Posudek oponenta k diplomové práci

předložené na Matematicko-fyzikální fakultě
Univerzity Karlovy v Praze

Autor: Peter Sabolčák

Název práce: SPAM Wars

Studijní program a obor: Informatika, softwarové systémy

Rok odevzdání: 2008

Jméno oponenta: Dan Lukeš

Pracoviště: SISAL MFF UK

Práce přehledově mapuje problematiku nevyžádané, obtěžující elektronické pošty a některé metody obrany. Oponentura práce na toto téma je obtížná věc – v této oblasti neexistuje mnoho všeobecně přijímaných pravd ani postupů. Dokonce neexistuje ani obecně uznávaná definice předmětu zkoumání – SPAMu. Přesto si dovolím tvrdit, že definice uvedená v prvním odstavci kapitoly 2 je daleko širší než je obvyklé a kdyby implementace filtru vycházela z této definice, byl by výsledný filtr v praxi principielně nepoužitelný. Naštěstí se při návrhu implementace filtru autor zabývá odstraňováním dopisů podle jiného klíče. V rešeršní části by dále bylo snadné polemizovat s výčtem a zejména hodnocením závažnosti nedostatků té-které popisované metody, nicméně, v tomto případě jde o hodnocení z velké části subjektivní. I když bych já v mnohých případech přisoudil informacím jinou váhu než autor, a tudíž dospěl k odlišným závěrům, je nutné na tuto část pohlížet jako na kvalitní práci.

V implementační části si autor vybral konkrétní metodu, kterou se bude pokoušet bojovat proti nežádoucím příchozím emailům. Jednou z metod "challenge/response" se snaží u nečekaných zpráv odlišit zprávy automaticky generované od zpráv odeslaných živým člověkem a to variantou Turingova testu. O (ne)vhodnosti užívání podobných metod se vždy vedly velké spory, aniž by dospěly k nějakému všeobecně přijatému závěru. Autor uvádí, že první implementace challenge/response systémů se objevily v roce 2000. Fakt, že ani po osmi letech se podobné metody významněji nerozšířily, přičemž šíření neomezují významné investiční nebo provozní náklady, počítám za argument ve prospěch odpůrců těchto metod. Tato práce však má svůj smysl i v případě, že tato implementace představuje slepou nebo okrajovou větev řešení.

Filtr má, z povahy zvolené metody, potíže s přeposílanými zprávami a také diskusními listy (protože ty jsou nejčastějším případem žádoucího dopisu, který neodeslal živý člověk). Práce s těmito problémy bojuje statečně, ale faktem je, že jediným známým uspokojivým řešením této potíže je celosvětově přijatá změna v tom, jak se dopisy přeposílají – a to já vidím jako hlavní překážku tomu, aby filtrace touto metodou mohla někdy masověji rozšířit. Další potíží tohoto filtru je jeho nekompatibilita s existujícími a používanými metodami boje proti nežádoucí poště. Již několik desítek dopisů s padělanou adresou odesílatele, po jejichž příchodu filtr na tyto adresy odešle "challenge", může způsobit, že vzdálený systém takovou aktivitu vyhodnotí jako útok a tuto síť zablokuje. Mezi těmito dvěma sítěmi, používajícími každá svůj druh filtru, tak bude komunikace efektivně znemožněna a to v obou směrech. Nicméně, obě zmíněné vady, ač pro praktické nasazení zásadní, nejsou důvodem snižovat význam této práce. V oblasti ochrany se zkouší vše možné a není důvod nezkusit i tohle. Co

ale lze práci vytknout objektivně je, že se autor těmito problémy v textu vlastně vůbec nezabývá. Ačkoliv se implementace některé problémy snaží různým způsobem zmírnit, bez nahlédnutí do zdrojových kódů se nedozvíme jak dalece se autorovi podařilo problémy eliminovat, ba co hůř, čtenář této práce z řad osob, které nemají vlastní hlubší znalost této problematiky se ani nedozví, že metoda tyto principiální problémy má.

Filtr je určen výhradně pro postfix a tak bych z hlediska portability považoval za jazyk první volby ten, ve kterém je napsán postfix. Měl bych tak jistotu, že filtr bude opravdu přeložitelný všude, kde lze přeložit tento MTA. U PERLu se můžeme dostat do situace, kdy na konkrétním systému bude postfix dostupný, ale implementace PERLu pro něj k dispozici nebude. Nutnost instalace poměrně velkého interpreteru navíc zvyšuje náklady na instalaci, následně je třeba tuto instalaci udržovat, případně se zabývat bezpečnostními konsekvencemi instalace tohoto interpreteru (což vyžaduje přítomnost odborníka znalého právě tohoto jazyka) a tak je na místě zmínit i zvýšené náklady provozní. V neposlední řadě jsem přesvědčen, že pro síťový server tohoto typu je interpreter vždy horší volba než kompilátor. Mrzutá je i těsná vazba systému na SASL. Nejen v implementaci, ale ani v teoretické části práce se autor nezabývá možností, že by uživatelé mohli být autentizováni jiným mechanismem, přičemž takové systémy se dnes reálně používají. Nasazení filtru v takovém prostředí a tudíž vynucený přechod na SASL může na sítích s velkým množstvím uživatelů znamenat tak velké vstupní náklady, že je nasazení filtru velmi nepravděpodobné, nemluvě o možném nepohodlí pro samotné uživatele (další autentizační systém, tudíž, pravděpodobně, další heslo). Filtr sice nabízí "workaround" autentizace pomocí IP, ten je ale navržen především pro použití serverem bez uživatelů (byl zamýšlen pro WWW server s webmailem).

Pokud ale na implementaci hledíme jako na výzkumný projekt, mající za cíl otestovat možnost nasazení podobného systému jsou výše zmíněné námitky prakticky nepodstatné. V každém případě autor kvalitně naimplementoval to, co si implementovat předsevzal.

Nemohu ještě nezmínit fakt, že podle mého názoru má dílo zásadní právní vady. Autor při jeho vytváření použil řadu cizích pomocných komponent, z nichž většina je uvolněna pod licencí GPL. To je možné jen v případě, že i výsledné dílo je pod GPL. Tuto podmínku měl autor v plánu dodržet, problém je ale v tom, že diplomová práce je ze zákona "školní dílo" a jako takové umožňuje užití, které je s ustanoveními GPL nekompatibilní. Výsledné dílo tudíž není "pod GPL" a to anuluje autorovo právo legálně užívat všechny GPL pomocné komponenty. Na obranu autora je ale třeba říci, že se tím tato škola při výuce programování a věcí souvisejících s touto problematikou nijak nezabývá a v zásadě se omezuje na "pokud to efektivně řeší zadání, pak jste to navrhli dobře". Podobnými vadami trpí celá řada odevzdávaných studentských prací a tak by, ač jde o připomínku poměrně zásadní, neměla mít vliv na obhajitelnost ani výslednou známku.

Závěrečné testy jsou zajímavé čtení, podstatnou a cennou součástí práce, nicméně, omezují se výhradně na technické a zátěžové parametry filtru. V práci tak zcela chybí zhodnocení vlivu filtru na (ne)příchod (ne)vyžádaných dopisů do cílových schránek, což je škoda.

Přes uvedené výtky a připomínky lze práci počítat mezi ty kvalitnější. Práce by rozhodně měla být obhájena, přičemž za vhodnou známku považuji "2".

Datum: 16. září 2008

Podpis: