly able to find a potential leak of five passwords (Figure 4.2).

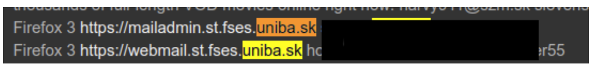
Hacker forums

Hackers often show their achievements on hacker's forums. We did not find any Slovak forum, but there is one Czech[4.11] . The forum has its bugtrack where hackers post their findings from Czech websites[4.12] .

Other useful source is archive of defaced webpages[4.13] . Hackers use it to brag about webpages they hacked. The red team can use it to find vulnerabilities that have already been found by black-hat hackers. Sometimes, the defacement is unnoticed by website administrators for a long time, and the vulnerability can be still exploitable.

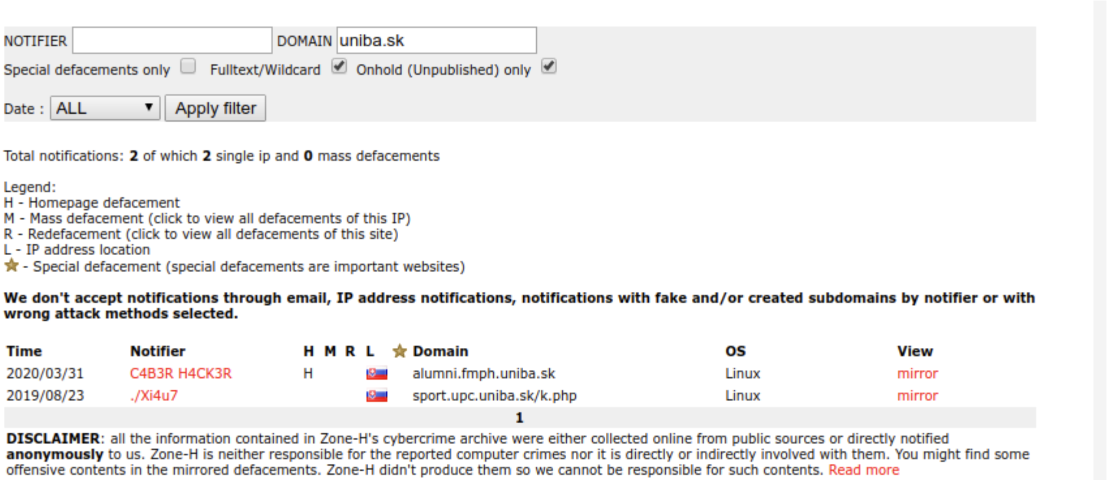
There is no tool known to us that can search all the hacker forums at once. We used a google dork `"uniba.sk" -site:.sk forum hack`, and among the results we found a hacker sharing hacked access to `mailadmin.st.fses.uniba.sk` (Figure 4.3). The post is 10 years old, therefore it is probably not significant for us.

We also searched the database of defaced websites. We were able to find defacements that were online at the time of writing (Figure 4.4). These defacements were probably



Firefox 3 https://mailadmin.st.fses.uniba.sk [redacted]
Firefox 3 https://webmail.st.fses.uniba.sk hc [redacted] er55

Figure 4.3: Hacked access found on hacker forums



NOTIFIER □ DOMAIN uniba.sk
Special defacements only □ Fulltext/Wildcard ☑ Onhold (Unpublished) only ☑
Date : ALL ▾ Apply filter

Total notifications: **2** of which **2** single ip and **0** mass defacements

Legend:
H - Homepage defacement
M - Mass defacement (click to view all defacements of this IP)
R - Redefacement (click to view all defacements of this site)
L - IP address location
★ - Special defacement (special defacements are important websites)

We don't accept notifications through email, IP address notifications, notifications with fake and/or created subdomains by notifier or with wrong attack methods selected.

