VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV INFORMAČNÍCH SYSTÉMŮ
DEPARTMENT OF INFORMATION SYSTEMS

CENZURA NA INTERNETU

CENSORSHIP IN THE INTERNET

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE — **MICHAL RAJECKÝ**
AUTHOR

VEDOUCÍ PRÁCE — **Mgr. Ing. PAVEL OČENÁŠEK, Ph.D.**
SUPERVISOR

BRNO 2020

Vysoké učení technické v Brně
Fakulta informačních technologií

Ústav informačních systémů (UIFS) Akademický rok 2019/2020

Zadání bakalářské práce

Student: **Rajecký Michal**
Program: Informační technologie
Název: **Cenzura na Internetu**
Censorship in the Internet
Kategorie: Web

Zadání:

1. Prostudujte aktuální problematiku cenzury na Internetu ve vybraných zemích. Po konzultaci s vedoucím práce vyberte jednu zemi a pro tuto řešte následující body.
2. Proveďte důkladnou studii aktuálně používaných cenzorských prostředků. Prostudujte také existující prostředky a projekty na zjišťování, analýzu a obejití cenzury.
3. Na základě výsledků analýzy navrhněte aplikaci umožňující praktické ověření, případně zmapování konkrétních cenzorských prostředků v dané oblasti.
4. Navrženou aplikaci implementujte a ověřte v praktických podmínkách.
5. Proveďte důkladnou diskuzi získaných výsledků.
6. Ve spolupráci s vedoucím práce shrňte výsledky do závěrečné zprávy, která bude publikována v zahraničních odborných zdrojích.

Literatura:

- Stallings, W.: Cryptography and network security. Prentice-Hall, 1999. ISBN 0-13-869017-0.
- Bishop, M.: Computer security: Art & Science. Addison-Wesley, Boston, 2003, ISBN 0-201-44099-7.
- Menezes, A. J., Oorschot, P.C. van, Vanstone, S.A.: Handbook of Applied Cryptography. CRC Press, 1996, ISBN 0-8493-8523-7 <http://www.cacr.math.uwaterloo.ca/hac/>.

Pro udělení zápočtu za první semestr je požadováno:

- Body 1 až 3.

Podrobné závazné pokyny pro vypracování práce viz https://www.fit.vut.cz/study/theses/

Vedoucí práce: **Očenášek Pavel, Mgr. Ing., Ph.D.**
Vedoucí ústavu: Kolář Dušan, doc. Dr. Ing.
Datum zadání: 1. listopadu 2019
Datum odevzdání: 28. května 2020
Datum schválení: 16. října 2019

Abstrakt

Cenzúra na internete je fenoménom tejto doby, ktorý výrazne obmedzuje slobodu slova. Cieľom tejto práce bolo zistiť aktuálny stav internetovej cenzúry v Číne. V rámci analýzy som navrhol a implementoval nástroj pre testovanie dostupnosti internetových stránok a detekciu cenzúry potencionálne blokovaných kľúčových slov. Tento nástroj overil rôzne úrovne prístupu k cieľovému serveru.

Analýza preukáza, že mesto Peking tvorí významnú časť čínskeho systému pre realizáciu cenzúry, nakoľko sa v ňom stráca veľká časť internetovej komunikácie, až 97% zo všetkých stratených dát. Testovanie tiež odhalilo premenlivosť cenzúry v čase. Priemerne 57,6% internetových stránok dostávalo počas testovacieho obdobia rozdielne výsledky. Prínosom tejto práce je podať aktuálny obraz o stave čínskej internetovej cenzúry a jej dopade na užívateľov. Cenzúra v Číne sa dá považovať za centralizovanú, veľmi rozsiahlu a časovo premenlivú.

Abstract

Internet censorship is a phenomenon of the time, which significantly restricts freedom of speech. The aim of this work was to find out the current state of Internet censorship in China. As part of the analysis, I designed and implemented a tool for testing website availability and detecting censorship of potentially blocked keywords. This tool verified different levels of access to a target server.

The analysis showed that the city of Beijing forms a significant part of China's system in the implementation of censorship, as it loses a large part of Internet communication, up to 97% of all lost data. Testing also revealed the variability of censorship over time. On average, 57.6% of websites received different results during the trial period. This article provides an up-to-date picture of the state of Chinese Internet censorship and its impact on users. Censorship in China can be considered centralized, very extensive and variable over time.

Kľúčové slová

cenzúra, Čína, firewall, filtrovanie, interent, proxy

Keywords

censorship, China, firewall, filtering, Internet, proxy

Citácia

RAJECKÝ, Michal. *Cenzura na Internetu*. Brno, 2020. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Mgr. Ing. Pavel Očenášek, Ph.D.

Cenzura na Internetu

Prehlásenie

Prehlasujem, že som túto bakalársku prácu vypracoval samostatne, pod vedením pána Pavla Očenáška. Ďalšie informácie mi poskytol pán Jiří Navrátil, lokálny manažér služby Planetlab.

Uviedol som všetky literárne pramene, publikácie a ďalšie zdroje, z ktorých som čerpal.

.....
Michal Rajecký
4. júna 2020

Poďakovanie

Poďakovanie patrí vedúcemu práce, pánovi Pavlovi Očenáškovi, za vedenie a jeho odborné postrehy. Poďakovanie patrí tiež lokálnemu manažérovi služby Planetlab, Jiřímu Navrátilovi, za cenné informácie v oblasti prístupu k zahraničným internetovým uzlom.

Obsah

Kapitola 1

Úvod

Internet, ako ho poznáme dnes v demokratických režimoch, nie je samozrejmosťou všade na svete. Existujú krajiny, ktoré obmedzujú slobodu slova zavádzaním rôznych foriem cenzúry. Dnes je doba rýchleho technického pokroku, kedy sa informácie stávajú iba jednotkami a nulami uloženými na serveroch. Dáta sa prenášajú medzi počítačmi rýchlosťou svetla a osobná komunikácia putuje do úzadia. To je dôvod prečo krajiny, realizujúce cenzúru, musia filtrovať informácie aj na internete. Rôzne technológie pre vykonávanie internetovej cenzúry majú rôzne nároky na zariadenia a systémy, na ktorých cenzúra prebieha. Tiež sa líši efektivita a úroveň, na ktorej k filtrovaniu dochádza.

Vývoj technológii pre realizáciu cenzúry je v dnešnej dobe veľmi rýchly a spolu s tým dochádza k vývoju prostriedkov pre jej obchádzanie. Cieľom práce je preskúmať aktuálny stav cenzúry v Číne, na akých úrovniach sa filtrovanie uskutočňuje a do akej miery efektívne blokuje neželané informácie. Predmetom práce je aj skúmanie, ako sa cenzúra prejavuje a aké dopady ma na užívateľov internetu.

Téma internetovej cenzúry je pre mňa veľmi zaujímavá, nakoľko sa jedná o zložitý proces, ktorý v podstate mení obraz internetu. V dnešnej dobre hrá internetová cenzúra kritickú rolu pri získavaní informácií a môže viesť ku skresleným alebo neúplným predstavám o historických udalostiach a k falošným domnienkam o aktuálnom dianí vo svete. Keďže sa jedná o tak závažný aspekt sieťovej prevádzky, že výrazným spôsobom ovplyvňuje každodenný život bežného človeka, rozhodol som sa bližšie preskúmať podobu cenzúry realizovanej v nejakej krajine s represívnym režimom, ktorou sa stala Čína.

Práca sa v následujúcej kapitole venuje popisu sieťovej komunikácie, aby bol čitateľ oboznámený so základnými mechanizmami prenosu dát, ktoré sú potrebné pri študovaní cenzúry. Kapitola 3 ponúka detailný pohľad na fungovanie firewallu a popis jednotlivých spôsobov, ktorými môže firewall fungovať. Podobou cenzúry a možnosťami jej realizácie sa zoberá kapitola 4, ktorá poskytne aj pohľad na rozsiahly systém, ktorým sa filtrujú dáta v Číne spolu s popisom jeho vedľajších účinkov a neúmyselných škôd. Kapitola 5 poskytuje pohľad na technológie, ktorými je možné cenzúru obísť a aj napriek obmedzeniam slobodne pristupovať k zakázaným informáciám. Návrhu nástroja, ktorým bude realizovaná analýza aktuálneho stavu cenzúry, sa venuje kapitola 6. Implementácia je popísaná v 7 a vyhodnocovanie nameraných dát v kapitole 8

Kapitola 2

Sieťový prenos dát na modeli OSI/ISO

Internet, ako ho poznáme dnes je bez pochyby najrozsiahlejší systém, aký bol kedy človekom vytvorený. Obsahuje stovky miliónov počítačov, komunikačných liniek a rôznych sieťových zariadení s miliardami užívateľov, ktorí sa pripájajú pomocou notebookov, tabletov, smartfónov a podobne. Koncové zariadenia sú spolu prepojené sieťou komunikačných liniek a sieťových prepínačov. Spojenie môže byť na fyzickej vrstve realizované rôznymi technológiami vrátane koaxiálnych káblov, medených drôtov, optických vlákien alebo na báze rádiového spektra. Každá metóda ponúka určitú prenosovú rýchlosť vyjadrovanú v bitoch za sekundu. Keď má koncový systém dáta na odoslanie, rozdelí ich na menšie celky, ku ktorým sú pridané bajty hlavičky. Výsledné pakety sú odoslané po sieti k cieľovému zariadeniu, kde je z nich poskladaná originálna správa. [19]

Pre zabezpečenie internetovej komunikácie sú potrebné sieťové protokoly, ktoré plnia dôležitú úlohu v odosielaní a prijímaní dát. Tieto protokoly sú organizované do siedmych vrstiev, z ktorých každá poskytuje určitú funkcionalitu. Rozhranie medzi jednotlivými vrstvami určuje, ako vyššia vrstva pristupuje k službám a informáciám nižšej vrstvy. [19] Referenčný model OSI je znázornený na obrázku 2.1.

Následujúci text sa venuje jednotlivým vrstvám, ich funkcii a významným sieťovým protokolom patriacich na danú vrstvu.

Aplikačná vrstva

Sieťové aplikácie sú dôvodom, prečo sú v dnešnej dobe potrebné sieťové protokoly podporujúce dané aplikácie. Od vzniku internetu bolo vyvinuté obrovské množstvo užitočných a zábavných aplikácii, ktoré predstavovali pre ľudí motiváciu začleňovať internet do každodenného života. Aplikačné dáta, ktoré majú byť prenesené do cieľovej stanice sú rozdelené na menšie celky a následne spracované a odoslané. Na tejto vrstve sa nachádzajú sieťové aplikácie a ich protokoly. Aplikačná vrstva internetu zahŕňa množstvo protokolov, medzi ktoré patria protokoly HTTP pre zasielanie webových dokumentov, SMTP pre doručovanie elektronickej pošty, FTP pre prenos súborov medzi dvoma koncovými zariadeniami alebo DNS pre preklad doménových mien na IP adresy. Zhluky dát prenášané na tejto vrstve sa nazývajú aplikačné správy. Často využívané protokoly, ktoré sa nachádzajú na tejto vrstve sú DNS (ktorý slúži napríklad pre preklad doménových mien na IP adresy), IMAP (pre správu elektronickej pošty) alebo DHCP (využívaný pre automatickú konfiguráciu zariadení pripojených do siete). [19]

Číslo vrstvy	Vrstva	Príklad protokolov
L7	Aplikačná vrstva	DNS, IMAP
L6	Prezentačná vrstva	NCP
L5	Relačná vrstva	SSL-TLS
L4	Transportná vrstva	TCP, UDP
L3	Sieťová vrstva	ICMP, IGMP
L2	Linková vrstva	Ethernet
L1	Fyzická vrstva	RS-422

Obr. 2.1: Referenčný model ISO/OSI popisujúci štruktúru sieťových a komunikačných protokolov [19].

V prípade sieťovej vrstvy je architektúra siete vopred daná a poskytuje predom stanovené služby aplikáciám. Ale pri návrhu sieťových aplikácii je architektúra určená programátorom. Vo väčšine prípadov sa jedná o jeden z dvoch najrozšírenejších prístupov; model klient – server alebo peer-to-peer. [19]

Prezentačná vrstva

Úlohou prezentačnej vrstvy je pomáhať aplikačnej vrstve pri riešení problémov spojených s rozdielnou reprezentáciou údajov. Prenášané dáta sú transformované a formátované tak, aby im rozumeli aplikácie. Táto vrstva zabezpečuje kompresiu a šifrovanie dát. [19]

Relačná vrstva

Relačná vrstva slúži na vytváranie, prevádzkovanie a ukončovanie spojení medzi aplikáciami komunikujúcich strán, inými slovami, definuje pravidlá pre prenosom dát medzi uzlami na sieti. Vrstva tiež rozhoduje o tom, či bude prenos prebiehať poloduplexne či duplexne. [19]

Transportná vrstva

Transportná vrstva zohráva kriticky dôležitú úlohu pri poskytovaní komunikačných služieb aplikačným procesom operujúcich na rôznych zariadeniach. Zabezpečuje logickú komunikáciu1 medzi týmito zariadeniami. Protokoly transportnej vrstvy sú implementované v koncových systémoch, nie v sieťových smerovačoch. Na strane odosielateľa je správa z aplikačnej vrstvy konvertovaná na paket transportnej vrstvy (tzv. segment) rozdeľovaním aplikačných dát na menšie celky, ku ktorým je pridávaná hlavička transportnej vrstvy. Takto vytvorené segmenty sú predané sieťovej vrstve, kde je pripojená hlavička sieťovej vrstvy a poslané k príjemcovi. Je dôležité poznamenať, že sieťové smerovače operujú výhradne na sieťovej vrstve a nepristupujú k obsahu segmentov transportnej vrstvy. Na strane príjemcu je zo sie-

ťového datagramu vybratý segment transportnej vrstvy, ktorý sa predá transportnej vrstve. Tu je datagram spracovaný a príslušné dáta sú sprístupnené cieľovej aplikácii. Dáta môžu byť na transportnej vrstve prenášané viac ako jedným protokolom. Internet disponuje dvomi protokolmi – TCP a UDP. [19]

Ani jeden zo spomínaných dvoch protokolov neposkytuje možnosť šifrovania prenášaných dát. Dáta, ktoré sú umiestnené do sieťovej schránky sú v rovnakej podobe odoslané po sieti. Ak by proces umiestnil citlivé informácie, napríklad heslo, v nezašifrovanej podobe do sieťovej schránky, nezašifrované heslo by putovalo od odosielateľa až k adresátovi. V prípade odchytenia týchto dát útočníkom je heslo nechránené a čitateľné. Vysoké nároky na bezpečnosť a dôvernosť prenášaných informácii prispeli k vytvoreniu kryptografického protokolu SSL (angl. Secure Socket layer), ktorý poskytuje šifrovanie, integritu dát a rovnako aj autentifikáciu koncového zariadenia. Upravená verzia SSL protokolu, označovaná ako TLS (angl. Transport Layer security) bola štandardizovaná organizáciou IETF (pridat referenciu na RFC 4346). [19]

Sieťová vrstva

Úloha sieťovej vrstvy je dopraviť pakety od odosielateľa k príjemcovi. K tomu sa využíva smerovacia tabuľka sieťových smerovačov, vďaka ktorej smerovač vie, na ktoré výstupné rozhranie je nutné prichádzajúci paket odoslať. Smerovanie je proces určovania cesty prenášaných paketov medzi dvoma stanicami. IP paket pozostáva z hlavičky (angl. header) a z dát (angl. payload). Pri preposielaní paketov musí smerovač prečítať obsah hlavičky kde sa nachádza informácia o cieľovej IP adrese aby vedel, kam ďalej paket posielať. [19]

Pri prenose paketov medzi odosielateľom a príjemcom môžu vznikať otázky ohľadom poskytovania rôznych služieb sieťovou vrstvou. Môže sa transportná vrstva spoľahnúť, že sieťová vrstva skutočne doručí paket adresátovi? Bude séria viacerých paketov doručená transportnej vrstve v rovnakom poradí, v akom boli odoslané? Bude sa čas medzi dvoma po sebe prijatými paketmi rovnať oneskoreniu, ktoré vznikne pri odosielaní týchto paketov? Poskytne sieť spätnú väzbu v prípade zahltenia? Odpoveď na tieto a mnohé ďalšie otázky určujú služby poskytované sieťovou vrstvou, medzi ktoré patrí garantovanie doručenia, zachovanie poradia paketov, garancia minimálnej šírky pásma alebo maximálny rozptyl oneskorenia. [19]

Linková vrstva

Sieťová vrstva smeruje pakty cez sériu smerovačov od zdroja k cieľu. Aby bolo možné dostať paket z jedného uzlu na druhý, musí sa spoliehať na linkovú vrstvu. V každom smerovači sieťová vrstva predá paket linkovej vrstve, ktorá sa postará o doručenie paketu k následujúcemu uzlu. Tam dochádza k predaniu paketu z linkovej vrstvy späť na sieťovú. [19]

Služby poskytované linkovou vrstvou závisia od protokol, ktorý na nej operuje. Niektoré protokoly linkovej vrstvy poskytujú spoľahlivé doručenie paketu zo zdrojového uzlu až k cieľu. Je nutné poznamenať, že sa nejedná o spoľahlivé doručovanie protokolu TCP, ktorý zabezpečuje spoľahlivé doručenie medzi dvomi koncovými systémami. Príklad protokolu operujúcom na linkovej vrstve je Ethernet alebo Wi-Fi[19].

