Death2Spam – Le a spammel!

Az előző cikkeimben bemutatott spamszűrők (postgrey, razor) mind hatékonyan képesek csökkenteni a felhasználók spam okozta frusztrációját. Azonban ezek mindegyike egy külső alkalmazás telepítését igényli vagy a levelező szerveren vagy a kliens oldalán. Bizonyos esetekben azonban ez nem járható út. Ebben az írásban egy olyan terméket mutatok be, amely semmilyen alkalmazás telepítését nem igényli, mégis rendkívül hatékony védelmet ad a kéretlen levelek ellen.

Választásom a *Death2Spam (D2S)* nevű spamszűrőre esett, amelyről csak jókat mondanak a felhasználói. A *D2S* valójában nem egy alkalmazás vagy program, hanem szolgáltatás, amelyet az előfizetők éves díjért vehetnek igénybe. Azonban senkinek nem kell a bizonytalanra kifizetni 35 dollárt, regisztráció után egy teljes hónapig ingyen lehet használni a *D2S*-t. Ha még ez sem győzte meg az Olvasót, tekintse meg a felhasználók véleményét a szolgáltatásról (➲ *http://www.death2spam.com/docs/feedback.html*). Olvastam itt olyan hozzászólást is, amelyben egy felhasználó átlag 300 levele közül mindössze 14 volt hasznos, a többi mind spam, de a *D2S* megfogta az összeset. Nem is csoda, hiszen a *D2S* minimum 99% (vagy még jobb) pontosságot ígér.

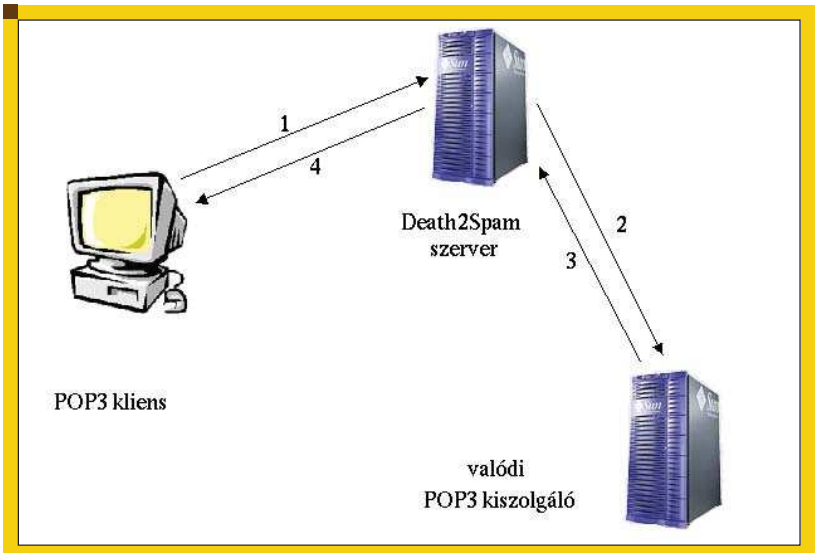
Hogyan működik?

A *D2S* valójában egy *POP3 proxy*, amely a felhasználó és az igazi *POP3* kiszolgáló között van.
A felhasználó úgy állítja be a levelező kliensét, hogy az a *D2S*-sel kommunikáljon, az pedig a felhasználó *POP3* parancsait továbbítja az igazi *POP3* szerver felé, ill. annak válaszait a kliens felé. Így minden levél átfolyik a *D2S*-en, és ha szerinte a levél biztosan spam, akkor azt megjelöli, a „Tárgy:" mezőbe beszúrja a `[spam]` szöveget, amelyre könnyen lehet szűrni a levelezőprogramban.
A felhasználó számára mindez transzparens módon, automatikusan történik (*1. ábra*).
A *D2S* az alábbi jelöléseket használja a tárgy mezőben:

- `[spam]` A levél biztosan spam
- `[spam?]` Nem biztos, de valószínűleg spam
- `[???]` Teljesen bizonytalan
- `[good?]` Nem biztos, de valószínűleg jó

Mi van benne?

