Um modelo de confiança entre servidores de e-mail

Leonardo B. de Oliveira
PUCPR, PPGIA,
Curitiba PR, Brasil, 80.215-901
bispo@ppgia.pucpr.br

and

Carlos A. Maziero
PUCPR, PPGIA,
Curitiba PR, Brasil, 80.215-901
maziero@ppgia.pucpr.br

Abstract

E-mail services are essential in the internet. However, the standard e-mail architecture presents problems that opens it to several threats. Alternatives have been proposed to solve some problems related with e-mail services, offering reliability and escalability to such systems. This work presents a distributed trust model, allowing to create dynamic and decentralized trusted server lists, through the temporary exclusion of servers used to spread malicious messages. Techniques like social network models, message filters, message management, and a trust information storage and propagation model were used for building it.

Keywords: E-mail services, SPAM, trust networks.

Resumen

O e-mail é uma ferramenta essencial na Internet. No entanto, a arquitetura dos sistemas de e-mail convencionais apresenta limitações que deixam o sistema vulnerável a diversas ameaças. Alternativas vêm sendo propostas para sanar essas deficiências e prover maior confiabilidade e escalabilidade aos sistemas de e-mail. Este trabalho apresenta um modelo de confiança que permite criar listas de servidores confiáveis e não-confiáveis de forma dinâmica e descentralizada, através da exclusão temporária de servidores propagadores de mensagens maliciosas. Na proposta são utilizadas diversas técnicas, como um modelo de redes de relacionamento, filtros de mensagens e um modelo de gerenciamento, armazenamento e propagação de confiança.

Palabras clave: Serviços de e-mail, SPAM, redes de confiança.

1 Introdução

Os sistemas de e-mail são muito utilizados, por sua simplicidade, flexibilidade e baixo custo de implantação e uso. Todavia, os sistemas de e-mail atuais sofrem de problemas causados por fragilidades nos protocolos de comunicação envolvidos. Dentre eles se destacam a falta de mecanismos robustos de autenticação, a falta de mecanismos de privacidade, confiabilidade e integridade das mensagens que trafegam pela rede e a falta de mecanismos de reputação de usuários e servidores de e-mail.

Vários estudos na área de sistemas de e-mail buscam corrigir essas fragilidades. Dentre eles destacam-se a filtragem e classificação do conteúdo das mensagens, o uso de chaves assimétricas nos cliente de e-mail para assinar e cifrar as mensagens que serão transportadas pela rede, a autenticação de servidores de e-mail e as ferramentas de classificação automática de mensagens maliciosas. Por mais que esses estudos tragam grandes avanços nos sistemas de e-mail, eles não provêem mecanismos capazes de mensurar o quão confiável é um servidor ou domínio de e-mail. Em outras palavras, os mecanismos de autenticação por si só não são capazes de minimizar o envio de mensagens não solicitadas (*spam*).

Este trabalho define um modelo de confiança entre servidores de e-mail que utiliza técnicas de classificação de mensagens, um modelo de autenticação de remetente e redes de relacionamento para criar um ambiente capaz de disponibilizar informações sobre servidores legítimos ou maliciosos de uma forma descentralizada. A seção 2 traz uma revisão das principais ameaças aos serviços de e-mail; a seção 3 descreve as principais técnicas usadas para autenticação de remetentes; a seção 4 descreve o uso de redes de confiança em ambientes distribuídos; a seção 5 apresenta a arquitetura desenvolvida e detalha seus aspectos funcionais; a seção 6 traz a implementação do modelo de confiança e a análise dos resultados; a seção 7 discute alguns trabalhos relacionados; por fim, a seção 8 traz a conclusão do trabalho e delineia algumas perspectivas de continuidade.

2 Ameaças ao serviço de e-mail

O sistema de e-mail foi inicialmente projetado para ser simples, pois seu público alvo era restrito a um ambiente pequeno e confiável, constituido basicamente pela comunidade acadêmica. O protocolo SMTP, responsável pela transferência de e-mails entre servidores [3, 8], não provia mecanismos robustos para autenticação e controle de acesso. Dentre os problemas atuais desse sistema destacam-se os *spams*, vírus e *scams* que comprometem o desempenho, robustez, segurança e usabilidade dos sistemas de e-mail atuais.

Várias técnicas vêm sendo usadas para controle de *spam*. As principais são baseadas em listas de servidores confiáveis e não-confiáveis, ou na filtragem das mensagens recebidas com base em seu conteúdo:

- *Listas Negras* (*Black Lists*): Servidores *RBL* (*Realtime Blackhole Lists*) distribuídos na Internet mantêm listas de endereços IP de geradores ou propagadores de *spam*, que podem ser consultadas por meio de DNS pelos servidores de e-mail para verificar a confiabilidade de um remetente [7].
- *Listas Brancas* (*White Lists*): cada servidor de e-mail pode manter uma lista de remetentes nos quais confia; essa lista é normalmente mantida por meio de pedidos de confirmação de envio que remetem a um formulário web ou outra abordagem similar [6].
- *Filtros antispam*: programas de filtragem que classificam os e-mails de acordo com seu conteúdo; podem ser usadas técnicas estatísticas, de classificação bayesiana, redes neurais, etc [9, 16].

Os vírus e os *worms* de e-mail são programas que têm como objetivo inutilizar sistemas, destruir arquivos, roubar informações de uma máquina ou simplesmente propagar-se indefinidamente. Um vírus de e-mail normalmente é constituído por um e-mail com um anexo executável (ou um código HTML que permita carregar um arquivo executável armazenado remotamente). Esse código executável pode ser ativado pelo usuário ou até mesmo se lançar de forma automática, com o objetivo de explorar vulnerabilidades no cliente de e-mail para se reproduzir e também causar algum tipo de dano ao sistema local.