Fyzická vrstva

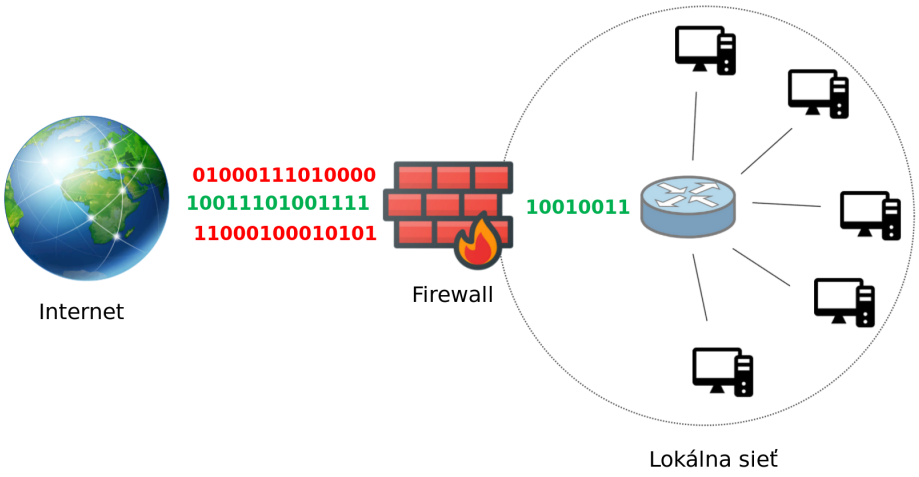
Zatiaľ čo úloha linkovej vrstvy je presúvať celé dátové rámce z jedného uzlu na druhý, fyzická vrstva zabezpečuje prenos jednotlivých bitov z dátových rámcov medzi uzlami. Charakteristiky prenosu závisia na typu prenosového média (môže sa jednať napríklad o optické vlákna či koaxiálne káble).

Kapitola 3

Firewall

Útočníci, ktorí poznajú rozsah IP adries nejakej siete, sú schopní bez problémov zasielať akékoľvek datagramy na adresy z daného rozsahu. Môže sa jednať o mapovanie siete pomocou nástroja ping, skenovanie portov, útoky na zraniteľné zariadenia pomocou upravených paketov, zahltenie vybraných zariadení veľkým množstvom ICMP[1] paketov alebo distribúciu škodlivého softwaru obsiahnutého v prenášaných paketoch. V dnešnej dobe sú rozšírené dve riešenia, ktoré pokrývajú opísanú problematiku. Firewall a IDS[2]. [19] Prvému z nich sa venuje táto kapitola.

Prvým firewallom bol počítač, ktorý mal za úlohu filtrovať sieťový prenos na základe analýzy paketov, pri čom dáta boli filtrované podľa zdrojovej a cieľovej IP adresy. V prípade, že firewall detekoval v prenášaných paketoch určené informácie, mohol ich zahodiť v súlade s implementovanými pravidlami. [10]



Obr. 3.1: Schéma znázorňujúca funkciu firewallu. Dáta, ktoré sú do lokálnej siete prenášane z internetu, prechádzajú cez firewall, ktorý blokuje priechod škodlivých dát [10].

[1]ICMP (Internet Control Message Protocol) operuje na sieťovej vrstve a slúži pre komunikáciu medzi zarideniami a smerovačmi. Bežne sa používa pre informovanie o vzniknutej chybe v komunikácii [19].

[2]IDS (Intrusion Detection System) je nástroj, typicky umiestnený na hranici siete, ktorý slúži na hĺbkovú analýzu paketov. Okrem hlavičiek prebieha tiež analýza dátového obsahu, vrátane dát aplikačnej vrstvy. IDS má databázu so známymi vzorcami škodlivých paketov, ktoré často predstavujú nebezpečenstvo [19].

Firewall je kombinácia hardwarových a softwarových prostriedkov slúžiaca na izoláciu vnútornej siete od zbytku internetu, čím sa priechod niektorých paketov povolí a iných zablokuje. Je to technológia, ktorá umožňuje kontrolovať spojenie medzi okolným svetom a sieťou na vnútornej strane firewallu, ako znázorňuje obrázok 3.1. Firewall má 3 primárne ciele [19]:

- Celý sieťový prenos prechádza cez firewall. Veľké organizácie môžu disponovať viacerými úrovňami firewallu, prípadne ho používať v distribuovanej forme. Avšak umiestnenie firewallu na jediný prístupový bod výrazne zjednodušuje jeho správu a dodržiavanie bezpečnostnej politiky.
- Iba autorizovanej sieťovej prevádzke bude umožnené prejsť z a do internej siete. Keďže všetky prichádzajúce a odchádzajúce dáta putujú cez brány firewallu, neautorizovaný obsah môže byť regulovaný.
- Firewall je zariadenie pripojené do siete. Ak nie je navrhnutý alebo nakonfigurovaný správne, môže to predstavovať kritický problém z pohľadu bezpečnosti siete. V takom prípade firewall poskytuje len falošný dojem bezpečia, čo je v mnohých ohľadoch horšie ako jeho úplná absencia.

Je možné filtrovať ľubovoľný obsah, ktorý je prenášný pod známym protokolom. Tým pádom je kriticky dôležité, aby bola sieťová prevádzka realizovaná pomocou nejakého štandartu (ako napríklad IPX, TCP/IP atď.). Ak dáta dodržujú určitý štandart, je možné podľa neho vykonávať aj filtrovanie, avšak len za predpokladu že zariadenie realizujúce triedenie dát podpruje dané protokoly. [10]

Toto filtrovanie je jednou možnosťou fungovania firewallu. Existujú celkom tri kategórie firewallovej technológie:

- Paketové filtre
- Proxy filtre
- Stavové paketové filtre

Následujúci text priblíži mechanizmus fungovania jednotlivých typov firewallu.

3.1 Paketové filtre

Firewally tejto kategórie operujú na transportnej vrstve. Analyzujú statické informácie v hlavičkách prenášaných paketov. Implementáciou paketových filtrov sa definujú pravidlá, podľa ktorých firewall zamedzuje priechod určitých paketov. Tieto pravidlá môže byť založené na jednom alebo viacerých z nižšie uvedených kritérií:

- IP adresa odosielateľa
- IP adresa príjemcu
- Zdrojový port
- Cieľový port
- Protokol

Medzi vlastnosti paketových filtrov patrí fakt, že si neukladajú žiadne stavové informácie. Pri spracovávaní paketu firewall rozhodne, či bude prijatý alebo odmietnutý a to na základe definovaných pravidiel, no pri ktoromkoľvek ďalšom paketu sa celý proces opakuje od začiatku. Nevýhodami týchto filtrov sú napríklad schopnosť paketov prejsť cez firewall v podobe fragmentov, náročná údržba prístupových zoznamov a neschopnosť filtrovať všetky služby. [10]

3.2 Proxy filtre

Jedná sa o zariadenie, ktoré analyzuje prenášané dáta na vrstvách 4 až 7 sieťového modelu OSI/ISO. Následkom širokého prístupu filtrov k sieťovej prevádzke sa výrazne zvyšuje účinnosť firewallu, no zároveň sa znižuje priepustnosť medzi odosielateľom a príjemcom. Takéto zariadenie ukrýva citlivé dáta pred vonkajšou sieťou, nakoľko užívateľ musí pre prístup k proxy serveru prejsť procesom autentizácie a ustanovenia spojenia. Užívatelia k službám,umiestnených mimo dôveryhodnú sieť, pristupujú pomocou proxy programov, ktoré sa nachádzajú na bráne spojujúcej vnútornú sieť s vonkajšou nechránenou zónou. [10]

Proxy filtre môžu fungovať dvomi spôsobmi. V prvom prípade užívateľ na vnútornej strane firewallu vytvoriť komunikačné spojenie k samotnému firewallu. Následne prebehne overenie užívateľa pomocou jeho autentizačných údajov a podľa toho mu systém poloví okruh prístupových práv pre oblasť mimo vnútornú sieť. V tomto modelu dochádza k vytvoreniu dvoch samostatných spojení, z ktorých prvé slúži pre prenos dát medzi užívateľom a proxy serverom, a druhé pre spojenie proxy s cieľom. [10]

V druhom prípade sa užívateľovi na vnútornej strane firewallu javí komunikácia s cieľom ako jediné priame spojenie. Proxy filter v skutočnosti do tohto procesu vstupuje podľa vopred definovaných pravidiel, napríklad podľa zdrojovej IP adresy môže vykonať proces overenia užívateľa, a vytvorí opäť dve spojenia, rovnako ako v prípade prvého mechanizmu. Z pohľadu užívateľa je táto varianta oveľa transparentnejšia, keďže o nej v podstave nevie.

Proxy filtrovanie nesie so sebou určité nevýhody. Nakoľko proxy server je prostredníkom medzi vnútornou sieťou a vonkajším svetom, stáva sa jediným zraniteľným miestom, ktorého napadnutie ohrozuje bezpečnosť celej siete vo za bránami firewallu. Okrem toho, proces dopĺňania ďalších služieb do firewallu je komplikovaný. Proxy firewall je zpravidla implementovaný na univerzálnom operačnom systéme z dôvodu nevyhnutného prístupu k určitým službám operačného systému, ktoré sú potrebné pre vykonávanie proxy procesov. To má za následok zvýšenú réžiu firewallu a tiež možné bezpečnostné riziká spojené so zraniteľnosťami známeho operačného systému. [10]

3.3 Stavové paketové filtre

Jedná sa o kombináciu tých najlepších vlastností z obidvoch predchádzajúcich technológii. Stavový paketový filter si ukladá všetky stavové informácie o každom spojení vedenom cez firewall. Pri ustanovení spojenia v ktoromkoľvek smeru sa informácie o spojení ukladajú do stavovej tabuľky.

V stavovej tabuľke naviazaných spojení existujú ku každému spojeniu TCP/UDP príslušné informácie, ako napríklad čísla portov, údaje o zdrojovej a cieľovej adrese, informácie o poradových číslach TCP a ďalšie doplňujúce príznaky. Pri ustanovení spojenia cez firewall dochádza k vytvoreniu objektu spojenia. Prenášané pakety sa následne porovnávajú

s údajmi v stavovej tabuľke naviazaných spojení a priechod paketov je umožnený len v prípade, že existuje v tabuľke príslušné spojenie.

Účinnosť tejto metódy spočíva v jej rýchlosti a tiež v porovnávaní jednotlivých paketov s naviazanými komunikačnými spojeniami. V prípade každého spojenia a rovnako aj v prípade nespojovanej transakcie firewall ukladá informácie do tabuľky. K nim sa môže neskôr vracať a rozhodnúť tak, či daný paket patrí niektorému existujúcemu spojeniu, prípadne či sa jedná o prenos paketov z neoprávneného zdroja. [10]

Kapitola 4

Cenzúra na internete

Vývoj a implementácia cenzúry na internete je realizovaná už dlhú dobu. Pôvodne boli cieľom filtrovania rôzne škodlivé druhy software, falošné správy alebo akýkoľvek nechcený či nevyžiadaný obsah za účelom zaručenia bezpečnosti a použiteľnosti služieb internetu a ochranu užívateľa. Hlavným cieľom cenzúry je zabránenie prístupu klienta k určitému obsahu. Realizovaná je pomocou hardwarových a softwarových prostriedkov, ktoré rozhodujú o tom čí vyžiadaný obsah má byť zablokovaný alebo nie. Systém cenzúry aplikovaný na verejný internet a užívateľov v celej krajine vyžaduje spoluprácu všetkých sprostredkovateľov internetu, ktorí vyhovejú požiadavkám na filtrovanie obsahu. To môže spôsobovať problémy komerčným organizáciám, pre ktoré toto opatrenie predstavuje prekážku v ich kreatívnej činnosti. Dodržiavanie cenzúry obmedzuje komplexnosť poskytovaných služieb. [7]

4.1 Technické aspekty cenzúry

Internet je rozsiahla a komplexná sieť sietí obsahujúca veľké množstvo hardwarových systémov, protokolov a služieb. Internetový obsah je možné distribuovať viacerými technológiami. Blokovanie len niektorých technológií môže spôsobiť zavedenie alternatívneho prístupu pri zdieľaní obsahu. Prvý krok pri realizovaní cenzúry je voľba miesta na internete, kde bude pokus o blokovanie obsahu realizovaný. Je nutné zaoberať sa aj znalosťami a vedomosťami, ktorými disponujú užívatelia a organizácie pripojené k sieti, aby bola cenzúra účinná a efektívna. Cenzúru je možné vykonávať viacerými metódami a na rôznej úrovni technickej náročnosti. V prípade politických režimov vyžadujúcich internetovú cenzúru je kľúčové stanoviť, kto má moc rozhodovať o blokovanom obsahu. V demokratických krajinách je nutná iba minimálna miera cenzúry. Cenzúra je zavedená v prípade bežných internetových služieb s cieľom chrániť užívateľa. Môže sa jednať napríklad o mailové služby, ktoré disponujú filtrovaním nevyžiadanej pošty alebo blokovanie pornografického obsahu poskytovateľmi rôznych internetových služieb. Výrazné filtrovanie obsahu je často využívané v represívnych politických režimoch, kde primárny cieľ je zabrániť prístupu užívateľov k neželaným informáciám. Pre vykonávanie tohoto typu cenzúry je nutná rozsiahla technická infraštruktúra. Vláda musí disponovať charakteristikami zakázaného obsahu, ako napríklad kľúčové slová alebo IP adresy. Ich zoznam spravuje bez výraznejšej účasti poskytovateľov internetu, čím sa výrazne znižuje riziko úniku týchto informácií.

V prípade detekcie takéhoto obsahu v sieťovej komunikácii je prístup k obsahu odmietnutý. Zároveň sa informácie o užívateľovi žiadajúceho o daný obsah môžu ukladať do logovacích systémov pre účel podrobnej analýzy, ktorá môže viesť k trestnému stíhaniu. [7]

Aby bolo možné bližšie skúmať používané techniky cenzúry, je dôležité určiť, čo za internetovú cenzúru možno považovať. Je to zamedzenie prístupu užívateľov k určitým webovým stránkam, bez účasti poskytovateľa obsahu a zariadenia, ktorým sa užívateľ pokúša pristupovať k týmto stránkam. [14] Rozličné metódy blokovania sú realizované v rozličných fázach prístupu na internet [16]. Následujúci text sa venuje spôsobom, ktorými je možné filtrovanie dosiahnuť.

4.1.1 Filtrovanie paketov

Filtrovanie paketov je technika používaná sieťovými smerovačmi a ďalšími zariadeniami, ktoré sa podieľajú na prenose dát. Keď prechádzajú jednotlivé pakety cez tieto zariadenie, dochádza k analýze hlavičky. Táto metóda môže byť realizovaná na vrstve L3 alebo vrstve L4, od čoho ďalej závisí jej granularita a spotreba zdrojov na filtrovacom zariadení. Funkcia filtrovania paketov je implicitnou súčasťou sieťových smerovačov patriacich poskytovateľom internetového pripojenia. Vďaka tomu, v rámci prevencie, sú schopní aplikovať istú mieru filtrovania vstupného a výstupného toku dát pre zaistenie bezpečnosti. To znamená, že teoreticky, na realizáciu cenzúry pomocou techniky filtrovania paketov už disponujú potrebným hardwarovým vybavením. Filtrovanie paketov neposkytuje nijakú možnosť informovania užívateľa ohľadom blokovania neželaného obsahu, a rovnako ani ohľadom blokovania bezproblémového obsahu, ktorý je neúmyselne filtrovaný následok vysokej granularity tejto metódy. [14]

4.1.2 Cenzúra na sieťovej vrstve

Na vrstve L3 prebieha inšpekcia IP hlavičky, ktorá obsahuje informácie ohľadom komunikujúcich zariadení. Na tejto úrovni je možné blokovať iba zdrojové alebo cieľové zariadenie na základe ich IP adresy, pri čom je možné odstaviť všetku komunikáciu ktorá odchádza od zariadenia, alebo k nemu prichádza. Keďže ani trojcestné ustanovenie TCP spojenia (angl. 3-way handshake), ktoré sa nachádza na začiatku webovej komunikácie, nie je možné zrealizovať, jedná sa o úplne zablokovanie akejkoľvek internetovej komunikácie. To znamená, že blokovaný nie je len prístup k webovým stránkam, ale aj ku všetkým ostatným službám. Cenzúra na tejto vrstve využíva pomerne málo zdrojov zo zariadení, ktoré ju realizujú. Smerovače aj tak majú za úlohu analyzovať adresy v IP hlavičkách, aby mohli rozhodnúť, kam budú pakety ďalej smerované. Nevýhodou je, že smerovače vedia spracovávať len fixné množstvo pravidiel. [14]

Blokovanie IP adries je možné jednoducho obísť pomocou proxy serverov umiestnených mimo krajiny (viď. 5.2). Ak internetová stránka zmení svoju adresu pričom ponechá svoje pôvodné doménové meno, užívateľ bude schopný bez zmeny pristupovať k danému webu. [4]

4.1.3 Cenzúra na transportnej vrstve

Vrstva L4 využíva všetky informácie z filtrovania na vrstve L3, a naviac analyzuje hlavičky dát obsiahnutých v IP paketoch. Hlavnou informáciou získanou týmto mechanizmom sú čísla sieťových portov. Čísla portov umožňujú používateľom internetu pristupovať k službám s rovnakou IP adresou. Každá známa internetová služba má štandardné číslo protokolu[1], ktorý slúži na smerovanie prichádzajúcich dát k príslušnému softwaru na cieľovom servery.

[1] Napríklad pre HTTP to je port číslo 80, pre DNS port číslo 53 a pre SSH port číslo 22 [19].

