Spam – co s ním?

Miroslav Bartošek, ÚVT MU

Elektronická pošta, e-mail, je jedním z obrovských fenoménů posledního desetiletí. Podstatně zefektivnila komunikaci a přispěla tak k výraznému zvýšení produktivity práce ve všech oblastech. Každý, kdo ještě zažil ve své práci předinternetovou éru, mi dá jistě za pravdu, že rozdíl mezi rychlou e-mailovou komunikací s kýmkoliv-kdykoliv-kdekoliv na světě a klasickou papírovou/dopisní korespondencí je nebetyčný. A to nejen při tak komunikačně intenzivních akcích, jako je například pořádání konferencí či odborná spolupráce rozsáhlých týmů, ale i v běžném každodenním životě. Bez e-mailu si lze dnes již stěží představit práci akademika – ale i úředníka, fungování univerzity, ale třeba i osobní komunikaci většiny z nás. E-mail přitom používáme nejen k vlastní komunikaci a výměně dokumentů; e-mailová schránka nám čím dál více slouží také jako pohotový osobní diář, adresář, poznámkový blok, úkolník a efektivní archiv. Prostě bez e-mailu se většina z nás již neobejde.

1 E-mail v ohrožení

Přesto je ale tento efektivní nástroj a každodenní nepostradatelný pomocník stovek miliónů lidí na celém světě v permanentním ohrožení. Tím, kdo na něj se stále zvyšující se intenzitou útočí, a hrozí až jeho totálním ochromením, je SPAM – nevyžádaná pošta. Podle některých odhadů tvoří spam dnes až 90 % veškeré e-mailové komunikace! Devět z deseti e-mailových zpráv nepřináší svým příjemcům žádnou užitečnou či očekávanou informaci. Zaplavuje je naopak spoustou obtěžujících nabídek, připravuje je o čas, peníze a energii. Většina z nás dostává stovky, ti komunikačně čilejší až tisíce, spamů denně! Naštěstí je většina z nich vyřazena antispamovými filtry, takže mezi skutečně doručenou poštu pronikne jen malá část, ale i ta dokáže pěkně komplikovat život. Někdy nám naopak důležitá zpráva nepřijde vůbec, protože ji antispamové filtry chybně vyhodnotily jako spam, a zpráva tak skončila v záplavě tisíců nevyžádaných zpráv na spamových skládkách.

Spam nekomplikuje život jen koncovým uživatelům. Stal se noční můrou počítačových správců a lidí zodpovědných za fungování komunikační infrastruktury organizace. Na centrální poštovní server Masarykovy univerzity přicházelo počátkem roku 2007 téměř milión e-mailových zpráv denně – a tento počet se stále rychle zvyšuje[1]. V průměru deset e-mailů za sekundu, nepřetržitě ve dne i v noci, v pracovní dny i o víkendu, musí být tímto serverem přijato, vyhodnoceno, zkontrolováno ohledně přítomnosti virů, otestováno v rámci ochrany proti spamům, a poté přeposláno na příslušná místa, buď dovnitř univerzity (na fakultní poštovní servery) nebo ven z MU. Pokud by server zkolaboval nebo přestal neustávající příval pošty zvládat, pocítí to okamžitě každý z nás. Proto je nezbytné infrastrukturu univerzitních poštovních serverů průběžně výkonnostně posilovat a vybavovat stále sofistikovanějšími ochrannými nástroji. To stojí peníze a zvyšuje nároky na kapacity specializovaných odborníků.

Mezi spammery a strážci užitečného fungování elektronické pošty zuří permanentní tichá válka. Na každý tah jedné strany reaguje okamžitě protitah druhé strany. Každé opatření proti spamům vede dříve či později k protiopatřením spammerů, která se snaží účinnost zavedených antispamových opatření eliminovat – buď novými technikami maskování spamů nebo razantním zvýšením jejich počtu. Přestože nová antispamová opatření zvýší procento zachycených spamů, vyšší počet útočících spamů znamená i vyšší počet spamů, které tato opatření překonají.

2 Nárůst počtu spamů

Řada uživatelů zřejmě zaznamenala v roce 2006 dramatický nárůst počtu spamů. Čím je tento nárůst způsoben? Zatímco v dřívějších letech byly hlavním zdrojem spamů buď poštovní servery spammerských organizací nebo zneužité nezabezpečené poštovní servery běžných organizací, v současnosti používají spammeři i nové, účinnější postupy. Využívají nedostatečně zabezpečené počítače koncových uživatelů, které infikují a vytváří z nich stroje k rozesílání vlastní

[1] Po zavedení greylistingu v březnu 2007 stoupl denní počet zpráv již na 1,7 miliónu.

pošty. Existují dokonce specializované skupiny hackerů, které nedostatečně zabezpečené osobní počítače cíleně vyhledávají, přeměňují na *zombie* (stroje pod svou vlastní kontrolou) a zapojují je do velmi rozsáhlých sítí *botnets* – některé z nich mají až stovky tisíc strojů. Ty pak prodávají spammerům. Aniž by to daný uživatel tušil, počítač, na kterém doma pracuje, rozesílá současně spamy do celého světa. Spolu se zvyšujícím se počtem domácností vybavených počítači a stále se zlepšujícími parametry jejich připojení na síť (vysokorychlostní Internet) tak roste i potenciál pro zvyšování počtu rozeslaných spamů.

Podle [3] je až 7 % počítačů na Internetu infikováno a hackeři jsou schopni celosvětově získat a přetvořit na zombie více jak 100 000 osobních počítačů týdně. Obrovské a výkonné sítě botnets pak pracují pro spammery téměř bezplatně (na náklady nevědomých majitelů PC) a chrlí do Internetu miliardy spamů denně. I kdyby byla jejich výtěžnost mizivá – řekněme jedna ku miliónu (jen na každý miliónty spam by některý z oslovených „zákazníků“ zareagoval a nabízené zboží/službu si koupil), znamená to pro spammery dostatečný příjem na to, aby se jim jejich „podnikání“ vyplácelo.

3 Vícestupňová ochrana

Jak jsme uvedli již v úvodu, bez soustavných spamových protiopatření by e-mail již dávno přestal být použitelnou službou. Problém je v tom, že žádné protiopatření nevydrží dlouho, není definitivní. Neustále je třeba přicházet s novými technologiemi. Protože žádná z nich nemůže být sama o sobě stoprocentně úspěšná, je třeba tyto technologie kombinovat a vytvářet vícestupňové ochrany.

Problém je i v samotném vymezení spamu; v řadě případů je objektivní identifikace spamu nemožná či nejednoznačná. Ten samý e-mail, který jeden uživatel považuje za obtěžující, může pro jiného uživatele představovat důležitou informaci. Proto musí být antispamové filtry na vyšších stupních ochrany (celouniverzitní, fakultní) nastaveny rozumně konzervativně, aby nenadělaly víc škody než užitku. A proto je také žádoucí doplnit tyto vyšší stupně ochrany osobními filtry přizpůsobenými konkrétním koncovým uživatelům – jejich osobním preferencím a charakteru jejich elektronické pošty. Teprve pak může být antispamová ochrana skutečně účinná.

4 Prevence

Nejefektivnější způsob, jak se vyhnout spamům, je neumožnit spammerům získat vaši e-mailovou adresu. To znamená zacházet s ní jako s citlivým údajem, který není radno sdělovat kdekomu na potkání. Tím je myšleno vyhnout se jak aktivnímu tak pasivnímu zveřejňování e-mailové adresy v prostředí Internetu. Pod aktivním zveřejňováním rozumíme uvádění e-mailové adresy v různých on-line formulářích, u nichž si nemůžete být jisti seriózností přijímající strany, v otevřených diskusních fórech a v inzerátech, nebo odpovídání na spamy. Stejně tak nebezpečné je ale i pasivní zveřejnění nechráněné e-mailové adresy na webových stránkách. Vyhledávání a sklízení e-mailových adres pomocí automatizovaných sběračů (spambots) z volně přístupných webových stránek a z diskusních fór je jeden z nejčastějších způsobů, jak spammeři přijdou na vaši adresu[2]. Podle výzkumů [4] potřebují dnes spammerské vyhledávače v průměru 19 dnů k tomu, aby odhalily e-mailovou adresu nově zveřejněnou na webu (nejkratší reakce byla do 1 sekundy po vystavení adresy).