Finalmente, os *phishing scams* ou simplesmente *scams* são mensagens falsificadas que utilizam fragilidades do protocolo SMTP para construir ataques de engenharia social, visando enganar os destinatários, convencendo-os a informar dados bancários, números de cartões de créditos e outras informações confidenciais. A solução para este problema é utilizar um sistema de assinatura/certificado digital ou então tecnologias que validem a autenticidade do servidor ou do remetente da mensagem.

3 Autenticação de remetentes

As assinaturas digitais são códigos anexados ao cabeçalho de uma mensagem, visando garantir a autenticidade de um remetente. Várias técnicas foram desenvolvidas com esse objetivo, como o *PGP* (*Pretty Good Privacy*), *SPF* (*Sender Policy Framework*), *SenderID* e *DomainKeys*. O *PGP* [15] é um pacote de cifragem que utiliza chaves públicas e chaves de sessão para cifrar documentos. Cada mensagem é cifrada utilizando uma chave de sessão, que por sua vez é cifrada usando a chave pública do destinatário; assim, somente ele conseguirá decifrá-la, garantindo então a privacidade e integridade da mensagem.

O *SPF* [13] faz uso do *DNS* para prover ao servidor destinatário informações sobre o servidor de origem de um e-mail. Essas informações são descritas em uma linguagem própria e armazenadas em um registro específico do servidor *DNS* do domínio de origem da mensagem. Ao receber uma mensagem, o servidor destinatário realiza uma consulta de *DNS* para obter as informações do domínio de origem e comprovar a autenticidade do remetente. O *SenderID* [12] é uma extensão do *SPF* que visa melhorar a linguagem *SPF* e estender o protocolo *SMTP* para resolver problemas relacionados à autenticação de mensagens encaminhadas (*forwarded messages*).

O *DomainKeys* [4] utiliza o sistema de chave pública para validar a autenticidade de um servidor de e-mail. A chave privada é armazenada localmente em cada servidor, sendo usada para gerar uma assinatura no cabeçalho do e-mail a ser enviado. A chave pública é armazenada em um registro *DNS* e será requisitada pelo servidor destinatário para validar a assinatura. Esta abordagem elimina a necessidade de autoridades certificadoras, pois o próprio servidor disponibiliza as chaves de seu domínio.

4 Relações de confiança

Nas relações humanas, as relações de confiança são fundamentais na construção e manutenção de um grupo de indivíduos. Por relação de confiança entende-se o quanto um indivíduo confia nos outros e como age em relação a desconhecidos. As relações de confiança podem ser classificadas em *Confiança hierárquica*, *Grupos sociais* e *Redes de relacionamento*.

A *confiança hierárquica* trata de todos os relacionamentos apresentados através de uma hierarquia, como por exemplo a confiança de um pai em um filho. Essas relações de confiança podem ser representadas através de uma árvore, onde os nós são os indivíduos e as arestas são o grau de confiança que cada indivíduo possui no outro. Através da transitividade, qualquer indivíduo da árvore pode definir um grau de confiança em outro indivíduo, mesmo que ambos não tenham conexão direta [10].

Um *grupo social* é um conjunto de indivíduos cujas atividades se relacionam mutuamente de forma sistemática para um determinado fim. Ou seja, um grupo deve ser concebido como um sistema cujas partes se inter-relacionam [5]. Integrantes de grupos sociais são capazes de compartilhar informações de interesse comum, propagando-as entre os integrantes. Os integrantes também utilizam a *confiança do grupo* para iniciar e manter relacionamento com outros indivíduos. Os grupos sociais podem ser representados a partir de um grafo, onde os nós são os integrantes e as arestas são as ligações entre os indivíduos [11].

A palavra *relacionamento* define uma interação entre dois indivíduos, baseada em percepções recíprocas. As *redes de relacionamento* são todos os relacionamentos estabelecidos por um indivíduo, ou seja, o convívio de um determinado indivíduo com pessoas que este conhece [14]. Essa teia pode ser formada também por conhecidos de conhecidos e assim por diante; a partir da teia é possível caminhar pelos nós conhecidos até chegar a um objetivo final. As redes de relacionamento são utilizadas pelas pessoas para interagir com outras pessoas ou então chegar até um indivíduo qualquer.

Os sistemas computacionais podem fazer uso dos mesmos princípios de relacionamento interpessoal para definir níveis de confiança entre seus diversos elementos. A seguir será apresentada uma proposta que aplica os conceitos de relações de confiança para tentar melhorar a qualidade de um sistema de e-mail, diminuindo a quantidade de e-mails maliciosos em circulação.

5 Um sistema de confiança entre servidores de e-mail

Este trabalho define uma arquitetura capaz de manter informações de confiança sobre servidores de e-mail de forma dinâmica e descentralizada, utilizando técnicas de classificação de e-mail e autenticação de servidores. Os servidores de e-mail são organizados em grupos de confiança $\mathcal{T}$, definidos por seus administradores.

Por exemplo, considerando o conjunto de servidores de e-mail apresentados na figura 1, os servidores do grupo de confiança de s_4 são s_2, s_3, s_5 e s_9 ($\mathcal{T}_4 = \{s_2, s_3, s_5, s_9\}$).

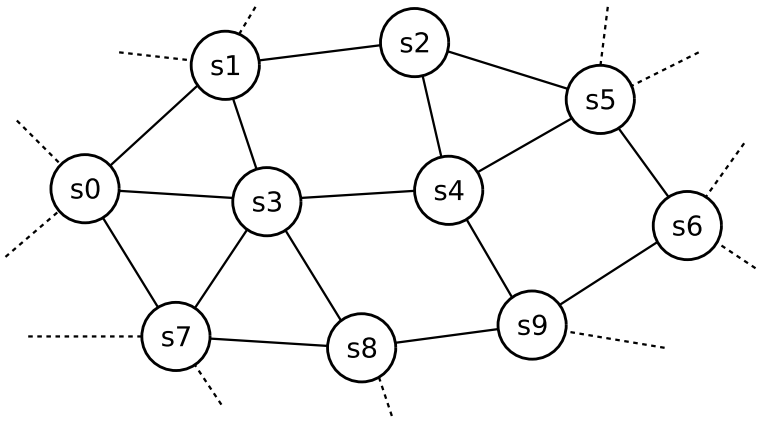

Figura 1: Grupos de confiança

Os servidores que constituem um determinado grupo de confiança cooperam entre si para difundir informações de confiança sobre servidores conhecidos pelos membros do grupo, usando técnicas de redes de confiança. Como os grupos de confiança dos diversos servidores podem ser distintos e se sobrepor parcialmente, as informações de confiança podem se propagar gradativamente pela rede, passando de um grupo de confiança a outro.

5.1 Arquitetura

A arquitetura do sistema utiliza conceitos de redes de confiança, em conjunto com ferramentas anti-spam, ferramentas anti-vírus e um modelo de autenticação para disponibilizar um sistema de confiança entre servidores de e-mail (ou *Mail Transport Agents* - MTAs). Na figura 2 estão representados os componentes que deverão ser implementados/integrados em cada MTA participante.

Servidor SMTP : responsável pela recepção das mensagens; implementa o protocolo SMTP.

Autenticar emissor : implementa um método de autenticação de domínio, como *SPF* ou *Domain Keys*.

Anti-spam e anti-vírus : analisam se uma mensagem é idônea ou maliciosa. Os resultados destes filtros são usados pelo sistema de confiança.

Sistema de confiança : analisa a confiança de um servidor de acordo com as ações deste servidor localmente e na rede.

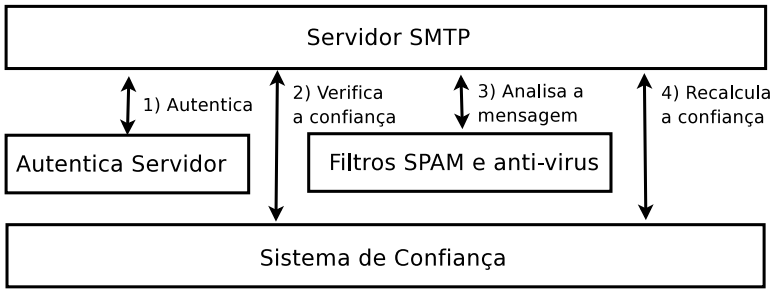

Figura 2: Modelo da arquitetura

Quando uma comunicação SMTP for estabelecida entre dois MTAs, o servidor de e-mail receptor deverá executar os seguintes passos para cada mensagem recebida (conforme indicado na figura 2):

1. Autenticar o servidor; se o servidor for corretamente autenticado, ir para o próximo passo, do contrário, recusar a mensagem e encerrar a conexão;
2. Verificar se o servidor emissor pode enviar mensagens, consultando o sistema de confiança (as regras para tal serão definidas na seção 5.2). Caso a resposta seja positiva o sistema recebe a mensagem, do contrário a conexão é encerrada;
3. Submeter e-mail recebido aos filtros anti-spam e anti-vírus. Esses filtros deverão indicar se o e-mail é legítimo ou malicioso;
4. Enviar resultado dos filtros ao sistema de confiança, que atualiza a base local de confiança;

A arquitetura foi dividida em três módulos: *gerência*, *armazenamento* e *propagação de confiança*, apresentados a seguir.

5.2 Gerência de confiança

Cada servidor de um grupo de confiança deve executar ações para a manutenção de informações de confiança sobre outros servidores de e-mail externos ao grupo. Para cada servidor externo é mantida uma *confiança local*, definida a partir de informações internas (mensagens recebidas), e uma *confiança global*, calculada a partir de informações provindas dos servidores membros do grupo de confiança. A média entre essas duas confianças define uma *confiança final*, que é usada para determinar a quantidade máxima de mensagens que podem ser recebidas do servidor em questão no ciclo corrente. Um servidor externo que exceda esse limite é *banido*, ou seja, impedido de enviar novas mensagens, até o início de um novo ciclo. A duração de cada ciclo é definida pelo administrador (em minutos, horas, dias, etc). A tabela 1 mostra as principais constantes e variáveis envolvidas no cálculo de confiança e uma descrição de sua finalidade. Todos os valores de confiança variam entre 0 e 1 (0% e 100%).

Tabela 1: Variáveis do cálculo de confiança

Variável	**Descrição**
$\mathcal{T}_i$	Conjunto de servidores SMTP pertencentes ao grupo de confiança de s_i.
$\mathcal{K}_i$	Conjunto de servidores SMTP externos conhecidos pelo servidor s_i. É importante observar que servidores do grupo de confiança não são considerados servidores externos: $\mathcal{K}_i \cap \mathcal{T} = \emptyset$.
$cl_i^j(x)$	Confiança local: indica o último valor conhecido por s_i da opinião individual de s_j sobre um servidor $x \in \mathcal{K}_j$ (cada servidor s_i armazena localmente os valores $cl_i^*(x)\ \forall x \in \mathcal{K}_i$.
$cg_i(x)$	Confiança global: indica a opinião do grupo de confiança sobre o servidor x, calculada por s_i a partir das confianças locais do seu grupo de confiança $\mathcal{T}$.
δ_c	Passo das variações de confiança.
c_{def}	Valor default para a confiança inicial a ser atribuída a novos servidores, definida empiricamente como 50%.
$age_i(x)$	"idade" das informações de confiança sobre x, em períodos.
age_{max}	idade máxima das informações de confiança sobre cada servidor.
$mm_i(x)$	Número de mensagens maliciosas recebidas de x por s_i no ciclo corrente.
$mi_i(x)$	Número de mensagens idôneas recebidas de x por s_i no ciclo corrente.
mm_{max}	Número de mensagens maliciosas necessárias em um ciclo para fazer a confiança local sobre um servidor reduzir de δ_c, caso ela esteja em 100% (valor definido pelo administrador do sistema).
lim	Número de mensagens idôneas/maliciosas necessárias para aumentar/diminuir a confiança local sobre um servidor em δ_c.
$conf$	Confiança instantânea, usada no cálculo de lim.
$banned_i(x)$	indica se o servidor x foi banido por s_i; caso seja verdadeiro, mensagens provindas de x serão recusadas por s_i até o final do ciclo corrente.

O procedimento 1 descreve as ações realizadas por um servidor $s_i \in \mathcal{T}_i$ ao receber uma conexão de um servidor SMTP $x \notin \mathcal{T}_i$ para a entrega de mensagens.

Procedimento 1 Ações em s_i ao receber uma conexão do servidor SMTP x:

1: **if** $x \notin \mathcal{K}_i$ **then**
2: $\mathcal{K}_i \leftarrow \mathcal{K}_i \cup \{x\}$ /* adiciona x ao conjunto de servidores conhecidos $\mathcal{K}_i$ */
3: $banned_i(x) \leftarrow$ FALSE
4: $cg_i(x) \leftarrow c_{def}$ /* ajusta para os valores default */
5: $cl_i^i(x) \leftarrow c_{def}$
6: $age_i(x) \leftarrow 0$
7: $mi_i(x) \leftarrow 0$
8: $mm_i(x) \leftarrow 0$
9: **end if**

A reavaliação dos valores de confiança local ocorre a cada recepção de mensagem. Caso o número de mensagens maliciosas recebido de um determinado servidor exceda um limite máximo, o servidor em questão é banido até o final do ciclo corrente. O procedimento 2 relaciona as ações realizadas por um servidor $s_i \in \mathcal{T}_i$ ao receber uma mensagem m de um servidor SMTP $x \notin \mathcal{T}_i$:

Procedimento 2 Ações de s_i ao receber uma mensagem m do servidor SMTP x:

1: **if** $banned_i(x)$ **then**
2: recusa a mensagem m
3: **else**
4: aceita a mensagem m
5: $age_i(x) \leftarrow 0$
6: $conf \leftarrow \sqrt{cg_i(x) \times cl_i^i(x)}$ /* confiança instantânea (média geométrica de cg e cl) */
7: analisa o conteúdo da mensagem m
8: **if** $msg\_idonea(m)$ **then**
9: $mi_i(x) \leftarrow mi_i(x) + 1$
10: $lim \leftarrow conf \times mm_{max}$
11: **if** $mi_i(x) \geq lim \wedge cl_i^i(x) < 1$ **then**
12: $cl_i^i(x) \leftarrow cl_i^i(x) + \delta_c$
13: $mi_i(x) \leftarrow 0$
14: envia $notify(x, cl_i^i(x))$ ao grupo de confiança $\mathcal{T}_i$
15: **end if**
16: **else**
17: $mm_i(x) \leftarrow mm_i(x) + 1$
18: $lim \leftarrow (1 - conf) \times mm_{max}$
19: **if** $mm_i(x) \geq lim \wedge cl_i^i(x) > 0$ **then**
20: $cl_i^i(x) \leftarrow cl_i^i(x) - \delta_c$
21: $mm_i(x) \leftarrow 0$
22: $banned_i(x) \leftarrow$ TRUE
23: envia $notify(x, cl_i^i(x))$ ao grupo de confiança $\mathcal{T}_i$
24: **end if**
25: **end if**
26: **end if**

No procedimento 2 pode-se observar que, que quando a confiança local sobre x diminui este é banido até o fim do ciclo. Além disso, a adoção de limites móveis (através das variáveis *conf* e *lim*) faz com que servidores pouco confiáveis percam sua confiança mais rapidamente do que a ganham e que servidores mais confiáveis demorem mais para perder sua confiança do que para ganhá-la.

Finalmente, ao encerrar cada ciclo de operação, cada servidor deve executar uma série de ações, descritas no procedimento 3. Elas consistem basicamente em re-habilitar servidores banidos, reiniciar contadores e "esquecer" das informações sobre servidores que há muito não se comunicam com s_i. Esse mecanismo de esquecimento é importante para que servidores externos não sejam beneficiados ou prejudicados por históricos muito antigos que não reflitam seu comportamento recente.

Procedimento 3 Ações de s_i ao encerrar um ciclo:

1: **for all** $x \in \mathcal{K}_i$ **do**
2: $banned_i(x) \leftarrow$ FALSE
3: $mi_i(x) \leftarrow 0$
4: $mm_i(x) \leftarrow 0$
5: $age_i(x) \leftarrow age_i(x) + 1$
6: **if** $age_i(x) = age_{max}$ **then**
7: $\mathcal{K}_i \leftarrow \mathcal{K}_i - \{x\}$ /* "esquece" do servidor x */
8: remove as informações locais sobre x
9: envia $notify(x, undef)$ ao grupo de confiança $\mathcal{T}_i$
10: **end if**
11: **end for**

5.3 Armazenamento de confiança

O módulo de armazenamento de confiança simplesmente faz a manutenção das informações locais sobre os servidores de e-mail conhecidos por cada servidor do grupo de confiança. As seguintes informações são mantidas em uma base de dados em s_i para cada servidor $x \in \mathcal{K}_i$:

- domínio de x;
- nome de domínio de x (*FQDN - Fully Qualified Domain Name*);
- endereços IP de x (pode haver mais de um endereço para o mesmo servidor);
- confianças locais $cl_i^*(x)$ e globais $cg_i(x)$;
- contadores de mensagens idôneas $mi_i(x)$ e maliciosas $mm_i(x)$ recebidas de x durante o ciclo corrente;
- validade temporal $age_i(x)$ dos dados referentes a x.

