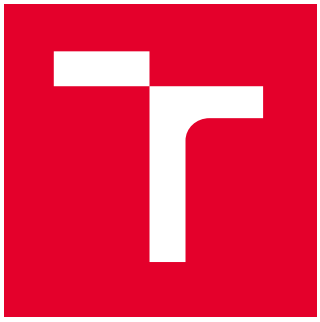


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

MODUL DO ANTISPAMOVÉHO SOFTWARE ZAJIŠŤUJÍCÍ FILTRACI E-MAILŮ ZA VYUŽITÍ GEOLOKACE ODESÍLATELE

ANTISPAM SOFTWARE PLUGIN FOR E-MAIL FILTERING BASED ON GEOLOCATION OF THE SENDER

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE — **Bc. Martin Kovařík**
AUTHOR

VEDOUCÍ PRÁCE — **Ing. Tomáš Caha**
SUPERVISOR

BRNO 2022

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Diplomová práce

magisterský navazující studijní program **Informační bezpečnost**

Ústav telekomunikací

Student: Bc. Martin Kovařík ***ID:*** 190001

Ročník: 2 ***Akademický rok:*** 2021/22

NÁZEV TÉMATU:

Modul do antispamového software zajišťující filtraci e-mailů za využití geolokace odesílatele

POKYNY PRO VYPRACOVÁNÍ:

Nastudujte technologii geolokace IP adresy a problematiku nevyžádané pošty. Vyberte antispamový software, do kterého implementujete rozšíření pro filtraci e-mailů na základě geolokace IP adresy odesílatele. Vyvinuté řešení otestujte a diskutujte výsledky. Vytvořený kód vystavte pod licencí MIT na GitHub.

DOPORUČENÁ LITERATURA:

[1] Linux Dokumentační projekt. 4. vyd. Computer Press, 2008. 1336 s. ISBN: 978-80-251-1525-1.
[2] KERNIGHAN, Brian W. a Dennis M. RITCHIE. Programovací jazyk C. 2. vydání. Přeložil Zbyněk ŠÁVA. Brno: Computer Press, 2019. ISBN 9788025149652.

Termín zadání: 7.2.2022 ***Termín odevzdání:*** 24.5.2022

Vedoucí práce: Ing. Tomáš Caha

doc. Ing. Jan Hajný, Ph.D.
předseda rady studijního programu

ABSTRAKT

Tato diplomová práce se zabývá vytvořením modulu do antispamového softwaru zajišťující filtraci e-mailů za využití geolokace odesílatele. V teoretické části práce se obecně popisuje spam a hrozby, které se šíří e-mailovou komunikací, jako je sociální inženýrství a malware. Následně je popsán princip e-mailu a e-mailových hlaviček, e-mailové filtry i jejich metody na filtraci zpráv a jako poslední je popisována geolokace IP adresy a způsoby zjištění polohy síťového zařízení. Vytvořený modul Geolock pro filtraci e-mailů byl vyvinut pro antispamový software SpamAssassin a používá geolokační databázi IP2Location. Je znázorněno zakomponování modulu do programu SpamAssassin a je popsán samotný modul a jeho jednotlivé části. Modul je otestován na datasetu a je naznačeno jeho reálné použití. Modul Geolock je zveřejněný na platformě GitHub pod licencí MIT (https://github.com/MartinKovarik/Geolock).

KLÍČOVÁ SLOVA

e-mailový filtr, geolokace IP adresy, Geolock, IP2Location, modul, Perl, plugin, spam, SpamAssassin

ABSTRACT

This thesis deals with the creation of a module for antispam software providing email filtering using sender geolocation. The theoretical part of the thesis describes spam in general and the threats that spread through email communication, such as social engineering and malware. Then the principle of email and email headers, email filters and their methods to filter messages are described and lastly IP geolocation and machine location methods are described.The created filtering module Geolock was designed for the SpamAssassin antispam software and uses the IP2Location geolocation database. The embedding of the module into SpamAssassin is shown and the module itself and its different parts are described. The module is tested on a dataset and a real-life application is outlined. The Geolock module is published on GitHub under the MIT license (https://github.com/MartinKovarik/Geolock).

KEYWORDS

e-mail filter, Geolock, IP2Location, module, IP address geolocation, Perl, plugin, spam, SpamAssassin

KOVAŘÍK, Martin. *Modul do antispamového software zajišťující filtraci e-mailů za využití geolokace odesílatele*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2022, 72 s. Diplomová práce. Vedoucí práce: Ing. Tomáš Caha

Prohlášení autora o původnosti díla

Jméno a příjmení autora:	Bc. Martin Kovařík
VUT ID autora:	190001
Typ práce:	Diplomová práce
Akademický rok:	2021/22
Téma závěrečné práce:	Modul do antispamového software zajišťující filtraci e-mailů za využití geolokace odesílatele

Prohlašuji, že svou závěrečnou práci jsem vypracoval samostatně pod vedením vedoucí/ho závěrečné práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno ....

podpis autora*

*Autor podepisuje pouze v tištěné verzi.

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu Ing.Tomáši Cahovi za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Obsah

Seznam obrázků

Seznam tabulek

Seznam výpisů

Úvod

Každému člověku jednou za čas do e-mailové schránky dojde podivná zpráva nebo e-mail nabízející různorodé produkty a služby. Pokud se však uživatel podívá do složky spam, uvidí, že takovýchto zpráv dostává pravidelně vícero, ale jsou automaticky filtrovány do této složky. Tuto práci provádí e-mailové filtry, které pomocí různých metod analyzují zprávy a hodnotí, zda se jedná o legitimní zprávu, nebo o nežádoucí e-mail. Následně se rozhodnou, zda zprávu přijmout či odeslat, nebo zda má být zpráva odfiltrována. Tímto způsobem zlepšují kvalitu používání e-mailové komunikace a zvyšují bezpečnost, jelikož zprávy nemusí obsahovat pouze „neškodné“ obchodní sdělení, ale mohou požadovat od uživatele jeho osobní či přihlašovací údaje, nebo dokonce obsahovat malware, který po stáhnutí nakazí počítač oběti.

Diplomová práce se zabývá geolokací IP adres a antispamovými programy. Do zvoleného filtrovacího programu SpamAssassin je vytvořen vlastní modul Geolock využívající geolokační databázi IP2Location pro blokaci zpráv na základě IP adresy odesílatele.

V kapitole 1 je v práci popsán obecně spam, sociální inženýrství a malware. Jak nevyžádané sdělení, tak malware a útoky využívající sociálního inženýrství se mohou šířit e-mailovou komunikací a ohrozit uživatele. V sekci zabývající se sociálním inženýrstvím jsou popsány často používané a aktuální hrozby, jako jsou phishing, spear phishing, sextortion nebo vishing. Podobným způsobem je v další sekci obecně popsán malware a podrobněji jeho podkategorie ransomware a cryptojacking, kvůli jejich aktuální využívanosti. V kapitole 2 je popsán e-mail obecně, jeho struktura a jak je posílán. V této sekci je také do detailu popsána e-mailová hlavička, její případné podvrhnutí a možné způsoby, jak se před ním bránit. Kapitola 3 se zabývá obecně e-mailovými filtry a metodami, jež využívají. Jsou zde i uvedeny a dále popsány dva příklady e-mailových filtrů, které se běžně používají. V následující kapitole 4 je popsáno využití geolokace IP adres. Geolokace se dělí podle toho, zda se využívají aktivní, nebo pasivní metody geolokace. Obě kategorie jsou zde popsány a v každé jsou používané metody blíže vysvětlené. Poslední část této kapitoly se věnuje geolokačním databázím a obsahuje několik příkladů používaných databází. V předposlední kapitole 5 se práce zabývá vývojem vlastního modulu *Geolock* pro e-mailový filtr SpamAssassin. Zde je popsána nutná konfigurace SpamAssassinu, popis samotného modulu využívající geolokační databázi IP2Location, ukázka funkce pro zpracování konfiguračního souboru a hlavní funkce odpovědné za vyhodnocení země odesílatele e-mailu. V této sekci je také popsáno zveřejnění modulu *Geolock* na platformě GitHub pod svobodnou licencí MIT. Ve finální kapitole 6 se práce zabývá testováním funkčnosti vytvořeného modulu. Nejdříve je ukázané správné chování na testovací zprávě a její výsledek. Následuje testování na datasetu spamu a legi-

timních zpráv neboli hamu českého uživatele. Výsledná data jsou prezentována ve formě tabulek a grafu. Na závěr jsou diskutovány výsledky a vhodné použití modulu *Geolock*.

1 Spam

Nevyžádané sdělení, které je také známé pod názvem Spam, je nepříjemnost, se kterou se setkal každý uživatel Internetu. Spam má mnoho forem. Nejčastěji se jedná o reklamní e-maily, ale v dnešní době se na něj dá narazit skoro na jakémkoliv místě, jako jsou příspěvky na diskuzních fórech, ve webových vyhledávačích, v SMS zprávách, na sociálních sítích nebo v mobilních aplikacích [1][2].

Spam není zaměřen na konkrétního člověka, ale je šířen na co největší počet možných uživatelů. Vznikem volně dostupných on-line překladačů, jako je např. Google Translate, se rozšířily možnosti spammerů neboli lidí vytvářejících spam, a zjednodušilo jim šíření nevyžádaných sdělení i do rozdílně mluvících zemí.

Šíření nevyžádaných reklamních zpráv je populární z důvodu velice nízké ceny. Často jsou spamy rozesílány z kompromitovaných mail serverů a uživatelských účtů. K těm se útočník může dostat např. zneužitím zranitelnosti v systému nebo využitím populárního sociálního inženýrství.

1.1 Sociální inženýrství

Se spamem úzce souvisí i sociální inženýrství, kde cíl již nemusí být kdokoliv, ale může se zaměřovat na konkrétní osobu. Oproti spamu však představuje větší hrozbu a má mnohem negativnější následky. V posledních letech každoročně roste počet útoků využívajících sociální inženýrství a dá se očekávat, že v nejbližší době tento trend nepřestane. Hlavním důvodem je fakt, že ve většině systémů je nejslabším článkem právě člověk. Stejně jako u spamu zde značně napomáhají online překladače. Naštěstí pro česky mluvící občany je čeština komplexní jazyk. Strojově přeložené texty jsou nedokonalé, a tedy v nich lze najít chyby, které můžou člověku napovědět, že se nejedná o legitimní zprávy [3].

Mezi nejpoužívanější a nejdůležitější typy útoků, které jsou v rámci této práce dále popsány, patří tyto:

- phishing,
- spear phishing,
- sextortion,
- vishing.

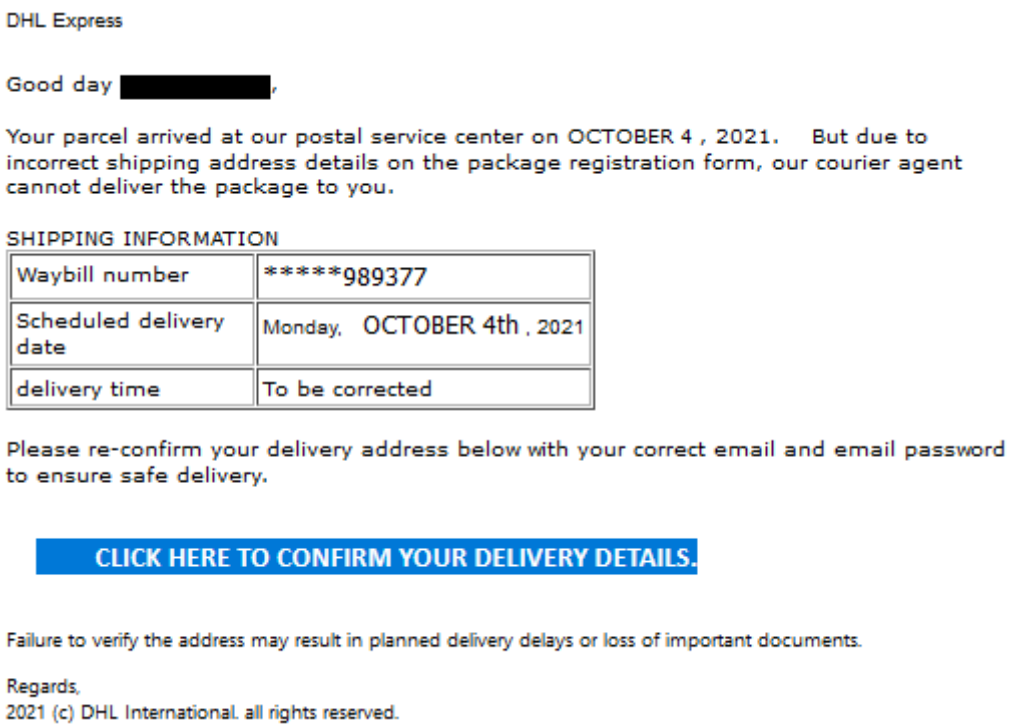
1.1.1 Phishing

Phishing je nerozšířenější typ útoku sociálního inženýrství. Jedná se o podvodné zprávy, které mají za cíl z oběti získat citlivé údaje, jakými mohou být čísla kreditních karet a přihlašovací údaje. Mezi časté phishingové zprávy patří, že útočník,

vydávající se za IT (*Information Technology*) správce, požaduje rychlé přihlášení do e-mailového účtu, jelikož probíhá pročištění systému od neaktivních uživatelů. Zároveň zpráva rovnou obsahuje pomocný odkaz do systému, kde se má uživatel přihlásit. Odkaz na podvodnou stránku bývá zamaskován v HTML (*HyperText Markup Language*) kódu tak, aby vypadal, že vede na legitimní stránku, nebo se používají služby URL (*Uniform Resource Locator*) zkracování, jako jsou například Bitly či Cuttly.

Důležitým faktorem je zde časový pres, který se snaží útočník vytvořit, aby oběť musela jednat co nejrychleji a pod tlakem neověřovala autentičnost zprávy. Běžná časová lhůta je 48 hodin. Pokud se odkazovaná stránka věrohodně podobá legitimní přihlašovací stránce do systému, tak důvěřivý a nezkušený uživatel odevzdá své přihlašovací údaje s dobrým pocitem, že zabránil vymazání své e-mailové schránky. V případě, že uživatel používá stejné heslo i v jiných internetových službách, jako je například internetové bankovnictví, může být taková chyba pro uživatele katastrofální.

Níže na obrázku 1.1 lze vidět phishingovou zprávu, ve které se útočník vydává za mezinárodní společnost DHL a požaduje od uživatele zadání své e-mailové adresy a hesla do formuláře na podvodné stránce, na kterou se uživatel dostane skrze odkaz ve zprávě.



DHL Express

Good day [redacted],

Your parcel arrived at our postal service center on OCTOBER 4 , 2021. But due to incorrect shipping address details on the package registration form, our courier agent cannot deliver the package to you.

SHIPPING INFORMATION

Waybill number	*****989377
Scheduled delivery date	Monday, OCTOBER 4th , 2021
delivery time	To be corrected

Please re-confirm your delivery address below with your correct email and email password to ensure safe delivery.

CLICK HERE TO CONFIRM YOUR DELIVERY DETAILS.

Failure to verify the address may result in planned delivery delays or loss of important documents.

Regards,
2021 (c) DHL International. all rights reserved.

Obr. 1.1: Phishing

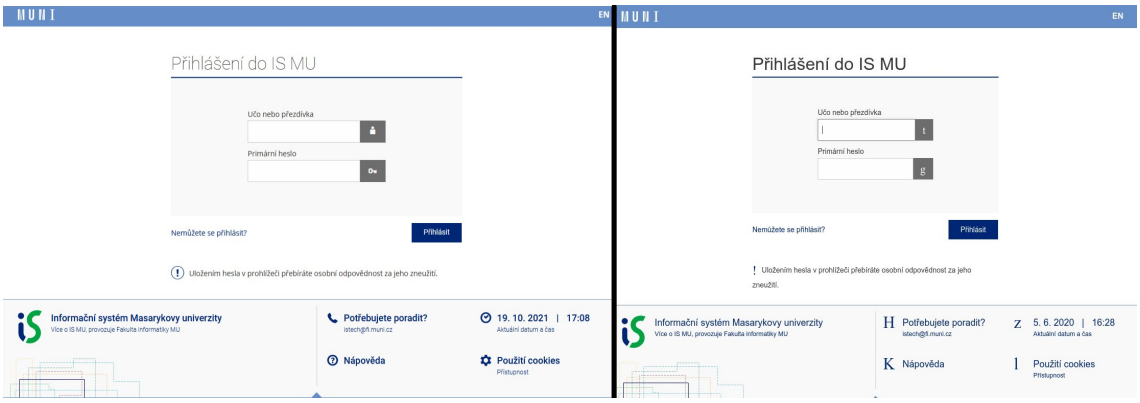
Podkategorie phishingu je také takzvaný reply-to phishing, jež nefunguje na principu vyplnění informací na zaslaném odkaze, ale o poslání informací přímo v e-mailu. Příkladem mohou být klasické zprávy o výhře v loterii nebo umírajícím arabském šejkovi bez potomka, který hledá, komu předat svůj majetek. Útočník se snaží oběť přesvědčit, že zrovna on je, čistě náhodou, šťastný jedinec a jediné co zbývá, je obratem zaslat svoje číslo účtu, jméno, příjmení, telefonní číslo, adresu bydliště, datum narození a přihlašovací údaje.

1.1.2 Spear phishing

Na rozdíl od běžného phishingu se jedná o cíleně zaměřené zprávy například na univerzitu či organizaci. Populární je použití techniky takzvané *typo squatting* [4], která spočívá ve špatně napsaném URL stránky. Na stránku může člověk narazit i omylem, tedy překlepem při psaní URL stránky, kde byly např. zaregistrované stránky foogle.com, hoogle.com, boogle.com, apod., které byly zvoleny kvůli blízkosti prvního písmene na qwerty klávesnici ke správnému znaku „g“ v doméně google.com. V podvodných zprávách se využívá spíše zaměnění vizuálně podobných znaků, jako je například „rn“ a „m“, nebo přidávání a odebírání znaků, jako je například facebok.com namísto správně napsané domény facebook.com.

