

Skoro všechno, co jste kdy chtěli vědět o spamu, ale báli jste se zeptat

Trocha historie

První zdokumentovaný nevyžádaný komerční e-mail¹ – pozvánka na prezentaci nových systémů DEC – se objevil v síti Arpanet 3. května 1978.ⁱ Sice se mu ještě neříkalo „spam“, ale reakce na něj byla většinou stejně odmítavá. Termín „spam“ se začal používat mnohem později, poprvé zřejmě 31. března 1993 v reakci na 200 příspěvků, zasláných do usenetové skupiny news.admin.policy díky chybě v programu, jehož účelem mělo být naopak spamu bránit.

Označení „spam“ vzniklo díky slavnému skeči Monty Python's Flying Circus.ⁱⁱ Ujalo se tak dobře, že se dnes v původním významu už téměř nepoužívá – dokonce ani jinak poměrně dobrý slovník Seznamu původní význam neuvádí vůbec a Merriam-Webster ho má na posledním třetím místě. Jako antonymum k výrazu „spam“ se někdy používá „ham“.

Navzdory odmítavým reakcím postižených se množství spamu s rostoucím počtem uživatelů Internetu stále zvyšuje a podle různých odhadů činí dnes zhruba 70-80% veškerého e-mailového provozu.ⁱⁱⁱ Zdá se, že přestože většina uživatelů spamy ignoruje a maže, stále se ještě najde nějaké to promile uživatelů nových a nepoučených nebo blbých, kteří zareagují. Právě a jen díky nim se spamování stále ještě vyplácí.

Spamy můžeme rozlišovat v zásadě na dva hlavní proudy:

- 1) zpravodaje – spammer se snaží vás přesvědčit, že jeho informace jsou pro vás užitečné, že je chcete, často dokonce tvrdí, že jste se k odběru těchto informací přihlásili sami; často se tváří jako pozvánky nebo pravidelné zpravodaje, chodí z existujících adres, odkazují na existující webové stránky patřící odesílateli a často také tvrdí, že se můžete odhlásit, což ovšem zdaleka není vždy celá pravda (viz e-mail harvesting níže).
- 2) nabídka produktů a propagace idejí – nabízejí nejčastěji Viagra, Valium a podobné medikamenty, případně informují o brzkém návratu Spasitele, o možnosti úžasné výhodné investice do akcií, které zakrátko jistě významně stoupnou apod. Odesílatel těchto spamů svou totožnost tají, odkazují na weby cizí nebo hacknuté.

Jakési rozdíly existují i v přístupu k cílové skupině. Zatímco někteří spammeři někde na konci spamu žádají o váš souhlas a slibují, že už vám nic dalšího nepošlou, pokud nebudete explicitně souhlasit (tzv. opt-in), ve velké většině vám odesílatel poměrně arogantně sdělí, že máte velké štěstí, že můžete tak zajímavé informace dostávat zdarma přímo do své mailové schránky a jestli je nechcete, musíte se odhlásit (tzv. opt-out). Podle českého práva (blíže viz §7 zákona č. 480/2004 Sb.^{iv}) jsou však oba způsoby protiprávní (druhý z nich je sice přípustný, ovšem jen za zákonem stanovených podmínek), váš souhlas musí totiž odesílatel získat ještě před zasláním prvního komerčního sdělení.

Spam dnes otravuje život prakticky každému uživateli Internetu. Návrhy, jak se mu lépe bránit, se objevují snad každý den, nicméně málokterý je skutečně aplikovatelný. Pokud také máte nápad, jak to dělat lépe, než ho někde publikujete, doporučujeme projít následující formulář standardní odpovědi na takové nápady, abyste pak nebyli zbytečně nepřijemně překvapeni (zdroj: /.):

Your post advocates a

technical legislative market-based vigilante approach to fighting spam. Your idea will not work. Here is why it won't work:

- Spammers can easily use it to harvest email addresses
- Mailing lists and other legitimate email uses would be affected
- No one will be able to find the guy or collect the money
- It is defenceless against brute force attacks
- It will stop spam for two weeks and then we'll be stuck with it
- Users of email will not put up with it
- Microsoft will not put up with it
- The police will not put up with it
- Requires too much cooperation from spammers
- Requires immediate total cooperation from everybody at once
- Many email users cannot afford to lose business or alienate potential employers
- Spammers don't care about invalid addresses in their lists
- Anyone could anonymously destroy anyone else's career or business