5.4 Propagação de confiança

A propagação de confiança é responsável por difundir informações sobre servidores conhecidos de s_i para os demais membros de seu grupo de confiança $\mathcal{T}_i$. Os grupos de confiança são definidos pelo administrador do serviço de e-mail e indicam quais servidores devem ser consultados para o cálculo da confiança global (nesta proposta, a escolha dos grupos é feita manualmente; métodos automáticos de estabelecimento de grupos de confiança ainda não foram abordados).

Conforme definido no procedimento 2, quando um servidor reajusta sua confiança local sobre um servidor externo, ele informa o fato aos demais servidores de seu grupo de confiança, através de uma mensagem *notify*. Ao receber uma mensagem $notify(x, c)$ vinda de s_j, o servidor s_i apenas atualiza suas informações locais sobre x, conforme descrito no procedimento 4. A confiança global é simplesmente a média das confianças locais sobre x dos servidores que compõe o grupo de confiança $\mathcal{T}_i$, sendo reavaliada a cada final de ciclo (procedimento 5). Caso algum servidor s_j não tenha informado sua opinião sobre x ($cl_i^j(x) = undef$), ele não entra no cálculo da média.

Procedimento 4 Ações de s_i ao receber uma mensagem $notify(x, c)$ de s_j:

1: **if** $j \in \mathcal{T}_i \wedge x \in \mathcal{K}_i$ **then**
2: $cl_i^j(x) \leftarrow c$
3: **end if**

Procedimento 5 Ações de s_i ao encerrar um ciclo:

1: **for all** $x \in \mathcal{K}_i$ **do**
2: $cg_i(x) \leftarrow media(cl_i^j(x) \neq undef, \forall s_j \in \mathcal{T}_i)$
3: **end for**

6 Implementação e resultados

O protótipo foi implementado em *Linux*, usando o *Postfix* como servidor de e-mail, o *SpamAssassin* como filtro anti-spam e o *Clamav* como filtro anti-vírus. Foi implementado um programa, denominado *TrustMail*, que implementa os procedimentos de gerência e propagação de confiança definidos na seção anterior. A seguir serão descritos a implementação do *TrustMail*, o sistema de comunicação entre o servidor de e-mail e os demais componentes e uma avaliação do funcionamento do sistema.

6.1 O protótipo *TrustMail*

O protótipo foi implementado na linguagem C e utiliza a base de dados SQLite para armazenar as informações locais de confiança. A comunicação do *TrustMail* com as outras ferramentas para recepção e classificação de e-mails é feita através de filas de mensagens POSIX.

O MTA escolhido foi o *Postfix* 2.2, pela facilidade de integração com outras ferramentas e por seu código ser simples e legível. O sistema de autenticação de servidor escolhido foi o SPF, por ter implementação simples e ser fácil de integrar com o *Postfix*. Tanto a comunicação do MTA com o SPF, quanto a comunicação com o *TrustMail* são feitas através do *policyd*, uma implementação SPF de código aberto. O código do *policyd* foi modificado para suportar a comunicação com o *TrustMail*. Este programa faz a autenticação de um servidor, tornando desnecessária a comunicação direta do *TrustMail* com o *Postfix*.

O filtro anti-spam escolhido foi o *SpamAssassin* 2.63, que possui vários critérios de classificação de mensagens. O anti-vírus utilizado foi o *Clamav* 0.70, que possui uma extensa lista de assinaturas de vírus. Ambos executam como *daemons*, facilitando a integração com outros programas. A comunicação do *Postfix* com os filtros de mensagens é feita através do script *Clamav-Filter*, modificado para suportar a notificação de mensagens maliciosas para o *TrustMail*. O protótipo da implementação atual está ilustrado na figura 3. Nessa figura é possível observar que o *TrustMail* independe do servidor, do sistema de autenticação e dos filtros de mensagens. Ou seja, este modelo pode ser facilmente integrado a outras ferramentas.

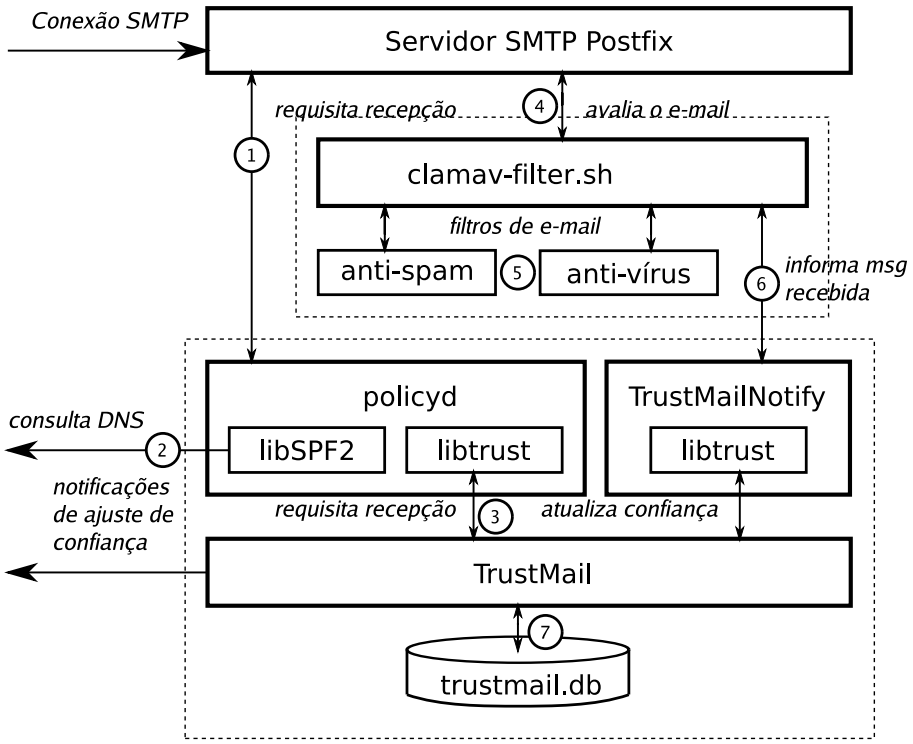

Figura 3: Protótipo implementado

O processo de recepção de um e-mail é constituido das seguintes etapas:

1. O *Postfix* invoca o *policyd* toda vez que receber um comando `RCPT TO`. Os parâmetros informados ao *policyd* são: endereço do emissor, ip do emissor e domínio informado no comando `HELO/EHLO`.