Další forma pokusu o obelstění oběti je zneužití domény nejvyššího řádu (TLD - *Top Level Domain*) a útočníci si registrující název domény s doménou nejvyššího řádu *.cm*, který je TLD pro Kamerun, *.co*, která je doména nejvyššího řádu pro Kolumbii, a *.om*, což je TLD pro Omán.

Na obrázku 1.2 je ukázáno porovnání legitimní přihlašovací stránky do informačního systému IS Masarykovy univerzity (vlevo) a podvodné přihlašovací stránky (vpravo) zaslané studentům a zaměstnancům v roce 2020.



Obr. 1.2: Spear Phishing

1.1.3 Sextortion

Název Sextortion je spojení slov *Sex* a *Extortion* neboli vydírání. Sextortion je forma vyhrožování přes zprávy, ve kterých typicky útočník tvrdí, že získal přístup ke stroji oběti a sleduje uživatele přes web kameru. Dále vyhrožuje, že nahrál videa, kde oběť sleduje pornografický obsah, a je připraven tyto videa zveřejnit zasláním jeho rodině, kolegům a kamarádům nebo publikováním na sociálních sítích, pokud nedostane zaplaceno v kryptoměně.

Pro zvýšení věrohodnosti těmto výhružným zprávám, útočníci oběti vysvětlují, proč antivirus na počítači nic nezachytil. Dále je běžné podvrhnutí e-mailových hlaviček, ve kterých jako odesílatele útočník nastaví e-mailovou adresu oběti, aby ji přesvědčil o přístupu k jejímu stroji. Pokud dříve uniklo heslo spojené s e-mailovou adresou v nějakém úniku dat společnosti, tak může být zasláno v rámci zprávy, aby se opět zvýšila věrohodnost. Kombinací těchto taktik a již zmíněného časového presu, dokdy musí oběť zaplatit, přesvědčí dostatek počítačově negramotných lidí, aby požadovanou částku zaplatili.

Ve výpisu 1.1 je ukázka vyděračské zprávy požadující zaplacení v kryptoměně. Dále pak na obrázku 1.3 lze vidět, že někdo útočníkovi uvěřil a zaplatil požadovanou částku.

Obr. 1.3: Sextortion

V roce 2019 FBI IC3 (Federal Bureau of Investigation Internet Crime Complaints Center) informovalo, že v předešlém roce bylo nahlášeno 51 146 případů vydírání, což je navýšení výskytu vyděračných zpráv o 242 % oproti roku 2017. Ty způsobily ztráty kolem 83 milionů dolarů. Z těchto nahlášených případů byla většina právě sextortion [5]. Tato kategorie zahrnuje:

- DoS (*Denial of Service*) útoky;
- sextortion;
- schéma nájemného vraha - vyhrožování, kdy útočník se vydává za nájemného vraha a požaduje finanční částku, jinak oběť a/nebo jeho rodinu zabije;
- napodobování vlády - útočník se vydává za vládního úředníka a pokouší se od oběti získat peníze;

- schéma půjčky - útočník se vydává za vymahače dluhů a požaduje zaplacení, aby se oběť vyhnula právním následkům;
- únik citlivých dat.

Výpis 1.1: Ukázka sextortion

```
Ahoj!
Už sis všiml, že jsem ti nedávno poslal z tvého účtu emailovou zprávu?
Ano, znamená to, že mám neomezený přístup k tvému zařízení.

Několik posledních měsíců tě sleduji.
Přemýšlíš, jak je to možné? No, tvůj systém byl infikován malwarem pocházejícím z
    erotického webu, který jsi navštívil. Možná těmhle věcem nerozumíš, a tak se ti
     to pokusím vysvětlit.

S pomocí viru trojského koně jsem získal kompletní přístup k tvému PC a všem dalším
     zařízením.
To jednoduše znamená, že tě můžu pouhým zapnutím tvé kamery a mikrofonu kdykoliv
    sledovat, aniž by sis čehokoliv všimnul. Kromě toho mám také přístup k seznamu
    tvých kontaktů a veškeré korespondenci.

Možná si teď říkáš: Vždyť mám na počítači aktivní antivirus, tak jak k tomu mohlo
    dojít. Jak to, že jsem nedostal žádné upozornění?
Odpověď je prostá: můj malware totiž používá ovladače, které každé čtyři hodiny
    aktualizují podpisy, díky čemuž je nezjistitelný a tvůj antivirový program o ně
    m nemá tušení.

Mám k dispozici video, na kterém si v jeho levé části honíš, zatímco v pravé
    polovině obrazovky se přehrává klip, při jehož sledování masturbuješ.
A chtěl bys vědět, co ti s ním můžu způsobit? Jediným kliknutím myši můžu tohle
    video odeslat na všechny používané sociální sítě a všem tvým emailovým kontaktů
    m.
Stejně tak mohu sdílet přístup k veškeré tvé emailové korespondenci a messengerům,
    jehož služby využíváš.

Pokud tomu chceš zabránit, jediné, co pro to musíš udělat,
je â€“ převést na moji bitcoinovou adresu bitcoiny v hodnotě 26000 Kč (pokud nevíš,
     jak to udělat, stačí si otevřít prohlížeč a do vyhledávače zadat výraz: Koupit
     bitcoiny).

Moje bitcoinová adresa (BTC Wallet) je: 1PGygSANCgDfrh5oyen5SboaPjmsJ8PDp2

Jakmile obdržím potvrzení o platbě, video okamžitě smažu a je hotovo. Už o mně
    nikdy víc neuslyšíš.
Na provedení transakce máš 2 dny (48 hodin).
Po otevření tohoto emailu dostanu upozornění a časomíra se spustí.

Jakýkoliv pokus o podání stížnosti k ničemu nevede, neboť tento email nelze zpětně
    dohledat a totéž platí i pro moje bitcoinové ID.
Na tomhle jsem totiž pracoval hodně dlouhou dobu a chybám jsem nedal žádný prostor.

Pokud bych se jakýmkoliv způsobem dozvěděl, že jsi tuhle zprávu ukázal nějaké třetí
     osobě, výše uvedené video okamžitě zveřejním.
```

1.1.4 Vishing

Vishing je hlasový phishing. Přestože není jako ostatní zmíněné kategorie šířen přes zprávy, je uveden, jelikož se jedná o velice aktuální hrozbu, která se ve větším počtu začala objevovat až roku 2020 a přetrvává do současnosti. Primárně se jedná o útočníka, který zavolá na telefon oběti a vydává se za zaměstnance bankovní instituce. Oběti je například sděleno, že je jí blokována karta nebo její bankovní účet byl napaden a její finance jsou v ohrožení. Tímto způsobem se pak dále snaží získat informace, jako jsou přihlašovací údaje k internetovému bankovnictví, číslo účtu, PIN kód, CVV (*Card Verification Value*) a CVC (*Card Verification Code*) kódy, platnost karty a podobně. Útočníci také můžou požadovat, aby uživatel přesunul své finance na bezpečný účet, jelikož ten jeho byl zrovna napaden a musí jednat rychle. Přestože je mu slibováno, že účet patří bance a následně mu budou peníze vráceny, tak účet patří útočníkovi a žádné peníze zpátky nedostane.

Stejně jako u klasického phishingu, zde útočníci podvrhují svá telefonní čísla, díky čemuž hovor vypadá, jako by doopravdy pocházel z legitimního čísla banky oběti. Někdy bývá tento hovor následován dalším podvrhnutým hovorem od Policie České republiky, aby se zvýšila věrohodnost a závažnost situace.

Tento rok na tyto podvody vydalo varování mimo jiné Policie České republiky [6], NÚKIB [7] (Národní úřad pro kybernetickou a informační bezpečnost), ESET [8] a komerční banky, jako jsou Česká spořitelna [9], ČSOB, AirBank nebo Česká národní banka [10].

1.2 Malware

Malware neboli škodlivý software je označení pro program, který byl vytvořen se záměrem škodit. Slovo malware se skládá z anglických slov *malicious*, tedy zákeřný či zlomyslný, a software. Stejně jako u sociálního inženýrství je zde hlavní cíl finanční obohacení útočníka. Objevuje se ale i škodlivý software, který poškodí nebo vymaže všechna data na stroji a nemá žádný jiný účel. Tedy nemusí se vždy jednat pouze o získání peněžních prostředků, ale může jít o útočníkovo dokazování si svých schopností a pro pocit moci.

Nakažení stroje může nastat mnoha způsoby, kde nejčastějšími jsou tyto:

- Zaslání škodlivého softwaru e-mailem.
- Stáhnutí si souboru, o kterém si oběť myslí, že je věrohodný - např. stáhnutí filmu zadarmo ze stránky nebo přes torrent.
- Škodlivý software je součástí jiného legitimního souboru - např. instalační soubor pro multimediální přehrávač.
- Útočník má fyzický přístup ke stroji.

- Oběť připojí ke stroji nalezený infikovaný USB flash disk.

Samotný pojem malware je dosti široký a je rozdělen do několika podkategorií, ale tyto podkategorie nejsou výlučné, a tedy daný škodlivý software může spadat pod více kategorií zároveň [11][12].

Nejdůležitější podkategorie jsou popsány níže:

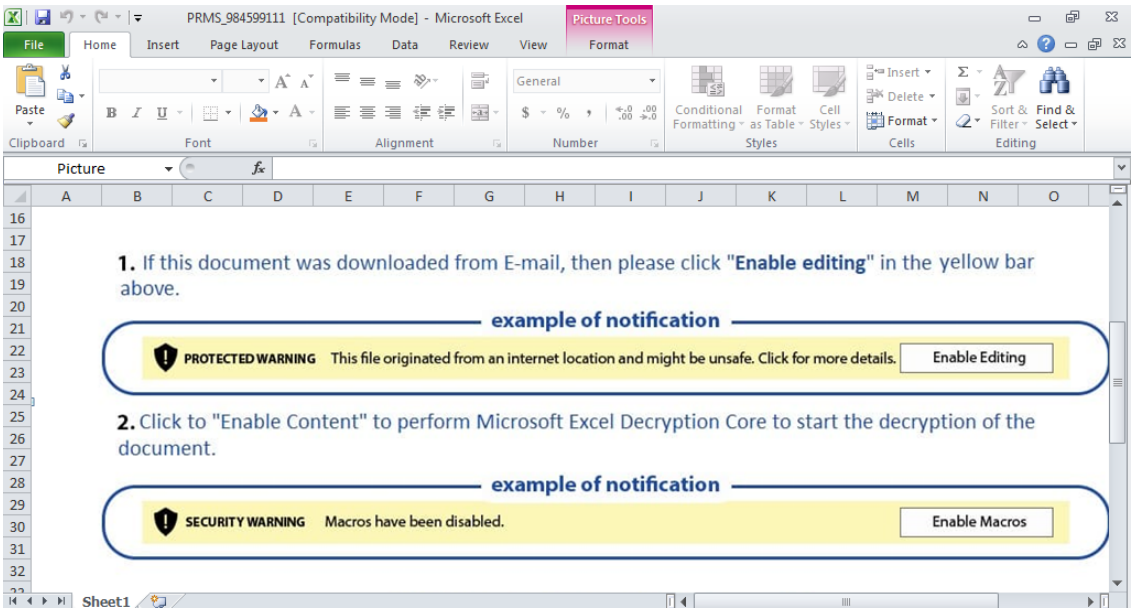
- **Počítačový virus** - jde o počítačový program, který, stejně jako biologický virus, potřebuje hostitele, aby mohl přežít a šířit se dále. Počítačový virus se schovává v rámci jiného programu a je schopen udělat kopie sám sebe. Při spuštění tohoto infikovaného programu vloží tyto kopie do ostatních programů a souborů, čímž je nakazí. Tímto způsobem může například zničit data obsažená v infikovaném souboru nebo celkově získat kontrolu nad strojem uživatele [13].
- **Červ** - jde o samostatný malware, který se sám aktivně snaží šířit po síti a pokouší se infikovat ostatní stroje využitím zranitelností. Na rozdíl od počítačového viru není spojen s dalším souborem a může se šířit bez spuštění od samotného uživatele. Nakažení stroje lze například zpozorovat v síťovém provozu stroje, kde skenuje v síti port 8291, který slouží pro komunikaci s MikroTik routerem, a snaží se zde využít zranitelností.
- **Trojský kůň** - známý také pod zkráceným názvem Trojan. Jedná se o program pojmenovaný podle příběhu z Antického Řecka, kde byl Trojský kůň použit pro infiltraci města Trója. Tento škodlivý software potřebuje pomoc uživatele, aby se mohl spustit. Proto se maskuje jako normální a užitečný program. Na rozdíl od počítačového viru a červa, se trojský kůň nesnaží infikovat další soubory nebo se dále šířit na jiné stroje.
- **Rootkit** - snaží se získat přístupová práva administrátora a sám sebe zamaskovat a schovat se tak před zrakem uživatele nebo antivirového softwaru.

Na obrázku 1.4 je ukázka malwaru stáhnutého z e-mailové zprávy, který je ve formě tabulky z programu Microsoft Excel a požaduje povolení maker. Pokud uživatel tyto makra povolí, soubor infikuje počítač.

Dále se dá malware rozdělit podle účelu, kde nejpopulárnější skupinu tvoří ransomware nebo cryptojacking.

1.2.1 Ransomware

Žhavým tématem posledních let je ransomware neboli vyděračský software. Ransomware zašifruje všechny soubory na stroji oběti a pokouší se šířit po síti dále. Pro dešifrování souborů je potřeba klíč, který je útočník ochotný poskytnout za výkupné. Tyto částky se v případě větších firem pohybují v řádech milionů korun. Avšak ani po zaplacení není jisté, že útočník poskytne klíč k dešifraci souborů nebo že dešifrace proběhne v pořádku pro všechny soubory [14].



Obr. 1.4: Příloha e-mailu obsahující škodlivý software

Útočník může vydírat společnost i bez zašifrování dat, a to výhrůžkou zveřejnění důvěrných souborů organizace, které často obsahují osobní údaje. Za únik osobních údajů v Evropské unii může Úřad pro ochranu osobních údajů udělit pokutu až 10 000 000 Eur, nebo až 2 % celkového celosvětového ročního obratu [15].

Podle zprávy „Stav ransomwaru 2021“ (*State of Ransomware 2021*) [16] společnosti Sophos, která se zabývá bezpečností, cena za napadení vyděračským softwarem se pohybuje okolo 1,85 milionů dolarů, což je v přepočtu 42 100 450 Kč (podle kurzu ke dni 25. 11. 2021). Tato částka zahrnuje nejenom výkupné, ale započítává i cenu prostoje, čas zaměstnanců, náklady na zařízení a síť, ušlý zisk a podobné. Taková částka může ochromit leckterou firmu. Znepokojivě je tato částka vyšší oproti roku 2020, kdy činila 0,76 milionů dolarů, tedy se během roku zvýšila na více jak dvojnásobek. Tyto statistiky pochází z 5400 společností po celém světě, které se pohybují od 100 do 5000 zaměstnanců. Z těchto společností nahlásilo 37 % , že bylo terčem ransomwaru, z toho v 54 % případů útočník uspěl v zašifrování dat. Po zaplacení výkupného ale bylo v průměru obnoveno pouze 65 % dat.

V České republice v roce 2019 napadl útočník ransomwarem Nemocnici Rudolfa a Stefanie v Benešově [17] a na začátku roku 2020 Fakultní nemocnici Brno [18]. NÚKIB (Národní úřad pro kybernetickou a informační bezpečnost) varoval, že má dojít k dalším významným útokům, a to směřovaným mimo jiné na zdravotnictví, energetický sektor a akademický sektor [19].

Obrázek 1.5 ukazuje okno, které se oběti objeví na počítači po nakažení a zašifrování dat vyděračským softwarem *WannaCry*, který v roce 2017 infikoval stroje po

celém světě. Celkově nakazil přes 300 000 počítačů a šířil se skrze nechvalně známý exploit *EternalBlue*.

Obr. 1.5: Výsledek nakažení ransomwarem WannaCry

1.2.2 Cryptojacking

S růstem cen a zájmu o kryptoměnu vzrostl i počet škodlivého kódu, který se snaží těžit kryptoměny na strojích obětí. Tento malware se na rozdíl od ransomwaru může spustit při návštěvě webových stránek, která mají v sobě Javascript kód. Není nutné tedy nic stahovat či instalovat. Stroj oběti pro útočníka těží kryptoměnu a vydělává mu. Tento kód může být vložen po útoku na stránky, nebo si ho majitel může dobrovolně do stránek dát sám. Příkladem je torrentová stránka *The Pirate Bay*, která těžila kryptoměnu Monero na strojích návštěvníků webových stránek [20]. Těžba na stroji oběti nemusí být pouze přes webové stránky, ale může se šířit stejně jako ransomware či jiný malware.

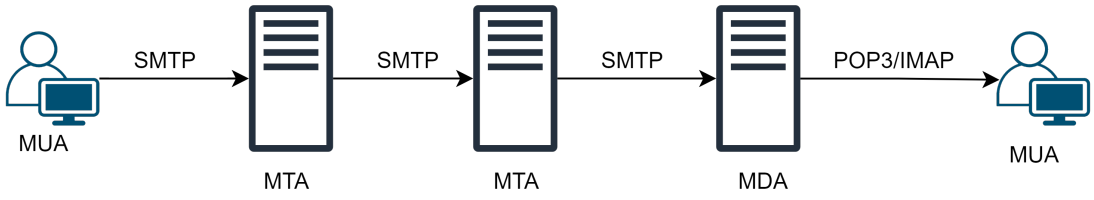
Těžba kryptoměny může vést k přehřívání stroje oběti, k poškození stroje, ke zkrácení životnosti stroje z důvodu neustálého vysokého zatížení, ke snížení výkonu stroje při jiných aktivitách a k vyšší spotřebě elektřiny [21].

2 E-mail

E-mail je elektronická zpráva přenesená od jednoho uživatele k druhému. Zpráva může mít pár řádků textu nebo obsahovat obrázky či jiné soubory. E-mail je tvořen dvěma částmi, a to hlavičkou a tělem. Tělem zprávy jsou právě ony řádky textu nebo přiložené soubory, zatímco hlavičky obsahují různé informace o e-mailu, jakými například jsou údaje o odesílateli a adresátovi či jakými poštovními servery zpráva prošla.