A *D2S* legfontosabb része a *Bayesian* elemző modul. Amikor a felhasználó letölt egy levelet, pl. kiadja a `RETR 3` utasítást, akkor a *D2S* kiszolgáló ezt továbbítja az igazi *POP3* szerver felé, majd fogadja a levelet tartalmazó *TCP* csomagokat. Ezt azonban nem továbbítja azonnal a felhasználó felé, hanem egy átmeneti fájlba írja, amíg a teljes levél meg nem érkezik. Ezután elvégzi rajta a *Bayesian* elemzést, szükség szerint módosítja a levélben a *Tárgy:* mezőt, és csak ezután továbbítja a levelet a felhasználóhoz.
A *Bayesian* szűrőt azonban használat előtt tanítani kell. A *D2S* több forrásból szerzi be ehhez a spam-et, épít a *D2S* felhasználók által történő kategorizálásra, de felhasználják a több év alatt számtalan postafiókból begyűjtött kéretlen levelek sokaságát is.
A *Bayesian* szűrő többféleképpen is tanítható. Ha elkészült a kezdeti adatbázis, akkor lehet választani, hogy csak hibás kategorizálás esetén tanítsuk (*Train-On-Error*), vagy amíg nem kategorizál hiba mentesen (*Train-Until-*



■ *1. ábra* D2S szerver a kliens és a POP3 kiszolgáló között

No-Errors). esetleg minden beérkező levél esetén (*Train-Everything*). Ez utóbbi azonban nagyobb adatbázist eredményez. *Zdziarski* módszere pedig csak addig tanítja az adatbázist, amíg az összes benne lévő token gyakorisága el nem ér egy fix értéket (*Train-Until-Mature*), azaz amíg nem vagyunk elég biztosak egy adott tokent illetően. *Yerazunis* azt állítja, hogy a *TOE* elfogadható teljesítményt és pontosságot ad (➲ *http://crm114.sourceforge.net/Plateau_Paper.html*). Ezért valószínűnek tartom, hogy a *D2S* az első megközelítést használja.

Amikor a *D2S* letölt egy levelet, azt a *Bayesian* modul szavakra, pontosabban szólva, tokenekre bontja.

Ezután lekérdezi az adatbázisból az összes token spam valószínűségét. Ha egy adott token nem szerepel az adatbázisban, akkor egy semleges alapértelmezett értéket (*Paul Graham* szerint 0,4) rendel hozzá. Vegyük példaképpen a következő részletet az egyik levelemből: *„the funds for further investment.”* Ezt tokenekre bontva az alábbi listát kapjuk, amelyhez a saját spam adatbázisom alapján az alábbi értékek tartoznak (*1. táblázat*).

A figyelmes olvasó bizonyára észrevette, hogy nem csak a szavak, hanem azok kombinációi is szerepelnek, mint tokenek, ezzel ugyanis növelhető a *Bayesian* döntés pontossága.

A *CRM114* például egy 5 szó hosszú csúszóablak segítségével képezi a tokeneket.

Ha minden tokennek megvan végre a spam valószínűsége, akkor ez alapján a tokeneket sorba rendezi, majd kiválasztja a „legérdekesebb” 15, esetleg 20 tokent. Azok a tokenek érdekesek, amelyek leginkább eltérnek a semleges középértéktől (0,5), azaz amelyek döntően vagy csak ham, vagy csak spam levelekben fordulnak elő. A kiválasztott tokenek spam valószínűsége alapján pedig egy képlettel kiszámítja az egész levélre érvényes valószínűséget.

A *Bayesian* képlet alapján a példabeli mondat 85,57% valószínűséggel spam, ami még alatta van a szokásos 90% feletti küszöbértéknek.

Itt szeretném megjegyezni, hogy a konkrét levél ennél bővebb volt, és a spamszűrőm 94,57% esélyt látott rá, hogy az említett levél spam, és valóban, a levél tényleg spam, a *junk folder*-emből vettem elő szemléltetés céljára.

Felmerülhet az Olvasóban a kérdés, hogy miért nem veszik figyelembe az összes tokent? Azért, mert ebben az esetben könnyű lenne átverni a szűrőt úgy, hogy egy sor véletlenszerű, semleges szót illeszt bele a spammer, amelyek csökkentik a levél összesített valószínűségét. Több olyan levelet is kaptam már, amelyben egy szöveges és egy *HTML* rész volt. Az utóbbi tartalmazta a spam-et, míg az előbbi több ártatlan, illetve halandzsa szót, amelyek egyértelműen a *Bayesian* szűrők megzavarása miatt voltak ott.

Hogyan használjuk?

