Česká zemědělská univerzita v Praze

Provozně ekonomická fakulta

Diplomová práce na téma:

Antispamové filtry – nasazení a srovnání účinnosti

Vypracovala: Magdaléna Matoušková

Vedoucí diplomové práce: Alexandr Vasilenko

Praha 2009

ČESTNÉ PROHLÁŠENÍ

Prohlašuji, že jsem diplomovou práci na téma Antispamové filtry – nasazení a srovnání účinnosti vypracovala samostatně a použila k tomu jen pramenů, které cituji a uvádím v přiloženém seznamu literatury, svých poznatků a konzultací s vedoucím práce.

_____ _____

Datum Magdaléna Matoušková

PODĚKOVÁNÍ

Děkuji tímto panu Ing. Alexandru Vasilenkovi za cenné rady a odborné vedení, které mi poskytoval v průběhu vypracování této diplomové práce.

ANTISPAMOVÉ FILTRY – NASAZENÍ A SROVNÁNÍ ÚČINNOSTI

ANTISPAM FILTERS – STOCKING AND CONFRONTATION OPERATION

SOUHRN:

Tato práce pojednává o rozšířeném jevu zvaném spam. V úvodu práce je vysvětlena funkční stránka elektronické pošty, poté následuje podrobné vysvětlení problematiky spamu, jeho původ, historie, druhy a výskyt. Dále jsou v průběhu této práce vylíčeny nejúčinnější způsoby ochrany proti spamu. V závěru je pak zvolen jeden z nejúčinnějších antispamových softwarů, je dopodrobna popsána jeho funkčnost, účinnost a modifikovatelnost a je navržena ukázková instalace na operační systém Windows.

SUMMARY:

This essay adresses the widespread phenomenon called spam. In the entery it discusses the way electronic email works, then it deals with detailed explanation of spam itself, its origin, history, kinds and occurences. Further on this essay describes the most effective ways of protection against spam. Finally, one of the most effective antispam software is chosen, it's functionality, effectiveness and modifyability is described and it's applicatory installation on OS Windows is provided.

1 ÚVOD

Pan Bill Gates se v roce 2004 opravdu zmýlil, když na ekonomickém fóru v Davosu prohlásil, že do dvou let bude spam vyřešen. Asi netušil, s kým má tu čest. Spameři jsou stále více vynalézaví a nezastaví se před ničím. Spam tedy rozhodně nezmizel, dokonce ani neklesají jeho počty, vývoj spamu má spíše rostoucí tendenci.

Spamem se proto zabývá stále větší množství subjektů, ať již jde o boj nás, koncových uživatelů, nebo organizací, pro něž je spam zdrojem živobytí. Vývojáři vyvíjí stále nová antispamová řešení, která čelí novým vynalézavým útokům spamerů. Různé organizace (MAAWG) sdružují uživatele emailů v boji proti spamu. Distributoři spořičů obrazovek se aktivně podílejí na boji proti spamu šířením svého speciálního spořiče, který generuje množství požadavků na známé servery spammerů a zahlcuje je tím. Řešení spamu je zkrátka denní chléb pro mnoho lidí. Důvod je jasný, spam nás otravuje a okrádá o čas.

Spam obtěžuje lidstvo již více než 30 let. V České Republice je v dnešní době 8 z 10 emailů nevyžádaných. Podle Serveru o českém internetu Lupa.cz je většina spamů zaměřena na propagaci farmaceutických výrobků, nebo na akciové trhy.

Tato práce klade otázky a hledá odpovědi týkající se nevyžádané elektronické pošty, zabývá se jejím původem, historií a právními aspekty, zkoumá její možnou filtrační metodiku, a hlavně navrhuje jeden z možných způsobů ochrany proti spamu. Tento způsob v podobě antispamového software dopodrobna popisuje, uvádí důvody pro jeho volbu a v neposlední řadě poskytuje ukázkovou instalaci.

Tato diplomová práce vychází z problematiky řešené v mé bakalářské práci na téma Spam a Despamové techniky.

2 CÍL PRÁCE A METODIKA

Cílem této práce je pokud možno komplexně uchopit rozšířený jev zvaný spam, systematicky popsat jeho typy, způsoby, jak se proti němu bránit a zejména pak popsat konkrétní zvolený antispamový software a poskytnout jeho ukázkovou instalaci.

K dosažení tohoto cíle bude nutné postupovat od základů, to zejména proto, aby každý čtenář, i laik v tomto oboru, porozuměl dané problematice.

Nejprve budou kladeny otázky, kde vlastně spam vzniká a jakým způsobem se k nám dostává. Aby byla tato práce pochopitelná i pro čtenáře nepříliš sběhlé v tématice elektronické pošty, bude nutné nejprve vylíčit základní charakteristiky a možnosti e-mailu.

Poté bude tato práce postupovat systematicky přes vysvětlení pojmu spam, jeho původu a etymologii slova, až k jeho historii, aby byla uchopena co nejrozsáhleji daná tématika a bylo tak čtenáři umožněno porozumění v této věci.

Teprve když bude důkladným rozborem zajištěno, že i laik je dostatečně poučen o tom, jakými způsoby nám spam hrozí a jakou cestou se k nám dostává, přejdeme k technikám, které nám různými způsoby umožňují spam co nejvíce eliminovat. Tyto techniky, vzhledem k jejich množství a rozmanitosti jejich působení, pak bude nutno co nejobjektivněji vysvětlit a umožnit tak čtenáři vlastní názor na to, jak asi ta která technika může v té které situaci pomoci.

Poté přijde řada na volbu konkrétního software, a poskytnutí důvodů pro jeho výběr. Následně bude důkladně popsáno, jak tento antispamový software funguje, jak ho lze modifikovat a v neposlední řadě, jak ho nainstalovat na klientský počítač s běžně používaným operačním systémem Windows.

3 SEZNÁMENÍ S FUNKČNÍ STRÁNKOU ELEKTRONICKÉ POŠTY

3.1. Co je Elektronická pošta?

Elektronická pošta, anglicky Electronic mail, ve zkratce tedy e-mail, je vzájemná komunikace uživatelů realizovaná pomocí počítačů, souborů, internetu a komunikačních linek. Původně byla e-pošta určena pouze pro textové soubory. V dnešní době umožňuje zasílání dalších, téměř jakýchkoliv, formátů, které jsou přikládány jako příloha. E-poštou se rozumí jak komunikace v rámci internetu tak v rámci intranetu (v rámci firmy). K práci s elektronickou poštou se obecně používají dva typy programů: klient elektronické pošty a internetový prohlížeč pro tzv. webový e-mailový systém.

Klient elektronické pošty, který je znám pod zkratkou MUA (z anglického výrazu Mail User Agent), je program používaný pro přijímání a odesílání e-mailů.

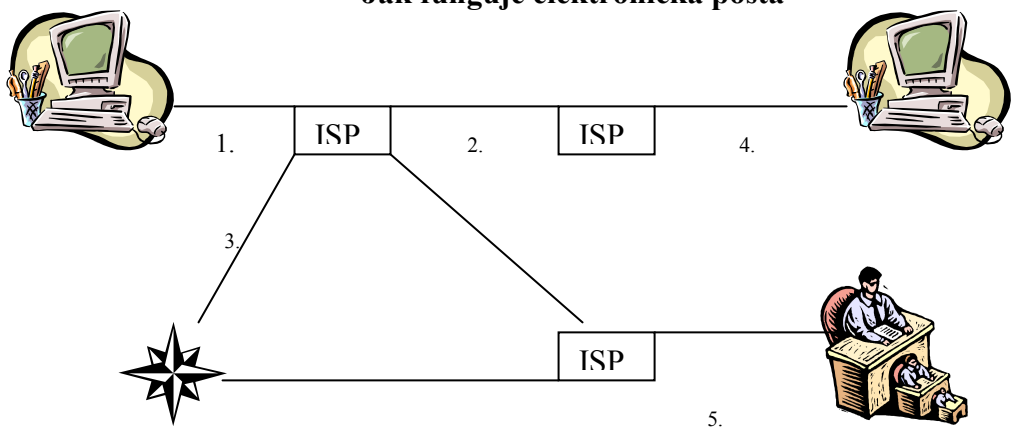
Internetový prohlížeč je softwarová aplikace, která umožňuje uživateli zobrazovat text, obrázky a ostatní informace, které jsou na webové stránce umístěny. Uživatel může krom jiného pomocí prohlížeče užívat poštovních služeb, které mu umožňuje poskytovatel jeho poštovní schránky, jakým je např. Seznam, Centrum, nebo Yahoo.

3.2. Jak funguje elektronická pošta?

Dejme tomu, že odesilatel posílá e-mail dvěma příjemcům. Následující diagram znázorňuje realizaci této komunikace.

Obrázek č. 1

Jak funguje elektronická pošta



Postup:

1. Odesilatel z Prahy posílá e-mail příjemci z Prahy na jeho firemní mail a současně kopii tohoto e-mailu posílá příjemci do Austrálie.
2. Firma, která poskytuje odesilateli připojení (ISP = Internet Service Provider) používá router (směrovač, který přeposílá datagramy k jejich cíli), aby našla spojení k oběma příjemcům zprávy
3. Do Austrálie jde zpráva přes satelit.
4. U příjemce v Austrálii se uchovává e-mail na serveru poskytovatele, dokud si ho příjemce nestáhne do počítače.
5. K příjemci v Praze jde pošta po drátech k serveru firmy, u které je připojen. Příjemce se připojí k internetu a stáhne si poštu.

3.3. Architektura elektronické pošty

Doručování elektronické pošty po Internetu se účastní tři druhy programů:

3.3.1 MUA (Mail User Agent) – poštovní klient:

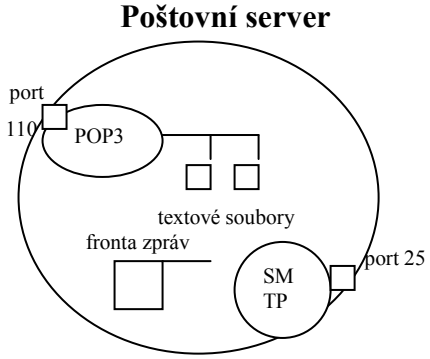
Zajišťuje odesílání zpráv a vybírání schránek, příkladem je Microsoft Outlook, Mozilla Thunderbird, Opera, Novell GroupWise a další. Je to v podstatě specializovaný editor, který umí kromě vytvoření zprávy také manipulovat se schránkami, odeslat zprávu nejbližšímu serveru, který se stará o doručování zprávy k cíli (tzv. MTA) a převzít zprávu ze serveru prostřednictvím POP3, nebo IMAP. Další nedílnou součástí MUA bývá také adresář, který udržuje přehled o adresách. Funkcím poštovního klienta se budu podrobně věnovat dále.

3.3.2. MTA (Mail Transport Agent) – poštovní server:

Běží obvykle jako démon (programová funkce, která běží „na pozadí“ a čeká na žádost o komunikaci) a naslouchá na portu TCP/25. K tomuto portu se může připojit buď poštovní klient nebo jiný server, který předá zprávu určenou k doručení. MTA zkontroluje, zda je zpráva určena

pro systém, na kterém běží. Pokud ano, předá ji programu MDA (lokální doručení). Pokud je zpráva určena jinému počítači, musí MTA navázat spojení se vzdáleným serverem, což učiní pomocí spolupráce se systémem DNS (systém doménových jmen). Od něho si vyžádá tzv. MX záznam pro cílovou doménu obsahující IP adresu počítače, který se stará o doručení pošty na této doméně. Poštovní server pro obsluhu pošty používá dva porty, port 110 pro příchozí poštu (protokol POP3), a port 25 pro odchozí poštu (protokol SMTP).

Obrázek č. 2



Protokoly Elektronické pošty

- SMTP: Internetový protokol určený pro přenos zpráv elektronické pošty mezi stanicemi. Tento protokol zajišťuje doručení pošty pomocí přímého spojení mezi odesilatelem a adresátem. Zpráva je doručena do tzv. poštovní schránky adresáta, ke které potom může uživatel kdykoliv pomocí protokolů POP3 nebo IMAP. Následuje popis, jak probíhá zaslání a přijetí e-mailu pomocí e-mailového klienta.

  1. Ve svém MUA napíšu e-mail a dám příkaz k odeslání. Můj poštovní klient byl předem nakonfigurován, aby věděl, ke kterému serveru se má připojit, a proto se k němu jednoduše bez dotazů připojí a komunikuje s ním následovně: sdělí mu adresu příjemce, všechny další informace z hlavičky e-mailu a pošle mu tělo zprávy.
  2. SMPT server vezme adresu příjemce a rozdělí ji na dvě části, před lokátorem (znak „@“) a za ním. Přes DNS službu (viz výše) si zjistí IP adresu SMTP serveru příjemce a spojí se s tímto serverem pomocí portu 25 a předá mu poštovní zprávu, kterou si nalezený server příjemce převezme a vloží ji do schránky.

- POP3, IMAP – internetové protokoly, sloužící pro stahování e-mailových zpráv ze vzdáleného serveru na klienta jsou rozdílné. Zatímco POP3 je určen pro časově omezené (např. vytáčené) připojení k Internetu a své zprávy si z poštovního serveru stáhne, a poté se odpojí, IMAP je spíše optimalizován pro práci v dlouhodobě připojeném režimu, kdy zprávy zůstávají uloženy na serveru a průběžně se stahují, když je potřeba, a tím poskytuje možnost pracovat se zprávami na serveru bez jejich stahování. Další rozdíly zahrnují podporu pro práci více připojených klientů zároveň, uchovávání stavů zpráv na serveru, podporu více složek a prohledávání zpráv na straně serveru.

3.3.3. MDA (Mail Delivery Agent) – program pro lokální doručování:

Poštovní server by mohl zprávy do schránek ukládat přímo, ale výhodnější je použití programu, který umožňuje při doručování ještě dále zprávy zpracovávat nebo filtrovat.

Tento program může například třídit zprávy do různých schránek dle obsahu, odesilatele, subjektu, apod., či odstraňovat nebo alespoň od kvalitních zpráv separovat nežádoucí zprávy jako je např. spam, jemuž se budu dále podrobněji věnovat.

4. DEFINICE, PŮVOD A HISTORIE SPAMU

4.1. Definice Spamu

Vzhledem k tomu, že v současné době žádná všeobecně uznávaná definice neexistuje, dovolím si spamming definovat jako hromadné rozesílání nevyžádaných zpráv. Tyto zprávy se nazývají spam a nemusí se vždy jednat pouze o rozesílání nevyžádaných zpráv emailem. Spam lze šířit i pomocí diskusních fór a chatových místností. Spam může nabývat jak charakteru komerčního, kdy se nám tvůrce nevyžádané pošty zašle svou reklamu, tak i charakteru nekomerčního, kdy se jedná o téma zprávy pro nás nezajímavé, nebo naopak o nebezpečné pobízení k zaslání nějakého obnosu kamsi, či vymazání obsahu určité složky v našem PC ze smyšleného nás ohrožujícího důvodu. Bohužel se v dnešní době můžeme se spamem setkat nejen na počítači, ale i v našem mobilním telefonu, ačkoliv se toto děje pouze zřídkakdy a charakter takového spamu bývá pouze komerční.

4.2. Původ slova Spam (Etymologie)

Původ slova Spam má zejména pro příznivce britské komediální skupiny Monthy Python poněkud komické zabarvení. SPAM je ve zkratce Spiced Ham and Pork, což je docela obyčejný americký „lančmít“v plechovce. V roce 1970 však britská skupina Monthy Python toto slovo použila v jednom ze svých skečů o tom, jak si manželé Brunovi chtějí objednat v restauraci jídlo a číšník jim stále dokola nabízí pokrmy ze Spamu. To samozřejmě náš pár dosti znepokojuje a obtěžuje.

Z celé scénky vyplývá, že spam je cosi, co je nám nabízeno stále dokola, ale co pro nás není vůbec zajímavé, ba naopak nás to otravuje a nechceme s tím mít nic společného.