E-mail je vytvořen klientem v poštovním programu neboli MUA (*Mail User Agent*). Příklad poštovního programu může být Microsoft Outlook nebo Mozilla Thunderbird. Po odeslání je e-mail směrován přes poštovní servery až k cílovému poštovnímu serveru odesílatele. Těmto poštovním serverům se říká MTA (*Mail Transfer Agent*) a jejich úkol je vzít zprávu od MUA, dekódovat její hlavičku a určit, kam má být dále směrována. V MTA převzetí zprávy od MUA zajišťuje MSA (*Message Submission Agent*), který je ve většině případů součástí MTA, ale může také být samostatným prvkem komunikace. Při této akci také MTA přidají informaci do hlavičky e-mailu, že zpráva šla přes tento poštovní server, odkud byla převzána a čas. Příklady MTA jsou Sendmail či Postfix. Další prvek v cestě je předání zprávy MDA (*Mail Delivery Agent*), který třídí a doručuje zprávy do schránky na serveru. K těmto zprávám pak uživatelé přistupují pomocí MUA. [22]

Pro přístup ke zprávám používají MUA protokoly IMAP (*Internet Message Access Protocol*) nebo POP3 (*Post Office Protocol version 3*). Hlavní rozdíl je, že POP3 zprávy stáhne a na serveru e-maily vymaže, zatímco IMAP je udržuje na serveru. Pro posílání zpráv se používá protokol SMTP (*Simple Mail Transfer Protocol*). [23] Cesta e-mailu je znázorněna na obrázku 2.1.



Obr. 2.1: Cesta e-mailu

2.1 E-mailová hlavička

Struktura e-mailu a jeho hlavičky jsou definovány v RFC 5322 [24]. Ukázka zkrácené hlavičky nevyžádané pošty lze vidět ve výpisu 2.1.

Výpis 2.1: Zkrácená hlavička podvodného e-mailu

```
Received: from maddog.nerdytechs.com (maddog.nerdytechs.com [108.170.35.130])
  by email-smtpd22.ng.seznam.cz (Seznam SMTPD 1.3.136) with ESMTP;
  Wed, 06 Apr 2022 19:40:59 +0200 (CEST)
Received: from [127.0.0.1] (port=57130 helo=urcbdplus.com)
  by maddog.nerdytechs.com with esmtp (Exim 4.95)
  (envelope-from <s6400cctje@urcbdplus.com>)
  id 1nbfel-0001Zl-Cn
  for anonymous@seznam.cz;
  Tue, 05 Apr 2022 05:39:22 -0400
Date: Tue, 5 Apr 2022 09:39:22 +0000
From: Scott Godfrey <scottgodfrey86@gmail.com>
Reply-To: scottgodfrey1000@gmail.com
Message-ID: <f601f3cbc9959365fa5fd347327ca1f2@urcbdplus.com>
MIME-Version: 1.0
Content-Type: multipart/alternative;
  boundary="b1_f601f3cbc9959365fa5fd347327ca1f2"
To: anonymous@seznam.cz
Subject: =?UTF-8?Q?M=C3=A1te_dar?=
```

Mezi nejdůležitější pole hlavičky patří tyto:

- **To:** Obsahuje e-mailovou adresu jednoho nebo více primárních příjemců zprávy.
- **From:** Specifikuje adresu odesílatele zprávy.
- **Cc:** Zkratka znamenající „Carbon Copy“, která má původ ve vytváření kopií na psacím stroji za využití uhlíkového papíru. Obsahuje adresy ostatních příjemců zprávy, na které obsah zprávy nemusí být přímo cílen.
- **Bcc:** Zkratka „Blind Carbon Copy“ znamenající skrytou kopii. Obsahuje adresy příjemců e-mailu, kteří nejsou ostatním adresátům zveřejněni. Při využití *Bcc* záleží na implementaci, jak se toto pole bude zobrazovat příjemcům zprávy. Jsou 3 různé způsoby:
  - Všem příjemcům, včetně adresátům ve skryté kopii, je toto pole z hlavičky odebráno.
  - Příjemcům je pole *Bcc:* skryto a adresát skryté kopie dostane celou hlavičku včetně *Bcc:* pole. V případě, že je více příjemců skryté kopie, tak je každému poslána hlavička obsahující pouze jeho adresu v tomto poli.
  - Hlavička zprávy zaslané příjemcům obsahuje prázdné pole *Bcc:*, kterým informuje příjemce, že byla někomu poslána skrytá kopie.
- **Reply-to:** Udává, na kterou adresu či adresy navrhuje autor zprávy, aby v případě odpovědi byla zpráva zaslána. Pokud toto pole není použito, tak by měla odpověď standardně jít na adresu nebo adresy uvedené v poli *From:*.
- **Subject:** Předmět zprávy. Obsahuje textový řetězec identifikující téma zprávy. V případě odpovědi může pole začínat textovým řetězcem „*Re:*“ následované původním předmětem zprávy.
- **Date:** Datum a čas. Typicky obsahuje den v týdnu, datum a místní čas uvedený s časovým pásmem, kdy byla zpráva odeslána.

- **Message-ID:** Automaticky vygenerované unikátní pole, které identifikuje konkrétní verzi zprávy. Další verze zprávy, jako je třeba odpověď na původní zprávu, mají své vlastní unikátní identifikátory.
- **Received:** Obsahuje cestu zprávy po jednotlivých serverech od odesílatele k adresátovi. Servery *Received:* sekce přidávají na vrchol a tedy chronologicky lze cestu zprávy sledovat od poslední položky k první. Servery nesmí nijak upravovat předešlé *Received:* řádky, které již v hlavičce jsou. Struktura *Received:* hlavičky je následující:
  - *From* - od jakého serveru byla zpráva přijata a jeho název a IP adresa.
  - *By* - který server zprávu přijal a přidal *Received:* řádek. V závorce je software, který stroj používá (například Postfix).
  - *With* - jaký mail protokol byl použitý (například SMTP nebo ESMTP).
  - *For* - obsahuje adresu adresáta.
  - *Id* - přiřazená identifikace pro logování.
  - *Časový údaj* - datum a čas většinou udávaný v místním čase stroje.

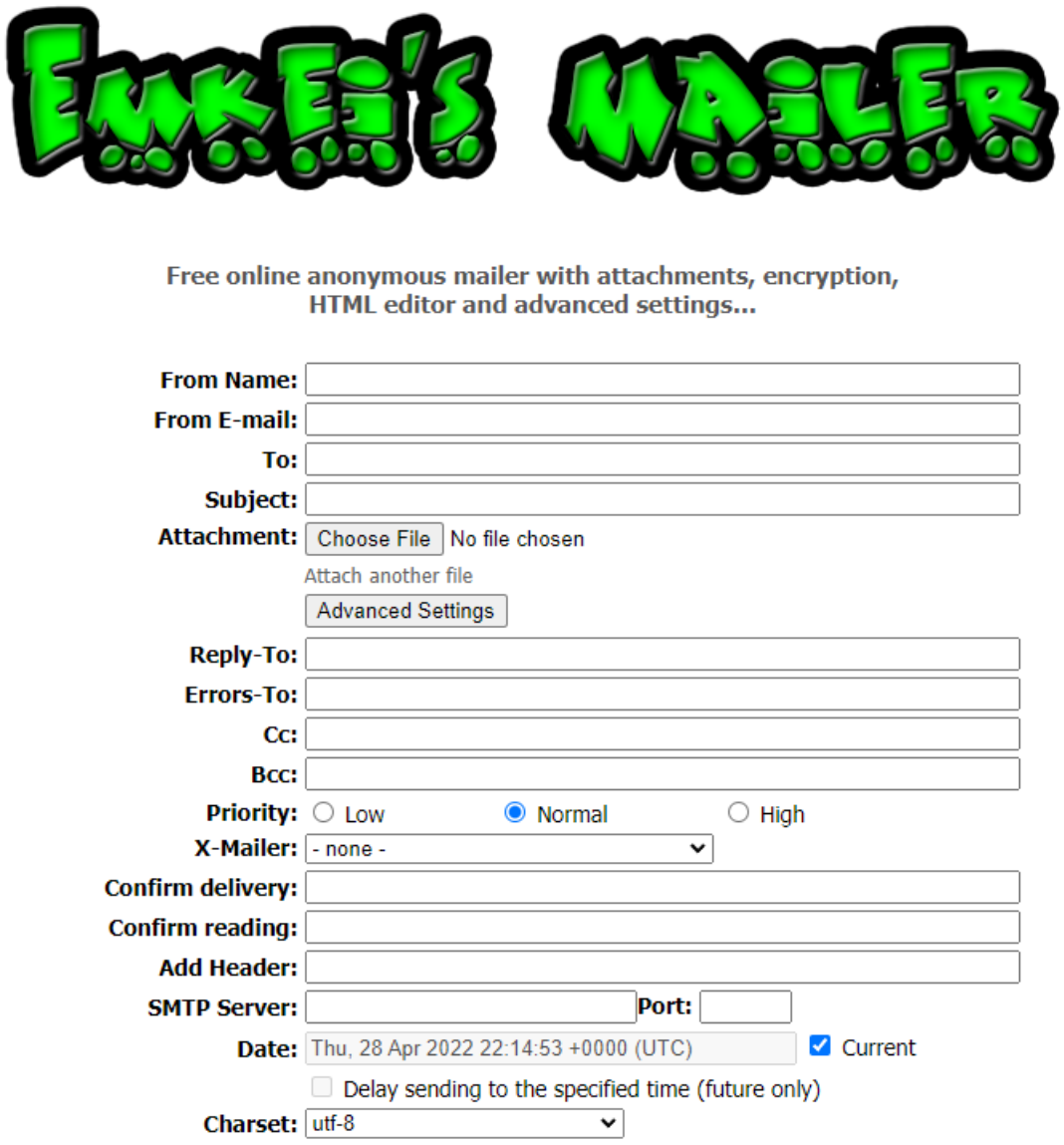
  Ne všechny tyto části se musí nacházet v *Received:* hlavičce nebo mohou mít jinou formu. Příklady na sebe nezávislých *Received:* hlaviček lze vidět ve výpisu 2.2.

Výpis 2.2: Ukázka nezávislých *Received:* hlaviček

```
Received: from msmtp-cc.iq.pl (unknown [86.111.240.236])
  by email-smtpd15.ko.seznam.cz (Seznam SMTPD 1.3.136) with ESMTP;
  Fri, 18 Mar 2022 12:05:03 +0100 (CET)
---------------------------------------------------------------------------
Received: from mailgw1.ro.vutbr.cz (147.229.1.78) by
  VE1EUR02FT048.mail.protection.outlook.com (10.152.13.177) with Microsoft SMTP
  Server (version=TLS1_2, cipher=TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384) id
  15.20.5102.18 via Frontend Transport; Tue, 29 Mar 2022 17:55:56 +0000
---------------------------------------------------------------------------
Received: from a6-100.smtp-out.eu-west-1.amazonses.com (a6-100.smtp-out.eu-west-1.
    amazonses.com [54.240.6.100])
  (using TLSv1.2 with cipher ECDHE-RSA-AES128-SHA256 (128/128 bits))
  (No client certificate requested)
  by mailgw1.ro.vutbr.cz (Postfix) with ESMTPS id 96BBD303B926
  for <xkovar79@stud.feec.vutbr.cz>; Tue, 29 Mar 2022 19:55:55 +0200 (CEST)
---------------------------------------------------------------------------
Received: from njmta-149.sailthru.com (173.228.155.149)
  by ny1mta-39.sailthru.com id hv7adm1qqbse
  for <anonymous@gmail.com>;
  Sat, 9 Jan 2021 11:05:44 -0500
  (envelope-from <delivery_20210109110544.22601888.127578@mx.sailthru.com>)
---------------------------------------------------------------------------
Received: from [2.57.169.135]
    by webmail.bellaliant.net with HTTP; Tue, 26 Apr 2022 04:32:17 -0400
---------------------------------------------------------------------------
Received: from User (localhost [127.0.0.1])
  by ns342529.ip-188-165-224.eu (8.15.2/8.15.2/Debian-14~deb10u1) with SMTP id 23
      F3bCOi025435;  Fri, 15 Apr 2022 05:37:13 +0200
```

2.2 Podvrhnutí e-mailových hlaviček

Podvrhnutí hlaviček e-mailu patří do běžného arzenálu člověka, jež rozesílá spam, phishing a jiné podvodné zprávy. Úprava hlavičky není složitá a existují i online nástroje, jako je například Emkei's Anonymous Mailer [25], který lze vidět na obrázku 2.2.



Obr. 2.2: emkei.cz

Primárně se upravuje pole *From:*, aby si oběť myslela, že jde zpráva od důvěryhodné osoby či autority. Upravit však lze libovolné pole. Nejvěrohodnější částí e-mailové hlavičky jsou pole *Received:*. Tato pole se dají také podvrhnout, ale jakmile je zpráva poslána do světa, legitimní poštovní servery začnou přidávat své vlastní

a korektní *Received:* hlavičky. Občas lze poznat podvrhnuté *Received:* pole podle nekonzistentních dat, jako je velká časová mezera mezi za sebou jdoucími *Received:* hlavičkami, nebo dokonce posunutí času do budoucnosti. Dalším znakem podvržení mohou být chyby ve formátování. Mnoho útočníků si na tyto věci dají pozor, a proto není vždy odhalení podvrhnuté hlavičky přímočaré.

Existují způsoby, jak se chránit před spamem a zprávami s podvrženými hlavičkami. Jedná se o tyto metody:

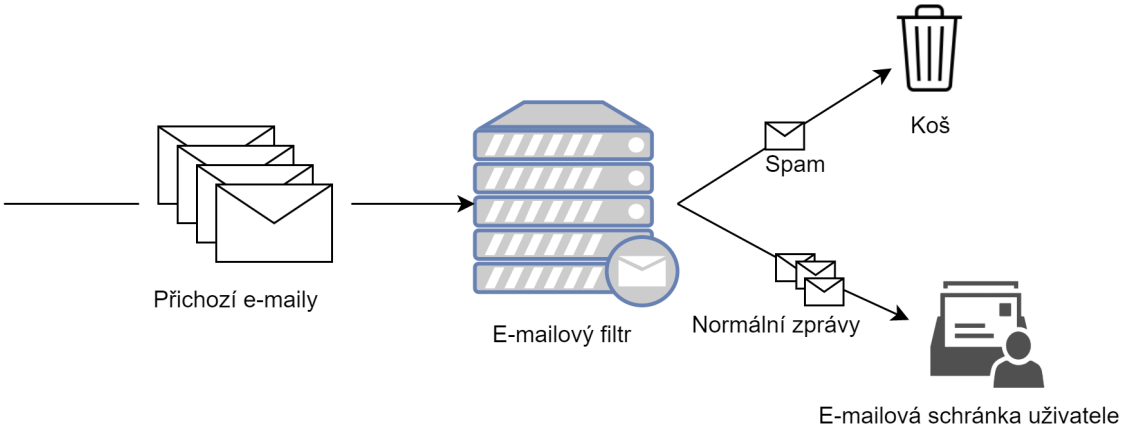
- **SPF** (*Sender Policy Framework*) - administrátor zveřejňuje, které IP adresy MTA mohou posílat zprávy z jeho domény. Pokud tedy útočník podvrhne adresu odesílatele, adresát si může ověřit, že zpráva nebyla odeslána z důvěryhodných IP adres, a tedy pravděpodobně se nejedná o legitimní zprávu. [26]
- **DKIM** (*DomainKeys Identified Mail*) - odesílatel vytvoří MD5 (*Message-Digest algorithm 5*) hash části odesílané zprávy. Následně tento hash zašifruje privátním klíčem a veřejný klíč zveřejní v DNS záznamu. Zašifrovanému hashi se říká DKIM podpis a ten je vložen do e-mailu. Příjemce použije veřejný klíč a dešifruje DKIM podpis. Následně si vytvoří ze stejných částí zprávy vlastní MD5 hash a porovná ho z hashem odesílatele. Pokud jsou totožné, příjemce ví, že e-mail opravdu pochází od odesílatele a že zpráva nebyla nijak po cestě změněna. [27]
- **DMARC** (*Domain-based Message Authentication, Reporting, and Conformance*) - používá kombinaci SPF a DKIM. Administrátor domény si nastaví pravidla DMARC, podle kterých příjemce ověřuje legitimnost zprávy. V rámci DMARC se kontroluje jak SPF, tak DKIM. Aby zpráva byla ověřena, musí úspěšně projít alespoň 1 z nich. [28]

3 E-mailové filtry

Kvůli množství spamu a nepříjemnostem, které způsobuje, vznikly způsoby, jak se mu bránit. Nejefektivnější a nejpopulárnější řešení jsou e-mailové filtry, které můžou být jak bezplatné, tak komerční. E-mailové filtry zajišťují třídění zpráv, kde lze filtrovat jak příchozí, tak odchozí zprávy. Tyto filtry pak rozhodují, zda jde o neškodnou zprávu a e-mail bude přijat či odeslán, nebo zda se jedná o nevyžádanou zprávu, která bude filtrem zahozena.

V této kapitole je popsáno několik metod používaných pro filtrování spamu v e-mailových filtrech [29][30]. Následně jsou krátce uvedeny příklady e-mailových filtrů používaných v praxi.

Na obrázku 3.1 je znázorněná funkce e-mailového filtru.



Obr. 3.1: E-mailový filtr

3.1 Metody používané pro filtrování spamu

Metod používaných na posouzení, zda je zpráva spam, nebo ne, je mnoho. Hlavním kritériem pro dobrou metodu je přesnost odhalení spamu a co nejmenší počet *false positives*, tedy legitimních zpráv, které byly označené jako spam. Zde je popsáno několik běžně používaných filtračních metod. Jde o metody jak primitivní, jako je blacklist a whitelist, tak o metody komplexní, kde jako příklad může posloužit Bayesovo filtrování. Platí zde pravidlo, že čím složitější je metoda, tím větší je náročnost časová a na systémové zdroje.

