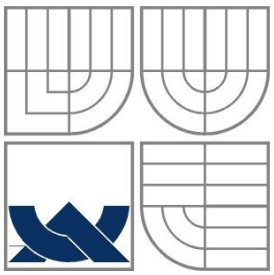
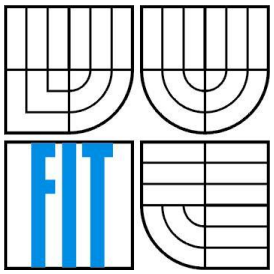


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÝCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER SYSTEMS

REALIZACE SPAMOVÉHO FILTRU NA BÁZI UMĚLÉHO IMUNITNÍHO SYSTÉMU

SPAM FILTER IMPLEMENTATION ON THE BASIS OF ARTIFICIAL IMMUNE SYSTEMS

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE
AUTHOR

BC. DAVID NEUWIRTH

VEDOUCÍ PRÁCE
SUPERVISOR

DOC. ING. LUKÁŠ SEKANINA, PH.D.

BRNO 2009

Vysoké učení technické v Brně - Fakulta informačních technologií

Ústav počítačových systémů Akademický rok 2008/2009

Zadání diplomové práce

Řešitel: **Neuwirth David, Bc.**
Obor: Informační systémy
Téma: **Realizace spamového filtru na bázi umělého imunitního systému**
Kategorie: Umělá inteligence

Pokyny:

1. Seznamte se s problematikou umělých imunitních systémů (UIS) a spamových filtrů.
2. Vypracujte studii na témata uvedená v bodě 1.
3. Navrhněte použití UIS pro filtraci spamů.
4. Navržený systém implementujte ve zvoleném programovacím jazyce a systému.
5. Ověřte účinnost navrženého systému.
6. Zhodnoťte dosažené výsledky.

Literatura:

- Dle pokynů vedoucího.

Při obhajobě semestrální části diplomového projektu je požadováno:

- První tři body zadání.

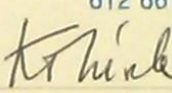
Podrobné závazné pokyny pro vypracování diplomové práce naleznete na adrese http://www.fit.vutbr.cz/info/szz/

Technická zpráva diplomové práce musí obsahovat formulaci cíle, charakteristiku současného stavu, teoretická a odborná východiska řešených problémů a specifikaci etap, které byly vyřešeny v rámci ročníkového a semestrálního projektu (30 až 40% celkového rozsahu technické zprávy).

Student odevzdá v jednom výtisku technickou zprávu a v elektronické podobě zdrojový text technické zprávy, úplnou programovou dokumentaci a zdrojové texty programů. Informace v elektronické podobě budou uloženy na standardním nepřepisovatelném paměťovém médiu (CD-R, DVD-R, apod.), které bude vloženo do písemné zprávy tak, aby nemohlo dojít k jeho ztrátě při běžné manipulaci.

Vedoucí: **Sekanina Lukáš, doc. Ing., Ph.D.**, UPSY FIT VUT
Datum zadání: 22. září 2008
Datum odevzdání: 26. května 2009

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
Fakulta informačních technologií
Ústav počítačových systémů a sítí
612 66 Brno, Božetěchova 2



doc. Ing. Zdeněk Kotásek, CSc.
vedoucí ústavu

Abstrakt

V dnešní době se v prostředí e-mailové komunikace setkáváme s velkým problémem s nevyžádanou poštou. Existuje několik metod, jak nevyžádanou poštu detekovat a jak ji rozlišit od pošty vyžádané. Tato diplomová práce popisuje, jakým způsobem lze využít umělé imunitní systémy k detekci nevyžádané pošty. Analyzuje a popisuje metody umělých imunitních systémů, které se dají využít k boji proti spamu. Druhá část této diplomové práce pojednává již o konkrétní realizaci spamového filtru na bázi umělých imunitních systémů a v závěru zhodnotí, zda je takto navržený spamový filtr úspěšnější nebo méně úspěšný než filtr využívající běžné metody detekce spamu.

Abstract

Unsolicited e-mails generally present a major problem within the e-mail communication nowadays. There exist several methods that can detect spam and distinguish it from the requested messages. The theoretical part of the master's thesis introduces the ways of detecting unsolicited messages by using artificial immune systems. It presents and subsequently analyses several methods of the artificial immune systems that can assist in the fight against spam. The practical part of the master's thesis deals with the implementation of a spam filter on the basis of the artificial immune systems. The project ends with comparison of effectiveness of the newly designed spam filter and the one which uses common methods for spam detection.

Klíčová slova

spam, nevyžádaná pošta, spamový filtr, detekce spamu, umělý imunitní systém

Keywords

spam, unsolicited e-mail, spam filter, spam detection, artificial immune system

Citace

NEUWIRTH DAVID: *Realizace spamového filtru na bázi umělého imunitního systému*, diplomová práce, Brno, FIT VUT v Brně, 2009

Prohlášení

Prohlašuji, že jsem tuto diplomovu práci vypracoval samostatně pod vedením doc. Ing. Lukáše Sekaniny, PhD.

Další informace mi poskytl Ing. Petr Matoušek, Ph.D.

Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

….....……

Bc. David Neuwirth

25. 5. 2009

Poděkování

Rád bych poděkoval vedoucímu mé diplomové práce doc. Ing. Lukášovi Sekaninovi, Ph.D. za cenné rady a trpělivost při konzultacích.
Dále děkuji Ing. Petrovi Matouškovi, Ph.D. za poskytnutí materiálů o filtraci nevyžádané pošty.
Speciální poděkování patří mé rodině, která mne všestranně podporuje ve studiu na vysoké škole a která mi byla oporou i při psaní této práce.

Obsah

1 Úvod

Dnes, na počátku 21. století, v době elektronické komunikace, bychom si život bez e-mailů nedokázali ani představit. Pomocí e-mailů komunikujeme se svými učiteli a profesory už od základní školy, odevzdáváme úkoly, projekty a seminární práce. E-mailem rezervujeme ubytování a letenky na naši vysněnou a dlouho očekávanou dovolenou a zanedlouho budeme pomocí e-mailů komunikovat i s naším budoucím zaměstnavatelem. E-mailová komunikace je dnes nepostradatelnou a nenahraditelnou formou komunikace a k našemu životu, ať už si to uvědomujeme nebo ne, neoddělitelně patří.

Každý z nás se ale už určitě setkal se situací, kdy si otevřel svého e-mailového klienta plný očekávání, že si přečte zprávy od svého příbuzného, kamaráda nebo obchodního partnera, ale místo těchto zpráv našel ve své poště spoustu nevyžádaných e-mailů, tzv. spamu[1].

S růstem rychlosti připojení k internetu firem rostou větší možnosti pro spamery. Podle zdrojů organizace MAAWG ([5] a [17]), která provedla průzkum 100 miliónů e-mailových schránek, bylo v obdobích od třetího čtvrtletí roku 2006 do druhého čtvrtletí roku 2007 81 až 87 procent e-mailů nevyžádaná pošta. Podle některých údajů tvoří nevyžádaná pošta dnes až 95% veškeré e-mailové komunikace a cena, kterou zaplatily americké firmy v roce 2007 za ztracenou produktivitu, za čas zaměstnanců potřebný k mazání spamu a za zařízení a spamové filtry, se vyšplhala na 13 miliard dolarů [6].

Problém spamu se týká nás všech a musíme se s tímto problémem tedy nějak vypořádat. Potřeba lepších a efektivnějších spamových filtrů je proto nutností. V prostředí e-mailové komunikace existuje spousta metod a technologií, pomocí kterých můžeme nevyžádanou poštu více či méně úspěšně detekovat a mazat. Mezi tyto technologie patří například kontrola klíčových slov, kontrola odesílatele e-mailů nebo kontrola zdrojového serveru, odkud e-mail pochází. Pro každou technologii detekce nevyžádané pošty najdeme na internetu spoustu komerčních ale také volně stažitelných programů a nástrojů. Některé tyto nástroje metody různě kombinují a jsou v boji se spamem méně či více úspěšné.

Umělé imunitní systémy jsou relativně nová oblast v informačních systémech a ve výpočetní technice obecně. Imunitní systémy jsou zajímavé svými vlastnostmi nejen z medicínského hlediska, ale obsahují spoustu vlastností, mechanismů a prvků, kterými se můžeme inspirovat také v technice, v našem případě při filtrování spamu. Mezi nejzajímavější vlastnosti biologických imunitních systémů patří například přizpůsobivost neznámým situacím, robustnost nebo také schopnost učit se [2]. V této diplomové práci zanalyzuju metody a principy umělých imunitních systémů a popíšu, jakým

[1] Spam je nevyžádané masově šířené sdělení, nejčastěji reklamní, přenášené internetem [1]. Mezi formy spamu patří jednak nevyžádané e-maily, ale také zaplavení diskusních fór, komentářů nebo zprávy posílané formou instant messagingu (např. přes ICQ).

způsobem by se prvky umělých imunitních systémů daly nasadit v boji proti spamu. V druhé části diplomové práce se pokusím vytvořit spamový filtr na bázi umělých imunitních systémů. Tento spamový filtr podrobím několika testům a experimentováním s nastavením určitých vstupních parametrů zanalyzuji, jak je tento navržený spamový filtr úspěšný v boji proti spamu.

2 Umělé imunitní systémy

V této kapitole si vysvětlíme principy imunitních systémů. Podíváme se, jak fungují imunitní systémy v biologii, a pokusíme se aplikovat určité poznatky na umělé imunitní systémy, které bychom mohli využít k filtraci spamu. K tomu, abychom vytvořili umělý imunitní systém, potřebujeme důkladně pochopit ten biologický a musíme z něj vybrat důležité principy a vlastnosti. Důkladný popis biologického imunitního systému najdete v mé bakalářské práci [2] nebo v [3], [4], [12], [13], [14] a [15]. V této kapitole se podíváme jen na ty principy a součásti biologického imunitního systému, které můžeme použít při filtraci spamu.

2.1 Biologický imunitní systém

Proč při boji se spamem vycházet právě z biologických imunitních systémů? V přírodě můžeme pozorovat spoustu případů, kdy zvířata přežívají různá zranění, léčí se z vážných i méně závažných nemocí a jsou pod neustálým útokem nejrůznějších bakterií a virů. To nás přivádí k myšlence a nápadu prozkoumat, proč jsou zvířata (a člověk) vůči těmto útokům a nehodám imunní a relativně snadno se s nimi vypořádávají, proč přežívají. Pokusíme se tedy zjistit, co činí imunitní systém tak výjimečným a co z něj dělá tak mocnou zbraň v boji proti infekcím, nemocím a zraněním.

Imunitní systém je soubor mechanismů v těle jedince, který zajišťuje identifikaci a eliminaci nepřátelských prvků. Imunitní systém je schopen detekovat útočníky od virů, přes nejrůznější bakterie až po různé plísně. Tito útočníci se obecně nazývají *patogeny* [4]. Samotná detekce je velice komplikovaná, neboť se patogeny v průběhu existence svého druhu vyvíjejí a různě mutují. Imunitní systém musí být tedy schopný rozpoznat prvky, se kterými se již dříve setkal, ale také prvky nové, které ještě nezná. Tato detekce je tedy důležitou, dokonce klíčovou vlastností, kterou každý dobrý spamový filtr potřebuje.

Imunitní systém disponuje řadou prostředků k identifikaci těchto cizích prvků. Od fyzických bariér, buněk pro všeobecnou ochranu, přes buňky se specifickým zaměřením na určité útočníky až po paměťové buňky, které si daný patogen pamatují.

Úkoly biologického imunitního systému bychom tedy mohli shrnout do dvou hlavních bodů. Prvním je identifikace (rozpoznání) cizího prvku a druhým důležitým bodem je reakce na tento cizorodý prvek, nejčastěji jeho eliminace. V případě filtrování spamu je pro nás důležitá právě identifikace cizích prvků. Z druhé částí se spíše než eliminací budeme zabývat zapamatováním si daného prvku a učením se.

2.1.1 Popis základních částí

Imunitní systém pracuje na několika vrstvách. První linií obrany je fyzická bariéra, která tělo chrání před vniknutím cizích látek. Pokud se patogenu podaří skrz tuto fyzickou bariéru proniknout (například drobnou kožní trhlinkou nebo díky zranění), pokusí se ho zastavit *vrozená imunita*. Tato vrozená imunita reaguje okamžitě, ale její reakce není specifická. Pokud se patogenu podaří proniknout i přes tuto část imunitního systému, čeká ho třetí vrstva - *adaptivní imunita*. Tato část imunitního systému se časem mění, přizpůsobuje se a učí se. Pokud tedy nějaký patogen zaútočí na naše tělo, adaptivní imunita si ho zapamatuje a při budoucím útoku bude reagovat daleko rychleji.

V této kapitole se tedy budeme zabývat jednotlivými částmi imunitního systému, podíváme se na důležité prvky a jakou úlohu v imunitním systému hrají.

2.1.1.1 Vrozená imunita

V této vrstvě hrají důležitou roli buňky, které hlídkují ve tkáních těla jedince. Na svém povrchu nesou řadu čidel, která reagují na určité specifické vlastnosti a struktury patogenů. Pokud zaznamenají útočníka, okamžitě vyšlou řadu signálů, které přilákají další buňky z krevního řečiště a které aktivují systém adaptivní imunity.

Pokud patogen vnikne do těla jedince, bude s určitou pravděpodobností vrozenou imunitou rozpoznán. Toto rozpoznání probíhá v prvotní fázi pomocí řady čidel, které se nazývají *Toll-Like receptory*. Tato čidla reagují na určité bílkoviny a proteiny specifické pro cizorodé prvky. Patogen, který je takto rozpoznán, je označen neboli *opsonizován* pomocí protilátek nazývaných *Imunoglobuliny*. Takto označený patogen pak putuje dále v těle jedince a může být označen dalšími protilátkami. Během označení patogenu je zároveň vyplavena řada *cytokinů*, z nichž některé spustí alarm nutící další buňky cestovat do místa zánětu a obecně řečeno uvádí imunitní systém do plné pohotovosti. Další buňky, tzv. *dendritické buňky*, naproti tomu stráví, neboli zničí, cizí prvky a pak putují do mízních uzlin, kde předloží fragmenty cizorodých prvků armádě buněk adaptivní imunity a také vyplaví cytokiny, což napomůže zahájit odpověď adaptivní imunity.

Důležitá vlastnost vrozené imunity je, že není specifická. Funguje na nejnižší biologicko-chemické úrovni a její chování by se dalo popsat jako „předprogramované“. Na rozdíl od adaptivní imunity reaguje na patogen jen obecnou reakcí a z dlouhodobého hlediska nemá vliv na vývoj schopností imunitního systému jako celku. Tato část imunitního systému hraje ale významnou roli.

2.1.1.2 Adaptivní imunita

Adaptivní imunita je oproti vrozené imunitě specifická. Její obranné buňky cíleně útočí na určitý typ patogenu. K reakci adaptivní imunity dochází pouze tehdy, je-li stimulována imunitou vrozenou. Adaptivní imunita se v průběhu životního cyklu jedince neustále vyvíjí a zdokonaluje.

Díky adaptivní imunitě je imunitní systém jedince obdařen *pamětí*. Jakmile je infekce potlačena, trénované *lymfocyty B a T* (nejdůležitější buňky adaptivní imunity) obklopí zbytky

patogenů a posléze se přemění na paměťové buňky. Díky této schopnosti máme možnost chránit se před chorobami očkováním. Pokud už je nepřítel známý (např. z dřívějšího útoku nebo právě díky očkování), je tvorba protilátek daleko rychlejší. Obranná reakce adaptivní imunity je tedy daleko rychlejší a efektivnější a jedinec často ani nepozná, že byl znovu nakažen.

Lymfocyty B a T

Lymfocyty typu B a lymfocyty typu T jsou nejdůležitější buňky adaptivního imunitního systému. Jsou to buňky, které dokážou detekovat patogen a určitým způsobem ho zajistit. Každý typ bakterie, viru nebo obecně jakékoli buňky je charakterizován svým *antigenem*, což je vlastně jeho charakteristický vzor nebo otisk. K tomu, aby bílé krvinky mohly rozpoznat patogen, mají na svém povrchu čidla, tzv. *receptory*, které dokážou daný antigen rozpoznat.

Lymfocyty typu B vznikají v kostní dřeni (B podle anglického názvu bone marrow) a lymfocyty typu T vznikají také v kostní dřeni, ale potom (zatím z neznámých důvodů) putují do brzlíku (T od anglického názvu thymus) a tam dozrávají. Každá bílá krvinka, když vstupuje do krevního řečiště, obsahuje právě jeden specifický antigenový receptor. Tato specifičnost je dána speciálními mechanismy, podle kterých se lymfocyty typu B a T utvářejí. Tyto mechanizmy mohou generovat milióny různých kombinací antigenových receptorů. V krvi jedince (v případě dospělého člověka) kolují milióny lymfocytů B a T a tudíž i milióny různých antigenových receptorů. Zde také nastává problém s tzv. *autoimunitou*, což je porucha imunitního systému, při které jsou buňky vlastního tělo označeny jako cizí a jsou postupně ničeny. Mezi nejznámější autoimunitní choroby patří cukrovka (kdy je eliminován inzulín) a revmatická artritida. Jak zabránit efektu autoimunity si řekneme v kapitole o algoritmu negativní selekce.

Pokud se bílá krvinka naváže svým receptorem na nějaký antigen patogenu, její metabolismus se prudce zrychlí a začne se množit. Tím se vytvoří stovky dalších lymfocytů, které jsou schopny detekovat určitý patogen. Tomuto množení se říká *klonální selekční expanze* (dále v textu si vysvětlíme a ukážeme, jak v umělých imunitních systémech funguje klonální selekční algoritmus).

Některé lymfocyty, potom, co se navážou na patogen, ho úplně zničí. Jiné lymfocyty patogeny jen rozlámou na malé kousky a s těmito fragmenty patogenu na svém povrchu odcestují do mízních uzlin. Podle nich se vytvoří během několika dní armáda lymfocytů B a T namířena přímo proti tomuto druhu patogenu.

