Vylaďte si svůj SpamAssassin (2)

Miroslav Bartošek, ÚVT MU

Antispamový filtr SpamAssassin instalovaný na většině poštovních serverů na univerzitě je velmi užitečným pomocníkem v boji proti nevyžádané poště – spamu. Přestože jeho účinnost při odhalování spamu je velmi vysoká (90-95%), může i relativně malá část prošlých spamů znepříjemňovat uživateli život. Přitom je poměrně snadné vyladit si svůj SpamAssassin téměř na 100% účinnost – aniž by k tomu uživatel potřeboval být programátorem a znal detailně principy fungování antispamového filtru. Stačí jen naučit se vytvářet jednoduchá doplňková filtrační pravidla a vědět, jak a kam tato pravidla zapsat. K tomu má posloužit tento článek.

Místo úvodu dovolte, abych posloužil příkladem z vlastní praxe: Asi před čtyřmi lety jsem v rámci jedné své přednášky o službách sítě Internet uváděl, že spam představuje velkou hrozbu pro e-mailovou komunikaci, protože běžný uživatel může být brzy zahlcen stovkami spamů denně a e-mail se tím stane pro něj prakticky nepoužitelným. Dnes přichází na mou e-mailovou adresu kolem 180 spamů denně – přesto to však efektivitu mé e-mailové komunikace nijak neohrožuje. To je však možné jen díky účinnému antispamovému filtru SpamAssassin [1], [2]. Ten při svém standardním nastavení automaticky zachytí (v mém případě) asi 92% veškeré nevyžádané pošty. Do poštovní schránky by tak proniklo zhruba jen 15 spamů denně. Jakkoliv jde již o množství ručně dobře zvládnutelné, představuje i ono po čase poměrně otravnou záležitost. Vytvořením několika jednoduchých osobních filtračních pravidel se podařilo dále snížit počet nezachycených spamů – v průměru na jeden denně (aniž by přitom docházelo k jakýmkoliv falešným indikacím).

Protože skladba spamů každého uživatele (stejně tak jako posuzování toho, co spamem je a co nikoliv) je do značné míry individuální, může být snaha o centrální vylaďování filtračních pravidel kontraproduktivní – zvyšuje se tím riziko selhání, kdy jako spam je označena i zpráva, o kterou uživatel nechce přijít. Řešením je vytváření osobních filtračních pravidel a nastavení šitých na míru každému jednotlivému uživateli, přesněji – na míru jeho neodhaleným spamům. Nejefektivnější je, pokud si uživatel dokáže vytvářet takováto pravidla sám, dle svých konkrétních potřeb.

1 Jak pracuje SpamAssassin – stručná rekapitulace

V článku [3] byly popsány základní principy práce a možnosti uživatelského nastavení filtračního programu SpamAssassin. Připomeňme, že SpamAssassin aplikuje pro rozpoznání případného spamu obsáhlou sadu pravidel – předdefinovaných statických testů a dynamických (Bayesovských) filtrů – přičemž každé pravidlo má určité bodové ohodnocení a pokud se skutečně uplatní, je toto bodové ohodnocení připočteno k celkovému skóre příslušné zprávy. Pokud celkové skóre zprávy dosáhne stanovené hranice (implicitně je nastavena na hodnotu 5), je zpráva označena jako spam. Bodová ohodnocení jednotlivých pravidel jsou nastavena tak, že k identifikaci spamu nestačí žádné pravidlo samo o sobě, vždy se jich musí současně uplatnit více. To snižuje riziko nesprávného označení spamu a komplikuje život spamerům, kteří se pokouší takový systém obelstít.

SpamAssassin může také spolupracovat s řadou mezinárodních služeb, které shromažďují od tisíců uživatelů po celém světě informace o zprávách považovaných za spamy nebo o serverech, které takovéto zprávy rozesílají. Zprávě indikované těmito službami přidělí SpamAssassin opět určité bodové ohodnocení. I toto ohodnocení je nižší než celkový limit pro spam, takže samo k prohlášení zprávy za spam nestačí (SpamAssassin se tak při využívání externích antispamových služeb chová podle hesla – „důvěřuj ale prověřuj“).

Vzhledem k tomu, že spameři se snaží antispamové filtry přelstít a přichází neustále s novými a rafinovanějšími typy spamů, je vhodné nespoléhat jen na sady statických testů (jejichž obměna vždy nutně za vývojem nových spamů zaostává), ale využívat i prvky „umělé inteligence“ implementované v programu SpamAssassin v podobě Bayesovských filtrů. Pro spolehlivou funkci Bayesovských filtrů je třeba SpamAssassin „doučit“.

Doučování správného rozpoznávání spamů probíhá tak, že uživatel čas od času předkládá programu vzory spamů (zejména těch, které program sám nedokázal jako spam odhalit) a současně také i vzory ne-spamů, tj. zpráv, které si uživatel v žádném případě nepřeje označovat za spam (pro tyto zprávy se vžilo označení „ham“). K tomu, aby program začal využívat Bayesovských filtrů, je třeba mu nejprve v rámci úvodního doučování předložit alespoň 200 zpráv typu spam a 200 zpráv typu ham (podrobněji k doučování viz [3]).

