Univerzita Karlova

Filozofická fakulta

Ústav informačních studií a knihovnictví / Studia nových médií

Diplomová práce

Bc. Pavel Schamberger

Způsoby využívání anonymizační sítě Tor

Ways of using Tor anonymous network

Praha, 2017

Vedoucí práce: Mgr. Vít Šisler, PhD.

Poděkování

Rád bych poděkoval vedoucímu této diplomové práce Mgr. Vítu Šislerovi, PhD. za velikou ochotu a trpělivost při řešení všech problémů a v neposlední řadě také za velmi podnětné připomínky.

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

V Praze, dne 31. července 2017

.....

Pavel Schamberger

Abstrakt

Tato diplomová práce se věnuje způsobům využívání anonymizační sítě Tor a to jak z pohledu užívání v pozitivním slova smyslu, tak ve smyslu negativním, tedy zneužívání. Cílem této práce je identifikovat různé způsoby využívání sítě a kvantifikovat zneužívání sítě na základě analyzovaných hlášení o zneužívání. Nejprve jsou definovány základní pojmy a stručně popsán vývoj internetu a identifikačních technologií v jeho prostředí. Následně je představen koncept Onion routingu, na jehož základě Tor síť funguje, spolu s větším technickým detailem o fungování sítě. V neposlední řadě je popsán Tor Browser, jakožto nejčastější nástroj pro využívání Tor sítě spolu s typy uživatelů, kteří jej využívají. V poslední kapitole teoretické části je pak představena problematika zneužívání sítě. Praktická část je řešena primárně formou kvantitativní analýzy hlášení o zneužívání a to z důvodu jejich počtu téměř 3 miliónů. Kvantitativní analýza je provedena za využití statistického jazyka R a základních data miningových, text miningových a statistických metod. Analyzovaná data se týkají několika velkých uzlů Tor sítě a sahají několik let do historie. Přestože celkový počet stížností v čase téměř exponenciálně narůstá, stížnosti na maligní chování uživatelů mimo ty copyrightové narůstají velmi mírně a to i při zvyšujícím se počtu uživatelů a potenciálu sítě.

Klíčová slova

anonymita, identita, internet, soukromí, web, lidská práva, zneužívání, Tor

Abstract

This diploma thesis deals with ways of using Tor anonymization network both from the point of view of the use in the positive sense and the use in the negative sense, the abuse. The aim of this work is to identify different ways of using the network and to quantify network abuse based on analyzed abuse reports. First, the basic concepts are defined and the development of the Internet and identification technologies in its environment are briefly described. Then, the concept of Onion routing, based on which the Tor network works, together with more technical details about the functioning of the network, is introduced. Last but not least, the Tor Browser is described as the most common tool for using the Tor network along with the types of users who use it. The last chapter of the theoretical part introduces the problem of network abuse. The practical part is primarily solved in the form of a quantitative analysis of abuse, due to their volume of almost 3 million. Quantitative analysis is done using the statistical language R and basic mining data, text mining and statistical methods. The analyzed data are related to several large nodes of the Tor network and contains several years of history. Although the total number of complaints increases almost exponentially over time, complaints about malicious behavior by users except those copyright related are growing very moderately, even with an increasing number of users and whole network potential.

Keywords

anonymity, identity, internet, privacy, web, human rights, abuse, Tor

Seznam použitých zkratek

CAPTCHA	Completely automated public Turing test to tell computers and humans apart
IP	Internet Protocol
Tor	The Onion Routing
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
TCP	Transmission Control Protocol
CERN	Conseil Européen pour la recherche nucléaire
HTML	HyperText Markup Language
URL	Uniform Resource Locator
ESR	Extended Support Release
FTP	File Transfer Protocol
DNS	Domain Name System
DDoS	Distributed Denial of Service
DMCA	Digital Millenium Copyright Act
P2P	Peer-to-peer
CMS	Content Management System
API	Application Programming Interface
SEO	Search Engine Optimization

Seznam použitých tabulek

Seznam použitých obrázků

Obsah

1. Úvod

Při čím dál větším prorůstání digitálních technologií do našich životů se setkáváme s tlakem na užívání naší reálné identity v kyberprostoru. Sociální sítě namísto uživatelského jména preferují jméno reálné, k ověřování účtů dochází pomocí telefonního čísla nebo zasláním fyzického dopisu na adresu trvalého bydliště. Tento tlak je do jisté míry pochopitelný a nepochybně souvisí s přerodem a dospíváním celého kyberprostoru z divokého, ničím neregulovaného místa, do média, které je přístupné široké veřejnosti a začínají zde platit pravidla podobná těm v prostoru fyzickém.

Celý tento vývoj by kromě ustanovení stabilního prostředí pro podnikání měl vyústit ve fungující e-government a volby po internetu, kdy je pravá identita uživatelů zajištěna do takové míry, že není nutné jejich fyzické ověření jinou osobou - úředníkem. V tomto ohledu je velmi progresivní Estonsko, které již v roce 2005 využilo volebního hlasování po internetu.[1]

Existují ale i případy, kdy je pro uživatele žádoucí, aby se skryl za pseudonymem, či působil na internetu zcela anonymně, což je v dnešní době z technického hlediska velmi nelehký úkol. Samotný prohlížeč Google Chrome varuje při otevření *anonymního* okna: “Anonymní režim neskryje vaši aktivitu před vaším zaměstnavatelem, poskytovatelem internetových služeb ani webovými stránkami, které navštívíte.” Jinak řečeno tento takzvaný anonymní režim nemá s anonymitou příliš společného.

Jednou z mála možností, jak si na internetu obstarat anonymitu, pak zůstává anonymizační software Tor[2], který je se svými dvěma milióny uživateli dominantním nástrojem pro poskytování anonymity v prostředí internetu. V širokém spektru uživatelů se velmi různí jejich motivace pro užívání této sítě. Těmito důvody mohou být například ochrana svého soukromí, obejití cenzury

[1] ISA. EGovernment in Estonia [online]. , 53 [cit. 2017-07-04]. Dostupné z: https://joinup.ec.europa.eu/sites/default/files/ckeditor_files/files/eGovernment%20in%20Estonia%20-%20February%202016%20-%2018_00_v4_00.pdf

[2] Tor Project: Anonymity Online [online]. [cit. 2017-03-11]. Dostupné z: https://www.torproject.org/

či zajištění svobody slova.[3] Naproti tomu existují uživatelé, kteří síť zneužívají pro počítačové útoky, podvody, nebo prodej ilegálních předmětů a látek.[4]

Bohužel z důvodu zmiňovaného zneužívání je často tento projekt vnímán v negativních konotacích, čímž trpí všichni uživatelé. Ti jsou pak obtěžováni častou nutností prokazovat svoji lidskost například za použití CAPTCHA (akronym pro Completely Automated Public Turing test to tell Computers and Humans Apart) a nebo úplnou blokací IP adres Tor serverů ze strany hostingových a internetových poskytovatelů. Tomuto trendu bohužel nepomáhají ani média, která si oblíbila slova jako je Darknet ve spojitosti s Tor sítí a pouze prohlubují negativní vnímání veřejností.

Jako Darknet jsou označovány takzvané Onion Services, dříve Hidden services, které umožňují uvnitř Tor sítě hostovat webové stránky, jenž nejsou mimo tuto síť dostupné. Poté, co se širší veřejnost dozvěděla o existenci virtuálních tržišť v rámci Onion Services, kde bylo dostupné ilegální zboží za úplatu v elektronické měně Bitocin, tak se velmi rychle označení Darknet ujalo. Tento populární název ale nereflektuje existenci kopie Wikipedie, Facebooku, nebo schránek pro whistleblowery a dalších prospěšných projektů, které jsou rovněž v rámci Onion Services provozovány. Z principu fungování softwaru Tor pak prakticky neexistují žádné pozitivní příběhy, které by sloužily jako přirozená protiváha. Z tohoto důvodu bych rád analyzoval data o zneužívání a na jejich základě vyhodnotil, jak velice závažným problémem zneužívání této sítě je.

Cílem této diplomové práce je prozkoumat způsoby využívání anonymizační sítě Tor s důrazem na maligní činnost uživatelů a to za účelem zjistit, jakými způsoby je síť nejčastěji zneužívána. Rovněž také prozkoumat, jak moc ovlivňuje množství protékajících dat počet stížností v čase a v neposlední řadě v jakém poměru dochází ke zneužívání sítě. To vše na základě hlášení o zneužívání, takzvaných *abuse reports*, vztahujících se k IP adresám Tor serverů. Tyto reporty jsou ve formě emailových zpráv zasílány operátorům serverů, ze kterých je detekována podezřelá aktivita a mohou být jak strojově generovány, tak ručně psány.

[3] ZAHORSKY, Ingmar. Tor, Anonymity, and the Arab Spring: An Interview with Jacob Appelbaum [online]. [cit. 2017-04-30]. Dostupné z: http://www.monitor.upeace.org/innerpg.cfm?id_article=816

[4] MOORE, Daniel a Thomas RID. Cryptopolitik and the Darknet. Survival [online]. 2016, 58(1), 7-38 [cit. 2017-05-27]. DOI: 10.1080/00396338.2016.1142085. ISSN 0039-6338. Dostupné z: http://www.tandfonline.com/doi/full/10.1080/00396338.2016.1142085

Teoretická část práce se v první kapitole zaměřuje na definování základních pojmů jako je pseudonymita, anonymita a identita. V kapitole druhé je popsána jejich aplikace a historický vývoj v prostředí internetu. V třetí kapitole je představen a popsán koncept a princip fungování Onion routingu, na jehož základě síť Tor funguje. Pomocí základních designových dokumentů Tor sítě a Tor browseru je přiblížena komplexnost a technické limitace anonymity na internetu. V závěru druhé části jsou představeny cílové skupiny uživatelů a popsány jejich motivace pro využívání softwaru Tor a anonymity na internetu obecně. Na samotném konci této části se pak nachází syntéza výzkumů zneužívání Tor sítě, které se liší jak metodikou, tak závěry.

Praktická část práce je postavena na analýze dat o zneužívání Tor sítě na základě již zmíněných hlášení o zneužívání (abuse reportů). V souladu s předem zvoleným přístupem postupuji přes explorativní analýzu dat, jejímiž výstupy jsou mimo jiné základní statistiky jako je časové rozložení a počty hlášení. Následně jsou data zpracována do strukturované podoby s extrakcí dalších informací jako jsou zmíněné IP adresy, názvy zneužívaných souborů, nebo mailové adresy odesílatelů stížností.

Za účelem přesnějších výsledků analýzy rozděluji analyzovaná data na dvě části. První polovina jsou stížnosti týkající se autorských práv, které jsou natolik odlišné a v celkovém množství natolik převažují nad druhou polovinu, že je bylo nutné zkoumat odděleně. Z těchto dat je primárně možné zjistit informace o zneužívání sítě ve smyslu nelegálního sdílení autorsky chráněného obsahu. V pozdější části analýzy se zabývám druhou polovinou dat, tedy hlášeními, které se autorských práv netýkají. Tato data jsou velmi rozmanitá a jejich analýza o to náročnější, nicméně právě z těchto dat je možné zjistit více o zneužívání sítě ve smyslu počítačových útoků a nelegálního obsahu.

Výzkumná část je řešena primárně formou kvantitativní analýzy hlášení o zneužívání, v určitých částech doplněnou analýzou kvalitativní. Kvantitativní analýza je provedena za využití statistického jazyka R a základních data miningových, text miningových a statistických metod.

Hlavním zjištěním práce je odvozený poměr benigních a maligních uživatelů části Tor sítě, která je provozována organizací Torservers.net. Tento poměr dokazuje, že většina uživatelů Tor síť nezneužívá. Dalším zjištěním je rapidní nárůst počtu stížností na nelegální sdílení autorsky chráněného obsahu v kontrastu ke zbylým stížnostem, které narůstají velmi mírně a to i při zvyšujícím se počtu uživatelů a potenciálu sítě.

2. Teoretická část

2.1 Definice anonymity, pseudonymity a identity

Třebaže má pojem anonymita kořeny v řečtině, kde slovo *anonymia* znamená beze jména či bezejmenný, hojněji se začíná objevovat až okolo 17. století v souvislosti s literární tvorbou. Ve zmíněné době dochází k nárůstu publikovaných literárních děl, u kterých je autor záměrně neuveden. V literárním a uměleckém světe obecně je rovněž poměrně často využíváno pseudonymity, kdy autor vystupuje pod pseudonymem, tedy jiným než vlastním jménem.

Důvodů k adopci pseudonymu literárním autorem může být hned několik. Od ryze praktického, kdy by mohlo dojít k záměně s již zavedeným autorem z důvodu identického jména, stejně tak v případě, že prvotní tvorba nebyla natolik úspěšná a kvalitní a autor s ní již nechce být nadále spojován. Ale i naopak, jak tomu bylo v případě J.K. Rowlingové, autorky úspěšné série knih o Harry Potterovi, která raději publikovala svůj detektivní román pod pseudonymem Robert Galbraith[5].

Podobně jako v případě J.K. Rowlingové, mnoho ženských autorek tvořilo a publikovalo pod mužskými pseudonymy hlavně v 19. století, kdy povolání spisovatele bylo téměř výhradní doménou mužů. A konečně v 20. a 21. století někteří slavní mužští autoři publikují pod ženským pseudonymem.

Poslední skupinou jsou pak autoři, kteří publikují pod pseudonymem či zcela anonymně z důvodu společensky nevyhovujícího obsahu, nebo radikálních myšlenek a názorů.

S rozmachem digitálních technologií a internetové sítě je najednou umožněno publikovat v elektronické formě obrovské mase lidí a tento fenomén nazýváme třetí informační explozí. Za první informační explozi je považován vznik písma ve starověké Mezopotámii a v údolí Nilu okolo roku 4 tisíce před našim letopočtem. Znamenala konec ústního předávání a zahájila proces zaznamenávání informací a jejich systematického zpracovávání a ukládání. Za druhou informační explozi je považována událost vynálezu knihtisku v 15. století, díky němuž byly

[5] Robert Galbraith [online]. [cit. 2017-03-11]. Dostupné z: http://robert-galbraith.com/about/

zpřístupněny informace i nižším vrstvám a položeny základy pro vzdělání širšího okruhu společnosti. Skončilo tak předčítání a naslouchání a lidé mohli konzumovat obsah napřímo. Třetí informační exploze je spjata s vývojem informačních technologií počínající rozvojem telegrafu v polovině 19. století až do současnosti, kdy do společnosti proniká internet, který zásadně mění práci s informacemi.[6]

Digitální technologie umožňují obrovské množství obsahu konzumovat a to ne pouze pasivně jak tomu bylo dřív, ale i aktivně obsah komentovat a s dalšími uživateli sítě komunikovat. Tyto principy jsou zastřešovány termínem Web 2.0, který je označením pro vývojovou etapu webu, ve které došlo k masivní adopci obsahu generovaným uživatelem. Příkladem může být oblíbená sociální síť Facebook, případně kolaborativně tvořená internetová encyklopedii Wikipedia.

Termín Web 2.0 je poprvé použit v roce 1999 Darcym DiNuccim v článku *Fragmented Future*, na který odkázal Tim O'Reilly na první konferenci o Webu 2.0 o pět let později:

"*Web, jak ho známe teď, který se jako statický text načte do okna prohlížeče, je jen zárodek webu, který přijde. První záblesky Webu 2.0 se již začínají objevovat a my sledujeme, jak se toto embryo začíná vyvíjet. Web bude chápán ne jako obrazovky plné textu a grafiky, ale jako prostředí, jako éter, jehož prostřednictvím dochází k interaktivitě. Objeví se na obrazovce počítače, na televizním přijímači, na palubní desce, na mobilním telefonu, na herní konzoli, a možná, že i na vaší mikrovlnné troubě.*"[7]

Právě v literární tvorbě je do jisté míry možné najít analogii s užíváním a publikováním své práce, nebo jen svých myšlenek na internetu. V některých částech světa mohou autoři statusů, tweetů a blogspotů čelit za svoje myšlenky obsažené v jejich dílech dotýkajících se politiky, víry a nebo sexuality perzekucím jak ze strany autorit, tak ze strany jinak smýšlejících osob. Jedním z řešení může být právě uchýlení se k publikování pod pseudonymem, či zcela anonymně.

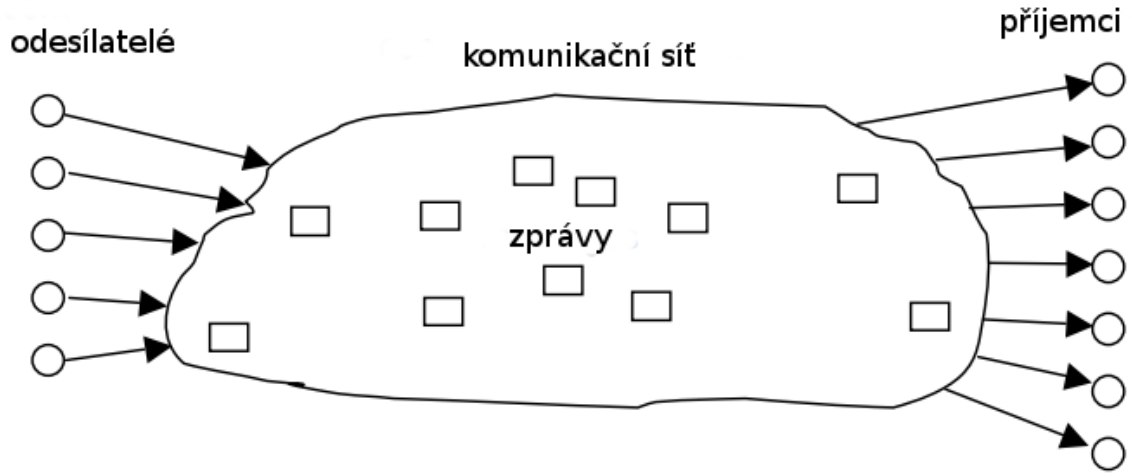
Gary Marx ve své obecné definici podmiňuje anonymitu neznalostí žádné ze sedmi dimenzí identifikačních informací. Mezi tyto dimenze patří: jméno osoby, umístění, pseudonym spojitelný se skutečným jménem nebo umístěním, pseudonym obsahující další informace, prozrazující

[6] Důsledky informačního zahlcení. WikiSofia [online]. [cit. 2017-05-07]. Dostupné z: https://wikisofia.cz/wiki/Důsledky_informačního_zahlcení
[7] DINUCCI, Darcy. Fragmented Future [online]. 1999 [cit. 2017-03-11]. Dostupné z: http://darcyd.com/fragmented_future.pdf

vzorce chování, příslušnost k některé sociální skupině nebo informace, předmět či dovednost prozrazující charakteristiku osoby.[8] Zmíněných sedm dimenzí je možné zařadit do takzvané sociální anonymity, která je aplikovatelná jak na svět skutečný, tak na svět digitální. V kyberprostoru je ovšem nutné myslet i na anonymitu technickou, kterou mezi prvními zmiňují autoři Hayne a Rice ve své práci z roku 1997, kde analyzují příspěvky uživatelů a na základě obsahu zkoumají identifikovatelnost autorů a efektivitu technické anonymity obecně.[9]

Andreas Pfitzmann s kolegy z Technické univerzity v Drážďanech vypracoval kompletní terminologii pro oblast soukromí z pohledu minimalizace dat.[10] Ve své práci *Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology* definuje a vysvětluje rozdíly mezi pseudonymitou, anonymitou, ale i nespojitelností (unlinkability) a nezpozorovatelností (unobservability).

Obrázek 1: Schéma komunikační sítě (Pfitzmann, 2005)



[8] MARX, Gary T. What's in a Name? Some Reflections on the Sociology of Anonymity. The Information Society [online]. 1999, vol. 15, issue 2, s. 99-112 [cit. 2017-03-11]. DOI: 10.1080/019722499128565. Dostupné z: http://www.tandfonline.com/doi/abs/10.1080/019722499128565

[9] HAYNE, Stephen C. a Ronald E. RICE. Attribution accuracy when using anonymity in group support systems. International Journal of Human-Computer Studies [online]. 1997, vol. 47, issue 3, s. 429-452 [cit. 2014-07-17]. DOI: 10.1006/ijhc.1997.0134. Dostupné z: http://linkinghub.elsevier.com/retrieve/pii/S1071581997901348

[10] PFITZMANN, Andreas a Marit HANSEN. Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology [online]. 2005, , 43 [cit. 2017-03-11]. Dostupné z: https://www.freehaven.net/anonbib/cache/terminology.pdf

Celá terminologie je zasazena v kontextu **odesílatelů** posílajících **zprávy příjemcům** za použití **komunikační sítě**. Toto schéma pokrývá scénáře uživatelů dotazující databáze, zákazníků nakupujících v e-shopech a mnoho dalších, ve kterých pouze dochází ke změně terminologie, která označuje odesílatele a příjemce v závislosti na kontextu. Útočník snažící se zjistit identitu uživatele může být jak vně sítě, tak i její součástí a zároveň může kontrolovat jednu, nebo více součástí sítě.

Jedním z hlavních předpokladů anonymity je vůbec existence množiny dalších subjektů, které jsou si svými atributy velmi podobní a mezi kterými je možné se ukrýt a takzvaně splynout s davem. Anonymita subjektu znamená, že subjekt není identifikovatelný v rámci této množiny subjektů, tzv. anonymitní množiny. V případě, že tato anonymitní množina neexistuje, může být anonymizační nástroj sebedokonalejší, ale při použití právě jedním uživatelem budeme vždy vědět, že je to jen a pouze on.

Anonymitní množina se může v čase měnit a to mimo jiné v závislosti na konání uživatele. Odesílatel zpráv může být anonymní pouze v množině dalších odesílatelů zpráv, příjemce v množině příjemců zpráv a tak dále. Zároveň mohou být tyto anonymitní (pod)množiny součástí větší množiny, například všech uživatelů používajících email, stále ale musí být dodrženo, že všechny subjekty v rámci množiny jsou neidentifikovatelné.

Anonymitu není možné vnímat izolovaně, ale je nutné na ni pohlížet v celé její šíři, neboť v případě skutečné anonymity je nutné myslet na mnoho jejích vlastností. V případě narušení byť jen jediné z vlastností popsaných výše se o anonymitu přestává jednat a často se mění v pseudonymitu, která je často nesprávně za anonymitu vydávána.

Anonymní stav je ale u většiny případů žádoucí pouze po přechodnou dobu a pouze pro specifické činnosti[11] a na základě výše popsaných principů anonymity by internet, jak jej známe dnes mohl fungovat jen velmi těžko. I v jeho počátcích fungoval spíše na principech pseudonymity. Pro obousměrnou komunikaci je totiž nutný adekvátní identifikátor, který může být třeba ve formě uživatelského jména, tedy pseudonymu.

[11] KANG, Ruogu, Stephanie BROWN a Sara KIESLER. Why do people seek anonymity on the internet?. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13 [online]. New York, New York, USA: ACM Press, 2013, s. 2657- [cit. 2017-07-04]. DOI: 10.1145/2470654.2481368. Dostupné z: http://dl.acm.org/citation.cfm?doid=2470654.2481368

Slovo pseudonym rovněž pochází z řečtiny, konkrétně ze slova *pseudonumon*, které znamená špatně pojmenovaný a správně indikuje pseudonym jako identifikátor subjektu, který je jiný než jeho reálné jméno. Pseudonymita je poté stav, kdy jsou pseudonymy používány jako identifikátory subjektů.

Anonymita i pseudonymita v sobě skýtá určité benefity, které částečně zmiňuji na začátku této kapitoly v souvislosti s literární tvorbou. Sociložka Sherry Turkle například zmiňuje možnost v rámci interakcí na internetu experimentovat s více verzemi vlastního já, při kterých se vracíme do dětských her, kdy nepociťujeme obavy z odmítnutí okolím.[12] Tuto myšlenku částečně reflektuje web 4chan,[13] diskuzní fórum zaměřené na sdílení obrazového obsahu, který je jedním z posledních velkých komunitních webů fungujících v anonymní rovině. Tuto stránku měsíčně navštíví 22 miliónů unikátních uživatelů a často bývá prvotním zdrojem vizuálního obsahu jako jsou memes, které se následně šíří po dalších sociálních sítích.[14] Na základě toho tvrdí Chris Poole, tvůrce této stránky, že anonymita vede k více autentickému obsahu a obhajuje tak anonymitu na internetu obecně.[15]

V protikladu k anonymitě a pseudonymitě stojí identita subjektu. Ta může být vysvětlena jako výhradní vnímání života, integrace do sociální skupiny a kontinuita, která je vázána k lidskému tělu a do určité míry utvářena společností, jak dále uvádí Pfitzmann ve své práci. S identitou v prostředí internetu se setkáváme na každodenní bázi při používání elektronické pošty, nebo sociálních sítí jako je Facebook. Tento fenomén je rovněž ilustrován zrodem takzvaných influencerů na sociálních sítích. Osob s velkým publikem, které skrze své názory formuje a jedná se o rozdílný princip oproti zmíněnému 4chanu, kde se na tomto podílí komunita jako

[12] TURKLE, Sherry. Life on the screen: identity in the age of the Internet. New York: Simon, c1995, 347 s. ISBN 978-068-4833-484. WARD, Mark. UK government tackles wrongly-blocked websites. BBC News [online]. 2014 [cit. 2017-07-04]. Dostupné z: http://www.bbc.com/news/technology-25962555

[13] 4chan [online]. [cit. 2017-07-04]. Dostupné z: http://www.4chan.org/

[14] BERNSTEIN, Michael S., Andres MONROY-HERNANDEZ, Drew HARRY, Paul ANDRE, Katrina PANOVICH a Greg VARGAS. 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community [online]. [cit. 2017-07-04]. Dostupné z: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2873/4398

[15] HALLIDAY, Josh. SXSW 2011: 4Chan founder Christopher Poole on anonymity and creativity [online]. 2011 [cit. 2017-07-04]. Dostupné z: https://www.theguardian.com/technology/2011/mar/13/christopher-poole-4chan-sxsw-keynote-speech

celek.[16] Děje se tak i díky reputaci a důvěryhodnosti, kterou tyto osoby disponují a které anonym nikdy mít nemůže.