Ak sa filtrovanie podľa IP adries rozšíri o informácie ohľadom sieťových portov, bude možné blokovať vybrané služby. Pri blokovaní internetovej komunikácie by stále boli dostupné služby, ktoré nie sú cieľom cenzúry a vďaka tomu by neboli filtrované ako vedľajší účinok. Ak užívateľ pristupuje k cenzurovanému obsahu, spomínané dve metódy budú v konečnom dôsledku znamenať rovnaký záver. Na vrstve L4 je možné zamedziť prístup na internet blokovaním portu 80[2], no stále ide o princíp prepúšťania všetkého alebo ničoho. Neexistuje možnosť kontroly nad mierou aplikovanej cenzúry na daný port. Cenzurovaný obsah bude blokovaný rovnako ako ten, ktorý nie je cieľom cenzúry ale nachádza sa pod rovnakým portom. Vo väčšine prípadov nie sú cieľom cenzúry úplne všetky stránky danej internetovej domény, no pri tomto filtrovaní nebude možné pristupovať ani k legálnemu obsahu dotknutej domény. [14]

4.1.4 Cenzúra na aplikačnej vrstve

Veľká časť internetovej komunikácie používa doménové mená, nakoľko číselné IP adresy sú pre užívateľa ťažko zapamätateľné. To platí najmä pre prehliadanie webov. Pri pokuse navštíviť webovú lokalitu je potrebné vykonať preklad zadaného doménového mena na IP adresu. Túto funkciu plní DNS server, ktorý obsahuje zoznam doménových mien a k nim príslušné adresy. [12] Pri prístupe na webovú stránku je najskôr odoslaná požiadavka na DNS server, ktorý má za úlohu poskytnúť užívateľovi preklad doménového mena na IP adresu. Je to prvá príležitosť pre cenzorskú stranu pre vykonávanie blokovania. [16]

DNS servery sa využívajú ako efektívny a veľmi častý prostriedok pre realizáciu cenzúry [12]. Jednou možnosťou je odmietnutie prekladu, a druhou presmerovanie na inú stránku [16]. Bez požadovanej IP adresy počítač nie je schopný pristúpiť na danú doménu [12]. Cenzúru realizovanú pomocou DNS serverov je možné obísť jednoduchým opatrením. Vzhľadom na to, že blokovaný je len preklad stránky a nie stránka samotná, je možné k nej pristupovať pomocou IP adresy, namiesto doménového mena. Ďalším riešením je posielať požiadavky na preklad doménového mena na iný DNS server, kde nie je implementovaná cenzúra danej webovej lokality. [7] Tento spôsob znázorňuje obrázok 4.1.

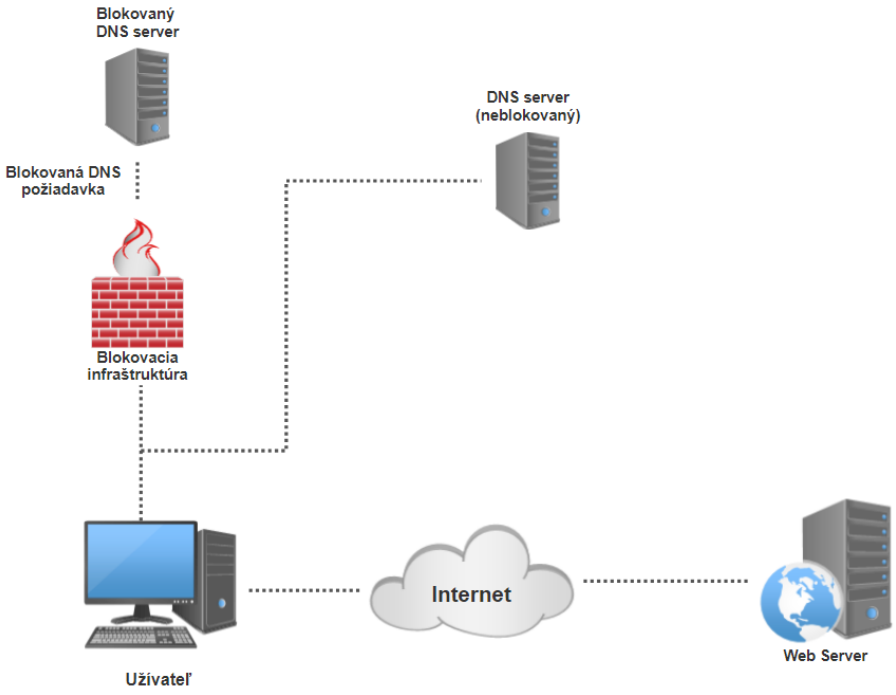
Pomocou manipulovania DNS záznamov nie je komunikácia filtrovaná priamo, ale užívateľovi je znemožnené získať informácie potrebné pre spojenie s cieľovým serverom [14]. V prípade, že z cenzorského DNS serveru príde užívateľovi odpoveď, môže sa jednať o dva typy [16]:

- Nepreložené doménové meno. Užívateľovi je podstrčená správa NXDOMAIN[3] pochádzajúca z middleboxov[4] alebo priamo zo serveru DNS.
- Presmerovanie. Blokované doménové meno sa preloží na nesprávnu IP adresu, ktorá je následne vrátená užívateľovi. Existujú dve možnosti presmerovania [16]:
  - Presmerovanie na súkromnú rezervovanú IP adresu. Požiadavky na preklad doménového mena môžu skončiť vrátením špeciálnej IP adresy, ktorá je rezervovaná pre určité funkcie v sieti, a tým zabraňuje užívateľovi aby opustil svoju lokálnu

[2]Port 80 je štandartným portom pre protokol HTTP. Okrem toho, internetová prevádzka môže využívať aj port 443 v prípade HTTPS a tiež sú bežné servery, ktoré operujú na portoch 8000 alebo 8080. Väčšinou to je prípad Unixových systémov [14].

[3]NXDOMAIN je negatívna odpoveď pochádzajúca z DNS serveru. Poukazuje na fakt, že požadované doménové meno neexistuje [5].

[4]Middlebox je sieťové zariadenie, ktoré analyzuje, upravuje a smeruje sieťovú prevádzku [8].



Obr. 4.1: Využitie neblokovaného DNS serveru pre preklad doménového mena.

sieť. Jedná sa o adresy z rozsahov 127.0.0.0/8, 0.0.0./8, 192.168.0.0/16 alebo 169.254.0.0/16.

– Presmerovanie na inú ako súkromnú rezervovanú IP adresu. Vrátená adresa je platná, no nepatrí cieľovému serveru. Užívateľ je následne smerovaný na internetovú stránku, ktorá je vopred určená cenzormi ako cieľ po neúspešnom pokuse o prístup k nepovoleným informáciám.

4.1.5 Cenzúra s využitím proxy serverov

Jedným zo spôsobov konfigurácie siete je smerovať sieťovú prevádzka cez tzv. proxy servery, Okrem spomenutých výhod sa HTTP proxy používa na zamedzenie prístupu k internetovým stránkam. Proxy server rozhoduje, či žiadosť smerujúca od klienta by mala byť povolená. Odpoveď zo strany serveru je pre proxy server dostupná v plnom rozsahu a vďaka tomu je možné vykonať analýzu a internetové stránky v prípade potreby blokovať. Výhodou je možnosť filtrovania individuálnych stránok namiesto domén či celých internetových serverov. [21]

Analýza obsahu sa väčšinou realizuje pomocou proxy serverov, cez ktoré je smerovaná internetová komunikácia a ktoré nepovolia priechod neželaného obsahu. Táto schéma filtrovania môže byť veľmi presná a vie blokovať jednotlivé internetové stánky alebo ich časti. Dôvodom, prečo mechanizmus proxy serverov nie je nasadený v širokom rozsahu je extrémne vysoká finančná náročnosť. [11]

4.1.6 Kombinácia TCP/IP a HTTP proxy

Cenzúra realizovaná pomocou proxy je založená na smerovaní komunikácie cez proxy servery, ktoré umožňujú jednoduchú analýzu prenášaných dát a ich regulovanie. [21]

Pri vykonávaní kontroly proxy serverom musia byť prenášané pakety spájané, dekódované a neskôr znova odoslané. Z hardwarového hľadiska ide o nákladnú operáciu a preto sa využíva šetrnejší prístup. Základom je zoznam blokovaných IP adries stránok obsahujúcich nepovolený obsah, ktorých dáta sú presmerované na transparentný HTTP proxy server. Ďalej je celá webová adresa skontrolovaná a v prípade pozitívnej zhody na zakázaný obsah dôjde k odmietnutiu prístupu. V opačnom prípade sa požiadavka spracuje bežným spôsobom. [12]

4.2 Cenzúra v Čínskej ľudovej republike

Rozsah a sofistikovanosť čínskeho programu pre obmedzovanie informácií a slobody prejavu nemá v histórii obdobu. Na rozdiel od Spojených štátov amerických, kde sú sociálne média centralizované, v Číne sú distribuované medzi stovky lokálnych stránok. Významný podiel zodpovednosti za filtrovanie informácii pripadá na poskytovateľov internetového obsahu, ktorým by v opačnom prípade hrozila finančná sankcia, prípadne zákaz činnosti. V snahe vyhovieť vládnym požiadavkám na cenzúru, každá internetová stránka zamestnáva až do 1000 cenzorov. [17] Oproti krajinám Stredného východu a Severnej Afriky, kde je užívateľom doručovaná varovná správa ohľadom filtrovania nepovoleného obsahu, Čína uprednostňuje výrazne nižšiu mieru transparentnosti v súvislosti s internetovou cenzúrou. Nová generácia obyvateľov Číny nepozná internet v takej podobe, v akej ho pozná zbytok sveta. Čínska vláda v posledných rokoch zablokovala, okrem iného, prístup na webové lokality sociálnych sietí a vyhľadávacích nástrojov vrátane stránok Google, Facebook, Twitter alebo Wikipédia. Väčšina mladých obyvateľov nevie o ich existencii či účelu. [29]

Dostupný obsah Wikipédie v Číne bol čiastočne a prerušovane regulovaný od roku 2004, s cieľom zamedziť prístup k nežiadúcim informáciám [23]. V roku 2015 došlo k zablokovaniu čínskej Wikipédie, jednej z najnavštevovanejších webových stránok v krajine [24]. Väčšina ostatných jazykov ostala aj naďalej prístupná, čo malo za následok zvýšený záujem o články v iných jazykoch [23]. V roku 2017 sa objavili informácie o vytváraní vlastnej verzie encyklopédie [26][23] bez účasti verených prispievateľov s cieľom regulovať verejne dostupné informácie [28]. V máji 2019 došlo zo strany čínskej vlády k rozšíreniu zákazu na všetky jazyky Wikipédie v plnom rozsahu [23].

4.2.1 Veľký čínsky firewall

Veľký čínsky firewall (angl. Great Firewall of China, skratka GFW) je jeden z najdôležitejších stavebných kameňov komplexného čínskeho systému filtrovania informácii. Jedná sa o najsofistikovanejší a najrozsiahlejší cenzorský systém na svete. Čínska vláda vie jednoducho odstrániť neželané informácie a potrestať tvorcov, ak sa jedná o obsah pochádzajúci z krajiny. V prípade, že sú informácie mimo krajiny, neostáva autoritám iná možnosť ako pokus o blokovanie alebo regulovanie prístupu. Bez cenzúry umiestnenej na medzinárodných sieťových bránach by bola klasická cenzúra, realizovaná v Číne mimo internetu, absolútne bezpredmetná. To je dôvod, prečo je Čínsky firewall kritickou súčasťou stability Čínskej ľudovej republiky. Veľký čínsky firewall je vládou kontrolovaný útočný systém využívaný na rušenie legitímnej komunikácie, ktorý zasahuje výrazne viac obetí ako kybernetický zločinci. Vďaka špeciálnym technológiám dokáže vláda úspešne zamedziť väčšine čínskych používateľov internetu prístup k väčšine internetových stránok a informácií, ktoré považujú čínske autority za nebezpečné. Avšak, nejedná sa o bezchybný systém, nakoľko odborníci v oblasti internetovej komunikácie, ale aj znalí používatelia internetu dokážu obchádzať

obmedzenia dané Veľkým čínskym firewallom. Obeťami agresívnej čínskej cenzúry nie sú iba čínsky užívatelia, ale tiež užívatelia z iných krajín, ktorí s Čínou nemajú nič spoločné. Okrem obmedzovania dostupných informácii slúži Firewall aj na ochranu ekonomiky, nakoľko blokovanie služieb ako Google, Facebook či Twitter predstavuje ekonomickú výhodu pre domácich poskytovateľov služieb. [4]

Čínsky firewall realizuje cenzúru symetricky. V praxi to znamená, že filtrovanie prebieha v oboch smeroch sieťovej prevádzky, čo umožňuje detekciu a blokovanie nežiadúcich požiadaviek zo strany užívateľa a tiež blokovanie odpovedí s nepovoleným obsahom. [11]

Štúdie [18] ukázali, že blokovanie prístupu k populárnym stránkam zo strany krajiny je dnes značne rozšírené. V prípade Číny sa jedná o ojedinelý typ filtrovania obsahu, tzv. stavová cenzúra. Filtrovaná je len prvá HTTP požiadavka z toku dát. Pre zaistenie efektivity sa uložia informácie o zdrojovej a cieľovej IP adrese, čísle portu a protokolu, aby mohla byť ďalšia komunikácia medzi týmito zariadeniami blokovaná, a to aj v prípade že predtým nebola. [27]

Veľký čínsky firewall je založený na 3 hlavných technológiách, ktoré sa starajú o filtrovanie obsahu: blokovanie IP adries, falšovanie DNS odpovedí a TCP RST pakety[5].

Firewall využíva pre vykonávanie cenzúry čiernu dieru, čo v kontexte počítačových sietí predstavuje miesto, kde dochádza k zahadzovaniu prenášaných dát potichu, bez informovania odosielateľa o neúspešnom doručení paketov. Smerovacie pravidlá, ktoré zabezpečujú smerovanie sieťovej prevádzky cez čierne diery, sú umiestnené do protokolu BGP[6], následkom čoho je možné komunikáciu s blokovanými stránkami odchytávať. Smerovanie do čiernej diery vie blokovať len odchádzajúcu komunikáciu z Číny, pričom nad prichádzajúcimi dátami nemá kontrolu. To ale nie je problém, postačuje filtrovať iba internetové stránky, nakoľko väčšina dnešnej internetovej komunikácie prebieha obojsmerne. Metóda čiernej diery predstavuje iba minimálnu záťaž na brány[7] ISP a nie sú potrebné žiadne ďalšie špeciálne zariadenia. [4]

GFW tiež využíva TCP RST pakety pre ukončenie spojení, v ktorých dochádza k nepovolenej komunikácii. Nie je potrebné poznať komunikujúce strany a ich IP adresy, filtrovanie sa spolieha výhradne na zoznam citlivých slov. V prípade šifrovanej komunikácie pomocou protokolu HTTPS je znemožnené pre GFW vidieť a detekovať kľúčové slová, ale vie zablokovať akúkoľvek komunikáciu naslepo zaslaním TCP RST paketu. Čínska vláda trvá na tom, že nestabilita internetu je spôsobená zlým stavom samotných webových stránok, čomu časť užívateľov skutočne verí. Pre zabezpečenie cenzúry je Čína nútená nahradiť služby ako Google ich vlastnými alternatívami (ako napríklad Baidu). [4]

4.2.2 Vedľajšie škody

Cenzorské technológie spomenuté v tejto kapitole - blokovanie IP adries, falšovanie DNS odpovedí a pakety TCP reset - spôsobujú vedľajšie škody. Keďže GFW nie je bezchybný systém, s cenzúrou prichádzajú aj neúmyselné efekty. [4]

Zasielanie TCP RST paketov je jedna z najúčinnejších metód, ktorá ukončí akékoľvek spojenie, v ktorom bolo detekované zakázané slovo. Vyhľadávanie môže byť výrazne ob-

[5]TCP RST, alebo TCP reset, je neočakávané ukončenie TCP spojenia, ktoré tým nahradí bežné štvorcestné ukončenie komunikácie. Dochádza k uvoľneniu všetkých zdrojov alokovaných pre dané spojenie a k odstráneniu všetkých informácii o spojení [3].

[6]BGP (Border Gateway Protocol) je dynamický smerovací protokol používaný v rámci centrálnych uzlov internetu. Každej podsieti umožňuje informovať zbytok internetu o jej existencii. Zabezpečuje, aby autonómne systémy na internete vedeli o danej podsieti, a tiež ako sa k nej dostať [19].

[7]Brána je aktívne sieťové zariadenie, ktoré prepojuje dve odlišné siete [19].

medzené ak hľadaný výraz v sebe obsahuje nejaké iné citlivé slovo, prípade skratku, ktorá absolútne nesúvisí s významom tohto výrazu. V tom prípade je spojenie zablokované, aj keď sa nejednalo o zakázané informácie. Blokovanie IP adries má výrazne neúmyselné efekty, keďže pod jednou IP adresou sa môže nachádzať viac webových stránok. Odporcovia cenzúry môžu využiť nízku granularitu tohto mechanizmu a cielene spôsobovať vedľajšie škody. [4]

Vedľajšie škody sa nevzťahujú výhradne na Čínu, ale môžu zasahovať aj užívateľov z iných krajín, ktorý pristupujú na stránky mimo Číny. Je to spôsobené smerovaním BGP a iteráciami pri DNS rezolúcii. [4] Technológia používaná Veľkým čínskym firewallom bola použitá aj v Pakistane v roku 2007 ako prostriedok pre blokovanie služby Youtube, čo malo za následok odoprenie služby na celom svete [2]. Modifikovanie DNS záznamov za účelom realizácie cenzúry môže znemožňovať prístup k webovej stránke aj v prípade, že sa užívateľ aj zdrojový server nachádzaju mimo Čínu. Spôsobujú to 3 faktory [4]:

- Niektorí čínsky ISP sú tranzitnými autonómnymi systémami, čo znamená, že si vymieňajú smerovacie informácie BGP s inými autonómnymi systémami, a smerujú získané informácie ďalej. Následne sú cenzorské pravidlá uložené aj mimo Číny.
- Niekoľko koreňových DNS serverov je umiestnených v Číne. Tí ISP, ktorí udržujú zrkadlové kópie koreňových serverov, oznámia svoje prefixy susedným ISP, napríklad v Južnej Kórei či Nemecku. To má za následok, že DNS servery v týchto krajinách budú delegovať svoje požiadavky na koreňové servery v Číne.
- Jeden preklad DNS pozostáva z viacerých iterácií. Ak v ktorejkoľvek časti putuje požiadavka cez ISP v Číne, celá rezolúcia môže byť ovplyvnená Veľkým čínskym firewallom a prístup bude presmerovaný alebo odmietnutý.