Společnost Hormel Foods Corporation, která vyrábí maso v konzervě a od které slovo SPAM pochází, nemá nic proti tomu, aby toto slovo bylo převzato a zobecněno pro účely označení nevyžádané pošty, ale dost se obává dne, kdy se lidé začnou ptát: „Proč by Hormel Foods pojmenovali konzervu s masem po nevyžádané poště?“.

4.3. Historie Spamu

4.3.1 Obecná historie

Vůbec první záznam o spamu jako takovém - ačkoliv se tomu tak tehdy ještě neříkalo - je z roku 1904, kdy se spam objevil ve formě telegramu, tedy nevyžádaného telegramu.

V roce 1980 se začalo již používat slovo SPAM a to na elektronických nástěnkách (zvaných bíbiesky), které umožňovaly uživatelům přispívat svými názory na jednotlivé nástěnky rozdělené podle témat. Často se stalo, že na nástěnku

začal psát své texty návštěvník, který nebyl pro účastníky již probíhajícího rozhovoru příliš vítaný.

Z tohoto důvodu tito účastníci zaplnili obrazovku slovy SPAM, aby z ní jeho příspěvky odsadili, a tím ho přiměli, aby dané téma opustil. Také se toto slovo s oblibou používalo k přerušení konverzace na nástěnce nepřátelského tématu, například mezi účastníky nástěnky Star Wars a Star Trek.

Další uplatnění našel spam na Usenetu, což byl soubor elektronických diskusních skupin. Zde se objevoval spam ve formě několikrát za sebou zveřejněné jedné a týž zprávy nebo komentáře a dále pak i nesmyslných, nic neříkajících, nebo často i škodlivých textů, které zahlcovaly diskusní skupiny. Jeden příklad škodlivého příspěvku se skrýval pod titulem „Vydělej peníze rychle!“ a lákal účastníky k zaslání malého příspěvku s příslibem značných výnosů do budoucna.

Prvního komerčního využití došel spam v roce 1994, kdy dva právníci a zároveň manželé, Laurence Canter a Martha Siegel, využili diskusních prostor na Usenetu k šíření reklamy na vlastní právně-imigrační služby. Tento spam vešel v povědomí veřejnosti pod názvem „Spam pro Zelenou kartu“ (Green Card Spam). Tito dva pak dále široce využívali Usenet a email k šíření svých reklam a jsou považování za průkopníky masového spamu, který dnes zahlcuje internet. Během několika let od tohoto incidentu se spamování přesunulo převážně do sfér emailů, kde setrvává dodnes.

4.3.2. Historie v Čechách

Jeden opravdu nevyžádaný e-mail, který zaskočil mnohé uživatele elektronické pošty, byl nasazen společností **Media Online.** Pomocí masivní reklamní kampaně šířené elektronickou poštou propagoval značku **tvujdum.cz**. Dosahoval četnosti až desítek tisíc reklamních dopisů.

Společnost se chtěla pojistit následující větou, uvedenou v e-mailu:

„Tento e-mail je Vám zasílán na základě pečlivého výběru a globální rešerše uživatelů, kteří své webové stránky věnují tématice bydlení, stavebnictví. Předem se omlouváme za nevyžádaný e-mail“ (1).

Česká internetová veřejnost se přesto rozhodla proti spamu bojovat a, jak uvádí server Lupa.cz, 30 lidí na společnost Media Online podalo stížnost. Magistrát hlavního města Prahy na toto zareagoval udělením pokuty spamující společnosti ve výši několika desítek tisíc korun.

Jeden z nejznámějších českých spamerů se jmenuje Ross Hedvíček a specializuje se
na šíření nekomerčních nevyžádaných e-mailů, ve kterých se snaží prosadit americkou republikánskou vládní stranu. Dalším nekomerčním spamem byl hromadný elektronický dopis zaslaný uživatelům emailového serveru mail.seznam.cz, který s podpisem Vladimíra Špidly propagoval hlasování pro vstup České Republiky do Evropské unie.

Další známý spamer, který neváhal ke svým nevyžádaným emailům přikládat dovětek o tom, že jeho hromadné e-maily rozhodně nejsou SPAMY, se jmenuje Jan Šinágl. Pan Šinágl posílal do mnoha schránek články, fotografie a scany se svým komentářem.

4.4. Rozdělení spamu

Spam má mnoho podob. Záleží na tom, komu je spam určen a za jakým účelem je rozesílán. Spam také reaguje na současné události. V posledních měsících se tedy spameři zabývali převážně novým americkým prezidentem, nominacemi na Oskary a letní dovolenou zdarma.

4.4.1. Rozdělení spamu z hlediska obsahu

Komerční spam má za úkol propagovat určitý produkt pomocí internetu a elektronické pošty. Je to nevyžádaná pošta s komerčním obsahem. Je stejně nežádoucí, jako spam nekomerčního charakteru, ale nemůže vyloženě ublížit. Spíše majitele cílových e-mailových schránek otravuje.

Nekomerční spam je naopak nevyžádaná pošta nekomerčního charakteru, která může mít mnoho podob.

Jednou z nich je například komentářový spam, neboli diskusní spam, kde jsou na volně přístupná diskusní fóra vkládány příspěvky, které s diskusí nikterak nesouvisí a jsou v dané diskusi nevyžádané.

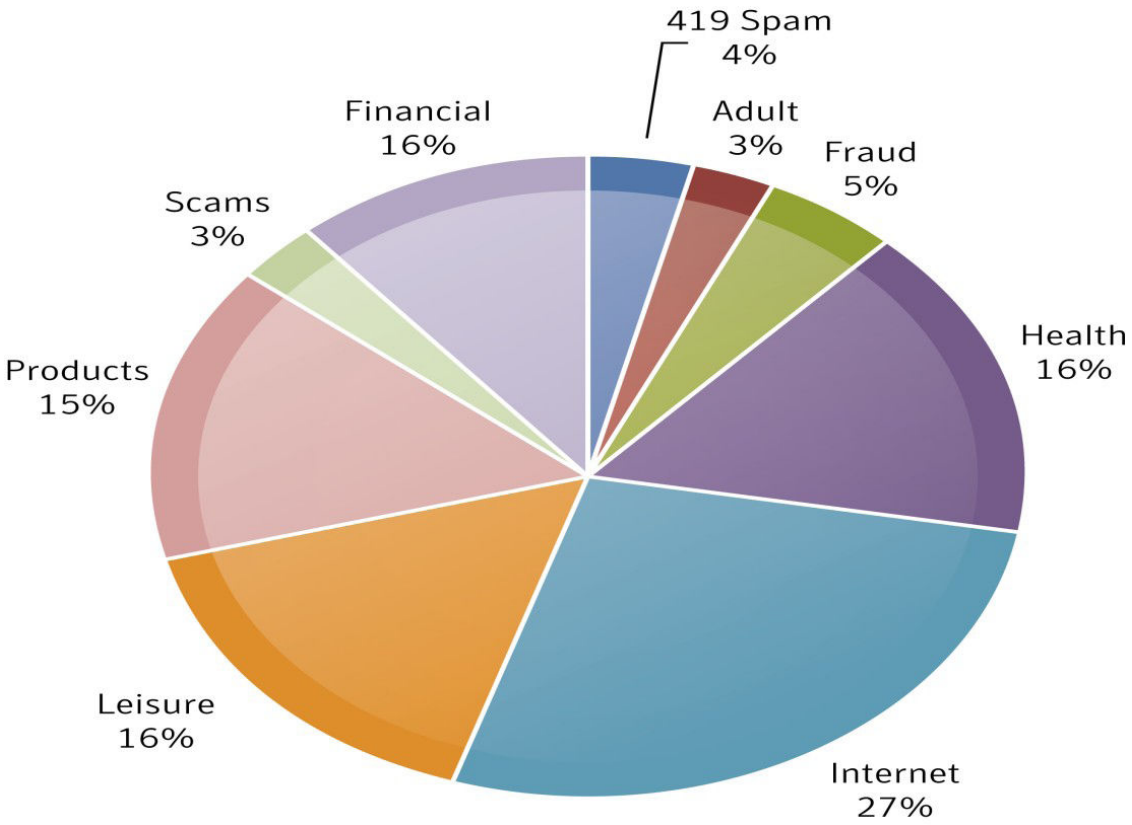
Dále nám pak většinou pouze prostřednictvím e-mailu hrozí spam požadující zálohu, což je v podstatě podvod založený na důvěře, kde je příjemce přesvědčován k zaslání relativně nízkého obnosu v naději na větší výnos. Neznámý autor potom zasílá spam, který díky své, často až nadprůměrné, důvěryhodnosti může být dost nebezpečný. Jako příklad tohoto spamu bych uvedla například „Nigérijský dopis", kdy odesilatel nabízí v dopise příjemci podíl na velkém obnosu, který se snaží převést ilegálně z Nigérie. Není ale schopen tento čin spáchat sám, a proto potřebuje prostředníka, který mu má zaslat peníze a později obdržet procento z daného velkého obnosu. Další příklad je spam nazývaný „Španělský vězeň", kde odesilatel e-mailu sdělí oběti, že je v kontaktu s velice zámožným Španělem, který je v současné době ve vazbě a přeje si zůstat v této záležitosti anonymní. Odesilatel tohoto spamu dále přesvědčí oběť, aby mu pomohla, samozřejmě finančně, dostat vězně z vazby, s příslibem velké odměny. A jako třetí příklad uvedu řetězový spam, se kterým jsem se setkala osobně, ale naštěstí ne přímo, nýbrž pouze z doslechu od známých, kteří dané záležitosti propadli. Jedná se o spam „Make Money Fast", česky: rychle vydělané peníze, který přesvědčoval příjemce, aby přeposlali 1 dolar v hotovosti na seznam adres uvedený v textu e-mailu, přidali poté své jméno a adresu
na konec seznamu, odmazali první jméno a adresu ze seznamu a celý email přeposlali dál. Pomocí teorie Pyramidy měl výsledný řetězec peněz tekoucích tam a zpět vydělat tisíce dolarů všem zúčastněným osobám.

Do třetice pouze zmíním spam zpravodajského charakteru, kde se nás spamer snaží přesvědčit o tom, že jeho názory a informace jsou pro nás nepostradatelné a opakovaně nám posílá novinky z ze svého oboru, zatímco nás tyto informace většinou vůbec nezajímají a nechápeme, proč by nás zajímat měly.

Následující obrázek ukazuje rozložení spamu z hlediska obsahu:

Obrázek č. 3

Global spam categories



Jak je vidět výše, nejvíce se spam týká různých internetových a počítačových produktů, dále jde pak o volný čas, nebo zdraví (medikamenty).

4.4.2. Rozdělení spamu z hlediska výskytu

Spam dnes zahlcuje prakticky všechny možné cíle. Může nám zahlcovat **počítač** a v něm e-mailové schránky, různá diskusní fóra, nebo instant messaging (např. ICQ, Skype), nebo **mobilní telefony**. U mobilních telefonů se jedná výhradně o spam šířený pomocí zpráv SMS a nazývá se m-spam. U nás se tento druh spamu objevuje pouze v podobě různých reklam od operátorů, kteří nám formou SMS

nabízí různé nové služby nebo tarify. Většinou je však možné tento spam zamítnout určitou odpovědí nebo voláním nějakého čísla.

4.4.3 Rozdělení spamu z hlediska přístupu k cílové skupině

Spam lze rozdělit z hlediska přístupu odesilatele k cílové skupině, které spam zasílá, a to konkrétně přístupu k možnosti příjemce zamítnout budoucí zasílání spamu.

Toto zamítnutí může spamer nabídnout příjemci dvěma způsoby.

První způsob, **opt-in** nastane, pokud nám někdo, většinou prostřednictvím svých webových stránek, nabídne možnost dostávat hromadný e-mail. Většinou se jedná o informační bulletin, či reklamu a my můžeme tuto nabídku přijmout tím, že se někde zaregistrujeme a uvedeme svou e-mailovou adresu.

Druhou metodou je **opt-out**, kdy nám odesilatel spamu na konci nevyžádaného e-mailu nabídne (většinou nejmenším možným fontem), abychom se v případě nezájmu o daný spam odhlásili na nějaké adrese.

4.4.4 Zvláštní druhy spamu

Spamdexing lze česky vyjádřit, jako manipulace relevance vyhledávaných WWW stránek, obvykle způsobem nesourodým se záměrem indexace vyhledávače. Vyhledávací prostředky (např. www.google.cz a mnoho jiných) zpravidla používají celou řadu algoritmů pro stanovení relevantnosti vyhledávaných odkazů a podle ní je poté řadí, neboli indexují. Spamdexing je opakovaná volba stejného méně relevantního odkazu, která zvyšuje prioritu tohoto odkazu.

Nevyžádaný reklamní tisk, který je umisťován do našich poštovních schránek lze taky nazvat spamem, v USA tomu říkají pouliční spam. Na většině místech světa je toto umisťování letáků na soukromý majetek bez souhlasu majitele nebo na veřejný majetek, bez písemného souhlasu příslušného státního orgánu, nezákonné. V České Republice existuje také zákaz nevyžádané reklamy podle zákona o regulaci reklamy, který se vztahuje právě na reklamní letáky vhazované do poštovních schránek. Jak uvádí IDNES.cz: „Pokud tedy nechceme, aby se nám ve schránce objevovaly nevyžádané reklamní letáky, mělo by například stačit vyjádřit svůj

nesouhlas na poštovní schránce. Tento způsob odmítnutí nevyžádané reklamy je již dnes často aplikován a v zahraničí je zavedenou praxí.“ (3).

4.5. Právní pohled

Je samozřejmostí, že pokud nám každý den zahlcují e-mailovou schránku „tuny“ spamu, máme vztek a často se ptáme sami sebe: „Tohle se smí?“ nebo „Tohle si můžou dovolit?“.

Odpověď na tyto otázky mi poskytl server EPRAVO:

„ Poslanecká sněmovna v úterý 29. června schválila ve třetím čtení návrh zákona o některých službách informační společnosti. Tento zákon by měl v souladu s evropskou legislativou nově upravit otázky týkající se zasílání nevyžádaných sdělení (tzv. spamming). Tyto otázky byly doposud řešeny především v zákoně o regulaci reklamy, podle něhož bylo zakázáno šířit nevyžádanou reklamu, pokud tato vede k výdajům adresáta nebo jej obtěžuje.

Návrh zákona se vztahuje na zasílání sdělení pouze obchodního charakteru. Těmito sděleními se rozumí všechny formy sdělení, která jsou určena k přímé či nepřímé podpoře zboží či služeb nebo image podniku a rovněž i reklama. Ačkoliv by jistě bylo užitečné rozšířit působnost zákona i na sdělení neobchodního charakteru (např. politická či náboženská sdělení), návrh zákona tak nečiní.

Obchodní sdělení, která jsou šířena e-mailem, hlasovou poštou, faxem nebo textovými zprávami (SMS), bude možno zasílat, pouze pokud s tím budou jejich adresáti předem souhlasit (tzv. metoda „Opt-in“). V každém takovém obchodním sdělení musí být adresátovi dána jasná a zřetelná možnost jednoduchým způsobem a zdarma nebo na účet zasílatele vzít zpět souhlas s využitím jeho elektronického kontaktu. Zasílatel bude povinen souhlas adresáta se zasíláním obchodních sdělení těmito prostředky prokázat.

Šířit obchodní sdělení jinými elektronickými prostředky než jsou výše uvedené (např. přímý telemarketing) bude možné také jen s předchozím a prokazatelným souhlasem jejich adresáta.