Time	Notifier	H	M	R	L	★	Domain	OS	View
2020/03/31	C4B3R H4CK3R	H			🇸🇰		alumni.fmph.uniba.sk	Linux	mirror
2019/08/23	./Xi4u7				🇸🇰		sport.upc.uniba.sk/k.php	Linux	mirror

1

DISCLAIMER: all the information contained in Zone-H's cybercrime archive were either collected online from public sources or directly notified **anonymously** to us. Zone-H is neither responsible for the reported computer crimes nor it is directly or indirectly involved with them. You might find some offensive contents in the mirrored defacements. Zone-H didn't produce them so we cannot be responsible for such contents. Read more

Figure 4.4: Currently defaced websites

still not found by the website's administrators, and the vulnerability that was used by the hacker is probably still not fixed.

Paste sites

Paste sites are interesting for OSINT for two main reasons:

- People can use them to share private data with others.
- Hackers often use them to leak private information.

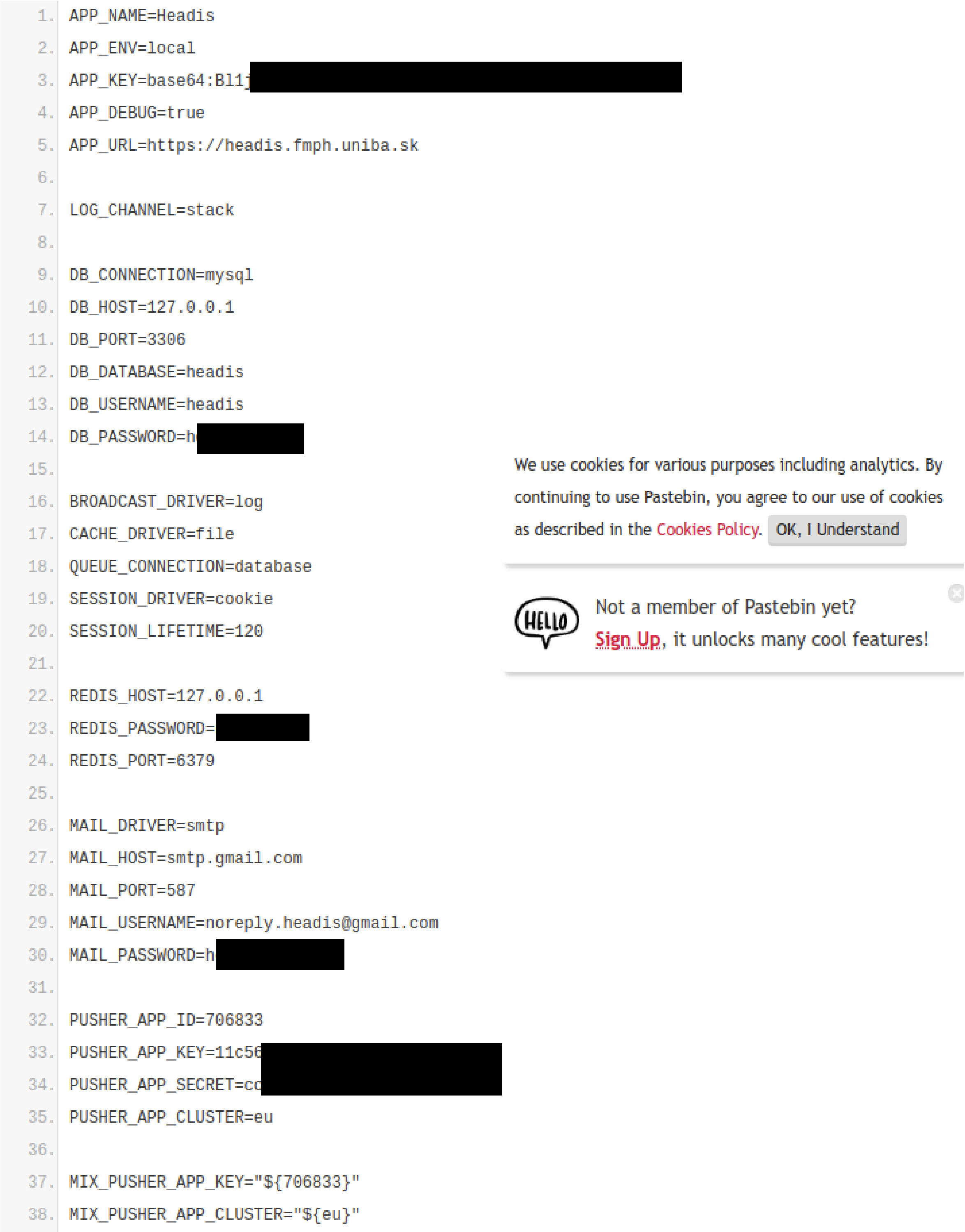
We used a paste search tool[4.14] to search for the url `uniba.sk`. The most interesting result we found was paste that contained multiple credentials, shown on Figure 4.5.

