Spamassassin 3.0

Neděle, 25. září 2005 |

tagy: Spamassassin

Tomáš Klos představuje spolehlivého zabijáka spamu.

Nevyžádaná pošta (tzv. spam) se stala v poslední době vážným problémem, který se dotýká prakticky každého, kdo používá ke komunikaci elektronickou poštu. Záplava nabídek povzbuzujících prostředků, skvělého softwaru za pár korun nebo úžasných výher zná asi každý majitel e-mailové adresy. Počet takovýchto zpráv často několikanásobně převyšuje množství normální (vyžádané) pošty, v případě firemní komunikace může jít i o několik tisíc zpráv denně. Proto vzniklo mnoho způsobů, jak se spamu bránit.

Open-source antispamový systém Spamassassin zahrnuje většinu těchto metod, čímž (spolu se schopností učit se) dosahuje jeho úspěšnost v odhalování spamu hodnoty blížící se 100 %. Spamassassin používá systém bodování podle výsledků jednotlivých částí testu. Výsledný součet bodů vede k označení zprávy jako vyžádané, či nevyžádané. Pokud je některá zpráva označena chybně, jednoduše lze Spamassassin "naučit", jak má tuto zprávu v budoucnu hodnotit.

Mezi základní způsoby obrany patří odmítání zpráv ze známých spamerských serverů, kontrola adresy odesílatele, vyhledávání klíčových slov v předmětu a těle zprávy, kontrola formátu (HTML formátování, obrázky, skrytý text) a bayesiánská analýza.

V současné době existuje Spamassassin ve verzi 3.0.2 a je možné si jej stáhnout ze stránek projektu. Většina hlavních linuxových distribucí obsahuje předkompilované balíčky, které již mají vyřešeny potřebné závislosti. Celý systém se skládá z několika částí, které používá ke svému chodu, mj. démon `spamd`, klient `spamc` a perlové třídy `Mail::SpamAssassin`.

Spamassassin lze používat i přímo, pomocí příkazu `spamassassin`. Vzhledem k tomu, že se jedná o perlový skript, je jeho činnost mnohem pomalejší, než využití démonu, který stále běží na pozadí a klientu, který s démonem spolupracuje.

Nasazení na firemní poštovní server

Předpokládejme, že spravujeme poštovní server středně velké firmy se 150 e-mailovými účty. Máme funkční poštovní doručovací systém, nakonfigurovaný pro spolupráci s procmailem a nainstalovánu poslední verzi Spamassassinu včetně spouštěcího skriptu. Nejprve zajistíme spouštění démonu po startu systému:

```
# chkconfig spamd on
```

případně vytvoříme symbolický link do jednoho z adresářů `/etc/rc.d/` (podle typu našeho systému).

Dále vytvoříme databázi pro bayesiánskou analýzu. K tomu potřebujeme minimálně 200 spamů a 200 vyžádaných zpráv (čím více, tím lépe). Ty uložíme do dvou adresářů, např `~/spam` a `~/ham`. Můžeme použít i klasický unixový mbox, v tom případě příkaz spustíme s parametrem `-mbox` a zadáme cestu k mbox souboru. Nejprve naučíme Spamassassin, které zprávy obsahují znaky spamu:

```
# sa-learn -spam ~/spam/*
```

Podobné je to v případě vyžádané pošty, které se označuje jako tzv. ham:

```
# sa-learn -ham ~/ham/*
```

Nyní upravíme konfigurační soubor `local.cf`, který se nachází v adresáři /etc/mail/spamassassin. Zadáme cestu k souborům s bayesiánskou databází, která je uložena vždy v domovském adresáři uživatele v podadresáři ./spamassassin. Jelikož v našem případě spouštíme Spamassassin pod rootem, je databáze umístěna v jeho adresáři. (Spamassassin ke svému plnohodnotnému běhu nepotřebuje být spouštěn pod účtem superuživatele. Pokud jej chcete spustit pod jiným uživatelem, použijte při spouštění démonu `spamd` parametr `-u` spolu se jménem tohoto uživatele.)

```
bayes_path /root/.spamassassin/bayes
```

Spamassassin umí spolupracovat s databází, což se hodí hlavně při provozu webhostingového nebo freemailového serveru, kdy uživatelé přistupují ke svým schránkám pomocí webového rozhraní a nemohou tedy ukládat nastavení do lokálních profilů. Podrobnosti o možnostech nastavení Spamassassinu s využitím databáze najdete v README.