Nad dodržováním pravidel pro šíření obchodních sdělení elektronickými prostředky by měl převzít dozor Úřad pro ochranu osobních údajů. Navržená výše pokuty za porušení pravidel pro šíření obchodních sdělení elektronickými prostředky činí až 10 milionů Kč, což je nepoměrně více než činí stávající výše pokuty za šíření nevyžádané reklamy podle zákona o regulaci reklamy. Ta dosahuje až 2 milionů Kč.“ (4)

Podle serveru ITprávo - serveru o internetovém a počítačovém právu byl výše zmíněný návrh zákona o službách informační společnosti schválen a nabyl účinnosti dne 7. září 2004 jako zákon č. 480/2004 Sb. Tento zákon pojednává kromě nevyžádané inzerce zejména o odpovědnosti poskytovatelů služeb (providerů) za obsah na jejich stránkách. Dříve totiž platilo, že tento poskytovatel nese odpovědnost stejnou, jako primární škůdce (ten, kdo závadný obsah na stránky umístil), protože obsah nezkontroloval. Tato odpovědnost se nyní omezila převážně na primárního škůdce, tedy umisťovatele obsahu. Tolik jen pro přesnost, vzhledem k tomu, že se daný zákon často nepříliš přesně a omezeně označuje ryze za antispamový.

Zákon č. 480/2004 Sb. však spamming povoluje v některých výjimkách a to zejména v případě regulovaných činností. „…osoby vykonávající regulované činnosti (lékaři, advokáti, daňový poradci, apod.) mohou za použití elektronických prostředků v rámci činností, které jsou obsahem regulované činnosti, šířit obchodní sdělení, a to v souladu s § 7 zákona a
v souladu s příslušnými pravidly vydávanými obchodními, profesními a spotřebitelskými sdruženími, která upravují zejména nezávislost, důstojnost, čest povolání a poctivý přístup k zákazníkům.Obchodní sdělení osob vykonávajících regulované činnosti musí obsahovat název profesní samosprávné komory zřízené zákonem, u níž je osoba vykonávající regulovanou činnost zapsána, odkaz na profesní pravidla uplatňovaná v členském státu Evropské unie, v němž je osoba vykonávající regulovanou činnost usazena, a způsob trvalého veřejného přístupu k informacím o příslušné profesní samosprávné komoře zřízené zákonem, jejímž je osoba vykonávající regulovanou činnost členem.“ (5).

Další skulina v zákoně, která povoluje určitý druh spamu je formulována takto:

„Za obchodní sdělení se nepovažují údaje umožňující přímý přístup k informacím o činnosti fyzické či právnické osoby nebo podniku, zejména doménové jméno nebo adresa elektronické pošty.“ (6).

Tento odstavec potvrzuje, že spamovat webovými adresami se smí, což umožňuje potenciálním spammerům zaplňovat naše e-mailové schránky zcela beztrestně odkazy na svoje servery. Přijde-li nám tedy například jinak prázdný e-mail pouze s odkazem: „www.internetovyobchod.cz/mikrovlnky_za_babku.html“, se stížností na UOOU neuspějeme, ačkoliv se dá říci, že takovýto odkaz už je svým způsobem reklamou. A vzhledem k tomu, že doména může mít až 256 znaků, spammer může pustit svou fantazii a inzertní duch na „špacír“ a vesele rozesílat svou doménu, čímž si zajistí poměrně solidní reklamu. Jako příklad zamítnuté stížnosti Úřadem pro ochranu osobních údajů, uvedu převzatý případ ze serveru http://papucha.cz, kde si autor blogu stěžuje na záporné vyřízení stížnosti, kterou podal
na níže uvedenou firmu, která mu zaslala e-mail znějící následovně:
„From: LevneTonery.cz

Dobrý den, rádi bychom Vás informovali o nových internetových stránkách

www.levnetonery.cz www.levneinkousty.cz

Na uvedených stránkách naleznete všechny další informace.“

Na žádost mu však UOOU odpověděl takto:

„Stížnost na nevyžádané obchodní sdělení – vyrozumění. K Vaší stížnosti na rozesílání nevyžádaného obchodního sdělení „www.levnetonery.cz“ sdělujeme, že na základě Vašeho podnětu bylo provedeno šetření za účelem zjištění zda je dodržován zákon č. 480/2004 Sb., o některých službách informační společnosti a o změně některých zákonů (dále jen „Zákon“) v souvislosti s rozesíláním elektronických zpráv s obchodní nabídkou. Kontrolou provedenou u podnikatelského subjektu Marek Krátký, Poljanova 3241/1, 143 00 Praha 4, IČ:70399565 a s ní související analýzou předmětné zprávy nebylo zjištěno porušení Zákona, neboť odeslané zprávy elektronické pošty nejsou obchodním sdělením ve smyslu tohoto Zákona. Vámi obdržená zpráva naplňovala výjimku dle § 2 písm. f) Zákona („Za obchodní sdělení se nepovažují údaje umožňující přímý přístup k informacím o činnosti fyzické či

právnické osoby nebo podniku, zejména doménové jméno nebo adresa elektronické pošty“).“ (7).

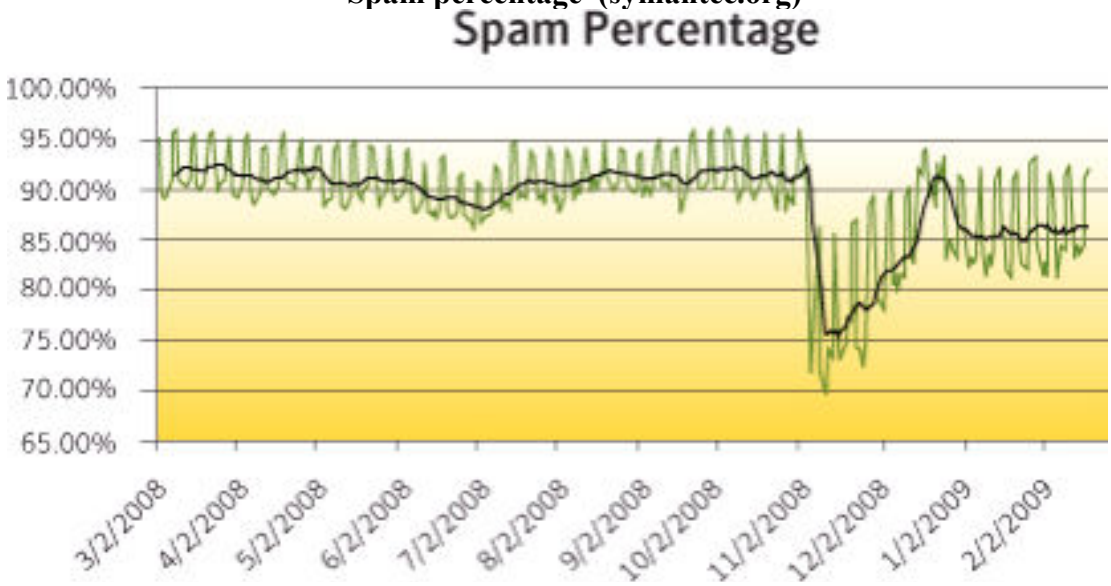
Dále je v této záležitosti poněkud znepokojující, že Úřad pro ochranu osobních údajů, který je zodpovědný za řešení podaných stížností na Nevyžádaný obsah sdělení a k tomuto účelu nabízí elektronický formulář, je ve své činnosti poměrně zaneprázdněný. Podle některých stížností na serveru BLOGUJE.cz odpovídá se značným zpožděním a pouze hromadným e-mailem, ve kterém oznamuje, že úřad nestíhá podněty vyřešit a že od podání podnětu do uzavření uplyne až jeden rok a vyzývá podávajícího, aby potvrdil, že na svém podnětu stále trvá.

Tato a jí podobné stížnosti mě utvrzují v názoru, že ačkoliv zde máme funkční úřad, který by měl situace nevyžádané elektronické pošty řešit a zabývat se dopadením těch, kteří se
na plnění našich e-mailových stránek podílí, plyne z toho všeho určitá beznaděj. Tu ale nezbývá než dávat za vinu opět spammerům a ne úřadu, protože bereme-li v úvahu to kvantum spamů, které se denně doručí, není divu, že úřad vyřizování každé stížnosti nestíhá.

5. FILTRAČNÍ METODIKA

Podle měsíčního reportu vydávaného společností Symantec je v dnešní době 86% emailů nevyžádaného charakteru (viz obrázek č. 3)

Obrázek č. 4
Spam percentage (symantec.org)



Každým dnem je po celém světě rozesíláno až na 10 miliard spamů, většina z nich však díky chytrým filtrovacím zařízením, jejichž druhům a funkčnostem se budu věnovat níže, nedosáhne svého cíle. Tyto filtry však nejsou stoprocentní a vždy existuje nějaká šance, že buďto vyfiltrují zcela korektní e-mail, nebo naopak propustí zcela evidentní spam. Často totiž nebývá jednoduché rozpoznat spam od „nespamu“.

Nejobecnější rozdělení ochrany naší elektronické schránky od spamů je dáno mírou našeho přičinění v tomto boji. Antispam je tedy buď naše aktivita plynoucí ze znalostí a zkušeností v tomto oboru, nebo pak automatizovaná protispamová technika, která bývá samostatným softwarem a chrání naší schránku po svém.

5.1. Jak se jim podařilo sehnat můj Email?

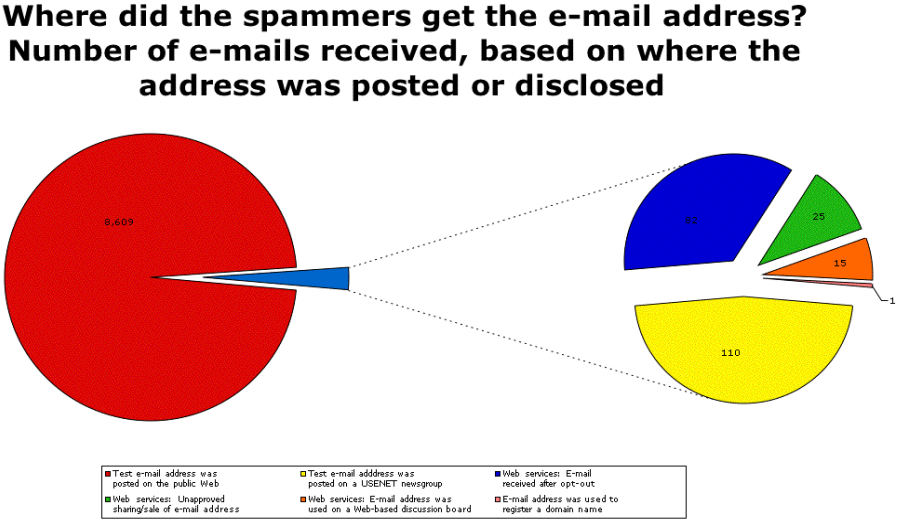
Vzhledem k tomu, že typický spammer cokoliv nerad kupuje, je tedy většinou na něm a jeho fantazii, jak a kde si emailové adresy obstarat. V první řadě se jedná o tzv. Sklizeň emailových adres, která je prováděna robotem vyhledávajícím emailové adresy, tedy zpravidla odkazy typu „mailto:jmeno@domena.xx“ a to kdekoliv na internetu, nejraději na různých diskusích, chatech. Zkrátka kdekoliv, kde je možné a obvyklé, že narazí na emailovou adresu. Dalším způsobem obstarávání adres je tzv. Slovníkový

útok, kdy spammer vkládá před zavináč všemožná jména a obvyklá slova ze slovníku v naději, že se spam doručí. Pokud má tedy někdo například emailovou adresu ve tvaru „petr@seznam.cz“, pravděpodobně bude dostávat do své schránky doslova haldy spamů. Je tedy více než žádoucí si svou emailovou adresu strukturovat nějakou méně pravděpodobnou posloupností, která nebude lákat spammery útočící slovníkovou sklizní. Méně pravděpodobné adresy však mohou přece jen být odhaleny, a to zejména Útokem hrubou silou, kdy spammer nepřemýšlí nad nějakými možnými jmény, názvy, či kombinacemi obojího, ale zkouší jednoduše všechny možné náhodné kombinace písmen a testuje, zdali se na takové adresy spam doručí. Proti tomuto útoku je možno bojovat výběrem emailové adresy, která je delší.

V březnu 2003 provedla americká společnost bojující za demokratické hodnoty a konstituční práva v digitalizované době CDT (Center for Democracy and Technology) jednu zajímavou studii. Šla za prostým cílem, pokusit se zjistit zdroj spamu. K těmto účelům zřídili stovky různých emailových adres a poté vyčkávali 6 měsíců, aby zjistili jaký druh emailů jim na tyto adresy chodí. Výsledky se velice lišily podle toho, kde byly tyto adresy použity. Nejvíce spamu lákaly veřejně publikované adresy na webových stránkách a v diskusních místnostech.

Obrázek č. 5

Počet přijatých e-mailů dle toho, kde byly adresy zveřejněny či odhaleny



LEGENDA

8609	Emailová adresa byla zveřejněna na webové stránce
82	Emaily přijaté po opt-out (odhlášení z mailinglistu)
25	Nepovolené sdílení/prodej emailových adres
15	Emailová adresa byla použita na diskuzním fóru
1	Emailová adresa byla použita k registraci domény
110	Emailová adresa bylo vyvěšena do diskusní skupiny na USENET

Celkem CDT během šesti měsíců na svých 250 adresách obdržela 10 000 emailů. Z toho asi 1600 bylo korektních emailů, 62 bylo neklasifikovatelných, a 16 jich bylo doručeno po odhlášení ze se seznamů pro zasílání informačních bulletinů. Zbývajících 8842 bylo označeno za spam.

5.2. Jak se bránit proti spammerům?

Naše přičinění v boji proti spamu může být dosti účinné. Není totiž nad to znát, jak který spammer k adresám přišel a podle toho postupovat při volbě emailové adresy či jejího zveřejňování kdekoliv na webových stránkách.

Způsobů, jak proti spamu bojovat vlastní silou bez pomoci automatizovaných antispamových softwarů je více a jsou aplikovatelné pouze na určité oblasti používání. Nejprve musíme opatrně vybírat, jak bude naše adresa vypadat a to hlavně na základě výše uvedených znalostí způsobů, jakými se naše adresa může dostat spammerovi do rukou. Dále jde o to abychom spamerovi znemožnili její odhalení. Toto lze provést následujícími metodami:

Maskování Emailů publikovaných veřejných elektonických prostranstvích:

Jedním z často užívaných způsobů je zamaskování adresy vložením nějakého evidentního výrazu, který později při vyhledání spamerem zabrání funkčnosti naší adresy. Tento výraz může být např. „NOSPAM“ a v praxi to pak vypadá asi takto: „majdaNO@SPAMseznam.cz“, dále lze jednoduše místo znaku pro zavináč (@) použít slovo „zavináč“. Tento způsob lze aplikovat například na webových diskusích či komentářích, zkrátka v místech, kde musíte svůj email zveřejňovat. Skrytí je na první

pohled vidět ale zároveň nepodléhá hromadnému vyhledávání emailových adres za účelem jejich pozdějšího zneužití.

Použití několika emailových adres:

Pro použití neznámé webové stránky nebo vyvěšení komentáře do diskuse je radno zřídit si novou emailovou adresu. Pro tento účel můžeme použít například „jednoúčelové emailové adresy“, které se sjednocují v jednom místě, ale umožňují okamžitě vypnout kteroukoliv z adres, která přitahuje spam.

Existují ještě nástroje, které jsou běžně zřízeny tvůrci některých internetových stránek a které zabraňují zneužití naší adresy spamerem. Takové nástroje se používají u internetových stránek existujících pro účely, pro něž se bez výměny emailových adres neobejdeme a byly by tedy snadným terčem spamerů. Jako nejúčinnější způsob se mi zdá být používání emailových formulářů, u kterých běžně nevíme, na jakou emailovou adresu píšeme, a námi vyplněná adresa se dostane zpravidla do rukou pouze tomu, komu je určena. Tyto formuláře se běžně používají například na serverech soustřeďujících nabídku a poptávku zaměstnání, tedy např. JOBS.cz. Dalším účinným nástrojem je například maskování emailových adres převodem do ASCII kódu.

