ESCOLA POLITÉCNICA
DE PERNAMBUCO

UMA ABORDAGEM PARA CONTROLE DE MENSAGENS INDESEJÁVEIS EM REDES UTILIZANDO O MAIL-POT

Trabalho de Conclusão de Curso

Engenharia da Computação

Antonio Felipe Costa de Almeida
Orientador: Prof. Msc. Wellington Pinheiro dos Santos

ANTONIO FELIPE COSTA DE ALMEIDA

UMA ABORDAGEM PARA CONTROLE DE MENSAGENS INDESEJÁVEIS EM REDES UTILIZANDO O MAIL-POT

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia da Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Recife, dezembro de 2008

Dedico este trabalho a minha mãe,
Zefinha.

Agradecimentos

Agradeço em primeiro lugar a Deus, a quem devo graças e que me deu a saúde e a força para enfrentar esses cinco anos de faculdade.

A minha família, em especial aos meus pais pela preocupação em me deixar como herança o conhecimento, minha tia Irene que sempre me ajudou financeiramente, nunca deixando de acreditar em mim e a minha irmã que amo muito.

A todos os professores do DSC, principalmente a Cristine Gusmão de quem sempre gostei de graça e ao professor Wellington Pinheiro que me orientou de forma exemplar, sempre paciente e educado. Todas as conversas de acompanhamento de projeto não acrescentaram conteúdo apenas à monografia, como também a mim.

A todos os amigos que fiz ao longo do curso, principalmente a Bruno Luigi, Juliane Botelho, Keldjan Alves, Marcos Alvares, Moisés Augusto, Murilo Pontes, Renato Morais, Thaysa Paiva e Vanessa Trajano. Cada um foi especialmente durante o tempo na faculdade, espero não perder o contato com nenhum de vocês. Desculpem-me os outros, mas não há espaço para citar todos.

A toda a equipe da Triforsec, em especial a Rodrigo Ramos pela compreensão e paciência durante a realização deste projeto e a Alejandro Flores pela atenção dispensada e agradável companhia de trabalho.

Agradeço a todos os amigos que fiz através do esporte, o pessoal do vôlei, da natação e da comunidade Meus SAIS.

A meus técnicos Sandro e Fátima que me ensinaram a gostar de esporte e fazê-lo com responsabilidade.

A professora Gilva Batista que sempre esteve ao meu lado, nas organizações de conclusões, viagens e lutas políticas em Arcoverde.

Resumo

Segurança da informação é uma área crítica para utilização de redes de computadores, tanto no âmbito profissional, quanto no pessoal. A crescente demanda no uso dos meios virtuais para comunicação, e a preocupação com a integridade, disponibilidade e confiabilidade das informações são de grande preocupação dos administradores de redes.

O aparecimento constante de pessoas mal intencionadas que tentam aproveitar de vulnerabilidades em redes de computadores é cada vez maior, aliada a falta de informação de como eles agem. Problemas como a disseminação de *spams* que causam prejuízos às empresas e indivíduos que utilizam o correio eletrônico como ferramenta de trabalho.

Este trabalho apresenta um estudo e o desenvolvimento de uma ferramenta, intitulada de Mail-pot, que se passa por um servidor de correio. O Mail-pot realiza todas as funções de um servidor de correio normal, porém ao invés de repassar o tráfego recebido, os *spams*, ele armazena esse conteúdo em banco de dados. Posteriormente, o conteúdo armazenado é analisado para identificação do país de origem do *spam*, conteúdo disseminado e outras particularidades.

A ferramenta foi desenvolvida em Java e os dados armazenados em banco de dados My-SQL. O Mail-pot foi instalado em um servidor de sistema operacional Linux e coletando *spams* por 30 dias. Em seguida, na fase de análise, os dados foram tratados e um relatório do conteúdo armazenado foi gerado. O tratamento dado proporciona a utilização desse tipo de ferramenta para gerar informações que auxiliam no desenvolvimento de políticas e implantações de segurança da informação.

Abstract

Information Security is a critical area for computer networking, either in professional or personal activities. The growing demand on the use of virtual communications and the need to ensure information integrity, availability and reliability are network administrators' major concerns.