Ucelený přehled výhod a nevýhod mezi anonymitou a identitou dle samotných uživatelů se objevuje ve studii *Why do people seek anonymity on the internet?*[17]*:*

Tabulka 1: Výhody mezi anonymitou a identitou

Kategorie	Výhody anonymity	Výhody identity
Sociální vazby	Vyhnout se ostatním Vyhnout se závazkům v rámci komunity Nižší bariéra pro nové vztahy Chránit blízké	Propojit se s reálnými přáteli Mít silnější sociální vazby Podporuje větší zapojení
Reputace a důvěra	Upřímnější hodnocení a doporučení	Dobré pro budování reputace Získat důvěru ostatních
Budování image	Mít kontrolu nad svojí image Vyhnout se ztrapnění / souzení / kritice	Vyhnout se tvrdé kritice Být konsistentní se svojí reálnou image
Emoční benefity	Pocit uvolněnosti a pohodlí Pocit pohody a sofistikovanosti	Pocit opravdovosti, integrovanosti Pocit blízkosti k ostatním
Vyjádření názoru	Nebát se vyjádřit názor	Vyvarovat se nezodpovědnému chování
Soukromí	Mít větší kontrolu nad zveřejňováním osobních údajů	Vypadat nevinně
Bezpečnost	Chránit osobní bezpečnost	Schovat se v davu

[16] FREBERG, Karen, Kristin GRAHAM, Karen MCGAUGHEY a Laura A. FREBERG. Who are the social media influencers? A study of public perceptions of personality. Public Relations Review [online]. 2011, 37(1), 90-92 [cit. 2017-07-04]. DOI: 10.1016/j.pubrev.2010.11.001. ISSN 03638111. Dostupné z: http://linkinghub.elsevier.com/retrieve/pii/S0363811110001207

[17] KANG, Ruogu, Stephanie BROWN a Sara KIESLER. Why do people seek anonymity on the internet?. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13 [online]. New York, New York, USA: ACM Press, 2013, s. 2657- [cit. 2017-07-04]. DOI: 10.1145/2470654.2481368.

	Vyhnout se právním důsledkům / spamu / stalkingu / ztráty vlastnictní	
Jednoduchost použití	Bez nutnosti přihlašování	Jednoduchost zapamatování účtu

2.2 Vývoj anonymity v prostředí internetu

Ve svých počátcích byl web respektive internet velmi anonymním místem. Primárně i díky stavebním kamenům celé sítě - protokolům IP a HTTP, které v sobě nenesou žádný stav a fungují pouze jako doručovatelé dat bez ohledu na doručovaný obsah, jeho souvislost s uživatelem nebo historičnost. Prakticky jediným identifikátorem byla IP adresa, nutná k přijímání a odesílání datových paketů. Postupem času však vzniká potřeba pro zachování stavu mezi klientem a serverem primárně ze strany e-commerce subjektů tak, aby bylo možné na webu nakupovat. Tyto technologie, počínající cookies a sběrem otisků uživatelských zařízení (fingerpriting) konče, dnes do značné míry identifikují uživatele napříč webem a dosažení anonymity na internetu, nebo jen ochrana vlastního soukromí se stává čím dál tím složitějším úkolem.

2.2.1 Vznik internetu

Jak ve své knize uvádí Christian Huitema[18], v Evropě byl v 90. letech protokol IP (Internet Protocol) často označován jako DoD IP. Toto označení odkazuje ke vzniku IP protokolu, který nevznikl jako tradiční mezinárodní standard, ale podobně jako software Tor vznikl protokol IP pod hlavičkou amerického ministerstva obrany (Department of Defense) a to skrze agenturu pro pokročilé výzkumné projekty DARPA.

[18] HUITEMA, Christian. Routing in the Internet [online]. Englewood Cliffs, N.J.: Prentice Hall PTR, c1995 [cit. 2017-07-13]. ISBN 978-0131321922.

Samotný vznik internetu je možné dohledat v prvotních experimentech se směrováním paketů, které DARPA financovala a dále měla vliv na jeho vývoj primárně v jeho počátcích na konci 60. a počátkem 70. let 20. století. Později se k vývoji přidávají i další výzkumné organizace z dalších zemí. Původní americká síť Arpanet velmi rychle expandovala přes Atlantický oceán díky satelitnímu spojení a zde je možné pozorovat oslabování vlivu amerického ministerstva obrany nad celou sítí. Ta se v 80. letech proměňuje z čistě výzkumné na síť akademickou a podobný přerod je možné sledovat i v letech devadesátých, kdy v akademické síti začínají převládat komerční motivy. Internet se rozrostl natolik, že jej žádná centrální skupina nemůže ovládat.

Souběžně s tím v první polovině 70. let postgraduální student Stanfordské univerzity Vinton G. Cerf představuje první verzi specifikací nového protokolu TCP, kterou v roce 1974 publikuje spolu s Robertem Kahnem v květném čísle časopisu IEEE Transactions on Communications.[19]

Zatímco protokol IP se stará o doručení paketů správnému příjemci, negarantuje, že budou doručeny bez chyb a kompletní. Naopak v protokolu TCP byl kladen hlavní důraz na spolehlivý přenos dat a nápravu případných chyb, což mělo za následek větší časovou náročnost. Z důvodu existence aplikací, které upřednostňovaly rychlost na úkor kvality dat, došlo k oddělení IP protokolu, který se primárně stará o přenos a směřování dat a TCP protokol, který se stará o kvalitu a kompletnost dat.

V 80. letech došlo ke kompletnímu přechodu na TCP/IP protokol a zároveň došlo k oddělení vojenské části Arpanetu do samostatné sítě Milnet, díky čemuž se síť stala více civilní. I tato skutečnost napomohla připojování dalších a dalších sítí, až se ke konci 80. let propojuje se sítí National Science Foundation - Nsfnet, která postupem času přebírá úlohu Arpanetu coby páteřní sítě. V březnu roku 1990 dochází k jeho úplnému odstavení, Nsfnet již kompletně převzal roli páteřní sítě.[20]

I samotný Nsfnet, který byl ve své době de facto synonym internetu, se postupně dostává do pozadí a uvolňuje místo komerčním poskytovatelům připojení, kteří v následujících letech plně přebírají funkcionalitu Nsfnetu.

[19] SPECIFICATION OF INTERNET TRANSMISSION CONTROL PROGRAM [online]. [cit. 2017-07-04]. Dostupné z: https://tools.ietf.org/html/rfc675
[20] PATERKA, Jiří. Na počátku byl ARPANET .. Computerworld [online]. 1995, (4) [cit. 2017-04-30]. Dostupné z: http://www.earchiv.cz/a95/a504c502.php3

Technologie je nadále vyvíjena dobrovolnickou organizací Internet Engineering Task Force (IETF)[21] a standardy jsou publikovány neziskovou organizací Internet Society.[22] Před nimi byl za vývoj odpovědný Internet Activities Board a jeho předchůdci, kteří se později zformovali v Internet Architecture Board. S rapidním růstem sítě ale přestala tato organizace efektivně zvládat své fungování, které bylo rovněž mnohými označováno za nedemokratické, proto jeho funkce v roce 1992 přebírá již zmiňovaná Internet Society.

2.2.2 Vznik webu a identifikačních technologií

Ve stejné době vzniká i World Wide Web, systém pro ukládání a prohlížení dokumentů, které na sebe odkazují hypertextovými odkazy v internetové síti. V roce 1989 jej vynalézá tehdejší zaměstnanec výzkumného střediska CERN Tim Berners-Lee. V roce 1990 představuje i první webový prohlížeč příznačně nazvaný WorldWideWeb a o rok později je celá technologie uvolněna mimo výzkumná střediska široké veřejnosti.[23]

Podobně jako dnes, byly webové stránky primárně textové dokumenty doplněné o multimediální obsah, vytvořené za pomocí jazyka HTML (HyperText Markup Language) s unikátním identifikátorem URL (Uniform Resource Locator), pomocí kterého je mezi stránkami odkazováno.

Prvotním a primárním identifikátorem na internetu je IP adresa, kterou každý uživatel a server musí disponovat, pro správou komunikaci paketů mezi nimi. Již na základě IP adresy je možné zjistit skrze kterého internetového poskytovatele uživatel přistupuje na internet a s určitou přesností i jeho geografickou polohu. Tyto informace jsou veřejně dohledatelné za použití veřejných i neveřejných databází, které disponují daty o rozsazích IP adres, které byli jednotlivým státům či poskytovatelům přiděleny. Tyto údaje ale nemusí být vždy zcela přesné, proto existují i další přístupy jak lokaci uživatele zjistit. Například za pomocí měření rychlosti

[21] The Internet Engineering Task Force [online]. [cit. 2017-07-13]. Dostupné z: https://www.ietf.org/
[22] Internet Society [online]. [cit. 2017-07-13]. Dostupné z: http://www.internetsociety.org/
[23] BERNERS-LEE, Tim. a Mark. FISCHETTI. Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor. San Francisco: HarperSanFrancisco, c1999. ISBN 978-0062515872.

odezvy, která se většinou pohybuje v jednotkách milisekund. Ketz-Bassett et al. představují takovýto přístup s přesností u většiny případů v řádech desítek kilometrů.[24]

Spolu s masovým rozšířením webu vzniká i potřeba přidání klientského stavu mezi uživatelem a serverem do jinak bezstavového protokolu HTTP. Tuto potřebu řeší v roce 1994 Lou Montulli, zaměstnanec společnosti Netscape Communications, producenta webového prohlížeče Mosaic Netscape, do kterého je podpora takzvaných cookies[25] přidána ke konci téhož roku a o rok později jsou cookies podporovány i v druhé verzi Internet Exploreru. Na internetu tak vzniká pomyslná druhá identifikační vrstva.

Slovem cookie je ve webové terminologii označován malý kousek dat, který webová stránka skrze webový prohlížeč ukládá na uživatelský počítač, nebo mobilní zařízení. Soubor cookie umožňuje, aby si webová stránka zapamatovala uživatelské akce a volby v průběhu času. Většina webových prohlížečů soubory cookie podporuje, ale uživatelé mohou nastavení prohlížeče změnit tak, aby cookie odmítli a zároveň je mohou smazat, kdykoli se jim zlíbí.

Díky technologii cookies je vyřešena implementace virtuálního nákupního košíku, jehož obsah je perzistentní vzhledem ke konkrétnímu uživateli do doby než cookie vyprší, nebo je smazána. S vývojem webu se rovněž vyvíjelo i užití cookies, které jsou v dnešní době velmi často využívané ke sledování uživatelů napříč webem za pomoci vloženého reklamního obsahu.

V případě, že uživatel obdrží cookie od webové stránky, která je vystavena na adresu z adresního řádku, jedná se o cookie první strany. Uživatel ale může rovněž obdržet cookie od třetí strany skrze vložený obsah, který je fyzicky hostován na jiném serveru - v případě takovéhoto obsahu se nejčastěji jedná reklamy. Díky tomu může pak provozovatel reklam sledovat uživatele napříč stránkami, na kterých se reklamy téhož provozovatele vyskytují a získat tak o uživateli mnohem více informací a to často bez jeho vědomí.

[24] KATZ-BASSETT, Ethan, John P. JOHN, Arvind KRISHNAMURTHY, David WETHERALL, Thomas ANDERSON a Yatin CHAWATHE. Towards IP geolocation using delay and topology measurements. Proceedings of the 6th ACM SIGCOMM on Internet measurement - IMC '06 [online]. New York, New York, USA: ACM Press, 2006, , 71- [cit. 2017-07-17]. DOI: 10.1145/1177080.1177090. ISBN 1595935614. Dostupné z: http://portal.acm.org/citation.cfm?doid=1177080.1177090

[25] D., Kristol a Montulli L. HTTP State Management Mechanism [online]. 2000 [cit. 2017-05-08]. Dostupné z: https://tools.ietf.org/html/rfc2965

Užívání technologie cookie se za více než dvacet let rozrostlo do takové míry, že Evropská unie přistoupila k její regulaci, která má za cíl omezit zneužívání této intruzivní technologie. Provozovatelé stránek jsou nuceni informovat uživatele o používání této technologie a zároveň je nutný jejich informovaný souhlas a to jak v případě uložení perzistentní cookie první strany, která zároveň nesmí mít expirační dobu více než 1 rok, tak rovněž v případě uložení všech cookies třetích stran. Jedinou výjimku tvoří dočasná cookie první strany, která je platná pouze po dobu trvání uživatelské návštěvy.[26]

Zatímco v roce 1999 označil New York Times cookies jako narušitele soukromí, které marketéři implantují do počítačů uživatelů.[27] O deset let později píše Edward Felten článek[28] s názvem „*If You're Going to Track Me, Please Use Cookies*". Volně přeloženo: V případě, že mě budete sledovat, použijte cookies prosím. V tomto článku popisuje, že přestože cookies narušují soukromí a nejsou ideální, oproti jiným identifikačním metodám poskytují daleko větší transparentnost a uživatelskou kontrolu.

Mezi takové metody se řadí například web-based device fingerprinting, technologie, která se začíná čím dál více namísto cookies prosazovat pro sledování uživatelů na internetu. Tato metoda využívá velké unikátnosti každého webového prohlížeče ve spojitosti se zařízením, na kterém je prohlížeč provozován.

Mezi prvními kdo na vzrůstající oblibu této technologie upozornil, byl v roce 2010 Peter Eckersley z Eletronic Frontier Foundation, organizace zabývající se ochranou soukromí v digitálním světě. Ten ve svém článku příznačně nazvaném *How Unique Is Your Web Browser?* (Jak unikátní je vás prohlížeč?)[29] blíže popisuje tuto technologii na základě výzkumného webu *panopticlick.eff.org*, kde sám sbíral otisky uživatelských zařízení.

Technologie web-based device fingerprinting využívá toho, že při prohlížení webu webový prohlížeč sdílí velké množství informací o zařízení, na kterém je provozován a rovněž i o svém

[26] Information providers guide The EU Internet Handbook: Cookies [online]. [cit. 2017-05-07]. Dostupné z: http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm
[27] ETZIONI, Amitai. Privacy Isn't Dead Yet [online]. 1999 [cit. 2017-07-17]. Dostupné z: http://www.nytimes.com/1999/04/06/opinion/privacy-isn-t-dead-yet.html
[28] E. W. Felten. If You're Going to Track Me, Please Use Cookies. https://freedom-to-tinker.com/blog/felten/ifyoure-going-track-me-please-use-cookies/, 2009.
[29] ECKERSLEY, Peter. How Unique Is Your Web Browser? [online]. , 1 [cit. 2017-05-08]. DOI: 10.1007/978-3-642-14527-8_1. Dostupné z: http://link.springer.com/10.1007/978-3-642-14527-8_1

nastavení a svých doplňcích. Mezi těmito informacemi, které prohlížeč sdílí, je například verze operačního systému, jazykové nastavení, časová zóna, rozlišení obrazovky, systémové fonty a mnoho dalšího. Paradoxem je, že například zapnutí volby Do Not Track (Nesledovat), která byla postupně implementována do všech populárních prohlížečů, mělo za následek, v případě jejího nerespektování, větší unikátnost daného uživatele, neboť u majority uživatelů byla tato volba vypnutá.

Závěrem experimentu Petera Eckersleyho je, že každý z prohlížečů v sobě nese alespoň 18.1 bitů identifikačních informací, což v praxi znamená, že pouze jeden prohlížeč z 286 777 může mít shodný otisk. A pro uvedení do širšího kontextu uvádí teorii dr. Latanya Sweeney, že pro identifikaci jedné osoby nám postačí znalost datumu narození, poštovního směrovacího čísla a genderu.[30] Tyto informace dohromady tvoří identifikační informaci o velikosti 33 bitů vzhledem k populaci sedmi miliard lidí. V tomto kontextu je pak otisk webového prohlížeče více unikátní informace, než je datum narození.

Novější metodou získání otisku uživatelského zařízení je takzvaný canvas fingerpriting, který funguje skrze extrakci obrazu z HTML5 canvasu. Prvek canvas je součástí HTML5, který dovoluje dynamicky vykreslovat 2D objekty a bitmapy[31] a pro získání unikátního otisku prohlížeče, tak stačí ve zlomku vteřiny vygenerovat WebGL objekt, font a barvu a následně získat hash takovéhoto obrázku. Minimální odlišnost v grafické kartě, fontu, verzích softwaru a ovladačů dovoluje vygenerovat stabilní identifikátor zařízení, který se kvalitou velmi blíží cookie.[32]

Gunas Acar et al. ve studii z roku 2014 zkoumá do jaké míry jsou podobné metody využívány v praxi. Zaměřili se právě na canvas fingerpriting, použití perzistentní cookie a na aplikaci synchronizace informací o cookies napříč více doménami, takzvaný cookies syncing. Již v roce 2014 se canvas fingerpriting objevoval na 5% z prvních 100 000 webů dle Alexa[33] žebříčku.[34]

---

[30] SWEENEY, Latanya. Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000. Dostupné z: https://dataprivacylab.org/projects/identifiability/paper1.pdf

[31] W3schools: HTML5 Canvas [online]. [cit. 2017-04-17]. Dostupné z: https://www.w3schools.com/html/html5_canvas.asp

[32] MOWERY, Keaton a Hovav SHACHAM. Pixel Perfect: Fingerprinting Canvas in HTML5 [online]. , 12 [cit. 2017-04-17]. Dostupné z: https://cseweb.ucsd.edu/~hovav/dist/canvas.pdf

[33] Alexa: Top Sites [online]. [cit. 2017-07-17]. Dostupné z: http://www.alexa.com/topsites

[34] ACAR, Gunes, Christian EUBANK, Steven ENGLEHARDT, Marc JUAREZ, Arvind NARAYANAN a Claudia DIAZ. The Web Never Forgets. Proceedings of the 2014 ACM SIGSAC Conference on Computer

Cookies odolné vůči smazání se pak objevili jen v několika jednotkách případů. Konečně synchronizace cookies se objevuje v několika stovkách případů včetně využívání cookies třetích stran.

Vyjma identifikačních technologií narušuje anonymity na internetu i sledování samotného obsahu, ke kterému uživatel přistupuje, nebo jeho komunikace. V případě, že má útočník přístup k datovému toku uživatele například při sdílení jedné sítě snadno může datový tok zachytávat a analyzovat. Takovému jednání by mělo být zamezeno skrze zabezpečený a šifrovaný protokol HTTPS, ale i ten má svoje slabiny jak popisuje Miller et al. Při sledování velikosti šifrovaného datového toku a jeho prodlev a specifik jednotlivých webů dosahuje 90% přesnosti v identifikaci jednotlivých stránek uvnitř populárních webů.[35]

Tento přehled více či méně důmyslných identifikačních technologií ilustruje nástrahy, se kterými se uživatel musí vypořádat, aby byl na dnešním internetu anonymní. Obrana proti některých z nich je postupně implementována v moderních prohlížečích, proti některým je možné se bránit vyztužením prohlížeče pomocí doplňků a proti některým z nich se brání velmi těžko. Nástroj, který v současné době poskytuje uživatelům největší bezpečí vůči zmíněným hrozbám je Tor Browser, který blíže popisuji v následující kapitole.

2.3 The Onion Routing

2.3.1 Co je to Tor?

The Onion Routing, zkráceně Tor, je metoda pro anonymní komunikaci skrze počítačovou síť. Jak popisuji detailně v další části práce, Onion routing vzniká v druhé polovině 90. let 20. století na půdě amerického námořnictva konkrétně v U.S. Naval Research Laboratory. Jedná se o kolektivní dílo autorů Paula Syversona, Michaela G. Reeda, a Davida Goldschlaga - v té době

---

and Communications Security - CCS '14 [online]. New York, New York, USA: ACM Press, 2014, , 674-689 [cit. 2017-07-17]. DOI: 10.1145/2660267.2660347. ISBN 9781450329576. Dostupné z: http://dl.acm.org/citation.cfm?doid=2660267.2660347

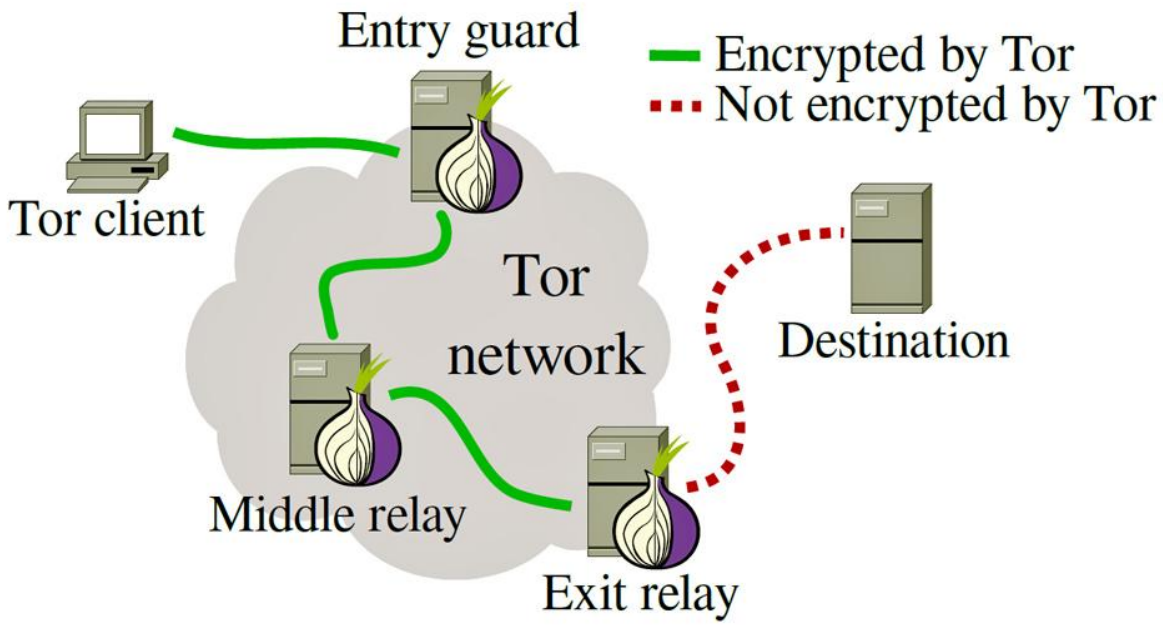
[35] MILLER, Brad, Ling HUANG, A. D. JOSEPH a J. D. TYGAR. I Know Why You Went to the Clinic: Risks and Realization of HTTPS Traffic Analysis [online]. , 143 [cit. 2017-07-17]. DOI: 10.1007/978-3-319-08506-7_8. Dostupné z: http://link.springer.com/10.1007/978-3-319-08506-7_8

zaměstnanců výzkumné laboratoře. Jejich hlavním úkolem bylo zabezpečit armádní důvěrnou komunikaci online.[36]

Později se k vývoji přidávají počítačoví vědci Roger Dingledine a Nick Mathewson, kteří spolu s Syversonem vyvíjí implementaci Onion routingu, ze které vzniká největší anonymizační síť Tor, The Onion Routing. Za účelem vývoje této sítě zakládají Dingledine a Mathewson spolu s dalšími pěti spoluzakladateli neziskovou organizaci The Tor Project, za finanční podpory Electronic Frontier Foundation a dalších dárců.[36]

Celý koncept Onion routingu je založen na směrování datového toku skrze síť dobrovolníků, datový tok je navíc obalen několika slupkami kryptografie, které jsou postupně odlupovány, až je původní datový tok komunikován s vnějším internetem. Ze zmíněného odlupování kryptografických vrstev je odvozen název onion, tedy cibule, která je rovněž k nalezení i v logu projektu.

Obrázek 2: Schéma Tor sítě (Tor Project, 2017)



[36] SYVERSON, Paul. Onion Routing: Brief Selected History [online]. [cit. 2017-07-04]. Dostupné z: https://www.onion-router.net/History.html

Clá síť se skládá z takzvaných *Entry guards*, *Middle relays* a *Exit relays* a *Bridges*. Role a funkcionalita Entry guard je v oficiální dokumentaci vysvětlena následovně[37]: „Tor, stejně jako ostatní nízko latentní anonymizační sítě, selhává v případě, že útočník kontroluje oba konce komunikačního kanálu. Předpokládejme například, že útočník kontroluje, nebo sleduje Tor relay, pomocí které vstupujete do sítě a zároveň kontroluje, nebo sleduje stránku, kterou navštěvujete. V tomto případě není známo jak zastavit útočníka ke spojení si informací o datovém toku a načasování z obou konců komunikačního kanálu."

V takovém případě, za předpokladu, že útočník kontroluje určitý počet uzlů **C** z celkového počtu všech uzlů sítě **N**. Když při každém použití sítě vybereme nový vstupní a výstupní bod sítě, útočník bude schopný korelovat všechen datový tok, který odešleme, s pravděpodobnostní okolo $\mathbf{(C/N)^2}$. Obecné profilování uživatelů ale může být pro většinu uživatelů stejně nežádoucí jako být sledován neustále a při častém použití se nebezpečí detekování uživatele zvyšuje. Proto využívání velkého počtu vstupních a koncových uzlů nedává uživateli možnost útěku před takovýmto útočníkem.

Řešením tohoto problému jsou *Entry guards*. Každý uživatel Tor sítě si vybere náhodně několik uzlů jako vstupních bodů a používá tyto relays pro připojení k dalším uzlům. Pokud tyto relays nejsou kontrolovány nebo pozorovány, útočník nemůže být nikdy úspěšný a uživatel je v bezpečí. V opačném případě, kdy jsou relays kontrolovány, nebo pozorovány útočníkem, může být uživatel vystaven profilování. Ne však většímu, než bez těchto opatření, kdy uživatel má reálnou šanci vyjádřenou uvedeným vzorcem (N-C)/N vyhnout se profilování úplně, zatímco předtím neměl žádnou.

Takováto limitace vstupních bodů do sítě skrze Entry guards rovněž brání potenciálnímu útočníkovi, který by provozoval menší počet relays a chtěl by zjistit všechny IP adresy uživatelů připojujících se k síti. V takovémto případě se dostane pouze ke zlomku uživatelů.

Po připojení do sítě přeposílá Entry Guard datový tok takzvané Middle relay. Middle relay se technicky příliš neliší od Entry Guards, což jsou v podstatě Middle relays náhodně vybrané, aby sloužily jako Entry Guard pro určité uživatele. Narozdíl od Entry Guard, ale nemá Middle relay

[37] Tor Project: What are Entry Guards? [online]. [cit. 2017-04-09]. Dostupné z: https://www.torproject.org/docs/faq.html.en#EntryGuards

přímé spojení s uživatelem, takže nezná jeho IP adresu ani jiné informace o datovém toku, který dále přeposílá do Exit relay.