2.2 Důležité principy imunitních systémů

Po pochopení biologických imunitních systémů, můžeme aplikovat tyto poznatky při návrhu a tvorbě umělých imunitních systémů. V biologickém imunitním systému je spousta prvků, které mají nějakou souvislost s detekcí a eliminací útočníků (T a B lymfocyty, dendritické buňky, cytokiny, TLR, …). Ačkoli jsou dobře popsány a prozkoumány, většinou stoprocentně nevíme, jaké přesně úlohy

v imunitním systému hrají. Při aplikaci principů z biologických imunitních systémů do umělých imunitních systémů si proto některé prvky zjednodušíme.

2.2.1 Prvky a principy

Mezi principy, které hrají v imunitních systémech důležitou roli, patří rozpoznávání a paměť. Na tyto dva principy se podíváme v následujících podkapitolách.

2.2.1.1 Rozpoznávání

Ve vrozené imunitě byly přítomny buňky jako neutrofily a dendritické buňky. V adaptivní imunitě hrály hlavní roli T a B lymfocyty. Z pohledu rozpoznávání je pro nás důležitá přítomnost receptorů, které mají schopnost rozpoznat a zachytit určitý vzor. Této části buňky říkáme *komplement* (anglicky antibody). Každý komplement je schopný rozeznat určitý vzor. Tomuto vzoru se říká *antigen*. Princip rozpoznání vzoru antigenem by se dal přirovnat ke klíči a zámku.

Komplementy budeme modelovat jako *řetězec bitů délky l*. Jako navázání komplementu na neznámý prvek budeme považovat shodu řetězce bitů na komplementu (detektoru) a řetězce bitů na antigenu (neznámém prvku). Komplement je tedy schopný rozeznat a navázat se jen na jeden určitý antigen. Jako příklad si můžeme uvést například často se vyskytující slova v těle nevyžádané pošty (např. viagra, rolex, software, discount).

K tomu, aby se komplement navázal na antigen, potřebujeme, aby se řetězce sobě navzájem rovnaly – aby byly shodné. To je ale v praxi těžko splnitelný požadavek – rozesílatelé spamu totiž začali určitá písmena nahrazovat číslicemi, které jsou daným písmenům podobná (např. v1agra, r0lex, s0ftware, d1sc0unt). Proto se využívá podobnosti řetězců. Existuje několik metod, pomocí kterých se dá určit, jak moc si jsou dva prvky podobné. Patří mezi ně například Hammingova vzdálenost, Euklidova vzdálenost nebo tzv. Manhattanská vzdálenost. Euklidova vzdálenost se vypočítá:

$$D = \sqrt{\sum_i (A_i - B_i)^2} \quad \text{a} \tag{1}$$

Manhattanská vzdálenost se vypočítá:

$$D = \sum_i |A_i - B_i|, \tag{2}$$

kde A_1 až A_n a B_1 až B_n jsou souřadnice dvou prvků (A a B). V případě dvou porovnávaných řetězců mohou být hodnoty A_1 až A_n a B_1 až B_n hodnoty jednotlivých prvků z daného řetězce. Například pokud uvažujeme dva řetězce písmen anglické abecedy: „abc“ a „eec“, kde hodnota písmene *a* je 1, písmene *b* je 2, písmene *c* je 3 a písmene *e* je 5, potom Euklidova vzdálenost těchto řetězců bude 5 a Manhattanská vzdálenost bude 7. Zabývat se budeme nejpoužívanější metodou, kterou je Hammingova vzdálenost. Hammingova vzdálenost má několik verzí, ale nejčastěji určuje, kolik odlišných prvků je ve dvou řetězcích nebo kolik změn musíme provést, abychom z jednoho

řetězce dostali ten druhý. Například Hammingova vzdálenost pro řetězce „software“ a „s0ftw4re“ je 2.

2.2.1.2 Paměť

Další důležitý princip, který můžeme najít v imunitních systémech, je schopnost adaptability. Adaptivní část imunitního systému disponuje pamětí.

Pokud se do systému (těla) jedince dostane nový neznámý patogen, který imunitní systém ještě nezná, dendritické buňky ho zachytí a dopraví do mízních uzlin. Zde je podle fragmentů přivezeného patogenu vytvořena a upravena armáda lymfocytů, které opustí mízní uzlinu a specificky vyrazí do boje proti tomuto patogenu. Po úspěšném zvládnutí nákazy se některé lymfocyty přemění v paměťové buňky imunitního systému (prodlouží se jejich životnost) a kolují po systému (těle) jedince pro případ, že by se nákaza vrátila. Pokud se do těla jedince dostane známý patogen, zachytí ho jeden z paměťových lymfocytů, který okamžitě spustí imunitní odpověď – sám lymfocyt se začne množit a vytvoří armádu lymfocytů namířenou proti tomuto patogenu. Fragmenty patogenu už tedy nemusí být dopraveny do mízních uzlin, aby tam pomohly vytvořit armádu lymfocytů, což výrazně urychlí imunitní odpověď a jedinec vůbec nemusí poznat, že byl znovu nakažen.

V biologických imunitních systémech nemůže být v těle jedince neomezený počet paměťových buněk imunitního systému. Životnost paměťových buněk je tedy určitým způsobem omezena, systém zapomíná. V imunitním systému pro usnadnění filtrace spamu budeme řešit kompromis mezi paměťovými nároky a rychlosti imunitní odpovědi. Další důvod, proč nepoužívat neomezenou a trvalou paměť spočívá v tom, že uživatel elektronické pošty potkává nové přátelé, komunikuje s novými obchodními partnery, mění své zájmy nebo se může dokonce naučit nový jazyk, ve kterém komunikuje. Trvalá paměť by v tomto případě byla tedy nepoužitelná.

2.2.2 Vlastnosti

Z modelu biologického imunitního systému bychom mohli abstrahovat několik významných vlastností, které najdou své uplatnění při návrhu umělého imunitního systému.

2.2.2.1 Paralelní činnost

Biologický systém dokáže fungovat a reagovat na více místech současně. Pokud se například patogen dostane do systému jedince na určitém místě, imunitní systém bude okamžitě reagovat. Během této reakce může jiný patogen zaútočit z jiného místa a imunitní systém musí být schopen reagovat i na tento a každý další souběžný útok.

2.2.2.2 Komplexnost

Imunitní systém je stavěn tak, aby byl schopný zareagovat na jakýkoli typ vetřelce (vir, bakterie, plíseň, …). Jakákoli buňka může být imunitním systémem napadena (včetně prvků samotného

imunitního systému). Zde musíme mít na paměti nebezpečí autoimunity, kdy jsou i vlastní buňky označeny jako cizí a je proti nim vedena imunitní odpověď. Pokud nám spamový filtr začne mazat korektní poštu je to velký problém, kterému musíme předcházet. Riziko autoimunity snižují algoritmy, které si popíšeme v následující kapitole.

2.2.2.3 Decentralizované řízení

Imunitní systém neobsahuje žádný centrální bod, který by fungování imunitního systému řídil či kontroloval. Absence tohoto prvku má velkou výhodu, protože dysfunkce centrálního řídícího bodu by mohla ovlivnit fungování celého systému. V biologickém imunitním systému je každý prvek zodpovědný sám za svou vlastní funkci. Tuto vlastnost bychom mohli v případě použití imunitních systému pro filtraci spamu využít asi až ve velkých korporátních sítích, kde máme spoustu poštovních serverů, které zpracovávají poštu samostatně, nezávisle na sobě.

2.2.2.4 Adaptabilita

Imunitní systém je schopen se přizpůsobovat a v průběhu životního cyklu jedince se učit efektivněji rozpoznávat nové nepřátele. Imunitní systém není schopen udržet si všechny paměťové buňky, protože nesmí přesáhnout maximální koncentraci lymfocytů v krevním řečišti. Funguje určitý kompromis mezi počtem prvků a dobou jejich existence v systému.

2.2.2.5 Distribuovaná struktura

Poslední, ale ne neméně cenná vlastnost, která má svůj význam při uplatnění poznatků z imunitních systémů například v bezpečnostních systémech, je distribuovaná struktura. Jsou dvě části imunitního systému, u kterých je tato vlastnost důležitá. První z nich jsou aktivní prvky, které se podílejí na eliminaci patogenu (nepřítele). V případě filtrace spamu nestačí mít spamový filtr na poštovním serveru, ale je výhodně mít určité součásti také na svém klientském počítači.

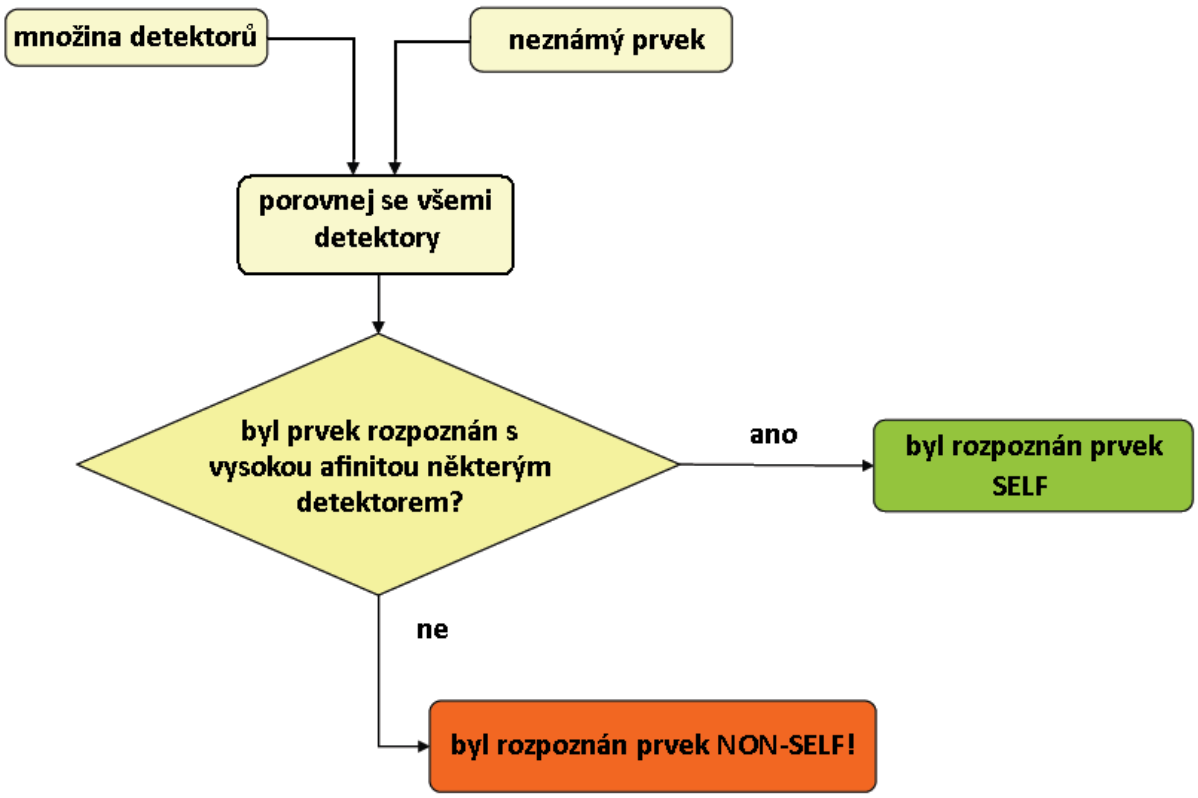
Druhá část imunitního systému, která by měla mít distribuovanou strukturu, je paměť. To znamená, že není jeden centrální bod, kde by byla uložena paměť systému, ale je rozložená po celé struktuře. Zde existují dva koncepty, buď mít identické kopie paměti na všech místech, nebo paměť rozložit a mít na každém místě jen určitou část (samozřejmě se zachováním určité redundance, abychom zajistili ještě větší bezpečnost).

2.3 Algoritmy používané v imunitních systémech

V této kapitole si vysvětlíme tři základní algoritmy, které se týkají imunitních systémů [3]. Patří mezi ně algoritmus pozitivní selekce, algoritmus negativní selekce a klonální selekční algoritmus. K tomu, abychom si mohli popsat, jak tyto algoritmy fungují, musíme si vysvětlit několik pojmů.

Jako *množinu S* nebo *SELF* budeme považovat množinu vlastních prvků. V biologických imunitních systémech patří do této množiny buňky vlastního těla. Je to množina, u které se budeme snažit zajistit konzistenci. Pokud se například do těla dostane cizorodá buňka, množina SELF je pozměněna a to musíme být schopni co nejrychleji detekovat.

Jako *řešení* (nebo také detektor) budeme označovat takový prvek, který je schopný rozpoznat určitý problém (např. cizí prvek, nepřátelskou buňku, vir nebo bakterii). V biologických imunitních systémech to byly imunoglobuliny, TLR receptory a lymfocyty typu T a B. Množinu všech možných řešení budeme značit jako *množinu P*. Množina P je tedy množina všech potenciálních detektorů. Zde nastává problém s autoimunitou, což znamená, že prvky množiny P jsou schopny označit za cizí i prvky z množiny SELF. Proto si zavedeme *množinu A*, do které za použití následujících algoritmů vložíme jen žádoucí prvky z P. To znamená, že v množině A budou pouze ty detektory, které rozpoznají a označí jen prvky, které neleží v množině SELF (tzv. non-SELF prvky). V opačném případě můžeme vytvářet takovou množinu A, ve které budou pouze ty detektory, které jsou schopny označit jen prvky z množiny SELF.



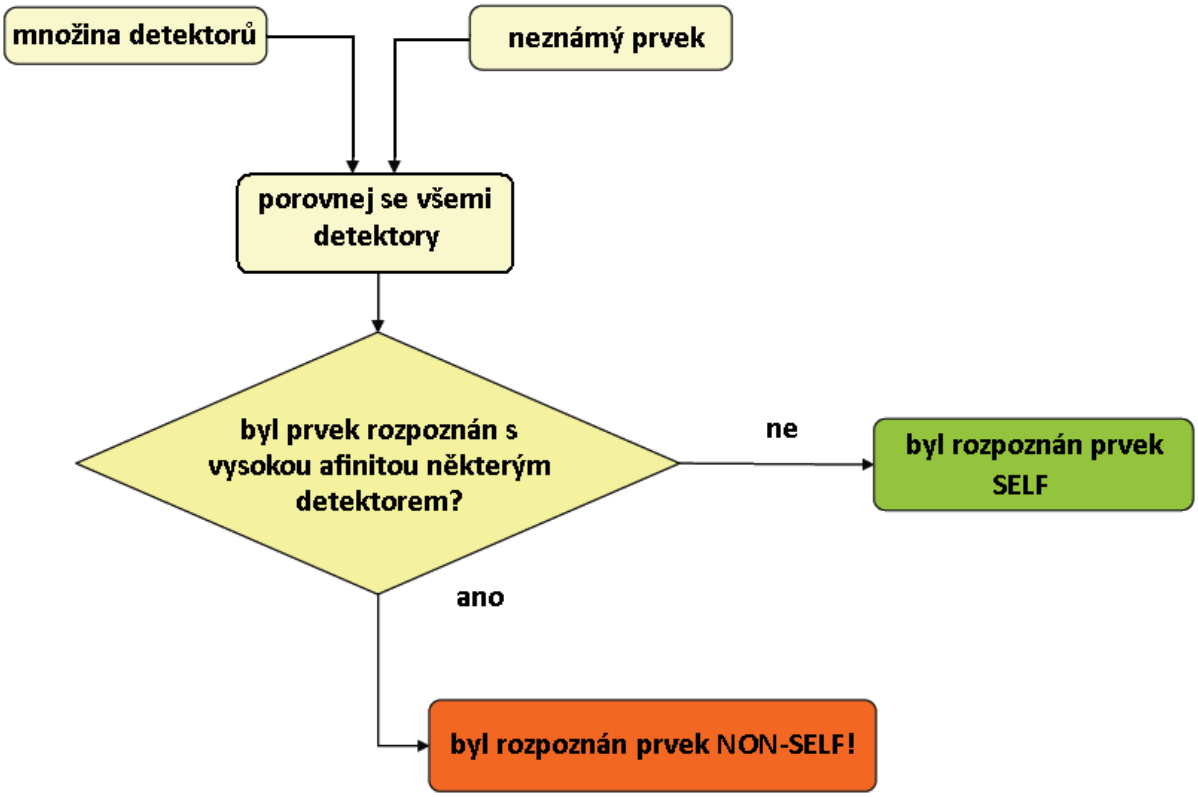
Obrázek 1 - Algoritmus pozitivní selekce

Afinita určuje podobnost, v našem případě úspěšnost (schopnost) rozeznání určitého prvku. Afinitu můžeme určit pomocí několika metod, např. Hammingovou vzdáleností.

2.3.1 Algoritmus pozitivní selekce

Princip pozitivní selekce se v imunitním systému využíval k odstranění zbytečných a neužitečných lymfocytů, které neměly žádné receptory nebo je měly nějakým způsobem poškozeny. Ve výsledku lymfocyty, které prošly pozitivní selekcí, byly ušetřeny zániku a mohly být použity v efektivní obraně systému jedince.

Algoritmus pozitivní selekce (obr. 1) slouží k odstranění prvků z množiny P, které nedokážou rozpoznat žádný z vlastních SELF prvků. Tento algoritmus uplatníme tam, kde vyžadujeme, aby množina řešení (detektorů) obsahovala jen detektory, které dokážou poznat prvky z množiny SELF. Například v případech, kdy množina P je mnohonásobně větší než množina SELF.