5 Způsoby boje proti spamu

V současnosti existuje a průběžně se rozvíjí řada přístupů k boji proti spamu. Z principiálních důvodů (hranice mezi spamem a chtěnou poštou jsou individuální a nelze je jasně definovat, technologie pro rozesílání a maskování spamů se průběžně zdokonalují) nemůže být žádný z nich sám o sobě zcela účinný. Dokáží však poměrně spolehlivě zachytit převážnou část spamu. Při kombinaci různých přístupů a jemném vyladění zohledňujícím jednotlivé uživatele se může dobrá antispamová ochrana blížit až ke stoprocentní účinnosti. Důležité je však nejen co nejvyšší procento zachycených spamů; ještě

[2] Z těchto důvodů jsou všechny e-mailové adresy na veřejném webu `http://www.muni.cz` maskovány, tj. zabezpečeny proti rozpoznání a sklízení automatizovanými roboty.

možná důležitějším kritériem z pohledu koncového uživatele je co nejnižší (nejlépe nulový) počet případů, kdy je jako spam vyhodnocena a nedoručena regulérní zpráva.

Hlavní směry v boji proti spamu jsou následující:

- filtrování podle obsahu zprávy;
- filtrování podle odesílatele;
- ekonomické přístupy;
- legislativní přístupy.

Uvedené směry se v praxi vzájemně kombinují a doplňují.

5.1 Filtrování podle obsahu

Podstatou tohoto přístupu je analýza obsahu těla zprávy, její hlavičky či obou těchto částí. Při analýze jsou vyhledávány určité známé charakteristiky spamů – v těch nejjednodušších přístupech jsou to výskyty podezřelých slov či znaků, v propracovanějších přístupech jsou hledány obecnější charakteristiky v širším kontextu, založené na statistikách a umělé inteligenci. Pokročilejší nástroje, jako je například spamassasin, nespoléhají pouze na jeden algoritmus či sadu pravidel. Aby se zpráva vyhodnotila jako spam, musí se současně sejít více podnětů, z nichž každý může mít nastavenu jinou váhu. Takovouto ochranu spammeři hůře překonávají a současně se snižuje riziko chybné identifikace dobré zprávy za spam.

Pro vlastní analýzu obsahu jsou používány různé technologie: jednoduchá filtrační pravidla, klasifikace využívající strojového učení, systémy založené na kompresních technikách, vyhledávání podobností v textu či obrazu.

Jako protizbraň proti filtrování obsahu používají dnes spammeři často obrazová sdělení, jejichž analýza je obtížnější a méně propracovaná než v případě textů. Vlastní tělo zprávy pak obsahuje žádný nebo jen neutrální text pro zmatení filtrů. Samotná informace je převedena do obrazové podoby a uložena v příloze e-mailu.

5.2 Filtrování podle odesílatele

U těchto způsobů filtrace nejde o to, co zpráva obsahuje, ale kdo ji poslal. Tradiční přístupy využívají seznamy špatných adres – blacklists – z nichž je příjem pošty zablokován, a oproti tomu seznamy důvěryhodných adres – whitelists – z nichž je naopak zpráva přijata vždy, bez ohledu na její obsah. Blacklisty mají tři zásadní slabiny: (a) adresu odesílajícího stroje lze zfalšovat; (b) s tím, jak se rozesílání spamů přesunulo z pevných serverů na zneužité koncové stanice zombie, není pro spammery problém rozesílající stroje rychle měnit; (c) jestliže se hackerům podaří zneužít některý počítač či adresu uvnitř důvěryhodné organizace, může se celá tato organizace dostat na blacklist a následně do velkých potíží, protože nevinné e-maily od jejích uživatelů mohou někteří příjemci odmítat (do této situace se čas od času dostává i MU). Proto je dobré používat blacklisty s rozumem – pouze jako pomocné doplňkové kritérium, nikoliv jako kritérium jediné a rozhodující. Řada poštovních správců využívání blacklistů odmítá zcela.

Mezi novější technologie filtrování podle odesílatele patří *greylisting* (viz článek Mirka Rudy v tomto čísle Zpravodaje). Vychází z toho, že zombie pro rozesílání spamů se nechovají jako standardní poštovní servery – rychle rozešlou kvantum zpráv a tím to pro ně končí, mizí aby nebyly odhaleny. Naproti tomu standardní poštovní servery jsou konstruovány tak, aby zajistily spolehlivé fungování pošty i při běžných provozních potížích, kdy se často nemusí podařit doručit příjemci zprávu hned napoprvé (například proto, že příjemcův poštovní server je dočasně mimo provoz). Při greylistingu jsou příchozí zprávy přijímajícím poštovním serverem nejprve vždy odmítnuty, a teprve ty zprávy, které po stanovené době přijdou znovu, jsou skutečně doručeny adresátům. Na rozdíl od blacklistů nemá greylisting žádný fixní seznam „špatných hochů“ – všechny považuje za rovnocenné a rozhoduje se až podle jejich konkrétního chování, nikoliv podle jejich adresy.