Exit relay slupuje poslední slupku kryptografie a má k dispozici datový tok uživatele, ale díky předchozím dvou skokům uvnitř sítě nemá žádné informace o původci těchto dat. Exit relay pak komunikuje uživatelský požadavek vnějšímu internetu a stejným způsobem posílá zpět odpověď.

Zatímco provozovatelům Middle relays nehrozí žádný postih za potenciálně maligní datový tok, protože jej pouze přeposílají uvnitř sítě, operátoři Exit relays již musejí incidenty a stížnosti řešit, neboť IP adresy jejich serverů jsou uvedeny pod uživatelskými požadavky a působí na první pohled jako původci tohoto datového toku.

Pro rekapitulaci uvádím, že princip Tor sítě je založen na směřování datového toku uživatele skrze tři uzly. První uzel zná IP adresu uživatele, nicméně neví nic o obsahu jeho požadavku, který je obalen několika slupkami kryptografie. Druhý uzel nezná informace o uživateli a i přestože slupuje jednu vrstvu kryptografie stále neví nic o datovém toku. Konečně třetí uzel slupuje i poslední vrstvu kryptografie a komunikuje uživatelský požadavek vnějšímu internetu a odpověď zasílá zpět, bez znalosti uživatelských informací.

Informace o všech relays v síti schraňují adresářové autority, provozované lidmi z blízkosti vývojového týmu a organizace Tor Project. Těchto serverů je v současné chvíli 9[38] a jsou rozmístěny ve Švédsku, Rakousku, Německu, ve Spojených státech amerických a Holandsku. Tyto adresářové autority sbírají informace o všech relays v síti a poskytují je uživatelům, aby se byli schopni připojit do sítě.

Jedinou výjimkou jsou takzvané *Bridges*. Jsou to relays, které jsou neveřejné z důvodu ztížení jejich zablokování z řad cenzorů například v Číně. Uživatel ve chvíli, kdy se připojuje skrze síť, kde je Tor blokovaný, si může speciálním mechanismem skrze Gmail nebo Twitter vyžádat tuto neveřejnou adresu a připojit se přes ni. Gmail, nebo Twitter je zde využit z důvodu jejich vlastních mechanismů ochrany před roboty, kterých tak nepřímo využívá i Tor síť.

[38] Tor Project: Atlas [online]. [cit. 2017-04-09]. Dostupné z: https://atlas.torproject.org/#search/flag:authority

2.3.2 Historie Tor sítě

Počátky softwaru Tor sahají do roku 1995, kdy začínají prvotní práce na návrhu a technických specifikacích *Onion routingu*, stavebního kamene celého řešení. Nutno podotknout, že se Onion routing ve svých počátcích poměrně výrazně lišil od Toru jak jej známe dnes. Současná verze je vytvořena na základě pozdější verze specifikace, která je často označována jako druhá generace a ve které jsou implementovány významné změny.[39] Tento prvotní vývoj od roku 1995 do roku 2004 popisuje na svém webu[40] Paul Syverson, jeden z původních tvůrců a dodnes zaměstnanec Naval Research Laboratory, kde Onion routing vznikl a byl zároveň financován. Fakt, že kořeny Toru jsou v Naval Research Laboratory, tedy americké armádě, pravidelně vyvolává paniku a živí konspirační teorie. Původní zadání a potřeba zabezpečit vojenskou komunikaci před odposloucháváním a analýzou provozu ze strany armády je ale vcelku logická.

V květnu roku 1996 je poprvé formálně představen tento koncept a předveden první prototyp na systému Solaris 2.5.1/2.6 skládá se z 5 uzlů umožňujících prohlížení internetu. Již v tento moment se objevuje potřeba otevřeného zdrojovém kódu z důvodů důvěryhodnosti a bezpečnosti. Jak Paul Syverson poznamenává, termín open-source se v této době běžně nepoužíval, nicméně veškerý kód, který v květnu existoval, byl uvolněn široké veřejnosti o dva měsíce později.[40]

O rok později začíná projekt spolufinancovat Defense Advanced Research Projects Agency, známá rovněž pod zkratkou DARPA, pro svůj velký podíl na vytvoření internetové sítě. Následně je publikován téměř finální design první generace na IEEE Symposium on Security and Privacy. Roku 1998 již funguje síť o 13 uzlech rozprostřených mezi Naval Research Laboratory, Naval Research And Development a University of Maryland. Další dvě sítě jsou vytvořeny jako proof-of-concept, jedna z nich na kanadském ministerstvu obrany.[40]

Jak dále Paul Syverson popisuje, vytvoření lokálního přesměrování téměř veškerého datového toku skrze Onion routing v rámci jejich organizace dosahovalo v pozdějších fázích milionu přístupů za měsíc. Později téhož roku Zero Knowledge Systems představují Freedom Network,

[39] Tor Project: About [online]. [cit. 2017-07-04]. Dostupné z: https://www.torproject.org/about/overview.html.en
[40] SYVERSON, Paul. Onion Routing: Brief Selected History [online]. [cit. 2017-07-04]. Dostupné z: https://www.onion-router.net/History.html

síť velmi podobnou principům Onion routingu ovšem s třemi hlavními odlišnostmi. Za prvé Onion routing fungoval skrze protokop TCP, kdežto Freedom Network skrze UDP. Za druhé Freedom Network byl komerční projekt financovaný ze soukromých peněz narozdíl od Onion routingu, který již v této době fungoval na dobrovolnické bázi. Za třetí pak Freedom Network využívá systému pro správu pseudonymů - jednak aby byla síť dostupná pouze platícím uživatelům a dále aby umožňovala perzistentní pseudonymní komunikaci. Freedom Network byla provozována od roku 1999 do 2001, kdy byla ukončena z důvodu komerčního neúspěchu.

V roce 1998 dochází k publikování nejvíce detailního popisu Onion routingu první generace ve vědecké publikaci *Anonymous Connections and Onion Routing.*[41] Oproti druhé generaci postrádá například dopřednou bezpečnost, nebo pro každou aplikaci vytváří nové spojení a v jejím designu chybí adresářová služba.[40]

Navzdory dramatickému vývoji v letech předchozích v roce 1999 dochází k pozastavení vývoje z důvodu chybějícího financování a odchodu většiny vývojářů za jinými příležitostmi. Nicméně výzkum a analýzy nadále pokračují. V lednu roku 2000 je ukončena i testovací síť skládající se z 5 uzlů, které, jak Paul Syverson, popisuje, běžely na jediném stroji Sun Ultra 2 2170 se dvěma procesory na frekvenci 167 Mhz a 256 MB operační paměti. Za dva roky provozu tato sít zpracovala přes 20 miliónů požadavků z více než 60 zemí světa. Na podzim téhož roku představují badatelé z Technické univerzity Drážďany *Java Anon Proxy.*[42] Na rozdíl od zmíněné Freedom network se nejedná o klon Onion routingu, ale o další anonymizační systém, který funguje dodnes.[40]

V roce 2001 dochází k obnovení vývoje za finanční podpory DARPA s cílem dopracovat kód do beta fáze a spustit novou síť. V témže roce je udělena cena Edison Invention Award za vynález Onion routingu.[40]

Následující rok je původní kód první generace zanechán z důvodu jeho zastaralosti a začíná vývoj takzvané druhé generace. V říjnu roku 2003 je Tor síť spuštěna a kompletní kód zveřejněn pod MIT licencí. Vývoj sítě i kódu je organizován pod stránkou www.torproject.org,

[41] REED, Michael G., Paul F. SYVERSON a David M. GOLDSCHLAG. Anonymous Connections and Onion Routing. IEEE Journal on Selected Areas in Communication Special Issue on Copyright and Privacy Protection. 1998.
[42] JAP -- ANONYMITY & PRIVACY [online]. [cit. 2017-04-08]. Dostupné z: https://anon.inf.tu-dresden.de/index_en.html

která slouží tomuto projektu do dnešních dní. Koncem roku 2003 má síť desítky uzlů z řad dobrovolníků, primárně na území Spojených států a jeden v Německu. To jsou již součástí vývojového týmu Roger Dingledine a Nick Mathewson, kteří jsou jedněmi z hlavních vývojářů dodnes.[40]

Těmito událostí se začíná psát novodobá historie Toru. Onion Services (dříve Hidden services), tedy technologie pro ukrytí fyzické polohy serveru uvnitř sítě, jsou spuštěny na jaře v roce 2003, nezisková organizace Electronic Frontier Foundation začíná spolufinancovat vývoj a koncem roku 2004 Tor síť čítá přes 100 uzlů rozesetých mezi třemi kontinenty.[40]

V prosinci roku 2006 základá Dingledine a Mathewson spolu s dalšími pěti spoluzakladateli neziskovou organizaci ve státě Massachusetts The Tor Project, Inc. zodpovědnou za udržování a vývoj Toru.[32]

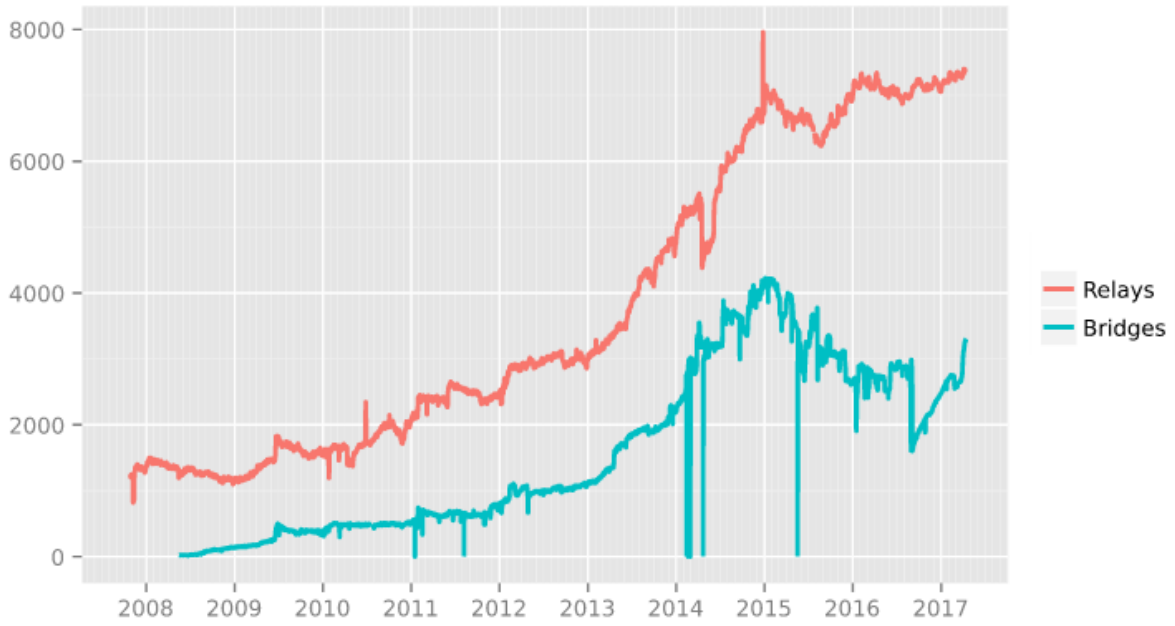
2.3.3 Tor v současné době

Od roku 2006 do současnosti se o vývoj neustále stará nezisková organizace The Tor Project, Inc. s Rogerem Dingledinem a Nickem Mathewsonem coby hlavními vývojáři a okolo softwaru Tor vyrostl celý ekosystém dalších nástrojů.

I po více než 10 letech vývoje, kdy bude brzy uvolněna první stabilní verze z vývojové větve 0.3[43], zůstává princip stále stejný a provoz sítě stojí na tisícovkách dobrovolníků, kteří přispívají výpočetním výkonem a hlavně konektivitou k fungování celé sítě.

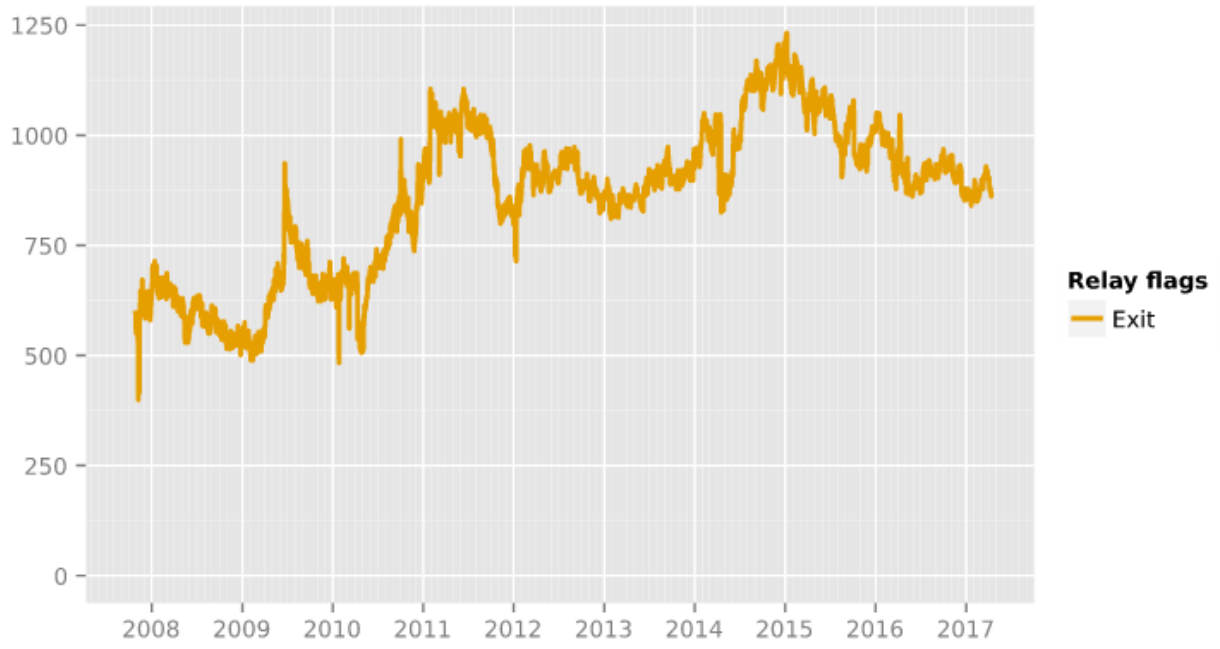
[43] Tor 0.3.0.5-rc is released: almost stable!. Tor Project [online]. [cit. 2017-04-15]. Dostupné z: https://blog.torproject.org/blog/tor-0305-rc-released-almost-stable

Obrázek 3: Počet Relays a Bridges v čase (Tor Project, 2017)



Jak je z přiloženého grafu patrné počet uzlů provozovaných dobrovolníky se blíží osmi tisícům a stejně tak počet Bridgů, tedy neveřejných uzlů, čítá přes tři tisíce.

Obrázek 4: Graf počtu Exit uzlů v čase (Tor Project, 2017)



Počet Exit uzlů komunikujících s vnějším internet je pak dlouhodobě okolo jednoho tisíce, jak dokládá přiložený graf.

2.3.3.1 Tor Browser Bundle

Kromě samotného softwaru Tor, který zůstává základním stavebním kamenem pro ostatní projekty, je jedním z těch hlavních takzvaný Tor Browser Bundle. Upravený prohlížeč, který se automaticky připojí do Tor sítě a zpřístupňuje tak díky jednoduchosti použití anonymizační síť širšímu publiku.

Tor Browser Bundle je v současné chvílí založen na webovém prohlížeči Firefox od společnosti Mozilla, přesněji na vývojové větvi s prodlouženou podporou ESR (Extended Support Release).[44] Oproti standardnímu Firefoxu je provedeno několik změn v kódu za účelem větší bezpečnosti a ochrany soukromí uživatelů. Chování prohlížeče je dále upraveno skrze doplněk TorButton, který se stará o zabezpečení na úrovni aplikace a rovněž blokuje mnoho typů

[44] Extended Support Release. Firefox [online]. [cit. 2017-04-15]. Dostupné z: https://www.mozilla.org/en-US/firefox/organizations/

aktivního obsahu. V neposlední řadě je provedeno několik změn oproti defaultnímu nastavení Firefoxu, rovněž za účelem většího bezpečí uživatelů.

O start aplikace a připojení do sítě se stará prvek zvaný Tor Launcher, který rovněž nabízí pohodlnou prvotní konfiguraci a v neposlední řadě dává uživateli možnost nepřipojovat se do sítě napřímo, ale skrze Bridge v případě, že se nachází v prostředí, kde je Tor aktivně blokovaný.

Z důvodu ochrany uživatelů před případným odposloucháváním ze strany Exit nodů je implementován doplněk HTTPS-Everywhere, který vynucuje komunikaci v zabezpečeném protokolu HTTPS namísto nezabezpečeného HTTP, všude tam, kde je to jen možné.[45]

Jako poslední doplňuje Tor Browser Bundle add-on NoScript, který chrání uživatele před nevyžádaným spuštěním JavaScriptu, Javy, Flashe a dalších zásuvných modulů v prohlížeči, které by mohly znamenat bezpečnostní hrozbu pro uživatele.[46]

Tor Browser Bundle rovněž integruje několik algoritmů souhrnně označovaných jako Pluggable Transports, které modifikují charakteristický datový tok Toru takovým způsobem, aby bylo možné se vyhnout blokování a cenzurním snahám ze strany poskytovatelů připojení.

Obecné nároky na bezpečné a anonymní prohlížení webu jsou rozděleny do dvou typů: bezpečnostních a těch s ohledem na soukromí. Tyto požadavky jsou v dokumentaci[47] Tor Browseru detailně popsány a zahrnují v sobě mnoho komplikací, které je nutno překonat pro zajištění uživatelské anonymity v co největší možné míře.

[45] Electronic Frontier Foundation: HTTPS Everywhere [online]. [cit. 2017-07-04]. Dostupné z: https://www.eff.org/https-everywhere
[46] NoScript [online]. [cit. 2017-07-04]. Dostupné z: https://noscript.net/
[47] Tor Browser: The Design and Implementation of the Tor Browser [DRAFT] [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/

2.3.3.1.1 Bezpečnostní požadavky

Bezpečnostní požadavky se týkají především zajištění bezpečného používání Toru. Porušení těchto požadavků má zpravidla za následek závažné riziko pro uživatele a často znamená okamžitou ztrátu anonymity.

1. **Dodržení proxy nastavení**

Jedním z hlavních bezpečnostních požadavků na prohlížeč je jeho schopnost komunikovat veškerý datový tok skrze Tor proxy, protože prohlížeč jednoduše nesmí obejít toto nastavení pro žádný obsah, neboť to znamená prozrazení reálné IP adresy uživatele a okamžitou ztrátu anonymity.[48]

V případě Firefoxu je veškerý datový tok směřován skrze SOCKS proxy, což znamená mimo jiné zakázání WebRTC protokolu, úpravu implementace DNS, vypnutí WebIDE, nebo mDNS, které používá UDP pakety.

Rovněž jsou blokovány standardní zásuvné moduly, mezi které patří například Flash, který má schopnost provádět libovolné systémové dotazy mimo prohlížeč a obejít tak nastavení proxy.

Vývojáři tak garantují, že díky těmto změnám je správně směrováno HTTPS, OCSP, HTTP, FTP, DNS, veškerý JavaScript včetně HTML5 audio a video objektů a mnoho dalších. Naopak jsou blokovány různé nestandardní protokoly, tak aby bylo potenciální riziko omezeno na minimum.

[48] Tor Browser: Proxy Obedience [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#proxy-obedience

2. Oddělení stavu

Druhým bezpečnostním požadavkem je vymanění se z veškerých předešlých stavů vzniklých předchozích použitím prohlížeče, což rovněž zahrnuje i stavy pluginů třetích stran, nebo podpůrných systémových knihoven jako je TLS.[49]

Tor Browser se snaží toto riziko eliminovat za použití vlastního Firefox profilu a nastavení domácí proměnné do kořenové složky Tor Browseru. Prohlížeč také nenačte žádná celosystémová rozšíření, což opět zabraňuje například v případě Flashe uniknutí Flash cookie.

3. Nezapisovat na disk

Třetí bezpečnostní požadavek je definován tak, že prohlížeč nesmí zapsat na disk a uchovávat jakékoliv informace spojené s prohlížením webu nad rámec dané relace, pokud si tak uživatel explicitně nepřeje a nechce uchovat svoji historii procházení na disku.[50]

Tohoto cíle je dosaženo za použití několika mechanismů. Tím hlavním je spouštění Tor Browseru v módu privátního okna, což samo o sobě řeší část požadavků a navíc je vypnuta mezipamět pro multimediální obsah, tak aby se HTML5 videa nemohla zapisovat na disk. V neposlední řadě je vypnuta knihovna asm.js, jejíž mezipamět se i v privátním módu zapisuje na disk.[51]

4. Izolace dat aplikace

Posledním z bezpečnostních požadavků je ten na izolování data aplikace.[52] Komponenty, které zajišťují poskytování anonymního prohlížení, musí být soběstačné, nebo musí poskytnout mechanismus pro rychlé a úplné odstranění všech důkazů o používání režimu. Jinými slovy,

[49] Tor Browser: State Separation [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#state-separation
[50] Tor Browser: Disk Avoidance [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#disk-avoidance
[51] Asmjscache: should not store cache entries when private browsing is enabled. BugZilla [online]. [cit. 2017-04-16]. Dostupné z: https://bugzilla.mozilla.org/show_bug.cgi?id=1047105
[52] Tor Browser: Application Data Isolation [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#app-data-isolation

prohlížeč nesmí zapsat nebo způsobit, že operační systém zapíše jakékoliv informace o používání anonymního prohlížení na disk mimo kontrolu aplikace.

2.3.3.1.2 Požadavky s ohledem na soukromí

Požadavky s ohledem na soukromí je soubor vlastností, které ovlivňují preference ohledně použití konkrétního prohlížeče, kterým je v současnosti již zmiňovaný Firefox.

1. **Nespojitelnost na základě identifikátorů**

Prvnímu z požadavků na soukromí je vyhověno skrze izolaci veškerých identifikačních dat napříč prohlížečem. Tato izolace znamená asociaci stavů a oprávnění s URL v adresním řádku prohlížeče tak, aby bylo zabráněno spojitelnosti uživatele na základě identifikačních dat třetími stranami.

Jak je dále uvedeno v dokumentaci,[53] mnoho z těchto dat může sloužit jako identifikátor a žádný z výrobců prohlížečů prozatím neinvestoval dostatečné úsilí k výčtu, nebo vypořádání se se sledováním uživatelů třetími stranami na základě cookies a dalších identifikátorů.

2. **Nespojitelnost na základě fingerprintingu**

Druhý z požadavků se týká takzvaného fingerprintingu, který odkazuje na anglické slovo fingerprint, neboli otisk prstu. Oproti předchozí kategorii, kde jsou identifikační informace ve větší či menší míře pod kontrolou uživatele, tato metoda funguje na základě digitálního otisku prohlížeče, do kterého se promítají všechna nastavení jak prohlížeče, tak celého systému.[54]

Ve větším detailu se problematikou fingerprintingu zabývám v předchozích kapitolách, nicméně specifikem Tor Browseru, do kterého vývojáři investovali velké úsilí, je ochrana před fingerprintigem skrze extrakci obrazu z HTML5 canvasu. Prvek canvas je součástí HTML5,

[53] Tor Browser: Cross-Origin Identifier Unlinkability [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#identifier-linkability
[54] Tor Browser: Cross-Origin Fingerprinting Unlinkability [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#fingerprinting-linkability

který dovoluje dynamicky vykreslovat 2D objekty a bitmapy[55] a pro získání unikátního otisku prohlížeče tak stačí ve zlomku vteřiny vygenerovat WebGL objekt, font a barvu a následně získat hash takovéhoto obrázku. Minimální odlišnost v grafické kartě, fontu, verzích softwaru a ovladačů dovoluje vygenerovat stabilní identifikátor zařízení, který se kvalitou velmi blíží cookie.[56]

Vývojáři proto v rámci Tor Browseru implementovali vyskakující okno, které se před extrakcí HTML5 canvasu dotazuje uživatele, zda si tak přeje učinit a zároveň je extrakce canvasu třetími stranami striktně zakázána.[57]

3. Dlouhodobá nespojitelnost

Posledním z požadavků na ochranu soukromí uživatelů je dlouhodobá nespojitelnost dle definice Adrease Pfitzmanna z úvodu práce. Tomuto požadavku se vývojáři snaží vyhovět formou tlačítka *New Identity*, které je součástí Tor Browseru. Tato funkce má za úkol smazat veškeré identifikátory a nastavení a skrze nové připojení do Tor sítě a otevřením nového okna tak de facto novou identitu uživateli opatřit.[58]

2.3.4 Užívání Tor sítě

Jako kteroukoliv jinou technologii včetně tužky nebo mobilního telefonu, je anonymitu možné využít za dobrým stejně tak jako za účelem špatným. Dennodenně se setkáváme s debatou o větší kontrole osob ve spojitosti s migrací a terorismem a s ní spojenými snahami o eliminaci anonymity jak z prostoru fyzického tak virtuálního. Jak tvůrci Toru uvádějí, Tor Project je

[55] W3schools: HTML5 Canvas [online]. [cit. 2017-04-17]. Dostupné z: https://www.w3schools.com/html/html5_canvas.asp
[56] MOWERY, Keaton a Hovav SHACHAM. Pixel Perfect: Fingerprinting Canvas in HTML5 [online]. , 12 [cit. 2017-04-17]. Dostupné z: https://cseweb.ucsd.edu/~hovav/dist/canvas.pdf
[57] Bug #6253: Add canvas image extraction prompt. [online]. [cit. 2017-04-17]. Dostupné z: https://gitweb.torproject.org/tor-browser.git/commit/?h=tor-browser-45.8.0esr-6.5-2&id=526e6d0bc5c68d8c409cbaefc231c71973d949cc
[58] Tor Browser: Long-Term Unlinkability [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#new-identity

založen na přesvědčení, že anonymita není dobrým nápadem pouze část času, ale že její existence je předpokladem pro svobodnou a fungující společnost.[59]

Jak je z povahy fungování tohoto nástroje patrné, v případě, že nástroj správně zafunguje, nikdo se o uživateli nedozví a neexistuje tak mnoho konkrétních pozitivních příběhů. Naopak média velmi rychle adoptovala označení Darknet pro Onion Services, kvůli ilegálním aktivitám, které se na nich rovněž mohou odehrávat.