Obrázek 2 - Algoritmus negativní selekce

2.3.2 Algoritmus negativní selekce

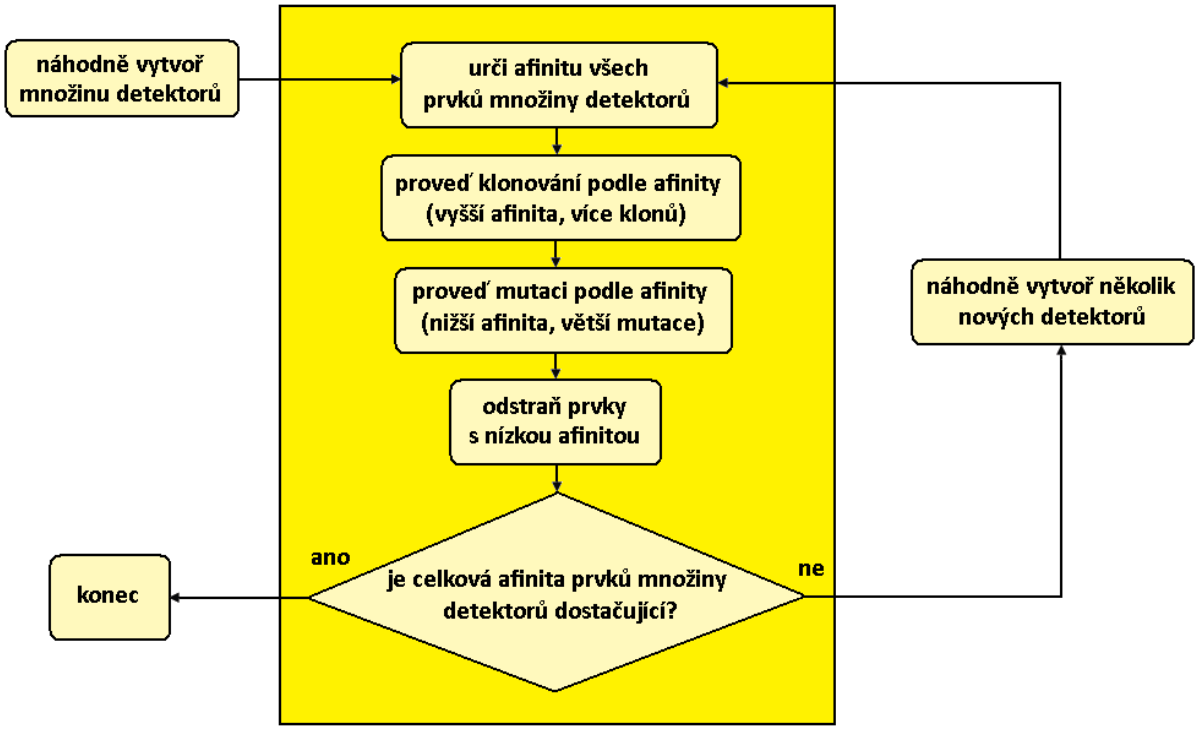
Princip negativní selekce se v biologických imunitních systémech aplikuje na lymfocyty T. Potom, co jsou lymfocyty T vytvořeny v kostní dřeni, cestují krevním řečištěm do brzlíku. Cestou je na ně aplikován algoritmus pozitivní selekce (jsou zachovány jen ty lymfocyty T, které mají v pořádku receptory a budou schopny dále fungovat – zabíjet buňky, na které se naváží). Když dorazí do brzlíku, dozrávají a je na ně aplikován algoritmus negativní selekce. Z celé populace lymfocytů T zůstanou pouze ty, které jsou schopny navázat se na cizí buňku (odstraní se takové lymfocyty T, které se

mohou navázat na prvky množiny SELF). Lymfocyty T se budou vázat jen na buňky, které nepatří do množiny SELF, a budou je zabíjet.

Algoritmus negativní selekce (obr. 2) slouží k vybrání těch řešení (detektorů), která jsou schopna rozpoznat pouze cizí prvky (non-SELF prvky). Tento algoritmus tedy odstraňuje ty prvky (ta řešení), které poznají SELF prvek. Používají se tam, kde je množina SELF daleko větší než její doplněk.

2.3.3 Klonální selekční algoritmus

Klonální selekční princip je způsob, jakým se lymfocyty T a B vypořádají s patogenem, pokud se na nějaký navážou. V případě, že se lymfocyty navážou na patogen, nastartuje se jejich velice aktivní metabolismus a začnou se rychle množit. Na nových lymfocytech se tvoří komplementy právě pomocí klonálního selekčního principu. V konečném důsledku potom, co se lymfocyt naváže na patogen a rozmnoží se, je vytvořena armáda lymfocytů, které mají schopnost rozpoznat tento určitý patogen a jeho nejbližší příbuzné (podobné) patogeny.



Obrázek 3 – Klonální selekční algoritmus

Klonální selekční algoritmus (nebo klonální selekční princip) je teorie, která popisuje, jakým způsobem vznikají řešení (detektory), která jsou schopna efektivně reagovat na určitý problém nebo množinu problémů. Prezentuje myšlenku, ve které jsou vytvářeny a generovány vlastně jen ta řešení, která eliminují a rozpoznají daný problém (problémy), a řešení, jejichž rozpoznávací schopnosti jsou menší, jsou eliminovány. Tento algoritmus nám pomáhá udržet množinu možných řešení co

nejefektivnější a nejpřesnější. V praxi můžeme jako problém považovat např. hledání podobných slov ke slovu viagra (např. v1agra, vi4gra, v1agr4,…).

3 Spamové filtry

Tato kapitola popisuje, jak fungují dnešní spamové filtry. Protože rozesílatele spamu používají moderní nástroje, je nutné, aby spamové filtry byly alespoň na stejné nebo vyšší technologické úrovni. Filtrování spamu můžeme rozdělit do několika kategorií – poštu můžeme kontrolovat na základě jejího původu, na základě formátu zprávy nebo na základě obsahu zprávy [1].

3.1 Filtrace podle původu

Tento způsob filtrace spamu využívá ke klasifikaci e-mailu IP adresu odesílatele. Patří mezi nejjednodušší na implementaci a není náročný na výpočetní prostředky. Když odesílatel zprávy naváže připojení na transportní a síťové vrstvě, příjemce zprávy okamžitě zná IP adresu odesílatele. Tuto informaci může na základě několika pravidel vyhodnotit a podle výsledku může spojení okamžitě přerušit nebo e-mail přijmout, označit a předat k dalšímu zpracování.

Mezi nevýhody tohoto filtrování patří případ, kdy odesílatel zfalšuje svou IP adresu. Falšování IP adres ale není jednoduchý proces, je třeba vynaložit určité úsilí. Jak popisuje [7] a [8], síť se dá jednoduše proti falšování IP adres chránit.

V následujících podkapitolách jsou popisovány metody využívající dynamické seznamy IP adres nebo kontrolu DNS (Domain Name System) záznamů. K filtraci podle původu zpráv nicméně můžeme ještě zařadit filtraci podle země původu. Pokud tedy nemáme žádné přátelé nebo obchodní partnery například v Číně, může zablokovat celý blok IP adres, který se používá k přístupu na internet v této zemi.

3.1.1 Blacklisting

Některé organizace, nejčastěji poskytovatelé internetového připojení (ISP), vytvářejí seznamy (tzv. *Blacklisty*) IP adres nebo rozsahů IP adres, které jsou používány k rozesílání nevyžádaných e-mailů. Několik organizací poskytují tyto seznamy široké veřejnosti na internetu, ostatní je používají jen pro vlastní potřeby.

Spamové filtry mohou při klasifikaci e-mailů využívat služeb těchto seznamů. Pokud najdou IP adresu odesílatele na seznamu, mohou e-mail rovnou zahodit nebo označit jako spam a předat k dalšímu zpracování. Blacklisty nemohou být uloženy offline přímo ve spamovém filtru, protože se údaje v nich často mění a musely by se aktualizovat a také proto, že by takový seznam byl vzhledem k množství IP adres velmi rozsáhlý. Dotazování do Blacklistu probíhá nejčastěji pomocí služby DNS [9].

Pokud se spamový filtr bude dotazovat např. Blacklistu *spamhaus.org*, vytvoří dotaz na DNS záznam typu A ve tvaru **14.176.229.147.sbl.spamhaus.org**, kde 147.229.176.14 je IP adresa serveru

(pozpátku), na který se ptá. Pokud tato IP adresa není na Blacklistu serveru *spamhaus.org*, odpověď nepřijde. Pokud tato IP adresa je na Blacklistu uvedena, server *spamhaus.org* pošle odpověď a v určitých případech v této odpovědi specifikuje, proč je IP adresa na Blacklist umístěna. Mezi nejznámější poskytovatele Blacklistu na internetu patří *SORBS (Spam and Open-Relay Blocking System)*, jehož bližší informace a specifikace naleznete zde: http://www.au.sorbs.net/

Za zmínku stojí nástroj na internetu, který projde nejznámější Blacklisty a ověří, zda se tam zadaná IP adresa nachází nebo ne. Odkaz na tento nástroj najdete v příloze č. 2.

3.1.2 Whitelisting

Whitelisting je oproti tomu metoda, která spočívá ve vytvoření seznamu bezpečných IP adres. Tedy IP adres důvěryhodných poštovních serverů, u kterých neočekáváme, že by mohly rozesílat spam.

Spamový filtr se pokusí IP adresu odesílatele vyhledat na Whitelistu a pokud ji nalezne, označí e-mailovou zprávu jako korektní a předá k dalšímu zpracování.

3.1.3 Greylisting

Greylisting je metoda, která pracuje na stejném principu, ale funguje dynamicky. Graylist obsahuje záznamy, které se dynamicky a podle určitých pravidel mění. Každý záznam je vlastně trojice: IP adresa, odesílatel a příjemce.

Tato metoda využívá protokolu SMTP, resp. chybových kódů tohoto protokolu. Pokud došlo při přenosu zprávy k neodstranitelné, trvalé chybě, a e-mail určitě nebude doručen, server odpoví chybou začínající číslem 5. Pokud došlo k chybě, která je ale dočasná (např. zdržení e-mailu z důvodu přetížení linky, vytížení serveru,…), server odpoví chybou začínající číslem 4. Odesílatel, který obdržel kód dočasné chyby, by si měl zprávu uložit do své fronty zpráv a po několika minutách se pokusit zprávu znovu odeslat. Protože rozesílatelé spamu posílají několik desítek tisíc zpráv, nepoužívají opakované pokusy o odeslání a zprávu, kterou nelze odeslat (i když jen kvůli dočasné chybě) zahodí a znovu neposílají.

Spamový filtr využívající tuto metodu zkontroluje při příchodu zprávy svůj Greylist (seznam trojic IP adres, odesílatelů a příjemců) a pokud danou trojici najde, zprávu přijme. Pokud tam tuto trojici ale nenajde, odpoví na tuto zprávu dočasnou chybou, uloží si trojici IP adresy, odesílatele a příjemce do svého seznamu a počká, pokud mu tato pozdržená zpráva přijde po určité době znovu. Pokud ano, zkontroluje Greylist, jestli se tam daný záznam nachází a pokud ano, zprávu přijme. Po určité době se záznamy z Greylistu odstraňují. K odstranění záznamu dojde také v případě, pokud se odesílatel pozdrženou zprávu nepokusí odeslat znovu.

Nevýhoda Greylistingu spočívá v tom, že jsou některé zprávy zdržovány. Tuto nevýhodu ale můžeme částečně eliminovat vhodným nastavením doby životnosti záznamů v Greylistu.

3.1.4 PTR a reverzní DNS záznamy

Pomocí PTR záznamu v DNS serveru se dá snadno zjistit, do jaké sítě nebo rozsahu IP adresa odesílajícího serveru náleží. Protože jsou korektní e-mailové servery k internetu připojeny pevným spojením s neměnnou IP adresou, mohou spamové filtry tyto kontroly provádět a pokud zjistí, že e-mail přichází z IP adresy, která patří uživateli připojující se k internetu přes vytáčené nebo dynamické spojení, mohou tento e-mail zamítnout.

Další kontrola spočívá v ověření reverzního DNS záznamu. Pokud chceme odeslat e-mail například do domény *@stud.fit.vutbr.cz*, musíme si u DNS serveru zjistit MX záznam pro tuto doménu. Po dotazování DNS serveru zjistíme IP adresu serveru, který přijímá a zpracovává poštu pro danou doménu (v případě *@stud.fit.vutbr.cz* je to *147.229.176.14*) a tomuto serveru můžeme e-mail odeslat. Pokud nám tedy přichází e-mail z adresy této domény (např. xneuwi00@stud.fit.vutbr.cz), provedeme dotaz na reverzní DNS záznam a pokud je výsledek jiný než IP adresa, z které nám e-mail přichází, tak tento e-mail jako spamový filtr zahodíme.

3.2 Filtrace na základě formátu zprávy

Tento způsob filtrace pracuje na základě formátu e-mailových zpráv. Mohli bychom ho rozdělit do dvou úrovní. První z nich je kontrola na úrovni protokolu *SMTP (Simple Mail Transfer Protocol)* a druhá úroveň pracuje nad protokolem *MIME (Multipurpose Internet Mail Extensions)* [18] [19]. První zmiňovaný protokol používají e-mailové servery při přenosu zpráv, zatímco druhý protokol používají e-mailoví klienti při vytváření zpráv.

V této kapitole se zároveň podíváme na dva komerční produkty a jak tento způsob filtrace využívají. První z nich je produkt *MailEssentials*[2] a druhý se jmenuje *Inteligent Message Filter (IMF)*[3].

3.2.1 Filtrace na úrovni protokolu SMTP

Mezi informace, které by nám měly usnadnit rozhodování, zda je určitá e-mailová zpráva spam nebo ne, patří především dvě pole tohoto protokolu.

Prvním z nich je adresa odesílatele e-mailu (pole FROM). Program Inteligent Message Filter je schopný zkontrolovat toto pole a na základě jeho hodnoty provést další akci. Můžeme nastavit seznam odesílatelů, které chceme blokovat, nebo můžeme blokovat všechny e-maily, které mají toto pole prázdné.

[2] GFI MailEssentials – spamový filtr společnosti GFI. Poslední verze v době psaní tohoto dokumentu byla verze 14. Více informací na http://www.gfi.com/mes/.

[3] Inteligent Message Filter – tvoří součást produktu Exchange Server společnosti Microsoft. V tomto dokumentu byla popisovaná verze 2003. Více informací na http://www.microsoft.com/exchange/default.mspx.

Druhé pole, které bychom při tvorbě spamového filtru měli brát v úvahu, je pole příjemce zprávy (pole RCPT TO). V tomto poli může být více e-mailových adres, klidně i několik stovek. Během SMTP komunikace běžný e-mailový server kontroluje příjemce a pokud některý příjemce neexistuje, okamžitě to oznámí odesílajícímu serveru. Této funkce ale rozesílatele spamu mohou zneužít, protože tak jednoduše zjistí, zda daná e-mailová adresa existuje nebo ne. Program *Inteligent Message Filter* je schopný omezit maximální počet příjemců a přijímat zprávy jen pro známé uživatele. Pokud tedy bude přicházet e-mail pro neznámou adresu, spojení může být okamžitě ukončeno.

Další z polí protokolu SMTP, které by se dalo využít pro filtraci spamu, je pole HELO/EHLO [20]. Po připojení odesílajícím server by se tento server měl tímto příkazem představit – měl by poslat příkaz ve tvaru *HELO/EHLO fully_qualified_domain_name* (např.: „HELO smtp.seznam.cz“) nebo *HELO/EHLO [ip address]*, kde ip adresa musí být ohraničená v hranatých závorkách (pro výše zmiňovaný server by tento příkaz měl tvar: „HELO [77.75.72.43]“). Spamový filtr může tuto hodnotu zkontrolovat a ověřit u DNS zda doménové jméno souhlasí s IP adresou odesílajícího serveru. Nicméně podle specifikace protokolu SMTP v RFC 5321 v sekci 4.1.4 stojí, že SMTP server může tuto hodnotu kontrolovat, ale nesmí e-mailovou zprávu zamítnout jen na základě této skutečnosti [21].

Někteří rozesílatele spamu se snaží co nejvíce ušetřit čas a šířku pásma připojení k internetu tím, že ignorují a nedodržují korektní ukončení SMTP spojení. SMTP spojení by se podle [21] mělo správně ukončit odesláním příkazu QUIT.

Pokud má e-mail více příjemců, SMTP protokol umožňuje vložit do jednoho TCP paketu více *RCPT TO:* polí. Tyto pole se spojí za sebe a umožní tím efektivnější využití šířky pásma. Toto zřetězení (neboli pipelining) nicméně v případě chybného příjemce omezí možnost odpovědět chybovou hláškou a protože se tím poruší specifikace SMTP protokolu, mohou spamové filtry takové spojení ukončit.

3.2.2 Filtrace na úrovni protokolu MIME

Filtrace na úrovni protokolu MIME je složitější než filtrace na úrovni protokolu SMTP. Protokol MIME je vlastně internetový standard, který popisuje e-mailovou zprávu. Dříve se používal jen jako rozšíření e-mailových zpráv pro podporu jiných znakových sad, ne-textových příloh a zpráv s více částmi. Dnes je obecně považován za standardní formát e-mailových zpráv, který podporuje drtivá většina e-mailových klientů [19].

V hlavičce zprávy se nachází několik polí, které jsou z hlediska filtrace spamu důležité. Patří mezi ně pole *From* – e-mailová zpráva může toto pole obsahovat prázdné, poškozené nebo adresa v tomto poli může být jiná než v poli FROM protokolu SMTP. V těchto případech můžeme zprávu označit jako spam.

Další pole se jmenuje *content-type*. V tomto poli je specifikován typ dat přenášených ve zprávě. Textové zprávy mají v tomto poli hodnotu *plain/text*, ale pokud je ve zprávě např. příloha nebo obrázek, je tato hodnota jiná. Program MailEssentials dokáže zkontrolovat poměr mezi textovou části e-mailu a obrázkovou části e-mailu. Spamy často neobsahují text, ale jen obrázek, na kterém je umístěn text zprávy. Pomocí tohoto pole tedy dokáže určit, zda se jedná o spam nebo ne.

Další pole se jmenují *content-type: charset* a *content-transfer-encoding*. V tomto poli je uvedeno kódování textu zprávy. Program MailEssentials je díky tomuto poli schopen filtrovat všechny zprávy, které jsou psány např. rusky nebo čínsky.

Protokol MIME obsahuje ještě další pole, nicméně tři výše zmíněné jsou pro filtraci spamu nejpodstatnější.

3.3 Filtrace podle obsahu

Automatická filtrace podle obsahu e-mailové zprávy nemůže z principu fungovat dokonale, protože to, jestli je zpráva korektní nebo se jedná o spam, nelze objektivně posoudit. Pro někoho může být e-mail s upozorněním na slevy v oblíbeném obchodě zajímavý, ale někdo jiný takovýto e-mail rovnou vymaže. I přesto filtrování podle obsahu poskytuje dobré výsledky a je využíváno nejčastěji. Existují dvě hlavní metody, které se často kombinují.

3.3.1 Pravidla

Během rozpoznávání spamu pomocí pravidel dochází ke kontrole určitých rysů e-mailů, které jsou pro spam typické. Mezi tyto rysy patří například kontrola časových údajů (doba mezi odesláním a přijetím e-mailu nebo zda-li byl e-mail odeslán v pracovní době nebo v noci), kontrola předmětu (je uveden, obsahuje jen velká písmena,…) a další [10].