2. O *policyd* recebe os parâmetros do *Postfix* e faz uma consulta ao servidor DNS do domínio do emissor, autenticando assim, a procedência do servidor. O retorno desta consulta pode ser:

   - *Pass*, *Soft Fail*, *Neutral*, *Unknown* ou *None*: O processo continua;
   - *Error*: Retorna ao *Postfix* a mensagem "450 Falha temporária";
   - *Fail*: Retorna um erro e a conexão deve ser finalizada pelo *Postfix*.

3. Se o processo prosseguir, o *policyd* invoca o *TrustMail* e pergunta se a conexão pode continuar. O *TrustMail* calcula a quantidade de mensagens que o servidor pode receber no ciclo corrente e responde ao *policyd*. O *policyd* repassa ao *postfix* a resposta recebida. Por fim, o processo de recebimento de mensagem do *Postfix* continua.

4. Quando uma mensagem for recebida, o *Postfix* chamará o script *Clamav-Filter*. Os parâmetros passados ao *Clamav-Filter* são: o endereço ip, o host de origem e o domínio da mensagem.

5. Em seguida, o *Clamav-Filter* recuperará a mensagem de um diretório, chamará o *Clamav* e o *SpamAssassin* e aguardará o retorno.

6. Depois, o *Clamav-Filter* chamará o *TrustMailNotify* que notificará ao *TrustMail* que uma mensagem maliciosa ou idônea foi recebida.

7. Por fim, o servidor armazenará as novas informações de confiança no *Trust_Mail.db*.

6.2 Resultados experimentais

O protótipo foi implantado e testado em um ambiente de máquinas virtuais UML (*User-Mode Linux*), em um experimento cuja topologia é mostrada na figura 4. Nele, quatro máquinas virtuais implementam o grupo de confiança ($s_0 \dots s_3$). Outra máquina virtual foi usada para simular um servidor de e-mail externo ao grupo de confiança (e_0), que vai enviar e-mails aos servidores do grupo.

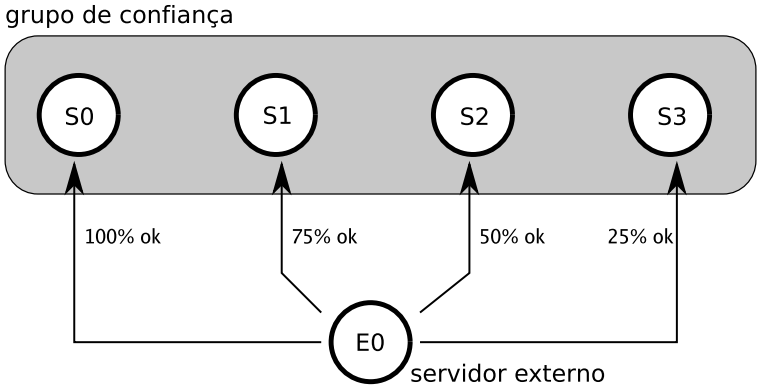

Figura 4: Topologia do experimento realizado

No experimento realizado, o servidor externo envia continuamente e-mails, de forma aleatoria (em média um e-mail a cada 15 segundos), mas uniformemente distribuída entre os servidores do grupo. O servidor externo se comporta de forma distinta em relação ao tipo de e-mail enviado a cada membro do grupo: ele nunca envia e-mails maliciosos a s_0, mas 25% dos e-mails enviados a s_1, 50% dos enviados a s_2 e 75% dos enviados a s_3 são maliciosos. Para os demais parâmetros, foi escolhida uma duração de ciclo de 30 minutos, $age_{max} = 10$, $c_{def} = 50\%$, $\delta_c = 10\%$ e $mm_{max} = 10$ mensagens. A duração total da execução foi de 24 horas.

Para avaliar a influência das informações do grupo nas decisões locais de cada servidor, foi avaliada a evolução das confianças locais e globais sobre e_0 ($cl_i^i(e_0)$ e $cg_i(e_0)$) em cada membro do grupo, em duas circunstâncias: sem as notificações de ajuste de confiança entre os membros do grupo (ou seja, sem cooperação

entre os membros do grupo) e com as notificações. Essa evolução é apresentada no gráfico da figura 5. Nele, é possível perceber que, caso não exista cooperação entre os membros do grupo (curvas $cl(s_i, x)$), cada servidor forma sua própria opinião (nível de confiança) sobre o servidor externo e_0. Além disso, a opiniões individuais de $s_{1,2,3}$ apresentam significativas variações ao longo do tempo, devido às alterações aleatórias de curto prazo no comportamento do servidor externo (como s_0 não recebe mensagens maliciosas de e_0, sua confiança nele sobe rapidamente a 100% e se mantém nesse nível). Por outro lado, ao incluir a cooperação entre os membros do grupo (curvas $cg(s_i, x)$), observa-se uma maior estabilidade temporal dos dados e também a convergência de todos os servidores do grupo para níveis similares de confiança sobre e_0.