5.3. Antispamové techniky

Existuje řada služeb a softwarů, které mohou freemailové servery i uživatelé emailů použít, aby snížili příval spamu do svých elektronických schránek. Některé z nich spoléhají na zamítnutí emailů, které jsou doručovány z obvyklých spamujících internetových stránek. Ostatní se řídí automatickou analýzou obsahu příchozí pošty, kterou pak buď vyřadí, nebo propustí. Tyto dva přístupy jsou známy jako blokování a filtrování.

Obě dvě metody snižují procento spamu doručené do našich schránek. Ale zatímco blokování snižuje cenu vynaloženou na spam, protože blokuje emaily dříve, než jsou odeslány, filtrování je zase důkladnější, protože zkoumá obsah zprávy.

V poslední době je z důvodu jejich výhodnosti zaznamenán narůstající trend v integraci despamových technik na poštovních serverech (MTA), které slouží pro přenos elektronické pošty z jednoho počítače na druhý a pracují na pozadí, zatímco

uživatel komunikuje s MUA (klientem elektronické pošty). Tato integrace znamená, že MTA, který obdrží zprávu, ji na pozadí nejprve zpracovává, postupujíce jí různými despamovými opatřeními, než se rozhodne, jestli ji přijme, nebo zamítne. Zprávy, které nejsou označeny za spam jsou doručeny původnímu příjemci, zatímco spam se zamítá během SMTP dialogu. Tento způsob zavržení spamu skýtá řadu výhod, které plynou z jeho porovnání s konvenční despamovou filtrací. V první řadě spamy nejsou nikdy odráženy zpět na odesilatele, což je výhodné, protože většina spamu přichází s podvrženou adresou odesilatele a odražením emailu zpět bychom tedy pouze spamovali nevinné. Další výhodou je, že spam v tomto případě neplýtvá místem na disku, jelikož se nikam neukládá. Spameři jsou informováni o tom, že jejich pokus o doručení spamu byl neúspěšný, budou tedy tuto adresu nadále považovat za nesprávnou. Výhoda spočívá také v tom, že je-li příchozí email neprávem klasifikován za spam, odesilatel je o tom také informován, a může se tedy pokusit to nějak napravit.

DNSBL (DNS-based Blackhole Lists):

DNSBL je prostředek, jímž může internetová stránka zveřejnit seznam IP adress, ze kterých pochází nevyžádaná pošta a to ve formátu, který je snadno dostupný programům na internetu. Tato technologie je postavená na Systému domén (DNS). Poštovní server se tak může nakonfigurovat, aby zamítal zprávy pocházející ze stránky zmíněné na jednom nebo více takovýchto seznamů.

Prosazování technických požadavků protokolu SMTP:

Protokol SMTP má své technické požadavky (RCF), kterými se řídí při jakékoliv komunikaci. Obyčejný spamer bývá nezřídka vybaven poměrně nekvalitním software nebo není schopen splnit všechny tyto standardy, protože nemá legitimní moc nad počítačem, ze kterého spam zasílá. To znamená, že nastavením restrikcí na poštovním serveru (MTA) může administrátor tohoto serveru značně snížit spam.

Kontrola HELO/EHLO:

Při komunikaci klienta elektronické pošty s poštovním serverem (MTA) si obě strany vyměňují různé informace. Poštovní server poslouchá na portu 25 a čeká na vyrušení, požadavek, klienta. Klient naváže komunikaci s poštovním serverem v okamžiku, kdy chce např. poslat e-mail. Tuto komunikaci zahajuje příkazem „HELO mydomain.com“. Server mu na to odpoví tímtéž a komunikace může dále pokračovat.

Spam může být omezen některými malými kontrolami, které se bohužel většina dodavatelů MTA neobtěžuje do těchto software implementovat. Jedná se na příklad o vyžadování platného FQDN (fully qualified domain name) v příkazu HELO, což je jednoznačná doména, která specifikuje svou pozici ve stromové hierarchii DNS. Jen tato kontrola může zredukovat spam o celých 25%.

Greylisting:

Greylisting dočasně odmítá zprávy od neznámých poštovních serverů. Dočasné zamítnutí je značeno chybovým kódem 4xx, který každý normální poštovní server chápe a opět iniciuje odeslání později. Tato metoda je založená právě na tom, že spammer většinou nepošle zprávu znovu, zejména z důvodu nákladů plynoucích ze strategie znovuodesílání. Jeho činnost je automatická a hromadná a všechno opakovat by ho tudíž stálo dost peněz.

Whitelisting

Jen malé množství organizací nabízí techniku Whitelistingu IP adresy nebo licence, kterou lze doplnit do e-mailu (za určitý poplatek), čímž systém ujišťuje příjemce, že zpráva je licencovaná, a tudíž se nejedná o spam. Tento systém vyžaduje zákonné vymáhání již zmíněné licence, což sebou nese značné potíže. Společnost, která vydělává tím, že tyto licence vydává, je především zaměřena na zisk a tedy co největší počet vydaných licencí a to většinou na úkor přísného vymáhání pravidel, která jsou s vydáváním licencí spojena.

Whitelisting se většinou využívá v kombinaci s Greylistingem, aby nedocházelo ke ztrátám korektních e-malů v případě, že by poštovní server odesilatele chybně zareagoval na 4xx error a nepokusil by se zprávu odeslat znovu, což se občas stává.

Filtrování obsahu dle určitých pravidel (heuristické skenování)

Tyto filtrovací techniky spoléhají na administrátora poštovního serveru a jeho specifikaci seznamů slov nebo výrazů, které by v příchozích e-mailech nebyly povolené. To znamená, pokud by například jeho webová stránka obdržela spam inzerující viagru, administrátor by toto slovo zakomponoval do filtrovacího seznamu a server by pak dále takovéto spamy zamítal. Tento způsob filtrování však nehledí pouze na obsah emailu, ale i na jeho hlavičku, z níž lze také leccos poznat. Hlavička obsahuje veškeré informace o e-mailu. Spamer obyčejně tyto informace nějak přepisuje a upravuje, aby skryl svou identitu, nebo aby e-mail vypadal korektně, toto se dá také

často zachytit při filtrování, jelikož tyto přepisy většinou neodpovídají parametrům RCF.

Tato statická metoda filtrování má však i své nevýhody. Za prvé je dosti časově náročné udržovat již zmiňovaný seznam aktuální. Dále pak filtrování dle obsahu může jednoduše vyfiltrovat nespamy z důvodu, že se v nich pouze objeví často spamované téma. Další nevýhodou je opět vychytralost spammerů, kteří ono často užívané slovo, většinou předmět jejich podnikání, jen nepatrně pozmění, aby ho filtry neviděli ale uživatelé e-mailu snadno přečetli (například užijí místo slova „viagra“ „V1agra).

Statistické neboli Bayesiánské filtrování obsahu:

Jakmile je tato technika spuštěna, nevyžaduje už žádnou úpravu ani údržbu. Namísto toho uživatelé označují příchozí zprávy za spam, nebo nespam a toto filtrovací software se těmito úsudky učí, jak to do budoucnosti dle obsahu příchozí zprávy rozpoznat za nás. Z těchto důvodů statistický filtr není ovlivněn autorem softwaru, ani administrátorem poštovního serveru, ale přímo námi. Hledí totiž na naše rozhodování a z toho se učí. Email je po příchodu rozložen na slova pro která se pak jednotlivě zjišťuje pravděpodobnost, že email obsahující toto slovo je spam. Jako příklad softwaru pro statistické filtrování mohu uvést například Bogofilter, DSPAM, SpamBayes, nebo emailové programy Mozilla Thunderbird, Mailwasher a nejnovější verze SpamAssassin.

Fuzzy logic filtrování

Je filtrování podobné Bayesiánskému filtrování. Jedná se o filtrování za pomoci fuzzy logiky. Fuzzy logika funguje na principu, kdy věci patří do různých skupin určitou mírou. Příkladem fuzzy logiky může být například tvrzení: „Pokoj patří ze 70% do skupiny ‘moc teplý‘, a ze 30% do skupiny ‘moc chladný‘“. Tímto způsobem se příchozí pošta řadí do skupin „spam“ a „ham“. Pokud hodnoty fuzzy logiky překročí u příchozí pošty určitý práh, přiřadí se zpráva do složky „Spam“. Rozdíl mezi Bayesovým filtrováním a Fuzzy logikou spočívá ve vstupním souboru dat, který je tvořen z příchozí pošty. Zatímco Bayesiánský detektor spamu tvoří soubor dat na základě striktně matematických poznatků o spamu, detektor využívající fuzzy logiku bude při tvorbě datové sady obsahovat v určité míře i data založená na lidských poznatcích, což by se dalo považovat za jeho velkou výhodu oproti statistickému filtrování.

Hybridní filtrování:

Tato metoda využívá všech výše zmíněných metod a každou metodu má ověnčenou určitým počtem bodů. Každá zpráva je pak podrobena všem těmto testům a jednotlivé výsledky testů jsou opět vyjádřeny číselně a sečteny. Pokud tento součet převyšuje určitou fixní hranici, zpráva je zamítnuta a označena za spam. Tím, že žádná samostatná metoda nestačí k zajištění toho, že je zpráva opravdu spam, se velice snižuje pravděpodobnost chybného označení e-mailu za spam.

Filtrování kontrolními součty:

Tento druh filtrování využívá toho, že každý spammer rozesílá pořád téměř stejné zprávy. Filtrace kontrolními součty z takových zpráv odebere vše, co se mezi jednotlivými zprávami může lišit. Zbytek zprávy zredukuje podle nějakého algoritmu do kontrolní sumy a poté nahlédne do databáze, která shromažďuje kontrolní součty patřící zprávám označeným za spam. Pokud se tato suma v databázi objeví, zpráva je pravděpodobně spam.

Autenticita a reputace:

Tato metoda je opět založena na existenci určitého seznamu, stejně jako techika DNSBL. Rozdíl je však v tom, že nejde o „černou listinu“, ale naopak o „bílou“. Na tomto seznamu, který také funguje pomocí DNS, se uchovávají naopak spolehlivé webové stránky a někdy i jejich reputace.

Sender Policy Framework (SPF)

Jedna z novějších technologií, jak se vypořádat se spamem se nazývá Sender Framework Policy a byla vydána společností *Anti-Spam Research Group (ASRG).* SPF funguje na velice jednoduchém principu. Definuje, které přístroje na síťi mohou posílat emaily.

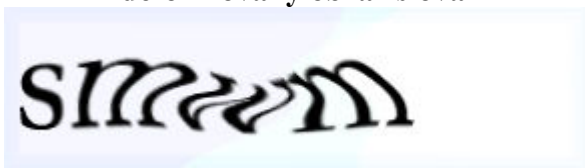
SPF funguje jako závora na protokolu SMTP, která využívá záznamů v DNS a umožňuje internetovým providerům identifikovat mailové servery na síťi, které mají povoleno rozesílat emaily z dané domény. Hostitel, který zasílá email a přitom není na seznamu povolených odesilatelů může být identifikován a buď odstraněn, přidán do karantény, nebo přezkoumán dodatečným antispamovým filtrem.

Některé další automatizované despamové techniky

Despamových software bylo za tu dobu, co nás spam obtěžuje, vyvinuto opravdu mnoho. Výše uvedený seznam vypisuje ty více užívané ať už v podobě výše popsané anebo poněkud pozměněné. Mezi další nástroje despamových technik patří například pomalá odezva, kterou využívají různé softwary z toho důvodu, že klasický nespamující server si počká, zatímco spamer je netrpělivý. Klient musí dle norem RFC při komunikaci se SMTP serverem počkat na úvodní pozdrav tohoto serveru, ale nedočkaví spameři nečekají a tudíž jsou odhaleni. Další metodou, která se neukazuje být právě tou nejúčinnější a nejoblíbenější je Systém výzvy/odezvy, kde se od neznámých serverů vyžaduje, aby podstoupily jakýsi test, než systém jejich zprávu doručí. Tato metoda je dosti kontroverzní z toho důvodu, že z ní samotné poněkud čiší spam.

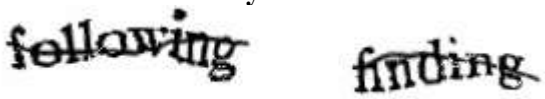
Doposud jsem zde probírala pouze despamové techniky, které probíhají na straně příjemce. Avšak i odesilatel e-mailů musí provádět řadu opatření, aby se ujistil, že neodesílá spam. Mnozí poskytovatelé internetu ověřují různými způsoby nové klienty, aby se ujistili, že nepřijali pod svou doménu spamera nebo jiného škůdce, a nepošpinili tak své jméno. Provideři tedy například často používají testování CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart), které bylo vynalezeno na Carnegie Mellon University. Podstata spočívá v tom, že počítač vygeneruje a vyhodnotí test, který sám není schopen vyřešit. Tímto testem se tedy snadno určí, zda uživatelem člověk, či program. Nejznámějším způsobem, s nímž se většina z nás běžně setkává je zdeformovaný obraz slova, který uživatel má za úkol opsat. Je to většinou slovo lehce čitelné, avšak pro počítač (robota rozesílajícího spam) už to znamená značnou námahu programátora. Dřívější styl obrazu (např. obrázek č. 6) používal například server Yahoo, nakonec se ale vývojem technologie podařilo ho programem přečíst.

Obrázek č. 6
Zdeformovaný obraz slova 1



Moderní CAPTCHA je oproti té původní, která byla založená na zdeformovaném obrazu, zaměřena na ztížení segmentace textu tím, že přes obraz vede křivá čára (Obrázek č. 7)

Obrázek č. 7
Zdeformovaný obraz slova 2



6. DOSTUPNÝ ANTISPAMOVÝ SW - POROVNÁNÍ

Každý emailový klient (Outlook, Eudore, Apple Mail, atd.) obsahuje filtry, které chrání naši schránku před přívalem spamu. Tyto antispamové filtry identifikují spam na základě předmětu, těla zprávy, nebo hlavičky emailu a to podle metod, které byly popsány v předchozí kapitole. Pokud filtr identifikuje příchozí email za spam, uloží ho do složky spam a my poté můžeme vyhodnotit, zda se skutečně jedná o spam či naopak.

Jak jsme již zmiňovali v kapitole 5, 86 % emailové pošty je nevyžádaného charakteru. Proto je zapotřebí příchozí poštu filtrovat, ještě než se stáhne a uloží do našeho klientského počítače, tedy na straně serveru.

Existuje celá řada antispamového software, jak pro klientskou stranu, tak stranu serveru, které jsou buď komerčního charakteru (většina sw), nebo s otevřenou licencí (např. SpamAssassin).

6.1. Antispamový sw na straně klienta

Klientské antispamy se instalují buď jako integrace emailového klienta, anebo jako samostatný software, který filtruje poštu, ještě než se dostane k emailovému klientovi. Samostatný software nám svou nezávislostí na emailovém klientovi dopřává svobodu při výběru emailového klienta, zatímco add-on vyžaduje použití takového emailového

klienta, který je tímto softwarem podporován. Na druhé straně integrátory emailových klientů většinou nevyžadují žádné velké úsilí, co se týče správy a nastavení a uživatel si vždy může nastavit filtrování spamů dle vlastních individuálních potřeb.

Jednotlivé antispamové programy se liší převážně přístupem k detekci spamu, metodami, jakých toho dosahují a samozřejmě také cenou.

6.2. Antispamový sw na straně serveru

Na straně serveru je to podobné, antispamový software je instalován buď přímo na emailový server, anebo do popředí tohoto serveru. Tato řešení umožňují filtraci emailů před jejich distribucí po síti a tím snižují dopad spamu na produktivitu koncových uživatelů a zároveň omezují zatížení síťových prostředků. Vzhledem k tomu, že serverové řešení je centralizováné, umožňuje jednoduché vykonávání antispamových a emailových procesů po síti, ale to bohužel většinou na úkor flexibility pro jednotlivé uživatele. Některá řešení umožňují prohlížení emailů označených za spam na serveru a tím také dávají možnost uživatelům upozornit síťové administrátory o výskytu legitimního emailu mezi spamy. Na základě takové stížnosti může administrátor opravit filtrační software k lepší přesnosti.