2 Kam zapisovat lokální nastavení

Pokud si chceme vytvářet své osobní filtry, musíme vědět, kam je zapisovat. Poté, co váš počítačový správce nainstaloval program SpamAssassin na váš poštovní server a aktivoval ho pro vaši e-mailovou schránku (viz opět [3]), byl ve vašem domovském adresáři na poštovním serveru vytvořen podadresář `.spamassassin` a v něm soubor `user-prefs` .

Právě do tohoto souboru můžete zapisovat svá lokální nastavení a osobní filtry – aniž byste tím ovlivnili, jak bude program SpamAssassin fungovat pro ostatní uživatele na témže serveru.

3 Základní nastavení

Z experimentů s programem SpamAssassin vyplynuly následující zkušenosti ohledně nastavení základních parametrů ovlivňujících citlivost detekce spamu:

3.1 Nastavení hodnoty skóre

Ačkoliv parametr `required_hits` umožňuje změnit limit celkového skóre potřebného k prohlášení zprávy za spam, spíše bych tuto variantu moc nedoporučoval. SpamAssassin používá obsáhlou sadu pravidel s velmi vybalancovaným bodovým ohodnocením. To je nastaveno konzervativně tak, aby minimalizovalo *chybně pozitivní indikace* (tj. filtr označil a zachytil jako spam zprávu, která ve skutečnosti spamem nebyla), i za cenu případně vyššího počtu *chybně negativních indikací* (tj. filtr nerozpoznal spam a propustil ho k uživateli jako normální zprávu). Nedoručení očekávané zprávy může mít totiž závažnější důsledky než propuštění spamu.

Snížení hranice vyžadované pro indikaci spamu (pod implicitních 5 bodů) může sice vést k zachycení více spamů, ale za cenu nežádoucího zvýšení chybně pozitivních indikací.

3.2 ohodnocení Bayesovských filtrů

I když Bayesovské filtry rozpoznají spam s nejvyšší možnou mírou pravděpodobnosti, tj. 99-100%, přiřadí příslušné pravidlo `BAYES_99` zprávě pouze 3.5 bodu. To znamená, že ačkoliv SpamAssassin pečlivě doučujete a ten si je v daném konkrétním případě spamem prakticky jistý, k označení zprávy za spam vůbec nemusí dojít! K tomu je třeba, aby „zapracovala“ ještě další pravidla s celkovým součtem ohodnocení alespoň 1.5 bodu. Získat přitom takovýto počet bodů (zejména u nových typů spamů) se nemusí vždy podařit. Pokud tedy SpamAssassin průběžně doučujete, je vhodné zvýšit bodové ohodnocení pravidla `BAYES_99` například na 4 body (přidáním řádku `score BAYES_99 4` do souboru `user_prefs`). Toto ohodnocení je již dostatečně vysoké na to, aby ve spojení například s indikacemi od externích antispamových služeb zachytilo řadu spamů, které by jinak prošly. Míra chybně pozitivních indikací se tím nezvýší.

4 Osobní filtrační pravidla

U spamů, které nebyly antispamovým filtrem zachyceny, lze často identifikovat nějakou jednoduchou společnou vlastnost – například text zprávy obsahuje vždy určité slovo nebo se ve zprávě vyskytuje webová adresa (URL) obsahující určitý řetěz znaků. Je velmi jednoduché vytvořit osobní filtrační pravidlo, které takové zprávy rozpozná a přiřadí jim stanovené bodové ohodnocení.

Například zjistíme, že přes SpamAssassin prochází spamy nabízející akcie různých firem, v nichž se obvykle vyskytuje slovo „stock“ (akcie). Pokud jsme si jisti, že námi očekávané zprávy takovéto slovo obsahovat nebudou, lze ho považovat za indikaci spamu. Do souboru `user_prefs`

(třeba na samý konec souboru) zapíšeme následující pravidlo:

```
body STOCK /stock/
describe STOCK Includes \
     string stock (**MBa**)
score STOCK 2
```

První řádek říká, že začíná pravidlo s názvem STOCK, které v těle e-mailu (body) hledá řetěz znaků „stock“. Druhý řádek popisuje text hlášení, které se při rozpoznání spamu zapíše do výpisu v záhlaví zprávy; toto záhlaví obsahuje seznam pravidel, na jejichž základě byl spam rozpoznán (viz níže). [1] Třetí řádek určuje bodové ohodnocení pravidla STOCK – pokud se pravidlo uplatní, přičtou se v tomto případě k celkovému skóre zprávy 2 body.

Pokud je SpamAssassin nastaven tak, aby zachycené spamy ukládal do vyhrazené poštovní schránky, je možné kontrolovat, podle kterých pravidel byla ta která zpráva identifikována jako spam. Začátek těla zprávy je na obrázku 1.