4.3 General information

Employees

Knowing the company structure and departments can help the red team to target phishing emails. Instead of sending the phishing email to everybody in the company, the red team can perform a more subtle approach and only target people who have interesting access:

- Managers usually have access to sensitive information. The highest access might be by C-level executives.

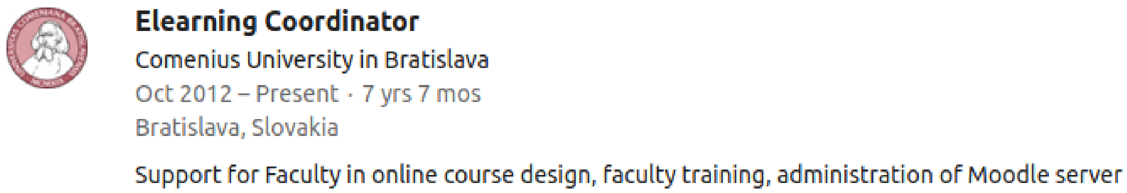


```
APP_NAME=Headis
APP_ENV=local
APP_KEY=base64:B1[redacted]
APP_DEBUG=true
APP_URL=https://headis.fmph.uniba.sk

LOG_CHANNEL=stack

DB_CONNECTION=mysql
DB_HOST=127.0.0.1
DB_PORT=3306
DB_DATABASE=headis
DB_USERNAME=headis
DB_PASSWORD=h[redacted]

BROADCAST_DRIVER=log
CACHE_DRIVER=file
QUEUE_CONNECTION=database
SESSION_DRIVER=cookie
SESSION_LIFETIME=120

REDIS_HOST=127.0.0.1
REDIS_PASSWORD=[redacted]
REDIS_PORT=6379

MAIL_DRIVER=smtp
MAIL_HOST=smtp.gmail.com
MAIL_PORT=587
MAIL_USERNAME=noreply.headis@gmail.com
MAIL_PASSWORD=h[redacted]

PUSHER_APP_ID=706833
PUSHER_APP_KEY=11c5[redacted]
PUSHER_APP_SECRET=c
PUSHER_APP_CLUSTER=eu

MIX_PUSHER_APP_KEY="${706833}"
MIX_PUSHER_APP_CLUSTER="${eu}"
```

We use cookies for various purposes including analytics. By continuing to use Pastebin, you agree to our use of cookies as described in the Cookies Policy.

Not a member of Pastebin yet? Sign Up, it unlocks many cool features!

Figure 4.5: Credentials leaked to Pastebin



Elearning Coordinator
Comenius University in Bratislava
Oct 2012 – Present · 7 yrs 7 mos
Bratislava, Slovakia

Support for Faculty in online course design, faculty training, administration of Moodle server

Figure 4.6: Moodle administration in LinkedIn

- IT department has elevated rights to every single computer in the company. However, they might also be the hardest to attack, as some of them might be also experts in computer security.
- Human resources department or financial department usually has access to sensitive information, for example, the salaries of all employees.