7. ANALÝZA ÚČINNOSTI ANTISPAMOVÝCH PROGRAMŮ

Posouzení kvality antispamového software se provádí převážně na základě jeho účinnosti. Účinnost antispamového software je jeho schopnost zachytit spam. Tato charakteristika se počítá pomocí následujícího vzorce:

1- (Počet false-negative klasifikací/počet spamů ve vzorku)

Počet false negative klasifikací je počet validních emailů označených za spam. Další charakteristikou, která se dá použít pro posouzení kvality antispamového nástroje je přesnost. Přesnost je přesným opakem účinnosti a jedná se o schopnost softwaru propustit legitimní poštu.

Některé antispamové nástroje jsou více účinné, některé méně. Jde o to jaký zaujímají přístup k detekci spamu a jaké despamové techniky používají. Dostupnost těchto softwarů je dána převážně jejich finanční náročností. Některé jsou bezplatné, většina však něco stojí.

Jeden z mála bezplatných antispamových řešení, který zároveň vede v žebříčcích hodnocení despamových software dle účinnosti je SpamAssassin. Toto řešení je navrženo pro síťovou architekturu klient-server, přičemž server by měl běžet na Unixovém operačním systému. SpamAssassin je spolehlivý, výkonný, snadno modifikovatelný pro naše potřeby a hlavně je zdarma. Proto je tato práce dále zaměřená pouze na tento SW, na jeho detailní popis, na principy na kterých funguje, na jeho využití a v neposlední řadě na ukázku jeho instalace.

7.1. SpamAssassin

Obrázek č. 8
Logo SpamAssassinu

SpamAssassin je program vydaný pod otevřenou licencí Apache Licence 2.0 a používaný pro filtrování spamů, na základě většiny dostupných filtračních technologií. SpamAssassin se obecně považuje za jeden z nejúčinnějších a nejpoužívanějších spamových filtrů a jeho úspěšnost v odhalování spamu se blíží 100%. Tento program je

určen pro síťovou architekturu klient-server, což znamená, že se skládá ze dvou částí (klientské a serverové)

SpamAssasin používá řadu technik pro obranu proti spamu ke kterým patří odmítání zpráv ze známých spamujících serverů, kontrola adresy odesilatele, filtrování podle klíčových slov v předmětu nebo těle zprávy, filtrování podle formátu a bayesiánské filtrování. K jednotlivým filtrům, které SpamAssassin implementuje, se dostaneme dále.

Spamassassin je oblíbeným nástrojem pro odhalování spamu zejména z těchto důvodů:

- Používá řadu různých pravidel a udává jim váhy podle jejich schopnosti diagnostikovat spam. Pravidla, která jsou účinnější při diskriminaci spamu od nespamu mají vyšší váhy.
- Váhy dané jednotlivým pravidlům lze modifikovat dle vlastního uvážení. Je také možné přidávat nová pravidla pomocí regulérních výrazů.
- SpamAssassin je adaptabilní s každým emailovým prostředím a učí se rozpoznávat, kteří odesílatelé jsou věrohodní a kteří nikoliv.
- Program umí reportovat spamy několika různým povolaným střediskům a lze u něho vytvořit *spamové pasti* – emailové adresy které se využívají pouze k forwardování spamů těmto střediskům.
- Je to software distribuovaný pod otevřenou GNU Public Licence, je tedy zdarma. Software se dá jakkoliv modifikovat a distribuovat dále.

7.1. Jak SpamAssassin funguje

SpamAssassin implementuje řadu filtračních pravidel, kterým udává určité bodové ohodnocení. Každé pravidlo testuje email nějakým způsobem a každému pravidlu je také přiděleno bodové ohodnocení.

Při příchodu nové zprávy je tato zpráva testována na základě všech těchto pravidel. Pokud se ukáže, že zpráva z nějakého důvodu splňuje podmínky daného pravidla je bodové ohodnocení tohoto pravidla přičteno k celkovému skóre daného emailu. Po otestování všech dostupných pravidel je celkové skóre porovnáváno s prahovou hodnotou. Pokud je skóre vyšší, než prahová hodnota, email je označen za spam.

SpamAssassin nefiltruje emialy, pouze je třídí na spam a ham. Nějaká další část systému doručování zpráv musí být pak nakonfigurována tak, aby se emaily filtrovaly dle příznaku, který jim SpamAssassin udělí.

7.2. Jak se dá SpamAssassin modifikovat

SpamAssassin je možné všelijak modifikovat pro naše potřeby. Systémový administrátor nebo i jednotliví uživatele mohou mít velkou kontrolu nad tím, jak SpamAssassin třídí příchozí poštu na spam a ham.

V první řadě je možné změnit prahovou hodnotu, která určuje bod, od kterého spadá email do spamu. Dále je možné měnit bodové ohodnocení jednotlivých pravidel podle našich potřeb. Například pokud nechceme, aby SpamAssassin využíval při skóringu emailu filtrování na základě klíčových slov, přiřadíme tomuto pravidlu hodnotu 0. V neposlední řadě nám SpamAssassin dovoluje užívat externích antispamových nástrojů, jako blacklisty open raley mail serverů a databází spamů.

Do SpamAssassinu můžeme také přidávat vlastní pravidla, podle kterých chceme filtrovat příchozí poštu. Software je psaný v Perlu a tedy veškeré integrace do něho musí být v tomto kódu. Řadu z nich můžeme najít na internetu od různých uživatelů, kteří již před námi cítili potřebu tento software trochu rozšířit pro vlastní potřeby.

7.3. Antispamové techniky SpamAssassinu

Spamassassin využívá následujících antispamových technik:

- **Statistické filtrování (Byesovo)** – tuto metodiku jsem již popisovala v kapitole 5. SpamAssassin obsahuje Byesovo filtrování spolu s nástroji, jak toto filtrování „vycvičit“. Tento filtr se dá také nastavit, aby se automaticky učil podle příchozích spamů a hamů.
- **Analýza hlavičky emailu –** odesílatele, předmětu, atd. SpamAssassin obsahuje řadu testů hlavičky emailu.

- **Negativní pravidla** – SpamAssassin umožňuje uživatelům vytvářet pravidla, která budou odečítat určitý počet bodů od celkového skóre. Některá pravidla s negativním bodovým ohodnocením jsou již v SW obsažena, například pokud se určitý řetězec v těle zpávy vyskytuje na whitelistu uživatele je celkové skóre snížené o 100 bodů. Administrátor určité společnosti může například přidávat negativní pravidla pro výskyt názvu produktu, prodávaným touto společností, v emailu.
- **Interní nebo lokální whitelisty a blacklisty –** seznamy povoleného nebo zakázeného odesilatele, řetězce v obsahu zprávy, atd.
- **Databáze obsahů emailu -** porovnává celý email s veřejně dostupnými databázemi na internetu, které udržují databáze spamů, protože většina spamů je odesílána mnoha uživatelům zároveň (např.: http://razor.sf.net, http://rhyolite.com/anti-spam/dcc, http://pyzor.sf.net).
- **SPF (Sender Policy Framework) –** Ověření, zda hostitel odesílající email z určité domény má povolení odesílat z této domény.

Nejnovější verze SpamAssin v3.2.5 implementuje celkem 746 různých testů, z nichž každý má své skóre. Testy fungují na principech popsaných výše a jsou různé pro každou z testovaných oblastí: hlavička, tělo zprávy, ostré tělo zprávy, URL nebo celá zpráva. Některé příklady těchto testů reprezentující jednotlivé antispamové techniky popsané výše jsou v následující tabulce.

Tabulka č. 1
Vybrané testy SpamAssassinu v3.2.x

AREA TESTED	DESCRIPTION OF TEST	TEST NAME	DEFAULT SCORES
uri	URI contains ".com" in middle	SPOOF_COM2OTH	2,044
header	Subject: contains string in the user's white-list	SUBJECT_IN_WHITELIST	-100
header	Subject: contains string in the user's black-list	SUBJECT_IN_BLACKLIST	100
header	From: address is in the user's black-list	USER_IN_BLACKLIST	100
header	From: address is in the user's white-list	USER_IN_WHITELIST	-100
header	SPF: sender matches SPF record	SPF_PASS	-0,001
body	Bayesian spam probability is 5 to 20%	BAYES_20	-0,74
body	Bayesian spam probability is 99 to 100%	BAYES_99	3,5
body	Message written in an undesired language	UNWANTED_LANGUAGE_BODY	2,8
rawbody	Sign often seen in spams	FR_MIDER	2,068
full	Listed in Pyzor (http://pyzor.sf.net/)	PYZOR_CHECK	3,7
full	Message has NUL (ASCII 0) byte in message	NULL_IN_BODY	2,425

V tabulce č.1 je 12 vybraných testů, které testují všechny možné oblasti testování.

První test SPOOF_COM2OTH prohlíží URL příchozí zprávu, zda neobsahuje „.com“ někde uprostřed. Pokud ano, přičte se k celkovému skóre 2,004 bodů.

Další testy se soustředí na hlavičku emailu. Zajímavé jsou například testy SUBJECT_IN_WHITELIST a SUBJECT_IN_BLACKLIST, které testují zda se předmět vyskytuje ve whitelistu, resp. blacklistu uživatele a na základě toho odčítají, resp. přičítají hodnotu 100 k celkovému skóre. Whitelisty a blacklisty mají tedy na celkové skóre značný vliv. Zatímco většina ostatních testů disponuje hodnotami od 0,001 do 5, testy porovnávající hlavičku, obsah, odesilatele, atd. s whitelisty nebo blacklisty mají bodové ohodnocení 100.

Další test SPF_PASS využívá filtrační metody SPF (Sender Policy Framework), kdy je prověřováno, zdali daná IP adresa má povolení odesílat z dané domény. Tento test má podle defaultního bodového ohodnocení velice malý vliv na celkové skóre.

Testy týkající se těla zprávy jsou převážně založeny na statistickém filtrování. Jak je vidět z tabulky č. 1, pokud je spamová pravděpodobnost od 5 do 20% (BYES_20) nabývá test negativní hodnoty, pokud se však chýlí ke 100% (BYES_99), bodové hodnocení je již positivní. Je tedy evidentní, že čím vyšší procento spamové pravděpodobnosti, tím větší vliv na finální skóre.

Druhým zmíněným testem, který se týká obsahu je UNWANED_LANGUAGE_BODY. Tento test přičte hodnotu 2,8, pokud se v obsahu zprávy vyskytuje nějaký cizí jazyk.

Testy, které hodnotí email na základě ostrého obsahu hodnotí především znaky v něm použité. Tento princip zachycuje další test FR_MIDER, který přičítá hodnotu 2,068 v případě, že email obsahuje některý ze znaků vyskytujících se obvykle ve spamech.

Poslední dva příklady testů SpamAssassinu třídí emaily na základě zprávy jako celku. Například, pokud se jedná o zprávu, která se vyskytuje na jednom z veřejně dostupných internetových zdrojů obsahujících databáze spamů, přičítá se hodnota 3,7. Dále se také dá uvažovat o spamu v případě, že velikost obsahu zprávy je NUL. Tedy zpráva je prázdná. V tomto případě se přičítá k celkovému skóre 2,45 bodů.

7.4. Příklad emailu vyhodnoceným SpmaAssassinem za spam

Následující pasáž (8) obsahuje email, který SpamAssassin správně vyhodnotil za spam. Zvýrazněné jsou pasáže doplněné SpamAssassinem.

```
From riverol5380503@jubii.dk Fri Nov  7 18:26:05 2003
Received: from localhost [127.0.0.1] by localhost
        with SpamAssassin (2.60 1.212-2003-09-23-exp);
        Sun, 09 Nov 2003 12:24:22 -0600
From: "brianj" <riverol5380503@jubii.dk>
To: <Undisclosed.Recipients@mailin-2.priv.cc.uic.edu>
Subject: Live your dream life!!                MPNWSTU
Date: Fri, 07 Nov 2003 15:32:41 -0800
Message-Id: <000016646728$00007347$00000042@mail3.mailnara.net>
X-Spam-Status: Yes, hits=12.9 required=5.0 tests=CLICK_BELOW,
        FORGED_MUA_EUDORA,FROM_ENDS_IN_NUMS,MISSING_OUTLOOK_NAME,
        MSGID_OUTLOOK_INVALID,MSGID_SPAM_ZEROES,NORMAL_HTTP_TO_IP,
        SUBJ_HAS_SPACES,SUBJ_HAS_UNIQ_ID autolearn=no version=2.60
X-Spam-Flag: YES
X-Spam-Checker-Version: SpamAssassin 2.60 (1.212-2003-09-23-exp)
X-Spam-Level: ************
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="----------=_3FAE8656.371BED4D"

This is a multi-part message in MIME format.

------------=_3FAE8656.371BED4D
Content-Type: text/plain
Content-Disposition: inline
Content-Transfer-Encoding: 8bit

Spam detection software, running on the system has
identified this incoming email as possible spam.  The original message
has been attached to this so you can view it (if it isn't spam) or block
similar future email.  If you have any questions, see
the administrator of that system for details.

Content preview:  Do you owe large sums of money? Are you stuck with high
  interest ra{tes? We can help! You can do what tens of thousands of
  americans have done, consolidate your high interest bills into one
  easy, low interest, monthly payment. [...]

Content analysis details:   (12.9 points, 5.0 required)

 pts rule name              description
---- ---------------------- --------------------------------------------------
 1.0 SUBJ_HAS_SPACES        Subject contains lots of white space
 4.3 MSGID_SPAM_ZEROES      Spam tool Message-Id: (12-zeroes variant)
 0.9 FROM_ENDS_IN_NUMS      From: ends in numbers
 0.2 NORMAL_HTTP_TO_IP      URI: Uses a dotted-decimal IP address in URL
 0.2 SUBJ_HAS_UNIQ_ID       Subject contains a unique ID
 4.3 MSGID_OUTLOOK_INVALID  Message-Id is fake (in Outlook Express format)
 0.1 MISSING_OUTLOOK_NAME   Message looks like Outlook, but isn't
 1.9 FORGED_MUA_EUDORA      Forged mail pretending to be from Eudora
 0.0 CLICK_BELOW            Asks you to click below

The original message was not completely plain text, and may be unsafe to
open with some email clients; in particular, it may contain a virus,
```

```
or confirm that your address can receive spam.  If you wish to view
it, it may be safer to save it to a file and open it with an editor.


------------=_3FAE8656.371BED4D
Content-Type: message/rfc822; x-spam-type=original
Content-Description: original message before SpamAssassin
Content-Disposition: attachment
Content-Transfer-Encoding: 8bit

Received: (qmail 25515 invoked from network); 7 Nov 2003 18:26:02 -0600
Received: from mailin-2.cc.uic.edu (HELO mailin-2.priv.cc.uic.edu) (128.248.155.213)
  by email0.cc.uic.edu with SMTP; 7 Nov 2003 18:26:02 -0600
Received: from mail3.mailnara.net (c-24-98-136-187.atl.client2.attbi.com [24.98.136.187])
        by mailin-2.priv.cc.uic.edu (8.12.10/8.12.9) with ESMTP id hA80PxJk011669;
        Fri, 7 Nov 2003 18:26:00 -0600
Message-ID: <000016646728$00007347$00000042@mail3.mailnara.net>
To: <Undisclosed.Recipients@mailin-2.priv.cc.uic.edu>
From: "brianj" <riverol5380503@jubii.dk>
Subject: Live your dream life!!                MPNWSTU
Date: Fri, 07 Nov 2003 15:32:41 -0800
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="----------=_1068251164-2528-687"
X-Priority: 3
X-MSMail-Priority: Normal
X-Mailer: QUALCOMM Windows Eudora Version 5.1
X-MimeOLE: Produced By Microsoft MimeOLE V5.00.3018.1300
Content-Length: 2290
Lines: 72

  Do you owe large sums of money? Are you stuck with
  high interest ra{tes? We can help!

  You can do what tens of thousands of americans have
  done, consolidate your high interest bills into one
  easy, low interest, monthly payment.

  By first reducing, and then completely removing your
  d+ebts, you will be able to start fresh. Why keep
  dealing with the stress, headaches, and wasted money,
  when you can consolidate your d+ebt and pay them off
  much sooner!

  Click below to learn more:

  http://61.186.254.9?affiliateid=mailer10


  hjeuubnfs

  ------------=_3FAE8656.371BED4D-
```

V první zvýrazněné části začínající „X-Spam-Status“ je vyhodnocení spamu a seznam testů s nenulovou hodnotou. Spam byl vyhodnocen jako „Yes“, tedy ano, jedná se o spam. Práh spamu je od 5,0 bodů. Tento email získal 12,9 bodů. Další údaje

obsahují informace o verzi SpamAssassinu na které byl spam vyhodnocován a další údaje.