Napríklad, v prípade, že užívateľ z Južnej Kórei chce pristúpiť na stránku www.sensitive.de, kde *sensitive* je blokované kľúčové slovo v GFW, uživateľov implicitný DNS server odošle požiadavky na koreňový server ("."), TLD server(".de") a autoritatívny server ("sensitive.de") v snahe získať preklad adresy www.sensitive.de. Ak užívateľov poskytovateľ internetu zvolí pre tento účeľ koreňový server v Číne, alebo bude smerovať požiadavku na TLD server či autoritatívny server cez územie Číny, potom bude prístup blokovaný Čínskym firewallom. [4]

Odhliadnuc od vedľajších účinkoch, GFW je veľmi úspešný v blokovaní veľkej časti politicky neprijateľného obsahu. Nikto nevie, aký je v skutočnosti počet zablokovaných stránok [4], presné špecifikácie Veľkého čínskeho firewallu nie sú známe, nakoľko Čína a jej poskytovatelia internetu nezverejňujú detaily týkajúce sa ich systémov [11] .

4.3 Cenzúra v ďalších krajinách sveta

Aj keď Čína disponuje najrozsiahlejším systémom na cenzurovanie internetového obsahu, ani zďaleka nie je jedinou krajinou. Väčšinou sa jedná o krajiny, ktorých režim nie je zlúčiteľný so slobodou slova ľudu a s neobmedzenými zdrojmi informácií. [12]

V Sýrii sa pre realizáciu cenzúry používa niekoľko cenzorských techník, medzi ktoré patrí aj filtrovanie na základe IP adries v snahe blokovať celé podsiete, filtrovanie domén pre obmedzenie prístupu k vytipovaným webovým stránkam a cenzurovanie kľúčových slov pre blokovanie neželanej komunikácie. Cenzúra vykonávaná pomocou zoznamu kľúčových slov ma výrazný negatívny dopad na vyhľadávacie nástroje, keďže ako vedľajší účinok je

blokovaný aj obsah, ktorý nie je cieľom cenzúry. Služby pre okamžité zasielanie správ (napríklad služba Skype) podliehajú v Sýrii prísnej cenzúre, zatiaľ čo v prípade sociálnych sietí sa cenzúra neprejavuje až tak výrazne [9]. Výskum z roku 2014 [9] ukázal, že používatelia internetu sa aktívne pokúšajú o obchádzanie cenzúry nie len klasickými prostriedkami, akými sú proxy servery, Tor či VPN software, ale tiež službami pre zdieľanie obsahu formou P2P.

Štúdia cenzúry v Južnej Kórei [27] odhalila realizáciu cenzúry pomocou sieťových smerovačov a systému DNS. V prípade, ak by cenzúra na úrovni DNS zasahovala iba jedinú internetovú stránku, tak je zvolená táto možnosť. Druhou alternatívou je viacej internetových stránok, ktoré zdieľajú rovnakú množinu IP adries, z ktorých sú blokované len niektoré. V tomto prípade je cenzúra vykonávaná na úrovni sieťových smerovačov v snahe minimalizovať vedľajšie účinky. Bolo tiež zistené, že stránky blokované pomocou DNS sú blokované aj na úrovni smerovačov. [27]

Pri cenzúre v Bangladéši a Indii dôjde k vypršaniu TCP spojenia, pričom užívateľ nie je oboznámený s dôvodom odmietnutia prístupu. Oproti tomu, v krajinách ako sú Bahrajn či Irán prebieha filtrovanie vrátením stavového kódu pre nepovolenú žiadosť (kód 403) spolu so správou, ktorá informuje užívateľa o vykonávaní cenzúry. V Thajsku je požiadavka GET povolená, následne však cenzorské zariadenie zabezpečuje opakované odoslanie RST paketu, čím dôjde k uzatvoreniu spojenia. V Bahrajne, Iráne aj Thajsku je transparentnosť cenzúry na vysokej úrovni, nakoľko v prípade vyžiadania nepovoleného obsahu je užívateľ patrične informovaný. Varovná stránka obsahuje tiež odkaz, ktorý môže užívateľ použiť na nahlásenie neopodstatneného blokovania, ak je toho názoru, že stránka by nemala podliehať cenzúre. V prípade Iránu je naviac užívateľovi poskytnutý aj zoznam neblokovaných stránok, na ktoré môže pokračovať. Po vypršaní časovača, ktorý sa nachádza na dolnom okraji varovnej stránky, bude jedna stránka vybraná a použitá ako cieľ presmerovania. [27]

Kapitola 5

Obchádzanie cenzúry

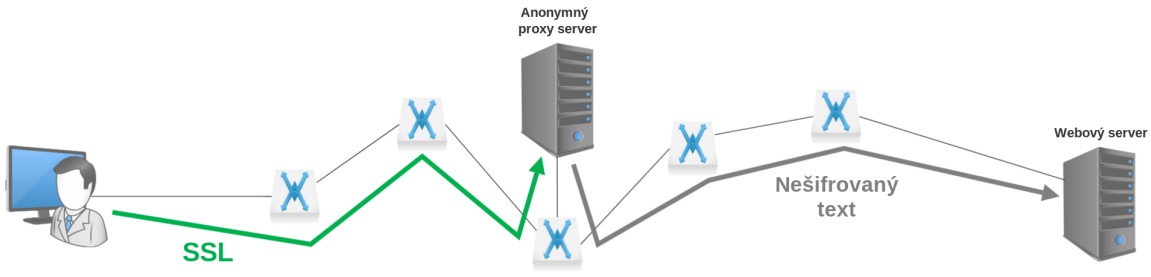
Efektívne blokovanie internetovej prevádzky vyžaduje hĺbkovú kontrolu dátových paketov a neustále sledovanie všetkej internetovej komunikácie.

5.1 Anonymita v prostredí internetu

V prípade, že chce užívateľ navštíviť stránku s citlivým alebo nepovoleným obsahom je, resp. malo by byť, predmetom jeho záujmu:

- Neodhaliť IP adresu webovému serveru.
- Utajiť túto aktivitu pred poskytovateľom internetového pripojenia.
- Zabrániť ISP v sledovaní internetovej komunikácie medzi klientom a serverom.

V prípade, že užívateľ zvolí klasické nešifrované spojenie priamo so serverom, nie je dodržaný ani jeden z vyššie spomínaných bodov. Aj v prípade použitia technológie SSL dochádza k porušeniu prvých dvoch bodov, keďže zdrojová IP adresa je viditeľná serverom v každom odoslanom datagrame a cieľová adresa každého paketu môže byť zachytená ISP. Pre zabezpečenie súkromia a anonymity má užívateľ možnosť využiť kombináciu SSL a dôveryhodných proxy serverov, čo je znázornené na obrázku 5.1.



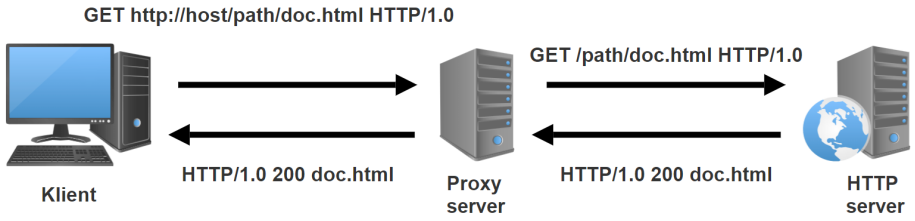
Obr. 5.1: Zabezpečenie anonymity a súkromia na internete prostredníctvom anonymného proxy serveru [19].

Vďaka tomuto prístupu dôjde na začiatku komunikácie k vytvoreniu šifrovaného spojenia s proxy serverom. Týmto kanálom je následne zaslaná HTTP požiadavka na cieľovú stránku. Po prijatí zašifrovanej HTTP požiadavky proxy serverom je správa dešifrovaná a preposlaná v nešifrovanej podobe webovej stránke. Tá zašle odpoveď v čitateľnej podobe na

proxy server, odkiaľ je odpoveď smerovaná užívateľovi cez zabezpečený kanál. Keďže webový server vidí iba IP adresu proxy serveru, užívateľ ostáva pred serverom v anonymite. Rovnako sa stáva anonymný aj obsah komunikácie, nakoľko prebieha na zašifrovanom kanálya tým pádom ISP nemá k prenášaným dátam prístup. Pri tejto metóde vybraný proxy server disponuje všetkými informáciami o komunikácii užívateľa so serverom vrátane IP adries a obsahu komunikácie v otvorenej podobe. Z tohto dôvodu účinnosť spomínaného prístupu závisí na dôveryhodnosti použitého proxy serveru. Ďalší, robustnejší spôsob využíva TOR, ktorý smeruje pakety cez sériu proxy server.

5.2 Proxy

Hlavným využitím proxy serverov je obídenie obmedzení firewallu za účelom slobodného prístupu na internet. Proxy je špeciálny HTTP server ktorý je väčšinou umiestnený na úrovni firewallu. Proxy server čaká na žiadosť zvnútra firewallu, po prijatí ju preposiela cieľovému serveru a následne doručí odpoveď prichádzajúcu zo strany serveru danému užívateľovi. [21]

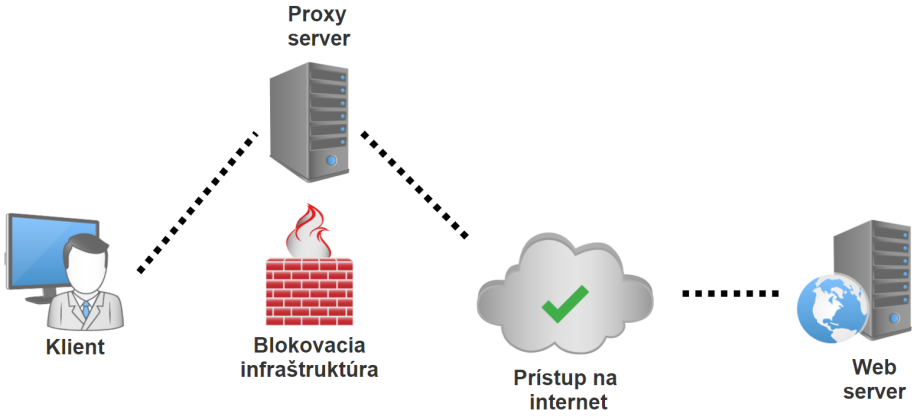


Obr. 5.2: Mechanizmus proxy serveru. Klient odošle požiadavku na proxy server pri čom špecifikuje celú adresu URL. Proxy sa následne pripojí k serveru a vyžiada si dokument pomocou relatívnej cesty.

Proxy servery si vedia dočasne uchovávať webové stránky v pamäti cache. Ak viacerí užívatelia jedného poskytovateľa internetového pripojenia (angl. ISP) majú záujem o rovnakú webovú stránku, vďaka proxy serverom sa na stránku pristupuje len prvý krát a ďalšie žiadosti sú obslúžené pomocou pamäte cache, odkiaľ je načítaný a vrátený obsah stránky. Do tejto pamäte je stránka uložená pri prvej žiadosti o prístup. Z pohľadu užívateľa je to výhodné vďaka rýchlejšiemu prístupu na stránku, keďže nie je nútený pripájať sa mimo sieť svojho ISP. Pre poskytovateľa internetového pripojenia výhoda spočíva v ušetrení prenosového pásma a tým pádom aj finančných prostriedkov. [12] Ďalšou výhodou je možnosť pristupovať k stránkam, ktoré sú uložené na serveroch ktoré nie sú momentálne aktívne. [4]

Obchádzanie cenzúry, ktorá blokuje priamy prístup k zahraničným webovým lokalitám je pomerne jednoduché. Postačuje pristupovať k želanej lokalite v mene proxy serveru, ktorý sídli mimo krajiny. Pre realizáciu tejto metódy nesmie byť použitý proxy server blokovaný. Nevýhoda tohto prístupu spočíva v nutnosti používanej aplikácie podporovať proxy. Taktiež je nevyhnutné, aby komunikačný kanál spojujúci užívateľa a proxy server nebol blokovaný. Pre účel vyššej úrovne súkromia existujú protokoly využívané proxy servermi, ktoré šifrujú prenášané dáta. [3]

Obrázok 5.3 znázorňuje prístup k nepovolenému obsahu obchádzaním firewallu. V tomto príklade vystupuje proxy server z pohľadu klienta ako webový server ku ktorému sa užívateľ pokúša pripojiť a zobrazuje požadované dáta. Aj napriek tomu, že väčšina proxy serverov komunikuje pomocou špeciálnych protokolov vyvinutých priamo na tento účel, určite proxy sú prístupné aj pri použití protokolu HTTPS. [3]



Obr. 5.3: Obchádzanie firewallu s využitím proxy serveru.

5.3 Sieťové tunelovanie

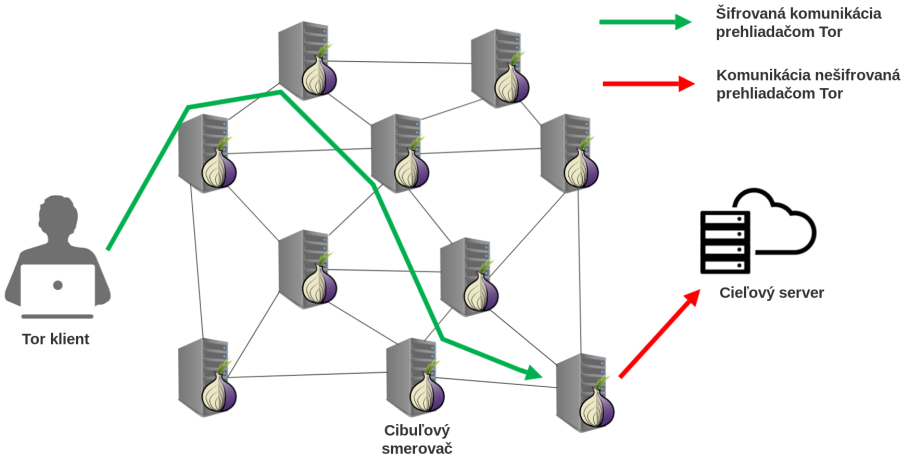
Software umožňujúci tunelovanie slúži na vytvorenie šifrovaného tunelu medzi odosielateľom a adresátom a znemožňuje tak filtrovacím technológiám prístup k prenášaným dátam. Po vytvorení tunelu sú všetky požiadavky prenášané týmto tunelom k cieľovému zariadeniu a následne na internet. Oproti proxy sa tunelovanie odlišuje v tom, že sa z pohľadu operačného systému jedná o samostatné pripojenie, čo odoberá povinnosť špecifických nastavení v aplikácii. VPN tunely sú neustále šifrované a tým pádom nehrozí odchytávanie komunikácie. [7]

5.4 Cibuľové smerovanie

Cibuľové smerovanie (angl. onion routing) využíva šifrovanie vereným kľúčom. Súkromný kľúč, ktorý je uchovávaný majiteľom na bezpečnom mieste, je určený na dešifrovanie. Odpovedajúci verejný kľúč môže byť distribuovaný kamkoľvek na svete a slúži pre zašifrovanie správ určených pre držiteľa súkromného kľúču. Jediná cesta, ako správu dekódovať je práve pomocou príslušného súkromného kľúča. Pri tomto mechanizme je internetová prevádzka šifrovaná zdieľaným súkromným kľúčom viacerých serverov cez ktoré sú dáta smerované. Keď cibuľový smerovač obdrží dáta, rozšifruje ich pomocou vlastného súkromného kľúču a následne odošle na ďalší server v zašifrovanej forme pomocou verejného kľúču nového cieľového serveru. Tento proces sa opakovane vykonáva do momentu, kedy dáta dorazia do cieľového uzlu, odkiaľ sú ďalej prenášané v nezašifrovanej podobe. [6]

Popísaný spôsob smerovania využíva webový prehliadač Tor (z angl. The Onion Router). Ponúka anonymitu aplikáciám založeným na TCP komunikácii (ako napríklad prehliadanie webu). Pre spustenie prehliadača nie sú potrebné žiadne zásahy do nastavení zariadenia ani povolenia jadra systému. Užívateľ má k dispozícii vďaka prehliadaču okruh (angl. circuit) zložený z niekoľkých cibuľových smerovačov, ktoré vždy poznajú predchádzajúci uzol, následujúci uzol a okrem toho žiadny ďalší v danom okruhu. [13] Schému internetovej komunikácie pomocou nástroja Tor je možné vidieť na obrázku 5.4

Každý cibuľový smerovač obsahuje užívateľský proces bez špeciálnych oprávnení a udržuje TLS spojenie s ostatnými smerovačmi v Tor sieti. Každý užívateľ má spustený proces cibuľového proxy (angl. onion proxy) ktorý vytvára sieťové okruhy a udržuje spojenia z užívateľských aplikácií. Tiež má za úlohu prijímať TCP dáta a smerovať ich cez vytvorené



Obr. 5.4: Schéma sieťovej komunikácie s využitím webového prehliadača Tor. Prenášané dáta sú smerované medzi sériou cibuľových smerovačov v šifrovanej podobe, až kým nedorazia do cieľového uzlu, odkiaľ sú v nešifrovanej podobe posielané na internet [13].

okruhy. Po tom, čo dáta dorazia do cieľového uzlu, sú ďalej prenášane na internet bez účasti cibuľového smerovania. [13]

5.5 Boj proti obchádzaniu cenzúry

Čínska vláda okrem realizácie cenzúry usiluje aj o minimalizovanie obchádzania obmedzení daných Veľkým čínskym firewallom. S rastúcou sofistikovanosťou nástrojov na obchádzanie filtrovania sa vylepšujú aj cenzorské nástroje. V snahe zabrániť obchádzaniu GFW sa realizuje metóda aktívneho snímania (angl. active probing), ktorá má dve hlavné úlohy. Pasívne monitoruje sieť a hľadá akúkoľvek podozrivú komunikáciu a aktívne sleduje dané servery, z ktorých blokuje práve tie, ktoré poskytujú služby pre obchádzanie cenzúry. Keďže nie je možné sledovať sieťovú prevádzku využívajúcu zabezpečený protokol HTTPS, nie je možné určiť, či sa jedná o prenos dát súvisiacich s vyhýbaním sa cenzúre alebo nie. Metóda aktívneho snímania vystupuje ako užívateľ a ustanovuje vlastné spojenia s podozrivým serverom. Ak server vykazuje chovanie charakteristické pre boj proti cenzúre (ako napríklad využívanie špeciálnych protokolov), cenzor vykoná príslušné opatrenie, akým je zaradenie IP adresy serveru do zoznamu blokovaných adries. [15]

Kapitola 6

Špecifikácia a návrh aplikácie

Cieľom je navrhnúť a implementovať nástroj `analyze.py` pre analýzu aktuálneho stavu cenzúry v ľubovolnej krajine. Program bude implementovaný v programovacom jazyku python 3 a cieleným operačným systémom bude Linux. S jazykom Python mám najrozsiahlejšie skúsenosti a dôvodom je tiež veľké množstvo modulov pre sieťovú komunikáciu a spracovávanie dát. Vybraný operačný systém je priateľský k programátorom a považujem ho za najlepšiu voľbu pre túto prácu. Program bude operovať na úrovni terminálu (príkazový riadok v Linuxu) bez grafického rozhrania.