The appearance of malicious people that try to exploit computer networks vulnerabilities is growing up constantly and is associated with lack of information about behavior patterns. Problems like spam dissemination that results on damages for companies and individuals that uses electronic mail as a work tool.

This work shows a research and the development of a tool, named by Mail-pot, which runs as an electronic mail server. The Mail-pot runs normally all mail server functions, but rather than forward the received traffic (spams) it stores the message content in databases. Subsequently, the stored content is analyzed for identification of spam source country, context and other variables.

The tool was developed in Java programming language and collected data was stored in a MySQL database. Mail-pot was installed in a server with Linux operating system and collected spams for 30 days. After this, in the analysis period, a data-mining was performed and a report of stored content was generated. Data-mining permits the use of this tool for generates information that helps the creation of security policies and implementations.

Sumário

Índice de Figuras

Índice de Tabelas

Lista de Abreviaturas e Siglas

ARPA – *Advanced Research Projects Agency*

WAN – *Wide Area Networks*

ISP – *Internet Service Provider*

IDS – *Intrusion Detection System*

CERT.br – Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil

SMTP – *Simple Mail Transfer Protocol*

JRE – *Java Runtime Environment*

LBL – *Lawrence Berkeley Laboratory*

AT&T – *American Telephone and Telegraph*

DTK – *Detection Toolkit*

INPE – Instituto Nacional de Pesquisas Espaciais

NBSO – *NIC BR Security Office*

PESC – Programa de Engenharia de Sistemas e Computação

COPPE/UFRJ – Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia da Universidade Federal do Rio de Janeiro

MCT – Ministério da Ciência e Tecnologia

ADSL – *Asymmetric Digital Subscriber Line*

TCP – *Transmission Control Protocol*

IP – *Internet Protocol*

DSL – *Digital Subscriber Line*

1 Introdução

Na Guerra Fria, durante um projeto de pesquisa militar na **Agência de Pesquisas em Projetos Avançados** (ARPA - *Advanced Research Projects Agency*)[1] [1], foi desenvolvida a rede que deu origem à Internet. Esse projeto surgiu como resposta do governo americano ao lançamento do Sputnik[2] pela União Soviética. Inicialmente a idéia era conectar os mais importantes centros universitários de pesquisa americanos com o Pentágono para permitir não só a troca de informações rápidas e protegidas, mas também para instrumentalizar o país com uma tecnologia que possibilitasse a sobrevivência de canais de informação no caso de uma guerra nuclear.

A tecnologia utilizada na época para transmissão de dados foi criada com o nome de WAN (*Wide Area Networks*), mas a linguagem utilizada nos computadores ligados em rede era muito complexa, por isso, na época, o potencial de alastramento da Internet não podia ser imaginado [2].

Durante a década de setenta, com a revisão das limitações dos programas utilizados nos computadores em rede, o *e-mail* (*eletronic mail*) tornou-se o primeiro uso da Internet entre os pesquisadores, porque possibilitava que a comunicação entre eles fosse facilmente acessível, e também para trocar informações dentro das universidades. As aplicações comerciais da Internet começaram a acontecer nos anos oitenta com os primeiros provedores de serviço da Internet (ISP – *Internet Service Provider*) possibilitando ao usuário comum a conexão com a rede mundial de computadores, a partir de sua casa [2].

Nos últimos 10 anos, a quantidade de computadores conectados à Internet cresceu mais de 18.000%, sendo atualmente cerca de 541,5 bilhões de dispositivos interconectados pela grande rede [3]. À medida que as conexões cresciam, surgiram

---

[1] Em março de 1972, foi renomeada para *Defense Advanced Research Projects Agency* – DARPA. Em seguida, mudou novamente pra ARPA em fevereiro de 1993, mas em março de 1996, voltou a se chamar DARPA.

[2] Em 4 de outubro de 1957, a antiga União Soviética lança o Sputnik, primeiro satélite artificial do mundo.

várias pessoas ou grupos denominados *blackhats*, quase sempre chamados de *hackers*, indivíduos que tentam usar a tecnologia da Internet para realizar atividades ilegais, destrutivas ou não autorizadas [4]. Essas atividades podem ir de um simples adolescente tentando atos de vandalismo em sites Web até uma tentativa sofisticada de comprometer as empresas administradoras de cartões de crédito, ou realizar ataques terroristas contra a infra-estrutura de um país.