V další zvýrazněné pasáži je již seznam testů a jejich bodové ohodnocení. V následujících testech vyšla nenulová hodnota:

SUBJ_HAS_SPACES – předmět obsahuje mnoho mezer

MSGID_SPAM_ZEROES- ID zprávy obsahuje 12 nul

FROM_ENDS_IN_NUMS –uživatelské jméno odesilatele končí číslicemi

NORMAL_HTTP_TO_IP – používá tečkovanou IP adresu v URL

SUBJ_HAS_UNIQ_ID – předmět obsahuje unikátní ID

MSGID_OUTLOOK_INVALID- falešné ID zprávy (ve formátu Outlook Express)

MISSING_OUTLOOK_NAME- Zpráva vypadá, jako z Outlooku, ale není

FORGED_MUA_EUDORA- falšovaný mail, tvářící se, že přišel od Eudory

CLICK_BELOW – Vyzývá ke kliknutí níže.

Email byl tedy označen za spam protože obsahoval mezery navíc v předmětu zprávy, Id zprávy obsahovalo moc nul, uživatelské jméno odesilatele zprávy končilo číslicemi, byla zde použitá tečkovaná IP adresa v URL, předmět obsahoval unikátní ID, ID zprávy bylo falešné, zpráva se tvářila, že pochází z Outlook Expresu, ale nepocházela, a autor vyzýval ke kliknutí níže.

Přestože zpráva obsahuje řadu triků k ošálení detektorů spamu, jako například vkládání znaků do slov „rates“ a „debt“, SpamAssassin jí identifikoval na základě jiných podezřelých charakteristik a přiřadil jí poměrně vysoké skóre.

7.5. Ukázková instalace SpamAssassinu na klientský počítač s OS Windows XP

SpamAssassin se skládá ze dvou částí: spamc a spamd, přičemž spamc tvoří klientskou část a spamd démona na serveru. Klientská část spamC vyžaduje, aby serverová část běžela na nějakém systému, se kterým může komunikovat pomocí TCP/IP. Serverová část SpamD běží pouze na Unixu.

Instalace SpamAssassinu (spamC) na klientský software tedy předpokládá, že uživatel bude přistupovat ke své poště přes emailový server, který má nainstalovanou serverovou část tohoto software, tedy spamd. Instalaci serverové části bohužel nemohu názorně ukázat, je totiž možná pouze na serveru s Unixovým operačním systémem, kterým nedisponuji. V této práci tedy pouze poskytnu návod, jakým se řídit při instalaci SpamAssassinu na klientské straně. Takovouto instalací docílím jakési odlehčené verze SpamAssassinu. Tato verze pak nebude číst poštu ze serveru a bude nutné jí pro otestování funkčnosti ukázkové zprávy manuálně předložit. To nám pro ozkoušení funkčnosti a pro případnou ukázku modifikovatelnosti bude stačit.

Jak jsem již nastínila výše SpamAssassin je software vyvinutý pro prostředí unixových operačních systémů. V jádře je SpamAssassin sadou modulů napsaných v programovacím jazyce Perl, spolu se skriptem v Perlu, který přijímá zprávu na standardním vstupu a testuje ji za použití těchto modulů. Z těchto důvodů je za potřebí před samotnou instalací SpamAssassinu na OS Windows vytvořit vhodné prostředí pro tento software. Samozřejmě jsou k dispozici různé distribuce SpamAssassinu, které jsou již uzpůsobené pro klasické prostředí operačního systému Windows, ale tyto distribuce nejsou ve většině případů úplné, neimplementují řadu testů a nejsou tím pádem tak účinné.

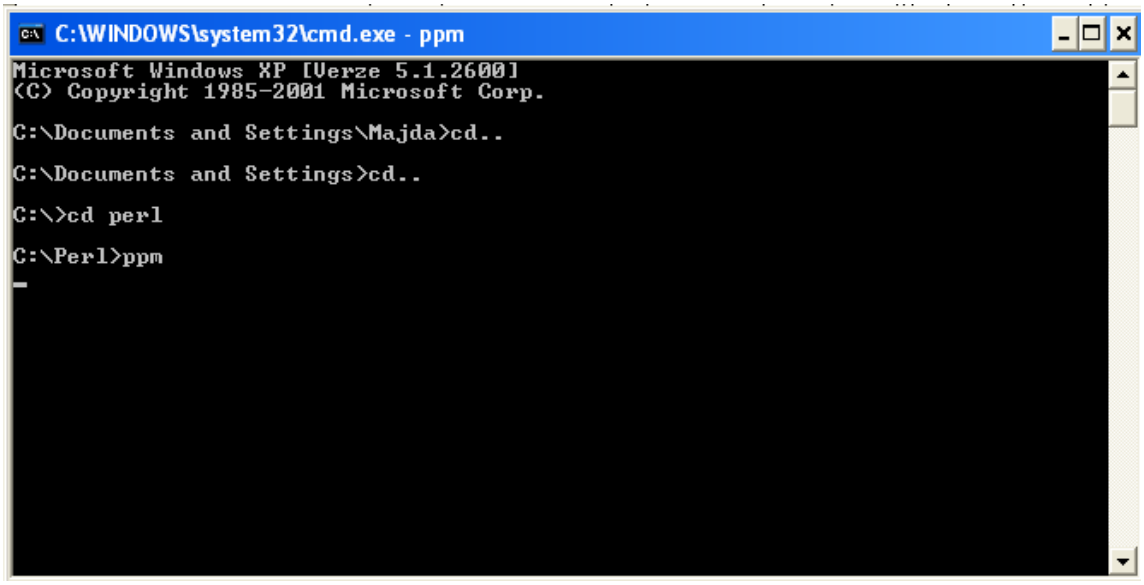
Průběh instalace bude tedy následující:

1) **Stažení a instalace ActivePerl** - prvním krokem instalace SpamAssassinu na klientský počítač s operačním systémem Windows bude instalace distribuce Perlu. Takovou distribucí je ActivePerl, což je otevřená distribuce Perlu zkompilovaná pro prostředí operačního systému Windows. Verze, která je potřebná pro instalaci SpmaAssassinu je ActivePerl 5.8.7.813 nebo pozdější, je dostupná z www.activestate.com

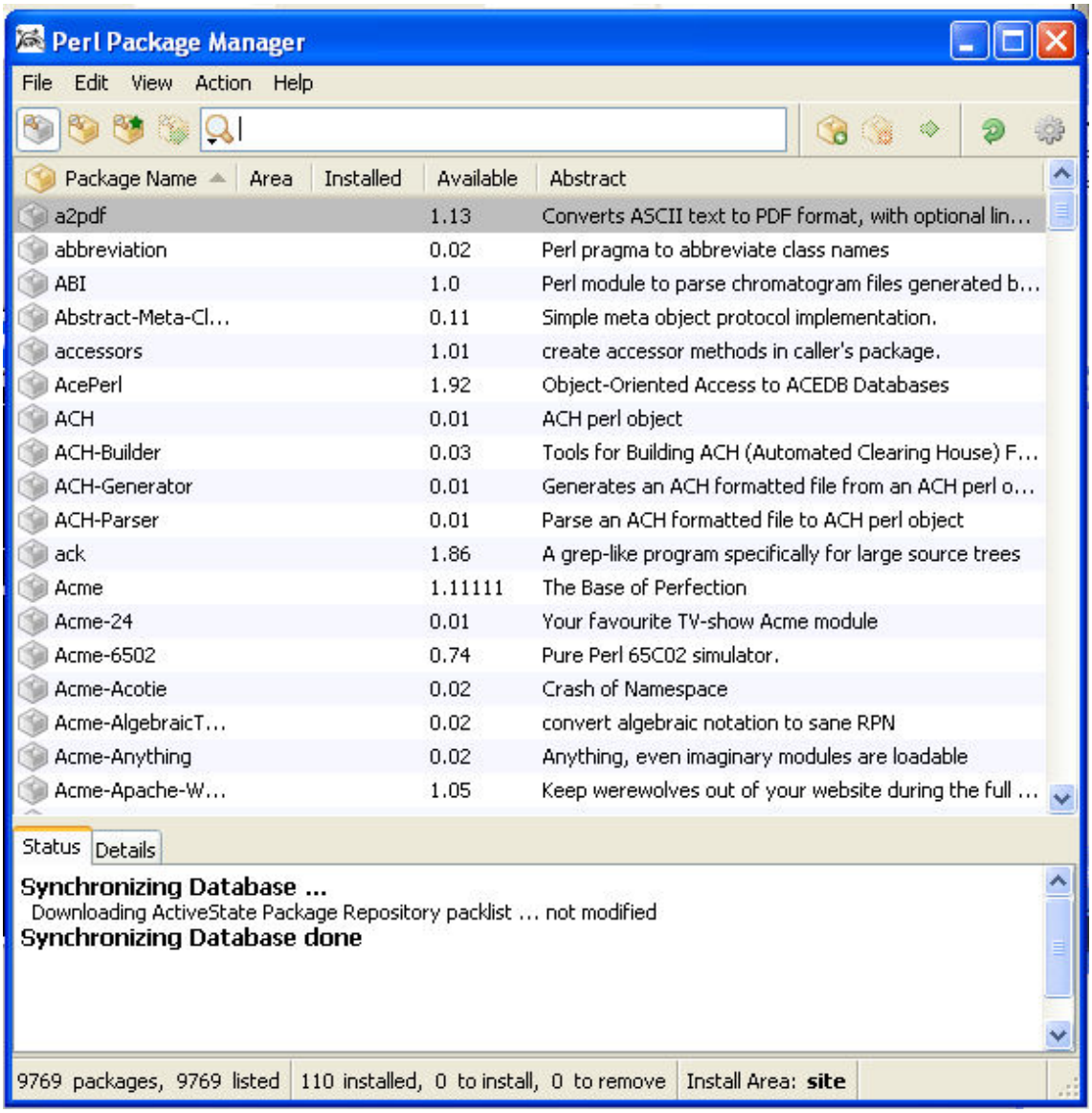
2) **Stažení a instalace NMAKE –** vzhledem k tomu, že OS Windows v sobě defaultně neobsahuje instalaci „make“ programu, nainstalujeme NMAKE. NMAKE je za potřebí pro stavbu projektů za použití cmd (příkazového řádku), je to program, který vytvoří vykonatelný kód na základě přečtení „make“ souboru (.mak), který je v podstatě skriptem, který specifikuje, co se má provést,

jaké soubory je za potřebí pro výstupy, atd. V našem případě nám NMAKE umožní stavbu dodatečných Perl modulů, kterých je za potřebí pro použití SpamAssassinu. NMAKE stáhneme z: http://download.microsoft.com/download/vc15/Patch/1.52/W95/EN-US/Nmake15.exe. Stažený NMAKE.exe rozbalíme. Vytvoří se nám tím tři soubory: NMAKE.exe. NMAKE.err, a README.txt, které nakopírujeme do /Perl/bin.

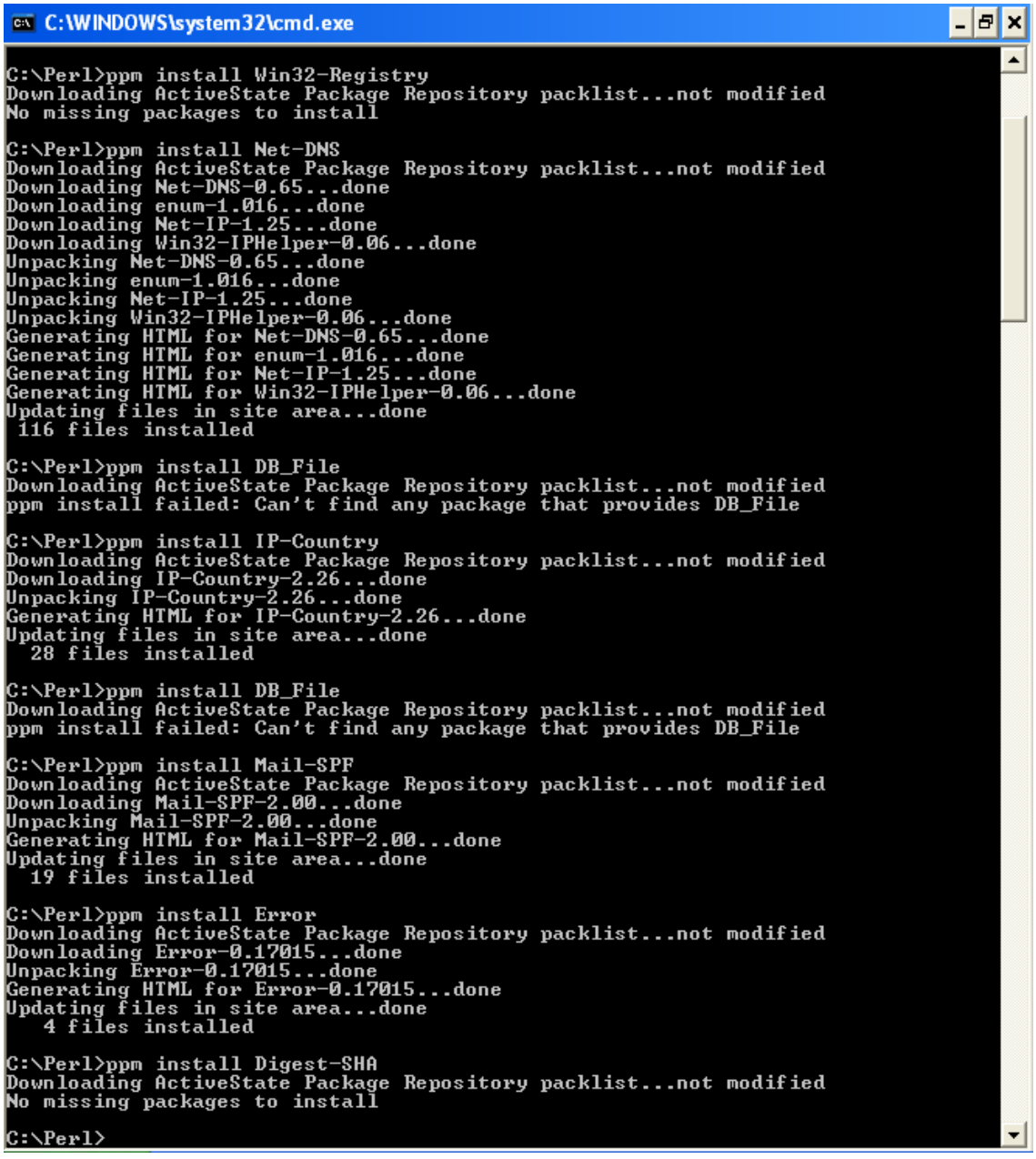
3) **Instalce modulů pomocí PPM (Perl Package Manager) –** pomocí manažéra pro hledání, instalaci a aktualizaci Perl modulů, PPM, budeme instalovat potřebné moduly. Manažer je dostupný přes příkazový řádek, nebo interaktivní prostředí PPM. Přes příkazový řádek je instalace snažší, spustíme tedy PPM následovně:



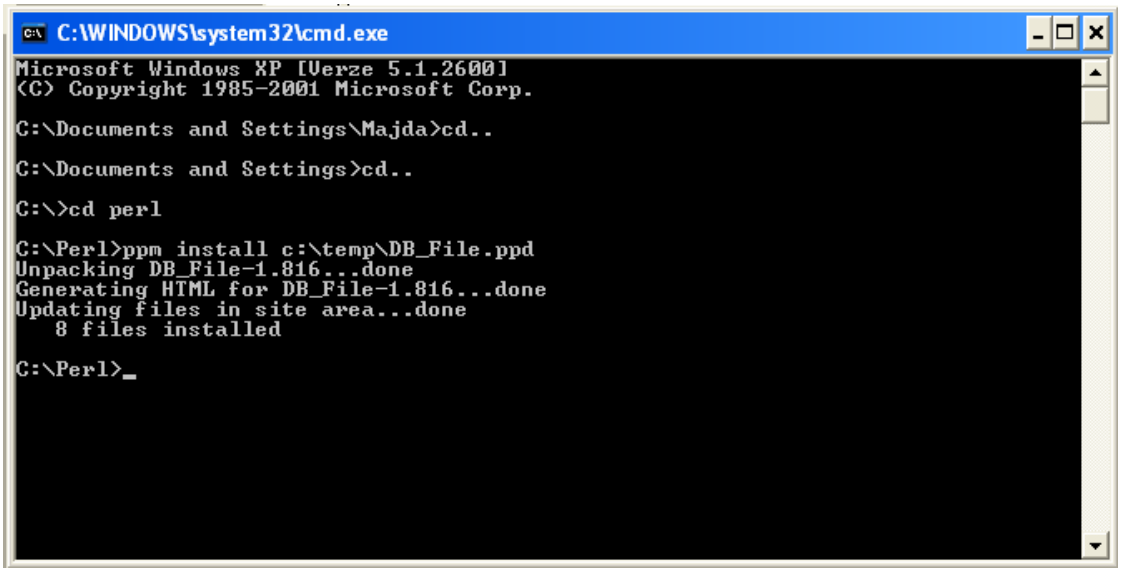
Spustí se nám prostředí PPM, který vypadá takto



Pokračujeme v příkazovém řádku instalací následujících modulů potřebných pro běh SpamAssassinu: Win32-Registry, Net-DNS, DB_File, IP-Country, Mail-SPF, Error, Digest-SHA.