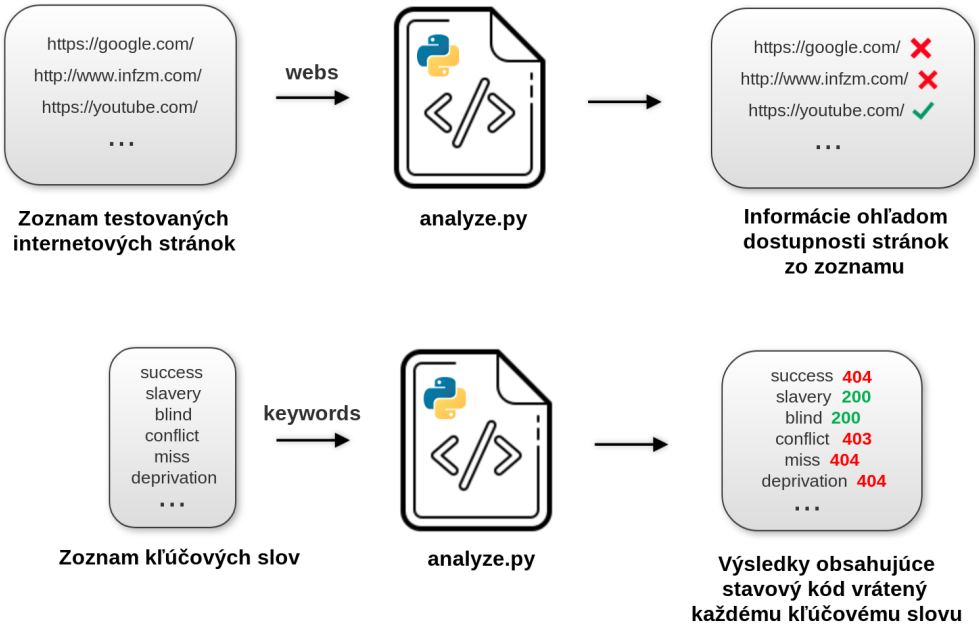
Vstupom bude zoznam potencionálne blokovaných stránok, prípade zoznam kľúčových slov, ktoré budú predmetom analýzy. Výstupom budú informácie o tom, či bol daný vstup blokovaný, prípadne na akej úrovni. Nástroj, vykonávajúci analýzu bude mať 2 základne funkcionality:

- Testovanie dostupnosti webových stránok. Ako argument programu bude vyžadovaný súbor s potencionálne blokovanými stránkami v podobe adries URL, prípadne ako IP adries. Jednotlivé ciele budú vo vstupnom súbore od seba oddelené znakom konca riadku. Pri spustení testu sa program pokúsi o prístup na danú lokalitu prostredníctvom vrstvy L7 (DNS preklad), vrstvy L4 (TCP spojenie), vrstvy L3 (odozva na PING) a ďalej o získanie dokumentu pomocou metódy GET. Získané výsledky budú uložené vo výstupnom súbore s časovým razítkom.
- Analýza filtrovania kľúčových slov. Vstupom programu bude zoznam slov, z ktorých každé bude použité v url adrese, na ktorú sa program pokúsi pristúpiť, a rovnako aj v obsahu získaného dokumentu. Predmetom záujmu je stavový kód, ktorý sa po požiadavke vráti.

Pri spustení bude nutné poskytnúť jeden z dvoch prepínačov podľa toho, ktorá funkcionalita je požadovaná: *webs* pre prvú z uvedených funkcionalít, alebo *keywords* pre druhú. Schému fungovania nástroja ukazuje obrázok 6.1.

6.1 Analýza cenzúry webových stránok

Ak bude program spustený s prepínačom *webs*, je nutné zadať cestu k zoznamu stránok. Postupne je testovaná každá stránka na rôznych úrovniach. Ak sa jedná o adresu URL (a nie IP adresu), ako prvý sa uskutoční pokus o DNS preklad. Ak rezolúcia nedopadne úspešne, cyklus sa ukončí a začne ďalšia iterácia pre následujúcu stránku. V prípade úspechu sa



Obr. 6.1: Princíp fungovania nástroja analyze.py. Pri prepínači *webs* sa vyžaduje súbor so zoznamom internetových stránok, ktoré budú predmetom testovania. V prípade prepínača *keywords* sa na vstupe očakáva zoznam kľúčových slov. V oboch prípadoch je výsledkom súbor s informáciami o cenzúre vstupných elementov.

pokračuj nadviazaním TCP spojenia pomocou 3-cestného ustanovenia spojenia (angl. 3-way handshake). Ďalej program otestuje odozvu na ping a v prípade neúspechu pomocou nástroja traceroute získa IP adresu uzlu, odkiaľ sa vrátila posledná odpoveď. Následne je zaslaná žiadosť o koreňový dokument uložený na adrese `"/"` a v prípade, že URL špecifikuje aj konkrétne umiestnenie dokumentu na serveri, program sa pokúsi pristúpiť aj k nemu. Celý postup je znázornený na obrázku 6.2.

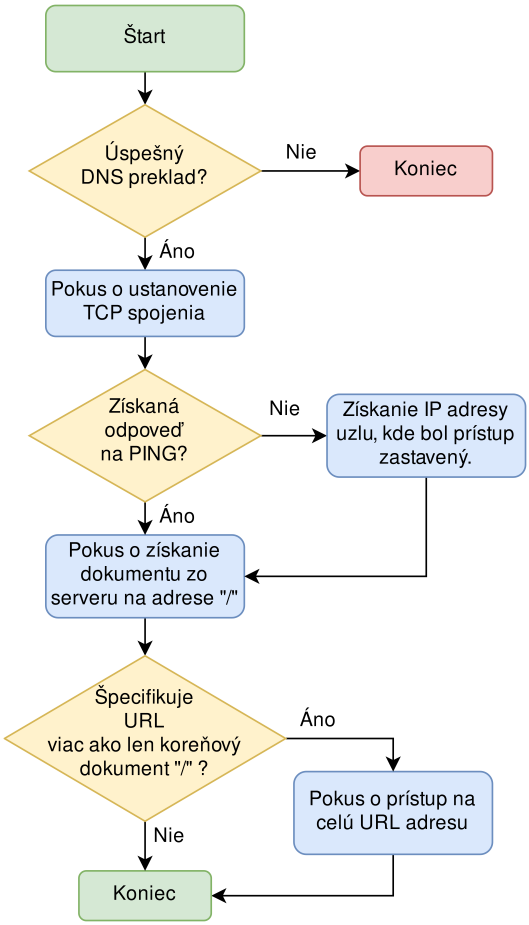
Ak to nie je nevyhnutné, analýza dostupnosti stránky nebude ukončená z dôvodu neúspechu v nejakom kroku. K ukončeniu dochádza len v prípade neúspešnej rezolúcie adresy (a samozrejme v prípade neplatného vstupu). Vďaka tomu bude možné získať čo najviac informácii o dostupnosti danej webovej stránky bez preskočenia ďalších testov. Mohlo by ísť o prípad, kedy prístup k celej URL adrese skončí úspechom aj napriek tomu, že nejaký predošlý krok nedopadol úspešne.

6.1.1 Výstupný formát

Pre zaznamenávanie výsledkov bude na začiatku behu programu pripravený prázdny slovník[1], do ktorého sú postupne vkladané záznamy o testovaných stránkach. Pre každú stránku bude vytvorený kľúč zodpovedajúci danej adrese, ktorého hodnota bude vnorený slovník obsahujúci vždy 7 položiek:

- TYPE - pre uloženie informácie, či je vstupný reťazec IP adresa alebo URL adresa, prípadne len doménové meno.

[1]Slovník je implementácia dátovej štruktúry v jazyku Python. Je to kolekcia dvojíc pozostávajúcich z kľúču a príslušnej hodnoty [25].



Obr. 6.2: Postup získavania informácii o internetovej stránke. Schéma predstavuje jednu iteráciu v z hlavného cyklu pre analýzu dostupnosti internetových stránok.

- DNS - pre zaznamenanie IP adresy z DNS prekladu.
- SYNADDR - pre uloženie informácie, či bola obdržaná odpoveď na SYN paket. Ak áno, uloží sa IP adresa komunikújúcej strany.
- PING - informácia ohľadom úspešnosti pingu.
- STOPPED - nadobudne hodnotu IP adresy, na ktorej sa zastavila komunikácia smerom k cieľovému serveru. Toto pole je využívané iba v prípade, ak sa odpoveď z nástroja ping nevráti.
- REQUEST - informácia ohľadom úspešnosti prístupu ku koreňovému dokumentu "/".
- EXTREQUEST - informácia ohľadom úspešnosti prístupu k celej URL adrese. Ak nie je špecifikovaný nijaký dokument (ako v prípade www.example.com/), toto pole nebude použité.

Program bude testovanie jednotlivých položiek zo vstupného zoznamu riadiť v hlavnom cykle. Na začiatku iterácie je ku každému zo spomínaných 7 kľúčov priradená hodnota "--", ktorá znamená, že daný aspekt zatiaľ nebol testovaný. Táto hodnota môže v určitých

prípadoch ostať nezmenená až do konca. V prípade, že test spojenia v akomkoľvek kroku nedopadne úspešne, je k príslušnému kľúču priradená hodnota XX. Ak dôjde k úspešnému testu, ku kľúčom DNS a SYNADDR bude priradená hodnota príslušnej IP adresy. Pre PING, REQUEST a EXTREQUEST platí, že je priradená hodnota OK. Špeciálnym prípadom je STOPPED, kde sa môže vyskytnúť jedine hodnota "--", prípadne IP adresa. Ak priamy DNS preklad stránky a ping nevrátia rovnakú IP adresu cieľového serveru, pole DNS bude obsahovať získané adresy v dvojprvkovom poli.

Slovník, obsahujúci informácie o všetkých testovaných stránkach bude uložený do aktuálneho adresára vo formáte JSON. Názov súboru bude obsahovať časové razítko z času realizácie testu.

6.2 Analýza cenzúry kľúčových slov

Pre overenie cenzúry kľúčových slov bude program očakávať 2 vstupné argumenty: súbor obsahujúci zoznam kľúčových slov a URL adresu. Zoznam slov bude prechádzaný v cykle a s každou iteráciou je nové slovo dosadené na koniec zadanej URL adresy. Po spustení sa najskôr overí dostupnosť zadanej stránky a v prípade neúspechu je program ukončený. Ak bude stránka dostupná, začne sa cyklicky prechádzať zoznam kľúčových slov. Pred začatím cyklu je vytvorený prázdny slovník, kam sa ukladá každé otestované kľúčové slovo spolu s návratovou hodnotu. Ten bude na konci behu programu uložený do výstupného súboru vo formáte JSON s časovým razítkom v názve súboru.

Pre dosiahnutie čo najvyššej presnosti je vhodné, aby výsledne URL adresy, ku ktorým bude program pristupovať, naozaj existovali. Pre tento účel bude implementovaný pomocný skript `generate.py`, ktorý na základe zoznamu kľúčových slov na vstupe vytvorí html dokumenty v aktuálnom pracovnom adresári. Pre každé kľúčové slovo bude vytvorený jeden súbor obsahujúci základnú kostru html spolu s daným slovom. Základná kostra HTML dokumentu obsahujúca značky `html`, `body`, `head` a `title`, je rozdelená do 3 premenných. Dve miesta, kde sa kostra rozdelila medzi 3 premenné, sú miesta pre vloženie kľúčového slova. Slovo bude umiestnené do stránky ako nadpis, a tiež ako názov samotnej stránky pomocou značky `title`. Skript `generate.py` očakáva existenciu súboru keywords.txt, ktorý je považovaný za zdroj kľúčových slov. Ak sa nebude nachádzať v rovnakom adresári, je užívateľ informovaný o neúspechu varovnou hláškou a program je ukončený. Pomocou zoznamu kľúčových slov bude možné v prípade potreby tiež vyčistiť aktuálny adresár od vygenerovaných html dokumentov. Tento skript je vhodné spustiť na serveri, kam bude testovací program pristupovať. Príklad vygenerovanej stránky je možné vidieť na obrázku 6.3. Zdrojový kód funkcie, pomocou ktorej dochádza ku generovaniu HTML dokumentov, sa nachádza na obrázku 6.4.

6.3 Prostriedky pre realizáciu testov

Aby bolo možné vykonať analýzu cenzúry v Číne, je potrebné splniť následujúce body:

- Získať prístup na čínsky internet. Proxy servery zverejnené na rôznych internetových stránkach sa ukázali byť vysoko nespoľahlivé. Cieľ je získať prístup k stabilnému uzlu, ktorý bude pripojený k čínskemu poskytovateľovi internetu.

```
<!DOCTYPE html>
<html>
  <head>
    <title>authority</title>
  </head>
  <body>
    <h1>Page about authority </h1>
  </body>
</html>
```

Obr. 6.3: Zdrojový kód internetovej stránky, ktorá bola vygenerovaná na základe kľúčového slova *authority*.

```
def add(keywords):
    for w in keywords:
        with open(w + ".html", "w") as f:
            f.write(part_1 + w + part_2 + w + part_3)
        os.chmod(w + ".html", 0o704)
```

Obr. 6.4: Funkcia, ktorá sa stará o generovanie HTML dokumentov na základe kľúčového slova. Pomocou argumentu je funkcii predaný zoznam kľúčových slov v liste, cez ktoré cyklicky prechádza a pre každé slovo vytvára samostatný dokument. Následne dochádza k zmene oprávnení pre prístup k súboru, aby bolo umožnený prístup pomocou protokolu HTTP (a nedochádzalo k chybe 403 - nelegálna požiadavka).

- Mať k dispozícii službu pre lokalizovanie internetových stránok a IP adries. Aj keď pre samotný beh testovacieho programu to nie je podmienkou, pri vyhodnocovaní získaných dát poskytne lokalizačná služba viac dôležitých informácií.
- Mať prístup na server, kde prebehne generovanie html dokumentov na základe zoznamu kľúčových slov. Bolo by možné tento krok vynechať a stavový kód 404 považovať za negatívny test na cenzúru konkrétneho kľúčového slova, ale v tom prípade hrozí výrazne skreslenie výsledkov. Je nutné predpokladať, že kód 404 môže byť vrátený aj z dôvodu prítomnej cenzúry.

V následujúcom texte je bližšie vysvetlený zvolený prístup k vyššie uvedeným bodom.

6.3.1 Použité technológie

Pre analýzu cenzúry v Čínskej ľudovej republike je nutné mať prístup na internet v danej krajine. Pre tento účel som zvolil platformu Planetlab[2], ktorá poskytovala prístup na internet v rôznych krajinách sveta. Táto technológia bola zvolená ako jediná platforma poskytujúca bezplatný prístup do Číny ako jediná dostupná akademická sieť. Ďalšie možnosti boli platené so záväzkami na niekoľko mesiacov, prípadne rok. Služba Planetlab vznikla v roku 2002 a postupne sa rozširovala o ďalšie uzly. [1] Počas písania práce, 1.mája 2020, došlo k ukončeniu projektu Planetlab a odstaveniu všetkých serverov, ktoré Planetlab tvorili. Táto služba postupom času starla a uzly odumierali. Veľká časť uzlov bola označená ako nefunkčná, prípadne vypnutá. Dokonca tiež mnoho uzlov, ktoré boli označené ako aktívne operujúce, v skutočnosti nefungovali. [22]

Pre prístup na uzly v Planetlabu bolo nutné vytvoriť si účet na oficiálnej internetovej stránke a nahrať verejný SSH kľúč do svojho profilu. Tento kľúč bol v priebehu niekoľkých hodín až dní distribuovaný na všetky zvolené uzly. Následne bolo možné prihlásiť sa do uzlu pomocou príslušného súkromného SSH kľúča. Pre zlý stav Planetlabu bolo veľké množstvo uzlov v Číne neaktívnych. Do väčšiny z fungujúcich uzlov sa ani po týždňoch nedostal verejný kľúč, vďaka ktorému by som na uzly vedel pristupovať. Čo sa týka Číny, skutočne fungujúce a spoľahlivé uzly boli dva, obidva na rovnakom servery, umiestnenom v Pekingu.