Každému rysu je přiřazena váha a po zkontrolování celého e-mailu se tyto hodnoty sečtou. Pokud je výsledek větší než určitá hranice, e-mail se označí jako spam.

3.3.2 Učení

Filtry založené na učení (často označované jako Bayesovské filtry) využívají prvky umělé inteligence. Spamové filtry využívající tuto metodu pracují ve dvou režimech. V režimu učení a v režimu rozpoznávání. V režimu učení jsou filtru předkládány e-maily, které jsou spam, a e-maily, které jsou uživatelem označeny jako korektní. Filtr každý takovýto e-mail rozdělí na fragmenty (často na slova) a každému fragmentu se statisticky přidělí pravděpodobnost, že e-mail obsahující tento fragment je spam. V režimu rozpoznávání se pak použijí nashromážděné informace, na jejichž základě se určí pravděpodobnost, že se jedná o spam. Pro výpočet pravděpodobnosti se nejčastěji používá vzorec, který navrhl matematik Bayes [11].

Přestože tyto filtry vyžadují, aby byly nejprve spuštěny individuálně uživatelem v režimu učení, jsou často využívány i na serverech [1].

3.4 Další metody v boji proti spamu

Kromě výše zmíněných tří možností filtrace e-mailových zpráv existuje ještě několik dalších způsobů, jak přijímání spamu zamezit nebo alespoň částečně omezit.

Spoustu moderních e-mailových klientů dokáže zobrazit e-mailovou zprávu v HTML formátu. Tento formát oproti běžnému textovému zobrazení skýtá spoustu zajímavých vylepšení, ať už v oblasti formátování zprávy nebo zobrazení obrázků a dalších grafických efektů. S těmito vylepšeními ale přichází také riziko, které bývá rozesílateli spamu často zneužíváno. Pokud rozesílatelé spamu vloží do zprávy specifický odkaz na obrázek, který se nachází na internetu, a příjemce tento e-mail jen zobrazí, obrázek se stáhne a rozesílatel ví, že e-mailová adresa příjemce je aktivní. Tuto informaci rozesílateli spamu mohou získat i jinými způsoby, ne jen vložením odkazu na obrázek. Stačí třeba vložit odkaz s textem: „Pokud nechcete dostávat reklamní e-maily, odhlaste se kliknutím na tento odkaz.“ Po kliknutí na odkaz, místo aby se uživatel odhlásil, začne dostávat daleko více spamu. Doporučení z [20] tedy říká: „Nepoužívat HTML formát zprávy“. Nebo používat ale jen u důvěryhodných odesílatelů.

Jedna z ne častých technik na boj proti spamu se nazývá *spamtrapping*. Tato technika spočívá v umístění určité e-mailové adresy na webové stránky společnosti způsobem, že si ji běžný návštěvník nevšimne, ale spamoví roboti, kteří procházejí internet za účelem sběru e-mailových adres, ji najdou. Spamový filtr ve společnosti se pak nastaví tak, aby s příchodem e-mailové zprávy na tuto adresu, zařadil odesílatele na blacklist a tím zamezil vstupu spamu do korektních e-mailových schránek uživatelů.

Tarpit networking je metoda, kterou bývá označována uměle zpomalovaná síťová komunikace, která se dá relativně efektivně využít i v boji proti spamu. Pokud e-mailový server vloží sekundové prodlevy po každém příkaze v SMTP komunikaci, korektní e-maily sice dorazí třeba s pětisekundovým zpožděním, ale pro spamové roboty je to tak velká časová ztráta, že spojení sami raději ukončí.

Další z metod, která se dá využít v boji proti spamu jsou tzv. *ham*[4] *hesla*. Na webových stránkách, kde je zobrazena e-mailová adresa (např. technické podpory nějakého výrobku), je v tomto případě také zobrazen určitý návod nebo postup, jak úspěšně odeslat e-mail na tuto adresu. Často to bývá formou vložení hesla do předmětu zprávy nebo na konec zprávy. Pokud je odesílatel pro

[4] Pojmem ham bývá často označován korektní e-mail (protiklad ke spamu), který vyplývá z historického hlediska této problematiky.

spamový filtr neznámý, zkontroluje, zda e-mail obsahuje toto heslo a pokud ne, označí e-mail za spam.

Žádná z těchto metod se nepoužívá pro rozpoznání spamu samostatně, ale vždy v kombinaci s jinými technikami.

4 Využití UIS pro filtraci spamu

V této kapitole se podíváme na konkrétní prvky a součásti biologického imunitního systému a na jejich mapování na umělý imunitní systém pro filtraci spamu.

Biologický imunitní systém rozlišuje mezi vlastními (*SELF*) a cizími (*non-SELF*) prvky, zatímco imunitní systém použitý pro filtraci spamu musí rozlišovat jako vlastní korektní e-maily a jako cizí bude považovat nevyžádanou poštu. V tomto ohledu má biologický imunitní systém výhodu, protože prvky množiny SELF se v průběhu času nemění tak často. V případě imunitního systému pro detekci spamu se množina SELF bude měnit daleko častěji – jak již bylo uvedeno dříve, uživatel si posílá e-maily s novými lidmi, má jiné záliby (píše si o jiných věcech) nebo se může naučit nový jazyk a psát e-maily v tomto jazyce. S tím musí umělý imunitní systém počítat.

4.1 Vrstvy UIS pro filtraci spamu

Stejně jako biologický imunitní systém bude i umělý imunitní systém pro filtraci spamu pracovat ve vrstvách. Ještě před těmito vrstvami bychom měli zmínit a určitě neměli opomenout jako jednu z vrstev aktuální zákony, které problematiku spamu upravují [1].

V této práci se ale budeme zabývat především technickým řešením, kde budou také obě vrstvy postupně provázány mezi sebou.

4.1.1 První vrstva

Na vrstvu vrozené imunity biologického imunitního systému bychom mohli namapovat principy jako jsou filtrace podle původu. Tedy metody jako jsou Blacklisting nebo Greylisting (viz kap. 3.1.1 a 3.1.3).

4.1.2 Druhá vrstva

Adaptivní vrstva biologického imunitního systému spoléhá především na lymfocyty. Ve druhé vrstvě umělého imunitního systému pro filtraci spamu by měl existovat model umělého lymfocytu. Tento umělý lymfocyt by měl nést právě jeden vzor (antigen) nevyžádané pošty (např. klíčové slovo, zdrojovou IP adresu,…) nebo několik vzorů (více klíčových slov, rozsah IP adres).

4.1.2.1 Databáze antigenů a lymfocytů

V případě filtrace spamu mohou být za antigen považovány nejen fragmenty e-mailových zpráv jako jsou slova, fráze nebo např. úseky HTML kódu ve zprávě, ale také hodnoty polí z hlavičky zprávy.

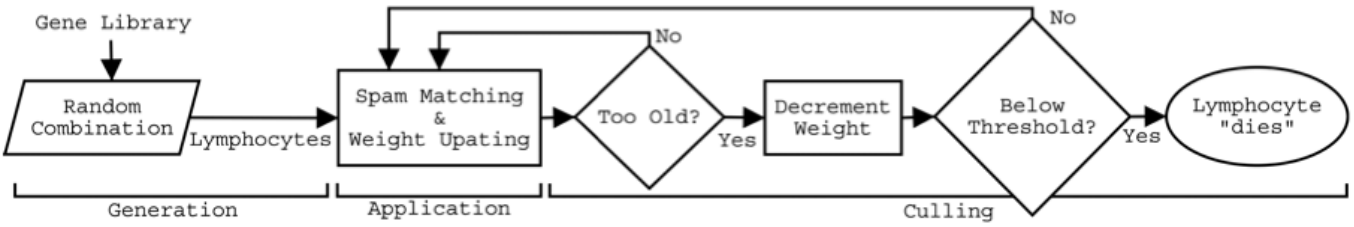
Databáze lymfocytů obsahuje umělé lymfocyty, kde každý z nich obsahuje právě jeden antigen. Vzhledem k rychlosti rozpoznávání nevyžádané pošty by databáze lymfocytů měla být co nejmenší.

Tvorba lymfocytů z databáze antigenů by měla probíhat podobně jako v biologickém imunitním systému. Tedy za pomocí klonálního selekčního algoritmu. Zde bych chtěl ale upozornit, že prvky množiny P (řešení) nebudou samotné antigeny, ale budou to zástupci skupiny antigenů, které se vztahují na konkrétní nevyžádané e-mailové zprávy.

4.1.2.2 Umělý lymfocyt

Podle [16] by měl dále každý lymfocyt obsahovat dvě hodnoty. První z nich (*spam_matched*) je číslo, které udává relativní počet nevyžádaných e-mailů, které byly tímto lymfocytem rozpoznány. To znamená kolik procent z nevyžádaných e-mailů tento lymfocyt označil jako spam. Druhé číslo (*msg_matched*) udává relativní počet všech e-mailů, které byly tímto lymfocytem rozpoznány. To znamená kolik procent ze všech zpráv tento lymfocyt označil jako spam.

Během chodu systému se lymfocyty průběžně učí na základě výsledků analýzy e-mailů, které procházejí systémem. Lymfocyty se postupně upravují, zastarávají a nové se vytvářejí. Životní cyklus umělého lymfocytu vidíte na obrázku č. 1 (převzato z [16]).



Obrázek 4 - Životní cyklus umělého lymfocytu

V první fázi je lymfocyt vygenerován a zařazen do databáze lymfocytů. V druhé fázi dochází k pokusu rozpoznání dané e-mailové zprávy tímto lymfocytem. Pokud je e-mail rozpoznán, je zvýšena hodnota *msq_matched* a pokud je e-mail po výstupu ze systému označen jako spam, upraví se hodnota *spam_matched* tohoto lymfocytu. Ve třetí fázi dojde k testu stáří lymfocytu a úpravě hodnoty *msq_matched*. V druhém větvení v tomto schématu dochází ke kontrole hodnoty *msq_matched* a pokud je tato hodnota pod určitou hranicí, lymfocyt se odstraní. Pokud bychom hranici stanovili například na 5%, tento proces rozhodování by vlastně odstranil všechny lymfocyty, které jsou již zastaralé a které jsou schopné se navázat na méně než 5% všech e-mailů procházejících systémem.

Touto cestou je umělý imunitní systém pro filtraci spamu schopen učit se a zapomínat informace během samotného běhu systému [16].

4.2 Návrh procesu filtrace

Když se narodí nový jedinec, jeho imunitní systém se spoléhá hlavně na vrstvu vrozené imunity. Tato vrstva je postupně vystavována cizím prvků a ničí nebezpečné patogeny. Na základě těchto procesů

vzniká postupem času druhá vrstva, adaptivní imunita. Adaptivní imunita se učí z výsledků vrozené imunity, zapamatovává si nebezpečné patogeny a postupně se zdokonaluje.

Stejně tak imunitní systém pro filtraci spamu bude na začátku svého životního cyklu spoléhat na první vrstvu. Druhá vrstva se bude na základě výsledků první vrstvy učit a postupně zdokonalovat. V určitých případech může druhou vrstvu učit také samotný uživatel, který explicitně označí určité e-maily jako spam.

Pokud systém již bude nějakou dobu pracovat a druhá vrstva bude obsahovat dostatečný počet umělých lymfocytů, rozhodnutí, zda daná e-mailová zpráva je nebo není spam bude záviset na dvou mezivýsledcích. Na výsledku analýzy první vrstvy a na výsledku analýzy druhé vrstvy umělého imunitního systému. Jaké váhy by měly mít jednotlivé mezivýsledky, aby systém pracoval dostatečně efektivně ale co nejspolehlivěji, ukážou až konkrétní testy nad implementací spamového filtru.

5 Realizace spamového filtru

V této kapitole se nachází popis a postup realizace spamového filtru na bázi umělého imunitního systému, který byl v rámci diplomové práce vytvořen. Před samotným návrhem a programováním spamového filtru bylo potřeba podrobně nastudovat prostředí, do kterého měl být spamový filtr zasazen. V kapitole 5.1 se tedy podíváme na principy přenosu a doručování zpráv v prostředí serveru *Microsoft Windows 2003 Server*. Pro vývoj aplikace byly využity převážně nástroje z dílny Microsoftu. Jakým způsobem byla tato aplikace vyvíjena a jak se integrovala do systému, je popsáno v kapitole 5.2. Dále si popíšeme funkcionalitu spamového filtru, jeho částí a jakým způsobem můžeme nastavit a ovlivnit chování programu. V kapitole 5.5 je detailně popsána implementace spamového filtru a na závěr této kapitoly několik postřehů a návrhu na zlepšení.

5.1 Microsoft Windows 2003 Server a poštovní služby

Spamový filtr se všemi jeho součástmi byl vyvíjen v prostředí operačního systému *Microsoft Windows 2003 Server*. Tento produkt není k dispozici volně ke stažení, nicméně jako student Fakulty Informačních technologií na VUT v Brně mám prostřednictvím služby MSDN AA[5] k dispozici jednu licenci.

Mezi hlavní důvody výběru tohoto operačního systému patří fakt, že se systémy rodiny produktů *Microsoft Windows Server* často využívají v podnikových sítích a jsou tudíž hodně rozšířeny. Z pohledu vývoje spamového filtru byl konkrétně produkt *Microsoft Windows 2003 Server* vybrán z důvodu širokých možností správy uživatelů a jejich poštovních schránek, možnosti konfigurace SMTP poštovního serveru, ale také z důvodu stability, bezpečnosti a relativně dobře přístupné dokumentace. K tomuto produktu existuje řada materiálů a dokumentace, která obsahuje spoustu užitečných informací a postupů, jak tento server nakonfigurovat, aby přesně splňoval všechny naše požadavky. Mezi nejužitečnější funkce systému, které byly v průběhu vývoje a testování spamového filtru využity, patří flexibilní nastavení virtuálních SMTP serverů, provázanost e-mailových schránek s účty uživatelů systému nebo výběr a doručení e-mailových zpráv z konkrétní složky na disku. Poslední funkce spočívá v tom, že se do určité složky na disku uloží soubory jednotlivých e-mailových zpráv ve formátu EML[6] a SMTP server si tyto zprávy sám vyzvedne a zpracuje – odešle na příjemce uvedené v souboru.

[5] MSDN Academic Alliance: http://msdn61.e-academy.com/elms/Storefront/Home.aspx?campus=vut_fit

[6] EML – Formát souboru obsahující e-mailovou zprávu. Podporovaný například klienty MS Outlook Express, Windows Mail nebo Mozzila Thunderbird.

5.1.1 Příprava prostředí

Abychom mohli testovat odesílání a přijímání e-mailových zpráv a funkčnost spamového filtru, bylo potřeba vytvořit několik uživatelských účtů a jejich schránek. Uživatelé byli vytvořeni na serveru a k odesílání a přijímání pošty byl použit klient *Microsoft Outlook Express*. Byly vytvoření tito uživatelé:

- Uživatel *Petr*, heslo *password#1*, e-mailová adresa petr@dip.local.
- Uživatel *Karel*, heslo *password#1*, e-mailová adresa karel@dip.local.
- Uživatel *Josef*, heslo *password#1*, e-mailová adresa josef@dip.local.
- Uživatel *Koš*, heslo *password#1*, e-mailová adresa kos@dip.local.

Poslední uživatel (*Koš*) slouží jako e-mailový doménový koš, do jehož schránky se doručují všechny e-maily, které mají neznámého příjemce.

5.1.2 SMTP služba

Spamový filtr byl implementován pro *SMTP službu* (anglicky SMTP Service), která tvoří součást komplexního balíku služeb *Internet Information Services 6.0* (IIS). Služby balíku IIS jsou volitelně doinstalovatelné součásti systému Microsoft Windows 2003 Server.

Po instalaci se služba automaticky spustí, ale před použitím se musí nakonfigurovat[7]. Pro potřeby implementace spamového filtru stačilo nakonfigurovat dva virtuální SMTP servery a doménu, pro kterou má tento server zpracovávat příchozí poštu:

- *Hlavní SMTP server* – k němu se připojovali testovací uživatelé a odesílali přes něj poštu. Nakonfigurovaný byl k poslechu na lokálních IP adresách všech síťových rozhraní na standardním portu 25.
- *Testovací SMTP server* – sloužil pro zátěžové testy a hromadné odesílání mailů, hlavně metodou vyzvednutí poštovních zpráv ze složky na disku. Naslouchal také na všech IP adresách ale na nestandardním portu 26.

Poštovní doména pro tento systém byla nakonfigurována na *@dip.local*.

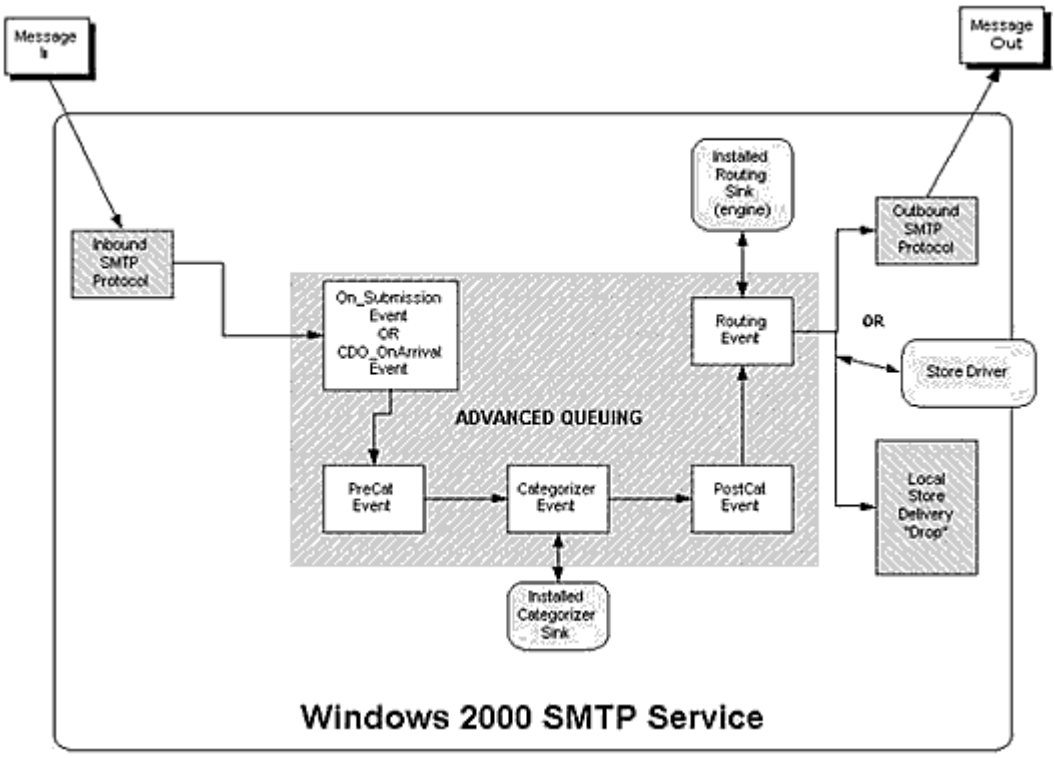
5.1.3 Doručování pošty

K tomu, abychom mohli vytvořit funkční spamový filtr, museli jsme dále dobře nastudovat a pochopit principy doručování e-mailových zpráv v prostředí systému *Microsoft Windows 2003 Server*. Dalším důležitým krokem bylo tedy nastudování způsobů doručování pošty v tomto prostředí [22].