Specifically, your plan fails to account for

- Laws expressly prohibiting it
- Lack of centrally controlling authority for email
- Open relays in foreign countries
- Ease of searching tiny alphanumeric address space of all email addresses
- Asshats
- Jurisdictional problems
- Unpopularity of weird new taxes
- Public reluctance to accept weird new forms of money
- Huge existing software investment in SMTP
- Susceptibility of protocols other than SMTP to attack
- Willingness of users to install OS patches received by email
- Armies of worm riddled broadband-connected Windows boxes
- Eternal arms race involved in all filtering approaches
- Extreme profitability of spam
- Joe jobs and/or identity theft
- Technically illiterate politicians
- Extreme stupidity on the part of people who do business with spammers
- Dishonesty on the part of spammers themselves
- Bandwidth costs that are unaffected by client filtering
- Outlook

And the following philosophical objections may also apply:

- Ideas similar to yours are easy to come up with, yet none have ever been shown practical
 - Any scheme based on opt-out is unacceptable
 - SMTP headers should not be the subject of legislation
 - Blacklists suck
 - Whitelists suck
 - We should be able to talk about Viagra without being censored
 - Countermeasures should not involve wire fraud or credit card fraud
 - Countermeasures should not involve sabotage of public networks
 - Countermeasures must work if phased in gradually
 - Sending email should be free
 - Why should we have to trust you and your servers?
 - Incompatibility with open source or open source licenses
 - Feel-good measures do nothing to solve the problem
 - Temporary/one-time email addresses are cumbersome
 - I don't want the government reading my email
 - Killing them that way is not slow and painful enough
- Furthermore, this is what I think about you:
- Sorry dude, but I don't think it would work.
 - This is a stupid idea, and you're a stupid person for suggesting it.
 - Nice try, asshole! I'm going to find out where you live and burn your house down!

1 Unsolicited Commercial E-Mail (UCE), popř. Unsolicited Bulk E-Mail (UBE) – běžné formální označení pro spam

Nicméně ať se jedná o opt-in nebo opt-out přístup, ve většině případů to je podvod – tzv. e-mail harvesting neboli sběr e-mailů. Pokusem o odhlášení dáváte spammerovi jednoznačnou informaci, že vaše adresa je platná a používaná. Takto ověřené adresy pak spammer používá pro další spamování, nebo je i prodává jiným spammerům.

Ochrana proti spamu

Čas od času se ozývají lidé, kteří nejsou o škodlivosti spamu přesvědčeni a obranu proti němu dokonce označují za formu cenzury. Osobní preference netřeba rozebírat, pokud někdo touží denně přečíst 150 nabídek na Viagra, jistě mu v tom není třeba bránit. Většině uživatelů spam vadí a mají pro to dobré důvody, ty ale nejsou předmětem tohoto elaborátu, na rozdíl od metod, jak jim od něj lze pomoci.

Každý antispamový produkt dnes používá celou řadu postupů, kterými postupně hodnotí procházející mail a přiděluje mu jakési kladné a záporné body. Spammeri zase na druhé straně analyzují dostupné antispamové produkty, z nichž se celá řada šíří s otevřenými zdrojovými kódy, a snaží se přijít na metody, kterými je lze obelstít. Tvůrci antispamů pak zase vylepšují své produkty, aby nové druhy spamů dokázaly odhalit a spammeri zase zkoumají, jak je znovu obelstít a tak se to děje stále znovu a znovu.

Podobně, jako v případě antivirů, je proto třeba antispamovou ochranu neustále aktualizovat a zvolit si takový produkt, který rychle reaguje na aktuální vývoj. Pro domácí uživatele a malé firmy většinou postačí lokální antispam na pracovní stanici, např. integrovaný v rámci poštovního klienta (poskytuje např. Mozilla Thunderbird), nebo ve formě pluginu (např. pro The Bat! nebo Microsoft Outlook). Větší organizace volí hlavně z důvodu zjednodušení a centralizování správy systémů takové antispamové produkty, které se instalují na poštovní brány nebo servery (např. SpamAssassin, QSF, nebo komerční produkty, např. IronPort, Barracuda, Brightmail apod.).

K dosažení co nejlepšího efektu při antispamové kontrole může pomoci také kombinace více antispamových produktů. Zejména komerční antispamy v sobě většinou integrují i mnohé další funkce, např. mohou provádět antivirové kontroly, mohou v procházejících mailech hledat předem definované výrazy, prepisovat nebo doplňovat hlavičky i obálky, duplikovat je, přidávat disclaimery a v neposlední řadě je na základě výsledků kontrol zařazovat do karantén a umožnit jednotlivým uživatelům snadnou správu karantény třeba prostřednictvím webového rozhraní.