[2] https://www.planet-lab.org

Pre analyzovanie dostupnosti webových lokalít bol vytvorený zoznam potencionálne blokovaných stránok, ktoré budú predmetom testovania. Zdrojom je projekt na GitHube[3] [20] obsahujúci možné cenzurované stránky z viacerých krajín sveta. Ide o stránky s politicky orientovaným obsahom, citlivými sociálnymi témami, sexuálnou tematikou, drogami a s ďalšími kontroverznými informáciami.

Na testovanie cenzúry kľúčových slov som použil školský server Fakulty informačných technológií VUT, kde som vo svojom priečinku vygeneroval html dokumenty zodpovedajúce zoznamu slov vďaka skriptu, ktorý je opísaný v kapitole 6.2. Vstupom pre `analyze.py`, ktorý bude spustený na čínskom uzle, bude url adresa `www.stud.fit.vutbr.cz/~xrajec01/`. Odkazuje na koreňový dokument môjho priečinka na serveri. Na koniec adresy budú postupne pridávané jednotlivé slová a program bude testovať, či Veľký čínsky firewall blokuje komunikáciu následkom prítomnosti týchto slov v požiadavkách.

Pre overenie správnej funkcionality bol program `analyze.py` spustený najskôr z územia Slovenskej republiky. Všetky pokusy o prístup k existujúcim dokumentom skončili podľa očakávaní, teda stavovým kódom 200, a neexistujúce dokumenty viedli na kód 404.

6.3.2 Výber vhodného lokalizačného nástroja

Pre účely lokalizácie webových stránok a IP adries bolo nutné vybrať službu s aplikačným programovým rozhraním, ktorá dokáže určiť mesto a štát, odkiaľ adresa či webová stránka pochádza. Kritériami boli bezplatná verzia, dostatočne veľký mesačný limit pre realizovanie požiadaviek a čo najlepšia presnosť. Po prieskume na internete som sa rozhodol pre 5 služieb, ktoré spĺňali prvé dve kritéria a z ktorých bude vybraná najlepšia možnosť pre túto prácu. Sú nimi Signals od spoločnosti Auth0[4], geoPlugin[5], ipgeolocation[6], Big Data Cloud API[7] a ipinfo[8].

Výber prebehol následovne. V pomocnom skripte je vytvorený krátky list náhodných webových stránok, ktoré budú lokalizované týmito službami. Z množstva údajov, ktoré služby vracajú, som vybral dve najdôležitejšie informácie s ohľadom na ciel tejto práce: štát a konkrétne mesto. Výsledné lokality som porovnával aj na stránke IP location[9] a závery boli pomerne jednoznačné. Signals a Geoplugin často zlyhávali v určovaní mesta, čo predstavuje značnú nevýhodu. Ipgeolocation ponúka výrazne nepresné výsledky, ktoré by mohli zkreslovať analyzovanie získaných dát v Číne. Big Data poskytovalo vcelku uspokojivé odpovede, no jednoznačne najlepšou službou bolo ipinfo. Pre programovací jazyk Python je tiež dostupný modul, ktorý slúži pre komunikáciu s touto službou. S ohľadom na všetky uvedené informácie bude pre prácu použitý nástroj ipinfo. Ukážku z testovania je možné vidieť na obrázku 6.5.

[3] https://github.com/
[4] https://auth0.com/signals/ip
[5] https://www.geoplugin.com/
[6] https://ipgeolocation.io/
[7] https://www.bigdatacloud.com/
[8] https://ipinfo.io/
[9] https://www.iplocation.net/

Obr. 6.5: Realizácia testov lokalizačných služieb.

Kapitola 7

Implementácia

Programovací jazyk Python poskytuje množstvo užitočných modulov, ktoré budú využité pri implementácii nástroja `analyze.py`. Napríklad modul `re` pre prácu s regulárnymi výrazmi, `json` pre prácu s dátami vo formáte JSON, `socket` pre sieťovú komunikáciu a mnoho ďalších, ktoré budú spomenuté v tejto kapitole. Nástroj je vyvíjaný pre operačný systém Linux, implementácia prebieha na distribúcii Kali[1] a použitý bude na čínskom serveri s distribúciou CentOS[2] a tiež na lokálnom počítači s Linux Kali.

Argumenty programu sú spracované na začiatku funkcie `main()` pomocou importovaného modulu `argparse`. Možnosti parametrov a prepínačov sú opísané ďalej v tejto kapitole.

Nástroj `analyze.py` potrebuje 3 pomocné triedy, ktoré musia byť umiestnené v priečinku *classes*, ich funkcia je vysvetlená v následujúcom texte:

- `Input` - slúži pre spracovanie jednotlivých webových stránok zo zoznamu. Ponúka metódy pre zistenie, či sa jedná o IP adresu, o doménové meno alebo o celú URL adresu a tiež obsahuje metódu, ktorá vráti iba doménové meno.
- `Uri` - poskytuje metódy pre internetové stránky, ktoré neboli zadané vo forme IP adresy alebo doménového mena. Obsahuje metódy pre odstránenie protokolu z adresy (http://), pre DNS preklad stránky a pre získanie TLD domény.
- `Trace` - trieda bude použitá len ak program neobdrží odpoveď z pingu. Pomocou nástroja traceroute s prepínačom -I zistí, z ktorého uzlu prišla posledná odpoveď.

Popísané metódy sú pre beh programu nevyhnutné, aby bolo možné meniť formu zápisu webovej stránky (odstrániť "https://", pridať "www", získať doménové meno, ...) podľa toho, aký modul a akú metódu program práve používa. Napríklad, pre získanie IP adresy na základe doménového mena pri použití metódy `gethostbyname()` z modulu `socket`, vstupná adresa nemôže byť zadaná s protokolom (http://www.google.com), keďže je vyžadované len doménové meno a došlo by tak k chybe.

V `analyze.py` je implementovaná funkcia `ping()`, ktorá má za úlohu otestovať odozvu serveru pomocou nástroja ping. To je realizované pomocou funkcie `Popen()` z modulu `subprocess` sa stará o vykonanie príkazu ping a uloženie odpovede do premennej. Tá je potom prehľadávaná pomocou regulárnych výrazov (modul `re`) na zhodu s reťazcom

[1] https://www.kali.org/
[2] https://www.centos.org/

100% packet loss. Ak sa daný reťazec v odpovedi nachádza, funkcia vráti hodnotu `False`, v opačnom prípade `True`.

Pre zisťovanie cenzúry kľúčových slov (argument *keywords*) je implementovaná jediná funkcia `keywords_censorship_test()`, ktorá po overení dostupnosti zadanej URL adresy pomocou funkcie `ping()` cyklicky prejde zoznam kľúčových slov. V každej iterácii dané slovo pripojí na koniec URL adresy a pomocou funkcie `get()` z modulu `requests` pošle GET požiadavku na server.

Čo sa týka druhej funkcionality programu, konkrétne testovanie dostupnosti webových stránok zo zoznamu (argument *webs*), je z `main()` zavolaná funkcia `webs_reachability_test()`. V jej tele sa nachádza cyklus, ktorý iteruje cez všetky internetové stránky a podľa toho, či sa jedná o doménové meno, IP adresu alebo URL adresu, vyberie jednu z troch ďalších funkcií, ktorá sa vykoná. To sa zisťuje metódami `is_domain()`, `is_URI()` a `is_IP()` z triedy `Input`. Popis spomínaných 3 cieľových funkcií je uvedený v následujúcom texte a postup ich volania je znázornený na obrázku 7.1. Dá sa povedať, že každá z funkcií je len rozšírenie inej funkcie z tejto trojice.

- `IP_process()` - pomocou funkcie `ping()` je zistená odozva na ping a v prípade neúspechu je s využitím triedy `Trace` zistená IP adresa uzlu, odkiaľ bola obdržaná posledná odpoveď. Ak ping dopadol úspešne, pomocou funkcie `requests.get()` je poslaná žiadosť o obsah stránky.
- `domain_process()` - dôjde k prekladu doménového mena na IP adresu pomocou funkcie `socket.gethostbyname()`. Následne je zavolaná funkcia `IP_process()`.
- `uri_process()` - na začiatku je zavolaná funkcia `domain_process()` a po jej ukončení sa program pokúsi pristúpiť k celej URL adrese.

Okrem súboru s výsledkami, nástroj `analyze.py` ponúka aj informácie o každej testovanej stránke počas behu na štandardnom výstupe. Ide o neformálny výstup poskytujúci približné informácie ohľadom cenzúry danej internetovej stránky. Formálny zápis výsledkov do súboru vo formátu JSON sa realizuje na konci každej iterácie, inými slovami, po otestovaní každej jednej stránky. Vďaka tomu sa pri prípadnej chybe behu programu nestratia všetky nazbierané dáta, ale budú zapísané v príslušnom súbore. Informácie budú chýbať len k stránkam, ktoré sa otestovať nestihli. Príklad obsahu štandardného výstupu počas behu programu je vidieť na obrázkoch 7.2 pre prepínač *webs* a 7.3 pre prepínač *keywords*.

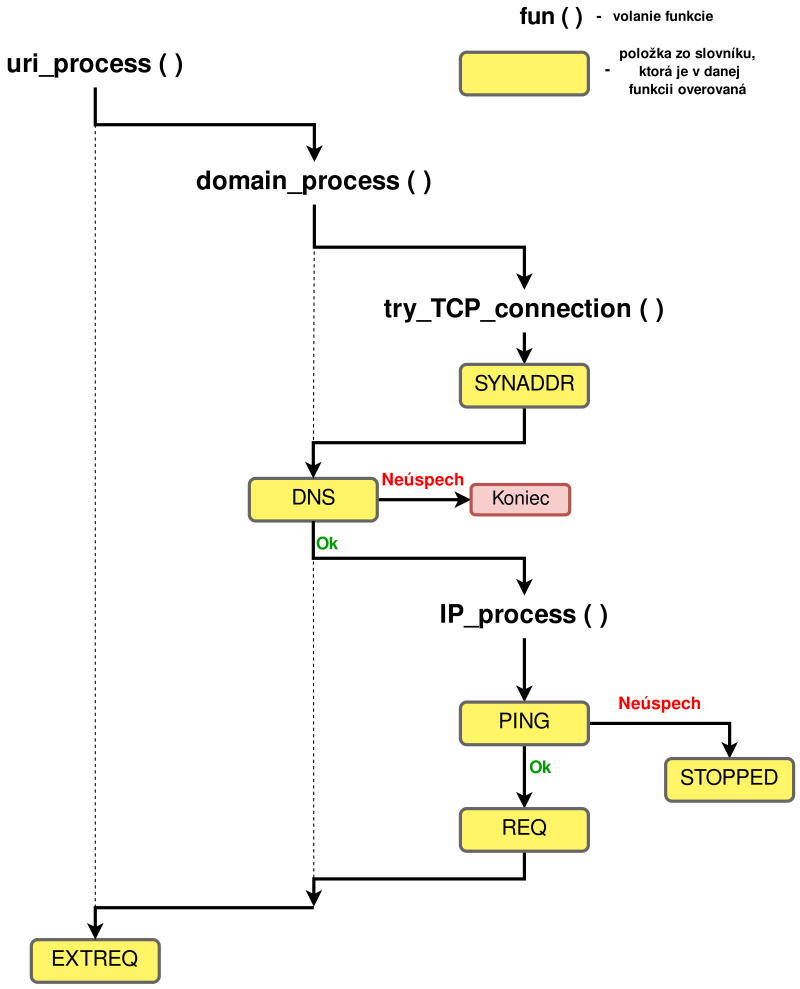
Názov výstupného súboru nie je možné ľubovolne zvoliť, ten vždy závisí na aktuálnom serverovom dátume a čase a má všeobecný tvar `xxx--dd-mm-rrrr--hh-mm.json`. Význam jednotlivých častí popisuje následujúci zoznam:

- `xxx` - určuje, ktorá funkcionalita programu bola zvolená, *output* ak bola analyzovaná dostupnosť webových stránok, alebo *keywords* v prípade, že program testoval cenzúru kľúčových slov.
- `dd-mm-rrrr` - dátum, ktorý bol na serveri v momente spustenia programu.
- `hh-mm` - konkrétny čas, ktorý bol na serveri v momente spustenia programu.

Príkladom možných názvov sú:

- output--02-03-2020--18-27.json
- keywords--24-04-2020--10-09.json

Ukážky obsahu výstupných súborov je možné vidieť na obrázkoch 7.4 a 7.5.



Obr. 7.1: Životný cyklus jednej iterácie. Schéma znázorňuje poradie, v ktorom sú funkcie volané a miesta, odkiaľ sú volané. Funkcia, ktorou sa začína, je daná typom zápisu internetovej stránky, ktorá sa práve spracováva. Pre URL adresy sa začína prvou funkciou, pre doménové meno príslušnou funkciou domain_process() a v prípade IP adresy sa začína až pri IP_process().

Argumenty programu

Chovanie programu je možné upraviť podľa potreby pomocou argumentov. Povinnosťou zostáva zvolenie jednej z dvoch základných funkcionalít pomocou pozičného argumentu *webs* alebo *keywords*. Okrem nižšie uvedených nepovinných argumentov je ďalej nutné dodať ako pozičný argument cestu k súboru, ktorý obsahuje prvky pre analýzu, teda zoznam internetových stránok alebo zoznam kľúčových slov.

Ak priebežné informácie na štandardnom výstupe nie sú želané, pomocou prepínaču -s užívateľ zabráni tomu, aby ho program vypisoval. Tento prepínač nijakým spôsobom neovplyvňuje obsah ani formu výsledného súboru.

```
1/3 -> https://www.fit.vut.cz/research/esf-projects/
1 - OK - DNS resolution 147.229.9.26(www.fit.vut.cz)
2 - OK - ping
3 - OK - "/" request content
4 - OK - "/research/esf-projects" request content

2/3 -> yahoo.com
1 - OK - DNS resolution 87.248.98.7(www.yahoo.com)
2 - OK - ping
3 - OK - "/" request content

3/3 -> google
** wrong input **

Done in: 00:00:03 (hh:mm:ss)
```

Obr. 7.2: Ukážka štandardného výstupu pri vykonávaní testu analýzy cenzúry s prepínačom *webs*

```
205)  200   http://www.stud.fit.vutbr.cz/~xrajec01/europe
206)  200   http://www.stud.fit.vutbr.cz/~xrajec01/revolution
207)  200   http://www.stud.fit.vutbr.cz/~xrajec01/cronyism
208)  200   http://www.stud.fit.vutbr.cz/~xrajec01/escape
209)  200   http://www.stud.fit.vutbr.cz/~xrajec01/aide
210)  200   http://www.stud.fit.vutbr.cz/~xrajec01/fakes
211)  200   http://www.stud.fit.vutbr.cz/~xrajec01/light-true
212)  200   http://www.stud.fit.vutbr.cz/~xrajec01/disney
213)  200   http://www.stud.fit.vutbr.cz/~xrajec01/man
214)  200   http://www.stud.fit.vutbr.cz/~xrajec01/armed
215)  200   http://www.stud.fit.vutbr.cz/~xrajec01/order
216)  200   http://www.stud.fit.vutbr.cz/~xrajec01/protests
217)  200   http://www.stud.fit.vutbr.cz/~xrajec01/289
218)  200   http://www.stud.fit.vutbr.cz/~xrajec01/heven
219)  200   http://www.stud.fit.vutbr.cz/~xrajec01/signals
220)  200   http://www.stud.fit.vutbr.cz/~xrajec01/fight
221)  200   http://www.stud.fit.vutbr.cz/~xrajec01/world
```

Obr. 7.3: Ukážka štandardného výstupu pri vykonávaní testu analýzy cenzúry s prepínačom *keywords*

```
    "tape": 200,
    "weapon": 200,
    "google": 200,
    "cops": 200,
    "blackout": 200,
    "down": 200,
    "arrested": 200,
    "bazoi": 200,
    "emigrate": 200,
246     "larceny": 200,
247     "explode": 200,
248     "great": 200,
249     "election": 200,
250     "victim": 200,
251     "incapable": 200,
252     "minister": 200,
253     "shadian": 200,
254     "LGBT": 200,
```

Obr. 7.4: Príklad záznamov z testovania prítomnosti cenzúry kľúčových slov.

Argument -p umožňuje smerovať komunikáciu programu cez zadaný proxy server. Očakáva sa tvar <server>:<port>, napríklad "-p 14.38.255.8:80". Tento proxy server je počas behu programu poskytnutý všetkým metódam `requests.get()` ako argument.

Po spustení programu je aktivovaný timer, ktorý pri ukončení vypíše na štandardný výstup dĺžku trvania testu. Príkladom je následujúci údaj:

```
Done in: 00:56:04 (hh:mm:ss)
```

Tento výstup sa vypíše vždy, keď program riadne ukončí svoju činnosť a nie je ho možné vypnúť ani prepínačom -s.

```
"https://cn.blizzard.com/zh-cn": {
    "STOPPED": "--",
    "PING": "--",
    "DNS": "XX",
    "REQUEST": "--",
    "TYPE": "URI",
    "SYNADDR": "XX",
    "EXTREQUEST": "OK"
},
209     "https://americanprogress.org": {
210         "TYPE": "URI",
211         "DNS": [
212             "13.224.197.120",
213             "13.224.197.115"
214         ],
215         "PING": "OK",
216         "REQUEST": "OK",
217         "EXTREQUEST": "--",
218         "STOPPED": "--",
219         "SYNADDR": "13.224.197.115"
220     },
```

Obr. 7.5: Príklad záznamov z testovania dostupnosti internetových stránok.

Pomocou argumentu -n je možné zabrániť vytváraniu výstupného súboru. Všetky informácie, ktoré sú v tom prípade k dispozícii sa nachádzajú na štandardnom výstupe.