Many sources can be used to find this information, for example:

- Company's webpage often lists some employees or contact emails.
- Data from leaked websites can be searched for emails associated with the company.
- Websites that contain emails. These sites are targeted at salesmen and recruiters[4.15]
- Many employees have the organization listed on LinkedIn. They often also have a specific description of their work listed there.

We can easily determine the organisation structure of university from its webpage[4]. The webpage also lists employees of faculty, for example, we can find the management[5] or the IT staff[6].

We can continue by finding social networks of employees and finding more information about their work. For example, we browsed through LinkedIn profiles of employees and found who is the administrator of the Moodle server (Figure 4.6).

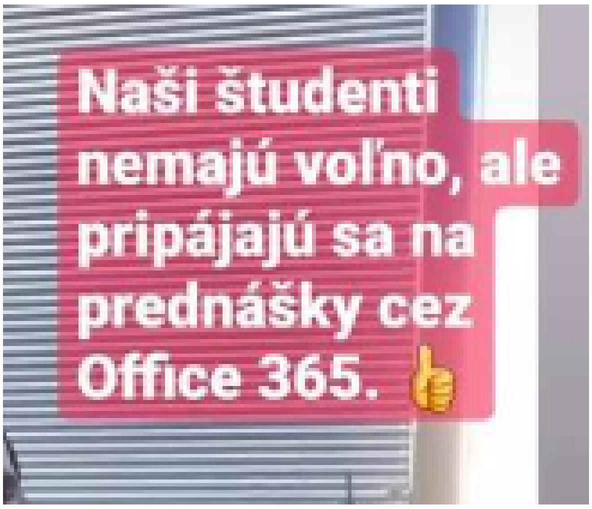
Basic information

Before crafting a phishing email, the red team might be interested in basic information about the organization. Such information can help in identifying potential issues specific to the company (for example, the red team can mention a recent merger in the phishing mail). We have shown how to gather this data in Slovakia in Chapter 2.

[4] https://uniba.sk/o-univerzite/fakulty-a-dalsie-sucasti/ and https://uniba.sk/o-univerzite/organy-uk/

[5] https://uniba.sk/vedenie/

[6] https://uniba.sk/en/about/faculties-and-units/cit/contact/



Naši študenti nemajú voľno, ale pripájajú sa na prednášky cez Office 365. 👍

Figure 4.7: Facebook post that proves usage of Office 365 in the University

For the University, the financial statement contains much information[7]:

- Many names
- That the University owns shares of the company UK VEDA, s.r.o.
- That the University owns seven historical buildings

Basic information usually does not contain anything critical, but it is still useful for attackers, for example when crafting phishing emails.

News mentions

The company's presence in media can reveal current issues. The red team can use current issues to create a believable phishing email.

An example of a current issue is the COVID-19 crisis. The University holds its first online exams[8]. A phishing email targeting the students and teachers could be built around this information.

Find social networks of organization

Many organizations have dedicated social media profiles. The information shared on these profiles is often a source of OSINT for the red team. An example of this is a photo of employees with visible employee cards. The red team can then create imitations of these cards.

We looked through University's Facebook page and found a photo telling that people are using Office365 for online education (Figure 4.7)

[7] http://www.registeruz.sk/cruz-public/domain/financialreport/attachment/7617442

[8] https://domov.sme.sk/c/22380531/univerzita-komenskeho-spustila-historicky-prve-online-statn
html

Job postings

The red team can find the following information in job postings:

- **Technologies used in company**
- **Contact information**
- **Roles and salaries**

There are multiple websites where Slovak companies list their job postings, for example:

- `profesia.sk`
- `kariera.zoznam.sk`
- `istp.sk`
- `pracovne-ponuky.eu`

If no information can be found, it is also possible to check older postings via Wayback Machine[4.16] .

We checked University's Profesia page[9]. We then looked at this URL through Wayback Machine to find old postings that might reveal more information. We found a page from 2015[10], and in there an old posting for a manager of IT centre[11]. From the posting, we can deduce that the IT centre uses Microsoft Office and Windows. No Linux administration skills are needed for this position.