3.1.1 Blacklist

Jedná se o seznam IP (*Internet Protocol*) adres, ze kterých jsou e-maily automaticky zahazovány. Tyto blacklisty neboli černé listiny může udržovat společnost jenom pro své účely, ale většinou se používají distribuované blacklisty spravované společnostmi, které seznam vlastní a udržují. Příklad takového blacklistu lze uvést Spamhaus nebo Spamcop [31]. Kdokoliv používající tyto blacklisty automaticky blokuje zprávy přicházející z nich. Může ale nastat, že i legitimní odesílatelé se dostanou na tyto listiny. Pokud je například kompromitován účet zaměstnance ve firmě, ze kterého následně útočník posílá spamy do světa, může se mailový server firmy dostat na černou listinu a nikdo z firmy není schopný odesílat zprávy. O odstranění z blacklistu lze spravující společnosti zažádat.

3.1.2 Whitelist

Jedná se o opak blacklistu, kde zprávy jsou přijímány pouze z IP adres uvedených ve whitelistu. Používání pouze této metody by bylo hodně limitující, a proto se používají variace této metody nebo kombinací s jinou metodou filtrace.

3.1.3 Greylist

Novější metoda filtrování je greylist, která využívá faktu, že odesílatelé spamu většinou posílají jenom jednu masivní vlnu a nesnaží se opakovaně kontaktovat stejného uživatele. Tato metoda automaticky odmítne e-mail od neznámého odesílatele. Pokud se pokusí odesílatel poslat další zprávu, tak greylist předpokládá, že se nejedná o spam a již zprávu doručí. Následně je odesílatel přidán na seznam povolených odesílatelů. Přestože tato metoda není náročná na systémové zdroje, tak zpomaluje přenos a může být nevhodná při komunikaci, která vyžaduje rychlou reakci.

3.1.4 Slovní filtry

Jedná se o metodu, která hledá v textu zprávy klíčová slova. Klíčová slova jsou zvolena taková, které se v normální komunikaci často nepoužívají. Tyto můžou být například *porn* nebo *free*. Pokud toto slovo zpráva obsahuje, tak je e-mail označen za spam. Kvalita této metody čistě závisí na zvolených klíčových slovech.

3.1.5 Bayesovo filtrování

Jde o jednu z nejefektivnějších metod detekce spamových zpráv. Vychází z Bayesovy věty z teorie pravděpodobnosti. Hlavní myšlenka této metody je, že většina událostí

je mezi sebou závislá a pravděpodobnost budoucího jevu se dá odvodit z předcházejícího jevu. Aby tato metoda ale správně fungovala, musí se nejprve učit na již předem klasifikovaných spamových a legitimních zprávách. Postupem času si vytvoří vlastní seznam slov a frází, které se objevují v normálních zprávách a stejný seznam pro spamové zprávy. V testovaných zprávách počítá četnost těchto frází a slov. Na základě nich pak vypočítává pravděpodobnost, že se jedná o spam. Výhoda této metody je vysoká přesnost a nízká šance *false positive*. Další výhoda je, že čím déle se tato metoda používá, tím přesnější a efektivnější by měla být. Nevýhodou metody je právě poměrně dlouhá doba učení na začátku používání a také vysoké náročnosti na systémové zdroje.

3.2 Příklady e-mailových filtrů

Níže jsou uvedeny dva příklady e-mailových filtrů, které se běžně používají. Jedná se o SpamAssassin a Rspamd. Oba tyto filtry jsou open-source, takže si je kdokoliv může volně vyzkoušet a implementovat. Existují i komerční řešení založené na těchto filtrech, kde není nutná žádná další konfigurace od zákazníka.

3.2.1 SpamAssassin

SpamAssassin je multiplatformní open-source program určený pro filtrování e-mailových zpráv. Existuje už přes 20 let a je vyvíjen Apache Software Foundation. Nejnovější verze je 3.4.6, která vyšla v dubnu roku 2021. Zároveň společnost pravidelně aktualizuje filtrovací pravidla. Jeho výhodou je vysoká flexibilita a jednoduchost uživatelského nastavení. Celý program je napsán v jazyce Perl [32].

SpamAssassin používá mnoho metod pro filtrování spamu, mezi které patří i tyto:

- kontrola blacklistů a whitelistů,
- Bayesovo filtrování,
- testování reputace odesílatele,
- testování autentizace.

Pokud uživatel potřebuje nějaké další funkce, může si vybrat z mnoha přídavných modulů. Nevýhoda SpamAssassinu je jeho poměrně vysoká náročnost na systémové zdroje a zpracování velkého množství zpráv může trvat delší dobu.

SpamAssassin je velice populární a využívá ho například Seznam.cz [33] nebo Masarykova univerzita. [34]

3.2.2 Rspamd

Rspamd je také open-source e-mailový filtr. Nejnovější verze je 3.0, která vyšla v srpnu roku 2021. Rspamd je na rozdíl od SpamAssassinu napsaný v jazyce C, a navíc podporuje přídavné moduly napsané v jazyce Lua. Používané metody a úspěšnost filtrování je u obou filtrů podobná. Rspamd dále podporuje greylisting, ale hlavně slibuje až 10x rychlejší zpracování e-mailů a menší náročnost na systémové zdroje [35][36]. Rspamd také disponuje mnohými funkcemi, které jsou již zabudované a není nutné, jako u SpamAssassinu, řešit moduly.

Oproti konkurenci je avšak Rspamd složitější na nastavení. Jelikož většina programu je napsána v jazyce C, je zde také otázka menší bezpečnosti [37][38].

4 Geolokace IP adresy

Geolokace je proces zjištění geografické polohy objektu. Technika známá jako IP geolokace se konkrétně zabývá geografickou lokalizací síťového zařízení s přiřazenou IP adresou. Tímto zařízením může být například mobilní telefon, notebook, stolní počítač nebo server. Dále se dá IP geolokace rozdělit na aktivní a pasivní geolokační metody, které budou blíže popsány v sekci 4.1 a 4.2. [39] [40]

Využití

Fyzická poloha zařízení se v dnešní době využívá v mnoha odvětvích a službách. S masivním rozšířením zařízení připojených k internetu a jeho uživatelů, webových stránek a aplikací na každou službu, se s geolokací v praxi potkává člověk denně.

Zde je uvedeno několik příkladů využití geolokace:

- Automatické zvolení jazyka při vstupu na internetovou stránku.
- Vyhledávač ukazuje více relevantní výsledky na základě polohy zařízení.
- Určení polohy nejbližší prodejny nebo výdejního místa při nakupování přes e-shop.
- Aplikace pro zjišťování jízdních řádů, která uživateli automaticky vyplní nejbližší zastávku.
- Streamovací služby jako Netflix upravující svůj nabízený katalog podle země uživatele.
- Ukládání záznamů geolokace přihlášení uživatele do služby, na základě které lze zjistit, že byl účet kompromitován.
- Zjištění polohy pachatele konající nelegální činnost přes Internet.
- Povolení přístupu na stránky pouze ze zvolené lokality - jako příklad lze uvést stránky, které jsou přístupné pouze z Číny nebo ukazují rozdílné obsahy pro uživatele přistupující z Číny a jiných zemí světa.
- Nabízení reklam uživateli na základě jeho fyzické polohy, jelikož jsou pro něho relevantnější.
- Automatická volba měny a dopravy při nakupování v mezinárodním e-shopu.
- Aplikace ukazující předpověď počasí na základě geolokace.
- Stránky ukazující správně čas události na základě časového pásma, ve kterém se uživatel nachází.
- Zobrazení uživateli aktuální zprávy a dění z jeho regionu.
- Automatické přesměrování na geograficky nejbližší server pro efektivní a nejrychlejší komunikaci uživatele se službou.

Tyto funkce geolokace nejsou relevantní, pokud uživatel využívá metodu, která maskuje jeho reálnou polohu. Mezi tyto metody patří VPN (*Virtual Private Network*) a Tor (*The Onion Router*), které tunelují provoz uživatele skrze VPN server

či Tor uzel, což má za výsledek změnu veřejné IP adresy uživatele. Toto způsobí, že aplikace a navštívené stránky si budou myslet, že uživatel se nachází právě v bodu VPN serveru nebo Tor uzlu a služby poskytované geolokací nebudou relevantní.

Tato funkce může být pro uživatele i výhodná, jelikož může být využita pro zpřístupnění obsahu, který nemá být dostupný pro zemi, kde se uživatel reálně nachází. Příkladem může být již zmiňovaný přístup k obsahu streamovacích služeb, které mají pro různé země odlišnou nabídkou, nebo přístup na stránky přístupné pouze z určité lokality.

4.1 Aktivní metody geolokace

Aktivní metody geolokace určují polohu cílové stanice na základě měření parametrů síťového přenosu mezi referenčními body neboli *landmarks* a zařízením, u kterého se snažíme získat geografickou polohu. Referenční bod je stanice, u které známe přesnou polohu a můžeme tuto informaci použít pro zjištění polohy hledaného stroje.

Nejčastěji používaný a měřený parametr je zpoždění neboli latence, což je čas mezi odesláním dat ze zdroje a jejich přijetím v cíli. Zpoždění vzniká na přenosových linkách, mezilehlých uzlech a koncových zařízeních. Dále může být latence ovlivněna výkonností a zatížením mezilehlých uzlů, omezenou rychlostí a zatížením přenosového média, a v neposlední řadě vzdáleností, jež musí data urazit [41].

K měření zpoždění se nejvíce používají nástroje *ping* a *traceroute*. Oba tyto prostředky používají protokol ICMP (*Internet Control Message Protocol*). *Ping* zašle dotaz na cílovou stanici a čeká na odpověď, čímž získá dobu přenosu ke stanici a zpět. Tomuto se říká obousměrné zpoždění (RTT - *Round Trip Time*). Nástroj *traceroute* zjišťuje uzly směrem k cílové stanici a zpoždění přenosu mezi těmito uzly [42].

Níže jsou stručně popsány vybrané aktivní metody pro zjišťování geolokace:

- Geoping,
- Shortest Ping,
- Constraint Based Geolocation,
- Topology Based Geolocation,
- Speed of Internet,
- Octant.

4.1.1 GeoPing

Metoda GeoPing byla představena v roce 2001 a jedná se o nejstarší metodu založenou na měření zpoždění. Funguje měřením latence k cíli z několika referenčních bodů a tímto odhaduje souřadnice cílové stanice. Z těchto dat se sestaví vektor zpoždění, který určuje vzdálenost od jednotlivých referenčních bodů. Poloha cílové

stanice je odhadnuta za použití Euklidovské vzdálenosti referenčních bodů. Vektory zpoždění jsou porovnány a jako výsledek je vybrán ten nejpodobnější. Nejedná se o nejpřesnější metodu, jelikož výstupem metody GeoPing jsou souřadnice nejbližšího referenčního bodu k cílové stanici. Přesnost se dá zvýšit počtem použitých referenčních bodů.

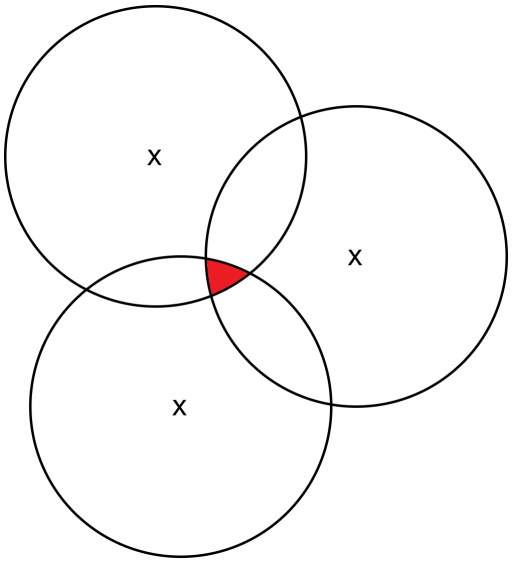
4.1.2 Shortest Ping

Shortest Ping je nejjednodušší aktivní metoda geolokace na základě měření zpoždění. Princip je založen na změření zpoždění přenosu od všech referenčních bodů k cílové stanici, kde výsledek metody je landmark s nejmenší hodnotou RTT. Stejně jako u metody GeoPing jsou výsledkem souřadnice referenčního bodu. Na základě použitých referenčních bodů se může jednat o velice nepřesný odhad geolokace cílové stanice. Přestože jde o velice jednoduchou metodu, v mnoha případech dosahuje stejných nebo i lepších výsledků než některé složitější metody, jako je například GeoPing.

4.1.3 Constraint Based Geolocation

Metoda Constraint Based Geolocation (CBG) je založena na principu nazvaném multilaterace, což je proces odhadování pozice cíle na základě vzdálenosti od dostatečného počtu známých bodů. Jako známé body se opět používají referenční body a od nich je měřeno zpoždění k cílové stanici. Z těchto hodnot se určí vzdálenosti, ze kterých se dostane oblast kolem referenčního bodu, ve kterém se může cílová stanice nacházet. Průsečíkem těchto oblastí kolem referenčních bodů je zóna, kde leží hledaný stroj [43].

Na obrázku 4.1 je znázornění fungování metody Constraint Based Geolocation, kde x jsou referenční body, kruhy kolem referenčního bodu je oblast, kde se může cílová stanice nacházet, a nakonec červeně zvýrazněný průsečík, kde leží hledaná cílová stanice.



Obr. 4.1: Princip metody CBG

4.1.4 Topology Based Geolocation

Jde o rozšíření předchozí metody CBG, kde se navíc používají informace o topologii sítě a směřování v síti. V metodě se používají i mezilehlé uzly, které lze zjistit pomocí nástroje *traceroute*. Tyto uzly pak napomáhají k přesnějšímu určení polohy cílové stanice.

4.1.5 Speed of Internet

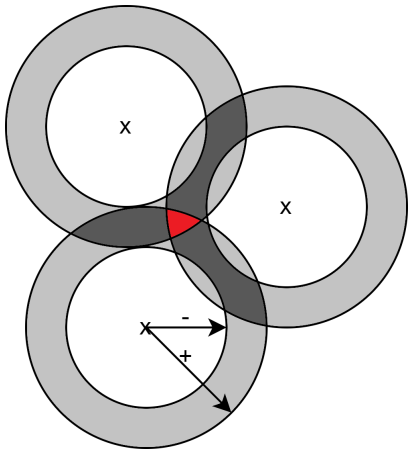
Metoda Speed of Internet (SOI) je založena na podobném principu jako metoda Constraint Based Geolocation. Při určování polohy se používají referenční body, od nichž se měří zpoždění k cílové stanici. Hranice kolem referenčních bodů je vypočítána za použití konstanty 4/9 rychlosti světla, která byla vypočítána za pomocí kombinace rychlosti, se kterou jsou data přenášena v optických kabelech, což je 2/3 rychlosti světla, a zpoždění, ke kterým dochází během přenosu dat. Oproti CBG je metoda rychlejší a méně výpočetně náročná, ale je méně přesná a je zde i šance, že se nepodaří určit pozici cíle.

4.1.6 Octant

Další rozšíření metody Constraint Based Geolocation je metoda Octant, která navíc udává oblast, kde se cílová stanice nemůže nacházet. Oblast, kde se stanice může nacházet, je označená jako pozitivní vzdálenost a oblast, kde se nemůže nacházet, se

nazývá negativní vzdálenost. Těmito 2 oblastmi vznikne mezikruží kolem referenčního bodu, kde se může cílová stanice vyskytovat. Průnikem mezikruží referenčních bodů může vzniknout nekonvexní oblast, která je dále popsána Beziérovými křivkami.

Na obrázku 4.2 lze vidět princip metody Octan, kde x jsou referenční body. Šipka se znakem „-“ značí negativní vzdálenost, šipka se znakem „+“ značí pozitivní oblast. Červeně je pak vyznačena oblast, kde se nachází cílová stanice.



Obr. 4.2: Metoda Octan

4.2 Pasivní metody geolokace

Pasivní geolokace je založena na statických databázích s informacemi o síťovém zařízení. Tyto databáze jsou buď veřejné nebo komerční. Vyhledání těchto informací je mnohem rychlejší a jednodušší než u aktivních metod geolokace, jelikož není nutné nic měřit či počítat. Nevýhoda je v nutnosti periodické aktualizace databáze, jelikož se v ní jinak mohou nacházet zastaralé a nepravdivé informace.

Pasivní metody geolokace se dále dají dělit podle toho, jakých údajů využívají:

- geolokace podle DNS,
- geolokace podle Wi-Fi,
- geolokace podle IP adresy.

4.2.1 Geolokace podle DNS

Pozice cílové stanice jde zjistit podle reverzních DNS (*Domain Name System*) záznamů. Reverzní překlad IP adresy nám dá doménové jméno, z něhož lze někdy

zjistit přibližnou polohu cílové stanice. V doménovém jméně se může nacházet například název města, státu nebo ISP (*Internet Service Provider*). Jelikož je struktura DNS hierarchická, jde většinou alespoň zjistit zemi cílové stanice podle TLD.

Na obrázku 4.3 lze vidět DNS záznam pro IP adresu 147.229.26.152, která má doménové jméno kam-c-ext-zavora.kamery.fce.vutbr.cz. Z tohoto jména lze zjistit, že TLD je *cz*, takže se pravděpodobně jedná o stroj na území České republiky. Dále je zde *vutbr*, takže se jedná o zařízení, které má spojitost s Vysokým učení technickém v Brně. Potom je v doménovém jméně *fce*, tedy je tu spojitost s Fakultou stavební VUT. Podle názvu lze odhadnout, že se jedná o kameru sledující závoru u Fakulty stavební na Vysokém učení technickém v Brně, tedy se dá odhadnout poloha s poměrně velkou přesností. Mnohdy ale IP adresy nemají přiřazené doménové jméno a v těchto případech je tato metoda nepoužitelná.

Experimentální rozšíření DNS LOC (zkráceno *location*) umožňuje zapsání geografické informace přímo do DNS záznamu. Tento záznam obsahuje zeměpisnou šířku, délku, nadmořskou výšku, fyzickou velikost a přesnost. Jelikož je tato funkce experimentální, je zřídka využívaná. Další nevýhodu tvoří fakt, že tyto informace nejsou nijak ověřované [44].

```
root@mail:~# nslookup 147.229.26.152
152.26.229.147.in-addr.arpa     name = kam-c-ext-zavora.kamery.fce.vutbr.cz.
```

Obr. 4.3: DNS dotaz

4.2.2 Geolokace podle Wi-Fi

Polohu cílové stanice lze zjistit i na základě okolích přístupových bodů a síly jejich signálů. Tuto metodu používá například Google ve svých Google Maps [45]. Jsou shromážděna data o okolních přístupových bodech a ta jsou následně porovnána s databází Google Maps. Jedná se o sílu signálu, SSID (*Service Set Identifier* - název bezdrátové sítě) a MAC (*Media Access Control* - jednoznačný identifikátor síťového zařízení) adresu přístupového bodu. Na základě toho je pak Google schopen určit, kde se stanice nachází. Tato databáze je neustále aktualizována, jelikož Google tyto informace o přístupových bodech shromažďuje kontinuálně pomocí chytrých telefonů svých zákazníků.