Běžné metody boje proti spamu

- 1) „Reputace“ odesilatele – databáze IP adres, z nichž v minulosti byly posílány spamy, popř. obecná databáze, která každé známé odesílací adrese přiřadí podle různých kritérií nějaké hodnocení pravděpodobnosti, zda z ní odeslaný mail bude legitimní, nebo spam. Příkladem mohou být služby RBL (Real-time Blackhole List), provozované často bezplatně, nebo komerční databáze, jako např. SenderBase spol. IronPort. Adresy se do těchto databází dostávají buď proto, že z nich v minulosti odcházela spam, nebo proto, že jsou na nich např. špatně zkonfigurované poštovní servery (tzv. open relays²). Na základě reputace poštovní brány mohou spojení z podezřelé adresy rovnou odmítnout, pokud poštu přesto přijmou, mohou jí přiřadit záporné hodnocení použité při dalším počítání pravděpodobnosti spamu, mohou také zároveň omezit přijímání pošty z dané adresy např. velikostí, nebo max. počtem přijatých zpráv za hodinu apod.
- 2) Greylisting – poštovní brána si udržuje databázi zdrojových IP adres mailů korelovaných se zdrojovou a cílovou e-mailovou adresou. Když přijde pošta, jejíž kombinace IP adresy a mailových adres v hlavičkách již byla zaznamenána, propustí ji hned, když ale přijde pošta, jejíž identifikace v databázi není, její přijetí odmítne způsobem, který odesílající server interpretuje jako tzv. soft error, tj. dočasný problém na straně příjemce. Zatímco standardní mailové servery v takovém případě poštu zkouší doručit po několika minutách znovu, spamovací systémy se často o další doručení nepokoušejí. Normální pošta je proto doručena (poprvé s mírným zpožděním), zatímco spamy přes greylist projít vůbec nemusí.
- 3) Uživatelské seznamy – uživatelé mají možnost definovat seznamy mailových adres nebo domén, které antispam vždy má zahodit (blacklist), nebo naopak dále nekontrolovat a nechat projít (whitelist).
- 4) Kontrola dodržování pravidel a konvencí – software používaný spammery často odesílá poštu trochu jinak, než to požadují příslušná RFC a nebývá zkonfigurovaný podle běžně užívaných konvencí (např. zpětně resolvable adresa odesilatele shodná s některým MX záznamem pro odesílatelovu doménu,

2 Open Relay – označení pro poštovní server, zkonfigurovaný tak, že přijímá a přeposílá dál poštu z jakékoli adresy na libovolnou jinou adresu. Takto nastavené servery bývaly často zneužívány spammery k rozepisování spamů, dnes už prakticky vymizely a objevují se jen krátkodobě, většinou díky chybě administrátora.

pipelining bez povolení přijímajícího serveru apod.). Zkoumáním dodržování pravidel a konvencí nelze odhalit spam úplně bezpečně, protože zdaleka ne každý poštovní server je správně zkonfigurovaný. Záleží na administrátorovi na straně příjemce, zda se rozhodne zprávy odmítat na základě takových chyb.

5) Analýza textu zprávy – lze ji rozdělit na následující části:

- Analýza pomocí pravidel – předdefinované kontroly na přítomnost známých identifikátorů spamů, např. struktury pošty, přítomnosti HTML tagů v textovém mailu, nadměrného počtu velkých písmen, křiklavých barev a zvláštních fontů, názvů produktů propagovaných spamy a celé řady dalších parametrů.
- Statistická srovnávací analýza – filtrování založené na databázi výskytů termínů v mailech, které uživatel v minulosti označil jako spam, nebo naopak jako ham.
- Databáze odkazů – spamy, které chtějí vydělávat na šíření nějakého produktu přímo, musí dát obětem možnost, jak kontaktovat prodejce. Ve spamu proto musí být nějaký odkaz na web, nebo adresa odesilatele musí být platná, aby bylo možno zaslat na ni odpověď. Zejména komerční produkty si vytvářejí databáze těchto odkazů. Když pak jejich přítomnost odhalí v textu zprávy, lze ji s velkou jistotou označit za spam.
- Textová a/nebo podobnostní analýza přiložených obrázků – analýzu textu lze provádět poměrně snadno, proto spammeři své maily často naplní náhodně vybranými texty (záměrně tím také snižují efektivitu zejména statistické analýzy textů) a vlastní spam pošlou ve formě obrázkových příloh. Proto v poslední době antispamy zahrnují i moduly, které se snaží pomocí OCR převést text z obrázků na jeho textovou reprezentaci a tu dále analyzovat běžnými způsoby, popř. srovnávají obrázky s databází obrázků z dříve zachycených spamů.