Další zajímavý parametr konfiguračního souboru je:

```
required_hits 5
```

který určuje, kolik bodů musí zpráva minimálně obdržet, aby byla označena jako spam. Díky úpravě tohoto parametru můžeme toto kritérium zpřísnit, nebo zmírnit. Výchozí hodnota je 5 bodů a z praxe mohu doložit, že je to hodnota optimální.

Spustíme démon `spamd`:

```
# /etc/init.d/spamd start
```

Nyní máme nastaveno vše podstatné pro rozpoznávání spamu, ale mailový systém o tom, že má použít spamfilter, zatím nic neví. Aby se začaly doručované zprávy filtrovat, upravíme soubor `procmailrc` v adresáři `/etc.` (Vzhledem k tomu, že se jedná o firemní poštu, pro kterou platí globální pravidla, použijeme pro všechny uživatele společný konfigurační soubor procmailu.)

Existují i jiné možnosti než použití procmailu, například maildrop nebo AMaViS.

Nejprve zjistíme, zda již byla na zprávě kontrola provedena, což poznáme podle zápisu v hlavičce zprávy. Pokud ne, pošleme ji klientovi `spamc`:

```
:0hbfw: spamassassin.lock

* !^X-Spam-Checker-Version:

| /usr/bin/spamc
```

Jakmile nám `spamc` zprávu vrátí, zkontrolujeme, jak byla ohodnocena. Pokud jako spam, uložíme ji do určeného adresáře, v našem příkladu do `/var/spool/mail/spamy`:

```
:0

* ^X-Spam-Flag: YES

/var/spool/mail/spamy
```

Jelikož se někdy může stát, že zpráva není označena jako spam, protože celkový součet bodů nebyl

dostačující, ale podle bayesiánské analýzy zpráva dostala maximum bodů, což znamená, že se jedná o naučený spam, přesuneme ji rovněž do archivu:

```
:0

* ^*tests=BAYES_99*

/var/spool/mail/spamy
```

Protože zprávy označené jako spam nezahazujeme, ale ukládáme je do adresáře, můžeme (například pomocí jednoduchého bashového skriptu) uložené zprávy prohledávat a v případě nesprávného označení upravit bayesiánskou databázi pomocí příkazu `sa-learn`.

Do spolupráce při vytváření kvalitní databáze spamu lze zapojit všechny firemní uživatele. Můžeme například vytvořit speciální e-mailovou adresu spam@vasedomena.cz, na kterou budou uživatelé přeposílat nevyžádanou poštu, která prošla do jejich mailboxu. Pak stačí nastavit cron tak, aby v pravidelných intervalech spouštěl příkaz `sa-learn` a učil se z tohoto mailboxu nové typy spamů, což může správci serveru ušetřit spoustu práce.

Pokud Spamassassin zprávu některého příjemce tvrdohlavě označuje jako spam, i když jde ve skutečnosti o vyžádanou poštu, můžeme ji z kontroly vyjmout pomocí parametru `whitelist_from` v konfiguračním souboru `local.cf`:

```
whitelist_from jmeno@domena.cz
```

Takto je možné vyjmout z kontroly jednu nebo více adres, nebo zprávy z celé domény:

```
whitelist_from *@domena.cz
```

Pokud chceme zadat více adres, je možné zapsat parametr vícekrát na samostatné řádky, nebo adresy napsat za sebou, oddělené mezerou.

Naopak - pokud chceme jako spam označit zprávy z konkrétní adresy nebo domény, i když zrovna typické známky spamu nenesou, můžeme použít volbu:

```
blacklist_from jmeno@domena.cz
```

Možnosti jsou stejné jako u parametru `whitelist_from`. Ostatní volby a možnosti naleznete v dokumentaci.