Preocupados com o aumento das ameaças das informações, surgiu a idéia de *honeynets* [5], processo que vai de encontro a filosofia da maioria das soluções para segurança da informação, soluções passivas como IDS (*Instruction Detected System*), *firewalls*, antivírus. As *honeynets* não funcionam como um tipo de implementação de segurança, mas seus resultados são utilizados como um guia para auxiliar na criação/manutenção/controle das políticas e implantações de segurança.

Este trabalho é o resultado de um projeto de seis meses, onde foi desenvolvido uma vertente dos projetos mantidos pelo CERT.br (Centro de Estudos, Respostas e Tratamento de Incidentes de Segurança no Brasil) [6], o Spampot [7] que segue a abordagem das *honeynets*. O Spampot faz um estudo dos tipos de ameaças a *proxies* abertos, analisando os *spams* recebidos.

O aplicativo desenvolvido, intitulado de Mail-pot, tem como objetivo analisar o tráfego recebido, os *spams*. Mas, diferentemente do Spampot, ele utiliza um servidor SMTP (*Simple Mail Transfer Protocol*) [8] emulado. O Mail-pot recebe e armazena todo o conteúdo que chega em banco de dados ao invés de repassar as mensagens. Nele não foi adotada nenhuma medida de segurança, nem tão pouco facilitação para os *spammers,* responsáveis por usar endereços de destinatários desconhecidos para o envio de mensagens não solicitadas em grande número, agirem.

A arquitetura é bastante simples. Necessita apenas de uma conexão com a Internet, uma máquina para hospedá-lo, um banco de dados MySQL [9] instalado e uma JRE *(Java Runtime Environment*) [10].

O Mail-pot está organizado em duas fases: **coleta dos dados** e **análise dos dados**. A fase de coleta durou 30 dias, sendo armazenados 501 *spams* no banco de dados. A fase de análise durou cerca de 150 dias onde foram usados outros

aplicativos, também desenvolvidos em Java, para fazer uma varredura no conteúdo dos *spams* e preencher dados, que os *spammers* tentaram camuflar, nas tabelas.

Os resultados obtidos permitiram uma melhor análise do problema *spam* e também um melhor entendimento de como o *spam* está sendo enviado. Foram identificados os países de origem mais freqüentes, as portas de acesso mais utilizadas, os domínios mais freqüentes, tanto dos remetentes, como dos destinatários e conteúdo disseminado.

1.1 Contribuições

Esse projeto vem para auxiliar no desenvolvimento de políticas e implementações de segurança da informação. O Mail-pot foi desenvolvido com o intuito de coletar dados para análise de mensagens indesejadas, gerando relatórios identificando, quando possível, a origem dos *spammers* e o conteúdo disseminado.

1.2 Estrutura da monografia

Está monografia está organizada da seguinte forma:

- Capítulo 1: Apresenta uma breve introdução do que é o trabalho.
- Capítulo 2: Descreve os conceitos básicos sobre elementos de segurança de informação, tipos de ataques e medidas de segurança.
- Capítulo 3: Descreve os conceitos de *honeynet* e *honeypot*, identificando suas características e arquiteturas, assim como aborda os conceitos de *spams* e suas particularidades.
- Capítulo 4: Apresenta a ferramenta desenvolvida, utilizada para a captura dos *spams,* assim como seu funcionamento e características.
- Capítulo 5: Conclusões e Trabalhos Futuros são apresentados as conclusões do trabalho e as sugestões de trabalhos futuros.