4.2.3 Geolokace podle IP adresy

Geolokace IP adresy se zjišťuje na základě vyhledání v geolokačních databázích, které můžou být volně dostupné, či komerční, nebo v databázi pro přidělování IP

adres organizací IANA (*Internet Assigned Numbers Authority*).

WHOIS je nejznámější veřejnou databází obsahující informace k IP adrese. Tato databáze je udržovaná odpovědným RIR (*Regional Internet Registry*) pro daný region. Tyto organizace spravují alokaci a registraci IP adres a AS (*Autonomous System*). RIR [46] spadají pod organizaci IANA. Je jich celkem 5, a to:

- AFRINIC (*African Network Information Center*) spravující oblast Afriky;
- ARIN (*American Registry for Internet Numbers*) spravující Antarktidu, Kanadu, USA a části Karibiku;
- APNIC (*Asia-Pacific Network Information Centre*) spravující Oceánii a východní, jižní a jihovýchodní Asii;
- LACNIC (*Latin America and Caribbean Network Information Centre*) spravující většinu Karibiku a Latinskou Ameriku;
- RIPE NCC (*Réseaux IP Européens Network Coordination Centre*) spravující západní a Střední Asii, Evropu a Rusko.

Informace z této databáze lze získat pomocí nástroje *whois* nebo na internetu existuje mnoho stránek poskytující vyhledávání dat v této databázi. Jako příklad je zde uvedena webová stránka *whois.domaintools.com*. Výsledek hledání IP adresy 147.229.26.152 [47] je vidět na obrázku 4.4. Z výstupu lze mimo jiné zjistit kontaktní e-mailovou adresu, v jakém IP rozsahu se adresa nachází, zemi, kdy byl záznam vytvořen a naposledy modifikován, komu IP adresa patří, telefonní číslo nebo město a směrovací číslo.

IP Location	Czech Republic Brno Brno University Of Technology
ASN	AS197451 VUTBR-AS, CZ (registered Dec 08, 2010)
Resolve Host	kam-c-ext-zavora.kamery.fce.vutbr.cz
Whois Server	whois.ripe.net
IP Address	147.229.26.152

```
% Abuse contact for '147.229.0.0 - 147.229.254.255' is ' abuse@vutbr.cz '

inetnum:        147.229.0.0 - 147.229.254.255
netname:        VUTBRNET
descr:          Brno University of Technology
country:        CZ
admin-c:        CA6319-RIPE
tech-c:         CA6319-RIPE
status:         ASSIGNED PA
mnt-by:         VUTBR-MNT
created:        2014-11-19T08:23:45Z
last-modified:  2015-01-30T08:37:07Z
source:         RIPE

role:           Brno University of Technology - Backbone Admins
address:        Brno University of Technology
address:        Antoninska 1
address:        601 90 Brno
address:        The Czech Republic
phone:          +420 541145453
phone:          +420 723047787
e-mail:          admin@cis.vutbr.cz
nic-hdl:        CA6319-RIPE
mnt-by:         VUT-BATCH-MNT
mnt-by:         VUTBR-MNT
created:        2015-01-30T08:31:35Z
last-modified:  2016-11-04T14:01:52Z
source:         RIPE
abuse-mailbox:   abuse@vutbr.cz

route:          147.229.0.0/17
descr:          VUTBR-NET1
origin:         AS197451
mnt-by:         VUTBR-MNT
created:        2014-12-04T19:07:00Z
last-modified:  2014-12-04T19:07:00Z
source:         RIPE
```

Obr. 4.4: Výsledek WHOIS dotazu na webové stránce *whois.domaintools.com*

4.3 Geolokační databáze

V této sekci bude stručně popsáno několik geolokačních databází. Ačkoliv tyto databáze sdílejí některé informace o jejich fungování a zjišťování geografické polohy cílových stanic, tak jsou tyto údaje záměrně vágní a nespecifické. Rozdíly v geo-

lokačních databázích spočívá v množství poskytovaných dat, jejich přesnosti a jak často jsou tato data aktualizována [48].

4.3.1 IP2Location

Společnost IP2Location nabízí 25 různých databází, označených DB 1 až DB 25, kde se zvyšujícím se číslem se zvyšuje i počet a kombinace poskytovaných dat. U DB 25 je možné u IP adresy zjistit tyto informace:

- země,
- region,
- město,
- zeměpisná šířka,
- zeměpisná délka,
- poštovní číslo,
- poskytovatel internetových služeb,
- doména,
- časové pásmo,
- rychlost připojení,
- telefonní předvolba,
- název nejbližší meteorologické stanice,
- kód nejbližší meteorologické stanice,
- MCC (*Mobile Country Code*),
- MNC (*Mobile Network Code*),
- mobilní operátor,
- nadmořská výška,
- typ použití,
- typ adresy,
- kategorie.

Společnost nabízí i bezplatnou verzi LITE [49], která má menší počet záznamů, menší přesnost a postrádá uživatelskou podporu. Podle statistik společnosti IP2Location LITE verze má přesnost 98 % na zemi IP adresy, 60 % na město a seskupuje adresy do bloku podle prefixu IPv4 /24. Placená verze má přesnost státu 99,5 %, 80 % města a IP adresy nejsou seskupovány do větších celků [50].

4.3.2 GeoIP2

GeoIP2 od společnosti MaxMind má volně dostupnou a komerční databázi. Volně dostupná databáze má název GeoLite2 a oproti placené verzi s názvem GeoIP2 je méně přesná. Zároveň u ní společnost neposkytuje žádnou uživatelskou podporu.

Pokud se uživatel rozhodne platit za komerční verzi, má na výběr z vícero databází, které se od sebe liší množstvím poskytnutých dat a cenou. Tyto databáze mají název GeoIP2 Country, City, Anonymous IP, ISP, Domain a Connection Type. Nejpoužívanější z těchto je GeoIP2 City, která obsahuje informace o kontinentu, zemi, regionu, městu, poštovním čísle a hrubé zeměpisné šířce a délce. Tato databáze je aktualizovaná 2x týdně [51].

4.3.3 HostIP

HostIP je komunitní projekt, kde uživatelé sami vkládají a upravují informaci o své IP adrese. Na vybudování databáze byla použita dále nespecifikovaná data z internetu. Aktuálně je v databázi přes 9,2 milionů záznamů. Tím, že uživatelé sami vkládají geolokační údaje, je možné, že řada těchto záznamů je nepřesná a zastaralá. O hledaných IP adresách poskytuje pouze informaci o zemi a městu [52].

Na obrázku 4.5 je ukázka z použití databáze na stránkách HostIP.

Obr. 4.5: Vyhledání adresy na stránkách HostIP

4.3.4 DB-IP

DB-IP nabízí placenou a bezplatnou verzi. V bezplatné verzi jsou základní informace, což je stát, město a číslo autonomního systému. V placené verzi jsou pak i další data, jako je např. měna státu, souřadnice, poskytovatel internetového připojení apod. Databáze obsahuje 32 milionů záznamů IPv4 a IPv6 bloků adres. Stejně jako u ostatních konkurenčních databází, bezplatná verze má menší přesnost a obsahuje méně záznamů [53].

5 Vývoj vlastního modulu *Geolock*

Jako platforma, pro kterou byl vyvíjen modul pro blokaci na základě geolokace IP adresy, byl zvolen po domluvě s vedoucím práce SpamAssassin, jako jedno z nejpopulárnějších a nejpoužívanějších antispamových řešení. Jako vývojové a testovací prostředí byl vybrán virtuální stroj s operačním systémem Debian 11 Bullseye [54][55]. Zde byl nainstalován SpamAssassin, konkrétně jeho nejnovější verze 3.4.6, která vyšla 12. dubna 2021. SpamAssassin je celý napsán v programovacím jazyku Perl a samotné moduly musí být také vytvořené v tomto jazyce.

Jako geolokační databáze byla po domluvě s vedoucím práce vybrána databáze od společnosti IP2Location, konkrétně IP2Location LITE DB 5, která obsahuje údaje o zemi, městu a souřadnicích cílové stanice. IP2Location má svůj vlastní CPAN (*Comprehensive Perl Archive Network*) modul [56], který lze použít pro práci s databázemi od této společnosti. Dalším faktorem, proč byla zvolena zrovna tato databáze, je její vysoká přesnost a menší chybovost oproti konkurenci. [57]

5.1 Zakomponování modulu do SpamAssassinu

SpamAssassin má nejčastěji konfigurační soubory ve složce */etc/mail/spamassassin/* nebo */etc/spamassassin/*. Pro načtení a použití modulu je nutné upravit konfigurační soubor *init.pre*, který SpamAssassinu říká, jaké moduly a v jakém pořadí je má načíst. Je tady nutné přidat následující řádek, který říka SpamAssassinu, aby načetl vytvořený modul *Geolock*.

Výpis 5.1: Načtení modulu v *init.pre*

```
loadplugin Mail::SpamAssassin::Plugin::Geolock Geolock.pm
```

Následně je nutné přesunout soubor *Geolock.pm*, což je samotný modul, a soubor *Geolock.cf*, který je konfigurační soubor pro modul, do složky SpamAssassinu.

5.2 Konfigurační soubor

Konfigurační soubor modulu *Geolock.cf* vypadá následovně:

Výpis 5.2: Konfigurační soubor *Geolock.cf*

```
header   BLOCKED_COUNTRY eval:get_country()
score    BLOCKED_COUNTRY 50.0
describe BLOCKED_COUNTRY The country of origin is blocked

rule 0 GB,US,SK

add_header all Country The country of origin is _MYTAG_
```

První řádek stanovuje, co se má provést. Zde *header* říká, že se bude modul zabývat hlavičkou e-mailové zprávy. Následuje *BLOCKED_COUNTRY*, což je název pravidla, a jako poslední je *eval:get_country()*. *Eval* znamená *evaluate*, tedy zhodnocení, jestli funkce z modulu *get_country()* vrací **1** jako **true**, nebo **0** jako **false**.

V případě, že vrátí funkce **1**, podle druhého řádku se zpráva ohodnotí skórem 50 bodů. Výchozí a oficiálně doporučená hodnota pro označení zprávy SpamAssassinem jako spam je 5 bodů. V případě, že správci podrobují e-maily velkému počtu testů, tak může být tato hranice navýšena například na 7 bodů. Ohodnocením zprávy 50 body by tedy mělo zaručit zablokování zprávy i v případě navýšené hranice. Pokud by správce chtěl tento modul využít například jenom jako doporučení, může si tuto hodnotu jednoduše změnit.

Třetí řádek popisuje, co se v případě zablokovaného zprávy zobrazí za popisek u názvu spuštěného testu modulu. Seznam všech spuštěných testů SpamAssassin přidává do zpracované hlavičky e-mailu.

Ve čtvrtém řádku si uživatel nastavuje, které země chce blokovat. V ukázkovém případě jsou zvoleny pro blokaci e-maily od odesílatelů ve Velké Británii, Spojených státech amerických a Slovenské republice. Zde se dá nastavit libovolný počet blokovaných zemí. Řádek musí začínat klíčovým slovem „rule“. Následuje mezer a číslo **0** nebo **1**. Toto číslo nastavuje, která IP adresa se bude brát jako adresa odesílatele pro vyhodnocení testu. Při zvolené hodnotě **0** se bere jako odesílatel poslední veřejná IP adresa v polích *Received:* v hlavičce e-mailu. V případě, že by správce měl podezření na podvrhnutí hlaviček, může si zde zvolit hodnotu **1**. Při této hodnotě se jako IP adresa odesílatele bere první veřejná IP adresa v polích *Received:*, která nepochází z *trusted_networks*, tedy sítí, kterým správce důvěřuje. Tyto sítě se dají nastavit v konfiguračních souborech SpamAssassinu a umožňují správcům zvolit sítě, kterým plně věří. U e-mailů pocházející z těchto sítí se plně věří hlavičkám e-mailu a provádí se méně testů na spam.

Poslední řádek nastavuje, že se každé zprávě přidá do hlavičky informace o proběhlém testu. Co přesně se přidává do hlavičky je dále popsáno v sekci se zjedno-

dušeným vývojovým diagramem 5.3.1.

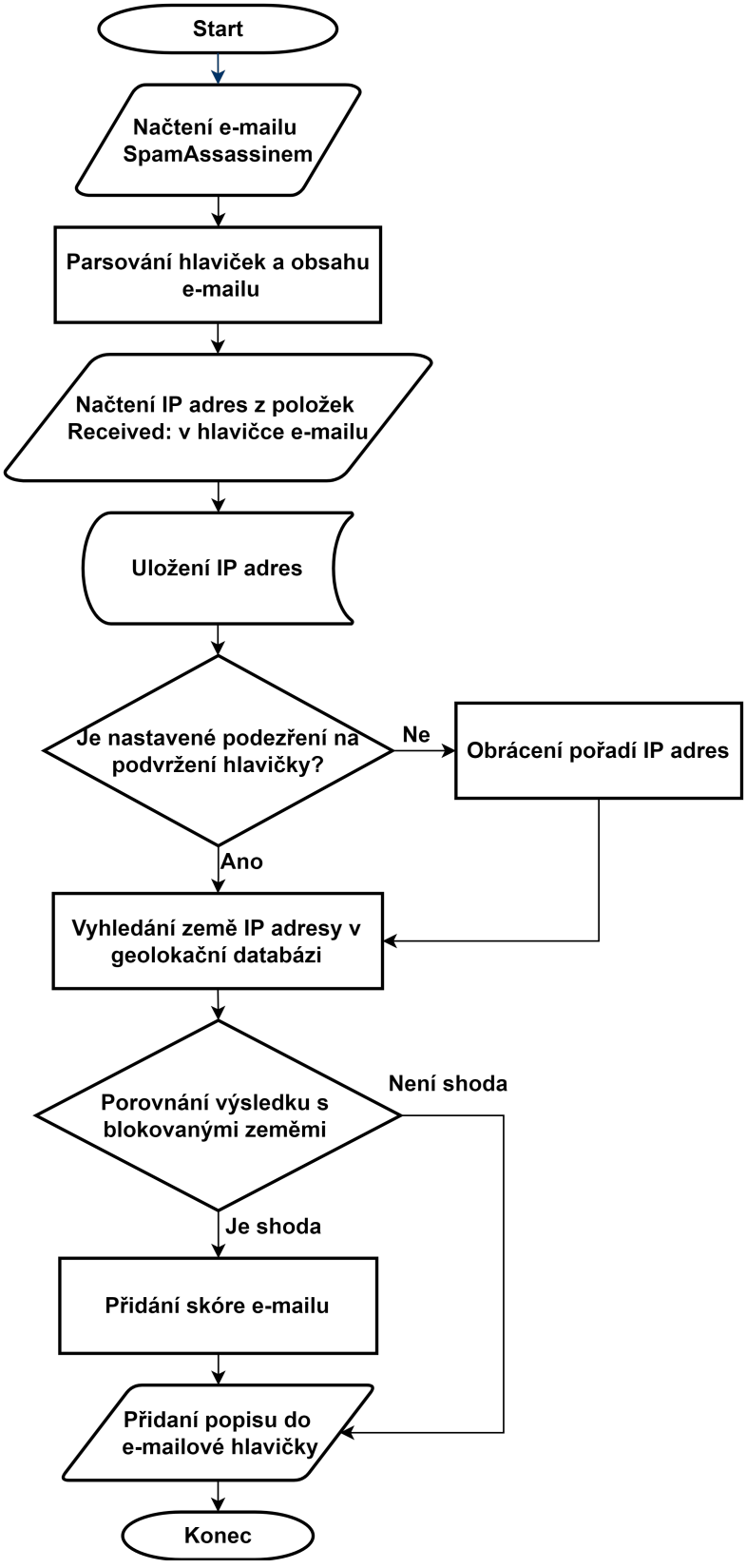
5.3 Popis modulu

V této sekci je popsán zjednodušený vývojový diagram vlastního modulu *Geolock*. Jako druhá část je ukázka funkce pro zpracování konfiguračního souboru a hlavní funkce odpovědné za získání IP adresy odesílatele z hlavičky e-mailu, zjištění země odesílatele pomocí geolokační databáze IP2Location a následné porovnání, zda se jedná o blokovanou zemi.

5.3.1 Vývojový diagram modulu

Na obrázku 5.1 je zjednodušený diagram fungování modulu. Postup fungování při testování, zda je zpráva spam, je následovný:

1. SpamAssassin načte zprávu a rozparsuje ji.
2. Modul pak přistupuje k IP adresám uvedeným v hlavičce v položce *Received:*.
3. Do pole si modul uloží IP adresy u poštovních serverů, kterým SpamAssassin nevěří.
4. Modul načte nastavení z konfiguračního souboru, zda správce má podezření na podvrhnutí e-mailových hlaviček.
   (a) Pokud správce nemá podezření na podvrhnutí hlaviček zprávy, tak modul pole obrátí, aby IP adresy byly po sobě chronologicky tak, jak zpráva procházela poštovními servery.
   (b) Pokud správce má podezření na podvrhnutí hlaviček zprávy, tak se s polem nic neděje.
5. Z pole si vezme první prvek a vyhledá v IP2Location databázi stát, ve kterém se IP adresa. nachází. IP2Location databáze musí být stažena na stroji s modulem. Databáze má zkratky zemí uložené podle normy ISO 3166-1 alpha-2, tedy dvoumístný kód pro zemi [58]. Přesný seznam podporovaných zkratek se dá najít na stránkách IP2Location [59]. V *Geolock.cf* je pro jednoduší použití vypsán seznam všech zemí, které databáze podporuje s plným a částečným pokrytím.
6. Tento výsledek porovná s blokovanými zeměmi.
   (a) Pokud se země původu e-mailu nachází v seznamu blokovaných zemí, je zpráva ohodnocena definovaným skórem 50 bodů.
   (b) Pokud je země původu e-mailu jiná než blokované země, není zprávě přidáno žádné skóre.
7. Do hlavičky je přidána informace, jestli se jedná o zprávu ze zablokované země, jaká je IP adresa odesílatele a z jaké země odesílatel je.