Tor ale často používají i **obyčejní lidé**, kteří nemají co skrývat, nicméně jsou opatrní ohledně svého soukromí. Například IP adresy je možné v dnešní době vystopovat až na úroveň jednotlivých ulic a s daty o uživatelích se běžně obchoduje. Tor totiž kromě posílení soukromí uživatelů rovněž šifruje veškerý datový tok a je tak možné se díky němu chránit před útočníky stojícími mezi uživatelem a serverem, s kterým komunikuje. I z tohoto důvodu je populární sociální síť Facebook dostupná uvnitř Tor sítě na adrese *https://facebookcorewwwi.onion*, kde je Tor použit pro větší zabezpečí uživatelů jako přidaná bezpečnostní vrstva.[60]

Další cílovou skupinou jsou **novináři** a jejich publikum. Francouzská organizace Reportéři bez hranic eviduje internetové vězně svědomí (*Prisoner of conscience* - termín použitý Peterem Benensonem může odkazovat se na každého, kdo je uvězněn kvůli své rase, sexuální orientaci, náboženskému vyznání nebo za své politické názory)[61] a doporučuje novinářům, jejich zdrojům, bloggerům a disidentům používat Tor, aby bylo zajištěno jejich soukromí a bezpečnost.[62] Vývoj Toru rovněž finančně podporuje americká organizace International Broadcasting Bureau[63] provozující mimo jiné Rádio Svobodná Evropa. A to z důvodu umožnění uživatelům internetu v zemích bez bezpečného přístupu k svobodným sdělovacím prostředkům přístup k globální perspektivě na kontroverzní témata, včetně demokracie, ekonomiky a náboženství.

[59] Tor Project: Users of Tor [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/about/torusers.html.en
[60] MUFFETT, Alec. Making Connections to Facebook more Secure [online]. [cit. 2017-04-17]. Dostupné z: https://www.facebook.com/notes/protect-the-graph/making-connections-to-facebook-more-secure/1526085754298237
[61] Prisoner of conscience. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-04-17]. Dostupné z: https://en.wikipedia.org/wiki/Prisoner_of_conscience
[62] Reporters sans frontières: Online survival Kit [online]. [cit. 2017-04-17]. Dostupné z: https://rsf.org/en/online-survival-kit
[63] Tor: Sponsors [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/about/sponsors.html.en

Státní orgány včetně policie, armády, nebo soudů rovněž Tor využívají a z historického hlediska byl právě za těmito účely vyvinut. Shromažďování důkazů a monitorování podezřelých online aktivit může být snadno prozrazeno třeba jen na základě několikanásobného připojení z IP adres asociovaných právě s vyšetřujícím orgánem. Ovie Carroll, ředitel kyberkriminální laboratoře na americkém ministerstvu spravedlnosti, nedávno vyzval federální soudce, aby Tor používali pro ochranu citlivých informací na soukromích i pracovních zařízeních.[64]

Aktivisté a **whistlebloweři** jsou další přirozenou uživatelskou skupinou Toru. Podobně jako u novinářů mohou díky anonymizační síti publikovat svoje myšlenky bez obav z represe. Přesně za tímto účelem Tor používali protivládní demonstranti v Egyptě během Arabského jara v roce 2011. V zemi byl během těchto událostí blokován Twitter a jeho případné použití monitorováno. Uživatelé tak použili Tor k obejití blokace a následně mohli anonymně sdílet informace, za předpokladu použití účtu, který s daným jedincem nebyl spojitelný.[65]

Ve velmi podobné situaci se teď ocitají aktivisté v Turecku, kde jsou dlouhodobě sociální sítě blokovány a zároveň monitorovány. Na seznamu blokovaných webů se nedávno ocitla i internetová encyklopedie Wikipedia,[66] jejíž blokování je zpravidla doménou nedemokratických zemí.[67]

Poslední z hlavních uživatelských skupin jsou **IT profesionálové**. Díky vlastnostem Toru jej mohou využít k ověření nastavení svých systémů, firewallů, nebo testovat možnosti jak obejít blokace. V případě výpadků DNS serverů na straně poskytovatelů internetu mohou Tor využít k navázání spojení, což může být v krizových situacích neocenitelné.

Výčet uživatelských skupin není definitivní a příslušnost uživatelů k jednotlivým skupinám se může překrývat a v čase měnit. IT profesionál může Tor v práci využít k testování firewallu, ale v

[64] Department of Justice Official Tells Hundred Federal Judges to Use Tor [online]. [cit. 2017-04-17]. Dostupné z: https://motherboard.vice.com/en_us/article/department-of-justice-official-tells-hundred-federal-judges-to-use-tor
[65] ZAHORSKY, Ingmar. Tor, Anonymity, and the Arab Spring: An Interview with Jacob Appelbaum [online]. [cit. 2017-04-30]. Dostupné z: http://www.monitor.upeace.org/innerpg.cfm?id_article=816
[66] Turkey blocks Wikipedia under law designed to protect national security [online]. [cit. 2017-04-30]. Dostupné z: https://www.theguardian.com/world/2017/apr/29/turkey-blocks-wikipedia-under-law-designed-to-protect-national-security
[67] Censorship of Wikipedia [online]. [cit. 2017-04-30]. Dostupné z: https://en.wikipedia.org/wiki/Censorship_of_Wikipedia

případě potřeby jej může využít za účelem získání přístupu k jinak blokovaným informacím a službám, nebo jen využívat zvýšené bezpečnosti plynoucí ze šifrování, které Tor využívá.

Tito uživatelé se ale bohužel často setkávají s druhořadým zacházením při prohlížení webu. Několik velkých hostingových společností přistoupilo k plošnému blokování přístupů z Tor sítě na základě zneužívání a útoků od části uživatelů, kteří Tor nepoužívají k bohulibým účelům. V lepším případě se uživatelé setkávají s ověřováním jejich lidskosti ve formě CAPTCHA testu, nebo je jejich přístup kompletně blokovaný.

Skupina výzkumníků okolo Sheharbano Khattaka[68] se pokusila ověřit, jak časté toto blokování anonymních uživatelů je a došla k závěru, že minimálně 1,2% všech webů blokuje Tor a z žebříčku Alexa Top 1000[69] je to dokonce 3,67%. Mezi subjekty komerčně nabízející blokování Toru se pak objevují anti spamové společnosti NForce a Voxility, které zmiňuji i v pozdější části práce jako časté původce stížností.

2.3.5 Zneužívání Tor sítě

Anonymní síť Tor je často skloňována ve spojitosti s ilegální činností. Ze samotné podstaty svého fungování poskytuje stejné služby aktivistům a novinářům na straně jedné a hackerům a obchodníkům s ilegálním zbožím na straně druhé. Není vůbec jednoduché kvantifikovat, v jaké míře je síť zneužívána k nelegálním činnostem, neboť spolehlivá data, která by takovéto závěry podložila, neexistují.

Společnost poskytující bezpečnostní webové služby CloudFare vydala v březnu roku 2016 prohlášení, ve kterém uvádí, že 94% veškerých požadavků pocházejících z Tor sítě je maligních. Dále uvádějí: “To neznamená, že navštěvují kontroverzní obsah, nýbrž jsou to automatizované požadavky, které mají poškodit naše zákazníky. Velké procento tvoří spamového komentáře, skenování zranitelností, podvodné klikání na reklamy, scrapování obsahu a skenování přihlášení, to vše přichází prostřednictvím sítě Tor. Pro uvedení do

[68] KHATTAK, Sheharbano, et al. Do you see what i see? differential treatment of anonymous users. In: Network and Distributed System Security Symposium. 2016. [cit. 2017-05-08]. Dostupné z: https://www.cl.cam.ac.uk/~sk766/publications/ndss16_tor_differential.pdf
[69] Alexa: Top sites [online]. [cit. 2017-05-08]. Dostupné z: http://www.alexa.com/topsites

kontextu, na základě údajů z projektu Honey Pot, 18% globálního e-mailového spamu nebo zhruba 6,5 bilionů nežádoucích zpráv za rok začíná automatizovanými sběrem e-mailových adres pomocí sítě Tor."[70]

Mike Perry z Tor Project s tímto tvrzením polemizuje, neboť 94% se zdá jako velmi vysoké číslo. Hádá, že k tomuto číslu mohli dojít chybnou metodologií, ve které je každá IP adresa, která kdy zaslala byť jediný spam nadále identifikována a datový tok z ní pocházející označen jako maligní.[71] Tuto domněnku potvrzuje již dříve zmiňovaný výzkum[45] zaměřený na dvojí zacházení s anonymními uživateli. Tento výzkum uvádí, že u nového exit serveru trvá v průměru 30 dnů, než dojde k události, při které získá špatnou reputaci. Jakmile se ale tak stane, po celou dobu studie se zlé reputace nezbaví a je blokován řadou služeb jakou je například CloudFare a jsou tak de facto trestání i nevinní uživatelé.

Konkurenční společnost Akamai ve svém reportu ohledně internetové bezpečnosti z roku 2015 zmiňuje velmi odlišná čísla.[72] V případě požadavků přicházejících z Tor sítě uvádí, že pouze jeden z 380 požadavků je maligní. Tento poměr je stále mnohonásobně vyšší v porovnání s požadavky mimo Tor síť, u kterých je přibližně pouze jeden požadavek z 11 500 závadný.

Tento report rovněž uvádí rozložení útoků z Tor sítě dle odvětví, do kterého spadá cíl.

Tabulka 2: Cíle útoků dle odvětví (Akamai, 2015)

Kategorie	Počet útoků	Frekvence
Maloobchod	212 189	35.60%
Finanční služby	156 760	26.30%
Technologie	123 442	20.71%
Média a zábava	49 834	8.36%
Veřejný sektor	34 800	5.84%

[70] PRINCE, Matthew. The Trouble with Tor [online]. [cit. 2017-05-08]. Dostupné z: https://blog.cloudflare.com/the-trouble-with-tor/
[71] PERRY, Mike. The Trouble with CloudFlare [online]. [cit. 2017-05-08]. Dostupné z: https://blog.torproject.org/blog/trouble-cloudflare
[72] AKAMAI. State of the Internet / Security Report [online]. [cit. 2017-05-08]. Dostupné z: https://media.scmagazine.com/documents/144/q2_2015_soti_security_report_-_35820.pdf

Ubytování a cestování	5 919	0.99%
Obchodní služby	5 241	0.88%
Automobilový průmysl	3 942	0.66%
Spotřební zboží	2 767	0.46%
Herní průmysl	813	0.14%
Různý	335	0.06%

Po dobu trvání studie byly nejpopulárnějším cílem stránky z kategorie retail, tedy maloobchod, následovány weby z kategorie finančních služeb a technologií.

Na závěr měřili autoři studie počty požadavků na klíčové aplikace související s obchodem jako jsou pokladny v e-shopech a platební brány. Stejně jako Tor může představovat bezpečnostní riziko, tak se rovněž může jednat i o obchodní příležitost neboť z Tor sítě bylo evidováno více než 35 milionů požadavků. Pokud toto číslo vyjádříme konverzním poměrem, který indikuje v jakém poměru se návštěvníci webu stanou zákazníky (provedou registraci, nákup apod.) jedná se o 1:895, který je velmi blízko konverznímu poměru mimo Tor síť, který je 1:834.

2.3.5. Onion Services (Hidden Services)

Navzdory tomu, že Onion Services (dříve Hidden Services) tvoří pouze velmi malou součást Tor sítě, jsou často s celou sítí zaměňovány pod označením Darknet. Toto označení se velmi rychle zpopularizovalo v souvislosti s kauzou nelegálního tržiště Silk Road, které bylo v rámci Onion Services provozováno.

Tato kauza začala v říjnu roku 2013, kdy americká FBI zablokovala virtuální tržiště Silk Road a zatkla údajného provozovatele této stránky Rosse Ulbrichta, který vystupoval pod přezdívkou Dread Pirate Roberts. Silk Road bylo anonymní internetové tržiště, do velké míry podobné například eBay, s tím rozdílem, že zde bylo nabízeno převážně nelegální zboží a platidlem byla kryptoměna Bitcoin. Prodejcem se mohl stát kterýkoliv návštěvník a z jeho prodejů plynula

provozovateli malá provize. Nejoblíbenějším artiklem z nelegálního sortimentu byly hlavně omamné látky, ale k dostání byly i zbraně, nebo služby jako DDoS útoky a hacking.[73]

Silk Road samozřejmě nebylo jediným tržištěm a dodnes jich existují desítky, nicméně velká medializace této kauzy, kdy se veřejnost dozvěděla o existenci tohoto zákoutí internetu, dostupného pouze v rámci Tor sítě, měla za následek velkou popularizaci i v negativním slova smyslu - vynesla Tor síti nálepku Darknet.

Nedávno publikovaná studie v British Journal of Criminology od sociologa Isaka Ladegaarda poskytuje důkazy o nárůstu prodejů na konkurenčním tržišti Agora v souvislosti s medializací kauzy Silk Road.[74] Díky tomu, že Agora vyžaduje uživatelskou zpětnou vazbu u každé transakce, Ladegaard je mohl zkombinovat s nabídkami a cenami, které scrapoval od listopadu roku 2014 a vypočítat tak reálný objem prodejů. Ty se hned po odsouzení Rosse Ulbrichta téměř ztrojnásobily v rámci denních tržeb z 40 tisíc dolarů na 100 tisíc pro prodeje v rámci USA a ze 100 tisíc na 250 tisíc dolarů pro mezinárodní transakce.

Ladegaard dále uvádí jasnou souvislost mezi medializací procesu a nárůstem tržeb. Domnívá se ale, že nárůst byl utlumen doživotním rozsudkem pro Rosse Ulbrichta, která měla na část potenciálních prodejců i zákazníků odstrašující účinek. Rovněž odhaduje, že v případě jeho osvobození by byl nárůst ještě masivnější.

Pro zasazení problému toho typu zneužívání zneužívání do kontextu je nutné kvantifikovat Onion Services jako celek a uvést jak velkou součástí Tor sítě jsou. Za tímto účelem sami autoři z Tor Project s pomocí několika dobrovolníků, kteří je nechali nahlédnout k datům ze svých relays a umožnili tak vidět aktivitu 2-5% ze všech Onion Services. Pomocí těchto dat a extrapolace došli autoři k závěru, že se denně 30 tisíc Onion Services ohlásí v Tor síti a společně vygenerují denní traffic o velikosti 5 terabytů. Tento datový tok tvoří přibližně 3,4% celkového datového toku a tedy majoritní většina 96,6% datového toku je mimo Onion Services.[75]

[73] Silk Road operator Ross Ulbricht sentenced to life in prison. The Guardian [online]. [cit. 2017-06-18]. Dostupné z: https://www.theguardian.com/technology/2015/may/29/silk-road-ross-ulbricht-sentenced
[74] LADEGAARD, Isak. We Know Where You Are, What You Are Doing and We Will Catch You. The British Journal of Criminology. 2017, , -. DOI: 10.1093/bjc/azx021. ISSN 0007-0955. Dostupné také z: https://academic.oup.com/bjc/article-lookup/doi/10.1093/bjc/azx021
[75] Some statistics about onions [online]. [cit. 2017-05-27]. Dostupné z: https://blog.torproject.org/blog/some-statistics-about-onions

Výzkumníci Daniel Moore a Thomase Rid se pokusili analyzovat, jaký obsah je v jakém počtu uvnitř Onion Services zastoupen.[76] Za tímto účelem postahovali data z 5 205 webů uvnitř Tor sítě a pro následnou kategorizaci si nadefinovali tyto kategorie:

Tabulka 3: Definice kategorií (Moore, 2016)

Kategorie	Popis
Zbraně	Obchod se zbraněmi.
Drogy	Obchod nebo výroba drog, včetně nelegálně získaných medikamentů na předpis.
Extremismus	Obsah přijímající extremistickou ideologii, včetně ideologických textů, vyjadřování podpory terorismu, návodů pro militanty a extremistická fóra.
Finance	Praní špinavých peněz, falešné bankovky, obchod s ukradenými kreditními kartami a bankovními účty.
Hacking	Nájemní hackeři, obchod nebo distribuce malwaru a DDoS útoků.
Nezákonná pornografie	Pornografické materiály znázorňující děti, násilí, zvířata, nebo materiály pořízené bez souhlasu účastníků.
Nexus	Stránky, které primárně odkazují na jiné nezákonné stránky a materiály v rámci darknetu.
Ostatní nelegálnosti	Materiály, které snadno nepasují do žádné z ostatních kategorií, ale zůstávají problematické. Například prodej ilegálního zboží, falešných pasů a občanských průkazů.
Sociální	Online komunity pro sdílení nelegálního obsahu skrze fóra, sociální sítě a jiné diskuzní platformy.
Násilí	Nájemné vraždy a instruktážní materiál pro spáchání násilného činu.
Ostatní	Legální ideologický a politický obsah, mrtvé schránky, informační repozitáře a legitimní služby.
Žádný	Weby, které byly buď nedostupné, nebo bez vizuálního obsahu,

[76] MOORE, Daniel a Thomas RID. Cryptopolitik and the Darknet. Survival [online]. 2016, 58(1), 7-38 [cit. 2017-05-27]. DOI: 10.1080/00396338.2016.1142085. ISSN 0039-6338. Dostupné z: http://www.tandfonline.com/doi/full/10.1080/00396338.2016.1142085

	včetně webů obsahujících pouze placeholder indikující, že obsah bude teprve vytvořen.

První sadu postahovaných webů pak ručně kategorizovali a tato data použili k natrénování statistického Support Vector Machines[77] modelu pro strojovou kategorizaci zbylého obsahu.

Tabulka 4: Výsledky výzkumu Moore (Moore, 2016)

Kategorie	Počet
Žádný	2 482
Ostatní	1 021
Drogy	423
Finance	327
Ostatní nelegálnosti	198
Neznámý	155
Extremismus	140
Nezákonná pornografie	122
Nexus	118
Hacking	96
Sociální	64
Zbraně	42
Násilí	17
Celkově	5 205
Celkově aktivních	2 723
Celkově nelegálních	1 547

Dle studie bylo z celkové počtu 5 205 webů 2 723 aktivních (kategorie *Žádný* indikuje absenci jakéhokoliv obsahu) a 1 547 z nich pak kategorizováno jako nelegální.

Takovýto typ výzkumu může být snadno zkreslený, neboť neexistuje žádný konečný seznam všech serverů uvnitř Onion Services a záleží na analytikovi, ke kterým webům se dostane a

[77] Support vector machine. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-05-27]. Dostupné z: https://en.wikipedia.org/wiki/Support_vector_machine

které bude analyzovat. Jak uvádějí autoři obdobné studie Clare Gollnick a Emily Wilson, kvalita takovéhoto výzkumu je založena na kvalitě náhodného vzorku.[78]

Gollnick et al. analyzovali náhodný vzorek 400 adres z množiny vzniklé kontinuálním procházením Onion Services a obdobně jako v předešlém průzkumu si nadefinovali 15 kategorií pro rozdělení webů na základě jejich obsahu.

Tabulka 5: Výsledky výzkumu Gollnick (Gollnick a Wilson, 2017)

Kategorie	Podíl
Legální	47,7%
Neaktivní	17,7%
Drogy	12,3%
Explicitní obsah (pornografie)	6,8%
Více nelegálních kategorií	6,5%
Farmaka	3,2%
Ostatní nelegální obsah	1,5%
Hacking	1,3%
Zneužívání	1,0%
Stažitelný obsah	0,5%
Extremismus	0,2%

Na základě tohoto průzkumu je možné tvrdit, že majoritní většina aktivních webů nese legální obsah a z těch nelegálních dominují ty s drogovým obsahem, primárně virtuální tržiště, následovány pornografií a těmi, kde se překrývá více typů nelegálního obsahu na jednom webu.

Podobné studie jsou v tomto případě limitovány a mohou být snadno zkresleny kvalitou náhodného vzorku, nebo velkou proměnlivostí Onion Services. Do velké míry lze ovšem usuzovat, že zastoupení nelegálního obsahu je řádově vyšší než na klasickém internetu. V kontextu celé Tor sítě se ovšem jedná o méně než 5% celkového datového toku, který je jen z

[78] GOLLNICK, Clare a Emily WILSON. The Truth About the Dark Web [online]. [cit. 2017-05-27]. Dostupné z: https://www.scribd.com/document/329783168/The-Truth-About-the-Dark-Web

části nelegální. Z těchto důvodů se domnívám, že označení Darknet pro celou síť je nesprávné, neboť vystihuje pouze její velmi malou část.

V negativních konotacích je často vnímána i Tor síť jako celek, což se každému uživateli potvrdí při prohlížení webu skrze Tor Browser. Velké množství populárních webových služeb proaktivně vystavuje uživatele opakovaným ověřením, nebo úplným blokacím ze strachu z potenciálního zneužití.[79] Míra a typy zneužívání jsou pak hlavními výzkumnými otázkami pro následující praktickou část práce, ve které bych za pomoci získaných dat ověřil hypotézy o převládajících typech zneužívání.

[79] KHATTAK, Sheharbano, et al. Do You See What I See? Differential Treatment of Anonymous Users [online]. [cit. 2017-07-04]. Dostupné z: http://www.cl.cam.ac.uk/~sk766/publications/ndss16_tor_differential.pdf. 2016.

3. Praktická část

3.1 Metodologie výzkumu

V této části práce je popsána metodologie kvantitativního výzkumu. Jako hlavní výzkumný problém bylo definováno prozkoumání způsobů využívání anonymizační sítě Tor s důrazem na maligní činnost uživatelů a to za účelem zjištění, jakými způsoby je síť nejčastěji zneužívána.

Jedním z mála datových zdrojů, které je možné analyzovat, jsou hlášení o zneužívání, takzvané abuse reports, vztahující se k IP adresám Tor serverů. Tyto reporty jsou ve formě emailových zpráv zasílány operátorům serverů, ze kterých je detekována podezřelá aktivita a mohou být jak strojově generovány, tak ručně psány.

Vzhledem k celkovému množství dat bylo nutné přistoupit ke strojovému zpracování a kvantitativní analýze pro zjištění počtu zneužívání v čase a kategorizaci stížností.

Před započetím výzkumu byly nadefinovány dvě výzkumné otázky, které jsou dále zpřesněny v jednotlivých hypotézách. Jedná se tedy o tyto výzkumné otázky a příslušné hypotézy:

1. Jaká je míra zneužívání Tor sítě?
    a. Zneužívání sítě v čase narůstá.
    b. Existuje vztah mezi velikostí datového toku a počtem stížností.
2. Jakými způsoby je síť Tor zneužívána?
    a. Síť je nejčastěji zneužívána za účelem počítačových útoků.

3.2 Analýza reportů o zneužívání

Jak je z fungování anonymizační sítě Tor patrné, největší potenciální riziko právního postihu hrozí provozovatelům takzvaných exit nodů, tedy serverům, které komunikují s vnějším internetem. Administrátoři těchto uzlů jsou pak nuceni reagovat na veškeré stížnosti a oznámení o zneužívání ze strany poškozených, pro které se koncový uzel sítě jeví jako původce maligního datového toku.

Tyto stížnosti bývají v naprosté většině generovány automaticky ve formě emailové zprávy adresované provozovateli serveru. Na tyto stížnosti je možné reagovat předpřipravenou odpovědí,[80] která se snaží osvětlit princip fungování Tor sítě, nebo v akutních případech může administrátor zamezit svému serveru v komunikaci s konkrétními IP adresami, nebo i celými porty pro zamezení specifických aplikací jakou může být například BitTorrent.

Za účelem získání dat pro praktickou část své diplomové práce jsem oslovil německou neziskovou organízací Zwiebelfreunde e.V., známou také pod názvem Torservers.net, která v současné době provozuje více než desítku velkých koncových uzlů v rámci Tor sítě. Tyto servery mají v současné době celkový potenciál datového toku až 478,37 MiB/s. Při celkovém potenciálu Tor sítě, který se pohybuje mírně nad 50 GiB/s, tak servery Torservers.net reprezentují téměř jedno procento celkové kapacity Exit uzlů celé Tor sítě.[81] [82]

Tabulka 6: Seznam serverů provozovaných Torservers.net

Jméno	Datový potenciál	Země
hessel0	50 MiB/s	Rumunsko
hessel1	25.42 MiB/s	Rumunsko
hessel2	50 MiB/s	Rumunsko
andregorz0	32.91 MiB/s	Rumunsko
edwardsnowden1	50 MiB/s	Rumunsko
edwardsnowden2	50 MiB/s	Rumunsko
criticalmass	22.87 MiB/s	Švédsko
RSF12thMarch	20.59 MiB/s	Švédsko
iVPN	19.35 MiB/s	Švédsko
RSFPressFreedom	19.55 MiB/s	Švédsko
HSLtor	20.84 MiB/s	Švédsko
dorrisdeebrown	16.84 MiB/s	Švédsko
amazonas	40 MiB/s	Holandsko

[80] Abuse Templates. Torservers.net [online]. [cit. 2017-02-05]. Dostupné z: https://www.torservers.net/wiki/abuse/templates
[81] Tor Network Status. Tor Status [online]. [cit. 2017-02-05]. Dostupné z: https://torstatus.blutmagie.de/index.php
[82] Tor Project: Metrics [online]. [cit. 2017-07-04]. Dostupné z: https://metrics.torproject.org/bandwidth-flags.html

politkovskaja	30 MiB/s	Holandsko
freeBogatov	30 MiB/s	Holandsko

Jak uvádí organizace na svém webu,[83] Zwiebelfreunde e.V. funguje v souladu se zákonem Telemediengesetz §15[84] (německý zákon o telemédiích), který zakazuje shromažďovat jakákoliv data umožňující osobní identifikaci, nebo data o používání, mimo případy, kdy jsou tato data nutná k vyúčtování služeb. A vzhledem k tomu, že veškeré své služby poskytuje zdarma, nejsou takováto data shromažďována.

I z toho důvodu je velmi složité vhlédnout do fungování sítě a přinést podložené statistiky o využívání respektive zneužívání sítě. Z povahy fungování za sebou nechávají datovou stopu pouze zneužívající uživatelé a to právě ve formě reportů o jejich maligních aktivitách.