This is not very useful information. University probably has job postings on other places then `profesia.sk`. We did not check all of them.

Contractors

It is sometimes possible to find a supplier with lower security standards. This supplier might have a certain level of trust established with the actual target. This trust can then be used to compromise the target organization.

Suppliers often show references for marketing purposes. The red team can search for these references to find suppliers.

We used the google dork `"referencie" "Univerzita Komenskeho" -site:uniba.sk`. Examples of contractors we found are:

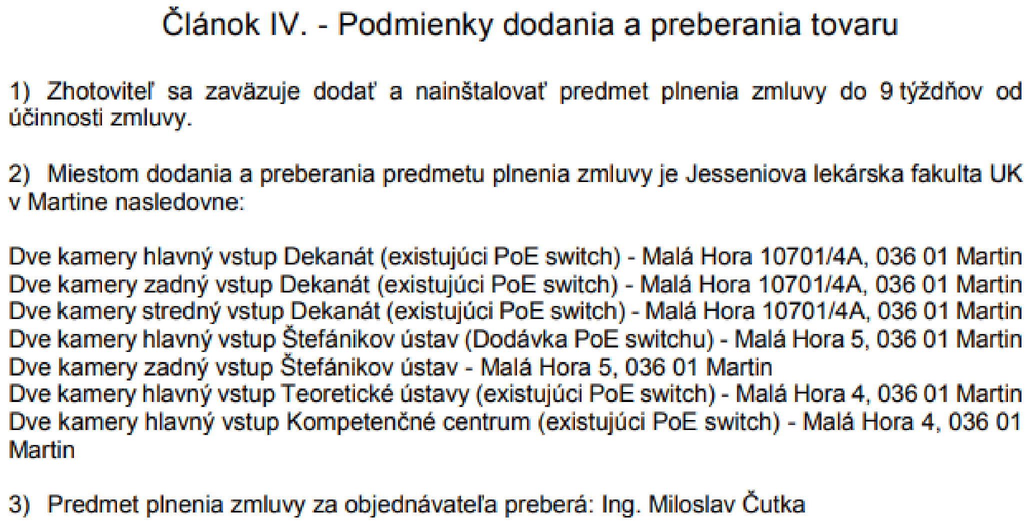
- VIPTel[12]

[9] https://www.profesia.sk/praca/univerzita-komenskeho-v-bratislave-rektorat/C827

[10] https://web.archive.org/web/20150413111738/http://www.profesia.sk:80/praca/univerzita-komenskeho-v-bratislave-rektorat/C827

[11] https://www.profesia.sk/praca/univerzita-komenskeho-v-bratislave-rektorat/O2075699

[12] https://www.viptel.sk/univerzita-komenskeho-v-bratislave



Článok IV. - Podmienky dodania a preberania tovaru

1) Zhotoviteľ sa zaväzuje dodať a nainštalovať predmet plnenia zmluvy do 9 týždňov od účinnosti zmluvy.

2) Miestom dodania a preberania predmetu plnenia zmluvy je Jesseniova lekárska fakulta UK v Martine nasledovne:

Dve kamery hlavný vstup Dekanát (existujúci PoE switch) - Malá Hora 10701/4A, 036 01 Martin
Dve kamery zadný vstup Dekanát (existujúci PoE switch) - Malá Hora 10701/4A, 036 01 Martin
Dve kamery stredný vstup Dekanát (existujúci PoE switch) - Malá Hora 10701/4A, 036 01 Martin
Dve kamery hlavný vstup Štefánikov ústav (Dodávka PoE switchu) - Malá Hora 5, 036 01 Martin
Dve kamery zadný vstup Štefánikov ústav - Malá Hora 5, 036 01 Martin
Dve kamery hlavný vstup Teoretické ústavy (existujúci PoE switch) - Malá Hora 4, 036 01 Martin
Dve kamery hlavný vstup Kompetenčné centrum (existujúci PoE switch) - Malá Hora 4, 036 01 Martin