Architektura Windows 2003 serveru a jeho SMTP služeb umožňuje jednoduché rozšíření a kromě vytvoření spamového filtru nabízí například možnosti snadné implementace antivirového

[7] Podrobný manuál k IIS a SMTP službě naleznete např. zde:
http://www.microsoft.com/windowsserver2003/iis/default.mspx

filtru, skenování a kontroly procházejících e-mailů na klíčová slova, přidání firemního razítka do e-mailů opouštějící naši společnost nebo například archivování veškeré e-mailové komunikace.



Obrázek 5 - Zpracování pošty [22]

Virtuální SMTP server může zařadit ke zpracování poštu přijatou standardním způsobem, tedy ze sítě naslouchánim na TCP portu číslo 25, nebo přečtením souborů e-mailových zpráv ze složky na disku. Potom, co e-mailová zpráva dorazí do SMTP serveru, je zpracovávána rozšířeným systémem zařazování (*advanced queuing system*). Tento systém během zpracování e-mailové zprávy volá různé funkce a formou odchytávání událostí (*sink události*) můžeme tyto funkce nahradit. Jak ilustruje obrázek č. 5, v prostředí IIS 6.0 existuje několik takových *sink událostí* a každá z nich je vyvolána v určitém okamžiku zpracování e-mailové zprávy. Existují dvě kategorie sink událostí – protokolové události a transportní události.

Protokolové události – e-mailová zpráva je posílána po síti pomocí SMTP protokolu na server. V této fázi jsou k dispozici tři typy událostí – *In*, *Out* a *Server Response*. K tomu, aby byla spuštěna konkrétní funkce u konkrétní události, se u protokolových *sink událostí* definují pravidla – pokud chceme např. odchytit okamžik, kdy odesílající server zadává jednotlivé e-mailové adresy příjemců[8], vytvoříme pravidlo s názvem RCPT [22].

Transportní události – po přijetí e-mailové zprávy serverem se spustí transportní událost *OnSubmission/OnArrival*. Tyto události zkontrolují, jestli je přiřazena nějaká *sink událost* a pokud

[8] V SMTP komunikaci se příjemci zadávají příkazem „RCPT TO: <adresa@prijemce.cz>".

ano, spustí ji. *Sink události* mohou být spouštěny asynchronně, což znamená, že se nečeká na ukončení zpracování zprávy, ale může se začít zpracovávat zpráva následující. Na obrázku č. 5 je znázorněno další zpracování e-mailové zprávy při průchodu systémem. Zpráva může být dále zpracována například z důvody zjištění konkrétních příjemců z distribuční skupiny, z důvodu kontroly počtu příjemců nebo velikosti zprávy a je dále buď uložena do schránky uživatelů na serveru nebo odeslána na jiný SMTP server.

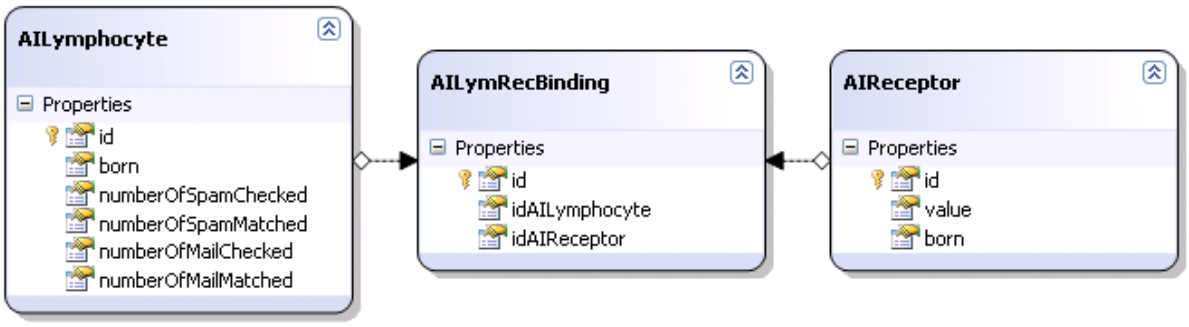
Pro realizaci spamového filtru je důležitá sink událost s názvem *OnArrival*. Tato událost se vyvolá vždy, když do systému vstoupí nová zpráva, a právě na tomto místě jsme do systému integrovali spamový filtr.

5.2 Použité nástroje a techniky

K vývoji spamového filtru na bázi umělého imunitního systému byl zvolen objektivně orientovaný jazyk *C#* ve verzi 3.0 v prostředí *.NET Frameworku 3.5*. Tento jazyk byl vybrán z důvodu širokých programátorských možností a technik, které tvorbu spamového filtru značně usnadňovaly. Mezi využité pokročilé programátorské techniky patří například generické kolekce, snadné ladění, optimalizace kódu nebo metoda přístupu k datům *linq*.

5.2.1 Jazyk C#

Kompilátor jazyka C# v prostředí .NET Frameworku vytváří tzv. řízený kód (managed code). Tento kód je vlastně určitým „mezikódem“ a není zpracováván přímo procesorem počítače, ale je interpretován a běží v prostředí .NET Frameworku[9]. S touto skutečností musíme počítat, protože spamový filtr budeme zabudovávat do prostředí, které běží v neřízeném kódu (unmanaged code). O této problematice více v kapitole 5.2.3.



Obrázek 6 – Modelování tříd v LINQ

Mezi nejzajímavější prvky tohoto jazyka a důvod k vybrání právě C# patří *linq*. *Linq* (anglicky Language Integrated Query) je nástroj nebo spíše jazyk, který je přímo integrován do prostředí jazyka

[9] Na stejném principu funguje např. Java v prostředí Java Virtual Machine.

C# a usnadňuje tvorbu, třídění a vyhledávání a vůbec jakoukoli další manipulaci s daty. Díky tomuto nástroji se podařilo ve spamovém filtru vytvořit abstrakci nad všemi daty způsobem, že se mohlo k obsahu e-mailových zpráv a k pravidlům v SQL databázi přistupovat jako by se jednalo o jeden typ dat. Pomoci tohoto nástroje bylo možno vytvořit třídy reprezentující jednotlivé tabulky v SQL databázi, pomoci nichž se k těmto tabulkám a datům v nich přistupovalo jako k běžným objektům v jazyce C#. Na obrázku č. 6 je zobrazena část datových tříd tohoto rozhraní.

5.2.2 Vývojové nástroje

Jako vývojové prostředí a nástroje byly použity volně dostupné nástroje společnosti Microsoft. Pro vytváření samotného spamového filtru byl použit editor *Microsoft Visual C# Developer 2008 Express Edition* a pro vytvoření webového uživatelského rozhraní spamového filtru to byl nástroj *Microsoft Visual Web Developer 2008 Express Edition*. Oba tyto nástroje jsou volně dostupné na stránkách Microsoftu a je možné je zdarma stáhnout[10].

Mezi další nástroje, které byly použity pro vývoj spamového filtru, patří *Microsoft SQL 2008 Server* také ve verzi *Express*. Tedy ve verzi, která je volně stažitelná z internetu. Tato verze SQL Serveru má oproti komerčně využívaným verzím omezení jen v podobě maximálního počtu souběžných připojení do databáze. Toto omezení ale vývoji spamového filtru nijak nebránilo, protože vyvíjenému filtru stačí do databáze přistupovat jen pomocí jednoho spojení. Použít jiný databázový server bychom museli v případě, kdybychom spamový filtr měli nainstalovaný na několika serverech a bylo by potřeba, aby jednotlivé spamové filtry sdíleli jednu společnou databázi.

5.2.3 Integrace do systému

Protože sink události jsou v systému implementovány pomocí neřízeného kódu a výsledek kompilátoru jazyka C# pracuje jako řízený kód v prostředí .NET Frameworku, museli jsme vytvořený spamový filtr korektně integrovat do systému. Samotné integraci předchází ještě řada operací, ve kterých je vytvořeno rozhraní mezi spamovým filtrem a SMTP službou, generování asembly klíčů a vytvoření referencí pro knihovny systému. Tento postup byl čerpán převážně z [23].

Samotná integrace do systému tedy probíhá v těchto krocích:

- **Registrace knihovny** – knihovna spamového filtru (*Spam1.dll*) se pomocí programu *RegAsm.exe*, který je součásti .NET Frameworku, zaregistruje do systému.
- **Připojení sink události** – pomocí skriptu *SMTPreg.vbs*, který je volně dostupný na stránkách Microsoftu[11], se funkce spamového filtru připojí k sink události.

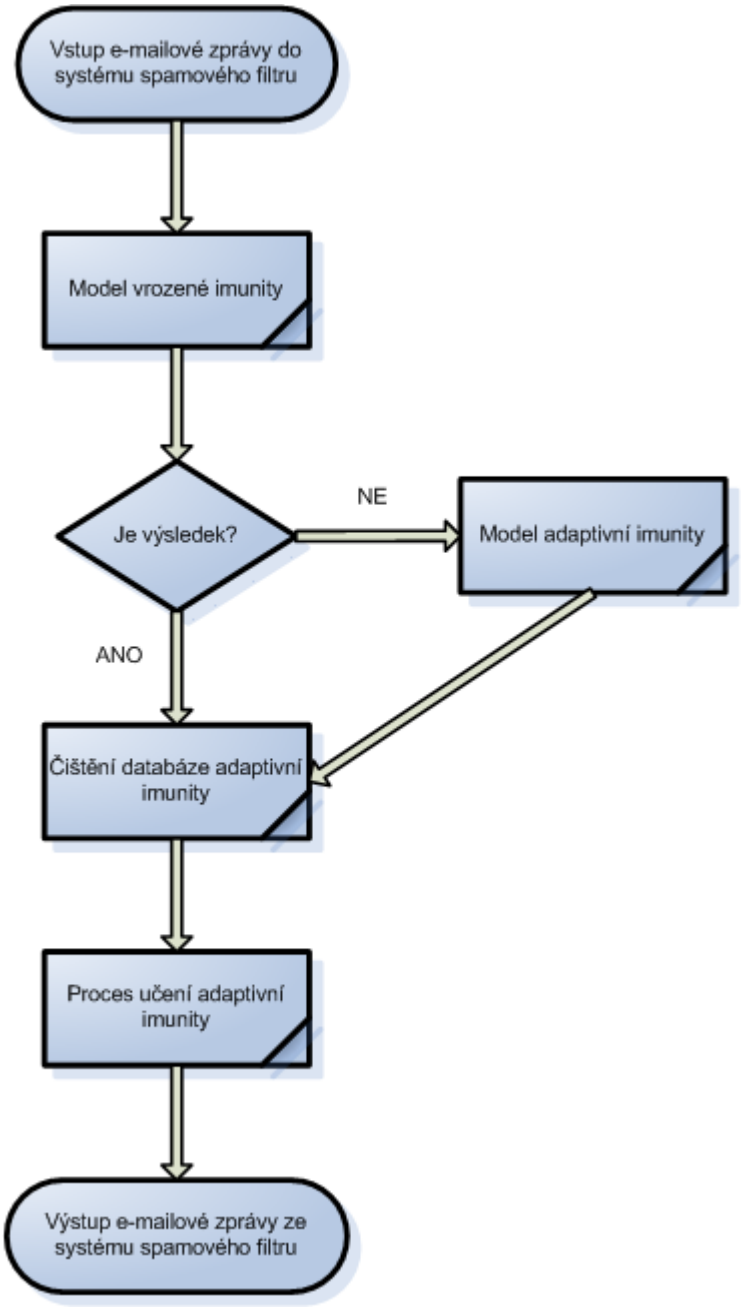
[10] Microsoft Visual Studio 2008 Express Edice: http://www.microsoft.com/Express/

[11] Event Management Script dostupný z WWW: http://msdn.microsoft.com/en-us/library/ms528023(EXCHG.10).aspx

Tyto dva kroky musí být provedeny při každé změně dll knihovny (kódu a kompilaci spamového filtru). Po korektní integraci spamového filtru do systému, se již po příchodu každého e-mailu spamový filtr aktivuje a spustí.

5.3 Funkčnost

V této podkapitole si detailně popíšeme samotnou funkčnost systému. Knihovna spamového filtru je rozdělena na dva logické celky – *model vrozeného imunitního systému* a *model adaptivního imunitního systému*. Tyto dvě části jsou spojeny aplikační logikou, kterou vystihuje obrázek č. 7.



Obrázek 7 - Aplikační logika spamového filtru

5.3.1 Model vrozeného imunitního systému

E-mailová zpráva vstoupí do systému spamového filtru a je zkontrolována *modelem vrozeného imunitního systému*. V této části spamového filtru dochází k testování e-mailové zprávy pomocí pevně nastavených pravidel, které reprezentují receptory biologického vrozeného imunitního systému. Tato pravidla může uživatel volně nastavovat a měnit, nicméně aplikace je vidí jako statickou a neměnnou databázi. Tato databáze se tedy za běhu systému automaticky nemění.

Pravidla

Pravidla mohou být několika typů. Každý typ odpovídá určitému poli nebo znaku e-mailové zprávy:

- **Předmět –** pole předmětu e-mailové zprávy.
- **Od –** pole odesílatele.
- **IP adresa –** IP adresa odesílatele zprávy.
- **Textová část** – textová část e-mailové zprávy:
  - **Tělo –** tělo zprávy. Pomocí tohoto pole můžeme zablokovat e-mailové zprávy, které budou obsahovat určitá klíčová slova.
  - **Znaková sada** – znaková sada e-mailové zprávy. Pravidlo tohoto typu nám umožní zablokovat všechny e-maily, které jsou například kódovány v čínštině.
  - **Kódování** – způsob kódování textu.
  - **Typ dat.**
- **HTML část –** pokud e-mail obsahuje HTML část, můžeme vytvářet tato pravidla:
  - **Tělo –** tělo zprávy.
  - **Znaková sada –** znaková sada.
  - **Kódování.**
  - **Typ.**

Každé pravidlo je vyhodnocováno jedním ze čtyř způsobů. Pole e-mailové zprávy může být testováno na shodu nebo neshodu s určitou hodnotou nebo testováno zda obsahuje nebo neobsahuje určitou hodnotu.

Pravidla modelu vrozeného imunitního systému jsou vyhodnocována postupně a pokud systém dojde k jednoznačnému závěru, předá tento výsledek nadřazené aplikační logice, která předá řízení funkci *čištění databáze modelu adaptivní imunity* (kap. 5.3.3). Pokud model vrozeného imunitního systému k jednoznačnému závěru nedojde (neexistují vhodné pravidla nebo si pravidla protiřečí), předá se tato skutečnost aplikační logice a ta předá řízení *modelu adaptivního imunitního systému.*

5.3.2 Model adaptivního imunitního systému

Tato část spamového filtru reprezentuje biologický adaptivní imunitní systém. Tento model již nepracuje s celou e-mailovou zprávou ale jen se slovy, které jsou obsaženy v textové nebo v HTML části e-mailu. V databázi tohoto modelu jsou obsaženy umělé receptory, kde každý z nich reprezentuje jedno slovo, a umělé lymfocyty s vazbami na libovolný počet receptorů. Kontrola, zda se jedná o spam, probíhá v těchto krocích:

- Z databáze modelu adaptivního imunitního systému se vyberou jen ty lymfocyty, které se už dokázaly navázat na určitý počet nevyžádaných e-mailových zpráv.
- Z tohoto výběru se provede další selekce a vyberou se jen ty umělé lymfocyty, od kterých se na e-mailovou zprávu navázalo určité procento receptorů (pokud má tedy lymfocyt 10 receptorů a hranice je nastavena na 25%, musí se navázat alespoň tři receptory).
- Provede se výpočet spamového skóre – tento výpočet se provede pomocí následujícího vzorce:

$$R = \frac{\sum r_s * r_m * p}{u}, \tag{3}$$

  kde proměnná r_s je hodnota *SpamRate* udávající kolik procent spamů umělý lymfocyt rozpoznal, proměnná r_m je hodnota *MailMatchedRate* udávající kolik procent všech e-mailů lymfocyt rozpoznal (navázal se na ně), proměnná p říká kolik procent receptorů daného lymfocytu se navázalo na e-mail a hodnota u je celkový počet úspěšných lymfocytů, které se na tento e-mail navázalo.
- V dalším kroku se provede uložení informací do historie a úprava data posledního použití lymfocytu.
- V posledním kroku se vyhodnotí výsledná hodnota R a pokud je nad určitou hranicí, je e-mail označen jako spam.

Pokud je e-mail modelem adaptivního imunitního systému označen jako spam, je tato skutečnost předána aplikační logikou modulu čištění databáze modelu adaptivního imunitního systému.

5.3.3 Čištění databáze modelu adaptivní imunity

Podle nastavení spamového filtru (viz kap. 5.4.1) se z databáze modelu adaptivní imunity odstraní v určitém rozsahu tyto typy lymfocytů:

- *Nejstarší* – odstraní se určitý počet těch lymfocytů, které jsou nejdéle nepoužívané.

- *Nejmenší spam-rate* – odstraní se určitý počet lymfocytů, které mají nejmenší schopnost vázat se na spam. Hodnota se počítá jako podíl rozpoznaných spamů a počtu spamů, které prošly systémem.
- *Největší false spam-rate* – pro uživatele je daleko přijatelnější mazat ze své schránky spamy, které projdou spamovým filtrem (false negatives), než aby korektní e-maily byly označovány jako spam (false positives). Hodnota false spam-rate tedy udává kolik procent korektních e-mailů tento lymfocyt označil jako spam. Z databáze se tedy odstraňují lymfocyty s největším hodnocením.
- *Nepoužívané lymfocyty* – z databáze se odstraní ty lymfocyty, které obsahují nejméně receptorů nebo které se nepoužívají vůbec.