Je nutné rovněž předeslat, že uzly provozované organizací Torservers.net mají aplikována pravidla pro omezení míry zneužívání. To v praxi znamená, že některé porty protokolu IP jsou kompletně blokovány a analyzovaná data jsou těmito nastavení postižena.

3.2.1 Přijetí a zpracování dat

1. května 2016 jsem po předchozí komunikaci s Moritzem Bartlem, administrátorem torservers.net, obdržel kompletní kopii emailové schránky **abuse@torservers.net** v zabaleném *.tar.xz*[85] archivu. Tato emailová adresa je primárním příjemce naprosté většiny všech stížností, jakožto oficiální adresa, kam mají být stížnosti směřovány a jsou do ní rovněž přeposílány stížnosti, které byly zaslány přímo některým z administrátorů.

Kopie schránky pochází z emailového klienta Icedove,[86] svobodné varianty programu Thunderbird od společnosti Mozilla, díky čemuž byla zachována struktura dat ve formě adresářů obsahujících přijaté, odeslané, archivované a filtrované zprávy - tento souborový systém je

[83] Abuse. Torservers.net [online]. [cit. 2017-02-05]. Dostupné z: https://www.torservers.net/abuse.html
[84] Telemediengesetz (TMG). Bundesministerium der Justiz und für Verbraucherschutz [online]. [cit. 2017-02-05]. Dostupné z: http://www.gesetze-im-internet.de/tmg/__15.html
[85] Komprimovaný dokument ve formátu *.xz* využívající algoritmus LZMA2 a zabalený pomocí svobodného formátu *.tar*.
[86] Icedove. Debian.org [online]. [cit. 2017-02-05]. Dostupné z: https://wiki.debian.org/Icedove

rovněž znám jako Maildir. Na české Wikipedii pak blíže specifikují: "*Maildir je široce používaný formát pro uložení e-mailových zpráv. K zachování integrity zpráv při manipulaci s nimi (vkládání, přesun, mazání) nevyžaduje zamykání souborů. Každá zpráva je uložena jako samostatný soubor s unikátním názvem. Všechny změny se provádějí pomocí atomických souborových operací, proto se o záležitosti související se souběhem (současným vícenásobným přístupem) stará souborový systém. Maildir je adresář (často nazvaný Maildir) se třemi podadresáři nazvanými tmp, new a cur.*"[87]

Mimoto byly zachovány rovněž konfigurační soubory a logy, které nebyly použity a pro následnou analýzu jsem se zaměřil výhradně na soubory s koncovkou .mail. Každý jednotlivý soubor s touto koncovkou reprezentuje jednu elektronickou zprávu a ve své textové podobě obsahuje jednak plné znění zprávy a rovněž i hlavičku emailové zprávy respektive metadata[88] o tom, kdy, odkud, komu byla zpráva zaslána.

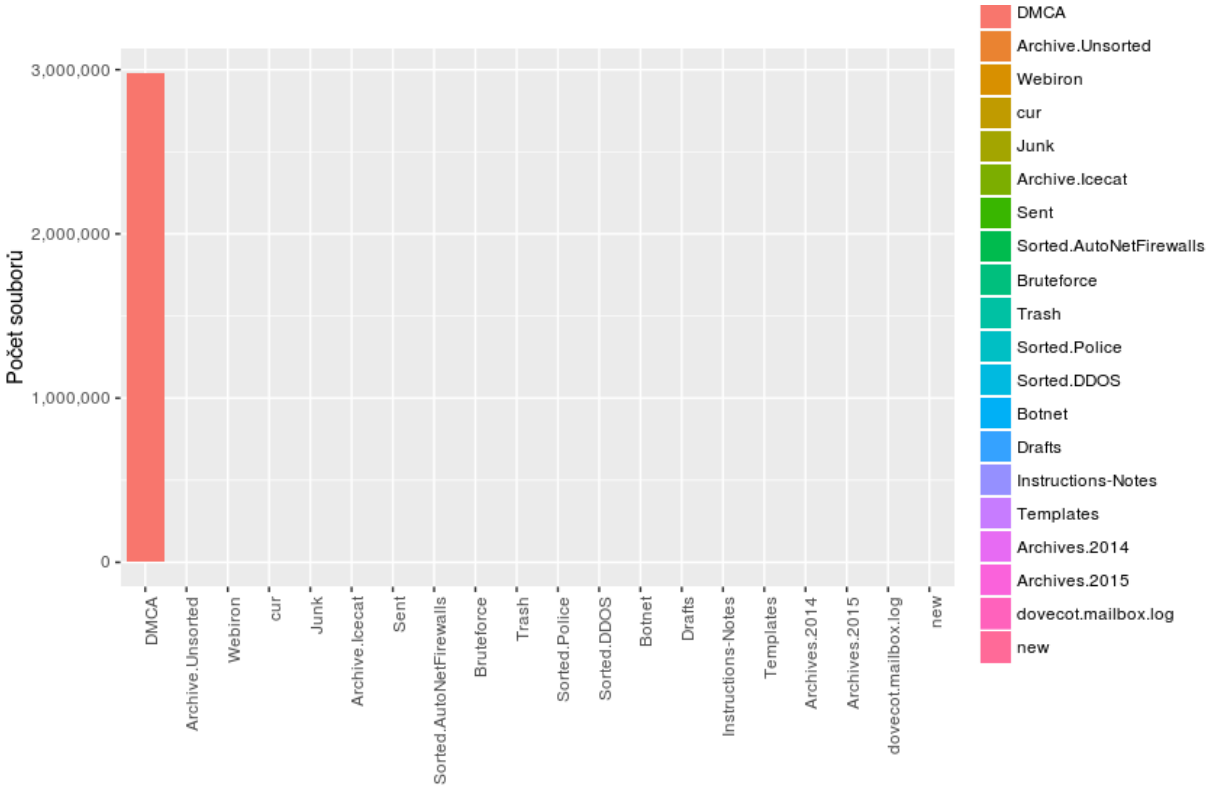
Velikost původního archivu činila 1 018 megabytů, která po rozbalení narostla na konečnou velikost 24 671 megabytů. Vzhledem k velkému množství dat bylo nutné přistoupit ke strojovému zpracování a data miningovým metodám hned z počátku analýzy. Prvním krokem bylo logicky prozkoumání obdržených dat a jejich následné zpracování do strukturované podoby.

V úvodním kroku jsem se zaměřil na vyčíslení celkového počtu zpráv a jejich rozložení v jednotlivých složkách.

[87] Maildir. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-02-05]. Dostupné z: https://cs.wikipedia.org/wiki/Maildir

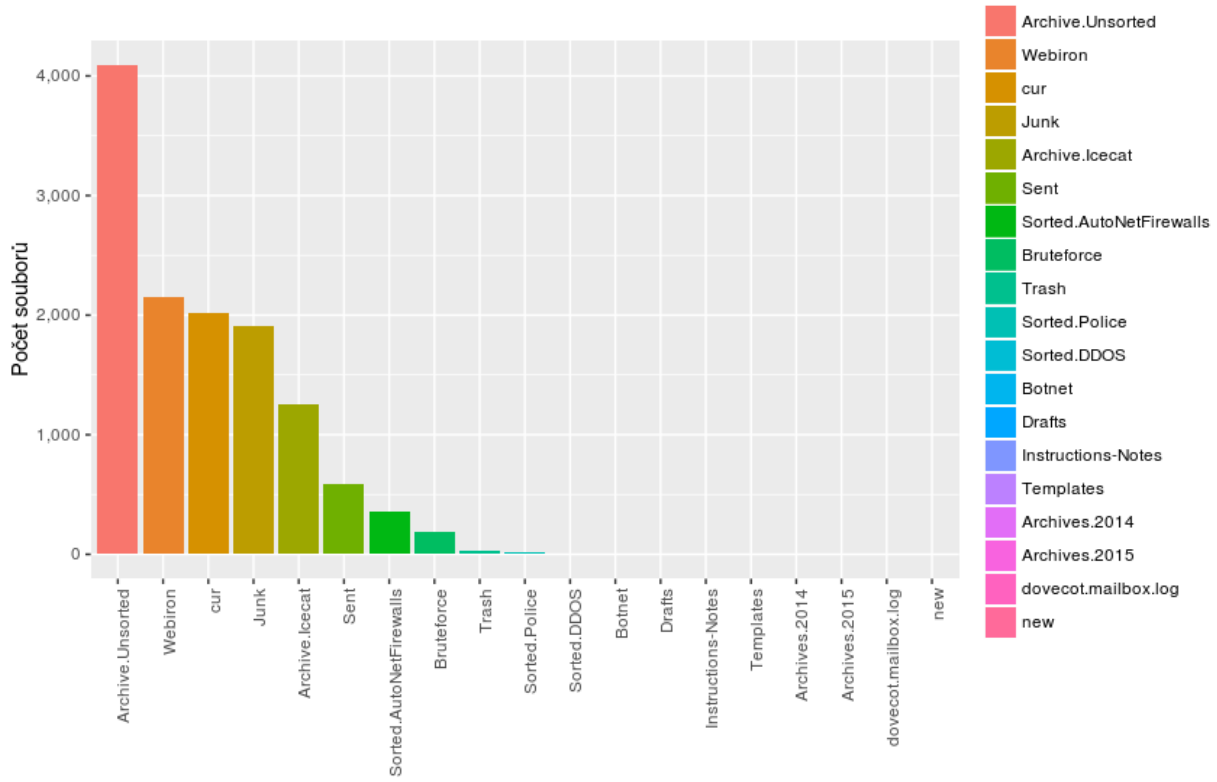
[88] Metada jsou takzvaná data o datech a často obsahují strojově vytvořené informace o časových údajích, kdy byla informace vytvořena či změněna.

Obrázek 5: Počet .mail souborů v jednotlivých složkách



Z přiloženého grafu je patrné, že drtivá většina, více než 99% souborů se nachází ve složce DMCA, do které bylo na základě předem nadefinového filtru přesunuto 2 979 381 z celkového počtu 2 992 013 analyzovaných souborů. Rozložení souborů ve zbylých složkách je následující.

Obrázek 6: Počet .mail souborů v jednotlivých složkách (bez DMCA)



Vzhledem k takto nerovnoměrnému rozložení počtu souborů jsem se rozhodl nadále analyzovat zprávy ze složky DMCA odděleně a to rovněž z důvodu, že se velmi pravděpodobně jedná o hlášení zneužívání na základě porušení autorských práv.

Více informací již ze surových dat nebylo možné získat a proto jsem přistoupil ke strojovému zpracování všech souborů s koncovkou .mail za účelem extrakce klíčových informací do strukturované a strojově čitelné podoby jako základ pro další analýzy. Pro tento úkol jsem si zvolil programovací jazyk R[89], za pomoci kterého jsem ve for cyklu[90] prošel každý jednotlivý soubor a s použitím regulárních výrazů extrahoval následující údaje:

[89] R. The R Project for Statistical Computing [online]. [cit. 2017-02-05]. Dostupné z: https://www.r-project.org/

[90] Cyklus for je řídící struktura, sloužící pro iteraci přes všechny prvky v kolekci - v tomto případě přes všechny soubory s koncovkou .mail.

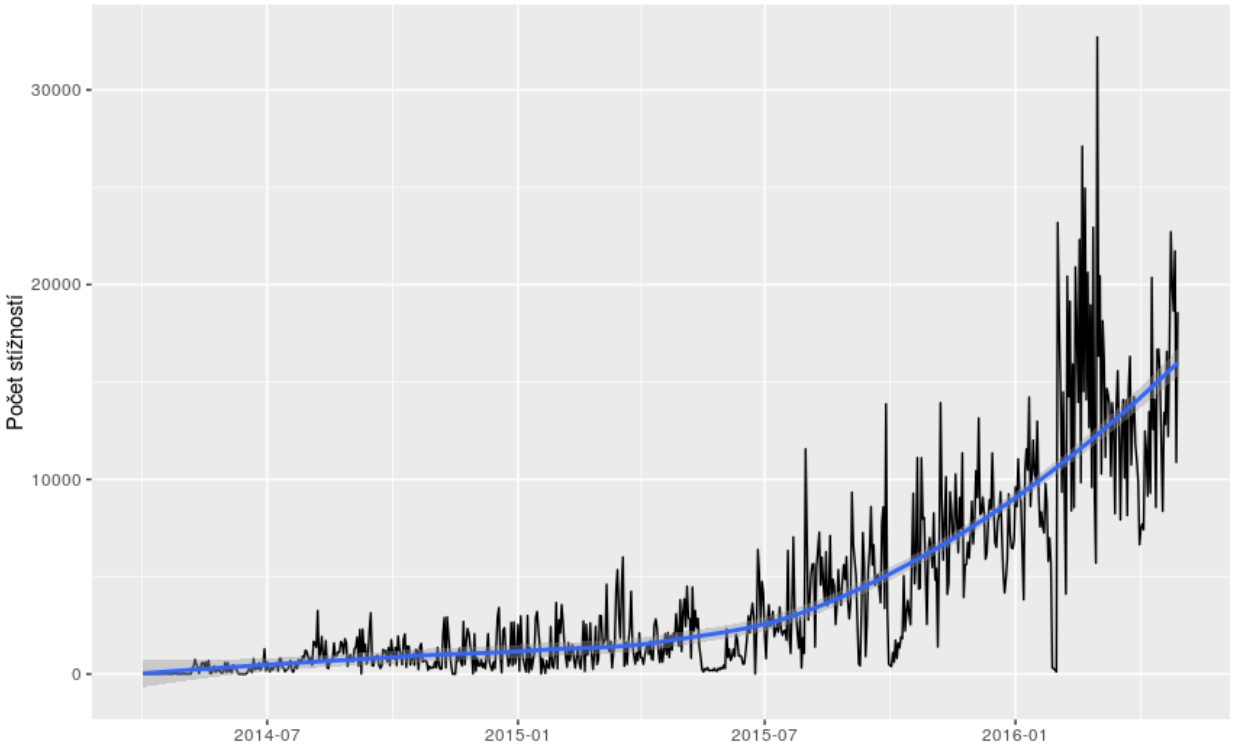
- **Název souboru** - původní název souboru pro zpětnou identifikaci.
- **Datum** - časový záznam z hlavičky emailové zprávy.
- **Odesílatel** - emailová adresa odesílatele stížnosti.
- **Příjemce** - emailová adresa příjemce stížnosti.
- **Předmět** - text uvedený v předmětu emailové zprávy.
- **Protokol** - informace o zneužívaném protokolu (primárně pro DMCA emailové zprávy).
- **Filename** - informace o zneužívaném souboru (primárně pro DMCA emailové zprávy).
- **Emailové adresy** - seznam všech emailových adres zmíněných v emailové zprávě.
- **IP adresy** - seznam všech IP adres zmíněných v emailové zprávě.

Takto vzniklý datový soubor sloužil jako primární zdroj informací pro všechny následné analýzy.

3.2.2 Analýza dat

Po následném vyčištění prvotního datového souboru a převedení časových záznamů na strojově čitelný formát bylo možné přistoupit k prvotní respektive explorativní analýze dat. Explorativní analýza dat je doporučeným prvním krokem při práci s daty a její součástí je prozkoumání dat bez jejich předchozí znalosti, odhalení jejich kvality a chybovosti za účelem zvolení dalšího analytického postupu.

Obrázek 7: Počet stížností v čase



Z časového pohledu je rozložení stížností na první pohled narůstající, je zde ale nutné brát v potaz proměnlivé vytížení serverů a i jejich narůstající celková kapacita. Touto hlubší analýzou se zabývám v pozdější části práce. Rekordní počet stížností je patrný ve dnech okolo 1. března 2016, který je co do počtu obdržených stížností absolutně nejvyšší s 32 708 stížnostmi obdržených tento den. Při průměrné hodnotě 3 908 stížností denně se pak blížíme téměř desetinásobku tohoto průměru.

Modrá linie protínající graf je generována vyhlazovací metodou LOESS a zobrazuje trend narůstání počtu denních stížností v čase. Metoda LOESS, původně navržena William S.

Clevelandem[91], patří mezi metody, které aplikují takzvanou váženou lokální regresi k získání neparametrického modelu, jak tuto metodu popisuje ve své bakalářské práci Anita Koncziová.[92]

3.2.2.1 DMCA stížnosti

Jak uvádím v úvodů této kapitoly, DMCA stížnosti tvoří majoritu všech analyzovaných zpráv a z toho důvodu jsem se rozhodl je analyzovat odděleně. Tyto stížnosti se odvolávají na Digital Millennium Copyright Act, americký zákon z roku 1998, který doplnil původní zákon o autorském právu s ohledem na nové technologie.[93] Tento zákon zavádí koncept takzvaného bezpečného přístavu, v originále "*safe harbour*", kdy poskytovatelé služeb při splnění určitých podmínek omezeně odpovídají za obsah dalších stran. Provozovatelé serverů tak nemusí aktivně monitorovat a vyhledávat obsah, který by autorská práva porušoval, avšak v případě nahlášení jsou nuceni konkrétní případy aktivně řešit.

3.2.2.1.1 Původci stížností

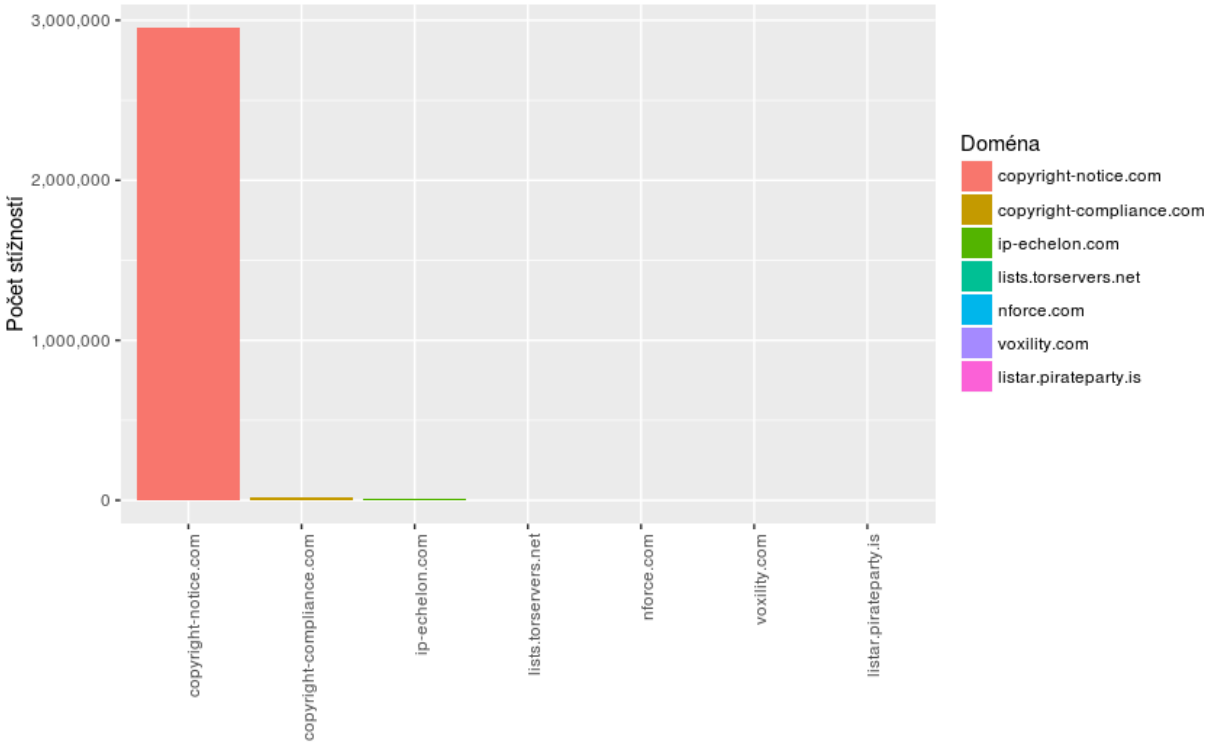
Veškeré stížnosti odvolávající se na Digital Millennium Copyright Act, tedy 2 979 381 emailových zpráv pochází celkově pouze ze 7 domén a v majoritě případů se jedná o specializované společnosti, které kontinuálně monitorují zneužívaný obsah a aktivně jej hlásí příslušným serverům, aby jej odstranily v souladu s legislativní úpravou.

[91] CLEVELAND, William S. Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association [online]. 1979, 74(368), 829-836 [cit. 2017-07-04]. DOI: 10.1080/01621459.1979.10481038. ISSN 0162-1459. Dostupné z: http://www.tandfonline.com/doi/abs/10.1080/01621459.1979.10481038

[92] KONCZIOVÁ, Anita. Neparametrické odhady regresní funkce. Brno, 2013. Dostupné také z: http://is.muni.cz/th/380229/prif_b/Bakalarska_praca.pdf

[93] U.S. COPYRIGHT OFFICE SUMMARY. THE DIGITAL MILLENNIUM COPYRIGHT ACT OF 1998 [online]. 1998 [cit. 2017-02-05]. Dostupné z: https://www.copyright.gov/legislation/dmca.pdf

Obrázek 8: Počet DMCA stížností z jednotlivých domén



1. p2p.copyright-notice.com

Doména copyright-notice.com suverénně vede v počtu zaslaných stížností a emailové zprávy zaslané z této domény tvoří 99% všech obdržených zpráv celkově. Tato doména je registrována na vcelku neznámou společnost Domains By Proxy, LLC.[94] Samotná doména společnosti Domains By Proxy, LLC je pak registrována na již výrazně známější Go Daddy Operating Company, LLC. Společnost GoDaddy Inc.[95] je největším registrátorem domén a poskytovatelem webhostingu a je v současné době obchodována i na New Yorské burze. Tato společnost evidentně využívá svých technických zdrojů k monitoringu porušování copyrightu primárně při sdílení souborů skrze P2P kanály.

2. copyright-compliance.com

[94] Whois. Domain Tools [online]. [cit. 2017-02-05]. Dostupné z: http://whois.domaintools.com/copyright-notice.com
[95] GoDaddy Inc. [online]. [cit. 2017-02-05]. Dostupné z: https://godaddy.com/

Druhým nejpočetnějším zasilatelem stížností v souladu s DMCA je s počtem 20 236 emailových zpráv doména copyright-compliance.com. Tato doména dle registračního záznamu[96] rovněž patří velkému poskytovateli webhostingu a registrátoru domén - Domain.com, LLC.[97]

3. ip-echelon.com

Třetím nejpočetnějším zdrojem je doména ip-echelon.com, kterou vlastní stejnojmenná společnost IP-Echelon. Dle jejich webové stránky se jedná o společnost specializující se na tuto problematiku: "*IP-Echelon je prémiový poskytovatel datové analytiky a měření sledovanosti na neautorizovaných distribučních kanálech. Spolupracujeme s předními tvůrci obsahu a sledujeme neoprávněné šíření ovlivňující jejich podnikání.*"[98]

4. torservers.net

Dalším ze zdrojů je na první pohled samotná doména torservers.net, respektive subdoména lists.torservers.net. Na této adrese je provozováno několik mailing listů a stížnosti byly nesprávně směřovány sem namísto analyzované abuse@torservers.net a následně ručně přeposlány na správnou destinaci. Vzhledem k nesignifikantnímu počtu 248 zpráv jsem dále neanalyzoval původního odesílatele.

5. nforce.com

Pátým zdrojem DMCA stížností je doména nforce.com, patřící stejnojmenné holandské společnosti NFOrce Entertainment B.V.. Zmíněná společnost je poskytovatelem dedikovaných a virtuálních serverů.[99]

6. voxility.com

[96] Whois. Domain Tools [online]. [cit. 2017-02-05]. Dostupné z: http://whois.domaintools.com/copyright-compliance.com
[97] Domain.com [online]. [cit. 2017-02-05]. Dostupné z: http://www.domain.com/
[98] IP-Echelon [online]. [cit. 2017-02-05]. Dostupné z: https://www.ip-echelon.com/
[99] NFOrce Entertainment B.V. [online]. [cit. 2017-02-05]. Dostupné z: https://www.nforce.com/

Předposledním zdrojem je doména voxility.com společnost Voxility, LLC. Shodně s předchozím případem se jedná o poskytovatele cloudové a hostingové infrastruktury.[100]

7. pirateparty.is

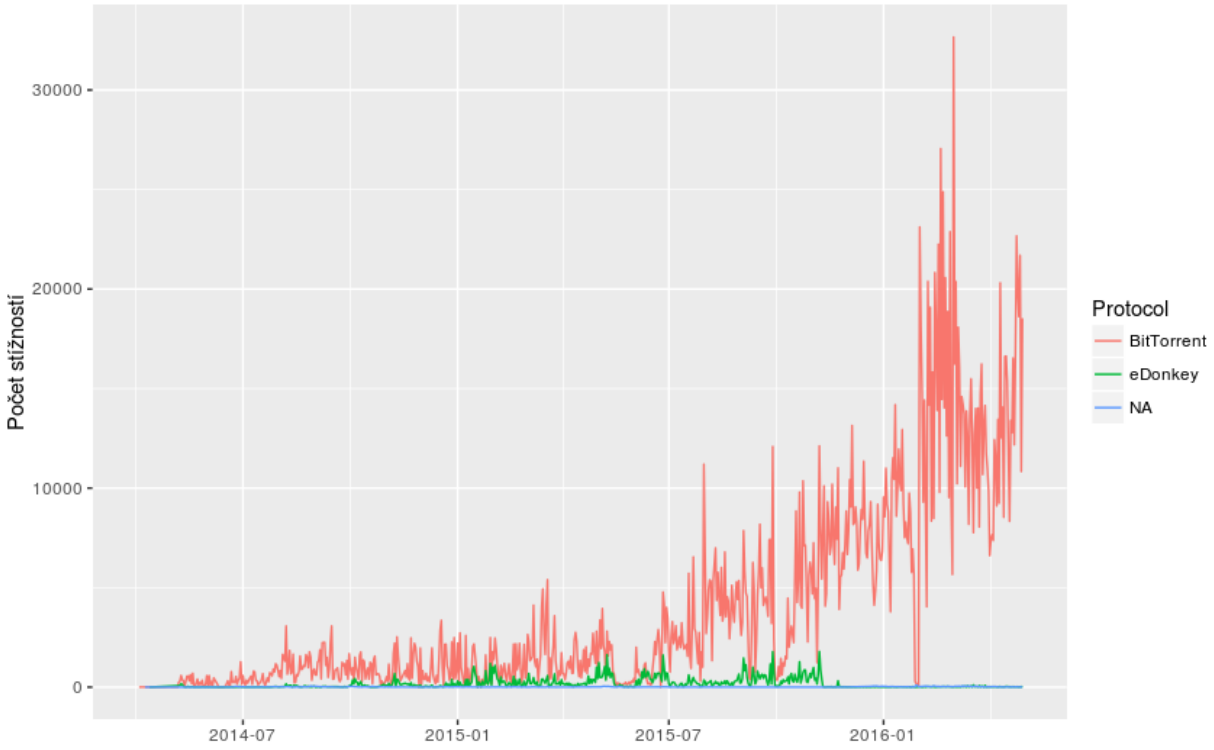
Zcela poslední doménou, odkud byla obdržena stížnost na porušování DMCA, je pirateparty.is, doména patřící islandské Pirátské straně.[101] Jedná se o jednu jedinou zprávu původně adresovanou veřejnému, v dnešní době již nefunkčnímu mailing listu istar.pirateparty.is do sekce Almennur - obecné. Zpráva je z důvodu islandského kódování znaků částečně nečitelná a dotazuje se ohledně kultury, načež dostává odpověď, že to není funkcí mailing listu a zpráva je přeposílána na abuse@torservers.net.