3) Predmet plnenia zmluvy za objednávateľa preberá: Ing. Miloslav Čutka

Figure 4.8: Cameras in one of the buildings

- CMSTypo3[13]
- Eagle Security[14]

Procurements and contracts

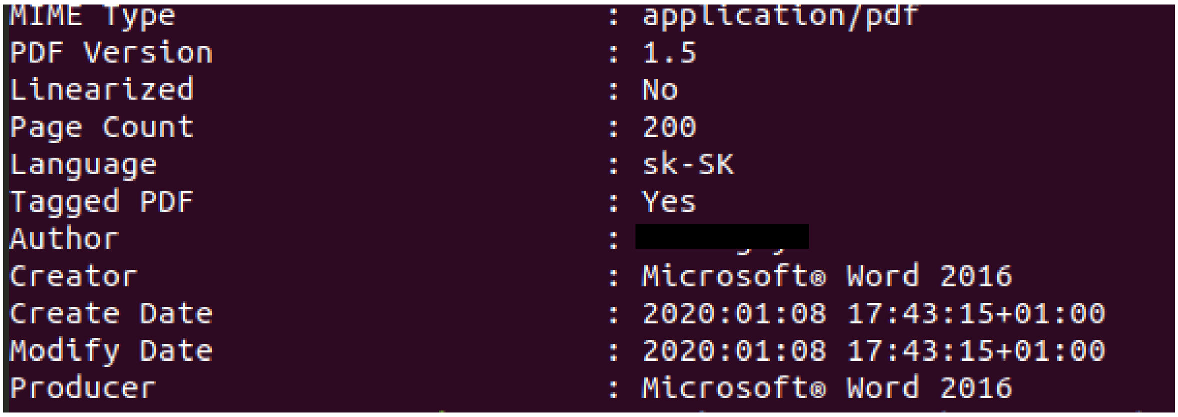
If the organization is public or conducts business with public organizations, the red team can find some information in published contracts and procurements. This information can include used software, contractors, or contact information.

Because the University is a public organization, almost all information can be found in public procurements. We searched the Journal of public procurements by CPV code 35000000-4 (Security, fire-fighting, police and defence equipment). We found multiple documents related to the physical security of buildings. We looked specifically at contracts related to cameras. After reading a few of the contracts, we find out that each contract is for cameras in different buildings of University. Some of the contracts list the exact location of the cameras (Figure 4.8). The whole process of finding these contracts is detailed in the Section 2.2.

Metadata from webpages

Metadata from documents disclose information about software versions used in the company. This information can be used to target exploits. With Google dorks or specialized tools, the red team can find recent documents in formats that contain metadata [6].

[13] https://www.cmstypo3.sk/referencie/referencie/detail/univerzita-komenskeho-v-bratislave/
[14] http://www.eaglesecurity.sk/referencie/



```
MIME Type                       : application/pdf
PDF Version                     : 1.5
Linearized                      : No
Page Count                      : 200
Language                        : sk-SK
Tagged PDF                      : Yes
Author                          : [redacted]
Creator                         : Microsoft® Word 2016
Create Date                     : 2020:01:08 17:43:15+01:00
Modify Date                     : 2020:01:08 17:43:15+01:00
Producer                        : Microsoft® Word 2016
```

Figure 4.9: Usage of Office 2016 by this employee

For University, we used the Google dork `site:uniba.sk ext:pdf from:2020`. The metadata of these documents shows who uses which version of Microsoft Office (Figure 4.9).

Physical locations

OSINT can help in physical red team exercises. The red team might find maps of the buildings or some other properties that could be targeted.

We found the locations of the University's buildings by their contact pages[15]. We also tried to find a map of the Faculty of Mathematics, Physics and Informatics, and we were able to find a very detailed map[16].