Jiný přístup využívají systémy založené na *spolehlivém prokázání identity odesílatele*. Využívají často kryptografické techniky, jejichž cílem je prokázat příjemci, že odesílatel zprávy je skutečně ten, za kterého se vydává. Přijímány jsou například pouze zprávy s ověřeným digitálním podpisem nebo jiným důvěryhodným znakem od odesílatele, počítače nebo domény. Problémem je zatím nižší míra standardizace a penetrace

potřebných technologií i globální bezpečnostní infrastruktury. Uvedený přístup však má vysokou potenciální účinnost; na rozdíl od jiných přístupů necílí na odstraňování následků, ale přímo na léčení základní příčiny spamu, kterou je nedostatečně zabezpečená infrastruktura elektronické pošty a Internetu obecně.

Další z přístupů spadají do kategorie tzv. *systémy úkol/odpověď* (challenge/response). Základní idea spočívá v tom, že pokud důvěryhodnost odesílatele není příjemci známa, zašle poštovní server příjemce odesílateli určitý úkol (challenge) a teprve po jeho úspěšném vyřešení je e-mail příjemci skutečně doručen. Úkol je navržen tak, aby byl snadno splnitelný člověkem, nikoliv však počítačem (čtenář se již asi setkal s takovými úkoly v podobě obrázků obsahujících slovo zapsané různě zpotvořenými písmeny). Tímto způsobem se příjemce může bránit proti zprávám rozesílaným softwarovými roboty.

5.3 Ekonomické přístupy

Tyto přístupy se snaží změnit ekonomickou základnu samotné existence spamu. Tou je fakt, že spammera dnes rozesílání i obrovského množství e-mailů téměř nic nestojí (na rozdíl třeba od klasických papírových letáků, jejichž netriviální cenu musí zaplatit inzerent – a navíc je tato cena vždy přímo úměrná množství letáků). Čili cílem je zavést taková opatření a technologie, aby se masové rozesílání spamů ekonomicky nevyplácelo. Jednou z možností je zavedení malých poplatků za odeslanou poštu (s případnou refundací prokazatelně seriózním zákazníkům). Tento model může být atraktivní zejména pro velké poskytovatele služeb elektronické pošty, naráží ale na různé politické a společenské bariery, a neřeší také problém spamů ze zneužitých počítačů.

Určitou ekonomickou zpětnou vazbu je možné zavést i jiným způsobem než přímými finančními platbami. Jednou z možností je modifikace systémů typu úkol/odpověď (viz výše), kdy přijímající poštovní server zvyšuje odesílajícímu serveru cenu za rozesílání pošty tím, že příjem zprávy podmiňuje vyřešením výpočetně náročného úkolu u odesílajícího serveru. Při malém počtu posílaných e-mailů tato dodatečná zátěž nevadí, při hromadném rozesílání pošty však již ano.

5.4 Legislativní přístupy

Řada vyspělých zemí světa přijala v posledních letech zákony, které hromadné rozesílání nevyžádaných zpráv omezují a ty nejhorší spammerské praktiky přímo zakazují. Antispamová legislativa existuje i u nás, ale například také v USA (CAN-SPAM Act 2003), odkud dnes pochází celosvětově nejvíce spamu. Již to samo o sobě demonstruje, nakolik jsou legislativní opatření v praxi skutečně účinná. Anti-spamové zákonodárství je určitě velmi potřebné a důležité, protože vymezuje právní rámec, umožňuje slušným firmám používat Internet pro reklamní účely v rámci daných pravidel a naopak trestně stíhat spammery. Je však zřejmé, že problém spamu nevyřeší, a že hlavní tíha boje proti spamu leží v oblasti technologií.

Literatura

[1] J.Goodman, G.V.Cormack, D.Heckerman. *Spam and the Ongoing Battle for the Inbox*. Communication of the ACM, Vol. 50, No. 2, February 2007. Pro uživatele MU dostupné elektronicky na `http://doi.acm.org/10.1145/1216016.1216017`

[2] B.Leiba, N.Borenstein. *A Multifaceted Approach to Spam Reduction*. Proceedings of the Conference on Email and Anti-Spam, 2004. Dostupné elektronicky na `http://www.ceas.cc/papers-2004/127.pdf`

[3] M.Furst. *Botnets. No. 1 emerging Internet threat.* CNN, January 31, 2006. `http://www.cnn.com/2006/TECH/internet/01/31/furst/`

[4] Project Honey Pot. `http://www.projecthoneypot.org` □