Posudek oponenta diplomové práce

Jméno diplomanta: Bc. Lubomír Jurečka
Téma práce: Distribuovaný systém pro detekci nevyžádaných zpráv
Oponent práce: ing. Ladislav Pešička, KIV

Diplomová práce Lubomíra Jurečky se zabývá problematikou detekce nevyžádaných zpráv. Cílem práce je vybrat vhodný způsob vytváření signatur e-mailových zpráv a dále navrhnout systém s centrálním úložištěm pro klasifikaci e-mailových zpráv dle četnosti výskytu.

Předložená diplomová práce má 75 stran a je rozdělena do deseti kapitol včetně dalších příloh. Diplomant se nejprve věnuje definici potřebných pojmů a rozdělení typů spamu. Následně popisuje dostupná anti-spamová opatření, včetně technik obcházení antispamových filtrů. Další kapitola zkoumá způsoby vytváření charakteristických signatur a hašovací techniky. Popis je dostatečně přehledný a vysvětluje nejdůležitější principy dané techniky. Drobnou výhradu bych měl k zařazení vlastního řešení 4.5, oddělil bych spíše přehled stávajících technik od vlastního návrhu novou kapitolou.

Diplomant navrhl vlastní způsob vytvoření otisku e-mailové zprávy na základě úvahy, cituji zkráceně autora *„Zpráva je ve skutečnosti pouze diskrétní signál o konstantní frekvenci a jednotlivé úrovně signálu tvoří znaky z ASCII tabulky; potom lze zprávu zkoumat metodami z oboru DSP“*. Tato myšlenka je dále podrobně rozvedena. Přesto zde postrádám analýzu, zda se někdo nepokoušel alespoň o podobný přístup k danému problému a dále (vzhledem k odlišnosti od běžných koncepcí) testování na rozsáhlejší množině e-mailových zpráv různého charakteru (spamy, regulární zprávy, virální reklamy) - zmiňován je korpus 8542 spamových zpráv. Přesto navržené řešení vítám jako inovativní a inspirativní pro další vývoj. Pouze rozsáhlé testy dokáží ověřit jeho účinnost.

V další části práce diplomant navrhl systém na architektuře klient-server pro shromažďování signatur zpráv a počítání četnosti podobných výskytů. Úkolem klienta je vlastní vytváření charakteristických signatur zpráv. Diplomant dále navrh vlastní komunikační protokol, vycházející z UDP protokolu pro komunikaci klient-server a využívající dále i HTTPS pro ověření identity. Přenášená data jsou šifrována blokovou šifrou AES-CBC.

Diplomant se správně zamýšlí nad tím, jaký význam má určení četnosti výskytu zprávy pro určení spamu v kapitole 5.4.1. Od čistě binárního členění přistupuje k stanovení dílčího spamového skóre, kdy se na výsledném ohodnocení může podílet více anti-spamových filtrů. S jeho závěry a přístupem souhlasím. Pro stanovení optimálních koeficientů převodních funkcí je nepochybně nutné praktické testování.

V další části práce diplomant realizuje programové vybavení navrženého systému. Pro tvorbu byl zvolen jazyk C++ a PHP pro webové rozhraní. Programy byly realizovány pro operační systém Debian GNU/Linux. Funkcionalita systému byla testována pro různé testové případy (výpadky linky, chybovost linky, zátěžové testy). V praxi bude systém sloužit jako další vrstva anti-spamové ochrany produktu Kerio Connect. Realizovaný systém je funkční. Diplomová práce představuje zajímavý způsob pro další možnosti boje s nevyžádanou poštou.

Na diplomanta bych měl následující otázku:

V případě nasazení systému se předpokládají centrální servery provozované Keriem,nebo i varianta, kdy případně každý klient bude moci nasadit vlastní instalaci? Šla by v takovém případě a měla by smysl synchronizace jednotlivých databází?

Diplomant splnil zadání diplomové práce, práci doporučuji k obhajobě a hodnotím ji klasifikačním stupněm:

Výborně

Ing. Ladislav Pešička
KIV, ZČU Plzeň

V Plzni, 13.6.2012