3.1.2.1.2 Protokol

Drtivá většina DMCA emailů rovněž obsahuje informaci, v případě, že se jedná o stížnost na neoprávněné šíření pomocí P2P sítě, jaký P2P protokol byl k těmto účelům využit. Celkově se jedná pouze o dva protokoly v čele s populárním protokolem BitTorrent, jehož využívání vytrvale klesá nicméně pořád má na svědomí téměř 4% celkového datové toku v Evropě, nebo 24%, tedy téměř čtvrtinu veškerého datového toku Asie, jak uvádí ve své výzkumné zprávě společnost Sandvine.[102]

[100] Voxility, LLC [online]. [cit. 2017-02-05]. Dostupné z: https://www.voxility.com/
[101] The Icelandic Pirate Party [online]. [cit. 2017-02-05]. Dostupné z: http://piratar.is/
[102] Global Internet Phenomena 2016 [online]. 2016 [cit. 2017-02-05]. Dostupné z: https://www.sandvine.com/downloads/general/global-internet-phenomena/2016/global-internet-phenomena-apac-mea.pdf

Obrázek 9: Počet DMCA stížností dle protokolu v čase



Opačný trend a nárůst využívání skrze Tor síť je možné přisoudit zvětšující se kapacitě sítě a rovněž její rychlosti, která sdílení dat pomocí P2P sítě umožňuje do větší míry, než tomu bylo dříve. Stále je ovšem využívání BitTorrentu skrze Tor nedoporučováno. Jednak nadbytečně zatěžuje síť a také se jedná o velké bezpečnostní riziko z pohledu anonymity uživatele. Jak shrnuje Roger Dingledine, jeden z vývojářů Toru, ve svém blogpostu *BitTorrent skrze Tor není dobrý nápad*[103] - někteří z populárních BitTorrent klientů prozradí uživatelskou IP adresu i přes směrování skrze Tor a jako příklad uvádí programy uTorrent, BitSpirit, nebo libTorrent.

Z přiloženého grafu lze rovněž zřetelně vyčíst strmý propad ve využívání protokolu eDonkey a nejpopulárnějšího klienta pro jeho použití eMule. Tento trend korespenduje s globálním poklesem ve využívání tohoto softwaru. Zatímco v roce 2015 obstarával 3,7% veškerého

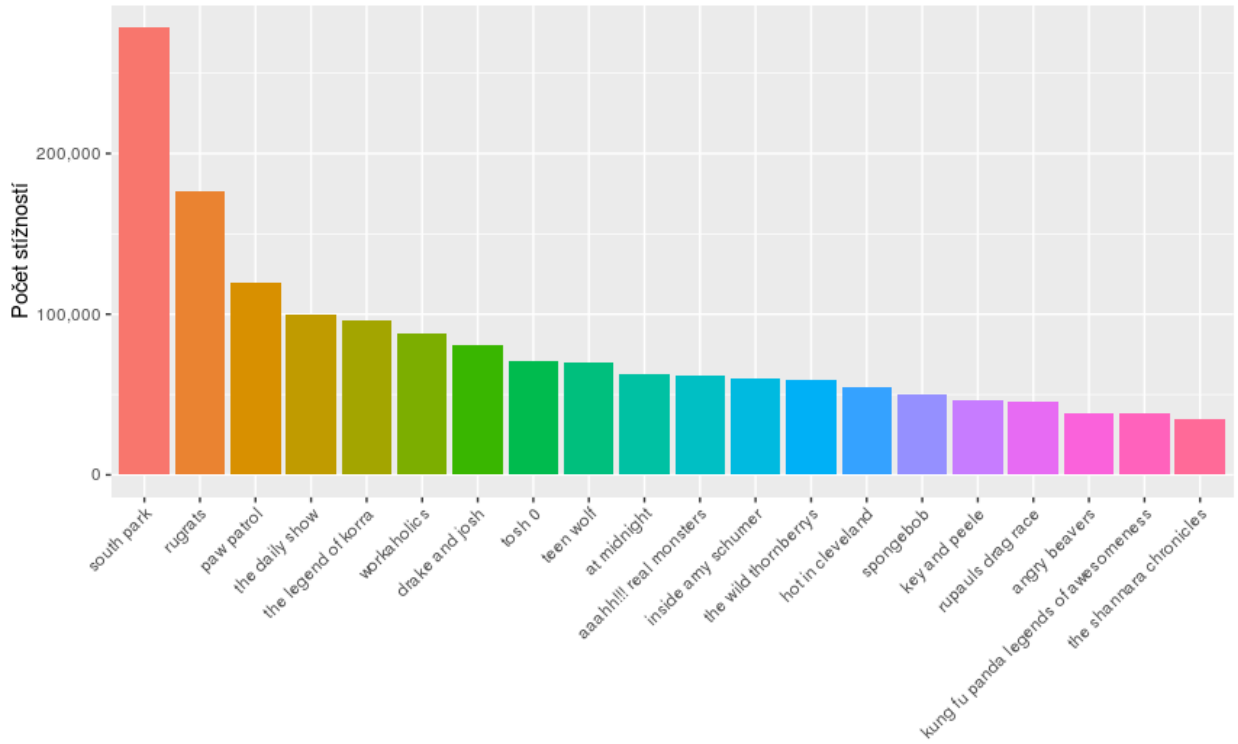
[103] DINGLEDINE, Roger. Bittorrent over Tor isn't a good idea [online]. 2010 [cit. 2017-02-05]. Dostupné z: https://blog.torproject.org/blog/bittorrent-over-tor-isnt-good-idea

evropského uploadu, za rok 2016 se již nevyskytuje v Top 10 přehledu a hovoří se o jeho strmém propadu ve využívání.[104]

3.1.2.1.3 Soubory

Spolu s informací o použitém protokolu obsahují DMCA stížnosti z velké části rovněž i název souboru respektive torrentu, u kterého je hlášeno porušení autorských práv. Po unifikaci nestandardizovaných názvů torrentů, které často obsahují další informace o souborech jako rok vydání, nebo údaje o sezóně v případě seriálu, se objevuje přibližně tři tisíce unikátních titulů.

Obrázek 10: Počet DMCA stížností na jednotlivé tituly

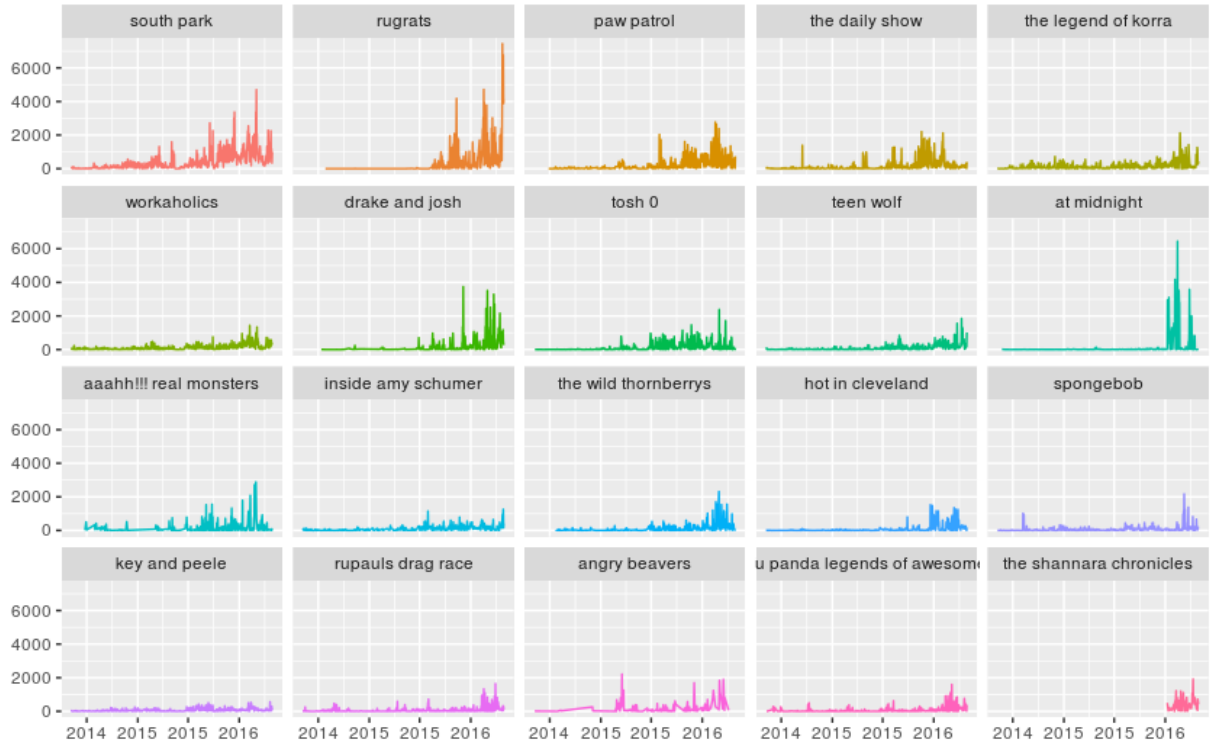


Přiložený graf zobrazuje dvacet nejčastěji zmiňovaných titulů v čele s populárním kresleným seriálem South Park. Těchto dvacet titulů reprezentuje téměř přesně jednu třetinu všech

[104] BitTorrent Traffic Share Drops to New Low [online]. 2015 [cit. 2017-02-05]. Dostupné z: https://torrentfreak.com/bittorrent-traffic-share-drops-to-new-low-150918/

stížností. Celkově mezi prvními sto tituly panuje převaha kreslených seriálů doprovázenými sitcomy, několika talk show a erotickými filmy.

Obrázek 11: Počet DMCA stížností na jednotlivé tituly v čase



Zatímco u některých titulů je zájem v čase rozložen, u dalších strmě vzrostl v posledním roce, kde je tento trend znatelný hlavně u soutěžního pořadu @midnight.[105] Jak název toho pořadu napovídá, je vysílán v pozdních nočních hodinách a je silně orientován na prostředí internetu a sociálních sítí.

[105] @midnight [online]. [cit. 2017-02-12]. Dostupné z: http://www.cc.com/shows/-midnight

Obrázek 12: Počet DMCA stížností dle formátu souboru v čase

Převahu seriálů je možné vysvětlit větším počtem souborů, epizod, které uživatele opakovaně stahují, zatímco u filmů se jedná o jednorázovou akvizici multimediálního souboru. Přiložený graf ilustruje zájem jak o audio (.aac), tak video soubory (.avi, .mkv, .mp4) v čase a převažující zájem o moderní formáty na úkor zastaralého formátu .avi.

3.1.2.1.4 Shrnutí

V rámci stížností týkajících se autorských práv můžeme pozorovat téměř exponenciální růst. Tento trend může být částečně vysvětlen tím, že celková propustnost Tor sítě neustále roste a i díky tomu je možné skrze ni streamovat video, nebo stahovat z P2P sítí, což ještě před několika roky možné nebylo, nebo jen s velikou dávkou trpělivosti.

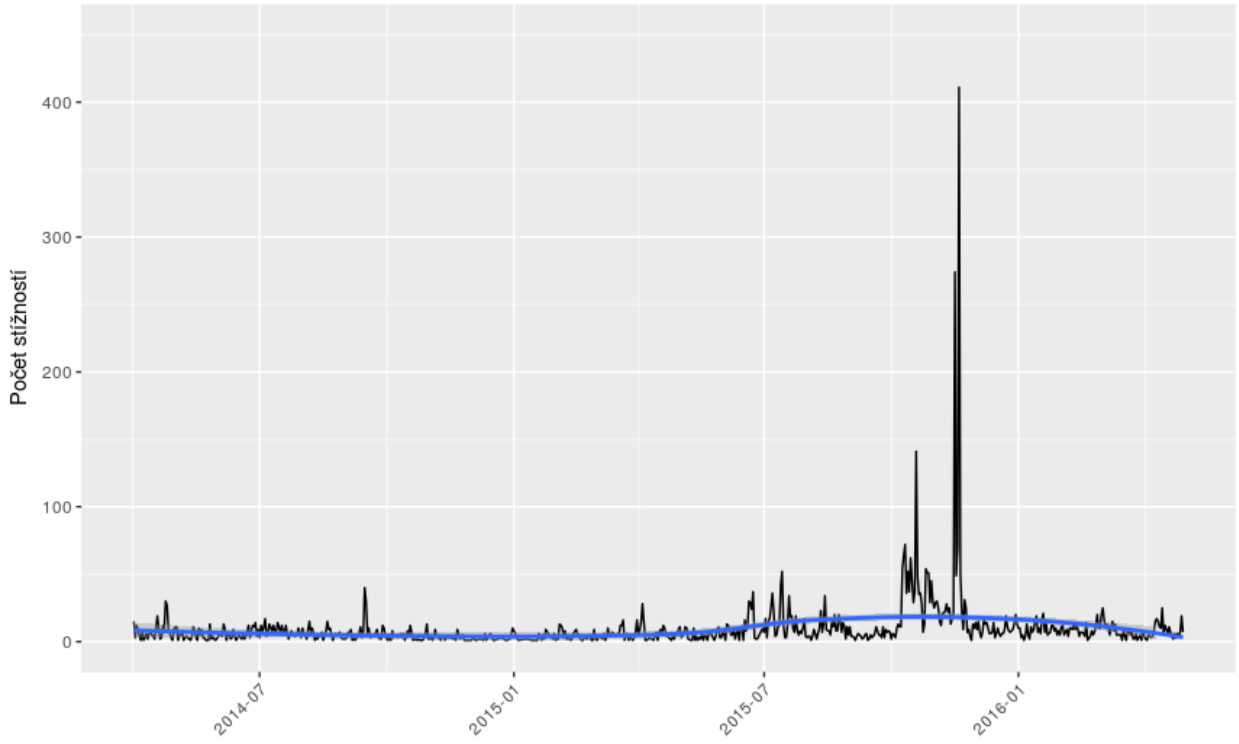
Analyzovaná data rovněž dokládají trendy v oblíbenosti P2P protokolů, kde i mimo Tor síť vítězí BitTorrent nad eDonkey, nebo souborových formátů, kde je viditelná vzrůstající oblíbenost formátu MP4 a MKV ve srovnání se stagnujícím AVI.

Téměr 3 milióny stížnosti bylo odesláno pouze ze 7 domén v drtivé převaze od společností, které evidentně provádí monitoring P2P sítí jako službu pro hollywoodské filmové společnosti včetně automatizovaného rozesílání stížností v souladu s Digital Millennium Copyright Act.

3.2.2.2 Ostatní stížnosti

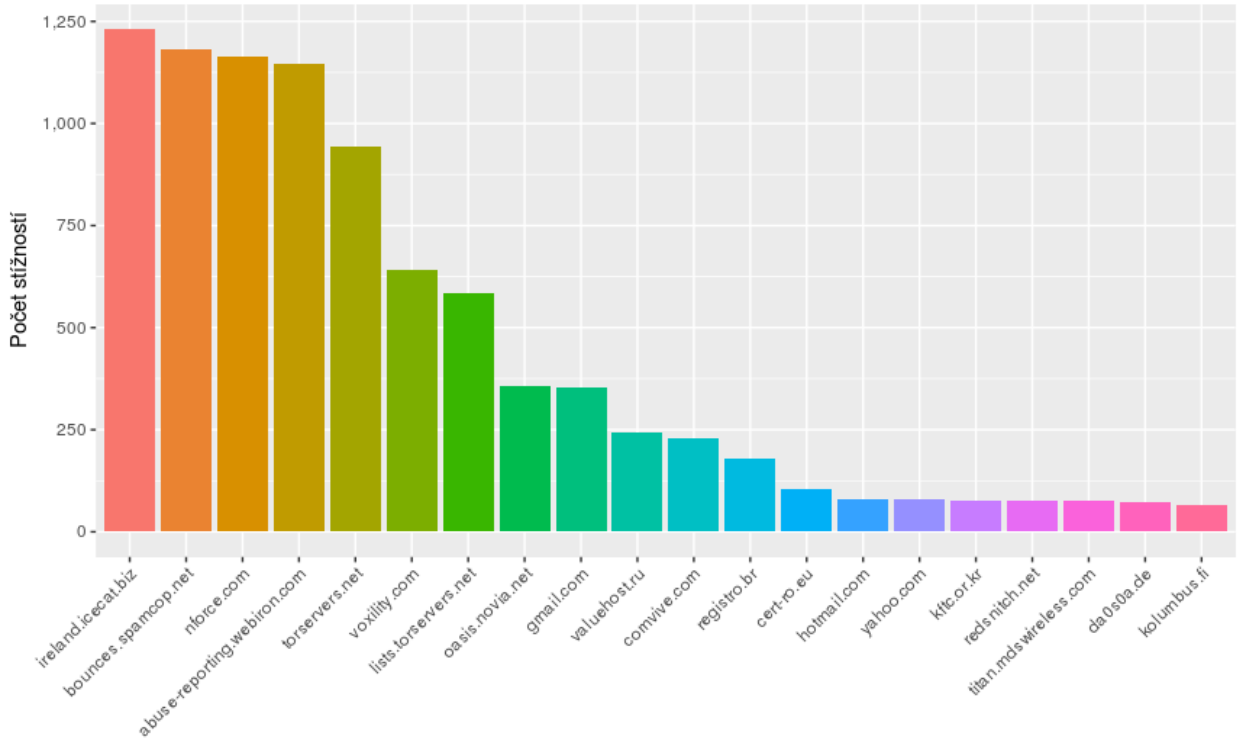
Po oddělení DMCA stížností stále zbývá 12 632 nerozřazených emailových zpráv. Jak jsem byl původně informován, mezi zprávami se pravděpodobně nacházejí stížnosti na počítačové útoky (včetně rozesílání spamu), na nenávistné písemné útoky (tzv. hate speech), několik žádostí od státních úřadů (policie, soud), ale i spam a šifrované, nebo nečitelné zprávy.

Obrázek 13: Počet ostatních stížností (bez DMCA)



Z časového hlediska jsou stížnosti rozloženy v čase rovnoměrněji než tomu bylo u DMCA stížností, které téměř exponenciálně narůstají, zde si naopak drží více či méně konstantní hodnotu okolo několika málo desítek stížností denně. 19. listopadu 2015 denní počet stížností dosahuje svého vrcholu s 411 stížnostmi obdrženými tento den. Jedná se o stížnosti od společnosti SpamCop, která informuje o ruském spambotovi šířícím spam skrze Tor servery. Definice spambota na české Wikipedii je následující: "Spambot je v informatice název počítačového programu, který slouží pro rozesílání spamu. *E-mailoví spamboti shromažďují emailové adresy z materiálů nalezených na Internetu za účelem vytvořit seznam emailových adres pro rozesílání spamu. Tito spamboti (web crawlers) vyhledávají emailové adresy na webových stránkách, v chatovacích místnostech, diskuzních serverech a dalších webových stránkách."*[106]

Obrázek 14: Počet ostatních stížností z jednotlivých domén



[106] Spambot. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-02-19]. Dostupné z: https://cs.wikipedia.org/wiki/Spambot

Veškeré stížnosti pocházejí z více jak 1 400 unikátních domén a i vzorek dvaceti nejčastějších ilustruje větší různorodost stížností než jak tomu bylo v předešlé kapitole.

- **ireland.icecat.biz**

Stížnosti z této domény se datují do roku 2011 a 2012, kdy bylo zasláno všech 1 230 zpráv. Všechny zprávy shodně odkazují na možný útok na Routing Information Protocol. "*Routing Information Protocol (RIP) je v informatice směrovací protokol umožňující směrovačům (routerům) komunikovat mezi sebou a reagovat na změny topologie počítačové sítě. Ačkoliv tento protokol patří mezi nejstarší doposud používané směrovací protokoly v sítích IP, má stále své uplatnění v menších sítích a to především pro svoji nenáročnou konfiguraci a jednoduchost. Poprvé byl definován v RFC 1058 (1988).*"[107]

RIP směřovač je možné identifikovat scanováním sítě na portu 520, který v UDP síti používá. Vzhledem k žádné autentizaci ve verzi 1 a clear textové autentizaci ve verzi 2 je snadné provoz odposlouchávat, přesměrovat a zároveň snadno zjistit heslo.

- **bounces.spamcop.net**

Druhou nejpočetnější skupinou, jsou stížnosti zaslané z domény spamcop.net. Společnost SpamCop založená roku 1998 Julianem Haightem se specializuje na reportování spamu a správou SpamCop Blocking Listu, nebo také SpamCop Blacklistu - zkráceně SCBL.[108]

Automatické zprávy zaslané ze subdomény bounces.spamcop.net pak pouze oznamují odesílateli, respektive jsou zasílány na IP adresu jevící se jako odesílatel zprávy (Tor Exit Node), že se IP adresa nachází na seznamu blokovaných stránek a důvody zablokování.[109] V tomto případě se velmi pravděpodobně jedná o reakci na zasílání spamu skrze Tor síť a následné zařazení IP adres na blacklist SpamCopu. Při manuální verifikaci několika zpráv se

[107] Routing Information Protocol. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-02-12]. Dostupné z: https://cs.wikipedia.org/wiki/Routing_Information_Protocol
[108] SpamCop. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-02-13]. Dostupné z: https://en.wikipedia.org/wiki/SpamCop
[109] Dispute Resolution: Bounce message recipients and end users. SpamCop.net [online]. [cit. 2017-02-13]. Dostupné z: https://www.spamcop.net/fom-serve/cache/405.html

opravdu jednalo o klasické nevyžádané zprávy slibující mimo jiné zázračné zbohatnutí, nebo zhubnutí.

- **nforce.com / voxility.com / valuehost.ru**

Nforce.com podobně jako voxility.com či valuehost.ru jsou klasické hostingové společnosti, na jejichž stroje, respektive na jejich klienty, mohou být vedeny počítačové útoky skrze Tor síť. DDoS[110] útoky počínaje a pokusy o zneužití slabin v populárních CMS[111] systémech konče.

- **webiron.com**

Webiron je specializovaná společnost, která poskytuje ochranu před boty a malwarem jako službu. Stížnosti pocházející z této domény tak adresují zmíněnou maligní činnost ve vztahu k jejich klientům. Zatímco na jedné straně tato společnost viní ve svém blogpostu nazvaném *Nechtěné důsledky posvátného proxy serveru* (The unintended consequences of the sacred proxy[112]) Tor síť za 99% veškerého maligního datového toku, na straně druhé si administrátoři exit serverů stěžují na nekorektní praktiky této společnosti a ošemetnou metodologii na jejímž základě opírá svá tvrzení o počtech útoku.[113]

- **torservers.net a lists.torservers.net**

Zprávy uvnitř mailové schránky zaslané z domén torservers.net jsou ve většině případů odpověďmi na doručené stížnosti (torservers.net) a přeposílané stížnosti či interní komunikace administrátorů (lists.torservers.net).

- **oasis.novia.net**

[110] Distributed Denial of Service je druh počítačového útoky, kdy dojde k přetížení serveru velkým množstvím požadavků načež je znepřístupněn ostatním uživatelům.
[111] Content Management System, česky systém pro správu obsahu, je software určený pro spravování zpravidla webového obsahu bez nutnosti úpravy kódu - například Wordpress.
[112] The unintended consequences of the sacred proxy. Webiron.com [online]. [cit. 2017-02-15]. Dostupné z: https://blog.webiron.com/index.php/2015/11/23/the-unintended-consequences-of-the-sacred-proxy/
[113] WUBTHECAPTAIN. Webiron requesting to block several /24 subnet [online]. 2015 [cit. 2017-02-15]. Dostupné z: https://lists.torproject.org/pipermail/tor-relays/2015-December/008259.html

Všech 357 stížností pocházejících z této domény se datují v rozmezí 2. září 2012 až 14. prosince téhož roku s téměř shodným obsahem a předmětem *Ongoing, repetitive, flooding and abuse on ba.broadcast*, který se postupem času mění na *Ongoing, repetitive, flooding, forgery, and abuse on Usenet newsgroup ba.broadcast from your site*. Jak je již z předmětů těchto zpráv patrné, administrátor usenetové skupiny ba.broadcast si stěžuje na zahlcování diskusí nerelevantním obsahem, který se nevztahuje k televiznímu a radiovému vysílání v Bay Area v okolí San Francisca. Dále uvádí, že jim toto konání rozvrací komunitu a podněcuje hádky, proto hodlají přejít na režim moderování příspěvků v této skupině.[114]

- **gmail.com / hotmail.com / yahoo.com**

Další skupinou domén, ze kterých přichází poměrně značná část stížností, jsou veřejné bezplatné emailové služby, které v majoritě případů používají jednotlivci pro osobní potřeby. Tuto hypotézu potvrzuje několik prozkoumaných zpráv, které došly z těchto domén. Zajímavým příkladem je jedna z posledních obdržených zpráv s předmětem Harrassement (Obtěžování), kde si student stěžuje, že byl udán jinými studenty, že při hodině španělštiny podváděl a prosí administrátory, zda nelze vypátrat původní oznamovatel.

Více časté jsou ale stížnosti na pokusy o proniknutí do osobní mailové schránky uživatelů, útoky na jejich webové stránky, spam a také nenávistné či výhružné zprávy.

3.1.2.2.1 Kategorizace stížností

Pro větší porozumění obsahu jednotlivých stížností jsem zvolil přístup strojové kategorizace všech stížností s následnou ruční validací na náhodném vzorku. Z plných textů 12 632 mailů jsem vytvořil korpus, na který jsem aplikoval text miningové transformace typu převedení všech charakterů na malá písmena, nahrazení všech mailových adres slovem mail, nebo nahrazení všech IP adres slovem IP.

[114] [PROPOSAL] Create moderated version of ba.broadcast, ba.broadcast.moderated. Google Groups [online]. [cit. 2017-02-15]. Dostupné z: https://groups.google.com/forum/#!topic/ba.broadcast/3fvYF14mKG4

Tento korpus jsem následně převedl na takzvanou *document-term matrix*, tedy velmi řídkou matici, ve které je každý dokument (mail) reprezentován jedním řádkem a každé unikátní slovo jedním sloupcem. Číselné hodnoty pak reprezentují frekvence, kolikrát se dané slovo v daném dokumentu vyskytuje. Matice je řídká z toho důvodu, že převažující hodnota je 0, protože s unikátností slova se snižuje jeho výskyt napříč dokumenty.

Pro zlepšení výsledků jsem celou matici převážil za pomoci metody term frequency–inverse document frequency.[115] Tato metoda využívá kombinace frekvencí slov a invertované frekvence dokumentů. Na jejich základě identifikuje klíčová slova uvnitř dokumentů za pomoci obecného předpokladu, že se klíčová slova v textu objevují v menším počtu než běžnější slova tvořící text. Výsledkem je pozměněná document-term matice, u které mají větší váhu unikátnější slova a naopak menší váhu slova nejfrekventovanější.

Dalším krokem bylo normalizování matice a vypočítání euklidovských vzdáleností jednotlivých dokumentů, respektive vektorů, které každý dokument reprezentují.[116] Vzdálenost v euklidovském prostoru odpovídá podobnosti dokumentů – čím více jsou od sebe dokumenty vzdáleny, tím více jsou odlišné.

Výsledná matice euklidovských vzdáleností slouží jako vstup do shlukovacího algoritmu k-means. Princip fungování algoritmu popisuje Jiří Kučera: "Jedná se nehierarchický algoritmus, který třídí data do k shluků na základě jejich vlastností. Počet shluků k se zadává na začátku algoritmu a je menší než počet objektů. Tento algoritmus pracuje tak, že přiřadí každý bod do shluku, jehož středu je nejblíže. Středy shluků se při každém běhu algoritmu znovu spočítají jako aritmetické průměry všech bodů shluku. Cílem je dosáhnout co nejmenších rozdílů uvnitř shluků."[117]

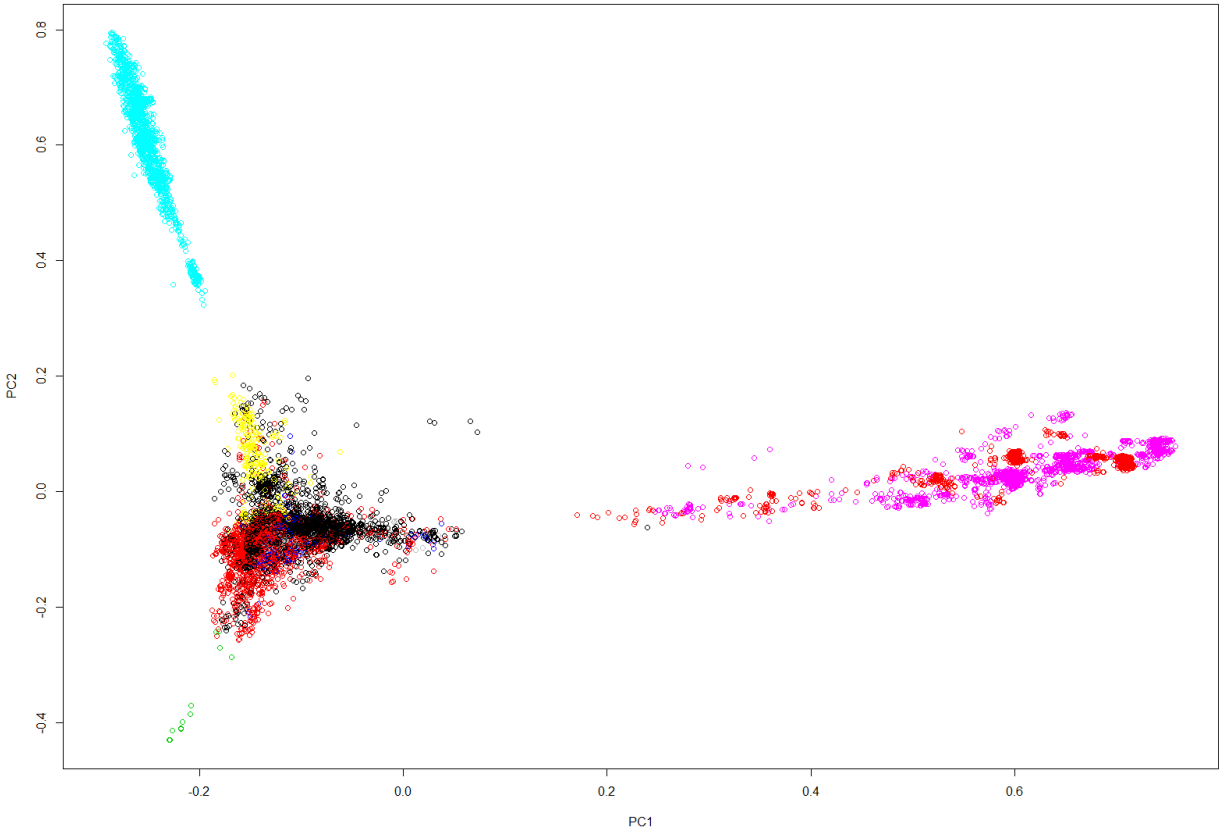
Standardní k-means nebylo možné použít vzhledem k velikosti matice a omezené výpočetní kapacitě. Alternativní metoda sférických k-means s opakovaným půlením adresuje tento

[115] SALTON, Gerard a Christopher BUCKLEY. Term-weighting approaches in automatic text retrieval. Information Processing [online]. 1988, 24(5), 513-523 [cit. 2017-06-17]. DOI: 10.1016/0306-4573(88)90021-0. ISSN 03064573. Dostupné z: http://linkinghub.elsevier.com/retrieve/pii/0306457388900210
[116] JAMES E. GENTLE. Matrix algebra theory, computations, and applications in statistics [online]. [Online-Ausg.]. New York, N.Y: Springer, 2007 [cit. 2017-06-17]. ISBN 978-038-7708-737.
[117] KUČERA, Jiří. Algoritmus k-means [online]. [cit. 2017-05-27]. Dostupné z: https://is.muni.cz/th/172767/fi_b/5739129/web/web/kmeans.html

problém a byla nakonec vybrána jako vhodná metoda. Od klasického k-means se algoritmus liší tím, že všechny body začínají v jednom shluku a za pomoci standardního k-means je rozdělen na dva. Tento krok je opakován dokud nedosáhne maximálního počtu iterací, nebo do doby kdy je dosaženo optimálního rozložení všech bodů do jednotlivých shluků.[118]

Při zvolení 8 shluků jako optimálního počtu došlo k rozřazení emailových zpráv následujícím způsobem:

Obrázek 15: Vizualizace shluků stížností



Velké samostatné shluky (tyrkysová, růžová) byly zpětně identifikovány buď jako spam 7 052 stížností, nebo jako šifrované nečitelné zprávy, které byly vyřazeny z další analýzy. Zbylý korpus 5 580 zpráv byl nadále kategorizován za pomoci extrakce klíčových slov z předmětu i textu zprávy.

[118] KUČERA, Jiří. Nehierarchické metody shlukování [online]. [cit. 2017-05-27]. Dostupné z: https://is.muni.cz/th/172767/fi_b/5739129/web/web/nehiermet.html#mqm

Tabulka 7: Počty stížností dle jednotlivých kategorií

Kategorie	Počet
WordPress útok hrubou silou	1 245
Spam	805
Ostatní	761
Škodlivý přístup k serveru	664
Semalt SEO referral spam	439
Porušení copyrightu	278
Útok hrubou silou POST	224
Google Groups spam	176
Urážlivé příspěvky na Usenetu	144
WordPress skenování zranitelností	104

1. WordPress útok hrubou silou

V největším počtu se ve stížnostech objevuje hacking ve spojitosti s redakčním systémem WodPress a na prvním místě konkrétně útok hrubou silou. Útok hrubou silou využívá výpočetního výkonu k náhodnému zkoušení kombinací jmen a hesel za účelem získání přístupu do administrace systému WordPress. Tento typ útoku může být úspěšný v případě využívání slabých hesel v kombinaci s běžnými uživatelskými jmény jako je *admin*.[119]

2. Spam

Druhou nejčetnější kategorií jsou stížnosti na spam, kde se jako původce těchto zpráv jeví Exit server Tor sítě. Zneužívání Tor sítě k rozesílání spamu je poměrně časté a velmi těžko se eliminuje, neboť blokace portů pro elektronickou poštu by měla dopad i na všechny ostatní uživatele. Právě kvůli rozesílání nevyžádané pošty se dostávají Exit servery na všemožné seznamy rozesílatelů spamu a jsou následně blokovány.

[119] WordPress: Brute Force Attacks [online]. [cit. 2017-05-28]. Dostupné z: https://codex.wordpress.org/Brute_Force_Attacks

3. Ostatní

Třetí nejpočetnější kategorií jsou stížnosti, které se nepodařilo jednoznačně zařadit a často se rovněž jedná o odpovědi a komunikaci nad zaslanými stížnostmi a dotazy.

4. Neoprávněný přístup k serveru

V této kategorii jsou zahrnuty stížnosti na neoprávněný přístup k serveru, zahrnující jak úspěšné, tak neúspěšné zpozorované pokusy. Jedná se jak o přístupy do osobních mailových schránek tak i o pokusy proniknout do firemních serverů.

5. Semalt SEO referral spam

Do této kategorie jsou zařazeny spamové zprávy od společnosti Semalt. Ta se svými falešnými referral požadavky snaží cílovou stránku posunout směrem nahoru v řazení populárních vyhledávačů a jedná se o nekalou optimalizační praktiku pro vyhledávače (SEO).[120]

6. Porušení copyrightu

Podobně jako v předchozí kapitole se zde nacházejí stížnosti v souvislosti s autorskými právy. Kromě nelegálního sdílení souborů přes P2P sítě se zde objevují i stížnosti na další file sharingové platformy a to v převážné většině ve jménu zákona DMCA.

7. Útok hrubou silou POST

Sedmá nejpočetnější kategorie shromažďuje stížnosti na útoky hrubou silou za využití POST metody u REST API. Podobně jako u první skupiny se zde útočník pokouší proniknout do uživatelských účtů za pomocí velkého počtu náhodných kombinací jmen a hesel, pouze namísto systému WordPress jsou zde cílem služby, které umožňují přístup skrze vlastní API.

[120] GAYER, Ofer. Semalt Hijacks Hundreds of Thousands of Computers to Launch a Referrer Spam Campaign [online]. [cit. 2017-05-28]. Dostupné z: https://www.incapsula.com/blog/semalt-botnet-spam.html

8. **Google Groups spam**

V této kategorii se nacházejí stížnosti na spamování komunitních fór Google Groups, případně i urážlivé a útočné příspěvky – takzvaný hate speech.

9. **Urážlivé příspěvky na Usenetu**

V předposlední kategorii jsou obsaženy stížnosti na urážlivé a nenávistné příspěvky, takzvaný hate speech, na diskusním systému Usenet.

10. **WordPress skenování zranitelností**

Desátou nejpočetnější skupinou jsou stížnosti na skenování zranitelností redakčního systému WordPress. Pro skenování zranitelností systému WordPress existují automatizované systémy prověřující známé zranitelnosti konkrétních verzí WordPressu, jeho pluginů ale i web serveru jako takového. V případě detekování některé ze zranitelností jsou pak dveře pro útočníka de facto otevřeny ke vniku do systému.

3.1.2.2.2 Geolokace IP adres

Ve stížnostech doručených na *abuse@torservers.net,* vyjma DMCA stížností, je zmíněno 8416 unikátních IP adres. Po odfiltrování IP adres samotných Tor serverů, které jsou historicky veřejně dostupné[121], zůstává 8170 unikátních IP adres, které jsou jak ve formě adres veřejných, tak i ve formě adres lokálních sítí.

Za účelem zprostředkování geografického pohledu, na jaké země je potenciálně útočeno skrze Tor, bylo nutné propojit IP adresy s jejich lokací. Pro dohledání geolokace k jednotlivým IP adresám jsem se rozhodl využít proprietární databázi MaxMind[122], která je v omezené verzi zdarma dostupná a snadno použitelná. Pro 5367 unikátních IP adres byla úspěšně dohledána geolokace ve formě státu, ve kterém se rozsah těchto IP adres používá, u zbylého počtu 2803

[121] CollecTor: Exit list [online]. [cit. 2017-03-11]. Dostupné z: https://collector.torproject.org/archive/exit-lists/
[122] MaxMind: IP Geolocation and Online Fraud Prevention [online]. [cit. 2017-02-19]. Dostupné z: https://www.maxmind.com/

adres nebyla lokace dohledána ať již z důvodu nekompletnosti databáze, tak i z důvodu častého uvádění neveřejných lokálních IP adres, které nejsou unikátní a proto je nelze lokalizovat.

Obrázek 16: Kartogram stížností

Z přiloženého kartogramu je patrné, že nejčastějším cílem respektive původcem stížností jsou dle IP adres Spojené státy americké s 12 851 zmínkami některých z IP adres, následovány Holandskem a Německem s počtem více než 5 000 zmínek a první pětici uzavírá Rumunsko a Spojené království s počtem přes 2 000 zmínek. Pro zajímavost, Česká republika je s 91 zmínkami na pomyslném 21. místě. Jak dokládají bílá místa na mapě, ve stížnostech se objevují IP adresy z téměř celého světa a z celkového počtu 190[123] suverénních států jsou zmíněny IP adresy 111 z nich.

[123] List of sovereign states. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-03-11]. Dostupné z: https://en.wikipedia.org/wiki/List_of_sovereign_states

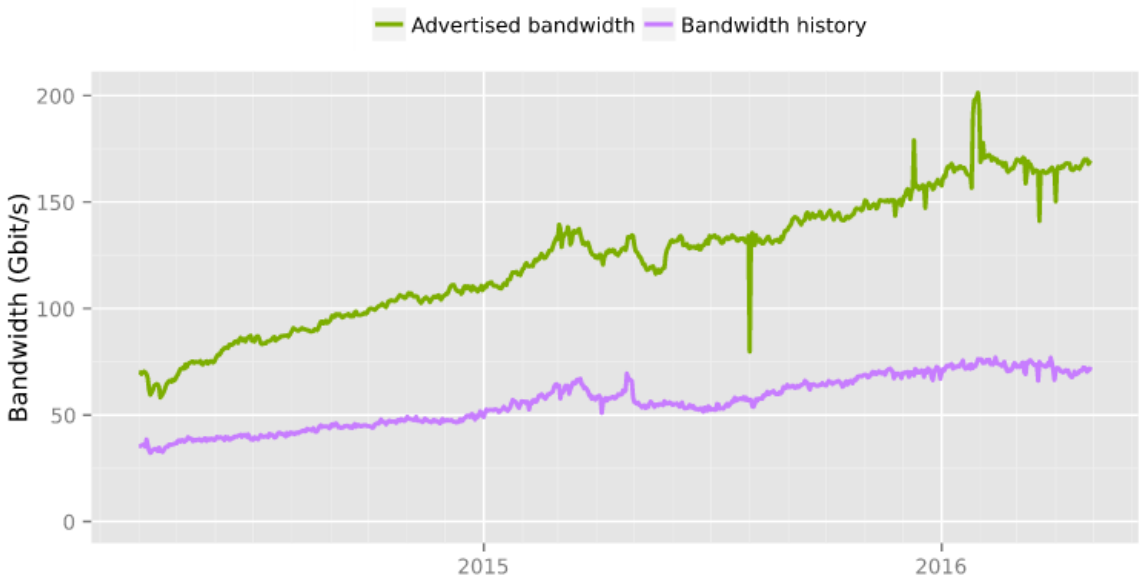
Tabulka 8: Počty výskytů cílů dle jednotlivých států

Stát	Počet výskytů
Spojené státy americké	12 851
Nizozemsko	5 776
Německo	5 189
Rumunsko	2 716
Spojené království	2 308
Ukrajina	680
Turecko	633
Malajsie	595
Španělsko	570
Čína	514

3.2.2.3 Korelační analýza

Jak uvádím v dřívější části práce, na celkové počty stížností se nedá nahlížet pouze jako na číslo absolutní, nýbrž je nutné míti na zřeteli, že počet uživatelů, kteří síť využívají se v čase mění a zrovna tak objem datového toku, který koncovými uzly protéká.

Obrázek 17: Celkový datový tok skrze Tor síť



Jak přiložený graf dokládá, využití sítě se v čase mění a neustále roste. Zatímco zelená křivka reprezentuje celkový potenciál sítě, růžová křivka dokumentuje skutečný objem datového toku, který byl sítí přenesen. Tato data reprezentují celou síť zatímco stížnosti na zneužívání se týkají pouze serverů provozovaných organizací Torservers.net. Tyto konkrétní servery historicky reprezentují kolem jednoho procenta všech Exit uzlů celé sítě a z toho důvodu se domnívám, že se jedná o reprezentativní vzorek a pro účely analýzy vzájemného vztahu mezi datovým tokem a počtem stížností použiji tato data v přímé souvislosti.

Historická data o datovém toku v celé síti jsou dostupné ve formátu .csv[124] na oficiálních stránkách projektu a jsou agregována po dnech. Po napojení informace o počtu stížností v daném dni je možné analyzovat vzájemný vztah těchto dvou proměnných. Za tímto účelem je vhodné použít korelační analýzu, která je používána jak doc. Bedáňová uvádí: "*Pro zjištění těsnosti závislosti (síly vztahu) dvou náhodných spojitých proměnných. V nejobecnějším smyslu, slovo „korelace" označuje míru stupně asociace dvou veličin. Dvě veličiny jsou*

[124] Tor Project: Metrics [online]. [cit. 2017-03-12]. Dostupné z: https://metrics.torproject.org/stats/bandwidth.csv

korelované (asociované), jestliže určité hodnoty jedné veličiny mají tendenci se vyskytovat společně s určitými hodnotami druhé veličiny. Jde tu tedy o dvoustranný reciproční vztah dvou náhodných proměnných X a Y, kdy nemá smysl uvažovat, že jedna z proměnných je závislá a druhá nezávislá; obě jsou závislé vzájemně."[125]

Korelační koeficient mezi počtem stížností a objemem přenesených dat koncovými uzly sítě se rovná **0.59** v rozmezí mezi -1 a 1 na vzorku 1734 pozorování respektive dnů. Obecně jsou korelace větší než 0.5 považovány za silné a dovolím si tedy tvrdit, že počet stížností je silně ovlivněn velikostí datového toku. Vzhledem k velmi výrazně převyšujícímu počtu stížností odvolávajících se na autorské právo nemusí velikost datového toku pouze indikovat počet uživatelů, ale objem stahovaného obsahu, který je v případě audiovizuálního obsahu v porovnání s tradičními webovými stránkami výrazně větší.

3.2.2.4 Shrnutí a limity výzkumu

Na rozdíl od copyrightových stížností je u těch zbylých, které obsahují stížností na hacking, nenávistné projevy, ale i spam, znatelný jen mírný nárůst v čase. Denní počty stížností jsou do velké míry korelovány s denním datovým tokem Tor sítě, který v čase rovněž narůstá, stejně tak jako počet uživatelů.

Podobně jako u copyrightových stížností je velké množství stížností zasláno společnostmi specializujícími se na kybernetickou bezpečnost, nebo na ochranu proti spamu. Druhá největší skupina obsahuje velké množství spamových zpráv stejně tak jako nečitelné zašifrované zprávy. V neposlední řadě je v datech i nezanedbatelné množství stížností psaných lidmi, kteří si často stěžují na nenávistné projevy a hackování v osobních rovinách.

Ručně psaným stížnostem by bylo rovněž možné přisuzovat větší váhu, neboť se dá usuzovat, že zde došlo k zneužití v takové míře, že jej poškozený chtěl aktivně řešit. Také je zde minimální riziko falešné detekce oproti strojově generovaným stížnostem.

[125] Korelační analýza [online]. [cit. 2017-03-12]. Dostupné z: http://cit.vfu.cz/statpotr/potr/teorie/predn5/linearni.htm

Jhaveri et al.[126] ve výsledcích výzkumu zaměřeného na reportování zneužívání potvrzuje dominantní roli specializovaných společností. Pod jejich ochranná křídla se uchylují provozovatelé jednotlivých serverů na místo řešení zneužívání vlastními silami. Dále zmiňují pozoruhodnou schopnost sebe regulace internetového prostředí, ve kterém jsou na základě hlášení odstavovány podvodné stránky a další maligní servery bez zásahu soudů či policie. Tuto skutečnost do určité míry potvrzují i analyzovaná data, neboť mezi milióny stížností se objevuje pouze okolo deseti případů, kdy se do řešení zneužívání zapojila státní moc.

Prezentovaná kategorizace stížností vychází z výsledků shlukovacího algoritmu k-means a z pojmenování výsledných shluků s ruční verifikací na omezeném vzorku. Tento algoritmus funguje čistě na základě podobnosti dokumentů a nepracuje s jejich obsahem na sémantické úrovni. Při použití pokročilejších analytických metod a strojového učení by zřejmě bylo možné kategorizovat stížnosti přesněji.

Analýza je rovněž ovlivněna zkoumanými daty, která pocházejí z jednoho zdroje a reprezentují signifikantní, ale pouze omezenou část sítě o velikosti přibližně jednoho procenta z celkového objemu dat, který skrze exit uzly protéká. Provozovatelé těchto konkrétních serverů jsou rovněž velmi aktivní v řešení stížností a šíření osvěty ohledně Tor sítě, což může mít za následek zkreslení dat oproti serverům, na kterých nejsou aplikovány žádné limity a pravidla.

[126] JHAVERI, Mohammad Hanif, Orcun CETIN, Carlos GAÑÁN, Tyler MOORE a Michel Van EETEN. Abuse Reporting and the Fight Against Cybercrime. ACM Computing Surveys [online]. 2017, 49(4), 1-27 [cit. 2017-07-24]. DOI: 10.1145/3003147. ISSN 03600300. Dostupné z: http://dl.acm.org/citation.cfm?doid=3022634.3003147

4. Závěr

Tato práce si kladla za cíl prozkoumat způsoby využívání, případně zneužívání, anonymizační sítě Tor. V praktické části poté na základě získaných dat kvantifikovat právě poměr mezi užíváním a zneužíváním Tor sítě a ověřit, zda-li se v případě zneužívání jedná o tak závažný problém, jak se často ozývá z médií a z úst kritiků. Ti se často zaměřují na takzvané Onion Services, které s oblibou označují jako Darknet. Onion Services tvoří pouze několik málo procent z celkového datového toku Tor sítě a obsah těchto stránek zdaleka není výhradně závadný.

Analýza dat na první pohled prokazuje dlouhodobý trend nárůstu stížností na maligní činnost za využití Tor sítě, na pohled druhý mohou za tento nárůst primárně stížnosti ve jménu Digital Millennium Copyright Act – amerického zákona, který doplnil dřívější zákon o autorském právu s ohledem na nové technologie. Z celkového počtu téměř 3 miliónů analyzovaných stížností adresuje méně než jedno procento jiný typ zneužívání než porušování autorského zákona za použití P2P sítí. Toto zjištění vyvrací hypotézu a převažujícím typu zneužívání, které jsem se domníval, že jsou počítačové útoky.

Hypotéza o narůstajícím počtu stížností v čase se potvrdila u celkových čísel, nicméně při samostatné analýze stížností, které se neodvolávají na DMCA, se trend nárůstu neprojevuje, nebo jen velmi mírně. Mezi těmito ostatními stížnostmi se nejčastěji objevuje tématika pokusů o neoprávněný přístup k redakčnímu systému, projevů nenávisti (tzv. hate speech), nebo spamu.

Absolutní počty stížnosti je ale nutné vnímat v souvislosti s narůstajícím počtem uživatelů sítě a její celkovou datovou propustností, která rovněž roste. Možnosti stahování velkých souborů nebo streamování videa nebyly dříve skrze Tor síť téměř myslitelné a i z tohoto důvodu je nárůst copyrightových stížností tak velký. Tento závěr potvrzuje i korelační analýza mezi historickým datovým tokem a počtem stížností v daných dnech za posledních 5 let, která potvrzuje silný vztah mezi těmito dvěma veličinami. Potvrzuje se tak vyřčená hypotéza o existenci vztahu mezi počtem stížností a velikostí datového toku.

Za předpokladu, že analyzovaná data reprezentují jedno procento celkového datového potenciálu exit serverů celé Tor sítě a obslouží tak jedno procento ze všech uživatelů Toru, lze odvodit poměr benigních a maligních uživatelů sítě. Rovněž za předpokladu, že jedna stížnost reprezentuje jednoho uživatele a stížnosti se neduplikují. Samotný poměr je při průměru 9,57 obdržených stížností na jedno procento průměrného počtu denních uživatelů[127] **1:2133**. V případě zahrnutí copyrightových stížností se průměr počtu stížností za den zvýší na 3 908 a rovněž poměr benigních a maligních uživatelů na **1:5**. Přestože je tento poměr vysoký a nepochybně vyšší než u běžných sítí, tak vyvrací častá tvrzení, že většina požadavků z Tor sítě je maligní.[128] Pro srovnání společnost Akamai uvádí 1:380 maligních požadavků z uvnitř Tor sítě oproti 1:11500 mimo Tor síť.[129]

Na základě analyzovaných dat a výše uvedených faktů si dovoluji tvrdit, že zneužívání Tor sítě, alespoň v případě serverů provozovaných organizací Torservers.net, není prováděno ve velkém měřítku a většina požadavků ze sítě do vnějšího internetu je nezávadná. Potvrzují se tak závěry výzkumu[130] o dvojím zacházení s uživateli Toru, kde je všichni uživatelé trestáni za prohřešky menšiny.

