SJEDNOCENÍ DATABÁZÍ PRO NAČÍTÁNÍ LOGŮ ZE SMTP SERVERŮ

Jiří Pospíšil[1]

[1]*Ostravská univerzita v Ostravě, Přírodovědecká fakulta, Katedra informatiky a počítačů, r12326@student.osu.cz*

Abstrakt

Cílem této práce je vytvořit aplikaci, která zajistí načtení vybraných údajů z logů do jednotné databáze. Logy jsou vytvářeny antivirovými a antispamovými programy pro kontrolu elektronické pošty na Ostravské univerzitě v Ostravě. Aplikace načítá vybrané údaje z logů tvořených antivirovými a antispamovými aplikacemi pro kontrolu elektronické pošty na Ostravské univerzitě v Ostravě. Umožňuje načítat logy jednotlivě nebo po adresářích a zobrazuje souhrnné statistiky, které může exportovat do souboru ve formátu CSV. Logy jsou tvořeny třemi soubory. První soubor s názvem **maillog** je tvořen výstupem z aplikací Postfix, Policyd a modulů pro blacklisting Spamhaus a Sophos. Dalším je pak **amavislog**, který obsahuje data z programu Amavis a jeho modulů SpamAssasin a ClamAV. Posledním logem je soubor s názvem **message_log** (dále jen „**messagelog**“) s výstupem z programu Sophos PureMessage.

Popis zpracování zprávy

Postfix po kontrole na přítomnost v blacklistech přijatou zprávu předá programu Policyd (nekontroluje vnitřní poštu z rozsahů IP adres OSU) a po navrácení zkontroluje hlavičku a tělo zprávy a předá ji programu Amavis. Ten jej dále po kontrole předá programu Sophos PureMessage, která ji po kontrole opět vrací Postfixu k doručení. Jednotlivé programy mohou zprávu propustit dále, označit jako spam, přesunou do karantény (pro kontrolu administrátorem) nebo zamítnout.

Popis jednotlivých programů

Postfix – jedná se o tzv. MTA (Mail Transfer Program), který zajišťuje přenos elektronické pošty. Může obsahovat další moduly nebo předávat zprávy ke zpracování dalším programům. V tomto případě kontroluje hlavičku a tělo zprávy na vybrané výrazy a dle nich může zprávu pozdržet (pro kontrolu administrátorem) nebo odmítnout. Součástí Postfixu jsou moduly pro blacklisting a to projektu **Spamhaus** a firmy **Sophos**. Je-li IP adresa odesílatele nalezena na některém z těchto seznamů, je zpráva automaticky zamítnuta. Postfix předává zprávy aplikaci Policyd a po navrácení je předává dalším programům (Amavis a pak PureMessage).

Policyd – program provádí kontrolu nevyžádané pošty metodou zvanou greylisting. Tato metoda vychází ze specifikace poštovního protokolu SMTP. Odesílající MTA se v případě neúspěšného doručení zprávy pokusí o opětovné doručení. Programy odesílatelů nevyžádané pošty však častokrát tuto specifikaci nedodržují a snaží se o odeslání co možná největší množství zpráv v krátkém čase. Aplikace si při prvním pokusu o doručení zprávy uloží kombinaci e-mailovou adresu odesílatele, příjemce a IP adresu odesílatel (dále jen „triplet“) a zprávu odmítne (resp. ohlásí odesílateli dočasnou chybu). Při opětovném pokusu program vyhledá uloženou trojici a zprávu propustí k další kontrole. Po pěti úspěšně doručených zprávách je IP adresa odesílatele umístěna na automatický whitelist a propouštěna bez zpoždění.

Amavis – Jedná se o program, který především kontroluje přílohy a blokuje přílohy s nepovolenou příponou. Obsahuje moduly **SpamAssasin** a **ClamAV**, které také kontrolují přílohu.

Sophos PureMessage – Komerční produkt, který kontroluje zprávu a přílohy na přítomnost virů, trojských koní a dalších hrozeb.