Obr. 5.1: Zjednodušený vývojový diagram modulu *Geolock*

5.3.2 Geolokace zdroje zprávy

Ve výpisu 5.3 je ukázaná funkce *parse_config* ze souboru *Geolock.pm*, která zpracovává konfiguraci ze souboru *Geolock.cf*. Funkce najde v konfiguračním souboru řádek začínající slovem „rule“ a pokud má správnou formu, uloží si blokované země a nastavení podezření na podvrhnutí hlaviček e-mailu.

Výpis 5.3: Funkce parse_config

```
sub parse_config {
    my ($self, $opts) = @_;
    my $key = $opts->{key};

    my $rule = "";
    my $spoof = "";

  # find the line starting with "rule" and
  # save blocked countries and Received header switch
    if ($key eq "rule") {
      if($opts->{value} =~ /(^(0|1)\s+(([A-Z]+(,[A-Z]+)+)|[A-Z]{2}))/){
          $spoof = $2;
          $rule = $3;
          $opts->{conf}->{"rule"}{rule}= $rule;
          $opts->{conf}->{"rule"}{spoof}= $spoof;
          $self->inhibit_further_callbacks();
      }else{
          dbg("Geolock: The rule is empty or not written correctly.");
          }
   }
   return 0;
}
```

Níže je uvedena hlavní funkce odpovědná za funkci modulu *Geolock*. Jednotlivé části jsou okomentovány v kódu. V této funkci *get_country* dochází k načítání IP adres do pole, zjištění země původu zprávy a následné vyhodnocení, zda se jedná o blokovanou zemi, nebo ne.

Tato část má na starosti načtení IP adres z *Received:* polí e-mailové hlavičky a uložení těchto adres do pole. V případě, že by se IP adresa objevila opakovaně, tak je ignorována. Tato funkcionalita byla přidána po objevení chyby v SpamAssassinu, kdy při jistých podmínkách detekoval IP adresu z jiného pole než *Received:* a v důsledku toho se zjišťovala geolokace jiné IP adresy než odesílatele.

Výpis 5.4: Hlavní funkce get_country

```
sub get_country{
  my ($self, $pms) = @_;
    my $msg = $pms->{msg};

    my @countries;
    my $ip;
  # load IP addresses from untrusted relays along the e-mail's path and save them
      into an array. If an IP address apears for a second time, it is ignored. This
       is because in some rare cases SpamAssassin incorrectly detects IP address
      not in a Received: field and this eliminates the error.
```

```
8
9       foreach my $relay (@{$msg->{metadata}->{relays_untrusted}}) {
10        if ( grep( /$relay->{ip}/, @countries ) ) {
11          dbg("Geolock: IP is already in array, ignoring it.");
12          }else{
13          push(@countries, $relay->{ip});
14          }
15    }
```

V další sekci se načítá konfigurace podezření na podvržení hlavičky a na základě tohoto nastavení se rozhodne, ze kterého konce pole se bude IP adresa brát. Pokud je pole prázdné, znamená to, že buď šla zpráva pouze ve vnitřní síti, nebo její zdroj je síť, které správce důvěřuje. Podle toho se i vypíše informace do zpracované hlavičky zprávy.

```
16      my @reverseC = reverse(@countries);
17    # load the spoofed Received header switch
18    my $spoof = $pms->{conf}->{"rule"}{spoof};
19    if(!defined($spoof)){
20      $pms->set_tag("MYTAG","The rule is empty or not written correctly.");
21        return 0;
22    }
23    # decide the direction of Received headers for comparison based on spoofed
          Received header switch
24    if($spoof){
25      $ip = $countries[0];
26    }else{
27      $ip = $reverseC[0];
28      }
29    my $countryshort;
30    my $i = 0;
31
32    # in the case there is no untrusted external IP address in the e-mail header
33    if(!defined($ip)){
34          $pms->set_tag("MYTAG","The e-mail did not go through any external non-
              trusted mail server.");
35          return 0;
36    }
```

Dále funkce kontroluje pomocí regex, zda IP adresa, která se má vyhodnotit, není z privátního nebo rezervovaného rozsahu. V takovém případě se posune na další adresu v uloženém poli. Následuje kontrola, která v případě zvolení dalšího prvku v poli zajišťuje, že se nejedná o prázdný prvek.

```
37    #in the case that the Received header contains a private or reserved IP address,
          move on the next Received header
38    while ($ip =~ /(^0\.)|(^10\.)|(^100\.6[4-9]\.)|(^100\.[7-9]\d\.)|(^100\.1[0-1]\
          d\.)|(^100\.12[0-7]\.)|(^127\.)|(^169\.254\.)|(^172\.1[6-9]\.)
          |(^172\.2[0-9]\.)|(^172\.3[0-1]\.)|(^192\.0\.0\.)|(^192\.0\.2\.)
          |(^192\.88\.99\.)|(^192\.168\.)|(^198\.1[8-9]\.)|(^198\.51\.100\.)
          |(^203.0\.113\.)|(^22[4-9]\.)|(^23[0-9]\.)|(^24[0-9]\.)|(^25[0-5]\.)|(^::1$)
          |(^[fF][cCdD])|(^[fF][eE][89aAbB][0-9a-fA-F]:)/){
39          if($i<scalar @countries){
40              if($spoof){
41                  $ip = $countries[$i];
42              }else{
```

```
43                  $ip = $reverseC[$i];
44              }
45              $i++;
46              }else{
47                  $pms->set_tag("MYTAG","The e-mail did not go through any external
                        non-trusted or non-reserved mail server.");
48                  return 0;
49              }
50      }
```

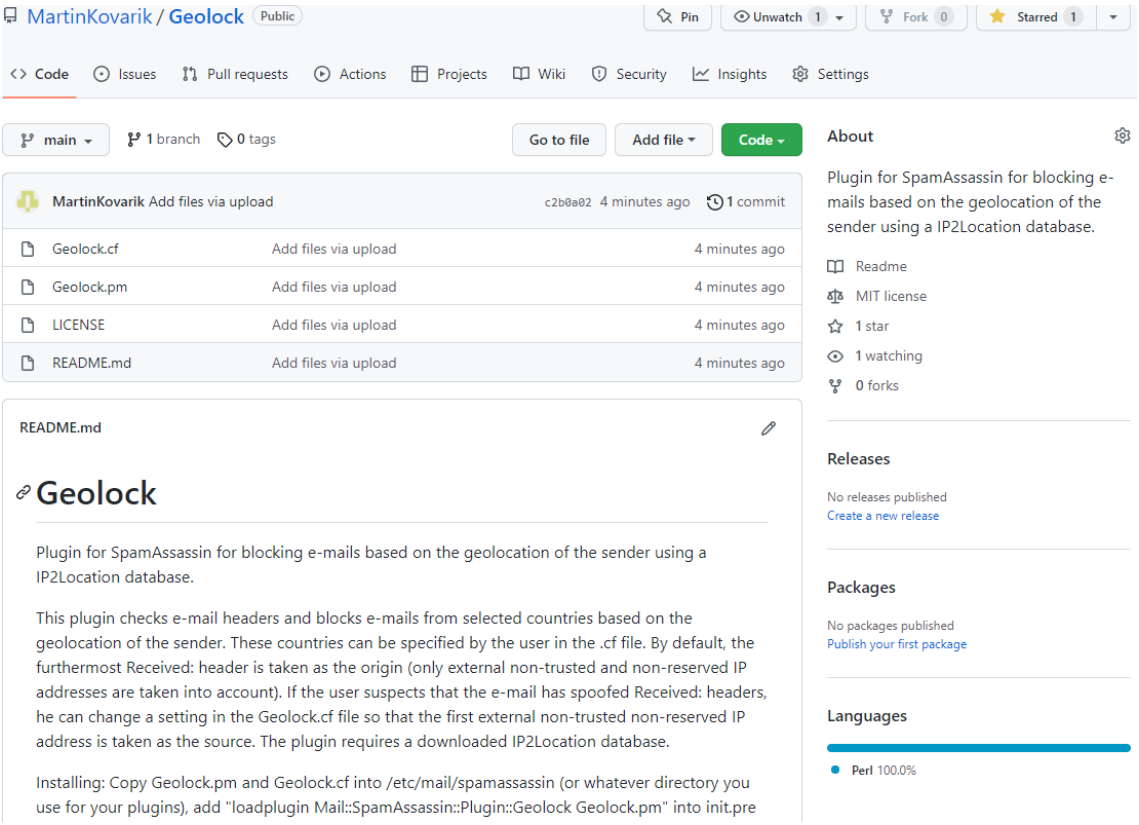
Následně se použije Perl knihovna od IP2Location pro práci s jejich databázemi a otevře se geolokační databáze, která musí být uložená na cestě */etc/mail/spamassassin/DB/ IP2LOCATION.BIN*. Databáze musí být tedy pojmenovaná „IP2LOCATION.BIN“ a může to být jakákoliv databáze od společnosti IP2Location. V této databázi se vyhledá země původu zprávy a porovná se zablokovanými zeměmi z konfiguračního souboru. V případě shody je vrácena **1** a tím je zpráva ohodnocena definovaným počtem bodů, jinak se bodové ohodnocení zprávy nijak neupravuje. V obou případech je přidaná do zpracované hlavičky e-mailu informace o IP adrese odesílatele, ze které země je a zda je blokována.

```
51    # open IP2Location database and look up country of origin
52      require Geo::IP2Location;
53      my $obj = Geo::IP2Location->open("/etc/mail/spamassassin/DB/IP2LOCATION.BIN");
54
55      if (!defined($obj)) {
56          print STDERR Geo::IP2Location::get_last_error_message();
57      }
58      $countryshort = $obj->get_country_short($ip);
59      $obj->close();
60
61      #load blocked countries
62      my $rule = $pms->{conf}->{"rule"}{rule};
63      if (! length $rule){
64          dbg("Geolock: No blocked country found in the configuration file.");
65          $pms->set_tag("MYTAG","No blocked country found in the configuration file."
                );
66          return 0;
67      }
68      my @array;
69      @array = split ' ',$rule;
70
71      # compare the country of origin with blocked countries from configuration file
            and decide if to block the e-mail or not
72      if ( grep( /$countryshort/, @array ) ) {
73          $pms->set_tag("MYTAG","The country of origin is BLOCKED : $ip :
                $countryshort");
74          return 1;
75      }
76      else{
77          $pms->set_tag("MYTAG","The country of origin is NOT BLOCKED : $ip :
                $countryshort");
78          return 0;
79      }
80  }
```

5.4 Zveřejnění modulu

Modul *Geolock* byl zveřejněn na platformě GitHub [60] a je komukoliv volně přístupný. GitHub je internetová služba, která podporuje vývoj softwaru a poskytuje hostování pro soukromé i veřejné repozitáře. Platforma má mnoho funkcí, mezi jehož hlavní patří verzování pomocí Git, systém řízení problémů, dokumentace, diskuzní fóra a podobné. Jedná se o jednu z nejpopulárnějších a největších platforem pro zveřejňování zdrojových kódů a open-source projektů, a proto byl zvolen pro hostování modulu *Geolock*.

Jak vypadá zveřejněný repozitář modulu *Geolock* na platformě GitHub je možné vidět na obrázku 5.2.



Obr. 5.2: GitHub repozitář modulu *Geolock*

Modul je zveřejněn pod svobodnou softwarovou licencí MIT [61] a je tedy možné modul *Geolock* bezplatně použít kýmkoliv pro jakékoliv účely. Libovolný uživatel může tento modul používat, dále šířit či používat v proprietárním softwaru. MIT licence je velice krátká a stručná. Její celé znění je ukázáno ve výpisu 5.5.

Výpis 5.5: Plné znění MIT licence

```
MIT License

Copyright (c) 2022 Martin Kovařík

Permission is hereby granted, free of charge, to any person obtaining a copy
of this software and associated documentation files (the "Software"), to deal
in the Software without restriction, including without limitation the rights
to use, copy, modify, merge, publish, distribute, sublicense, and/or sell
copies of the Software, and to permit persons to whom the Software is
furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all
copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,
FITNESS FOR A~PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE
AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER
LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM,
OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE
SOFTWARE.
```

6 Ověření funkčnosti vytvořeného modulu

V této sekci je nejdříve ukázáno otestování modulu *Geolock* na testovací zprávě. Přidaná vlastní hlavička modulem je zvýrazněná ve finální hlavičce zprávy a správný výsledek je ověřen i jiným způsobem geolokace IP adresy. V druhé části je modul otestován na datasetu, který obsahuje 1200 zpráv spamu a 300 legitimních zpráv neboli hamu. Tyto zprávy pochází z poštovní schránky českého uživatele. Dataset je rozdělen na 2 stejně velké části a je na něm otestován modul *Geolock*.

6.1 Testování modulu na jedné zprávě

Zprávu uloženou v textové podobě v souboru *test.txt* lze otestovat SpamAssassinem následujícím příkazem, kde možnost *-t* značí, že chceme danou zprávu otestovat:

Výpis 6.1: Otestování zprávy SpamAssassinem

```
#spamassassin -t  < test.txt
```

Výsledek tohoto testu lze vidět níže, kde je ukázaná přidaná hlavička od SpamAssassinu oznamující, které všechny testy e-mail ohodnotily. Pro přehlednost je červeně vyznačen test *BLOCKED_COUNTRY*, který je výsledek modulu *Geolock*.

Výpis 6.2: Zkrácená přidaná SpamAssassin hlavička

```
Received: from unknown (HELO rly-xr02.nikavo.net) (75.249.246.124) by
    rly-xw05.oxyeli.com with asmtp; 28 May 0102 16:50:33 +0300
From: <Sheila7316x53@hotmail.com>
To: Preferredfriends@aol.com
Subject: *****SPAM***** Learn How To Make $8,000 within 7-14 days! 3341iYjl5-964K
X-Spam-Checker-Version: SpamAssassin 3.4.6 (2021-04-09) on mail.kovar.com
X-Spam-Flag: YES
X-Spam-Level: **********
X-Spam-Status: Yes, score=64.8 required=5.0 tests=BLOCKED_COUNTRY,
  DATE_IN_FUTURE_06_12,FREEMAIL_ENVFROM_END_DIGIT,FREEMAIL_FROM,
  FREEMAIL_REPLYTO_END_DIGIT,FSL_HELO_FAKE,INVALID_DATE,MISSING_MIMEOLE,
  MSGID_OUTLOOK_INVALID,MSOE_MID_WRONG_CASE,RATWARE_MS_HASH,RDNS_NONE,
  SPF_HELO_NONE,SPF_SOFTFAIL,SPOOFED_FREEMAIL,SPOOFED_FREEMAIL_NO_RDNS,
  TO_NO_BRKTS_MSFT,TRACKER_ID autolearn=no autolearn_force=no
  version=3.4.6
X-Spam-Country: The country of origin is BLOCKED : 75.249.246.124 : US
Content analysis details:   (64.8 points, 5.0 required)
 pts rule name              description
---- ---------------------- --------------------------------------------------
 0.4 INVALID_DATE           Invalid Date: header (not RFC 2822)
 0.0 FSL_HELO_FAKE          No description available.
 0.2 FREEMAIL_REPLYTO_END_DIGIT Reply-To freemail username ends in digit
                            [sheila7316x53[at]hotmail.com]
 3.9 MSGID_OUTLOOK_INVALID  Message-Id is fake (in Outlook Express format)
 0.2 FREEMAIL_ENVFROM_END_DIGIT Envelope-from freemail username ends in digit
                            [sheila7316x53[at]hotmail.com]
 50.0 BLOCKED_COUNTRY        The country of origin is blocked
 0.0 DATE_IN_FUTURE_06_12   Date: is 6 to 12 hours after Received: date
```

```
 0.0 SPF_HELO_NONE           SPF: HELO does not publish an SPF Record
 0.0 FREEMAIL_FROM           Sender email is commonly abused enduser mail provider
                             [sheila7316x53[at]hotmail.com]
 1.0 SPF_SOFTFAIL            SPF: sender does not match SPF record (softfail)
 0.1 TRACKER_ID              BODY: Incorporates a tracking ID number
 1.3 RDNS_NONE               Delivered to internal network by a host with no rDNS
 3.4 MSOE_MID_WRONG_CASE     No description available.
 1.8 MISSING_MIMEOLE         Message has X-MSMail-Priority, but no X-MimeOLE
 1.0 RATWARE_MS_HASH         Bulk email fingerprint (msgid ms hash) found
 0.0 SPOOFED_FREEMAIL_NO_RDNS From SPOOFED_FREEMAIL and no rDNS
 0.0 TO_NO_BRKTS_MSFT        To: lacks brackets and supposed Microsoft tool
 1.4 SPOOFED_FREEMAIL        No description available.
```

Tučně vyznačený řádek začínající s *X-Spam-Country:* je také vložen modulem. Tento řádek obsahuje informaci, zda je země původu zprávy blokovaná, IP adresu odesílatele a země původu. V případě, že zpráva pochází z vnitřní sítě nebo neprošla sítí, které správce nevěří, tak tato informace je zde uvedena místo informace o zemi původu.

V případě podezření na podvrhnutí *Received:* hlaviček je výsledná hlavička pozměněna na první externí IP adresu, které správce nevěří. V ukázkové spam zprávě je pak výsledek vidět v přidané hlavičce ve výpisu 6.3.

Výpis 6.3: Přidaná hlavička modulem *Geolock*

```
X-Spam-Country: The country of origin is NOT BLOCKED : 213.105.180.140 : GB
```

Níže je uvedená e-mailová hlavička spamové zprávy, která se testuje. Modul si z této hlavičky vezme IP adresy ve zvýrazněných položkách *Received:*. Postup je více do detailu popsán dříve v sekci 5.3.1. V tomto ukázkovém příkladu je v modulu nastavená země na zablokování *US* a správce nemá podezření na podvrhnutí cesty zprávy. IP adresa odesílatele je 75.249.246.124, což IP2Location databáze vrátí jako *US*, a tedy je zpráva označena jako spam.