V databázi se také udržuje maximální počet umělých receptorů – jejich přebytek se tedy v tomto kroku odstraní.

Po odstranění neužitečných umělých lymfocytů se databáze doplní. Pro vytváření nových umělých lymfocytů se používá klonální selekční algoritmus (viz kap. 2.3.3), při kterém se několikrát zkopírují (naklonují) ty nejlépe hodnocené umělé lymfocyty. Čím je hodnocení umělého lymfocytu lepší, tím vícekrát se zkopíruje. Každý takto nově vytvořený umělý lymfocyt se dále mutuje – jsou mu částečně měněny spojení s umělými receptory. Klonální selekční algoritmus řídí velikost této mutace, platí však, že čím má umělý lymfocyt větší hodnocení, tím méně se mutuje.

Zařazením nově vytvořených umělých lymfocytů do systému se nastartuje algoritmus pozitivní (kap. 2.3.1) a negativní selekce (kap. 2.3.2). Tyto algoritmy spočívají v tom, že se nově vytvořené umělé lymfocyty nepoužívají pro filtraci spamu, ale určitou dobu jsou aplikovány na příchozí e-mailové zprávy a učí se. Stejně jako lymfocyty typu T, které nějakou dobu cestují po těle jedince a jsou z nich zachovány jen ty, které se neváží na vlastní prvky (negativní selekce) a zároveň se dokážou navázat na cizí prvky (pozitivní selekce).

5.3.4 Učení

Poslední částí modelu adaptivní imunity je modul učení. Jako vstup této funkci jsou předána slova e-mailové zprávy a výsledek, k jakému došel model vrozené nebo adaptivní imunity. Proces učení se zase probíhá v několika krocích:

- Z databáze umělých lymfocytů se vyberou jen ty, jejichž receptory se navázaly na některá slova e-mailu.
- Z tohoto výběru se dále vyberou jen ty lymfocyty, od kterých se na e-mail navázalo alespoň určité procento receptorů.
- Těmto umělým lymfocytům se zvýší hodnota *numberOfMailMatched* a pokud se jedná o spam, zvýší se také hodnota *numberOfSpamMatched*.

- V dalším kroku se aktualizuje datum posledního použití jednotlivých receptorů a lymfocytů.
- V posledním kroku modulu učení se u všech lymfocytů v celé databázi zvýší hodnota *numberOfMailChecked* a pokud se jednalo o spam, zvýší se také hodnota *numberOfSpamChecked*.

Tato část je v procesu filtrování spamu poslední, tímto modulem průchod mailu spamovým filtrem končí.

5.4 Ovládání

Ovládání aplikace je velmi jednoduché a zahrnuje nastavení klíčových parametrů, které jsou důležité pro fungování spamového filtru a které ovlivňují jeho běh. Většina tohoto nastavení se provádí přes webové rozhraní aplikace, které je blíže popsáno v podkapitole 5.4.1. Druhý typ nastavení, které není potřeba měnit tak často nebo se nemění vůbec, se nastavuje před kompilací spamového filtru v souboru *Konstanty.cs*. V tomto souboru se dají změnit tyto hodnoty:

- **maxDelkaSlova** – hodnota typu integer udává, jaká je maximální délka zpracovávaného slova z e-mailové zprávy. Každé delší slovo bude ořezáno. Výchozí hodnota tohoto nastavení je 16.
- **wordSeparators** – pole obsahující znaky (char), které reprezentují oddělovače slov v textu e-mailové zprávy. Výchozí hodnota je mezera, konec řádku, interpunkční znaménka a znak „větší než“ (>) a „menší než“ (<).
- **LogFile** – hodnota typu string, která udává složku, do které se v serveru budou ukládat soubory s logovacími informacemi.
- **ErrorFile** – hodnota typu string, která obsahuje soubor, do kterého se budou ukládat chybová hlášení programu.
- **urovenLogovani** – hodnota typu integer, která udává úroveň logování informací. Hodnoty mohou být od 0 (nelogovat žádné informace) až po úroveň 5 (logovat vše).

5.4.1 Nastavení

Hlavní nastavení spamového filtru se provádí přes webové rozhraní. Po spuštění prohlížeče a přechodu na adresu http://server/SpamUI/Settings.aspx se zobrazí dialog pro úpravu nastavení spamového filtru (viz Příloha č. 6). V tomto dialogu můžeme změnit tyto hodnoty:

- **Počet lymfocytů adaptivní imunity** – jaký maximální počet modelů lymfocytů adaptivního imunitního systému bude aplikace udržovat v databázi.
- **Počet receptorů adaptivní imunity** – jaký maximální počet modelů receptorů adaptivního imunitního systému bude aplikace udržovat v databázi.

- **Procento navázaných receptorů nutných pro označení e-mailu** – hodnota zadávaná v procentech, říká kolik je potřeba procent receptorů daného lymfocytu, aby se navázalo na e-mailovou zprávu, abychom ji označili jako podezřelou ze spamu.

V sekci čistění databáze jsou hodnoty, které udávají, kolik daných lymfocytů se má z databáze odstranit během procesu čištění databáze:

- **Nepoužívané lymfocyty** – počet nejméně využívaných lymfocytů.
- **Malý spam-rate** – počet lymfocytů, které mají nejmenší schopnost navázat se na spam.
- **Velký false spam-rate** – počet lymfocytů, které nejvíce označují korektní e-mailové zprávy jako spam.
- **Staré lymfocyty** – počet nejstarších lymfocytů, které během procesu čištění databáze, budou odstraněny.

Další nastavení:

- **Kolik dobrých lymfocytů klonovat** – udává počet nejkvalitnější lymfocytů, které se budou množit.
- **Kolikrát daný lymfocyt klonovat** – udává násobek, kolikrát se kvalitní lymfocyt bude množit.
- **Minimální počet rozpoznaných spamů, aby se lymfocyt použil pro detekci** – tato hodnota udává minimální počet spamů, které musí nové lymfocyty rozpoznat, aby se začaly používat pro detekci spamu.
- **Minimální skóre pro spam** – udává minimální skóre pro spam, aby byl označen. Hodnota je v procentech.

Další nastavení spamového filtru spočívá v definici pravidel modelu vrozené imunity spamového filtru. Po kliknutí na tlačítko receptory v sekci menu *Vrozená imunita* se zobrazí tabulka modelů receptorů vrozeného imunitního systému (viz Příloha č. 3). Jednotlivé pravidla se dají upravit, smazat nebo vytvořit nové pomocí kliknutí na příslušná tlačítka.

Samotné pravidlo reprezentující receptor vrozené imunity může být vybráno z několika typů. Každý typ kontroluje určité pole v e-mailové zprávě a patří mezi ně: *předmět zprávy, odesílatel, IP adresa odesílatele, textové tělo, textová znaková sada, textové kódování, textový typ, HTML tělo zprávy, HTML znaková sada, HTML kódování a HTML typ zprávy*. Každé toto pole může být kontrolováno a porovnáváno jedním ze čtyř způsobů – pravidlo může kontrolovat, jestli se pole e-mailu **shoduje** s danou hodnotou, jestli **obsahuje** danou hodnotu, **nerovná se** nebo **neobsahuje** danou hodnotu. Po kliknutí na tlačítko **uložit**, se pravidlo uloží do databáze a ihned se spamovým filtrem začne používat.

5.4.2 Výstupy

Spamový filtr během svého běhu ukládá výstupy v několika formách:

- *Databáze modelů receptorů a lymfocytů* – během zpracování e-mailových zpráv si spamový filtr vytváří a udržuje databázi modelů lymfocytů a receptorů adaptivního imunitního systému. Obsah této databáze se zobrazí kliknutím na příslušná tlačítka webového rozhraní v sekci menu *Adaptivní imunita* (viz Příloha 4. a 5.).
- *Historie* – po kliknutí na tlačítko **Historie** ve webovém rozhraní se zobrazí poslední 50 zpracovaných e-mailových zpráv s hodnocením a modely lymfocytů a receptorů, které se na e-mailovou zprávu navázaly.
- *Logovací soubory* – v závislosti na konstantě urovenLogovani se ukládají informace o průběhu zpracování e-mailových zpráva spamovým filtrem.

Poslední, ale pro koncového uživatele nejdůležitější forma výstupu spamového filtru, je zajisté označení e-mailové zprávy příznakem, zda se jedná o spam nebo zda jde o korektní e-mailovou zprávu. Pokud je e-mailová zpráva označena za spam, vloží se před předmět zprávy příznak ve tvaru „**[castIS SPAM]**“, kde **castIS** nabývá hodnoty **Innate** (popř. **Adaptive**), což udává, která část spamového filtru označila e-mailovou zprávu za spam.

5.5 Implementace

V této kapitole se podíváme, jakým způsobem byl spamový filtr implementován. Aplikace vytvořeného spamového filtru na bázi umělých imunitních systémů se skládá ze dvou částí. První z nich je samotný spamový filtr, který se stará o filtraci příchozí pošty. Druhá část aplikace je webové uživatelské rozhraní, pomocí kterého může uživatel sledovat stav databáze spamového filtru nebo změnit výchozí nastavení filtru a ovlivnit tak jeho chod.

5.5.1 Spamový filtr

Samotný spamový filtr je implementován jako dynamická knihovna, která se naváže na SMTP službu, jak již bylo popsáno v předchozích kapitolách. Tato část aplikace se skládá ze šesti souborů s kódem:

SMTPOnArrival.cs

Soubor *SMTPOnArrival.cs* obsahuje výchozí funkce aplikace. Funkce *ISMTPOnArrival.OnArrival(Message msg, ref CdoEventStatus EventStatus)* je vstupní funkce spamového filtru. Tato funkce je volána SMTP službou vždy, když do systému vstoupí nová e-mailová zpráva. První z parametrů je struktura obsahující daný e-mail, který se bude zpracovávat, a druhý je návratová hodnota, pomocí které nastavíme, zda se e-mail bude dále zpracovávat nebo ne.

Tento parametr nemá na výsledek spamového filtru vliv. Funkce *OnArrival* ve svém těle postupně volá funkce části vrozené imunity a adaptivní imunity.

dataDC.dbml

Soubor *dataDC.dbml* obsahuje datové třídy jednotlivých objektů reprezentující databázi spamového filtru. Každá třída koresponduje s jednou tabulkou v relační databázi v SQL serveru. Více o datovém modelu je níže v kapitole 5.3.3.

InnateImmunity.cs

Soubor InnateImmunity.cs obsahuje statickou třídu InnateImmunity, která reprezentuje model vrozené části imunitního systému. V této třídě jsou tři funkce, které řídí průběh zpracování e-mailu v části modelu vrozené imunity spamového filtru. Jsou to:

- `public static int check(Message msg, string logFile)` – vstupní funkce, která zpracuje e-mail. Jako výsledek vrací konstanty *IsSPAM*, *NotSPAM* nebo *Nothing* v případě, že vrozená imunita nedošla k výsledku.
- `private static int IIcheck(Message msg, int typ, int porovnani, string hodnota, int vysledek)` – funkce, která použije konkrétní pravidlo a podle zadaných parametrů ho vyhodnotí. Jako výsledek vrací konstanty *IsSPAM*, *NotSPAM*, *Nothing* (pokud nemá pravidla, podle kterých by určil výsledek) nebo *DontKnow* v případě, že existují pravidla, která si protiřečí.
- `private static bool porovnej(string CastMailu, string HodnotaVdb, int Porovnani)` – pomocná funkce, která porovná dvě hodnoty různými typy porovnání. Tyto typy jsou *Shoda*, *Obsahuje*, *Neobsahuje* a *Liší se*.

AdaptiveImmunity.cs

Soubor AdaptiveImmunity.cs obsahuje statickou třídu AdaptiveImmunity reprezentující model adaptivní části imunitního systému. Implementace této části byla nejsložitější a je nejrozsáhlejší. Třída AdaptiveImmunity obsahuje tyto funkce:

- `public static double check(int logIdHistorie, string logFile, ref string[] GlobalSlova)` – tato funkce reprezentuje a stará se o testování e-mailové zprávy pomocí adaptivní částí spamového filtru.
- `public static void DatabaseLearn(int Status, string logFile, ref string[] GlobalSlova)` – funkce DatabaseLearn vyhodnocuje aktuální e-mail a na základě výsledku (Status), ke kterému dospěl model vrozené nebo adaptivní části imunitního systému, upraví databázi.
- `public static void DatabaseModify(string logFile, ref string[] GlobalSlova)` – tato funkce se stará o správu a úpravu databáze modelu adaptivní části

imunitního systému. Obsahuje úkony pro vymazání určité části umělých lymfocytů a vytvořením nových. Podrobnější popis algoritmů naleznete v kapitole 5.4.

Routines.cs

Tato statická třída obsahuje podpůrné funkce, které využívají ostatní části spamového filtru. Patří mezi ně například ukládání informací do log souborů, logování chyb nebo čtení a práce s nastavením.

Konstanty.cs

Třída Konstanty je statická třída, ve které jsou uloženy všechny konstanty programu. Pro tyto hodnoty je vytvořena záměrně samostatná třída (a soubor), aby bylo snadné jednoduchým zásahem tyto konstanty v případě nutnosti změnit. Mezi tyto hodnoty patří *oddělovače slov*, *maximální zpracovávaná délka slova v mailu*, umístění *logovacích souborů* nebo *úroveň logování*.

5.5.2 Webové rozhraní

Druhá část aplikace spamového filtru je webové rozhraní. Toto webové rozhraní je implementováno jako dynamické webové stránky s využitím technologií ASP .NET, Linq a jazyka C#. Webové rozhraní se skládá z těchto stránek:

- InnateReceptors.aspx – webová stránka zobrazující modely receptorů z vrozené imunity.
- AdaptiveReceptors.aspx – webová stránka zobrazující modely receptorů z adaptivní imunity.
- AdaptiveLymphocytes.aspx – webová stránka zobrazující modely bílých krvinek z adaptivní imunity.
- Settings.aspx – webová stránka s nastavením spamového filtru.
- History.aspx – webová stránka zobrazující posledních 25 e-mailových zpráv, které prošly spamovým filtrem, a výsledky analýzy.

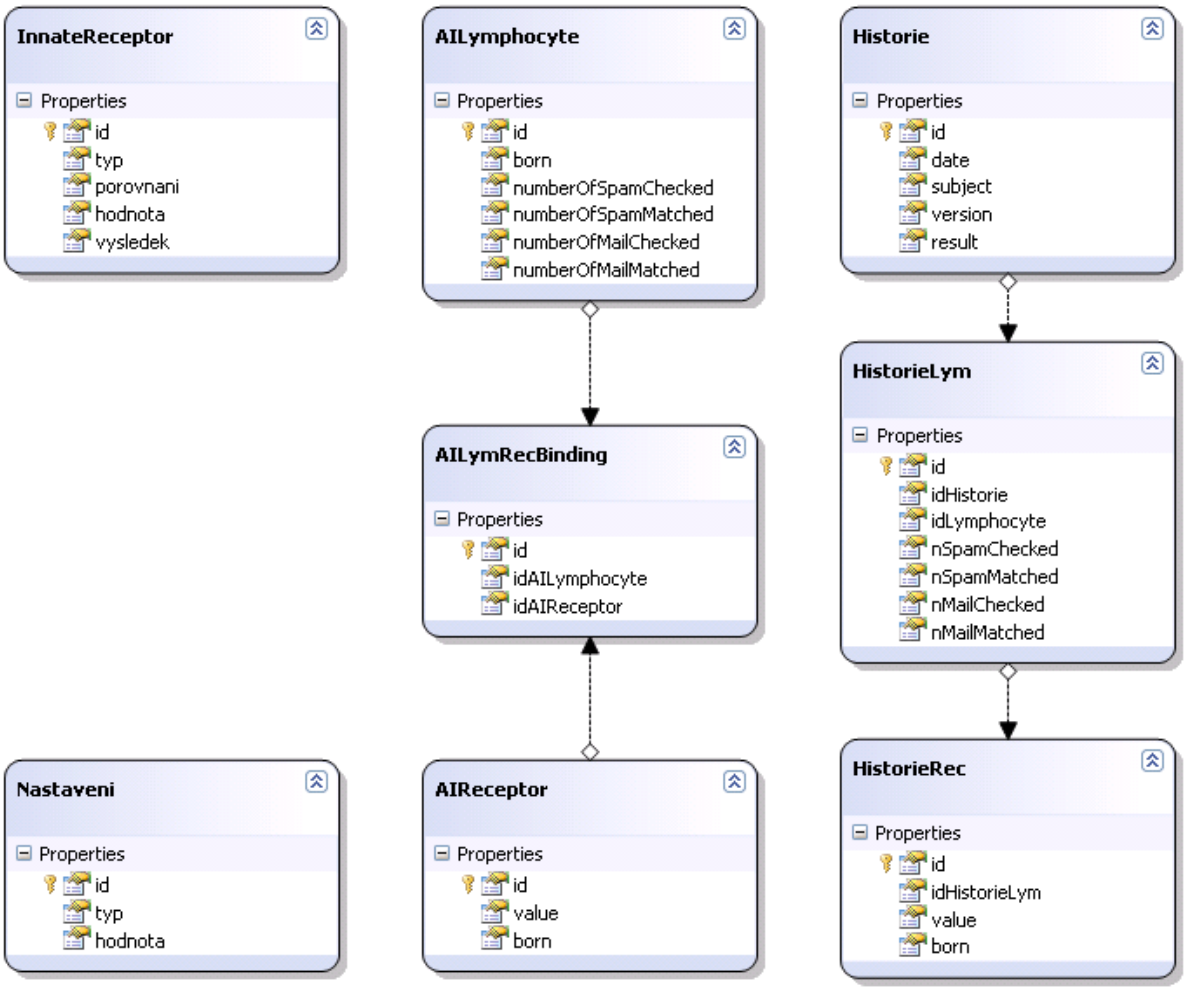
5.5.3 Datový model

Data spamového filtru jsou uložena v SQL databázi. Tato databáze obsahuje několik tabulek, které bychom si mohli rozdělit do těchto skupin:

- Data modelu vrozené imunity – tabulka InnateReceptor obsahuje model receptorů vrozené imunity.
- Data modelu adaptivní imunity – tabulky AILymphocyte, AIReceptor a AILymRecBinding obsahují modely receptorů a bílých krvinek adaptivní imunity. Poslední zmiňovaná tabulka váže tyto dva prvky vztahem M:N.