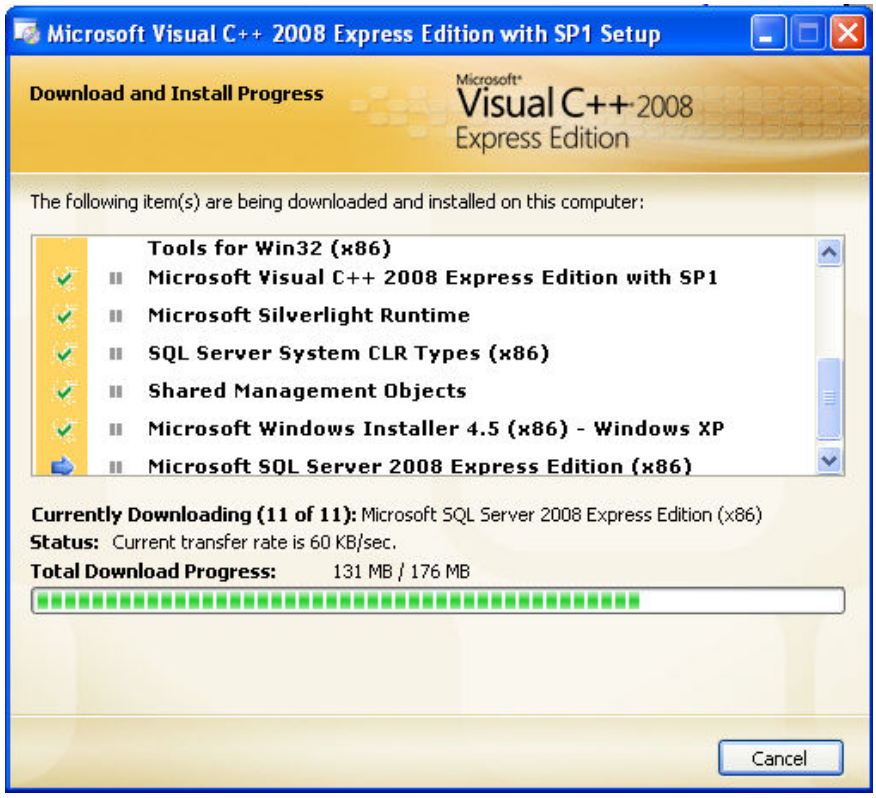
Jak je vidět, podařilo se nainstalovat všechny potřebné balíčky až na Win32-Registry, který je již nainstalován, a DB_File, který PPM nenalezl. Modul DB_File musíme tedy zkusit nainstalovat jinou cestou. DB_file stáhneme z: http://ppm.activestate.com/PPMPackages/zips/, kde vybereme verzi na základě čehož budeme přesměrováni na stránku se seznamem dostupných modulů pro tuto verzi. V seznamu vyhledáme DB_File a stáhneme. Po stažení tento .zip rozbalíme někam na C, např. do C:\Temp. Poté můžeme pokračovat s instalací DB_File v příkazovém řádku. Instalaci spustíme příkazem *ppm install C:\temp\DB_File.ppd.* Zde je průběh instalace:



Jak je vidět, tímto způsobem se nám podařilo nainstalovat poslední chybějící modul.

4) **Instalace Microsoft Visual C++ 6.0, nebo pozdější verze:** Tento nástroj instalujeme z toho důvodu, že pro instalaci SpamAssassinu potřebujeme kompilátor C. Tento nástroj ve verzi 9.0 je zdarma k dispozici zde: http://www.microsoft.com/express/vc/#webInstall. Instalace je jednoduchá přes průvodce. Je dobré změnit původní cestu, kam se bude instalovat na kratší cestu bez mezer, tedy například C:\VC2008. Je to z toho důvodu, že NMAKE by si s dlouhou cestou neporadil.

Průvodce instalací vypadá následovně:



5) **Instalace SDK (Software developement kit) –** SDK je sada nástrojů, které nám umožňují vytvářet aplikace například pro určitý softwarový balík, hardwarovou platformu, nebo operační systém. Vhodný devkit stáhneme z: http://www.microsoft.com/downloads/details.aspx?FamilyId=E6E1C3DF-A74F-4207-8586-711EBE331CDC&displaylang=en.
Opět je žádoucí uložit instalační soubory do zkrácené cesty, například C:\SDK, aby si s tím NMAKE poradil.

6) **Instalace SpamAssassinu-** Nyní již máme sestavené vhodné prostředí pro instalaci SpamC (tedy klientského SpamAssassinu). Stáhneme instalační balíček SpamAssassinu z: http://spamassassin.apache.org/downloads.cgi?update=200806121419. Balíček rozbalíme na disk C. Spustíme si Visual Studio 2008 Command Prompt přes START/Programy/Microsoft Visual C++ 2008 Express Edition/Visual Studio Tools/ Visual Studio 2008 Command Prompt. V příkazovém řádku Visual Studia si otevřeme složku SpamAssassinu příkazem: *cd Mail-SpamAssassin-3.2.5*. Pokračujeme spuštěním souboru Makefil.pl příkazem *Makefile.pl* a

odpovídáme na otázky kladené v příkazovém řádku. Dále pokračujeme trojicí příkazů: *NMAKE, NMAKE test, NMAKE install.* Proběhne kompletní instalace, ověří se dostupné funkce a testovací technologie a vytvoří se dokumentace.

Touto instalací bylo dosaženo jakési odlehčené verze. SpamAssassin v této podobě dokáže rozpoznat Spam od Hamu na základě všech dostupných testů. Zprávy se mu však musí podávat. Chybí nám totiž démon, který čeká na příchozí zprávy a následně s nimi manipuluje, popřípadě přepisuje předmět zprávy, pokud je email vyhodnocen za spam. Jak již bylo řečeno spamd neběží na operačním systému Windows. Toto by se dalo obejít například pomocí Cygwinu, speciálního unixového emulovaného prostředí pro Windows. Instalace Cygwinu je však složitá a není předmětem této diplomové práce. Důležité je, že nyní máme SpamAssassin v takovém stavu, ve kterém ho můžeme otestovat a zkusit si některé z možných modifikací.

7.6. Testování SpamAssassinu:

A) Podáním Spamu –GTUBE : Můžeme použít například testovací mail GTUBE (Generic Test for Unscoliceted Bulk Email). Test spustíme jednoduše příkazem : *spamassassin –D < sample-spam.txt.* V souboru sapmle-spam.txt je obsah GTUBE, který vypadá následovně.

Subject: Test spam mail (GTUBE)
Message-ID: <GTUBE1.1010101@example.net>
Date: Wed, 23 Jul 2003 23:30:00 +0200
From: Sender <sender@example.net>
To: Recipient <recipient@example.net>
Precedence: junk
MIME-Version: 1.0
Content-Type: text/plain; charset=us-ascii
Content-Transfer-Encoding: 7bit

This is the GTUBE, the
Generic
Test for
Unsolicited
Bulk
Email

If your spam filter supports it, the GTUBE provides a test by which you can verify that the filter is installed correctly and is detecting incoming spam. You can send yourself a test mail containing the following string of characters (in upper case and with no white spaces and line breaks):

XJS*C4JDBQADN1.NSBN3*2IDNEN*GTUBE-STANDARD-ANTI-UBE-TEST-EMAIL*C.34X

You should send this test mail from an account outside of your network.

Jde zde o tělo zprávy, konkrétně dlouhý řetězec v něm obsažený. Tento řetězec by měl SpamAsssassin poznat a vyhodnotit testem GTUBE. Tento test ohodnotí email 1000 body, to znamená, že daleko překročí prahovou hodnotu pěti bodů. Pokud tedy necháte SpamAssassin ohodnotit GTUBE email, mělo by Vám vyjet následující hodnocení:

Jak je vidět z tohoto okna příkazového řádku. SpamAssassin vyhodnotil tento email za spam (červená šipka). Poté vypsal skóre, podle kterého vyhodnotil email za spam (zelená šipka). A následovně vypsal testovací metody, které použil pro vyhodnocení tohoto emailu (žlutá šipka). Test GTUBE vyšel positivní, a tedy tento email dostal 1000 bodů.

B)Podáním Hamu: Použijeme klasický validní email na jakékoliv téma.

V tomto případě například následující zprávu od garanta této DP, Alexandra Vasilenka, která zve na sjezd diplomantů:

Tue, 17 Mar 2009 13:45:51 +0100 (CET)
Received: from Pefgw-MTA by mail.pef.czu.cz
with Novell_GroupWise; Tue, 17 Mar 2009 13:45:47 +0100
Message-Id: <49BFA986.6270.0055.0@pef.czu.cz>
X-Mailer: Novell GroupWise Internet Agent 7.0.3
Date: Tue, 17 Mar 2009 13:45:42 +0100 (CET)
From: =?us-ascii?Q?Alexandr=20Vasilenko?= <vasilenko@pef.czu.cz>
To: jirihodulik@centrum.cz,
=?us-ascii?Q?David=20Polomis?= <d.polomis@elektrizace.cz>,
honza.srp@email.cz,
=?iso-8859-2?Q?Jakub=20Mal=FD?= <maly.jakub2@gmail.com>,
=?us-ascii?Q?Michal=20Wachfaitl?= <wachulik@gmail.com>,
cerninv@seznam.cz,
cz775170@seznam.cz,
=?us-ascii?Q?Jirka=20Podhorny?= <Jirka.Podhorny@seznam.cz>,
=?iso-8859-2?Q?Luk=E1=B9=20Komberec?= <Lukas.Komberec@seznam.cz>,
madla.matouskova@seznam.cz,
=?iso-8859-2?Q?Tom=E1=B9=20Kopa?= <tomas.kopa@seznam.cz>,
xhulv105@studenti.czu.cz,
=?iso-8859-2?Q?Luk=E1=B9=20Pavl=EDk?= <XPAVL104@studenti.czu.cz>
Subject: =?iso-8859-2?Q?Semin=E1=F8=20diplomant=F9=20v=20Kostelci=20nad=20=C8ern=FDmi=20lesy?=
Mime-Version: 1.0
Content-Type: multipart/mixed; boundary="=__Part634B9866.4__="
X-Smtpd: 1.1.9@14043
X-Nod32result: clean (vdv=3942)

This is a MIME message. If you are reading this text, you may want to
consider changing to a mail reader or gateway that understands how to
properly handle MIME multipart messages.

--=__Part634B9866.4__=
Content-Type: text/plain; charset=UTF-8
Content-Transfer-Encoding: 8bit
Content-Disposition: inline

Dobrý den,

pro všechny i pro Ty, kteří nemohou přijet. Návod na sestavení
prezentace DP - ať už na seminář nebo na finální prezentaci.

Osnova prezentace:

1.Název práce, jméno a školitel
2. Cíle práce
3. Postup řešení - metodika
4. Vlastní postup práce s dílčími poznatky (počet slidů dle potřeby)
5. Zhodnocení získaných dat
6. Porovnání s cíli práce.

Toť vše - nesnažte se popisovat něco, co v práci pouze používáte, například vyjmenovávat standardy wi-fi je zbytečnost. Soustřeďte se hlavně na to, co jste dělali Vy - Váš přínos a zbytku věnovat minimum času.

Prezentaci máte mít na 10-15 minut, tak důraz na to, co je podstatné a co je opravdu Vaše dílo.

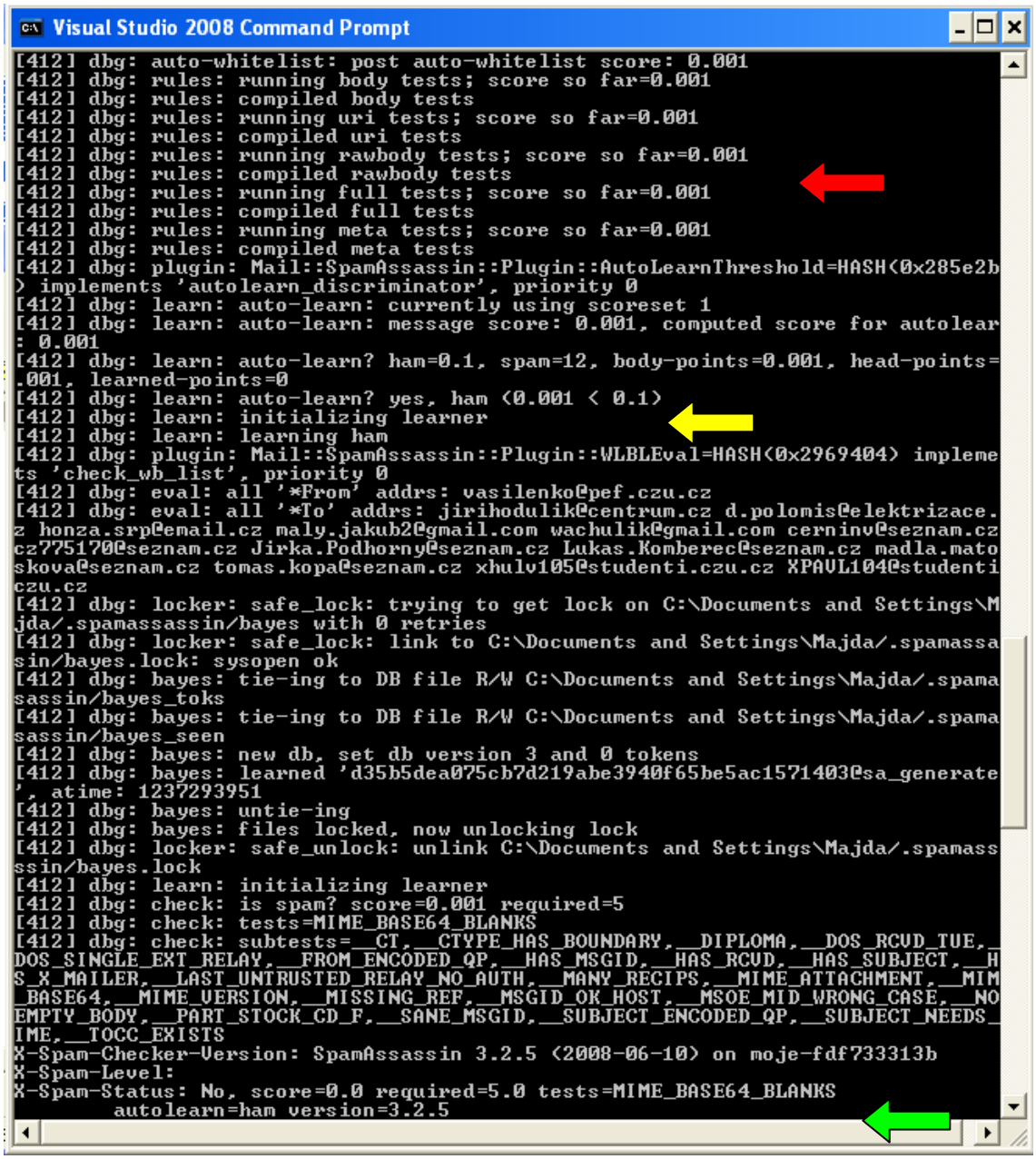
P.S. Většina lidí buď potvrdila účast nebo se omluvila - kromě Vostřela a Hulína - u těch se bude sitace řešit dále.