Ve výpisu 6.4 je pak hlavička původní analyzované spam zprávy, včetně celé cesty zprávy přes poštovní servery, kterou lze vidět v polích *Received:*.

Výpis 6.4: Hlavička spam zprávy

```
Return-Path: Sheila7316x53@hotmail.com
Delivery-Date: Tue May 28 21:01:35 2002
Received: from mandark.labs.netnoteinc.com ([213.105.180.140]) by
    dogma.slashnull.org (8.11.6/8.11.6) with ESMTP id g4SK1YO24988 for
    <jm@jmason.org>; Tue, 28 May 2002 21:01:34 +0100
Received: from hotmail.com (kbl-mdb6237.zeelandnet.nl [62.238.24.141]) by
    mandark.labs.netnoteinc.com (8.11.2/8.11.2) with SMTP id g4SK1R731575 for
    <jm@netnoteinc.com>; Tue, 28 May 2002 21:01:28 +0100
Received: from sparc.zubilam.net ([146.172.86.49]) by
    a231242.upc-a.zhello.nl with esmtp; Tue, 28 May 0102 11:05:33 -0300
Received: from unknown (118.186.232.151) by sparc.zubilam.net with NNFMP;
    28 May 0102 08:00:33 +0700
Received: from unknown (75.72.44.176) by rly-xw05.oxyeli.com with asmtp;
    28 May 0102 14:55:33 +0200
Received: from unknown (HELO rly-xr02.nikavo.net) (75.249.246.124) by
```

```
    rly-xw05.oxyeli.com with asmtp; 28 May 0102 16:50:33 +0300
Reply-To: <Sheila7316x53@hotmail.com>
Message-Id: <003b45b76c1e$6862a6d3$3da65ae7@dnvdoi>
From: <Sheila7316x53@hotmail.com>
To: Preferredfriends@aol.com
Subject: Learn How To Make $8,000 within 7-14 days! 3341iYjl5-964K
```

Pro kontrolu lze výsledek geolokace IP adresy 75.249.246.124 z webových stránek IP2Location lze vidět na obrázku 6.1.

Obr. 6.1: Výsledek geolokace na stránkách IP2Location

6.2 Testování na datasetu

V této sekci je modul *Geolock* otestován na datasetu reálných zpráv z českých poštovních schránek. Dataset obsahuje:

- 1200 spam zpráv.
- 300 legitimních zpráv neboli ham.

Tento dataset byl náhodně rozdělen na 2 stejně velké části, tedy 2 skupiny po 600 zprávách spamu a 150 zprávách hamu.

Prvně je v tabulce 6.1 vidět globální statistika zemí a počtu IP adres šířící spam vzata ze stránek Spamhaus [62]. Spamhaus je mezinárodní nezisková organizace, která se zaměřuje na bezpečnost na Internetu a sledování spamu. Spravují 1 z nejpoužívanějších IP blacklistů. Na webových stránkách Spamhaus je pravidelně aktualizovaná tabulka top 10 zemí s nejvíce zdroji spamu. Z těchto dat byla vytvořena tato tabulka, která obsahuje data aktuální k datu 22. 4. 2022. Země s nejvíce zdroji spamu jsou tedy Spojené státy americké s 3393 IP adresami, následované Čínskou lidovou republikou, Ruskou federací, Spojenými státy mexickými, Dominikánskou republikou, Saudskou Arábií, Uruguayskou východní republikou, Indickou republikou, Brazilskou federativní republikou a jako poslední je Japonsko s 351 IP adresami.

Tab. 6.1: Spamhaus globální statistika

Země	**Počet unikátních IP adres**
US	3393
CN	2432
RU	763
MX	638
DO	588
SA	537
UY	438
IN	400
BR	399
JP	351

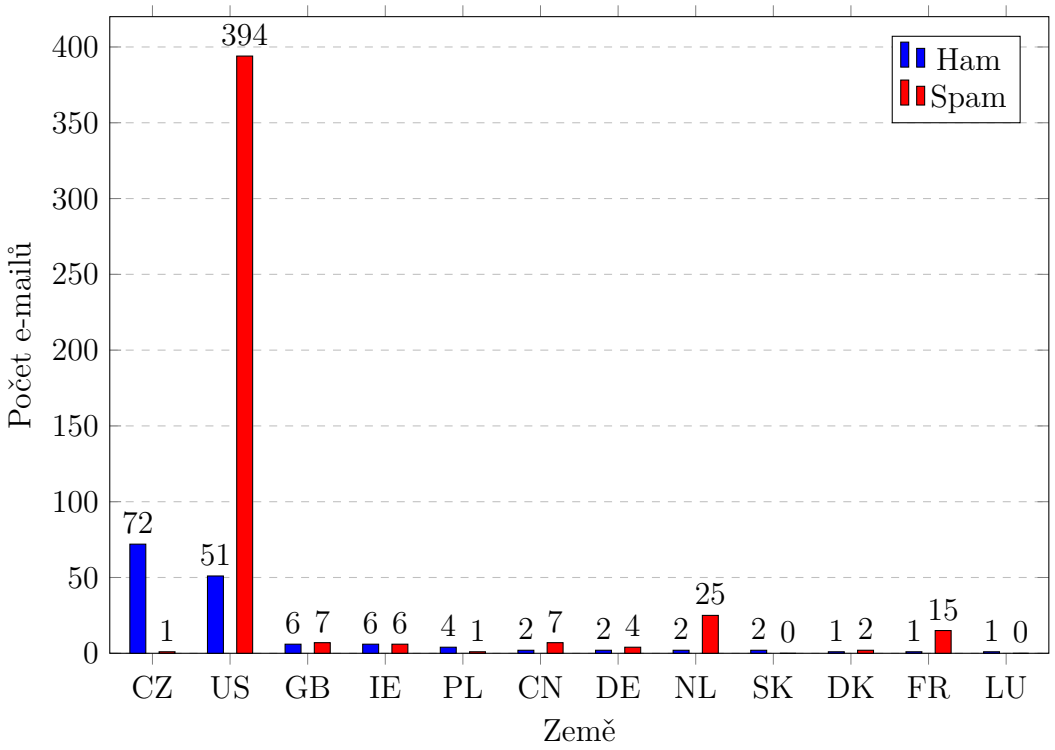
Jako první je otestována a vytvořena statistika z první části datasetu. Z výsledků tohoto testu jsou pak zablokovány země, ze kterých šel pouze spam a žádný ham. Tento postup reflektuje, jaké by mohlo být skutečné použití modulu *Geolock* v praxi.

V tabulce 6.2 lze vidět rozložení prvních 600 zpráv spamu na země původu. Na prvním místě s 394 spam zprávami se nachází Spojené státy americké, následované Finskem s 58 zprávami a Nizozemím s 25 zprávami. V porovnání se Spamhaus statistikami lze vidět, že i zde jsou Spojené státy americké na prvním místě a objevuje se zde i Ruská federace, Čína, Indie, s 2 zprávami Brazílie a s pouhou 1 zprávou Japonsko. Překvapivý byl vyšší počet spamu z Finské republiky, Nizozemska, Francouzské republiky a Nigerijské federativní republiky, kde tyto země se neobjevují v tabulce od Spamhaus. Celkově spam byl obdržen z 39 různých zemí.

Tab. 6.2: Rozložení první poloviny spamu

Země	**Počet spamu**
US	394
FI	58
NL	25
FR	15
RU	13
NG	12
IN	8
CN, GB	7
IE	6
CA, ID	5
DE, IT	4
VE, VN	3
BG, BR, CH, DK, KR, MY, SG, TG	2
AU, BF, BJ, CZ, EC, HK, JP, KH, LT, LV, PL, PR, SI, TW, ZA	1

Pouze z těchto dat by se dalo usoudit, že by bylo vhodné zablokovat například top 5 zemí s nejvíce rozeslaným spamem. Tímto by se zablokovalo celkově 505 spamových zpráv. Nejdříve je ale potřeba se podívat na původ ham zpráv. Na obrázku 6.2 lze vidět graf rozložení původu ham zpráv v kombinaci s dříve zjištěným rozložením spamu. V případě použití původního návrhu, který by zablokoval top 5 zemí šířících spam, došlo by k nežádoucí blokaci 54 legitimních zpráv.Kdyby byl použit původní návrh zablokování top 5 zemí šířících spam, tak by došlo k blokaci 54 legitimních zpráv, což není žádoucí.



Obr. 6.2: Rozložení první poloviny hamu s uvedeným počtem spamu

Na základě těchto dat byly zablokovány všechny země, ze kterých došel spam a zároveň nedošla jediná legitimní zpráva. Země, které byly zablokované, lze vidět níže:

Výpis 6.5: Zablokované země

```
AU, BF, BG, BJ, BR, CA, CH, EC, FI, HK, ID, IN, IT, JP,
KH, KR, LT, LV, MY, NG, PR, RU, SG, SI, TG, TW, VE, VN, ZA
```

S nově blokovanými zeměmi byla otestována druhá část datasetu. Rozložení druhé poloviny hamu lze vidět v tabulce 6.3. Oproti první části hamu se zde objevuje méně zemí a opět dominuje Česká republika s 97 zprávami a Spojené státy americké s 45 zprávami. Ze zablokovaných zemí se zde objevila pouze jediná zpráva, a to z Ruské federace.

Tab. 6.3: Rozložení druhé poloviny hamu

Země	**Počet hamu**
CZ	97
US	45
GB	2
PL	2
BE	1
HU	1
RU	1
SK	1

Rozložení spamu v druhé části datasetu lze vidět v tabulce 6.4. Země s největším počtem spamu jsou velice podobné jako v první části datasetu, kdy znovu tabulku vedou Spojené státy americké následované Finskem. Celkově spam v druhé části datasetu přišel z 41 různých zemí.

Tab. 6.4: Rozložení druhé poloviny spamu

Země	**Počet spamu**
US	358
FI	50
FR	27
NL	21
IN	16
RU	14
JP	11
CA, NG	10
GB	9
ZA	8
CN, DE	6
PL, VN	5
ES, LT, UA	4
BJ, CH, TW	3
AZ, BR, ID	2
AE, CL, HU, IR, IT, KH, KR, MU, NZ, PH, PK, RO, SG, SI, TG, TH, UZ	1

Na základě blokace zemí z první části spamu, se zablokovalo celkově 132 zpráv, z čehož bylo 131 spamu a 1 zpráva byla ham. Rozložení zablokovaných spam zpráv lze vidět v tabulce 6.5.

Tab. 6.5: Počet zablokovaných zpráv v druhé polovině spamu

Země	**Počet spamu**
FI	50
RU	14
JP	11
CA, NG	10
ZA	8
VN	5
LT	4
BJ, CH, TW	3
BR, ID	2
IT, KH, KR, SG, SI, TG	1

V tabulce 6.6 je obecná matice záměn, kde zkratky v jednotlivých polích značí následovné:

- **TP** (*True Positive*) - skutečně pozitivní, počet správně zablokovaného spamu.
- **FN** (*False Negative*) - falešně negativní, počet nezablokovaného spamu.
- **FP** (*False Positive*) - falešně pozitivní, počet špatně zablokovaných zpráv.
- **TN** (*True Negative*) - skutečně negativní, počet správně nezablokovaných zpráv.

Tab. 6.6: Matice záměn obecně

	Předpokládaný pozitivní stav	**Předpokládaná negativní stav**
Pozitivní stav	TP	FN
Negativní stav	FP	TN

V tabulce 6.7 je matice záměn vyplněná s daty z testování druhého datasetu, kde **TP** je 131 zpráv, **FN** je 469 zpráv, **FP** je 1 zpráva a **TN** je 149 zpráv.

Tab. 6.7: Matice záměn modulu *Geolock*

	Předpokládaný pozitivní stav	**Předpokládaná negativní stav**
Pozitivní stav	131	469
Negativní stav	1	149

Z matice záměn je následně vypočítaná specificita, která měří podíl podíl negativů, které jsou správně identifikovány. Rovnice a výpočet lze vidět níže, kde výsledek je $0{,}99\overline{3}$:

$$\text{Specificita} = \frac{TN}{TN + FP} = \frac{149}{149 + 1} = 0,99\overline{3}. \tag{6.1}$$

Je vypočtena i přesnost, která udává, jaká je pravděpodobnost, že je pozitivní výsledek správně klasifikován. Přesnost vyšla $0{,}99\overline{24}$ a její rovnice s výpočtem lze vidět níže:

$$\text{Přesnost} = \frac{TP}{TP + FP} = \frac{131}{131 + 1} = 0,99\overline{24}. \tag{6.2}$$

Následně je vypočtena správnost, která udává, kolik zpráv modul *Geolock* klasifikuje správně. Výsledek tohoto výpočtu je $0{,}37\overline{3}$. Rovnice a výpočet správnosti lze vidět níže:

$$\text{Správnost} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{131 + 149}{131 + 149 + 1 + 469} = 0,37\overline{3}. \tag{6.3}$$

Matthewsův korelační koeficient, zkratkou MCC, udává rozhodovací sílu binárního klasifikátoru a nabývá hodnot v intervalu od -1 do 1, kde -1 je nejhorší rozhodovací schopnost a 1 je perfektní predikce. Obecný vzorec MCC je vidět níže:

$$\text{MCC} = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(FP + FN)(TN + FP)(TN + FN)}}. \tag{6.4}$$

Na rozdíl od dříve vypočtené správnosti je MCC vhodnější na skupiny dat různých velikostí [63]. V tomto případě jsou různé velikosti skupin, jelikož je v druhé části datasetu 150 ham zpráv a 600 spam zpráv. Kvůli tomu by tato hodnota měla být reprezentativnější reálnému použití modulu *Geolock*. Vypočítaný MCC s hodnotami je roven 0,2223 a tento výpočet lze vidět níže:

$$\text{MCC} = \frac{131 * 149 - 1 * 469}{\sqrt{(131 + 1)(1 + 469)(149 + 1)(149 + 469)}} \approx 0,2223. \tag{6.5}$$

Výsledná hodnota je pozitivní a modul přispívá k ochraně proti spamu. Tento výsledek čistě závisí na správném nastavení blokovaných zemí a nepředpokládá se, že by byl používán pouze tento modul samotný pro klasifikaci spamu.

V tabulce 6.8 je vidět porovnání již ukázané tabulky se Spamhaus statistikou z 22. 4. 2022, která je vlevo, a tabulky 10 zemí nejvíce posílající spam v celém testovaném datasetu z poštovní schránky českého uživatele, která je vpravo.

Tab. 6.8: Spamhaus statistika (vlevo) a celkový počet spamu v datasetu (vpravo)

Země	Počet unikátních IP adres
US	3393
CN	2432
RU	763
MX	638
DO	588
SA	537
UY	438
IN	400
BR	399
JP	351

Země	Počet spamu
US	752
FI	108
NL	46
FR	42
RU	27
IN	24
NG	22
GB	16
CA	15
CN	13

V obou tabulkách jsou na prvním místě jako největší zdroj spamu Spojené státy americké, ale jinak je v mezi tabulkami značný rozdíl. Čína, Rusko a Indie se nachází v obou tabulkách, ale na jiných pozicích, kde například Čína je v datech od společnosti Spamhaus na druhé pozici, zatímco v datasetu spamu českého uživatele se nachází až na poslední pozici. Země jako jsou Mexiko nebo Saudská Arábie, se ve statistice spamu v datasetu nevyskytují, ale naopak vysoce spamující země Finsko, Nizozemsko a Francie se vůbec nenachází v datech od společnosti Spamhaus. Celkově se v obou tabulkách zároveň nachází 4 země a ty jsou v tabulkách barevně vyznačené.

Toto porovnání ukazuje, že správce nemůže blokovat čistě podle Spamhaus či podobných globálních statistik a je nutné si prvně udělat vlastní relevantní statistiku, podle které lze zvolit blokované země. Vhodnou volbou blokovaných zemí může tedy modul *Geolock* snížit počet doručených spamů a je na uživateli, zda ho použije adekvátním způsobem.

Závěr

V rámci práce byla popsána problematika spamu a hrozeb spojených s nevyžádanou a škodlivou poštou. Konkrétně se jednalo o spam obecně, sociální inženýrství, kde byly blíže analyzovány kategorie jako je phishing nebo sextortion, a malware. Následně byl popsán e-mail, jeho součásti a cesta sítí. Je zde i bližší pohled na e-mailovou hlavičku a její podvržení. Dále byly rozebrány metody filtrování e-mailů a zvolené filtrovací programy. Potom byly popsány možné aktivní a pasivní metody geolokace a vybrané geolokační databáze.

Cílem diplomové práce bylo vytvořit modul do antispamového softwaru zajišťující filtraci e-mailů za využití geolokace odesílatele. Tento cíl byl splněn, kde v rámci práce byl vytvořen vlastní modul s názvem *Geolock* do antispamového programu SpamAssassin, který využívá geolokační databázi IP2Location. V práci byl popsán modul *Geolock* a jeho implementace do antispamového softwaru. Modul napsaný v jazyce Perl umožňuje filtrovat zprávy pocházející ze zvolené země. Uživatel si může zvolit libovolný počet zemí, ze kterých chce zprávy blokovat. Funkce zajišťující zpracování konfiguračního souboru a hlavní funkce odpovědná za získání IP adresy odesílatele z hlavičky e-mailu, vyhledání země v geolokační databázi a následně porovnání, zda se jedná o blokovanou zemi, jsou v práci popsány. Modul umožňuje uživateli nastavit podezření na podvržení e-mailové hlavičky, při kterém se IP adresa odesílatele z hlavičky získává jiným způsobem.

V poslední části je ukázáno správné chování modulu *Geolock* na testové zprávě a je uvedená konečná hlavička e-mailu po zpracování. Následně byl modul otestován na datasetu spamu a hamu z poštovní schránky českého uživatele. Tento dataset byl rozdělen na dvě stejně velké části, kde z výsledků první části byly zablokovány země, ze kterých šel pouze spam a zároveň nepřišel žádný ham. Výsledky zablokování v druhé části datasetu ukázaly, že blokace na základě země odesílatele je přínosná.