Kapitola 8

Zber a vyhodnotenie dát

S využitím prostriedkov uvedených v predošlej kapitole bola vykonaná séria testov a získané dáta boli prenesené na lokálny počítač pre ďalšiu analýzu. Presný mechanizmus zberu dát približuje táto kapitola.

8.1 Širokospektrálna analýza cenzúry

Nástroj `analyze.py` je spustený na čínskom uzle Planetlabu v Pekingu. Vstupom je zoznam 566 stránok, ktoré pochádzajú z projektu na GitHube [20], ako bolo uvedené v 6.3.1. Pre každú webovú stránku z pripraveného zoznamu sa postupne vykonajú skúšky spojenia, ako znázorňuje obrázok 6.2. Po otestovaní všetkých stránok je výsledný súbor s dátami prenesený na lokálny počítač. Pre každú stránku existuje v súbore samostatný záznam obsahujúci informácie o dostupnosti, resp. nedostupnosti danej webovej stránky. Výsledky sa vyhodnocujú v opačnom poradí, než boli jednotlivé úrovne prístupu testované. Pre každú analyzovanú stránku sa vykonajú následujúce kroky v konkrétnom poradí:

1. Ak bola v adrese špecifikovaná cesta k inému, než len ku koreňovému dokumentu ("/") a pokus o prístup skončil úspešne, nejedná sa o cenzúru. Ak prístup zlyhal, ide o nezískanie odpovede na plnú URL adresu. V prípade, že adresa obsahuje iba koreňový dokument ("/"), čo znamená že sa v poli EXTREQUEST nachádza hodnota "--", pristupuje sa k bodu 2.

2. Ak prístup k zadanej adrese, obsahujúcej cestu ku koreňovému dokumentu ("/"), skončí úspešne, nejedná sa o cenzúru. Zároveň to implikuje prítomnosť hodnoty "--"v poli STOPPED a tiež úspešné získanie odpovede na ping. Ak sa nepodarilo k dokumentu pristúpiť, ide o nezískanie koreňového dokumentu. V prípade, že sa v poli REQUEST nachádza hodnota "--", pristupuje sa k bodu 3.

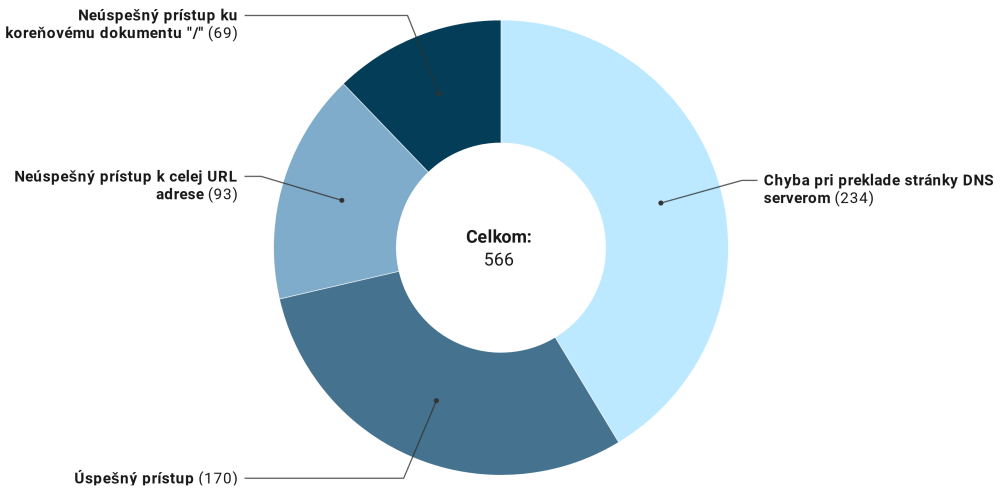
3. Ak ustanovenie TCP spojenia nebolo úspešné, ide o blokovanie IP adresy. V prípade úspechu sa pokračuje na bod 4.

4. Ak DNS preklad nebol úspešný, jedná sa o filtrovanie na úrovni protokolu DNS. Tento krok sa uplatňuje len pre internetové stránky, ktoré nie sú zadané v podobe IP adresy.

Ako je uvedené v kapitole 4.2, z dôvodu vysokej komplexnosti čínskeho systému na realizáciu cenzúry dochádza k spomaľovaniu a obmedzovaniu celej sieťovej prevádzky, čo môže mať za následok nezvyčajne chovanie siete pri prístupe na internet. Toto špecifické

poradie je zvolené pre dosiahnutie čo najvyššej presnosti pri vyhodnocovaní dát. Aj keď nejaký z predošlých krokov nedopadol úspešne, hodnota OK v poli EXTREQUEST znamená negatívny test na filtrovanie danej stránky.

Test spomínaného zoznamu 566 stránok na čínskom serveri trval takmer 7 hodín. Toto zdržanie bolo spôsobené vysokou mierou nedostupnosti stránok a častým realizovaním príkazu traceroute. Získané výsledky naznačujú isté obmedzenia v používaní čínskeho internetu, keďže až 70% stránok vykazovalo nejaký typ problému pri nadväzovaní spojenia. Rozdelenie stránok podľa úrovne, kde došlo k odopreniu prístupu, zobrazuje obrázok 8.1.



Obr. 8.1: Výsledky prístupu k testovaným stránkam.

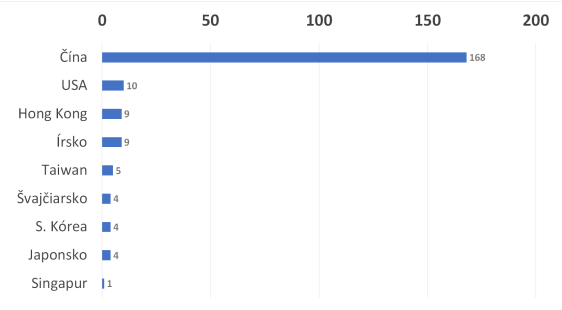
Zo všetkých testovaných stránok nie je v grafe uvedený ani jediný prípad neúspešného ustanovenia TCP spojenia. Je to z toho dôvodu, že vždy tomuto javu predchádzalo zlyhanie DNS prekladu a preto sú stránky zaradené do kategórie neúspešného prekladu.

V prípade, že program neobdržal odpoveď na ping, pomocou nástroja traceroute zisťuje, na ktorom uzle na ceste k cieľovému serveru dochádza k prerušení komunikácie. Inými slovami, z ktorého uzlu príde posledná odpoveď. Pri spustení nástroja traceroute je použitý prepínač -I, aby dochádzalo k zasielaniu ICMP správ miesto UDP, tak ako to realizuje aj nástroj ping (UDP správy sú pre traceroute implicitné nastavenie). Ak by sa zastavovala komunikácia na rovnakom uzle intenzívnejšie ako na ostatných, mohlo by to naznačovať prítomnosť Veľkého čínskeho firewallu.

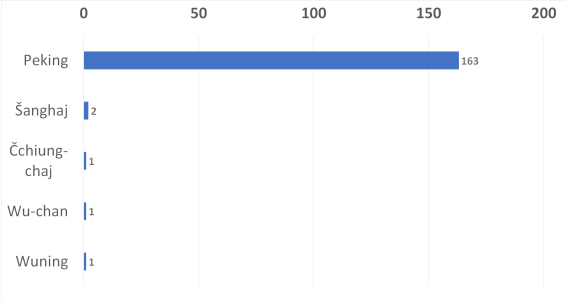
Na obrázku 8.4 je možné vidieť, že v drvivej väčšine prípadov, kedy došlo k zastaveniu komunikácie pri vykonávaní príkazu traceroute, sa jednalo o uzly umiestnené práve v Číne. Po bližšej lokalizácii je zrejmé, že v meste Peking sa pravdepodobne nachádza uzol, kde končí časť internetovej komunikácie. Údaje k všetkým čínskym uzlom je možné vidieť na obrázku 8.3 a reprezentáciu na mape zobrazuje 8.4.

Získané dáta naznačujú, že Peking, prípadne nejaký hardwarový či softwarový prostriedok umiestnený v tomto meste, je súčasťou Veľkého čínskeho firewallu. Je pravdepodobné, že sieťová komunikácia je úmyselne smerovaná cez tento uzol, aby cenzorom bolo umožnené analyzovať prenášané dáta a zahadzovať pakety s neželanými informáciami.

Pri interpretácii výsledkov je nutné brať do úvahy premenlivosť prostredia internetu. Na príklad, z celkového počtu 161 testovaných URL adries, ktoré špecifikovali viac ako



Obr. 8.2: Počet prípadov, kedy došlo k ukončeniu komunikácie, zoradené podľa početnosti na jednotlivé štáty.



Obr. 8.3: Lokalizácia čínskych uzlov, na ktorých dochádza k ukončeniu komunikácie. V grafe je uvedený počet prípadov na každé mesto.

koreňový dokument, bolo 12 prípadov, kedy prístup ku koreňovému dokumentu dopadol úspešne ale prístup k celej URL nie. Skúmaniu tejto premenlivosti sa venuje následujúca sekcia práce, pri čom je použitý rozdielny zoznam slov.

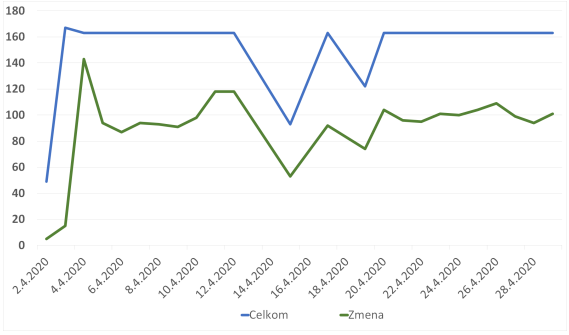
8.2 Vývoj cenzúry v čase

Ako bolo uvedené v teoretickej časti práce, realizácia cenzúry v Číne je veľmi komplexný proces pozostávajúci z manuálnych aj automatizovaných procesov, ktoré vedú k nekonzistentným reakciám na pokusy o prístup k neželaný informáciám. Následkom vedľajších účinkov filtrovania sa rovnaký efekt uplatňuje aj na niektoré bezpečné informácie, ktoré nie sú predmetom cenzúry. Tejto problematike sa bližšie venuje kapitola 4.2.2. Následujúca analýza sa zameriava na premenlivú dostupnosť vybraných internetových stránok.

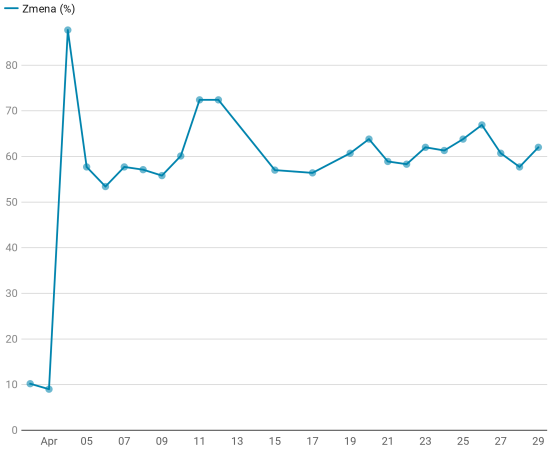
Za účelom získania informácií o premenlivosti čínskej cenzúry v čase som zostavil zoznam 160 stránok, ktoré boli testované po dobu jedného mesiaca na čínskom uzle. Zoznam netvoria len stránky s citlivým či kontroverzným obsahom, ktorý by mohol byť v Číne cenzurovaný, ale aj webové lokality, ktoré neobsahujú žiadne neželané informácie. Táto rôznorodosť vybraných stránok má za cieľ poskytnúť čo najpresnejší obraz o cenzúre. Nástroj `analyze.py`, s prepínačom *webs* pre analýzu dostupnosti webových stránok, bol spúšťaný v intervaloch každého druhého dňa až každý deň. Pôvodným plánom bolo vytvoriť proces pre automatické vykonávanie testu v pevnom intervale, no uzol Planetlabu neposkytoval úplnú funkcionalitu. Jedným z obmedzení bola nemožnosť naplánovať takýto proces. Náhradné riešenie, ktoré sa ponúkalo bol server, ktorým by sa každý deň pomocou zabezpečenej komunikácie pristupovalo k uzlu a vykonával by testy. Avšak, pre prístup k uzlu sa používal zabezpečený SSH kľúč, ktorý znemožňoval uvedenú automatizáciu. Pri vytváraní spojenia sa zakaždým vyžaduje účasť užívateľa pre zadanie hesla, ktorým je kľúč chránený. Z uvedených dôvodov boli testy vykonávane manuálne. Cieľom bolo analyzovať dostupnosť 163 stránok po dobu jedného mesiaca, no premenná stabilita uzlu neumožňovala otestovať všetky internetové stránky pri každom spustení programu. Preto existujú záznamy z určitých dní, kedy bola otestovaná iba časť zoznamu.

Predmetom sledovania boli prípady, kedy pokus o prístup na internetovú stránku priniesol iné výsledky ako v naposledy vykonanom teste. Za takúto zmenu sa považuje akákoľvek odchýlka vrátane rozdielnej IP adresy vrátenej DNS serverom. Získané výsledky sú znázornené na obrázku 8.5.

Obr. 8.4: Umiestnenie čínskych blokujúcich uzlov na mape spolu s frekvenciou výskytu.



Obr. 8.5: Graf znázorňujúci celkový počet testovaných stránok na každý testovací deň po dobu jedného mesiaca (modrá krivka) a počet stránok, ktoré zaznamenali rozdielne výsledky ako v predchádzajúcom teste (zelená krivka).



Obr. 8.6: Podiel stránok s rozdielnymi výsledkami z testovania dostupnosti v porovnaní s predošlým testom vzhľadom k danému dňu.

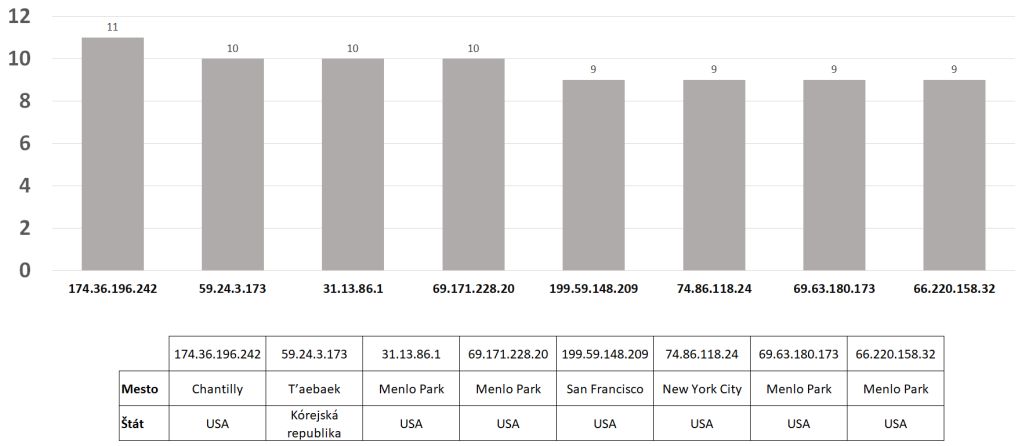
Z každého testovacieho dňa počínajúc druhým dňom bola vypočítaná miera zmeny v percentách oproti predošlému testovaniu, čo znázorňuje graf 8.6. Po spriemerovaní hodnôt zo všetkých testovacích dní vyšla výsledná miera nekonzistencie odpovedí 57,6%.

Celkom 30 internetových stránok z daného zoznamu obdržalo v každom teste identické odpovede. Medzi tieto stránky patria aj weby Masarykovej univerzity, Fakulty informačných technológií VUT a tiež stránky sociálnych sietí ako Instagram, Tiktok či Snapchat. Stránky

českých univerzít mali konštantné výsledky z testovania dostupnosti, ktoré umožňujú bezproblémový prístup z miesta realizácie testov. Stránky spomínaných sociálnych médií, na rozdiel od tých univerzitných, sú neprístupné a filtrovanie sa vždy prejavuje identicky.

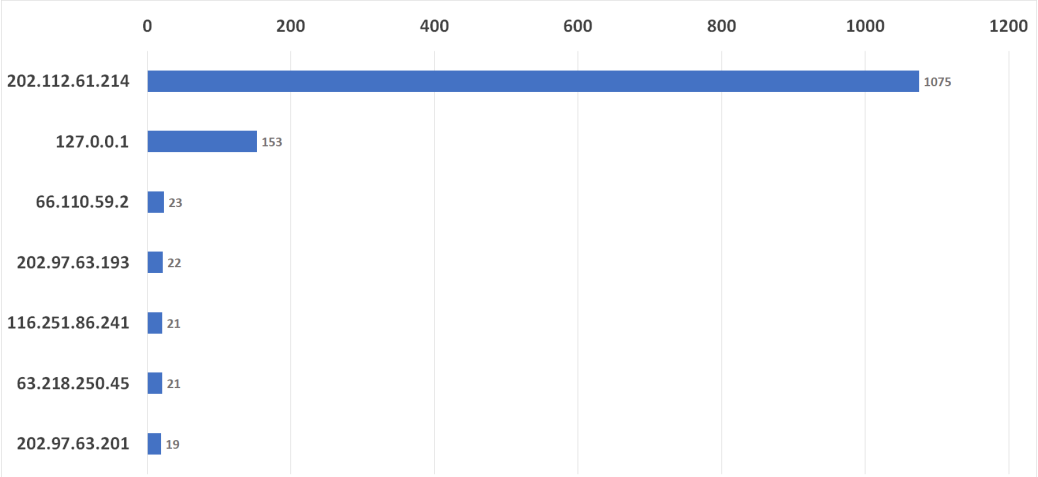
Testy odhalili IP adresy, ktoré počas testovacieho obdobia boli vrátené ako výsledok DNS prekladu viacerým stránkam. Vo výsledkoch sa vyskytujú adresy, ktoré počas mesiaca boli preložené z 9-11 rozdielnych stránok. Adresa 174.36.196.242 bola 5.4.2020 vrátená ako odpoveď na `badoo.com` a tiež `plurk.com`. Zoznam IP adries aj s počtom webových stránok, na ktoré boli preložené, zobrazuje graf na obrázku 8.7. Pre kontrolu bol test realizovaný aj zo Slovenska. Predmetom záujmu boli IP adresy, ktoré sa v poli DNS objavili viac ako pri jednom webe. Výsledky ukázali, že ak takáto situácia nastala, vždy sa jednalo o rovnakú doménu pričom sa líšila len doména nižšie radu, prípadne cesta k dokumentu. Kontrolné testy z územia Slovenska dopadli podľa očakávaní a dáta získané v Číne tým pádom skutočne naznačujú podozrivú aktivitu.