A felhasználónak semmit egyebet nem kell tennie, mint ellátogatnia a *D2S* honlapjára és regisztrálnia egy fiókot az alábbi *URL*-en: ➲ *https://www.death2spam.com/d2s/NewAccount*. Meg kell adnunk a *POP3* szerver nevét, *POP3* felhasználó nevünket és jelszavunkat (*2. ábra*). A biztonságra kényes felhasználók megnyugtatására közli velünk az oldal, hogy nem tárolják sehol a jelszavakat, hanem csak a *POP3* fiók létezését ellenőrzik.

A személyes adatok kezeléséről szóló oldalon pedig ígérik, hogy a felhasználókról semmilyen adatot nem gyűjtenek, csak amit a szolgáltatás működtetéséhez feltétlen szükséges, és a jelszavakat sem tárolják, hanem csak egyszerűen továbbküldik az igazi *POP3* szerver felé.

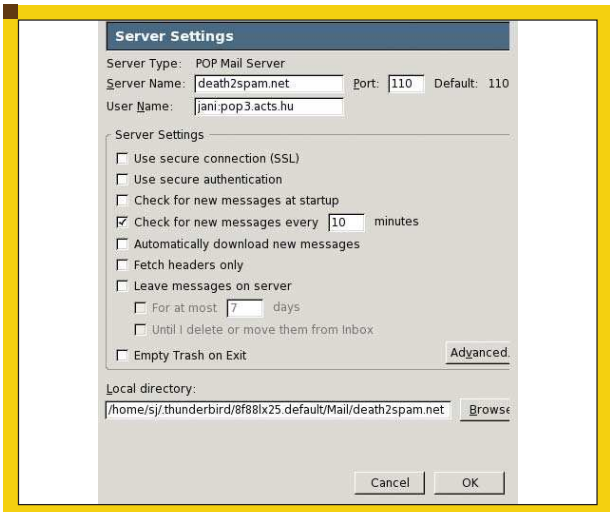
A *D2S* honlapján beszédes ábrák segítik a felhasználókat, hogy a legkülönbözőbb levelező programokkal is tudják használni a *D2S* szolgáltatását. Egyébként a regisztráció után csak be kell állítani a *POP3* szerver nevét, ez kötelezően *death2spam.net*, illetve felhasználói nevünket, amit úgy képezhetünk, hogy vesszük az igazi *POP3* kiszolgálóhoz tartozó felhasználó nevünket, majd kettőspont után folytatjuk azt a kiszolgáló nevével, például *jani:pop3.acts.hu* (*3. ábra*). És már tölthetjük is le a leveleinket. A problémás leveleket a *D2S* a már fentebb említett módon megjelöli, ezeket tehetjük külön mappába.

Esetenként a *D2S* is hibázhat, vagy bizonytalan lehet egy adott levelet illetően, és egy-egy spam átcsúszhat,

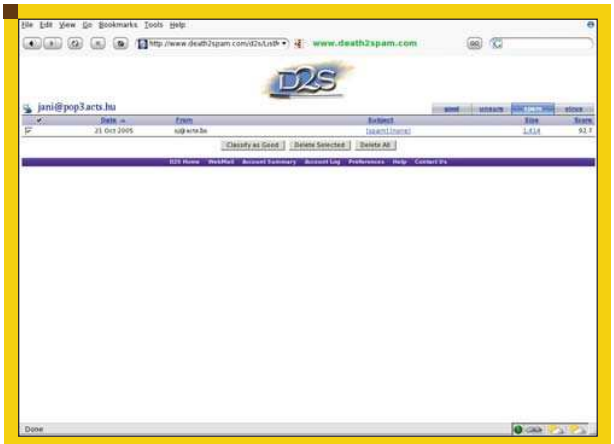
1. táblázat

Token	Spam valószínűség [%]
the	78,95
further	84,22
investment.	99,99
funds	99,99
for	91,80
the+funds	99,99
funds+for	99,99
for+further	78,73
further+investment.	40,00

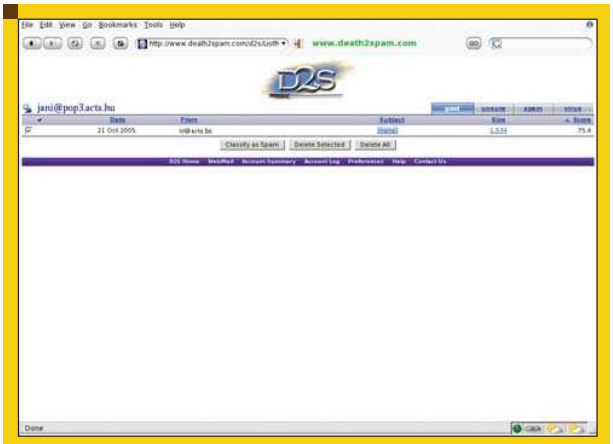
2. ábra Felhasználói fiók készítése



3. ábra Thunderbird beállításai



4. ábra Téves kategorizálás történt!



5. ábra Javítás után

ill. némely leveleket tévesen spamként azonosíthat. Ebben az esetben érdemes tanítani a *D2S*-t, azaz korrigálni a hibát, amit egy könnyen kezelhető web felületen tehetünk meg. A *4. ábrán* egy levél tévesen spamként lett értékelve. Jól látható, hogy az elküldött teszt levél a *D2S* szerint 92.7% valószínűséggel spam. Ez azonban tévedés! Ezért gyorsan korrigáltam, azaz tanítottam, amit a *„Classify as Good”* gombra kattintva tettem meg. Az *5. ábrán* jól látható, hogy a tanítás után már csak 75.4% valószínűséggel gondolja a *D2S* azt, hogy a levél spam, azaz ez a levél így már jó. Hadd tegyem gyorsan hozzá, hogy a levél mindössze a *„teszt”* szót tartalmazta a fejlécen kívül, és még tárgya sem volt.

A *D2S* honlapján egy alap- illetve bővebb listában szereplő opciókat állíthatunk be (*6. és 7. ábra*). Pl. mi alapján rendezze sorba a leveleket; hány levelet mutasson meg egy oldalon; hány % spam valószínűség esetén törölje automatikusan a levelet; törölje-e a vírusos leveleket azonnal; a kategorizálás módját, ami lehet engedékeny, alapértelmezett vagy agresszív.

A *D2S* felhasználói statisztikát is vezet. Számon tartja, hogy hány levelet kaptunk, abból hány volt spam, illetve ham, továbbá a bizonytalan esetek számát (*8. ábra*). A biztonságra sokat adó felhasználók számára jó hír, hogy támogatja a titkosított s*POP3* kapcsolatokat is.

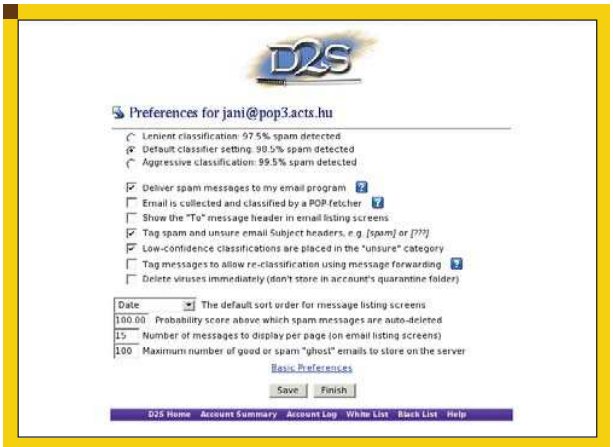
A próba

Itt az ideje, hogy kipróbáljuk a *D2S*-t valódi levelekkel. Ehhez fogtam az utolsó hónap spam termését, 677 levelet, amit a *D2S* segítségével töltöttem le. Két levéllel kapcsolatban teljesen bizonytalan volt.

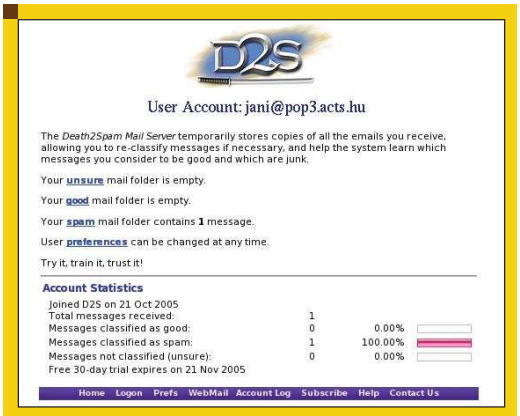
Az egyik egy konferenciára hívta fel a figyelmem. A levél nyilvánvalóan spam, legalábbis abban az értelemben, hogy kéretlen, azonban elolvasva mégis csak ártatlannak ítéltem. A másik egy kínai nyelvű levél, biztosan spam. További 3 levélre azt mondta, hogy valószínűleg spam, és igaza volt. Néhány levéllel azonban gondban voltam. Míg némely leveleket a web felületen határozottan spam-ként jelölt meg, addig ezen levelek „Tárgy” mezejében – miután azokat letöltöttem – nem szerepelt a `[spam]` sztring.