- Data uchování historie – tabulky Historie, HistorieLym a HistorieRec obsahují informace o již zpracovaných e-mailových zprávách. Tyto tabulky slouží spíše pro zpětné monitorování než po běh aplikace. Díky těmto tabulkám dochází k redundanci určitých dat (receptorů a lymfocytů), ale kvůli dynamičnosti obsahu databází modelu adaptivní imunity je to potřeba.
- Data nastavení spamového filtru – jsou uložena v tabulce Nastavení.

Na následujícím obrázku je zobrazen diagram datového modelu spamového filtru.



Obrázek 8 - Datový model

6 Experimentální vyhodnocení

V této kapitole popíšeme výsledky experimentování s vytvořeným spamovým filtrem. Spamový filtr byl podroben několika testům, kde každý z nich obsahoval jiné hodnoty v nastavení spamového filtru. Jednotlivé testy se lišily v nastavení čištění databáze a v nastavení vytváření nových umělých lymfocytů.

6.1 Výchozí prostředí

Testování probíhalo nad vzorkem 100 e-mailových zpráv, z nichž 80% tvořil spam. Tyto zprávy byly uloženy do složky, ze které si SMTP služba zprávy vyzvedávala a předávala ke zpracování.

Pro testování bylo měněno několik faktorů:

Faktor výchozích stavů databází

Tento faktor měl dvě možné varianty. V první z nich byly do databáze modelu vrozené imunity vloženy tyto umělé receptory:

- pokud textové tělo obsahuje **replica**, jedná se o spam,
- pokud textové tělo obsahuje **meds**, jedná se o spam,
- pokud textové tělo obsahuje **pills**, jedná se o spam,
- pokud textové tělo obsahuje **viagra**, jedná se o spam,
- pokud se znaková sada rovná **Cyrilice(koi8-r)**, jedná se o spam

a do databáze modelu adaptivní imunity byl vložen jeden umělý lymfocyt s desíti receptory (slova: replica, sex, viagra, enhancement, pills, growth, offer, weight, price a enlarging). V druhé variantě byla smazána databáze modelu vrozené imunity a spamový filtr se spoléhal jen na již vytvořenou databázi modelu adaptivní imunity.

Faktor velikosti databáze modelu adaptivní imunity

Druhý faktor se opět skládá ze dvou variant. V první variantě jsme omezili maximální počet umělých lymfocytů na 24 a receptorů na 100 a v druhé variantě jsme v databázi povolili maximálně 100 umělých lymfocytů a 1000 receptorů.

Faktor změny databáze modelu adaptivní imunity

Tento faktor má opět dvě varianty. V první variantě spamový filtr mazal čtvrtinu umělých lymfocytů z databáze při každém procesu čištění databáze. V druhé variantě mazal spamový filtr polovinu záznamů.

Faktor způsobu čištění databáze

Poslední faktor spočíval ve způsobu promazávání databáze modelu adaptivní imunity. Databáze maže určitý počet umělých lymfocytů podle daného nastavení (viz 5.3.3).

V první variantě tohoto faktoru mažeme z databáze převážně ty umělé lymfocyty, které se nejméně používají. Tato varianta testu tedy lépe simuluje prostředí, ve kterém chodí na SMTP server hodně různých typů spamu.

Druhá varianta tohoto faktoru z databáze maže především ty lymfocyty, které vykazují malou hodnotu spam-rate. Tato varianta testu se tedy hodí více pro simulaci prostředí, kde se vyskytuje větší množství podobného typu spamu.

6.2 Testy

Číslo testu	1	2
Výchozí stav DB	s pravidly vrozené imunity	jen adaptivní imunita
maximální počet lymfocytů	24	24
maximální počet receptorů	100	100
počet klonovaných lymfocytů	3	3
kolikrát byly klonovány	2	2
smazat nepoužívaných lymfocytů	3	3
smazat s malým spam-rate	1	1
smazat s velkým false spam-rate	1	1
smazat staré lymfocyty	1	1
průměrný čas zpracování jednoho e-mailu	11,389 s	10,608 s
Rozpoznaných spamů	**35**	**49**
Chybně označeno za spam	**0**	**2**

Číslo testu	3	4
Výchozí stav DB	s pravidly vrozené imunity	jen adaptivní imunita
maximální počet lymfocytů	100	100
maximální počet receptorů	1000	1000
počet klonovaných lymfocytů	5	5
kolikrát byly klonovány	5	5
smazat nepoužívaných lymfocytů	13	13
smazat s malým spam-rate	4	4
smazat s velkým false spam-rate	4	4
smazat staré lymfocyty	4	4
průměrný čas zpracování jednoho e-mailu	43,157 s	42,393 s
Rozpoznaných spamů	**42**	**59**
Chybně označeno za spam	**6**	**6**

Číslo testu	5	6
Výchozí stav DB	s pravidly vrozené imunity	jen adaptivní imunita
maximální počet lymfocytů	24	24
maximální počet receptorů	100	100
počet klonovaných lymfocytů	4	4
kolikrát byly klonovány	3	3
smazat nepoužívaných lymfocytů	4	4
smazat s malým spam-rate	3	3
smazat s velkým false spam-rate	3	3
smazat staré lymfocyty	2	2
průměrný čas zpracování jednoho e-mailu	16,673 s	15,960 s
Rozpoznaných spamů	**37**	**47**
Chybně označeno za spam	**0**	**3**

Číslo testu	7	8
Výchozí stav DB	s pravidly vrozené imunity	jen adaptivní imunita
maximální počet lymfocytů	100	100
maximální počet receptorů	1000	1000
počet klonovaných lymfocytů	10	10
kolikrát byly klonovány	5	5
smazat nepoužívaných lymfocytů	25	25
smazat s malým spam-rate	9	9
smazat s velkým false spam-rate	8	8
smazat staré lymfocyty	8	8
průměrný čas zpracování jednoho e-mailu	44,439 s	43,715 s
Rozpoznaných spamů	**44**	**58**
Chybně označeno za spam	**5**	**7**

Číslo testu	9	10
Výchozí stav DB	s pravidly vrozené imunity	jen adaptivní imunita
maximální počet lymfocytů	24	24
maximální počet receptorů	100	100
počet klonovaných lymfocytů	3	3
kolikrát byly klonovány	2	2
smazat nepoužívaných lymfocytů	1	1
smazat s malým spam-rate	3	3
smazat s velkým false spam-rate	1	1
smazat staré lymfocyty	1	1
průměrný čas zpracování jednoho e-mailu	9,315 s	8,579 s
Rozpoznaných spamů	**40**	**53**
Chybně označeno za spam	**1**	**4**

Číslo testu	11	12
Výchozí stav DB	s pravidly vrozené imunity	jen adaptivní imunita
maximální počet lymfocytů	100	100
maximální počet receptorů	1000	1000
počet klonovaných lymfocytů	5	5
kolikrát byly klonovány	5	5
smazat nepoužívaných lymfocytů	4	4
smazat s malým spam-rate	13	13
smazat s velkým false spam-rate	4	4
smazat staré lymfocyty	4	4
průměrný čas zpracování jednoho e-mailu	48,338 s	47,598 s
Rozpoznaných spamů	**47**	**57**
Chybně označeno za spam	**5**	**6**

Číslo testu	13	14
Výchozí stav DB	s pravidly vrozené imunity	jen adaptivní imunita
maximální počet lymfocytů	24	24
maximální počet receptorů	100	100
počet klonovaných lymfocytů	4	4
kolikrát byly klonovány	3	3
smazat nepoužívaných lymfocytů	3	3
smazat s malým spam-rate	3	3
smazat s velkým false spam-rate	3	3
smazat staré lymfocyty	3	3
průměrný čas zpracování jednoho e-mailu	12,910 s	12,310 s
Rozpoznaných spamů	**37**	**41**
Chybně označeno za spam	**0**	**1**

Číslo testu	15	16
Výchozí stav DB	s pravidly vrozené imunity	jen adaptivní imunita
maximální počet lymfocytů	100	100
maximální počet receptorů	1000	1000
počet klonovaných lymfocytů	10	10
kolikrát byly klonovány	5	5
smazat nepoužívaných lymfocytů	9	9
smazat s malým spam-rate	25	25
smazat s velkým false spam-rate	8	8
smazat staré lymfocyty	8	8
průměrný čas zpracování jednoho e-mailu	44,720 s	43,916 s
Rozpoznaných spamů	**43**	**57**
Chybně označeno za spam	**4**	**6**

Číslo testu	17
Výchozí stav DB	s rozšířenými pravidly vrozené imunity (celkem 7 pravidel[12])
maximální počet lymfocytů	24
maximální počet receptorů	100
počet klonovaných lymfocytů	4
kolikrát byly klonovány	3
smazat nepoužívaných lymfocytů	4
smazat s malým spam-rate	3
smazat s velkým false spam-rate	3
smazat staré lymfocyty	2
průměrný čas zpracování jednoho e-mailu	19,691 s
Rozpoznaných spamů	**51**
Chybně označeno za spam	**0**

[12] K původním pěti přídány dvě: „pokud předmět obsahuje **Diet**, jedná se o spam“ a „pokud textové tělo obsahuje **weight**, jedná se o spam“.

6.3 Zhodnocení naměřených výsledků

Naměřené hodnoty jednotlivých testů vykazují nízké hodnoty počtu rozpoznaných spamů. Spamový filtr rozpoznal z 80 nevyžádaných e-mailový zpráv v průměru 46,625 spamů. Ve srovnání s komerčním spamovým filtrem z programu Microsoft Outlook 2003, který rozpoznal 70 spamů a 2 korektní e-maily označil za spam, je tento naměřený počet nízký. Tento výsledek je dán převážně ze dvou důvodů. První z nich je malá velikost startovací množiny pravidel v modelu vrozené imunity. Počet těchto pravidel byl zvolen záměrně tak nízký, aby se dala lépe prezentovat schopnost spamového filtru učit se nová pravidla. Počet těchto pravidel je ve srovnání se spamovým filtrem v programu Microsoft Outlook 2003 nesrovnatelně nižší (podle [10] obsahuje databáze pravidel spamového filtru v programu Microsoft Outlook 2003 přibližně 93 000 záznamů). Jak ukazuje test s číslem 17, kterému jsme do startovacích pravidel přidali 2 další pravidla, množství těchto startovacích pravidel významně ovlivňuje schopnost spamového filtru rozpoznat spam od korektních e-mailových zpráv.

Druhý důvod je nastavení malé citlivosti na spam (minimální skóre pro spam, kap. 5.4.1). Změnou této hodnoty, by se dalo dosáhnout zvýšení počtu rozpoznaných spamů, nicméně jako vedlejší efekt by stoupl počet chybně označených e-mailů. To je ale z pohledu uživatelů méně přijatelné, protože uživatelé raději vymažou přijatý spam ze své schránky, než aby jim byl korektní e-mail smazán spamovým filtrem.

Průměrná doba potřebná pro zpracování e-mailové zprávy je asi 28,501 sekund. Tato hodnota je ve srovnání se spamovým filtrem v programu Microsoft Outlook 2003 vysoká. Hlavní důvod je použité úložiště databáze pravidel. Spamový filtr v programu Microsoft Outlook 2003 [10] má databázi uloženou v binárním souboru *outlfltr.dat*, který má uložen a přímo do něj přistupuje v operační paměti počítače. Navržený spamový filtr na bázi umělého imunitního systému používá databázi pravidel, kterou ukládá na SQL Server. Hlavní prodlevy tedy způsobují dotazy do databáze pravidel a optimalizace této části systému by zpracování značně urychlilo. Tento způsob uložení pravidel byl nicméně zvolen z důvodu reprezentace vlastnosti biologických imunitních systému, kterou je distribuovaná struktura (viz kap. 2.2.2.5). Pokud implementuje navržený spamový filtr na bázi umělých imunitních systémů do rozlehlejší sítě na několik serverů, mohou navzájem paměť sdílet. Tuto vlastnost spamový filtr v programu Microsoft Outlook 2003 nemá.

6.4 Návrhy na zlepšení

Během implementace spamového filtru na bázi umělého imunitního systému a psaní této diplomové práce jsem narazil na několik možností úprav v různých oblastech, které by mohly úspěšnost spamového filtru ještě zvýšit. Spamový filtr nyní prezentuje využití prvků umělých imunitních

systému v boji proti spamu, nicméně tyto úpravy by mohly zvýšit jeho schopnosti a dosáhnout lepších výsledků v rozpoznávání spamu.

První oblastí je rychlost zpracování příchozích e-mailových zpráv. Zrychlení analýzy e-mailů by se dalo dosáhnout, kdyby spamový filtr neměnil databázi modelu adaptivní imunity při každém nově příchozím e-mailu. V biologickém imunitním systému dochází k průběžnému obnovování lymfocytů typu B a T, nicméně v případě spamového filtru by stačilo tuto změnu (zánik neužitečných umělých lymfocytů a naklonování nových) provádět například v 30 minutových intervalech.

Další oblast, která by mohla zvýšit schopnost rozpoznávat nevyžádané e-mailové zprávy, spočívá ve změně pohledu na umělý receptor. Nyní tento receptor představuje detektor jednotlivého slova vyskytujícího se v těle zprávy. Umělé receptory by se dále mohly rozšířit na detekci např. slova v předmětu zprávy, ip adresy odesílatele, kódování textu, použitou znakovou sadu, čas přijetí e-mailu, atd.

Poslední rozšíření, které spočívá také v možném zvýšení schopnosti rozpoznat spam, je změna vzorce, pomocí kterého se počítá spamové skóre (kap. 5.3.2). Do tohoto vzorce by se mohla zavést záporná hodnota r_f, která představuje hodnotu *False spam-matched rate* (podíl počtu korektních e-mailů označených jako spam a celkového počtu kontrolovaných e-mailů). Nicméně tato změna ovlivní dobu výpočtu, což se negativně projeví na rychlosti zpracování e-mailových zpráv.

7 Závěr

V této diplomové práci jsme se zabývali možností využití principů a prvků umělých imunitních systémů v boji proti spamu.

Abychom mohli vytvořit úspěšný umělý imunitní systém pro filtraci spamu, museli jsme nejdříve pochopit, jak pracuje biologický imunitní systém a museli jsme z něj vybrat jen ty části, prvky a principy, které můžeme aplikovat a implementovat ve spamovém filtru. V první části této práce jsme se tedy dotkli vlastností, jako jsou vícevrstvost, adaptabilita a distribuovaná struktura. Prozkoumali jsme schopnosti rozpoznávání nežádoucích prvků imunitním systémem a jeho schopnosti učit se. Dále jsme si shrnuli, jaké algoritmy se v imunitních systémech vyskytují a jak fungují.

Další část této práce zkoumá a studuje metody, které používají dnešní spamové filtry k detekci nevyžádané pošty. Prošli jsme tři základní způsoby, pomocí kterých jsou rozpoznávány korektní e-maily od spamu. V další částí práce se nachází návrh využití prvků umělých imunitních systémů při tvorbě systému pro filtraci spamu.

V praktické části diplomové práce jsme se věnovali vývoji spamového filtru, který napodobuje a využívá prvky a principy biologických imunitních systémů. Dále jsme se pokusili zhodnotit dosažené výsledky a pomocí experimentování s vytvořeným spamovým filtrem jsme zjistili, jak je tento přístup k filtrování spamu efektivní a úspěšný. Došli jsme k závěru, že tento způsob filtrování elektronické pošty je svým způsobem zajímavý, nicméně pro reálné využití je potřeba ještě několik aspektů zlepšit a dále studovat možnosti tohoto, v informačních technologiích nového, přístupu k filtrování elektronické pošty.

Literatura

[1] *Spam* [online]. 2008 [cit. 2008-12-27].
Dostupný z WWW: <http://cs.wikipedia.org/wiki/Spam>.

[2] NEUWIRTH, D. *Umělé imunitní výpočetní systémy*. bakalářská práce. Brno. FIT VUT v Brně. 2007

[3] CASTRO, L. N. de, TIMMIS, J. *Artificial Immune Systems : A New Computational Intelligence Approach*. [s.l.] : Springer, 2002. 380 s.

[4] O'NEILL, L. A. J. *Imunitní systém včasné výstrahy. Scientific american*. únor 2006, s. 44-51.

[5] MAAWG Issues First Global Email Spam Report. *News & Events - MAAWG Release* [online]. 2006 [cit. 2008-12-30]. Dostupný z WWW: <http://www.maawg.org/news/maawg060308>.

[6] *Spam* [online]. 2008 [cit. 2008-12-27].
Dostupný z WWW: <http://en.wikipedia.org/wiki/Spam_(electronic)>.

[7] *IP Spoofing* [online]. 2009 [cit. 2009-01-03].
Dostupný z WWW: <http://en.wikipedia.org/wiki/Ip_spoofing>.

[8] *Network Layer Firewall* [online]. 2008 [cit. 2009-01-03].
Dostupný z WWW: <http://en.wikipedia.org/wiki/Network_layer_firewall>.

[9] SCHRYEN, G. *Anti-Spam Measures: Analysis and Design*. Springer, 2007. 209 s.
ISBN 978-3-540-71748-5.

[10] MAPILAB LTD. *MS Outlook 2003 Spam Filter: Under the hood* [online]. 2003.
Dostupný z WWW: < http://www.spamhelp.org/articles/underthehood.pdf>.

[11] *Bayesian probability* [online]. 2009 [cit. 2009-01-02].
Dostupný z WWW: <http://en.wikipedia.org/wiki/Bayesian_probability>.

[12] BARTŮŇKOVÁ, J., ŠEDIVÁ, A., HÖLZELOVÁ, E. *Primární imunodeficience : Příručka pro pacienty a jejich rodiny*. FN Motol, Praha, 2. LF UK : Ústav imunologie. 1999.
Dostupný z WWW: <http://www.tigis.cz/Knihy/imuno/>.

[13] *Antibody* [online]. 2007 [cit. 2008-12-26].
Dostupný z WWW: <http://en.wikipedia.org/wiki/Immunoglobulin>.

