Ještě ke spamu

Spam je dnes tak aktuálním tématem, že jsme předcházející rozsáhlý článek převzatý z německého Chipu rozšířili i o příspěvky našich autorů (některé další najdete i v příštím Chipu).

Spam obecne a v Unixe

Začiatky spamu boli pomerne zvládnuteľné a nevyžiadané správy sa rozširovali najmä cez konferencie USENET. Keď začiatkom druhej polovice 90. rokov prestal byť internet obmedzený najmä na akademickú sféru, dostali sa do popredia jeho dve služby:

* WWW založená na HTTP protokole;
* e-mail fungujúci cez protokol SMTP.

Pre koncového užívateľa je tiež dôležitý protokol POP3, umožňujúci vzdialený prístup k poštovým schránkam. Neskôr sa pretlačil do popredia IMAP4 ako náhrada pre POP3. Obidve služby pracujú na pomerne jednoduchých, ale flexibilných protokoloch. Nanešťastie poskytujú e-mailové protokoly aj dostatok flexibility pre spamerov. V čase ich vzniku nikto na problém spamu nemyslel, a preto je teraz nevyhnutné obzerať sa po prostriedkoch, ktoré tieto protokoly doplnia o účinnú obranu proti spamu.

Ako pracujú spameri

Spameri pri svojej činnosti používajú čoraz sofistikovanejšie techniky. Princíp ich práce je ale zakaždým rovnaký - musia si zaobstarať e-mailové adresy, získať e-mailové servery, z ktorých sa spam bude posielať, a potom už len treba spam poslať. E-mailové adresy sa dajú najľahšie zaobstarať prehľadávaním webu. Prakticky každá stránka obsahuje nejakú kontaktnú informáciu. Na poslanie spamu sú treba vhodné mailservery, najlepšie sú tie, ktoré majú povolený relaying - akceptujú správy z hocakej domény a pošlú ich ďalej. Tieto sa dajú pomerne ľahko získať systematickým skenovaním IP adries v internete. Zaujímavý je fakt, že na pozberanie e-mailových adries a nájdenie vhodných serverov na poslanie správ stačí skript s veľkosťou zopár riadkov. Na internete je dokonca dostupný software vykonávajúci túto činnosť. Poslanie správ na zoznam adries je potom triviálnou úlohou.

Ako sa dá spam zastaviť

Kontaktne informácie

Hocakým poskytnutím kontaktných informácií (na webe) sa vystavujeme riziku zaradenia do Spam nerozoznavá OS, je problémom pod každým z nich databázy spamera. Medzi známe triky patri zverejnenie kontaktu prostredníctvom obrázku či "znetvorenie" adresy tak, aby bol problém s jej automatickým rozpoznávaním.

Software

Na automatické triedenie spamov od normálnych e-mailov existuje niekoľko kategórií softwaru. Pri ich výpočte skúsim spomenúť ich najznámejšie unixovské implementácie. Všetky by sa mali dať nájsť cez vyhľadávač www.freshmeat.net.

Obsahové filtre. Na základe obsahu e-mailov sa dá pomerne ľahko usúdiť, či je správa spam, alebo či je hodná čítania. Správy obsahujúce určité textové reťazce (napr. "Make Money Fast") tak môžu byť zaradené do špeciálnych adresárov.

Funkcionalitu filtra môže vykonávať buď priamo doručovací agent, ako napr. procmail, alebo môže byť prenechaná klientovi.

Whitelist. Veľmi spoľahlivo fungujúca, značne agresívna technika, ako zabrániť spamom dosiahnuť poštovú schránku, je nechať prísť len správy od explicitne povolených odosielateľov. Keď príde správa z neznámej adresy, pošle sa na túto adresu verifikačný e-mail, vyzývajúci odosielateľa odpovedať naň, ináč nebude pôvodná správa doručená. Najznámejší nástroj implementujúci túto techniku pre Unixy je asi TMDA.

Blacklist. Koncept blacklistov je veľmi jednoduchý - počíta s tým, že jeden a ten istý spam je doručený viacerým užívateľom. Jedná sa o distribuovanú sieť, v ktorej stačí, aby jeden užívateľ označil správu za spam, a správa bude u každého iného užívateľa siete klasifikovaná ako spam. Známe nástroje sú pyzor či razor.

Blacklisty nie sú tak spoľahlivé, ako by bolo vhodné, odporúča sa ich používať v kombinácii s inými metódami.

Hodnotenie správ - kritériá. Každú správu je možné klasifikovať na základe určitých kritérií. Aj spamy spĺňajú určité kritériá, napríklad často sú ich subjekty písané veľkými písmenami, obsahujú Javascript či sú písané v HTML. Ak správa spĺňa určitý počet týchto kritérií, dá sa považovať za spam. SpamAssassin, veľmi známy software tejto kategórie, robí presne spomínanú činnosť - kontroluje prichádzajúce správy a pomerne presne vie odhadnúť, či sa jedná o spam. Jeho činnosť sa dá skombinovať s využívaním blacklistov.

Bayesian. (viď ďalej)

Martin Užák

HISTÓRIA SPAMU

Hoci sa na spam pozerá ako na problém až od konca minulého desaťročia, prvé nevyžiadané správy sa datujú do dôb, keď sa internet len formoval.

1971 - V duchu hippies poslal administrátor MIT cez CTTS hromadnú správu "THERE IS NO WAY TO PEACE, PEACE IS THE WAY".

1978 - Prvý internetový (vtedy ešte vlastne arpanetový) spam. Všetkým adresám zo západného pobrežia USA bolo poslané pozvanie na prezentáciu nového počítača DEC-20 do Kalifornie.

1988 - Na Usenete sa objavila správa, v ktorej žiadal autor o zaslanie peňazí.
Od začiatku 90. rokov sa začali rozširovať správy typu "Make Money Fast".

1993 - Chybný software poslal do usenetovej konferencie news.admin.policy asi 200 správ po sebe. Nevyžiadané správy boli vtedy pravdepodobne prvýkrát označené pojmom spam.

1994 - Tento rok bol pre históriu spamu mimoriadne zaujímavý. Začiatkom roka bol každej usenetovej konferencii poslaný e-mail so subjektom: "Global Alert for All : Jesus is comming soon". Jednalo sa o prvé hromadné zneužitie siete tohto druhu. O niekoľko mesiacov neskôr nasledovala správa "Green Card Lottery - Final one". Takisto bola poslaná do každej konferencie a je pomerne známa.