Hezký den

--=__Part634B9866.4__=
Content-Type: application/msword; name="Seminar_diplomantu_KIT_program09.doc"
Content-Transfer-Encoding: base64
Content-Disposition: attachment; filename="Seminar_diplomantu_KIT_program09.doc"

--=__Part634B9866.4__=--

SpamAssassin ohodnotil tento spam následovně:



Jak je vidět, SpamAssassin označil tento email za HAM (zelená šipka). Score bylo vypočítáno na 0,001 (červená šipka). Toto skóre nám vyšlo na základě positivního testu MIME_BASE_64_BLANKS, který testuje, zda se v emailu nevyskytují nadbytečné mezery v kódování BASE64.

Zajímavé zde je, jak SpamAssassin určuje, podle kterých emailů se bude učit Byesovskému filtrování, viz žlutá šipka. Testuje tu, zda je skóre těla zprávy nižší než hodnota 0,1. Pokud ano spustí tzv. „studenta HAMU“. Opačně tento princip funguje v případě, že je skóre těla zprávy vyšší než 12. Pokud se skóre pohybuje

někde mezi 0,1 a 12 student se nespustí a SpamAssassin se podle zprávy neučí. Tyto prahové hodnoty a jiné testovací skóre se dají u SpamAssassinu snadno modifikovat pro naše potřeby. Pokud například usoudíme, že spodní hranice pro učení spamu (12 bodů) je moc vysoká, můžeme ji snížit. Modifikovatelnost SpamAssassinu si ukážeme podrobně v další kapitole.

7.7. Ukázka modifikovatelnosti SpamAssassinu

Jak již bylo několikrát zmíněno, SpamAssassin je velice pružný nástroj ve smyslu uživatelské modifikovatelnosti pro individuální potřeby. Pokud máme pocit, že některé bodové ohodnocení určitého testu není dostatečné, můžeme to snadno změnit podle našich představ. Všechny změny, které provádíme, musí být zaznamenány v souboru local.cf, který se nachází v Mail-SpamAssassin-x.x.x\rules. V této složce nalezneme také všechna dostupná pravidla seřazená podle kategorií, co testují a jak. Samotné soubory s pravidly nikdy neměníme, protože by se nám změny přepsaly při každé aktualizaci SpamAssassinu. Pokud jsou překopírované v souboru local.cf, zůstanou i po upgradu.

Ukažme si tedy pár příkladů modifikace SpamAssassinu v souboru local.cf:

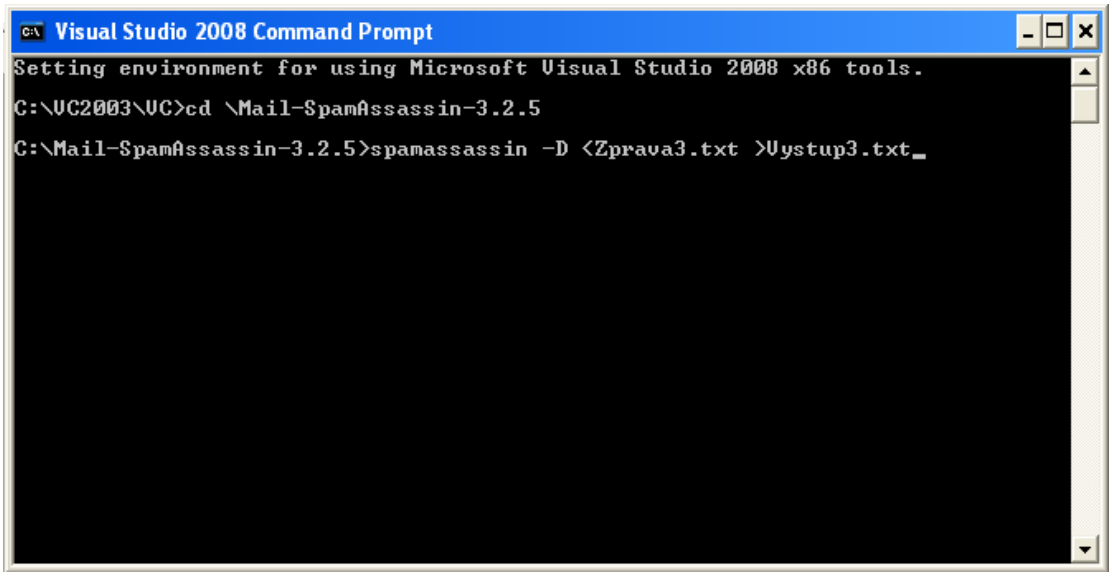
- Chceme-li například přidat kamarádku Petru do Whitelistu, protože emaily od ní zaručeně nejsou spam:
  whitelist_from petrah@seznam.cz

- Nebo se nám zdá prahová hodnota spamovosti moc mírná a chceme ji trochu snížit, aby filtrovala veškeré spamy:
  requiered_hits 3

- Nebo chceme změnit maximální/minimální skóre za tělo zprávy, při kterém se SpamAssassin učí Hamu/Spamu. Původně je zde nastavená maximální hodnota 0,1 pro Ham a minimální hodnota 12 pro Spam.
  bayes_auto_learn_threshold_nonspam 0.3

bayes_auto_learn_threshold_spam 10.0

Nyní si můžeme vyzkoušet, zda SpamAssassin bere naše změny na vědomí. Následujícím příkazem mu předložíme email, který je evidentně spam. Tentokráte použijeme i výstupní .txt, kam SpamAssassin uloží výsledek testování. Testování spustíme následujícím příkazem.



Vyhodnocení tohoto emailu v souboru vystup3.txt je pak následující:

From Enrike Lares Tue Mar 31 12:35:25 2009
Received: from localhost by moje-fdf733313b
with SpamAssassin (version 3.2.5);
Sat, 04 Apr 2009 09:16:05 +0200
From:
Message-Id: <BLU108-W533D81AA96453216C9A958B88A0@phx.gbl>
X-Spam-Flag: YES
X-Spam-Checker-Version: SpamAssassin 3.2.5 (2008-06-10) on moje-fdf733313b
X-Spam-Level: ****************
X-Spam-Status: Yes, score=16.1 required=3.0 tests=FH_FROMEML_NOTLD,
MIME_HEADER_CTYPE_ONLY,MISSING_DATE,MISSING_HB_SEP,MISSING_HE
ADERS,
MISSING_MIME_HB_SEP,MISSING_SUBJECT,TVD_RCVD_IP,TVD_RCVD_IP4
autolearn=unavailable version=3.2.5
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="----------=_49D70935.792A0000"

This is a multi-part message in MIME format.

----------=_49D70935.792A0000
Content-Type: text/plain; charset=iso-8859-1
Content-Disposition: inline
Content-Transfer-Encoding: 8bit

Spam detection software, running on the system "moje-fdf733313b", has identified this incoming email as possible spam. The original message has been attached to this so you can view it (if it isn't spam) or label similar future email. If you have any questions, see madla.matouskova@seznam.cz for details.

Content preview: Enrike Lares <enrikelares@hotmail.com> Add sender to Contacts Subject: Dear potential partner, Date: Tue, 31 Mar 2009 19:35:25 +0000 Importance: Normal MIME-Version: 1.0 Bcc: Content-Length: 2828 Compact Headers Dear potential partner, Do you need famous brand of electronic products with original quality and international warranty? Do you want to start your own business career for money making ? What ever you are a small personal business or largest wholesale entity we also can provide your support to be our stable customers or agent. We are largest wholesale business on consumming electronic products between America&China, laptops, Digital camera Videos,GPS,cellphone,mp4,game console and many other electronic products.which market is mainly in Europe,America,south Asia,Australia and Southen America. There is much profit for you if you are our stable customer or agent. [...]

Content analysis details: (16.1 points, 3.0 required)

pts	rule name	description
3.3	TVD_RCVD_IP4	TVD_RCVD_IP4
1.6	TVD_RCVD_IP	TVD_RCVD_IP
2.2	FH_FROMEML_NOTLD	E-mail address doesn't have TLD (.com, etc.)
0.0	MISSING_DATE	Missing Date: header
2.5	MISSING_HB_SEP	Missing blank line between message header and body
1.6	MISSING_HEADERS	Missing To: header
2.7	MISSING_MIME_HB_SEP	BODY: Missing blank line between MIME header and body
0.9	MIME_HEADER_CTYPE_ONLY	'Content-Type' found without required MIME headers
1.3	MISSING_SUBJECT	Missing Subject: header

The original message was not completely plain text, and may be unsafe to open with some email clients; in particular, it may contain a virus, or confirm that your address can receive spam. If you wish to view it, it may be safer to save it to a file and open it with an editor.

Jak je vidět, email byl správně vyhodnocen za spam a naše modifikace jsou již platné. Prahová hodnota se změnila na 3 a ostatní změny také vstoupili v platnost.

Celkem 9 testů vyšlo positivních. Mezi nimi například chybějící předmět zprávy nebo chybějící prázdný řádek mezi hlavičkou a tělem zprávy. Narazili jsme zde také na jeden problém SpamAssassinu, který se týká testů TVD_RCVD_IP4 a TVD_RCVD_IP. Tyto testy jsou positivní, pokud hlavička zprávy obsahuje v **doručeno: od** (Recieved: from) numerickou IP adresu. Toto je však problém například pokud někdo používá mail na yahoo.com. Tento server se totiž neobtěžuje se zpětným vyhledáním DNS a ponechá

adresu numerickou. Znamená to tedy, že přiřadí hodnotu 4.9 každému emailu (3.3(TVD_RCVD_IP4) + 1.6(TVD_RCVD_IP)).

8. ZÁVĚR

Cíl této práce kladl důraz na systematičnost a srozumitelnost objasnění spamu a technik jak se proti němu bránit, a převážně pak na výběr kvalitního antispamového software a jeho detailný popis a návod k instalaci.

Na základě zvážení funkčnosti, efektivity a cenové dostupnosti nabízených antispamových software byl zvolen SpamAssassin. Mezi primární důvody vedoucí k jeho výběru patří jeho bezplatnost, snadná modifikovatelnost, a hlavně jeho efektivita. Jak již bylo řečeno, jeho účinnost se blíží 100%.

Tento software byl dopodrobna popsán, co se týče funkčnosti, efektivity a modifikovatelnosti. Pro potenciální uživatele tohoto antispamového řešení byla poskytnuta ukázková instalace klientského sw na běžně užívaný operační systém Windows (32-bit). Instalace se ukázala být celkem bezproblémová a neměla by potenciálnímu uživateli zabrat déle než půl dne.

Po instalaci SpamAssassinu následovalo otestování tohoto software na několika emailech, jak spamů tak hamů. Testováním byla ověřena funkčnost a schopnost odhalovat spam a propouštět validní emaily.

Dále byla navržena možná modifikace pro potřeby uživatele, při které byl odhalen i jeden z problémů tohoto software týkající se neopodstatněného označení emailů za spamy od emailového poskytovatele Yahoo. Na základě tohoto nálezu bylo navrženo řešení, jakým tento problém eliminovat v podobě deaktivace pravidel, které chybně označují email za spam.

Samozřejmě, že žádné antispamové řešení se neobejde bez nedostatků. Důležité ale pak je, aby daný software poskytoval možnost se s nedostatky vypořádat. V našem případě je tato možnost poskytnuta ve snadné modifikovatelnosti SpamAssassinu.

V závěru tedy mohu konstatovat, že SpamAssassin splnil mé předpoklady a předpoklady na něho kladené v popisu tohoto produktu a je to tedy vhodný produkt pro implementaci při síťové architektuře klient-server.

9. SEZNAMY

9.1. LITERATURA

[1] Matoušková, Magdaléna: Spam a despamové techniky. Praha. 2004

[2] Hlavenka, Jiří: Používáme, využíváme [a zneužíváme] e-mail. Brno, Computer Press. 2002. ISBN 807226-606-3

[3] Kocman, Rostislav. Lohinský, Jakub: Jak se bránit vyrům, spamu a spyware. Brno, Computer Press. 2005. ISBN 80-251-0793-0

[4] Wolfe P., Schott Ch., Erwin M. W. : Antispam. Metody, nástroje a utility pro ochranu před spamem. Computer Press. 2005. ISBN 80-251-0497-6

[5] WikipediA, www.wikipedia.org

[6] LUPA – Server o českém internetu, www.lupa.cz

[7] IKAROS – Elektronický časopis o české informační společnosti, www.ikaros.cz

[8] IT PRAVO – Server o internetovém a počítačovém právu, www.itpravo.cz

[9] MESSAGING ANTI-ABUSE WORKING GROUP, www.maawg.org

[10] THE APACHE SPAMASSASSIN PROJECT, http://spamassassin.apache.org/

[11] Schwartz, Alan: SpamAssassin. O'Reilly. 2004. ISBN 0596007078.

[12] McDonald, Alistair: SpamAssassin: A practical guide to integration and configuration. Packt Publishing Ltd. 2004. ISBN 1904811124

[13] SpamHelp, www.spamhelp.cz

[14] Symantec – State of Spam Report, http://www.symantec.com/business/theme.jsp?themeid=state_of_spam

9.2. CITACE

1. LUPA.cz [online], 2003-01-22, [cit. 24.5.2007]
2. Spamie, Hubálek, Zdeněk, Ikaros [online], 2006-01-02, [cit. 23.5.2007],URL<http://www.ikaros.cz/node/2087>
3. Za spamming bude hrozit pokuta až 10 milionů korun, EPRAVO [online], 2004-07-16, [cit. 04.5.2007]

URL<http://www.epravo.cz/v01/index.php3?s1=Y&s2=4&s3=C&s4=0&s5=0&s6=0&m=1&recid_cl=27228&typ=clanky>

4. Za spamming bude hrozit pokuta až 10 milionů korun, EPRAVO [online], 2004-07-16, [cit. 10.5.2007] URL<http://www.epravo.cz/v01/index.php3?s1=Y&s2=4&s3=C&s4=0&s5=0&s6=0&m=1&recid_cl=27228&typ=clanky>
5. Matejka Ján, K zákonu č. 480/2004 Sb. o službách informační společnosti, Regulované činnosti (§8 zákona), ITpravo [online], 2004-11-16, [cit. 28.5.2007].)
6. Zákon č. 480/2004 Sb §2f
7. Mailový SPAM legálně?, Papucha [online], 2006-06-15, [cit. 30.5.2007] URL http://papucha.cz/blog/2006/06/spam-legalne/
8. A message tagged by SpamAssassin, Schwartz, Alan: SpamAssassin. O'Reilly, 2004, str.3.

9.3. OBRÁZKY

8.4. TABULKY