Dáta získané pri analyzovaní premenlivosti cenzúry som využil aj na lokalizovanie uzlov, na ktorých najčastejšie dochádzalo k ukončeniu komunikácie (pole STOPPED). S rozdielnym zoznamom stránok je možné získať odlišné informácie než v predchádzajúcej časti práce. Je nutné podotknúť, že rovnaký súbor so stránkami bol opakovane testovaný počas mesiaca. Z toho plynie predpoklad, že výrazne jednoznačné údaje sú znásobené opakovaným testovaním, a tým pádom môžu vyvolávať falošný pocit jednoznačnosti. Na obrázku 8.8 je vidieť IP adresy uzlov, na ktorých najčastejšie dochádzalo k ukončeniu komunikácie. Inými slovami, odkiaľ sa vrátila posledná odpoveď pri spustení nástroja traceroute. Najčastejšia adresa, na ktorej sa zastavovala komunikácia na úrovni ICMP správ je 202.112.61.214, ktorá sa nachádza v Číne, konkrétne v Pekingu. Druhá najčastejšia adresa je 127.0.0.1, čo je rezervovaná adresa pre označenie vlastného zariadenia v sieti. To naznačuje, že Veľký čínsky firewall zabraňuje tomu, aby sme pri prístupe na určité webové stránky opustili lokálnu sieť. Ďalšie adresy sú výrazne menej početné, avšak veľká časť z nich sa nachádza rovnako v meste Peking.



	174.36.196.242	59.24.3.173	31.13.86.1	69.171.228.20	199.59.148.209	74.86.118.24	69.63.180.173	66.220.158.32
Mesto	Chantilly	T'aebaek	Menlo Park	Menlo Park	San Francisco	New York City	Menlo Park	Menlo Park
Štát	USA	Kórejská republika	USA	USA	USA	USA	USA	USA

Obr. 8.7: Graf znázorňujúci počet internetových stránok, ktoré boli preložené na jednu IP adresu počas testovacieho obdobia jedného mesiaca.



	202.112.61.214	127.0.0.1	66.110.59.2	202.97.63.193	116.251.86.241	63.218.250.45	202.97.63.201
Výskyt	1075	153	23	22	21	21	19
Mesto	Peking		Los Angeles	Peking	Peking	Peking	
Štát	Čína		USA	Čína	Čína	Čína	

Obr. 8.8: Zoznam uzlov, odkiaľ najčastejšie prichádzala posledná odpoveď nástroja traceroute. Ku každej adrese je uvedená aj konkrétna lokalita, ak bolo možné adresu lokalizovať.

8.3 Skúmanie cenzúry kľúčových slov

Pre odhalenie filtrovania kľúčových slov bol vytvorený zoznam obsahujúci viac ako 400 slov a slovných spojení, ako je uvedené na za začiatku tejto kapitoly. Najskôr sa test vykonal na lokálnom počítači na Slovensku, aby sa overila funkcionalita programu. Vstupom bola rovnaká stránka, aká je použitá aj pri testoch v Číne (teda www.stud.fit.vutb.cz/ xrajec01/) a identický zoznam kľúčových slov. Program podľa očakávaní vrátil stavové kódy 200 na všetky slová zo zoznamu a kód 404 na pár neexistujúcich dokumentov, ku ktorým sa program pokúšal pristúpiť len v rámci overovania správnosti.

Po vykonaní testu na čínskom serveri Planetlabu bol výsledný súbor s dátami prenesený na lokálny počítač pre bližšiu analýzu. Tento test ponúkol pomerne prekvapivé výsledky, konkrétne nula blokovaných slov a tým pádom 100% výskyt stavového kódu 200. Každá jedna žiadosť o prístup k URL adrese skončila úspechom a dokument bol vrátený. Pre overenie, že program dostal dokument v nezmenenej podobe, boli všetky vrátené stránky uložené pre kontrolu. Po ich analýze je možné konštatovať, že čínsky systém pre filtrovanie dát ani v jedinom prípade nezasiahol a neobmedzil komunikáciu.

8.4 Zhodnotenie výsledkov

Získané výsledky naznačujú, že Veľký čínsky firewall smeruje sieťovú prevádzku cez uzly Pekingu, kde dochádza k analýze prenášaných dát za účelom blokovania neželaného obsahu. Ak systém vyhodnotí komunikáciu ako zakázanú, jednoducho ju zablokuje. Pravdepodobne v tomto meste existuje tzv. čierna diera v sieti, ktorá zahadzuje prenášané pakety bez informovaní odosielateľa. To potvrdzuje teóriu systému realizácie cenzúry v Číne, ktorá je opísaná v kapitole 4.2.1.

Podľa získaných výsledkov je hlavná úroveň filtrovania úroveň DNS serverov, ktoré neposkytujú správny preklad určitých stránok. Väčšinu neúspešných spojení s cieľovým serverom tvoria dva prípady: buď nie je vrátená žiadna odpoveď na DNS požiadavku, alebo je stránka preložená na neplatnú, prípadne falošnú IP adresu. To implikuje existenciu zoznamu zakázaných stránok, podľa ktorého sa cenzúra riadi. Tento zoznam je nutné pravidelne aktualizovať aby nedochádzalo k chybnému povoleniu nepovolenej komunikácie. Taktiež je pravdepodobné, že GFW obsahuje zoznam IP adries, ktoré sú falošné, a ktoré sa vracajú ako odpoveď na preklad zakázanej stránky. Proces prístupu k webovej stránke je tým znemožnený hneď v prvej fáze.

Množstvo zdrojov uvedených v tejto práci prezentuje prítomnosť rozsiahleho systému pre vykonávanie cenzúry, ktorý operuje na čínskom internete, ktorý je tiež známy pod názvom Veľký čínsky firewall. Tieto informácie však odporujú získaným výsledkom z testov na prítomnosť cenzúry výhradne kľúčových slov. Všetky dostupné informácie spolu naznačujú istú mieru neobjektívnosti, resp. skreslenia, pri zbere dát. Poukazuje to na snahu čínskej vlády utajovať existenciu cenzúry.

Jednou z možností, prečo nebol zistená cenzúra kľúčových slov môže byť aj spôsob využívania siete Planetlab. Ten poskytoval prístup užívateľom z celého sveta, vďaka čomu cenzúra na konkrétne slová nemusela byť aplikovaná. Štúdia z roku 2015 [16] potvrdzuje, že akademické siete neposkytujú reprezentatívny pohľad na stav cenzúry v krajine. Predovšetkým v prípade štátov s rozsiahlou internetovou cenzúrou môže byť merateľná miera filtrovaného obsahu až o 26% nižšia oproti skutočnosti.

Kapitola 9

Záver

Cieľom práce bolo preskúmať aktuálny stav cenzúry vo vybranej krajine, ktorou bola Čína. Samotnému testovaniu prechádzala štúdia cenzorských prostriedkov využívaných vo svete a dostupných materiálov týkajúcich sa cenzúry na území Číny. Ďalej bola práca zameraná na možnosti obchádzania cenzúry, čo je rozsiahla oblasť obsahujúca rôzne technológie, od triviálnych po sofistikované. Cenzori okrem realizácie cenzúry bojujú aj proti jej obchádzaniu. S ohľadom na všetky dostupné štúdie a texty zaoberajúce sa danou problematikou bol navrhnutý a implementovaný nástroj na praktické overenie cenzúry v ľubovolnej krajine. Získané výsledky jednoznačne poukazujú na existenciu jedného hlavného bodu čínskeho systému pre výkon cenzúry, ktorým je Peking. Taktiež pravdepodobne existuje určitá snaha čínskej vlády o utajovanie cenzúry.

Realizovanie meraní bolo výrazne sťažené zlým stavom služby Planetlab, ktorá bola vybraná pre zabezpečenie prístupu k čínskemu poskytovateľovi internetu, nakoľko bolo možné pripojenie len do jedného mesta z celej Číny. Ďalšou prekážkou bolo náhle ukončenie služby Planentlab uprostred písania práce.

Aj napriek spomenutým nepríjemnostiam mi práca rozšírila obzory v oblasti filtrovania dát na internete, výraznou mierou prispela k mojim vedomostiam ohľadom sieťovej komunikácie a rozvila znalosti programovacieho jazyku Python.

Výsledky práce sú pripravené pre publikovanie v rámci odbornej správy, ktorej spracovanie je súčasťou projektu bakalárskej práce.

Prácu by bolo možné rozšíriť o dáta získane z bežných uzlov v Číne, ktoré by nepatrili do akademickej siete. S vyššou stabilitou uzlov by bolo možné vykonať rozsiahlejšie testovanie a získať podrobnejší obraz ohľadom cenzúry v Číne.

Literatúra

[1] *PlanetLab History* [online]. [cit. 2020-05-12]. Dostupné z: http://www.planet-lab.org/history.

[2] YouTube Hijacking: A RIPE NCC RIS case study. [online]. RIPE Network Coordination Centre. Marec 2008, [cit. 2020-05-19]. Dostupné z: https://www.ripe.net/publications/news/industry-developments/youtube-hijacking-a-ripe-ncc-ris-case-study.

[3] *Troubleshoot TCP/IP connectivity* [online]. Jún 2018 [cit. 2020-04-20]. Dostupné z: https://docs.microsoft.com/en-us/windows/client-management/troubleshoot-tcpip-connectivity.

[4] Anderson, D. Splinternet Behind the Great Firewall of China. *Queue*. New York, NY, USA: Association for Computing Machinery. november 2012, zv. 10, č. 11, s. 40–49. DOI: 10.1145/2390756.2405036. ISSN 1542-7730. Dostupné z: https://doi.org/10.1145/2390756.2405036.

[5] Andrews, M. *Negative Caching of DNS Queries (DNS NCACHE)* [Internet Requests for Comments]. RFC 2308.

[6] Biryukov, A., Pustogarov, I. a Weinmann, R.-P. *Trawling for Tor Hidden Services: Detection, Measurement, Deanonymization* [online]. Luxembourg: University of Luxembourg, 2013 [cit. 2020-04-05]. Dostupné z: https://www.ieee-security.org/TC/SP2013/papers/4977a080.pdf.

[7] Callanan, C., Dries Ziekenheiner, H., Guerra, R. et al. *Leaping Over the Firewall: A Review of Censorship Circumvention Tools*. Washington, D.C.: Freedom House, 2011.

[8] Carpenter, B. a Brim, S. *Middleboxes: Taxonomy and Issues* [Internet Requests for Comments]. RFC 3234.

[9] Chaabane, A., Chen, T., Cunche, M. et al. Censorship in the Wild: Analyzing Internet Filtering in Syria. In: *Proceedings of the 2014 Conference on Internet Measurement Conference*. New York, NY, USA: Association for Computing Machinery, 2014, s. 285–298. IMC '14. DOI: 10.1145/2663716.2663720. ISBN 9781450332132. Dostupné z: https://doi.org/10.1145/2663716.2663720.

[10] Chapman, D. W. *Zabezpečení sítí pomocí Cisco PIX Firewall*. Vyd. 1. Brno: Computer Press, 2004. ISBN 80-7226-963-1.

[11] Clayton, R., Murdoch, S. J. a Watson, R. N. M. Ignoring the Great Firewall of China. In: Danezis, G. a Golle, P., ed. *Privacy Enhancing Technologies*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, s. 20–35. ISBN 978-3-540-68793-1.

[12] Deibert, R., Rohozinski, R. a Palfrey, J. *Access Denied - The Practice and Policy of Global Internet Filtering*. MIT Press, 2008. ISBN 02-620-4245-2.

[13] Dingledine, R., Mathewson, N. a Syverson, P. *Tor: The Second-Generation Onion Router*. The USENIX Association, 2004. Dostupné z: https://www.usenix.org/legacy/events/sec04/tech/full_papers/dingledine/dingledine.pdf.

[14] Dornseif, M. Government mandated blocking of foreign Web content. Júl 2003.

[15] Ensafi, R., Fifield, D., Winter, P., Feamster, N., Weaver, N. et al. Examining How the Great Firewall Discovers Hidden Circumvention Servers. In: *Proceedings of the 2015 Internet Measurement Conference*. New York, NY, USA: Association for Computing Machinery, 2015, s. 445–458. IMC '15. DOI: 10.1145/2815675.2815690. ISBN 9781450338486. Dostupné z: https://doi.org/10.1145/2815675.2815690.

[16] Gill, P., Crete Nishihata, M., Dalek, J., Goldberg, S., Senft, A. et al. Characterizing Web Censorship Worldwide: Another Look at the OpenNet Initiative Data. *ACM Trans. Web*. New York, NY, USA: Association for Computing Machinery. január 2015, zv. 9, č. 1. DOI: 10.1145/2700339. ISSN 1559-1131. Dostupné z: https://doi.org/10.1145/2700339.

[17] King, G., Pan, J. a Roberts, M. E. How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review*. 2013, zv. 107, 2 (May), s. 1–18.

[18] Kührer, M., Hupperich, T., Bushart et al. Going Wild: Large-Scale Classification of Open DNS Resolvers. In: *Proceedings of the 2015 Internet Measurement Conference*. New York, NY, USA: Association for Computing Machinery, 2015, s. 355–368. ISBN 9781450338486. Dostupné z: https://doi.org/10.1145/2815675.2815683.

[19] Kurose, J. F. *Computer networking : a top-down approach*. 6th ed., International. Boston ; London: Pearson, 2013. ISBN 978-0-273-76896-8.

[20] Lab, C. a Others. *URL testing lists intended for discovering website censorship*. 2014. https://github.com/citizenlab/test-lists. Dostupné z: https://github.com/citizenlab/test-lists.

[21] Luotonen, A. a Altis, K. *World-Wide Web Proxies* [online]. Finland: [b.n.], apríl 1994 [cit. 2020-04-05]. Dostupné z: http://courses.cs.vt.edu/~cs4244/spring.09/documents/Proxies.pdf.

[22] Navrátil, J. [elektronická pošta]. Message to: rajecky.m@gmail.com 19.feb 2020 [cit. 2020-05-12]. Osobná komunikácia.

[23] Siegel, R. *Search result not found: China bans Wikipedia in all languages* [online]. The Washington Post, máj 2019 [cit. 2020-04-10]. Dostupné z: https://www.washingtonpost.com/business/2019/05/15/china-bans-wikipedia-all-languages/.

[24] Smith, C. *We Had Our Arguments, But We Will Miss You Wikipedia* [online]. HuffPost, jún 2015 [cit. 2020-04-10]. Dostupné z: https://www.huffpost.com/entry/we-had-our-arguments-but-_b_7610130.

[25] Sturtz, J. *Dictionaries in Python* [online]. Real Python [cit. 2019-05-20]. Dostupné z: https://realpython.com/python-dicts/.

[26] Toor, A. *China is building its own version of Wikipedia* [online]. The Verge, máj 2017 [cit. 2020-04-10]. Dostupné z: https://www.theverge.com/2017/5/4/15541016/china-wikipedia-encyclopedia-online-censorship.

[27] Verkamp, J.-P. Inferring Mechanics of Web Censorship Around the World. In: *Presented as part of the 2nd USENIX Workshop on Free and Open Communications on the Internet*. Bellevue, WA: USENIX, 2012. Dostupné z: https://www.usenix.org/conference/foci12/workshop-program/presentation/Verkamp.

[28] Watt, L. *China is launching its own version of Wikipedia – without public contributions* [online]. Independent, máj 2017 [cit. 2020-04-10]. Dostupné z: https://www.independent.co.uk/news/world/asia/china-wikipedia-chinese-version-government-no-public-authors-contributions-communist-party-line-a7717861.html.

[29] Yuan, L. *Young people in China don't know the internet we do – and they like it that way* [online]. September 2018 [cit. 2020-04-10]. Dostupné z: https://www.independent.co.uk/life-style/gadgets-and-tech/features/china-internet-social-media-great-firewall-of-china-censorship-apps-a8510036.html.

Príloha A

Spustenie aplikácie

Nástroj `analyze.py` je implementovaný v jazyku Python 3 a používa následujúce moduly, ktoré musia byť dostupné pri spustení:

- requests
- subprocess
- re
- sys
- argparse
- socket
- json
- time
- os
- ipinfo
- urllib

Zároveň program musí mať k dispozícii pomocné triedy, ktoré používa pri realizovaní testov. Tie sa povinne nachádzajú v priečinku nazvanom *classes*. Príkaz, ktorým je možné v terminály spustiť nástroj `analyze.py` vyzerá napríklad následovne:

```
./analze.py -s webs ../LIST.txt
```

Tým je zvolená funkcia analýzy dostupnosti webových stránok, ktoré sa nachádzajú v súbore LIST.txt umiestnenom v nadradenom adresári. Na štandardnom výstupe nebudú vypisované žiadne informácie okrem údaju o dĺžke trvania testu, ktorý sa zobrazí na konci.

Súbor s výsledkami vo formáte JSON sa bude nachádzať v rovnakom adresári ako spustený skript.