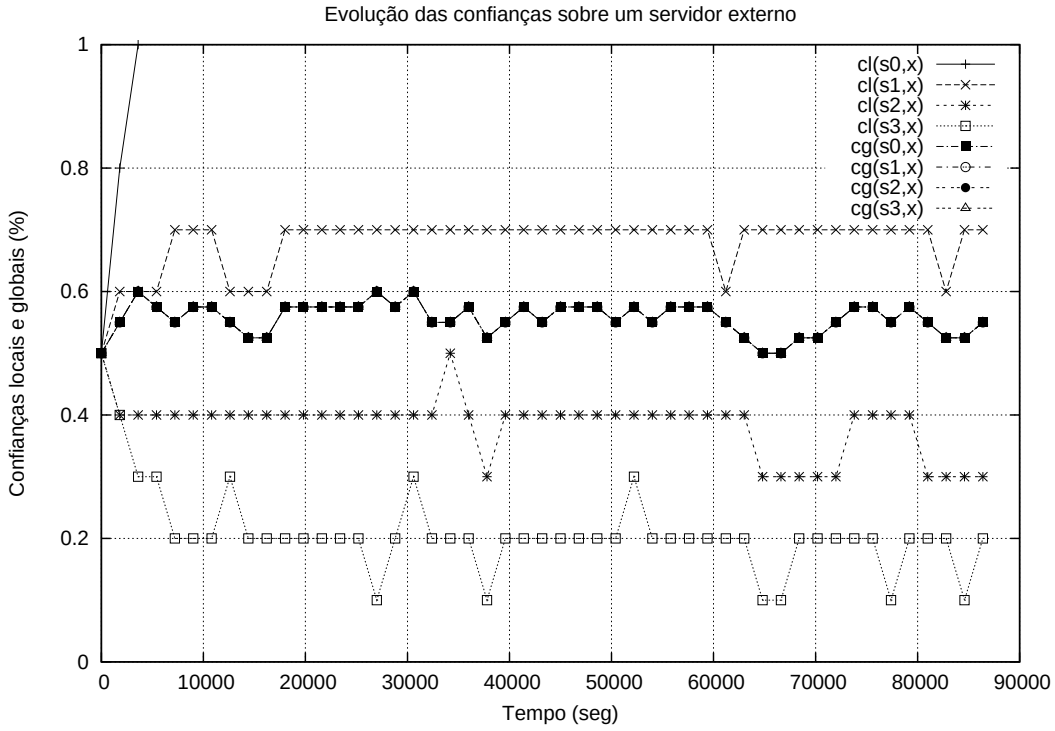

Figura 5: Evolução das confianças local e global nos servidores

No mesmo experimento foi avaliada a quantidade de mensagens recusadas através do mecanismo de banimento de servidores pouco confiáveis (linha 22 do procedimento 2). Os resultados obtidos estão apresentados no diagrama da figura 6. Pode-se observar que, com exceção de s_0 (que não recebeu spam), os demais servidores recusaram parcelas significativas dos e-mails que lhes foram enviados (em proporções inversas às respectivas confianças sobre o servidor externo). Isso mostra que o banimento limita a quantidade de mensagens que um servidor externo pode enviar a cada membro do grupo, caso a confiança sobre ele não seja 100%.

Finalmente, foi avaliada a quantidade de mensagens de notificação $notify(x, c)$ geradas, em relação ao número de e-mails recebidos do servidor externo. No experimento, foram enviados por e_0 19250 e-mails, e que provocaram o envio de 2072 notificações de ajuste de confiança entre os servidores do grupo de confiança. Essa quantidade representa cerca de uma notificação para cada 9,3 e-mails recebidos, o que é um valor bastante baixo.

7 Trabalhos correlatos

Poucos trabalhos têm sido propostos especificamente na área de combate ao spam usando redes sociais. Dentre estes, os dois aqui apresentados são considerados por nós os mais relevantes.

O sistema *MailRank* [2] é um sistema colaborativo para a construção de *white-lists* globais. Dados das atividades dos usuários envolvidos são coletados e agrupados em uma rede de relacionamento global. Cada