Stejně jako globální statistiky vyšly jako největší zdroj spamu Spojené státy americké, ale zároveň byly zdrojem velkého počtu legitimních zpráv, proto nemohly být zablokovány. Pro správné použití modulu je důležité, aby si uživatel udělal vlastní statistiku hamu a spamu, aby nedocházelo k blokaci legitimních zpráv. Modul *Geolock* byl vydán na platformě GitHub pod svobodnou licencí MIT a může ho kdokoliv použít. Práce byla také prezentována na konferenci STUDENT EEICT 2022.

Literatura

[1] VEZINA, Luc. *9 Ways to Eliminate Spam in Your Community Forum*. Vanilla Forums [online]. 29. 8. 2020 [cit. 6. 12. 2021]. Dostupné z: <https://blog.vanillaforums.com/product/9-ways-to-eliminate-spam-in-your-community-forum>

[2] RAO, Justin M a David H REILEY. *The Economics of Spam*. Journal of Economic Perspectives [online]. 2012, 29. 8. 2020, 2012(Vol. 26 3), 87-110 [cit. 6. 12. 2021]. ISSN 0895-3309. Dostupné z: doi:10.1257/jep.26.3.87

[3] FUKÁRKOVÁ, Barbora a Petra MIKULOVÁ. *Techniky sociálního inženýrství: Příběhy sociálního inženýrství*. MUNI CSIRT-MU: Kyberbezpečnost na univerzitě [online]. [cit. 6. 12. 2021]. Dostupné z: <https://security.muni.cz/socialni_inzenyrstvi>

[4] *What is Typosquatting? – Definition and Explanation*. Kaspersky [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.kaspersky.com/resource-center/definitions/what-is-typosquatting>

[5] *2018 Internet Crime Report*. Internet Crime Complaint Center (IC3) [online]. 22. 4. 2019 [cit. 6. 12. 2021]. Dostupné z: <https://www.ic3.gov/Media/PDF/AnnualReport/2018_IC3Report.pdf>

[6] *Vishing a spoofing*. Policie České republiky [online]. červen 2021 [cit. 6. 12. 2021]. Dostupné z: <https://www.policie.cz/clanek/vishing-a-spoofing.aspx>

[7] *Upozornění na vishing zneužívající identitu bankovních institucí*. Národní úřad pro kybernetickou a informační bezpečnost [online]. 20. 4. 2021 [cit. 6. 12. 2021]. Dostupné z: <https://nukib.cz/cs/infoservis/hrozby/1705-upozorneni-na-vishing-zneuzivajici-identitu-bankovnich-instituci/>

[8] *Vishing: Jak ho rozeznat a vyhnout se mu?* ESET [online]. 25. 8. 2021 [cit. 6. 12. 2021]. Dostupné z: <https://www.eset.com/cz/blog/hrozby/vishing-jak-ho-rozeznat-a-vyhnout-se-mu/>

[9] *Vishing: Buďte obezřetní při sdělování údajů přes telefon*. Česká spořitelna [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.csas.cz/cs/o-nas/bezpecnost-ochrana-dat/vishing>

[10] *Vishing: Upozorňujeme na telefonáty zneužívající jméno ČNB*. Česká národní banka [online]. 11.8.2021 [cit. 6. 12. 2021]. Dostupné z: <https://www.cnb.cz/cs/dohled-financni-trh/ochrana-spotrebitele/upozorneni/Vishing-Upozornujeme-na-telefonaty-zneuzivajici-jmeno-CNB/>

[11] *Počítačové viry, antivirová ochrana a bezpečnost na internetu*. Katedra technické a informační výchovy [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.ped.muni.cz/wtech/u3v/iepp/05.pdf>

[12] ANUAR, Nor Badrul, Rosli SALLEH a Ahmad FIRDAUS. *The rise of "malware": Bibliometric analysis of malware study*. Journal of Network and Computer Applications [online]. 2016, 75(9), 58-76 [cit. 6. 12. 2021]. ISSN 10848045. Dostupné z: doi:10.1016/j.jnca.2016.08.022

[13] MOIR, Robert. *Defining Malware: FAQ*. Microsoft [online]. 4. 1. 2009 [cit. 6. 12. 2021]. Dostupné z: <https://docs.microsoft.com/en-us/previous-versions/tn-archive/dd632948(v=technet.10)>

[14] *Ransomware attacks: detection, prevention and cure*. Network Security [online]. 2016, 2016(9), 5-9 [cit. 6. 12. 2021]. ISSN 13534858. Dostupné z: doi:10.1016/S1353-4858(16)30086-1

[15] *Porušení povinností při zpracování osobních údajů*. Úřad pro ochranu osobních údajů [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.uoou.cz/poruseni-povinnosti-pri-zpracovani-osobnich-udaju/ds-1487/archiv=0&p1=1483>

[16] *The State of Ransomware 2021*. SOPHOS [online]. duben 2021 [cit. 6. 12. 2021]. Dostupné z: <https://secure2.sophos.com/en-us/content/state-of-ransomware.aspx>

[17] MAGDOŇOVÁ, Jana. *Na nemocnici v Benešově útočil ruský virus Ryuk. Jermanová odmítá, že by někdo požadoval výkupné*. iROZHLAS [online]. 14 .1. 2020 [cit. 6. 12. 2021]. Dostupné z: <https://www.irozhlas.cz/zpravy-domov/nemocnice-benesov-kyberneticky-utok-ransomware-vykupne-ochrana-osobnich-udaju_2001140615_cha>

[18] *Fakultní nemocnice Brno čelí kybernetickému útoku. Denně prověřuje na 20 podezření na koronavirus*. iROZHLAS [online]. 13. 3. 2020 [cit. 6. 12. 2021]. Dostupné z: <https://www.irozhlas.cz/zpravy-domov/fakultni-nemocnice-brno-kyberneticky-utok_2003130756_zit>

[19] *Zpráva o stavu kybernetické bezpečnosti České republiky za rok 2019*. Národní úřad pro kybernetickou a informační bezpečnost [online]. 18. 9. 2020 [cit. 6. 12. 2021]. Dostupné z: <https://www.nukib.cz/download/publikace/zpravy_o_stavu/NUKIB_ZSKB_2019.pdf>

[20] CLULEY, Graham. *The Pirate Bay is cryptomining for Monero with your CPU again: Always read the <small> print*. Graham Cluley: Computer security news, advice, and opinion [online]. 6. 7. 2018 [cit. 6. 12. 2021]. Dostupné z: <https://grahamcluley.com/pirate-bay-cryptomining-monero/>

[21] *Cybercriminals can unknowingly use your computer to generate cryptocurrency*. INTERPOL [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.interpol.int/Crimes/Cybercrime/Cryptojacking>

[22] GUO, Hong, JIN, Bo, a Wei QIAN.*Analysis of Email Header for Forensics Purpose*. 2013 International Conference on Communication Systems and Network Technologies [online]. [cit. 1. 5. 2022]. Dostupné z: doi: 10.1109/CSNT.2013.78

[23] *O poštovních programech*. Seznam Nápověda [online]. [cit. 1. 5. 2022]. Dostupné z: <https://napoveda.seznam.cz/cz/o-postovnich-programech/>

[24] RESNICK, Peter W. *RFC 5322: Internet Message Format*. IETF Datatracker [online]. říjen 2008 [cit. 1. 5. 2022]. Dostupné z: <https://datatracker.ietf.org/doc/html/rfc5322>

[25] Emkei's Anonymous Mailer: Free online anonymous mailer with attachments, encryption, HTML editor and advanced settings... [online]. 2009 [cit. 1. 5. 2022]. Dostupné z: <https://emkei.cz/>

[26] WONG, Meng Weng a Wayne SCHLITT *RFC 4408: Internet Message Format*. IETF Datatracker [online]. duben 2006 [cit. 1. 5. 2022]. Dostupné z: <https://datatracker.ietf.org/doc/html/rfc4408>

[27] CROCKER, Dave, Tony HANSEN a Murray S. KUCHERAWY *RFC 6376: Internet Message Format*. IETF Datatracker [online]. září 2011 [cit. 1. 5. 2022]. Dostupné z: <https://datatracker.ietf.org/doc/html/rfc6376>

[28] KUCHERAWY, Murray S. a Elizabeth ZWICKY *RFC 7489: Internet Message Format*. IETF Datatracker [online]. březen 2015 [cit. 1. 5. 2022]. Dostupné z: <https://datatracker.ietf.org/doc/html/rfc7489>

[29] WAKCHAURE, Sushma L, Shailaja D PAWAR, Ganesh D GHUGE a Bipin B SHINDE. *Overview of Anti-spam filtering Techniques*. International Research

Journal of Engineering and Technology (IRJET) [online]. leden 2017(Volume: 04 Issue: 01), 429-434 [cit. 6. 12. 2021]. ISSN 2395-0056. Dostupné z: <https://www.irjet.net/archives/V4/i1/IRJET-V4I169.pdf>

[30] SAHU, Vikram. *Spam filters, techniques and how to avoid email hitting spam filters?* Netcore Cloud [online]. 31. 5. 2021 [cit. 6. 12. 2021]. Dostupné z: <https://netcorecloud.com/tutorials/spam-filter/>

[31] HUCKABY, Jeff. *The 8 Email Blacklists You Should Actually Care About.* RackAID: Linux Server Support & Management [online]. 3. 1. 2019 [cit. 6. 12. 2021]. Dostupné z: <https://www.rackaid.com/blog/email-blacklists/>

[32] *Apache SpamAssassin* [online]. [cit. 6. 12. 2021]. Dostupné z: <https://spamassassin.apache.org/>

[33] GABRIELOVÁ, Rita. *Emailovou poštu na Seznamu chrání nový antispamový filtr.* Protext: PR služby ČKT [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.protext.cz/zprava.php?id=7151>

[34] *Antispamová ochrana.* FI MU [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.fi.muni.cz/tech/unix/spams-and-viruses.html.cs>

[35] *Compare Rspamd with other spam filters.* Rspamd: Fast, free and open-source spam filtering system [online]. [cit. 6. 12. 2021]. Dostupné z: <https://rspamd.com/comparison.html>

[36] CORBET, Jonathan. *Spam filtering with Rspamd.* LWN.net [online]. 1. 9. 2017 [cit. 6. 12. 2021]. Dostupné z: <https://lwn.net/Articles/732570/>

[37] HAMWORTH, Jessica. *C is 'least secure' programming language, study claims.* The Dailg Swig: Cybersecurity news and views [online]. 20. 3. 2019 [cit. 6. 12. 2021]. Dostupné z: <https://portswigger.net/daily-swig/c-is-least-secure-programming-language-study-claims>

[38] KERNIGHAN, Brian W. a Dennis M. RITCHIE. *Programovací jazyk C.* 2. vydání. Přeložil Zbyněk ŠÁVA. Brno: Computer Press, 2019. ISBN 9788025149652

[39] VERNER, Lukáš a Dan KOMOSNÝ. *Geolokace síťových zařízení v internetových sítích.* Elektrorevue [online]. 2011, 13(3) [cit. 6. 12. 2021]. ISSN 1213-1539. Dostupné z: <http://www.elektrorevue.cz/cz/clanky/komunikacni-technologie/0/geolokace-sitovych-zarizeni-v-internetovych-sitich/>

[40] BALEJ, Jiří. *Srovnání přesnosti aktivních geolokačních technik*. Access server [online]. 11. 7. 2012 [cit. 6. 12. 2021]. Dostupné z: <http://access.fel.cvut.cz/view.php?cisloclanku=2012070001>

[41] BALEJ, Jiří a Dan KOMOSNÝ. *Zdroje zpoždění při komunikaci v Internetu*. Elektrorevue [online]. 2010, 12(3) [cit. 6. 12. 2021]. ISSN 1213-1539. Dostupné z: <http://www.elektrorevue.cz/cz/clanky/komunikacni-technologie/0/zdroje-zpozdeni-pri-komunikaci-v-internetu/>

[42] BALEJ, Jiří. *Aktivní IP geolokace pro verifikaci pozic stanic v Internetu*. Brno, 2017, 95 s. Dizertační práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: doc. Ing. Dan Komosný, Ph.D.

[43] CROVELLA, Mark a Serge FDIDA. *Constraint-based geolocation of internet hosts*. Proceedings of the 4th ACM SIGCOMM conference on Internet measurement - IMC '04 [online]. New York, New York, USA: ACM Press, 2004, 25. 10. 2004, 4, 288-293 [cit. 6. 12. 2021]. ISBN 1581138210. Dostupné z: doi:10.1145/1028788.1028828

[44] DAVIS, C, P VIXIE, T GOODWIN a I DICKINSON. *Rfc1876: A Means for Expressing Location Information in the Domain Name System*. IETF Datatracker [online]. leden 1996 [cit. 6. 12. 2021]. Dostupné z: <https://datatracker.ietf.org/doc/html/rfc1876>

[45] *Geolocation API*. Google Maps Platform [online]. [cit. 6. 12. 2021]. Dostupné z: <https://developers.google.com/maps/documentation/geolocation/overview>

[46] *The Number Resource Organization* [online]. [cit. 6. 12. 2021]. Dostupné z: <https://nro.net/>

[47] *IP Information for 147.229.26.152*. DomainTools [online]. [cit. 6. 12. 2021]. Dostupné z: <https://whois.domaintools.com/147.229.26.152>

[48] SHAVITT, Yuval a Noa ZILBERMAN. *A Study of Geolocation Databases*. ArXiv.org e-Print archive [online]. 31. 5. 2010 [cit. 6. 12. 2021]. Dostupné z: <https://arxiv.org/abs/1005.5674>

[49] *IP2Location Commercial vs. IP2Location LITE comparison*. LITE IP2Location: Free IP Geolocation Database [online]. [cit. 6. 12. 2021]. Dostupné z: <https://lite.ip2location.com/edition-comparison>

[50] *IP2Location IP Geolocation and IP2Proxy Databases*. IP2Location: IP Address to IP Location and Proxy Information [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.ip2location.com/database>

[51] *GeoIP2 Databases*. MAXMIND: IP Geolocation and Online Fraud Prevention [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.maxmind.com/en/geoip2-databases>

[52] *IP Address Lookup Hostip.info*. HostIP [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.hostip.info/>

[53] *Comprehensive IP Geolocation Database Download*. DB-IP: IP Geolocation API & Free Address Database [online]. [cit. 6. 12. 2021]. Dostupné z: <https://db-ip.com/db/>

[54] *Linux Dokumentační projekt*. 4. vyd. Computer Press, 2008. 1336 s. ISBN: 978-80-251-1525-1.

[55] *Debian – The Universal Operating System* [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.debian.org/>

[56] *Geo::IP2Location*. Meta::cpan: A search engine for CPAN [online]. [cit. 6. 12. 2021]. Dostupné z: <https://metacpan.org/pod/Geo::IP2Location>

[57] JANOUŠEK, J. *Pozice stanic v síti Internet*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2015. 47 s. Vedoucí bakalářské práce doc. Ing. Dan Komosný, Ph.D.

[58] *ISO 3166: Country Codes*. ISO: International Organization for Standardization [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.iso.org/iso-3166-country-codes.html>

[59] *Area Code Coverage*. IP2Location: IP Address to IP Location and Proxy Information [online]. [cit. 6. 12. 2021]. Dostupné z: <https://www.ip2location.com/area-code-coverage>

[60] *GitHub*. GitHub [online]. [cit. 1. 5. 2022]. Dostupné z: <https://github.com/>

[61] *The MIT License*. Open Source Initiative [online]. [cit. 1. 5. 2022]. Dostupné z: <https://opensource.org/licenses/MIT>

[62] *The Top 10 Worst Countries*. The Spamhaus Project [online]. [cit. 1. 5. 2022]. Dostupné z: <https://www.spamhaus.org/statistics/countries/>

[63] Chicco, D., Jurman, G. *The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation.* BMC Genomics 21, 6 (2020)[online]. [cit. 1. 5. 2022]. doi:10.1186/s12864-019-6413-7

Seznam symbolů a zkratek

AFRINIC African Network Information Centre

APNIC Asia Pacific Network Information Centre

ARIN American Registry for Internet Numbers

BCC Blind Carbon Copy

CBG Constraint Based Geolocation

CC Carbon Copy

CPAN Comprehensive Perl Archive Network

DKIM DomainKeys Identified Mail

DMARC Domain-based Message Authentication, Reporting, and Conformance

DoS Denial of Service

DNS Domain Name System

FN False Negative

FP False Positive

HTML HyperText Markup Language

IANA Internet Assigned Numbers Authority

ICMP Domain Name System

IMAP Internet Message Access Protocol

IP Internet Protocol

ISP Internet Service Provider

IT Information Technology

LACNIC Latin America and Caribbean Network Information Center

MAC Media Access Control

MCC Mobile Country Code

MD5 Message Digest algorithm 5

MDA Mail Delivery Agent

MNC Mobile Network Code

MSA Message Submission Agent

MTA Mail Transfer Agent

MUA Mail User Agent

POP3 Post Office Protocol version 3

RIPE NCC Réseaux IP Européens Network Coordination Centre

RIR Regional Internet Registry

RTT Round-Trip Delay

SMTP Simple Mail Transfer Protocol

SOI Speed of Internet

SPF Sender Policy Framework

SSID Service Set Identifier

TLD Top-Level Domain

TN True Negative

TOR The Onion Router

TP True Positive

URL Uniform Resource Locator

VPN Virtual Private Network

A Obsah elektronické přílohy

V příloze lze nalézt 5 souborů, a to hlavní modul *Geolock.pm*, konfigurační soubor modulu *Geolock.cf*, soubor s MIT licencí *LICENCE*, *README.md* soubor použitý na platformě GitHub a testovací e-mail, který byl použit v práci, *test.txt*.

Modul byl vyvíjen ve virtuálním stroji Debian 11 Bullseye za použití programovacího jazyku Perl verze 5.32.1 a SpamAssassin verze 3.4.6.

- `/` ..... kořenový adresář přiloženého archivu
  - `Geolock.pm` ..... Modul napsaný v Perlu
  - `Geolock.cf` ..... Konfigurační soubor pro modul Geolock
  - `LICENCE` ..... MIT licence
  - `README.md` ..... Readme obsahující informace o programu
  - `test.txt` ..... Testovací e-mail použitý v práci v textové podobě