Samotná aplikace

Je napsána v programovacím jazyce C++ s využitím frameworku Qt. Jako databázový systém používá server MySQL. Je možné ji spustit jak z konzole (s parametry), tak s grafickým uživatelským rozhraním (dále jen „GUI“). Oba přístupy umožňují zvolit jednotlivé soubory ke zpracování nebo celý adresář. GUI umožňuje navíc zobrazení souhrnných číselných statistik za vybrané časové období a výsledky exportovat do souboru ve formátu CSV s rozdělením po měsících, dnech nebo hodinách. Jednotlivé statistiky jsou rozděleny v záložkách dle jednotlivých logů. Kromě nich také obsahuje tři další záložky. První z nich umožňuje zobrazit deset nejčastějších e-mailových adres odesílatelů, příjemců a IP adres odesílatelů. Ve druhé pak lze zobrazit celkový počet zpráv, celkový počet zpráv označených jako spam nebo virus, počty označení jednotlivými programy a moduly, průměrnou dobu mezi prvním pokusem o doručení a průměrný počet pokusů v době úvodní prodlevy. Poslední záložka pak rovněž zobrazuje dobu zpoždění a počet pokusů v době úvodní prodlevy, tentokrát však s rozdělením do intervalů.

Pořadí načítání logů je maillog, amavislog a messagelog. U maillogu se rozpoznávají dva typy řádků, první pro Policyd a druhý pro Postfix. V případě Policyd se řádky se stavy whitelist_sender=update, whitelist=update, whitelist=bypass, greylist=new a greylist=bypass vloží ihned do databáze. U ostatních se aplikace pokusí vyhledat v databázi záznam pro shodný e-mail (na základě shody odesílatele, příjemce, IP adresy a dle vybraných stavů policyd a datumu a času). U řádků programu Postfix se nejdříve vyhledává podobný řádek (na základě shody odesílatele, příjemce, IP adresy a dle datumu a času). U obou typů řádků platí, že pokud je hledaný řádek nalezen, aplikace v databázi aktualizuje údaje, v opačném případě vloží nový záznam do databáze.

V amavislogu je situace podobná. Nejdříve se program pokusí najít podobný řádek v databázi (na základě shodných e-mailových adres odesílatele a příjemce, dále dle chybějícího amavis statusu a dle datumu a času). V případě nalezení takové řádku se aktualizují jeho údaje, jinak se vloží nový záznam do databáze.

V messagelogu se zpráva vyhledává na základě dvojice kritérií. Buď dle shodného identifikátoru zprávy (pmx_id) nebo dle e-mailových adres odesílatele a příjemce, dále dle chybějícího pmx_id a podle datumu a času. V případě nalezení takové řádku se aktualizují jeho údaje, jinak se vloží nový záznam do databáze.

V maillogu se pro řádek programu Policyd ukládají následující údaje: datum a čas, status, IP adresa serveru, hostname serveru, e-mailová adresa odesílatele a příjemce.

Pro řádek programu Postfix se ukládá datum a čas, e-mailová adresa odesílatele a příjemce, IP adresa serveru, status (zde důvod nedoručení zprávy), a server helo odesílatele.

V amavislogu se ukládá datum, čas, e-mailová adresa odesílatele a příjemce, IP adresa serveru, oba statusy, skóre programu Amavis a velikost v bytech.

V messagelogu se ukládá datum, čas, e-mailová adresa odesílatele a příjemce, IP adresa serveru, identifikátor zprávy (pmx_id), první dvě akce provedené se zprávou (pmx_action1, pmx_action2), odůvodnění akce (pmx_reason), pravděpodobnost spamu a velikost zprávy.

Poděkování

Děkuji vedoucímu práce RNDr. Tomáši Sochorovi, CSc. za cenné rady a pomoc při tvorbě práce a RNDr. Jarmile Holcové za technické konzultace a podporu. Tento příspěvek vznikl jako součást řešení projektu Studentské grantové soutěže (SGS15/PřF/2015), který je podporován Ministerstvem školství, mládeže a tělovýchovy.