4.1 Regulární výrazy

V pravidlech se nemusí vyskytovat jen jednoduché řetězce znaků, lze do nich zapisovat i regulární výrazy, pomocí nichž lze specifikovat podstatně složitější podmínky pro hledání v textu. Například po čase zjistíme, že spameři propagující akcie často maskují slovo „stock“ řetězem „st0ck“. V tomto případě vylepšíme první řádek našeho pravidla tak, že v něm uvedeme regulární výraz specifikující řetěz „stock“ nebo řetěz „st0ck“:

```
body STOCK /stock|st0ck/
```

Často se také v daném řetězu znaků může vyskytovat různá kombinace malých a velkých písmen (Stock, STOCK, sToCk, aj.). Přidáme-li na konec regulárního výrazu znak „i“, pravidlo bude necitlivé na velká či malá písmena. Dalším vylepšením může být přidání příkazu \b na místa, kde se má vyskytnout začátek a konec slova. Umožní nám to zadat regulární výraz identifikující slovo „stock“ ale již nikoliv třeba slovo „stocking“. Vylepšené pravidlo necitlivé na malá/velká písmena a zachycující pouze celá slova „stock“ (nikoliv řetězy začínající tímto podřetězem znaků) bude mít následující úvodní řádek:

```
body STOCK /\bstock\b|st0ck/i
```

O tom, jak vytvářet regulární výrazy, se lze dočíst například v seriálů článků pro začátečníky [4].

4.2 Hledání řetězu znaků v URL

Jiným typem spamu, který mne dlouho obtěžoval, byly e-maily nabízející nákup hodinek po internetu. Spamer v tomto případě neustále rafinovaně měnil texty a zejména kritické slovo „hodinky“. Používal varianty jako „watch“, „chronometr“, „clockwork“, „wrist jewelery“ a jiné, takže nebylo možné vytipovat spolehlivě řetěz znaků, pomocí něhož by bylo možné tento typ spamu identifikovat. Přesto měly všechny spamy něco společného – odkaz na webovou stránku pro nákup vnucovaného zboží. Ačkoliv i tyto odkazy se postupně měnily, vždy obsahovaly doménu 2.úrovně „geocities“ (měnit domény je pro spamera podstatně složitější než obměňovat slova v textu). Pak již bylo hračkou vytvořit osobní pravidlo hledající výskyt daného řetězu znaků v URL vyskytujícím se v těle zprávy (ať již v jeho textové nebo HTML-verzi):

```
uri GEOCITIES /geocities./
describe GEOCITIES Includes \
   URI with geocities. (**MBa**)
score GEOCITIES 3
```

Možností a variant pro vytváření osobních filtračních pravidel i způsobů jejich využití je spousta. Například přidělením záporného skóre danému pravidlu je možné naopak určitý druh zpráv preferovat a řešit tak problém chybně pozitivních indikací u příslušného typu zpráv. Velmi čtivý, nenáročný a přitom užitečný úvod do tvorby osobních filtračních pravidel pro SpamAssassin lze nalézt na [5]. Nicméně již znalost samotných základů uvedených v tomto článku může uživateli výrazně pomoci řešit problém nezachycených spamů. Tak jako v mém případě, kdy již pouhá dvě pravidla STOCK a GEOCITIES uvedená výše pomohla na dlouhou dobu návrat do šťastných dob, kdy e-mail sloužil lidem výhradně k užitečné komunikaci a nikoliv k obtěžování nevyžádanou poštou.

[1] Jako součást hlášení je uveden i sufix (**MBa**), který v tomto případě naznačuje, že se jedná o vlastní pravidlo autora článku.

```
Content preview:  On the Rise Newsletter - November Issue, 2005 In ths
  issue we are going to profile a company involved in the Red Hot homeland
  security sector. Also recently entering the Oil/Energy Industry! This
  company's stock is very much undervalued considering [...]

Content analysis details:   (7.2 points, 5.0 required)

 pts rule name              description
---- ---------------------- --------------------------------------------------
 2.0 STOCK                  BODY: Includes string stock (**MBa**)
 4.0 BAYES\_99              BODY: Bayesian spam probability is 99 to 100\%
                            [score: 1.0000]
 1.2 RCVD\_IN\_BL\_SPAMCOP\_NET RBL: Received via a relay in bl.spamcop.net
```

Obrázek 1: Záhlaví zachyceného spamu

Literatura

[1] SpamAssassin. Domovská stránka programu. http://spamassassin.apache.org/

[2] M. Kolaja, M. Bartošek. *Jemný úvod do (anti)spamové problematiky*. Zpravodaj ÚVT MU. ISSN 1212-0901, 2002, roč.12, č.5, s.1-6

[3] B. Moučka. *Vylaďte si svůj SpamAssassin.* Zpravodaj ÚVT MU. ISSN 1212-0901, 2005, roč.15, č.5, s.8-12

[4] Regulární výrazy. Seriál článků na root.cz. Dostupné na http://www.root.cz/clanky/regularni-vyrazy-1/

[5] M.Kettler. *A straightforward guide to writing your own add-on rules for SpamAssassin*. Dostupné na http://mywebpages.comcast.net/mkettler/sa/SA-rules-howto.txt

□