2 Elementos de Segurança da Informação

A necessidade de proteger as informações não é uma realidade apenas dos dias atuais. Ela existe desde os primórdios da espécie humana. Em tempos antigos, a informação era representada, por exemplo, através de manuscritos, objetos, cartas. Junto com a informação surgiu também à necessidade de compartilhá-la com outras pessoas de forma segura. Atualmente, as informações constituem o objeto de maior valor para as empresas e organizações, assim como para as pessoas. O progresso da informática e das redes de comunicações nos apresenta um novo cenário, no qual os objetos do mundo real estão representados através de sistemas, *e-mails*, documentos, planilhas, *etc*. Os objetos do mundo virtual podem ter valor igual ou maior aos dos objetos do mundo real. A proteção das informações engloba três aspectos:

- **Confidencialidade** – garantir que os dados enviados a um indivíduo qualquer cheguem a ele diretamente, ou seja, sem qualquer intervenção de qualquer outro individuo mesmo que seja apenas para leitura;
- **Disponibilidade** – garantir que a informação esteja sempre disponível para o uso legítimo, ou seja, por aqueles usuários autorizados pelo proprietário da informação;
- **Integridade** – garantir que a informação não seja destruída ou corrompida e o sistema tenha um desempenho correto. Ela deve ter assistido que os dados não foram alterados por um indivíduo não autorizado.

A segurança da informação tem como propósito proteger as informações, independente de onde elas estejam situadas. Ela é um processo contínuo onde o objetivo não é garantir total segurança, pois isto é impossível, mas sim medir e gerenciar os riscos relacionados aos negócios e aos ativos da informação, para que eles sejam mantidos em níveis aceitáveis [11].

2.1 Conceitos

Os principais conceitos de Segurança da Informação utilizados neste trabalho são os que seguem:

2.1.1 Ativos de Informação

Em segurança da informação, um ativo é qualquer elemento no qual são armazenadas informações e que representa valor para a organização, empresa ou pessoa, ou seja, aquilo que se deseja proteger. Eles podem ser classificados por quatro tipos:

- **tecnologia** – computadores, sistemas, mídia de *backup*, *etc*;
- **pessoa** – usuários, gestores, custodiantes, *etc*;
- **ambiente** – escritórios, sala de equipamentos, prédio da organização, *etc*;
- **processo** – procedimento de *backup*, normas de segurança, contratos com terceiros, *etc*.

2.1.2 Ameaça

Uma ameaça é um perigo em potencial que pode explorar uma falha de segurança na proteção dos ativos. Exemplos de ameaça são: fraude ou espionagem; roubo de informações; perda de dados; pirataria, *etc*. A consolidação de uma ameaça é feita por um agente de ameaça. Ameaças podem se concretizar por razões intencionais (*e.g.*, ação de um vírus na rede e roubo de senhas) ou acidentais (*e.g.*, falha de um computador). Uma boa prática para reduzir o número de incidentes é o treinamento e conscientização de usuários sobre a segurança da informação.

2.1.3 Vulnerabilidade

A vulnerabilidade pode ser qualquer falha em um sistema, procedimento, estrutura de rede, estrutura física, ou qualquer exposição indevida da informação, que possa ser explorada por uma ameaça. Elas podem ser geradas devido a, por exemplo, falta de atualizações de sistemas, ausência de controle de permissões, inexistência de *backups*, falta de preparação dos usuários.

2.1.4 Risco

O risco é a probabilidade de que algum incidente ocorra, provocando algum dano às informações. Em segurança da informação, o risco é entendido como a probabilidade de alguma ameaça se concretizar, gerando impacto nos negócios da organização. Para quantizar o risco relacionado com um ativo, é necessário medir a severidade das vulnerabilidades encontradas, a relevância deste ativo para os negócios da organização e probabilidade destas vulnerabilidades serem exploradas por uma ou mais ameaças.

2.2 Ataques

Ameaças podem utilizar diversas técnicas para obter sucesso na exploração de vulnerabilidades, ou seja, efetivar um ataque bem sucedido. A seguir serão apresentados alguns exemplos de técnicas utilizadas para ataques aos ativos de informação.

2.2.1 Engenharia Social

A engenharia social, no contexto da Segurança da Informação, consiste na arte de enganar as pessoas, e não as máquinas. O termo é utilizado para descrever um método de ataque, onde um agente de ameaça faz uso da persuasão, muitas vezes abusando da ingenuidade ou confiança do usuário, para obter informações que podem ser utilizadas para obtenção de acesso não autorizado a computadores ou informações. O meio mais eficaz para mitigar os riscos gerados pela engenharia social é a capacitação dos usuários em segurança da informação.