Klasický spam ale není jediný nešvar, se kterým si Spamassassin dokáže poradit. Jako vedlejší produkt jeho činnosti je i odfiltrování některých typů zpráv, kterými se šíří viry. V případě virů jde vlastně rovněž o nevyžádanou poštu, a proto také obsahuje některé její znaky. Je tedy možné, že se v archivu nevyžádané pošty objeví zprávy s viry, i když spolu se Spamassassinem běží na serveru některý antivirový systém. Existují dvě možnosti, jak se to může stát: buď ještě antivirový systém neobsahoval aktualizaci, která by mu umožnila virus zachytit, nebo je řetězec kontroly pošty nastaven tak, že první na řadu přichází Spamassassin, který zavirovanou zprávu označí jako spam a antivirový systém se tudíž vůbec nedostane ke slovu. Každopádně díky této příjemné vlastnosti Spamassassin zvyšuje účinnost antivirové ochrany.

Nasazení na pracovní stanici

Na běžné linuxové pracovní stanici se dostáváme do odlišné situace. Systém může využívat více

uživatelů, zprávy budou stahovány přímo z externího POP nebo IMAP serveru. Bude tedy třeba využít schopnosti spolupráce našeho poštovního klientu se Spamassassinem.

Samozřejmě je možné provádět stahování pošty například fetchmailem a dále postupovat podobně jako u firemního poštovního serveru. Nelze však předpokládat, že pracovní stanici budou vždy využívat natolik pokročilí uživatelé, aby si mohli sami konfigurovat fetchmail a procmail, případně že bude vždy po ruce někdo, kdo to provede za ně.

Mezi nejvyužívanější poštovní klienty patří zcela jistě KMail (v prostředí KDE) a Novell (dříve Ximian) Evolution (v prostředí GNOME). Oba tyto majoritní programy ve svých aktuálních verzích podporují spolupráci se Spamassassinem pomocí vlastních filtrů. Princip filtrování je obdobný, příchozí zpráva se pošle Spamassassinu, který ji zkontroluje, označí a pošle zpět poštovnímu klientu. Ten se zprávou naloží podle výsledku testu, pokud nebyla jako spam označena, doručí ji do inboxu. Pokud jako spam označena byla, provede předvolenou akci - buď zprávu vymaže nebo ji uloží na určené místo. Nakonec se spustí `sa-learn`, který aktualizuje bayesiánskou databázi. Pokud je jako spam označena vyžádaná pošta, je možné ji označit jako ham, a tím naučit bayesiánský filtr, že se v tomto případě jedná o vyžádaný typ zprávy. Stejně tak v opačném případě, když není spam označen jako nevyžádaná zpráva. Z dalších klientů, které podporují kontrolu spamu ve spolupráci se Spamassassinem, jmenujme například Sylpheed-Claws, tuto funkčnost je však třeba nainstalovat jako plugin. Nastavení a ovládání filtrů je ve jmenovaných programech natolik intuitivní, že není třeba je detailně popisovat.

Poštovní klient Mozilla Thunderbird používá vlastní antispamový systém, který funguje na podobných principech jako Spamassassin. Využívá rovněž principů bayesiánské analýzy, takže je schopen učit se, jeho úspěšnost při rozpoznávání nevyžádané pošty je velmi vysoká. Je součástí programu, takže není třeba nic zvlášť instalovat. Rovněž ovládání je jednoduché a intuitivní. Navíc je možné nastavit jej tak, že nikdy nebude považovat za spam poštu zaslanou z adres obsažených v adresáři, což je jakási obdoba použití whitelistu. Dá se také naučit, aby jako spam odfiltrovával zavirované zprávy.

Instalace Spamassassinu na pracovní stanici je prakticky stejná, jako instalace na poštovní server. Pokud půjde o některou z rozšířených desktopových distribucí, například Mandrakelinux, stačí zadat:

```
# urpmi spamd spamc
```

Nainstalují se všechny potřebné balíčky. Poté zajistíme spouštění démonu po startu systému:

```
# chkconfig spamd on
```

a nakonec jej spustíme:

```
# service spamd start
```

Nyní nám již nic nebrání nastavit svůj poštovní klient, začít učit Spamassassin rozpoznávat nevyžádanou poštu, a tím ji definitivně odstranit ze světa naší elektronické komunikace.

Odkazy

- Stránky projektu
- Popis spolupráce Spamassassinu s databází
- Popis konfiguračních voleb
- On-line generátor konfiguračního souboru Spamassassinu