Így a levelezőkliensem nem értékelte azokat spam-ként. Jellemzően orosz nyelvű levelekkel jártam így. A *D2S* tehát 99,23% pontossággal dolgozott, azonban ha szigorú vagyok,

6. ábra Alap beállítások



7. ábra Részletes beállítások



8. ábra Felhasználói fiók statisztika

és a problémás orosz spam-eket is hibának tekintem, akkor 97,04% eredményt ért el.
A *D2S* az utólagos kategorizálás céljára tárolta mind a 677 levelet, azonban a web felületen egy kattintás a *„Purge Folder”* gombra, és törölte az adott folder tartalmát. A *D2S* teljesítménye (*qemu*-ban futtatott *Outlook*-kal vizsgálva) másodpercenként átlag 1 levél volt.

További gondolatok

A *Bayesian* adatbázis felhasználói szempontból lehet egyedi, amikor minden felhasználó saját adatbázissal rendelkezik, illetve lehet globális, amikor a felhasználók egy központi adatbázison osztoznak. Véleményem szerint az előbbi a jobb megoldás, mert ebben az esetben rendelkeznek a felhasználók a legnagyobb szabadsággal, hogy eldönthessék számukra mi a spam. Bizonyos esetekben azonban ez az út nem célravezető. Lehet, hogy a felhasználók nagyobbik része nem igényel ilyen fokú szabadságot, vagy csak a lehető legkevesebb időt akar testreszabásra fordítani. Esetleg a rendszergazdák döntenek úgy, hogy nem éri meg a felhasználónkénti külön adatbázis. A D2S ez utóbbi megközelítést használja. Csak tájékoztatásképpen: a saját fejlesztésű spamszűrőmhöz készített, 2-3000 ham-ből és körülbelül ugyanennyi spam-ből készített redukált adatbázis mérete eléri a 13 MB-t.

Fentebb már utaltam arra, hogy a spammerek igyekeznek olyan szavakat a levélbe illeszteni, amelyek megtalálhatóak a felhasználók ham leveleiben. Könnyen belátható, hogy ez meglehetősen reménytelen vállalkozás, ha mindenkinek saját adatbázisa van.
A *Bayesian* spamszűrők nagy előnye, hogy rendkívül kevés hibával dolgoznak, jellemzően 99% fölött van a pontosságuk, ami egyetlen téves kategorizálást jelent 100 levelenként. Csak az összehasonlítás végett, az *RBL* listák jó ha 20-30% pontosságot elérnek, de ezt is óriási téves pozitív hibaszázalék mellett teszik.
Eddigi tapasztalataim szerint a *Bayesian* szűrőkön az olyan kéretlen levelek csúszhatnak át, amelyekben az üzenet helyett csak egy web cím van – ahol a szöveg valójában olvasható – és néhány teljesen semleges, az adatbázisban nem található szó. Azonban ezek ellen sem reménytelen a védekezés, például egy robottal, amelyik letölti az adott *URL*-t, és átadja annak tartalmát *Bayesian* elemzésre.

Záró gondolatok

A *D2S* az eddig ismertetett spamszűrők közül (*postgrey*, *razor*) a legjobb hatékonysággal fogta meg a kéretlen leveleket. Többek – például *Paul Graham* – szerint a *Bayesian* szűrők jelentik a végső megoldást a spam-re. *Jonathan Zdziarski* – a *DSPAM* készítője – *„Ending spam”* címmel írt egy könyvet, amelyben szintén a *Bayesian* elv előnyeit ecseteli. Magam is egy *Bayesian* spamszűrőn dolgozom, és meg tudom erősíteni, az ilyen elven működő szűrők valóban rendkívül hatékony megoldást nyújtanak a kéretlen levelekre, ellentétben például az *RBL* listákkal, amelyekről *Josh Mehlman* csak úgy nyilatkozott (➲ *http://techupdate.zdnet.com/techupdate/stories/main/0,14179,2873539,00.html*), hogy az ma már – mint megoldás – több problémát okoz, mint maga a spam. A *D2S* azonban egy hónap ingyenes teszt után olcsó éves díjért óriási segítséget ad ahhoz, hogy végre valóban a munkánkra koncentrálhassunk a haszontalan levelek olvasása és törlése helyett. Egy próbát mindenképpen megér.

Sütő János
(jsuto@freemail.hu)
1997 óta használ Slackware Linux-ot. Szabadidejében a postfix clapf nevű vírus- és spamszűrőjét polírozza.