2.2.2 Softwares Maliciosos

Software malicioso ou simplesmente *malware* (*malicious software*) é um termo genérico que abrange todos os tipos de programas especificamente criados para executar ações mal intencionadas em um computador. Podem ser citados como exemplos:

- **vírus** – programa de computador que se propaga inserindo cópias de si mesmo e se tornando parte de outros programas e arquivos de um computador. O vírus depende da execução do programa ou arquivo hospedeiro para que possa se tornar ativo e dar continuidade ao processo de infecção;
- ***worms*** – programa capaz de se propagar automaticamente através de redes, enviando cópias de si mesmo de um computador para outro. Ao contrário do vírus, o *worm* não embute cópias de si mesmo em outros programas ou arquivos e não necessita ser explicitamente executado para se propagar. A sua propagação se dá através da exploração de vulnerabilidades existentes nos *softwares* instalados nos computadores;
- ***backdoors*** – software que invasores utilizam para garantir uma forma de retornar a um computador comprometido, sem precisar recorrer aos métodos utilizados na realização da invasão. Na maioria dos casos, também é intenção do invasor poder retornar ao computador comprometido sem ser notado.
- **cavalos de tróia (*trojans*)** – em segurança da informação, um cavalo de tróia é um programa, normalmente recebido como um "presente" (*e.g.*, cartão virtual, álbum de fotos, *etc*) que além de executar funções para as quais foi aparentemente projetado, também executa funções maliciosas de forma imperceptível;
- ***keyloggers*** – programa capaz de capturar e armazenar as teclas digitadas pelo usuário no teclado de um computador. Os *keyloggers* mais modernos também são capazes de capturar a região da tela onde o usuário clica com o *mouse*;
- ***phishing*** – técnica utilizada por invasores, onde estes enviam *e-mails* induzindo o usuário a clicar em algum *link* malicioso, levando à instalação de algum *software* malicioso em seu computador.

2.3 Medidas de Segurança

Nesta seção, serão apresentadas algumas medidas de segurança que devem ser utilizadas para diminuir os riscos relacionados às informações e os ativos.

2.3.1 Gestão de Riscos

O processo de gestão de riscos torna possível a identificação e a correta avaliação dos riscos associados aos ativos de informação que sustentam os negócios das organizações. Com um processo sistemático de identificação, análise, avaliação, tratamento, comunicação e revisão dos riscos, é possível traçar a evolução do nível do risco nos ativos, priorizando, desta forma, os investimentos e iniciativas para a redução dos riscos.

Para que os riscos sejam corretamente tratados, a primeira ação a ser tomada é o conhecimento adequado sobre estes riscos pertinentes às atividades cotidianas das organizações. O processo utilizado para isso é a **Análise de Riscos** [12].

Uma vez tendo os riscos mapeados, é possível ter a avaliação destes com base nos critérios de riscos estabelecidos pela direção da empresa. Estes critérios refletirão o apetite de risco da organização, ou seja, o quanto ela tolerará os riscos encontrados. O resultado desta ação de avaliação de riscos será o plano de ação com as medidas necessárias para tratamento dos riscos, que tem como objetivo reduzir os riscos encontrados até níveis aceitáveis através de implementações de controles de segurança nos ambientes analisados.

Destaca-se no processo de gestão de riscos as atividades de comunicação e revisão. Paralelamente a cada ação citada anteriormente, deve-se realizar uma comunicação eficaz para todas as partes envolvidas.

Adicionalmente, apenas um ciclo de análises e tratamento de riscos não produz um nível de excelência na organização. É fundamental que o processo de gestão de riscos seja cíclico, monitorando-se a evolução dos riscos [11].

2.3.2 Política de Segurança

A política de segurança é um conjunto de diretrizes, normas e procedimentos que permitem que as pessoas possam desempenhar suas atividades diárias dentro de um padrão de segurança que esteja alinhado com as necessidades da organização [13].

2.3.3 Implementações de Segurança

São ações tomadas para proteger a informação. Pode ser na forma de: uso de tecnologias; criação de procedimentos; treinamentos; uso de medidas de segurança física. A definição das ações mais urgentes pode ser feita a partir dos resultados de análises de riscos.