We also looked for at the real estate registry. The University owns many buildings so we only searched Bratislava Ruzinov. We found two buildings:

- University Pharmacy[17]
- Faculty of Social and Economic Sciences[18]

4.4 Application of Slovak sources

We used Slovak sources in following steps of assignment:

- Searching for domains
- Searching for buildings
- Procurements and contracts

[15], For example, `https://fmph.uniba.sk/kontakt/`

[16] `https://www.mapa.matfyzjein.sk/`

[17] `https://zbgis.skgeodesy.sk/mkzbgis/sk/kataster?bm=orto&z=17&c=17.155524,48.155273&sc=n&it=point&dt=parcelsC#/detail/kataster/stavba/805556/4813/12/15294_65/LEKAREN`

[18] `https://kataster.skgeodesy.sk/Portal/sk/Detail/Participant/3915767493`

When targeting private companies, procurements and contracts do not contain as much information. Therefore, the most useful source for red teamers available in Slovakia is the list of the publicly accessible domains.

Conclusion

We have shown how the application of Slovak specific sources makes a difference in the ability to perform relevant OSINT. In Slovakia, OSINT is well known to journalists, but we did not find any extensive research that would be useful in the area of information security. This thesis is the first that discusses sources in Slovakia from a malicious viewpoint. We believe it can help people and organisations better understand attackers and help defend against malicious OSINT.

In the chapter about the personal data, we show that even if a person has a small digital footprint, sources that the person does not control can reveal a lot of information. It is almost certain that some information about the person is online. We found that the most challenging part of the OSINT process was to process the amount of information, especially when there were multiple people with the same name and we had to often rely on social media profiles. This limitation directly shows one option to hide from investigators performing OSINT and enhance privacy. People who want to remain hidden on social media can create multiple fake profiles of themselves with conflicting information. This method would not stop an investigation from law enforcement, but it would make the investigation very difficult for journalists or stalkers.

For organisations, there is always a certain amount of information available online. The biggest danger in OSINT is for public institutions, as they share much more data than private. The best thing they can do is to know what information is publicly available and use this during threat modelling. Public institutions can also make sure not to show more information then is needed (for example, the exact locations of cameras do not need to be specified in the contract). All organisations should understand that they can not register a secret Slovak domain. We think that our domain enumeration tool provides new options to red teamers or journalists.

There are possibilities for future research based on our work:

- Many of the sources we found could be indexed with the application of more advanced automation, such as Optical Character Recognition. This would enable investigators easier search. For example, the sources that are difficult to search (like the CRZ) can be fully scraped and indexed with the use of OCR. In the case of CRZ, this would be very useful for red teaming, as red teamers could search even the contracts that do not have the correct CPV codes.

- Perform similar research in other countries. It is complicated to find good sources for OSINT in other smaller countries. Most resources that we found written in the English language are about the USA, Great Britain or Russia. We believe that it would be beneficial to create a compilation of OSINT sources for most countries.
- Updating this research will be important in the future. From experience, we know that many of the sources we provided will be inaccessible in a few years, and many new sources will emerge. Open sources are changing every day and so should change the methods of researchers.
- There is a possibility of a deeper analysis of some of the topics we discussed. We did not dive deep into some topics that could provide interesting results, for example, the usage of social networks in Slovakia or the contracts and procurements. It might be possible to mine more information from such sources.

Bibliography

[1] List of all slovak resolved domains. Accessed on 4.5.2020 at `https://github.com/redlab-sk/osint-sk-data`.

[2] Atiyat, M. A. Github recon and sensitive data exposure, 2019. Accessed on 21.5.2020 at `https://www.bugcrowd.com/resources/webinars/github-recon-and-sensitive-data-exposure/`.

[3] Bazzell, M. *Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information*, 7 ed. Independently published, 2019.

[4] Bodo, L. How to trace social media users across multiple platforms, 2020. Accessed on 14.3.2020 at `https://osintcurio.us/2020/03/02/how-to-trace-social-media-users-across-multiple-platforms/`.