Jakkoliv se motivace uživatelů pro použití anonymizačního nástroje Tor různí a oproti běžným uživatelům internetu je tato motivace násobně častěji spíše maligní, nelze přehlížet v současnosti nezastupitelnou roli toho nástroje z pohledu základních lidských práv jako svoboda slova, listovní tajemství, nebo právo na soukromí.

[127] Tor Project: Users - Tor Metrics [online]. [cit. 2017-07-25]. Dostupné z: https://metrics.torproject.org/userstats-relay-country.html
[128] PRINCE, Matthew. The Trouble with Tor [online]. [cit. 2017-05-08]. Dostupné z: https://blog.cloudflare.com/the-trouble-with-tor/
[129] AKAMAI. State of the Internet / Security Report [online]. [cit. 2017-05-08]. Dostupné z: https://media.scmagazine.com/documents/144/q2_2015_soti_security_report_-_35820.pdf
[130] KHATTAK, Sheharbano, et al. Do you see what i see? differential treatment of anonymous users. In: Network and Distributed System Security Symposium. 2016. [cit. 2017-05-08]. Dostupné z: https://www.cl.cam.ac.uk/~sk766/publications/ndss16_tor_differential.pdf

Seznam použitých zdrojů

1) ISA. EGovernment in Estonia [online]. , 53 [cit. 2017-07-04]. Dostupné z: https://joinup.ec.europa.eu/sites/default/files/ckeditor_files/files/eGovernment%20in%20Estonia%20-%20February%202016%20-%2018_00_v4_00.pdf
2) Tor Project: Anonymity Online [online]. [cit. 2017-03-11]. Dostupné z: https://www.torproject.org/
3) ZAHORSKY, Ingmar. Tor, Anonymity, and the Arab Spring: An Interview with Jacob Appelbaum [online]. [cit. 2017-04-30]. Dostupné z: http://www.monitor.upeace.org/innerpg.cfm?id_article=816
4) MOORE, Daniel a Thomas RID. Cryptopolitik and the Darknet. Survival [online]. 2016, 58(1), 7-38 [cit. 2017-05-27]. DOI: 10.1080/00396338.2016.1142085. ISSN 0039-6338. Dostupné z: http://www.tandfonline.com/doi/full/10.1080/00396338.2016.1142085
5) Robert Galbraith [online]. [cit. 2017-03-11]. Dostupné z: http://robert-galbraith.com/about/
6) Důsledky informačního zahlcení. WikiSofia [online]. [cit. 2017-05-07]. Dostupné z: https://wikisofia.cz/wiki/Důsledky_informačního_zahlcení
7) DINUCCI, Darcy. Fragmented Future [online]. 1999 [cit. 2017-03-11]. Dostupné z: http://darcyd.com/fragmented_future.pdf
8) MARX, Gary T. What's in a Name? Some Reflections on the Sociology of Anonymity. The Information Society [online]. 1999, vol. 15, issue 2, s. 99-112 [cit. 2017-03-11]. DOI: 10.1080/019722499128565. Dostupné z: http://www.tandfonline.com/doi/abs/10.1080/019722499128565
9) HAYNE, Stephen C. a Ronald E. RICE. Attribution accuracy when using anonymity in group support systems. International Journal of Human-Computer Studies [online]. 1997, vol. 47, issue 3, s. 429-452 [cit. 2014-07-17]. DOI: 10.1006/ijhc.1997.0134. Dostupné z: http://linkinghub.elsevier.com/retrieve/pii/S1071581997901348
10) PFITZMANN, Andreas a Marit HANSEN. Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology [online]. 2005, , 43 [cit. 2017-03-11]. Dostupné z: https://www.freehaven.net/anonbib/cache/terminology.pdf
11) KANG, Ruogu, Stephanie BROWN a Sara KIESLER. Why do people seek anonymity on the internet?. Proceedings of the SIGCHI Conference on Human Factors in Computing

Systems - CHI '13 [online]. New York, New York, USA: ACM Press, 2013, s. 2657- [cit. 2017-07-04]. DOI: 10.1145/2470654.2481368. Dostupné z: http://dl.acm.org/citation.cfm?doid=2470654.2481368

12) TURKLE, Sherry. Life on the screen: identity in the age of the Internet. New York: Simon, c1995, 347 s. ISBN 978-068-4833-484. WARD, Mark. UK government tackles wrongly-blocked websites. BBC News [online]. 2014 [cit. 2017-07-04]. Dostupné z: http://www.bbc.com/news/technology-25962555
13) 4chan [online]. [cit. 2017-07-04]. Dostupné z: http://www.4chan.org/
14) BERNSTEIN, Michael S., Andres MONROY-HERNANDEZ, Drew HARRY, Paul ANDRE, Katrina PANOVICH a Greg VARGAS. 4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community [online]. [cit. 2017-07-04]. Dostupné z: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2873/4398
15) HALLIDAY, Josh. SXSW 2011: 4Chan founder Christopher Poole on anonymity and creativity [online]. 2011 [cit. 2017-07-04]. Dostupné z: https://www.theguardian.com/technology/2011/mar/13/christopher-poole-4chan-sxsw-keynote-speech
16) FREBERG, Karen, Kristin GRAHAM, Karen MCGAUGHEY a Laura A. FREBERG. Who are the social media influencers? A study of public perceptions of personality. Public Relations Review [online]. 2011, 37(1), 90-92 [cit. 2017-07-04]. DOI: 10.1016/j.pubrev.2010.11.001. ISSN 03638111. Dostupné z: http://linkinghub.elsevier.com/retrieve/pii/S0363811110001207
17) KANG, Ruogu, Stephanie BROWN a Sara KIESLER. Why do people seek anonymity on the internet?. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13 [online]. New York, New York, USA: ACM Press, 2013, s. 2657- [cit. 2017-07-04]. DOI: 10.1145/2470654.2481368.
18) HUITEMA, Christian. Routing in the Internet [online]. Englewood Cliffs, N.J.: Prentice Hall PTR, c1995 [cit. 2017-07-13]. ISBN 978-0131321922.
19) SPECIFICATION OF INTERNET TRANSMISSION CONTROL PROGRAM [online]. [cit. 2017-07-04]. Dostupné z: https://tools.ietf.org/html/rfc675
20) PATERKA, Jiří. Na počátku byl ARPANET .. Computerworld [online]. 1995, (4) [cit. 2017-04-30]. Dostupné z: http://www.earchiv.cz/a95/a504c502.php3
21) The Internet Engineering Task Force [online]. [cit. 2017-07-13]. Dostupné z: https://www.ietf.org/

22) Internet Society [online]. [cit. 2017-07-13]. Dostupné z: http://www.internetsociety.org/
23) BERNERS-LEE, Tim. a Mark. FISCHETTI. Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor. San Francisco: HarperSanFrancisco, c1999. ISBN 978-0062515872.
24) KATZ-BASSETT, Ethan, John P. JOHN, Arvind KRISHNAMURTHY, David WETHERALL, Thomas ANDERSON a Yatin CHAWATHE. Towards IP geolocation using delay and topology measurements. Proceedings of the 6th ACM SIGCOMM on Internet measurement - IMC '06 [online]. New York, New York, USA: ACM Press, 2006, , 71- [cit. 2017-07-17]. DOI: 10.1145/1177080.1177090. ISBN 1595935614. Dostupné z: http://portal.acm.org/citation.cfm?doid=1177080.1177090
25) D., Kristol a Montulli L. HTTP State Management Mechanism [online]. 2000 [cit. 2017-05-08]. Dostupné z: https://tools.ietf.org/html/rfc2965
26) Information providers guide The EU Internet Handbook: Cookies [online]. [cit. 2017-05-07]. Dostupné z: http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm
27) ETZIONI, Amitai. Privacy Isn't Dead Yet [online]. 1999 [cit. 2017-07-17]. Dostupné z: http://www.nytimes.com/1999/04/06/opinion/privacy-isn-t-dead-yet.html
28) E. W. Felten. If You’re Going to Track Me, Please Use Cookies. https://freedom-to-tinker.com/blog/felten/ifyoure-going-track-me-please-use-cookies/, 2009.
29) ECKERSLEY, Peter. How Unique Is Your Web Browser? [online]. , 1 [cit. 2017-05-08]. DOI: 10.1007/978-3-642-14527-8_1. Dostupné z: http://link.springer.com/10.1007/978-3-642-14527-8_1
30) SWEENEY, Latanya. Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000. Dostupné z: https://dataprivacylab.org/projects/identifiability/paper1.pdf
31) W3schools: HTML5 Canvas [online]. [cit. 2017-04-17]. Dostupné z: https://www.w3schools.com/html/html5_canvas.asp
32) MOWERY, Keaton a Hovav SHACHAM. Pixel Perfect: Fingerprinting Canvas in HTML5 [online]. , 12 [cit. 2017-04-17]. Dostupné z: https://cseweb.ucsd.edu/~hovav/dist/canvas.pdf
33) Alexa: Top Sites [online]. [cit. 2017-07-17]. Dostupné z: http://www.alexa.com/topsites
34) ACAR, Gunes, Christian EUBANK, Steven ENGLEHARDT, Marc JUAREZ, Arvind NARAYANAN a Claudia DIAZ. The Web Never Forgets. Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS '14 [online].

New York, New York, USA: ACM Press, 2014, , 674-689 [cit. 2017-07-17]. DOI: 10.1145/2660267.2660347. ISBN 9781450329576. Dostupné z: http://dl.acm.org/citation.cfm?doid=2660267.2660347

35) MILLER, Brad, Ling HUANG, A. D. JOSEPH a J. D. TYGAR. I Know Why You Went to the Clinic: Risks and Realization of HTTPS Traffic Analysis [online]. , 143 [cit. 2017-07-17]. DOI: 10.1007/978-3-319-08506-7_8. Dostupné z: http://link.springer.com/10.1007/978-3-319-08506-7_8

36) SYVERSON, Paul. Onion Routing: Brief Selected History [online]. [cit. 2017-07-04]. Dostupné z: https://www.onion-router.net/History.html

37) Tor Project: What are Entry Guards? [online]. [cit. 2017-04-09]. Dostupné z: https://www.torproject.org/docs/faq.html.en#EntryGuards

38) Tor Project: Atlas [online]. [cit. 2017-04-09]. Dostupné z: https://atlas.torproject.org/#search/flag:authority

39) Tor Project: About [online]. [cit. 2017-07-04]. Dostupné z: https://www.torproject.org/about/overview.html.en

40) SYVERSON, Paul. Onion Routing: Brief Selected History [online]. [cit. 2017-07-04]. Dostupné z: https://www.onion-router.net/History.html

41) REED, Michael G., Paul F. SYVERSON a David M. GOLDSCHLAG. Anonymous Connections and Onion Routing. IEEE Journal on Selected Areas in Communication Special Issue on Copyright and Privacy Protection. 1998.

42) JAP -- ANONYMITY & PRIVACY [online]. [cit. 2017-04-08]. Dostupné z: https://anon.inf.tu-dresden.de/index_en.html

43) Tor 0.3.0.5-rc is released: almost stable!. Tor Project [online]. [cit. 2017-04-15]. Dostupné z: https://blog.torproject.org/blog/tor-0305-rc-released-almost-stable

44) Extended Support Release. Firefox [online]. [cit. 2017-04-15]. Dostupné z: https://www.mozilla.org/en-US/firefox/organizations/

45) Electronic Frontier Foundation: HTTPS Everywhere [online]. [cit. 2017-07-04]. Dostupné z: https://www.eff.org/https-everywhere

46) NoScript [online]. [cit. 2017-07-04]. Dostupné z: https://noscript.net/

47) Tor Browser: The Design and Implementation of the Tor Browser [DRAFT] [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/

48) Tor Browser: Proxy Obedience [online]. [cit. 2017-04-17]. Dostupné z:

https://www.torproject.org/projects/torbrowser/design/#proxy-obedience

49) Tor Browser: State Separation [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#state-separation

50) Tor Browser: Disk Avoidance [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#disk-avoidance

51) Asmjscache: should not store cache entries when private browsing is enabled. BugZilla [online]. [cit. 2017-04-16]. Dostupné z: https://bugzilla.mozilla.org/show_bug.cgi?id=1047105

52) Tor Browser: Application Data Isolation [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#app-data-isolation

53) Tor Browser: Cross-Origin Identifier Unlinkability [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#identifier-linkability

54) Tor Browser: Cross-Origin Fingerprinting Unlinkability [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#fingerprinting-linkability

55) W3schools: HTML5 Canvas [online]. [cit. 2017-04-17]. Dostupné z: https://www.w3schools.com/html/html5_canvas.asp

56) MOWERY, Keaton a Hovav SHACHAM. Pixel Perfect: Fingerprinting Canvas in HTML5 [online]. , 12 [cit. 2017-04-17]. Dostupné z: https://cseweb.ucsd.edu/~hovav/dist/canvas.pdf

57) Bug #6253: Add canvas image extraction prompt. [online]. [cit. 2017-04-17]. Dostupné z: https://gitweb.torproject.org/tor-browser.git/commit/?h=tor-browser-45.8.0esr-6.5-2&id=526e6d0bc5c68d8c409cbaefc231c71973d949cc

58) Tor Browser: Long-Term Unlinkability [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/projects/torbrowser/design/#new-identity

59) Tor Project: Users of Tor [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/about/torusers.html.en

60) MUFFETT, Alec. Making Connections to Facebook more Secure [online]. [cit. 2017-04-17]. Dostupné z: https://www.facebook.com/notes/protect-the-graph/making-connections-to-facebook-more-secure/1526085754298237

61) Prisoner of conscience. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-04-17]. Dostupné z: https://en.wikipedia.org/wiki/Prisoner_of_conscience

62) Reporters sans frontières: Online survival Kit [online]. [cit. 2017-04-17]. Dostupné z: https://rsf.org/en/online-survival-kit
63) Tor: Sponsors [online]. [cit. 2017-04-17]. Dostupné z: https://www.torproject.org/about/sponsors.html.en
64) Department of Justice Official Tells Hundred Federal Judges to Use Tor [online]. [cit. 2017-04-17]. Dostupné z: https://motherboard.vice.com/en_us/article/department-of-justice-official-tells-hundred-federal-judges-to-use-tor
65) ZAHORSKY, Ingmar. Tor, Anonymity, and the Arab Spring: An Interview with Jacob Appelbaum [online]. [cit. 2017-04-30]. Dostupné z: http://www.monitor.upeace.org/innerpg.cfm?id_article=816
66) Turkey blocks Wikipedia under law designed to protect national security [online]. [cit. 2017-04-30]. Dostupné z: https://www.theguardian.com/world/2017/apr/29/turkey-blocks-wikipedia-under-law-designed-to-protect-national-security
67) Censorship of Wikipedia [online]. [cit. 2017-04-30]. Dostupné z: https://en.wikipedia.org/wiki/Censorship_of_Wikipedia
68) KHATTAK, Sheharbano, et al. Do you see what i see? differential treatment of anonymous users. In: Network and Distributed System Security Symposium. 2016. [cit. 2017-05-08]. Dostupné z: https://www.cl.cam.ac.uk/~sk766/publications/ndss16_tor_differential.pdf
69) Alexa: Top sites [online]. [cit. 2017-05-08]. Dostupné z: http://www.alexa.com/topsites
70) PRINCE, Matthew. The Trouble with Tor [online]. [cit. 2017-05-08]. Dostupné z: https://blog.cloudflare.com/the-trouble-with-tor/
71) PERRY, Mike. The Trouble with CloudFlare [online]. [cit. 2017-05-08]. Dostupné z: https://blog.torproject.org/blog/trouble-cloudflare
72) AKAMAI. State of the Internet / Security Report [online]. [cit. 2017-05-08]. Dostupné z: https://media.scmagazine.com/documents/144/q2_2015_soti_security_report_-_35820.pdf
73) Silk Road operator Ross Ulbricht sentenced to life in prison. The Guardian [online]. [cit. 2017-06-18]. Dostupné z: https://www.theguardian.com/technology/2015/may/29/silk-road-ross-ulbricht-sentenced
74) LADEGAARD, Isak. We Know Where You Are, What You Are Doing and We Will Catch You. The British Journal of Criminology. 2017, , -. DOI: 10.1093/bjc/azx021. ISSN 0007-0955. Dostupné také z: https://academic.oup.com/bjc/article-

lookup/doi/10.1093/bjc/azx021

75) Some statistics about onions [online]. [cit. 2017-05-27]. Dostupné z: https://blog.torproject.org/blog/some-statistics-about-onions

76) MOORE, Daniel a Thomas RID. Cryptopolitik and the Darknet. Survival [online]. 2016, 58(1), 7-38 [cit. 2017-05-27]. DOI: 10.1080/00396338.2016.1142085. ISSN 0039-6338. Dostupné z: http://www.tandfonline.com/doi/full/10.1080/00396338.2016.1142085

77) Support vector machine. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-05-27]. Dostupné z: https://en.wikipedia.org/wiki/Support_vector_machine

78) GOLLNICK, Clare a Emily WILSON. The Truth About the Dark Web [online]. [cit. 2017-05-27]. Dostupné z: https://www.scribd.com/document/329783168/The-Truth-About-the-Dark-Web

79) KHATTAK, Sheharbano, et al. Do You See What I See? Differential Treatment of Anonymous Users [online]. [cit. 2017-07-04]. Dostupné z: http://www.cl.cam.ac.uk/~sk766/publications/ndss16_tor_differential.pdf. 2016.

80) Abuse Templates. Torservers.net [online]. [cit. 2017-02-05]. Dostupné z: https://www.torservers.net/wiki/abuse/templates

81) Tor Network Status. Tor Status [online]. [cit. 2017-02-05]. Dostupné z: https://torstatus.blutmagie.de/index.php

82) Tor Project: Metrics [online]. [cit. 2017-07-04]. Dostupné z: https://metrics.torproject.org/bandwidth-flags.html

83) Abuse. Torservers.net [online]. [cit. 2017-02-05]. Dostupné z: https://www.torservers.net/abuse.html

84) Telemediengesetz (TMG). Bundesministerium der Justiz und für Verbraucherschutz [online]. [cit. 2017-02-05]. Dostupné z: http://www.gesetze-im-internet.de/tmg/__15.html

85) Komprimovaný dokument ve formátu .xz využívající algoritmus LZMA2 a zabalený pomocí svobodného formátu .tar.

86) Icedove. Debian.org [online]. [cit. 2017-02-05]. Dostupné z: https://wiki.debian.org/Icedove

87) Maildir. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-02-05]. Dostupné z: https://cs.wikipedia.org/wiki/Maildir

88) Metada jsou takzvaná data o datech a často obsahují strojově vytvořené informace o

časových údajích, kdy byla informace vytvořena či změněna.

89) R. The R Project for Statistical Computing [online]. [cit. 2017-02-05]. Dostupné z: https://www.r-project.org/

90) Cyklus for je řídící struktura, sloužící pro iteraci přes všechny prvky v kolekci - v tomto případě přes všechny soubory s koncovkou .mail.

91) CLEVELAND, William S. Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association [online]. 1979, 74(368), 829-836 [cit. 2017-07-04]. DOI: 10.1080/01621459.1979.10481038. ISSN 0162-1459. Dostupné z: http://www.tandfonline.com/doi/abs/10.1080/01621459.1979.10481038

92) KONCZIOVÁ, Anita. Neparametrické odhady regresní funkce. Brno, 2013. Dostupné také z: http://is.muni.cz/th/380229/prif_b/Bakalarska_praca.pdf

93) U.S. COPYRIGHT OFFICE SUMMARY. THE DIGITAL MILLENNIUM COPYRIGHT ACT OF 1998 [online]. 1998 [cit. 2017-02-05]. Dostupné z: https://www.copyright.gov/legislation/dmca.pdf

94) Whois. Domain Tools [online]. [cit. 2017-02-05]. Dostupné z: http://whois.domaintools.com/copyright-notice.com

95) GoDaddy Inc. [online]. [cit. 2017-02-05]. Dostupné z: https://godaddy.com/

96) Whois. Domain Tools [online]. [cit. 2017-02-05]. Dostupné z: http://whois.domaintools.com/copyright-compliance.com

97) Domain.com [online]. [cit. 2017-02-05]. Dostupné z: http://www.domain.com/

98) IP-Echelon [online]. [cit. 2017-02-05]. Dostupné z: https://www.ip-echelon.com/

99) NFOrce Entertainment B.V. [online]. [cit. 2017-02-05]. Dostupné z: https://www.nforce.com/

100) Voxility, LLC [online]. [cit. 2017-02-05]. Dostupné z: https://www.voxility.com/

101) The Icelandic Pirate Party [online]. [cit. 2017-02-05]. Dostupné z: http://piratar.is/

102) Global Internet Phenomena 2016 [online]. 2016 [cit. 2017-02-05]. Dostupné z: https://www.sandvine.com/downloads/general/global-internet-phenomena/2016/global-internet-phenomena-apac-mea.pdf

103) DINGLEDINE, Roger. Bittorrent over Tor isn't a good idea [online]. 2010 [cit. 2017-02-05]. Dostupné z: https://blog.torproject.org/blog/bittorrent-over-tor-isnt-good-idea

104) BitTorrent Traffic Share Drops to New Low [online]. 2015 [cit. 2017-02-05].

Dostupné z: https://torrentfreak.com/bittorrent-traffic-share-drops-to-new-low-150918/

105) @midnight [online]. [cit. 2017-02-12]. Dostupné z: http://www.cc.com/shows/-midnight

106) Spambot. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-02-19]. Dostupné z: https://cs.wikipedia.org/wiki/Spambot

107) Routing Information Protocol. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-02-12]. Dostupné z: https://cs.wikipedia.org/wiki/Routing_Information_Protocol

108) SpamCop. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-02-13]. Dostupné z: https://en.wikipedia.org/wiki/SpamCop

109) Dispute Resolution: Bounce message recipients and end users. SpamCop.net [online]. [cit. 2017-02-13]. Dostupné z: https://www.spamcop.net/fom-serve/cache/405.html

110) Distributed Denial of Service je druh počítačového útoky, kdy dojde k přetížení serveru velkým množstvím požadavků načež je znepřístupněn ostatním uživatelům.

111) Content Management System, česky systém pro správu obsahu, je software určený pro spravování zpravidla webového obsahu bez nutnosti úpravy kódu - například Wordpress.

112) The unintended consequences of the sacred proxy. Webiron.com [online]. [cit. 2017-02-15]. Dostupné z: https://blog.webiron.com/index.php/2015/11/23/the-unintended-consequences-of-the-sacred-proxy/

113) WUBTHECAPTAIN. Webiron requesting to block several /24 subnet [online]. 2015 [cit. 2017-02-15]. Dostupné z: https://lists.torproject.org/pipermail/tor-relays/2015-December/008259.html

114) [PROPOSAL] Create moderated version of ba.broadcast, ba.broadcast.moderated. Google Groups [online]. [cit. 2017-02-15]. Dostupné z: https://groups.google.com/forum/#!topic/ba.broadcast/3fvYF14mKG4

115) SALTON, Gerard a Christopher BUCKLEY. Term-weighting approaches in automatic text retrieval. Information Processing [online]. 1988, 24(5), 513-523 [cit. 2017-06-17]. DOI: 10.1016/0306-4573(88)90021-0. ISSN 03064573. Dostupné z: http://linkinghub.elsevier.com/retrieve/pii/0306457388900210

116) JAMES E. GENTLE. Matrix algebra theory, computations, and applications in statistics [online]. [Online-Ausg.]. New York, N.Y: Springer, 2007 [cit. 2017-06-17]. ISBN 978-038-7708-737.

117) KUČERA, Jiří. Algoritmus k-means [online]. [cit. 2017-05-27]. Dostupné z: https://is.muni.cz/th/172767/fi_b/5739129/web/web/kmeans.html

118) KUČERA, Jiří. Nehierarchické metody shlukování [online]. [cit. 2017-05-27]. Dostupné z: https://is.muni.cz/th/172767/fi_b/5739129/web/web/nehiermet.html#mqm

119) WordPress: Brute Force Attacks [online]. [cit. 2017-05-28]. Dostupné z: https://codex.wordpress.org/Brute_Force_Attacks

120) GAYER, Ofer. Semalt Hijacks Hundreds of Thousands of Computers to Launch a Referrer Spam Campaign [online]. [cit. 2017-05-28]. Dostupné z: https://www.incapsula.com/blog/semalt-botnet-spam.html

121) CollecTor: Exit list [online]. [cit. 2017-03-11]. Dostupné z: https://collector.torproject.org/archive/exit-lists/

122) MaxMind: IP Geolocation and Online Fraud Prevention [online]. [cit. 2017-02-19]. Dostupné z: https://www.maxmind.com/

123) List of sovereign states. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-03-11]. Dostupné z: https://en.wikipedia.org/wiki/List_of_sovereign_states

124) Tor Project: Metrics [online]. [cit. 2017-03-12]. Dostupné z: https://metrics.torproject.org/stats/bandwidth.csv

125) Korelační analýza [online]. [cit. 2017-03-12]. Dostupné z: http://cit.vfu.cz/statpotr/potr/teorie/predn5/linearni.htm

126) JHAVERI, Mohammad Hanif, Orcun CETIN, Carlos GAÑÁN, Tyler MOORE a Michel Van EETEN. Abuse Reporting and the Fight Against Cybercrime. ACM Computing Surveys [online]. 2017, 49(4), 1-27 [cit. 2017-07-24]. DOI: 10.1145/3003147. ISSN 03600300. Dostupné z: http://dl.acm.org/citation.cfm?doid=3022634.3003147

127) Tor Project: Users - Tor Metrics [online]. [cit. 2017-07-25]. Dostupné z: https://metrics.torproject.org/userstats-relay-country.html

128) PRINCE, Matthew. The Trouble with Tor [online]. [cit. 2017-05-08]. Dostupné z: https://blog.cloudflare.com/the-trouble-with-tor/

129) AKAMAI. State of the Internet / Security Report [online]. [cit. 2017-05-08]. Dostupné z:

https://media.scmagazine.com/documents/144/q2_2015_soti_security_report_-_35820.pdf

130) KHATTAK, Sheharbano, et al. Do you see what i see? differential treatment of anonymous users. In: Network and Distributed System Security Symposium. 2016. [cit. 2017-05-08]. Dostupné z: https://www.cl.cam.ac.uk/~sk766/publications/ndss16_tor_differential.pdf