A utilização de técnicas de emulação de serviços para monitorar as ações de invasores, também vem sendo bastante difundida. São soluções baratas e servem para uma infinidade de aplicações, desde o monitoramento de invasões em uma rede privada ou pessoal ao monitoramento de *spams* que inundam um servidor de correio. A partir dos dados coletados é feito um levantamento das vulnerabilidades e origem de ataques, onde a organização adota uma política de segurança e a implementa baseada nesse estudo preliminar.

3 O Projeto Honeynet

Com a falta de informação de como agiam os *blackhat*, quase sempre chamados de *hackers*, foi fundada em 1999 a primeira *honeynet*, uma rede criada para ser comprometida. O projeto foi formado por 30 profissionais da segurança dedicados a aprender as ferramentas, táticas e os motivos dos *blackhats* e compartilhar das lições aprendidas. O grupo aprende criando sistemas de produção e, em seguida, monitorando toda a atividade de e para esses sistemas.

3.1 Honeynets

Antes das *honeynets*, todos os outros recursos para combater falhas de segurança eram passivos, *Firewalls*, *IDS*, *Proxies*, todos utilizados para proteger passivamente, o uso de mecanismos para observação das atividades de invasores em redes conectadas à internet é utilizado na prática há um bom tempo no mundo da tecnologia da informação.

Contudo, as primeiras experiências na área datam de 1988, quando o especialista Clifford Stoll, em seu livro: ***The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer Espionage*** [14]***,*** faz um relato completo sobre a história da invasão (origem do ataque, motivos e redes-alvo) nos sistemas do *Lawrence Berkeley Laboratory* (LBL).

Quatro anos depois, em 1992, seria a vez do especialista Bill Cheswick explicar no artigo ***An Evening with Berferd In Which a Cracker is Lured, Endured, and Studied*** [15] os resultados do acompanhamento de invasões em um dos sistemas da AT&T (American Telephone and Telegraph), projetado especialmente para este fim.

O termo *honeypot* só surgiria em meados de 1998, quando Fred Cohen desenvolveu a ferramenta *Detection Toolkit* (DTK) [16], mas as intenções eram as mesmas: configurar um ou mais sistemas que parecem atraentes para os invasores de redes, mas que também podem monitorar com um grau mais alto de precisão o que está acontecendo. A ferramenta foi a primeira utilizada para emulação de diversas vulnerabilidades e coleta de informações sobre os ataques sofridos.

Mas é em 1999, quando um grupo de especialistas em segurança da informação liderado por Lance Spitzner lança o ***Honeynet Project*** [5], uma rede projetada exclusivamente para ser comprometida por ataques. O conceito de *honeynets* ganha repercussão mundial e demonstra a importância do estudo do comportamento dos invasores de uma rede para o desenvolvimento de novas ferramentas e sistemas de defesa.

3.1.1 Arquitetura

As *honeynets* têm seus projetos baseados no tipo de vulnerabilidade que quer estudar, logo são personalizadas de acordo com necessidades específicas. Por isso, os projetos de *honeynets* variam quanto a arquitetura, tendo cada projeto sua própria topologia, sistemas operacionais, ferramentas usadas, desenvolvimento de outras ferramentas para análise e contenção do tráfego.

Uma *honeynet* utiliza o conceito de camadas para a captura dos dados. Quanto mais camadas de informações tiverem, mais fácil será analisar um ataque e manter um nível de segurança aceitável caso alguma camada falhe. Ela é formada basicamente por três áreas - a Internet, a *honeynet* e a rede administrativa - separadas por um *firewall*. A Internet é uma rede não confiável porque, por padrão, o tráfego não é criptografado e é dela que vem qualquer tráfego mal intencionado.

A *honeynet* é o conjunto de *honeypots* que se destinam a serem comprometidos. Cada dispositivo da rede é considerado um *honeypot*, uma vez que eles também podem ser atacados. A rede administrativa é uma rede confiável na qual se coletam remotamente os dados e administra a rede *honeynet*. Todo o tráfego deve passar primeiro pelo *firewall* e ter cuidado com a segmentação e o controle de acesso.