[5] Commission, E., 2008. Accessed on 23.5.2020 at `https://simap.ted.europa.eu/documents/10184/36234/cpv_2008_guide_en.pdf`.

[6] da Veiga, M. S. Osint metadata collecting for reconnaissance, 2019. Accessed on 23.5.2020 at `https://medium.com/hacker-toolbelt/osint-metadata-collecting-for-reconnaissance-6b3ff18ddbfe`.

[7] Das, R. Red teaming overview, assessment & methodology, 2019. Accessed on 23.5.2020 at `https://resources.infosecinstitute.com/red-teaming-overview-assessment-methodology/`.

[8] Dushantha, S. Sherlock project. Accessed on 23.5,2020 at `https://github.com/sherlock-project/sherlock`.

[9] Gonzales, C. Two europol stopchildabuse images geolocated, 2020. Accessed on 23.01.2020 at `https://www.bellingcat.com/news/2019/12/05/two-europol-stopchildabuse-images-geolocated-part-i-madagascar/`.

[10] Hacek, J. *Práca novinára s otvorenými zdrojmi a dátami.* Bratislava, Stimul, 02 2020.

[11] Hudák, P. Asset discovery: Doing reconnaissance the hard way, 2018. Accessed on 14.3.2020 at `https://0xpatrik.com/asset-discovery/`.

[12] Hudák, P. Osint primer: Domains (part 1), 2018. Accessed on 23.5.2020 at `https://0xpatrik.com/osint-domains/`.

[13] Hudák, P. Asset discovery: Doing reconnaissance the hard way, 2019. Accessed on 14.3.2020 at `https://0xpatrik.com/subdomain-enumeration-2019/`.

[14] Kerner, R. Reconnaissance: A walkthrough of the "apt" intelligence gathering process, 2015. Accessed at on 14.3.2020 `http://www.kerneronsec.com/2015/10/a-walkthrough-of-apt-intelligence.html`.

[15] Lowenthal, M. M. Osint: The state of the art, the artless state. *Studies in Intelligence*. Accessed at `https://www.cia.gov/library/readingroom/docs/DOC_0006122548.pdf`.

[16] Mercado, S. Sailing the sea of osint in the information age. *Studies in Intelligence 48*, 3 (7 2004). Available at `https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol48no3/pdf/v48i3a05p.pdf`.

[17] MITRE. *PRE-ATT&CK.* Accessed on 23.5.2020 at `https://attack.mitre.org/tactics/pre/`.

[18] Office of the Chairman of the Joint Chiefs of Staff. *DOD Dictionary of Military and Associated Terms*, 2019. Accessed on 23.01.2020 at `https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf`.

[19] Seitz, J. Automatically discover website connections through tracking codes, 2015. Accessed on 23.01.2020 at `http://www.automatingosint.com/blog/2015/08/osint-discover-shared-tracking-code-between-domains/`.

[20] Swayze, S. Using open source intelligence to analyze the competition, 2016. Accessed on 14.3.2020 at `https://www.stewartswayze.com/blog/2016/6/25/using-open-source-intelligence-to-analyze-the-competition`, accessed.

[21] Tange, O. Gnu parallel - the command-line power tool. *;login: The USENIX Magazine 36*, 1 (Feb. 2011), 42–47.

[22] Vigo, M. From email to phone number, a new osint approach, 2019. Accessed on 29.04.2020 at `https://www.martinvigo.com/email2phonenumber/`.

[23] Voľanská, P. Breaking the public procurement act and open sources. Master's thesis, Comenius University in Bratislava. Faculty of Arts, 2019.

[24] Wu, S., and Manber, U. Agrep: A fast approximate pattern-matching tool. In *Proc. of the Winter 1992 USENIX Conference* (San Francisco, California, 1991), pp. 153–162.

Appendix A: Tools

In the thesis, we described two tools that we created to automatize some process:

- We added new capabilities to the tool Sherlock
- We created our tool that can find correlated Slovak websites

Source code of both tools is available on the electronic appendix to the thesis.