6) Speciální dočasná pravidla – někteří výrobci poskytují zatím poměrně unikátní službu, která spočívá v tom, že využívají zdroje informací o rozesílání mailů v reálném čase k identifikaci podezřelých mailů, které by mohly být např. důsledkem šíření nové virové nákazy. Tyto informace co nejrychleji zpracují a dají k dispozici systémům svých zákazníků tak, že tyto systémy mohou začít karanténovat podezřelé maily dlouho před tím, než jsou k dispozici aktualizace antivirových signatur. S tím, jak pokračuje zkoumání podezřelých mailů, zpřesňují se zároveň dočasná pravidla a dříve zadržené maily, které aktualizovaným pravidlům již nevyhovují, jsou průběžně z karantény uvolňovány. Obvykle se počítá s tím, že zhruba během 12 hodin je již aktualizována většina antivirovýchází a uvolní pak k dalšímu zpracování i zbytek podezřelých mailů.

Méně používané metody

Na rozdíl od metod popsaných výše, existují i metody další, možná dokonce účinnější, ale jsou málo rozšířené, protože vyžadují určitou nezvyklou aktivitu na straně odesílatelů, a to se ne vždy setkává s pochopením (pro důvody viz rámeček).

Dobrym příkladem takové metody je TMDA (Tagged Message Delivery Agent),^v který, když přijde e-mail od neznámého odesílatele, uloží tento mail do karantény a odesílateli odešle zprávu se speciální hlavičkou a vysvětlením, co je třeba udělat, aby původní pošta byla z karantény vyňata a doručena. V podstatě stačí, aby původní odesílatel na tento mail jen odpověděl, TMDA na straně příjemce rozpozná speciální hlavičku, kterou před tím použil a automaticky doručí původní mail z karantény do uživatelského mailboxu. Existuje i další podobné řešení (komerční), které odesílateli místo mailu se zvláštní hlavičkou zašle URL, na které je třeba kliknout, aby byl původní mail doručen.

Přístupy popsané výše znamenají pro příjemce mnohem menší zátěž, na druhou stranu ale z vlastní zkušenosti vím, že mnozí odesílatelé měli potíž přečíst si celý mail od TMDA a pochopit, co je třeba udělat, aby jejich původní mail byl doručen. Proto bych jejich použití doporučil jen jako doplněk k jinému antispamovému řešení, kde by se žádost o potvrzení zasílala nikoli každému novému odesílateli, ale jen těm, jejichž maily se zdají být spamy.

Drobnou vadou tohoto systému může být, že protože většina spammerů odesílá maily z adres, na které se nedá poslat pošta zpátky, odpovědi od TMDA zůstanou delší dobu „viset“ v odchozí frontě a pokud jich je opravdu hodně, mohly by způsobit zpomalení nebo zahlcení místní poštovní brány.

Jaký antispam zvolit

Spektrum dostupných produktů je velmi široké, od volně dostupných přes relativně levná až po velmi drahá řešení. Každé z nich má své výhody a nevýhody, proto se občas vyplatí je i zkombinovat (např. některé

nekomerční řešení s komerčním). Velkou výhodou komerčních řešení je, stejně jako u antivirů, kontinuální a rychlé aktualizování antispamových pravidel, která si tato řešení stahují běžně v pěti až desetiminutových intervalech. Díky tomu jsou schopny reagovat rychleji a kvalitněji jak na konkrétní nové spamy, tak i na novátorské přístupy spammerů, kteří neúnavně hledají stále nové způsoby, jak přes antispamy projít.

Hanuš Adler

autor je ředitelem společnosti

Actinet Informační systémy s.r.o.

- i Spam History: <http://www.templetons.com/brad/spamterm.html>
- ii Monty Python's Spam <http://www.youtube.com/watch?v=cFrtpT1mKy8>
- iii Většina mailů jsou spamy: <http://news.bbc.co.uk/1/hi/technology/5219554.stm>
- iv Zákon č. 480/2004 Sb. <http://www.uoou.cz/index.php?l=cz&m=left&mid=11:01&u1=&u2=&t=>
- v Tagged Message Delivery Agent <http://tmda.net/>