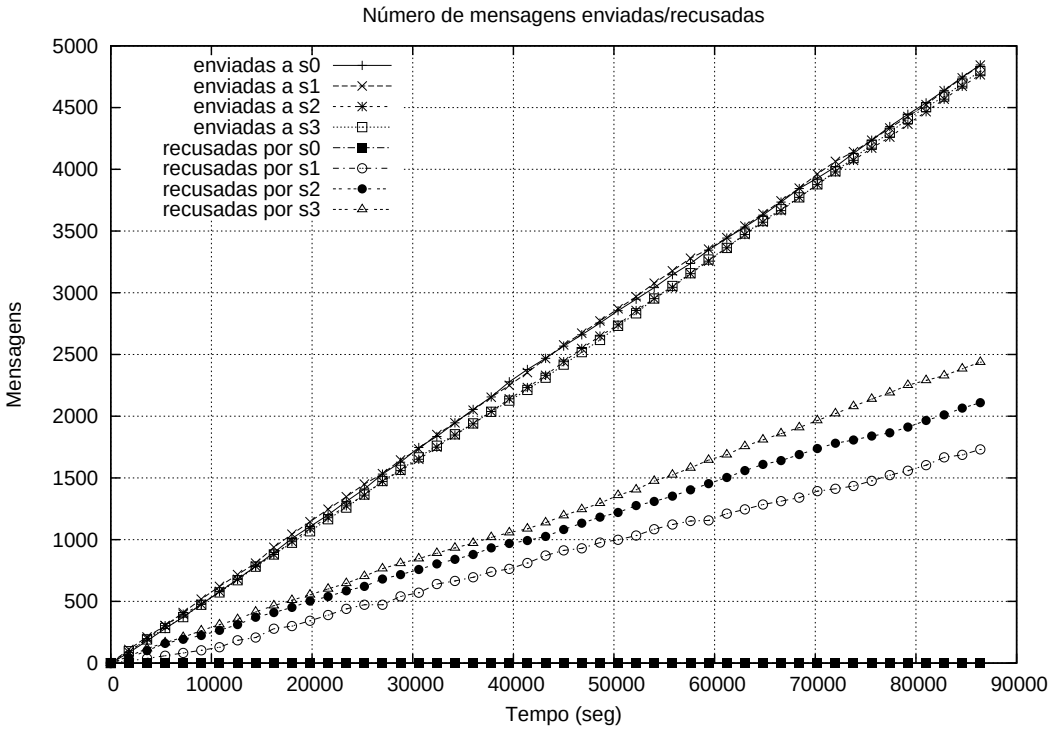

Figura 6: Número de mensagens enviadas/recusadas pelos servidores

mensagem enviada por um usuário a outro é transformada em um voto de confiança, usado na construção da rede. O sistema é compatível com a estrutura de e-mail atual, sendo composto por dois elementos básicos: o *MailRank Proxy*, que intermedia a comunicação de cada MUA (*Mail User Agent*, ou cliente de e-mail) com os servidores de e-mail, extraindo informações dos e-mails enviados e recebidos por cada usuário do sistema, e o *MailRank Server*, um servidor central que coleta dados de todos os MailRank proxies, visando criar uma classificação global dos usuários baseada nos e-mails enviados e recebidos.

Por outro lado, o trabalho apresentado em [1] discute a construção de uma ferramenta antispam que extrai as informações de relacionamento dos usuários de e-mail através da análise dos campos de endereço dos cabeçalhos de e-mail (*From*, *To*, *Cc*, *Bcc*, etc), que são usados para construir um grande grafo de relacionamento existentes entre os usuários. Em seguida, esse grafo é usado para construir *white-lists* de usuários, usando uma propriedade das redes sociais conhecida como *tendência de aglomeração*. Finalmente, os grupos aglomerados assim construídos são analisados em busca de alguns padrões de relacionamento que caracterizam o spamming.

Nosso trabalho difere de ambos em alguns aspectos: enquanto aqueles visam classificar usuários como spammers ou não, nosso trabalho analisa o comportamento de servidores de e-mail, o que permite uma maior escalabilidade; além disso, nosso trabalho apresenta uma arquitetura completamente descentralizada, enquanto aqueles dependem de alguma centralização na construção e análise da rede de relacionamentos.

8 Conclusão e trabalhos futuros

Este trabalho descreveu um modelo de confiança em servidores de e-mail, que define um grupo de servidores que confiam mutuamente entre si e trocam "opiniões" sobre servidores externos. Cada um dos componentes do grupo usa as opiniões dos demais membros para contruir uma confiança global e usa essa informação para limitar a quantidade de e-mails recebidos de servidores externos. A confiança é propagada através de um modelo de redes de relacionamento, descentralizando a manutenção e evolução das informações de confiança.

Outros aspectos deste trabalho que podem ser explorados em trabalhos futuros seriam a) a definição de

reputações entre os membros de cada grupo de confiança, permitindo a inclusão ou exclusão automática de membros no grupo; b) a definição dos valores ótimos para as constantes do modelo, de acordo com as condições reais do ambiente de operação; e c) um estudo mais aprofundado dos mecanismos de propagação das confianças entre grupos distintos através dos servidores em comum, de forma escalável.

Referências

[1] Boykin, P. O., and Roychowdhury, V. P. Leveraging social networks to fight spam. *IEEE Computer 38*, 4 (2005), 61–68.

[2] Chirita, P. A., Diederich, J., and Nejdl, W. Mailrank: Using ranking for spam detection. *ACM International CIKM Conference* (2005).

[3] Crocker, D. RFC 822: Standard for the format of arpa internet text messages, Aug. 1982.

[4] Delany, M., and Yahoo. Domain-based email authentication using public-keys: Advertised in the DNS (DomainKeys). Internet Draft, 2004.

[5] Grahagan, J. *Comportamento interpessoal e de grupo*. Rio de Janeiro: Zahar, 1976.

[6] Hall, R. J. How to avoid unwanted email. *Communications of the ACM 41*, 3 (1998), 88–95.

[7] Jung, J., and Sit, E. An Empirical Study of Spam Traffic and the Use of DNS Black Lists. In *Internet Measurement Conference* (Taormina, Italy, October 2004).

[8] Klensin, J. RFC 2821: Simple mail transfer protocol, Apr. 2001.

[9] Sahami, M., Dumais, S., Heckerman, D., and Horvitz, E. A bayesian approach to filtering junk E-mail. In *Learning for Text Categorization: Papers from the 1998 Workshop* (Madison, Wisconsin, 1998), AAAI Technical Report WS-98-05.

[10] Vaz, W., and Magalhães, G. C. Um modelo para derivação de relacionamentos espaciais em equivalentes semânticos relacionais. *IV Simpósio Brasileiro de GeoInformática - Caxambú, MG* (2002).

[11] Weil, P. *Relações humanas na familia e no trabalho*. ed. Petropolis: Vozes, 1982.

[12] Wong, Microsoft, and Lentczner. The SenderID record: Format interpretation. http://www.ietf.org/internet-drafts/draft-ietf-marid-protocol-02.txt, 2004.

[13] Wong, M. Sender Policy Framework (SPF): A convention to describe hosts authorized to send SMTP traffic. http://db.org/drafts/internet/mengwong/spf/00/, 2004.

[14] Woods, D. R., Omerod, S. D., and Black, M. D. *Como tecer uma rede de relacionamentos e se valer dela*. São Paulo :Nacional, 1976.

[15] Zimmermann, P. R. *The Official PGP User's Guide*. The MIT Press, 1995.

[16] Zou, C. C., Gong, W., and Towsley, D. Code Red worm propagation modeling and analysis. In *9th ACM conference on Computer and Communications Security* (2002), pp. 138–147.