[14] Immune System. *Science Aid* [online]. 2006 [cit. 2008-12-27].
Dostupný z WWW: <http://www.scienceaid.co.uk/biology/humans/immunesystem.html>.

[15] *Patient and Family Handbook of the Immune Deficiency Foundation.* USA : The Immune Deficiency Foundation. 1993.

[16] ODA, T., WHITE, T. *Spam Detection using an Artificial Immune System* [online]. 2004 [cit. 2008-12-28]. Dostupný z WWW: <http://terri.zone12.com/doc/academic/crossroads/>.

[17] MAAWG Issues First Global Email Spam Report. *E-mail Metrics Report* [online]. 2007 [cit. 2008-12-30].

Dostupný z WWW: <http://www.maawg.org/about/MAAWG20072Q_Metrics_Report.pdf>.

[18] *Simple Mail Transfer Protocol* [online]. 2009 [cit. 2009-01-05]. Dostupný z WWW: <http://en.wikipedia.org/wiki/Smtp>.

[19] *MIME* [online]. 2009 [cit. 2009-01-06]. Dostupný z WWW: <http://en.wikipedia.org/wiki/MIME>.

[20] *Anti-spam techniques* [online]. 2009 [cit. 2009-05-10]. Dostupný z WWW: <http://en.wikipedia.org/wiki/Anti-spam_techniques_(e-mail)>.

[21] KLENSIN, J. *RFC 5321 Simple Mail Transfer Protocol : Order of Commands* [online]. 2008 [cit. 2009-05-10]. Dostupný z WWW: <http://tools.ietf.org/html/rfc5321#section-4.1.4>.

[22] LEMSON, D. *Microsoft Windows 2000 SMTP Service Events* [online]. Microsoft Corporation. 2000 [cit. 2009-05-01]. Dostupný z WWW: <http://msdn.microsoft.com/en-us/library/ms998608.aspx>.

[23] SCALES, G. *Glen's Exchange Dev Blog : C# Catchall Onarrival Event Sink* [online]. 2005 [cit. 2009-05-09]. Dostupný z WWW: <http://gsexdev.blogspot.com/2005/05/c-catchall-onarrival-event-sink.html>.

Přílohy

Příloha č. 1 – CD s elektronickou verzí diplomové práce a zdrojovými kódy spamového filtru.

Příloha č. 2 – Test umístění IP adresy v nejznámější Blacklistech

Příloha č. 3 – Náhled obrazovky, Receptory vrozené imunity

Příloha č. 4 – Náhled obrazovky, Receptory adaptivní imunity

Příloha č. 5 – Náhled obrazovky, Lymfocyty adaptivní imunity

Příloha č. 6 – Náhled obrazovky, Nastavení parametrů

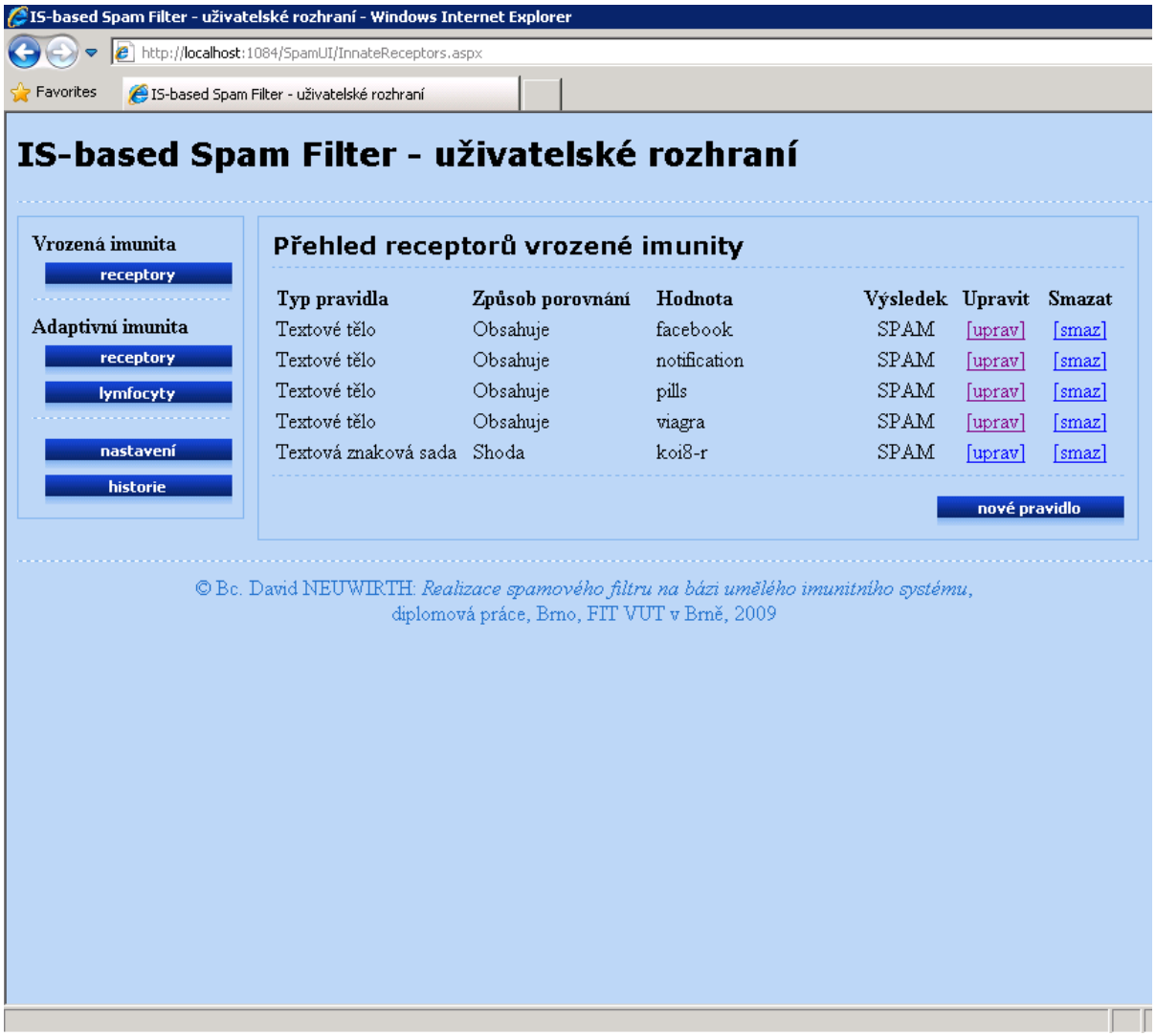
Příloha č. 7 – Náhled obrazovky, Historie

Příloha č. 2 – Test umístění IP adresy v nejznámější Blacklistech

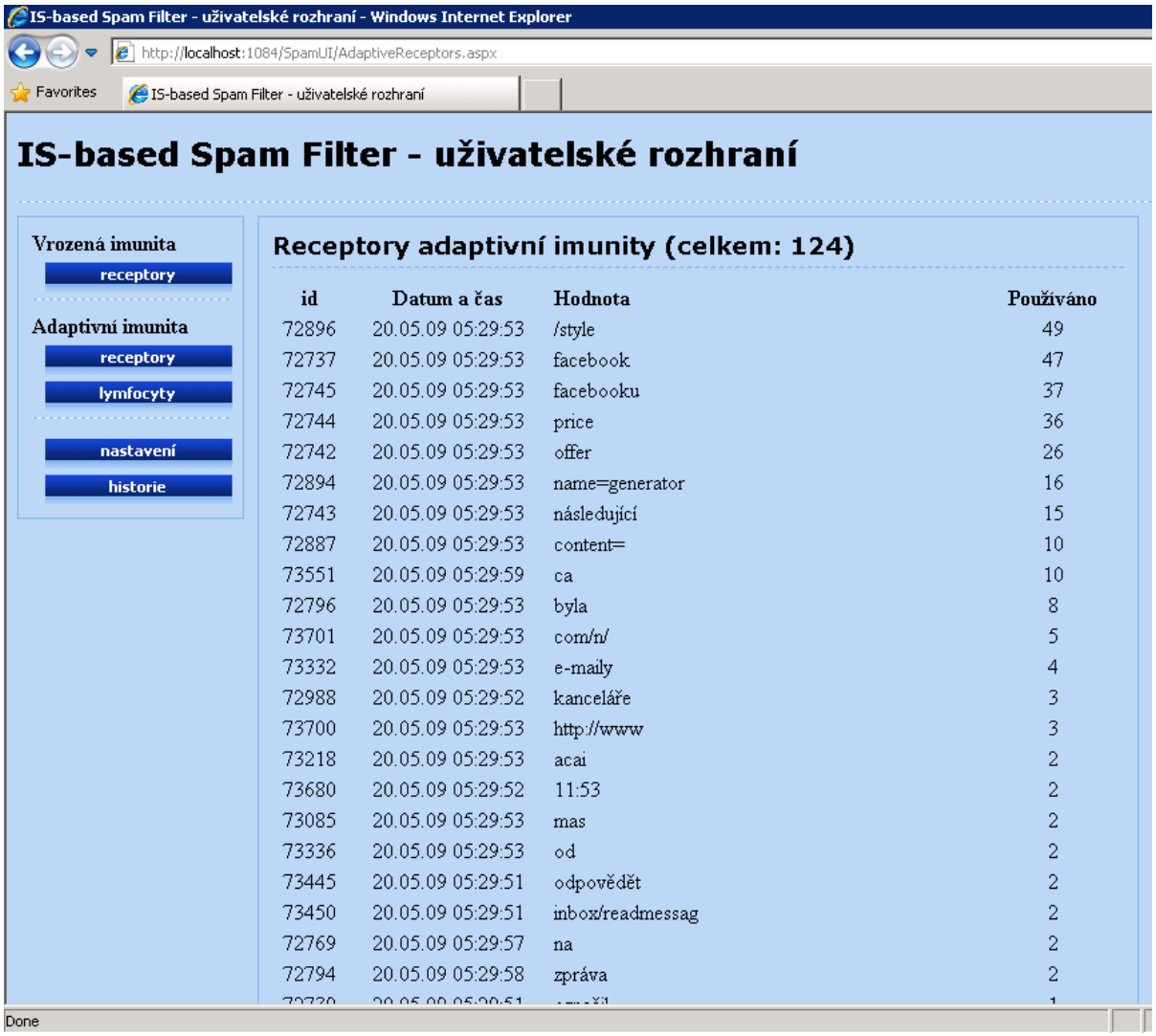
Adresa: http://openrbl.org/client/

Návod: Zadejte ověřovanou IP adresu a klikněte na **Lookup**.

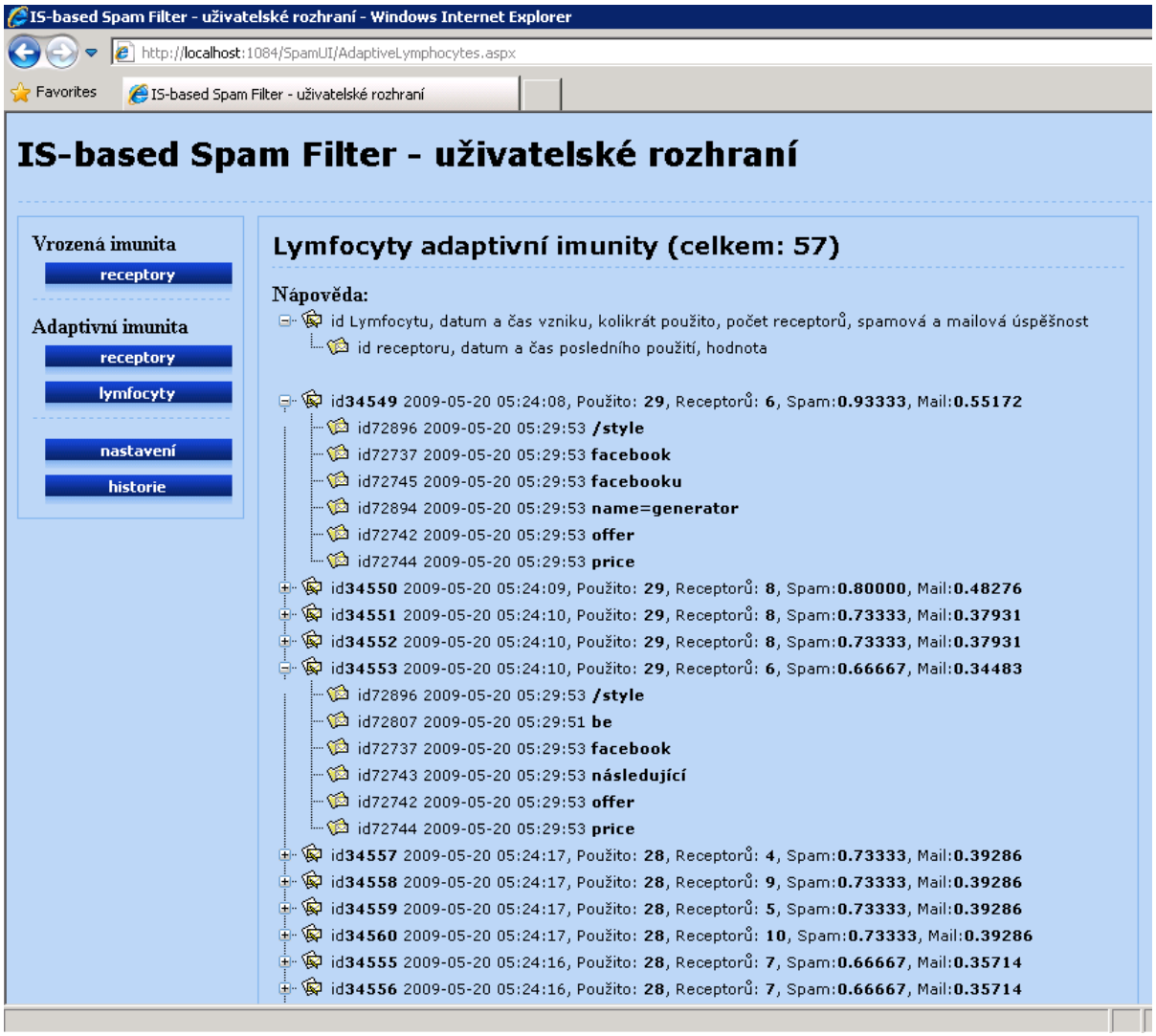
Příloha č. 3 – Náhled obrazovky, Receptory vrozené imunity



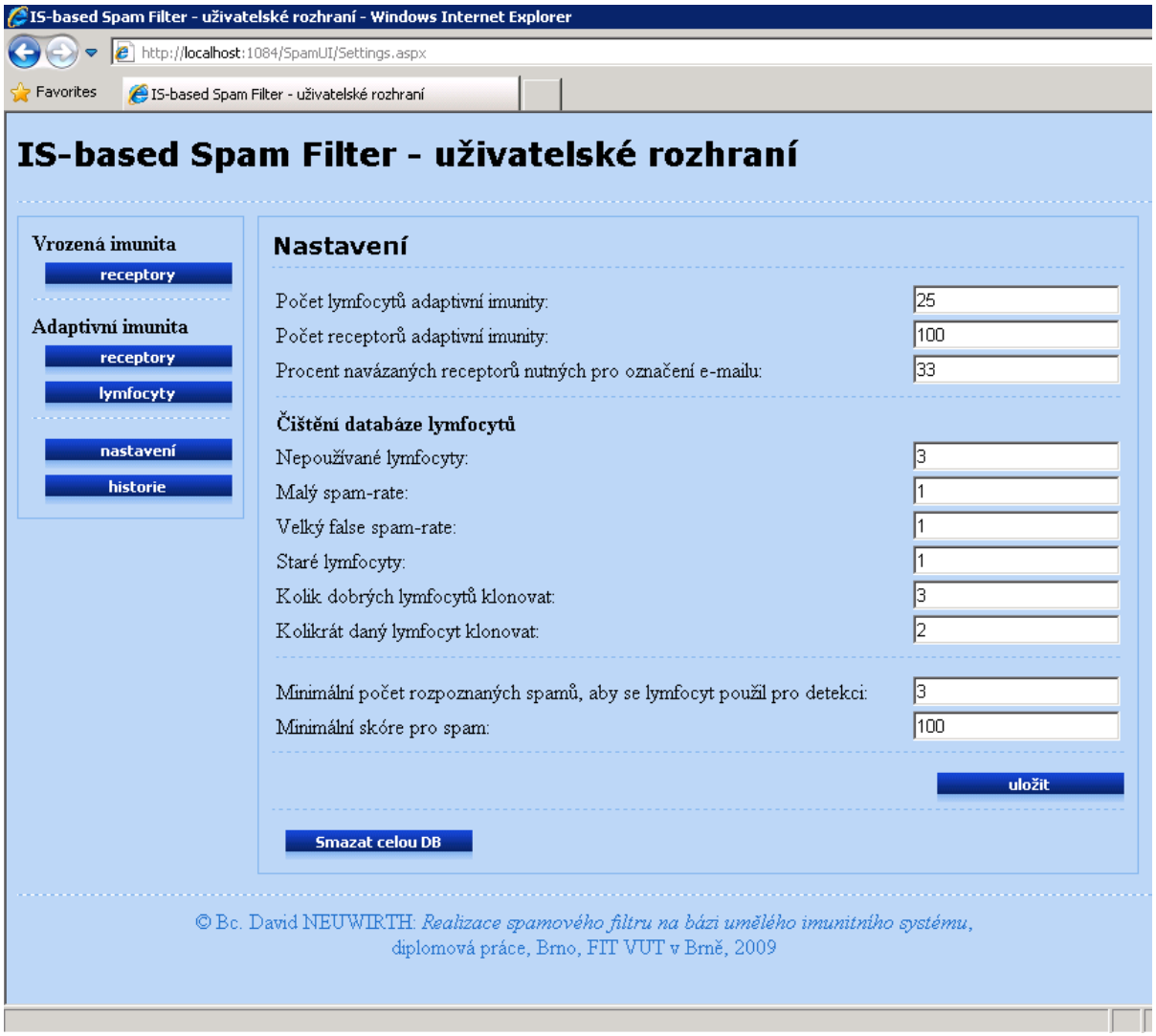
Příloha č. 4 – Náhled obrazovky, Receptory adaptivní imunity



Příloha č. 5 – Náhled obrazovky, Lymfocyty adaptivní imunity



Příloha č. 6 – Náhled obrazovky, Nastavení parametrů



Příloha č. 7 – Náhled obrazovky, Historie