Uma das maiores preocupações dos Projetos de *Honeynet*s espalhados pelo mundo é não deixar que sua rede de *honeypots* seja usada por *blackhats* para fins escusos, disseminando *malwares*, *spam*, *etc*.

A Figura 1 é um exemplo de topologia de uma h*oneynet*, com sistemas de alta interação, sistemas não emulados, prontos para serem comprometidos e um *Honeywall Gateway,* servidor dedicado, situado entre o roteador e os *honeypots*.

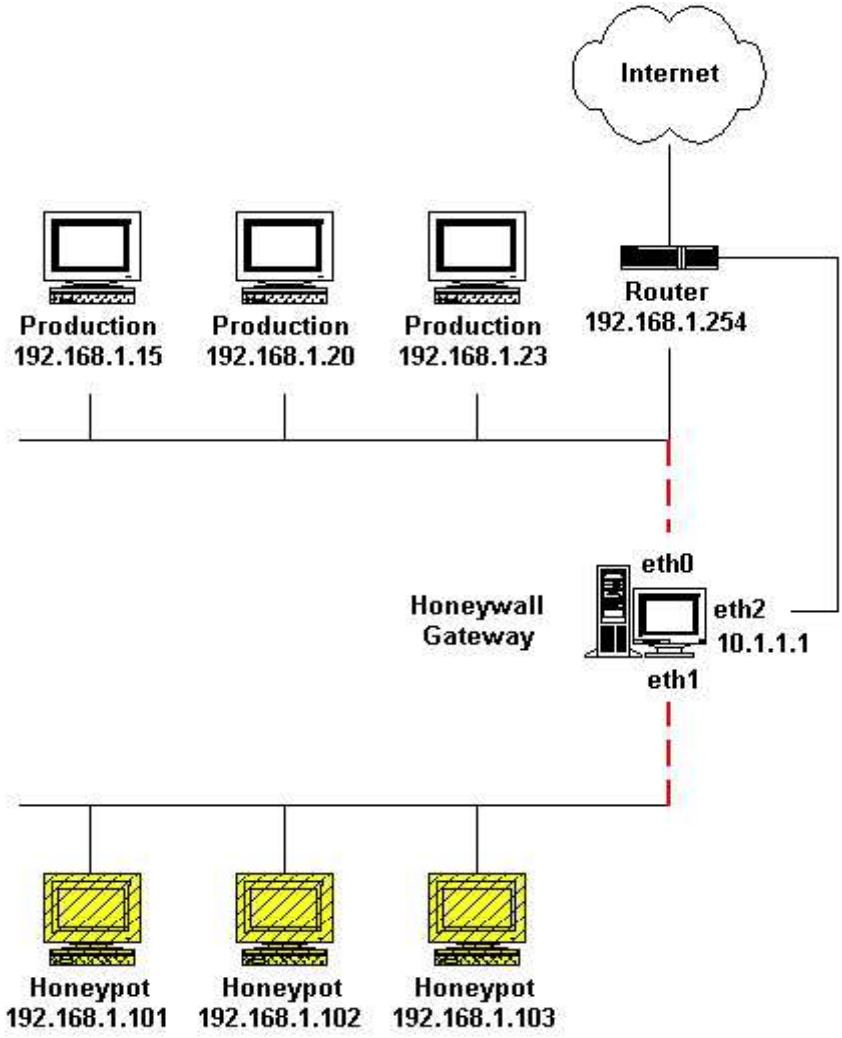

Figura 1. Topologia de uma *Honeynet* [5]

Dentre as funções do *Honeywall*, além de atribuir um servidor *Sebek* [18] *(*capturador de teclas e atividades), estão a de armazenar *logs* do *firewall* (*IPTables*) [19], do sistema (*syslog*) [20], do detector de intrusões (*Snor*t) [21] e do analisador de tráfego (*tcpdump*) [22], e de realizar o controle de conexões externas (*session limit*), através da ferramenta *Snort-inline* [23]. A união de todos esses mecanismos permite a realização do **controle** e **captura dos dados**.

Devido à versatilidade desses modelos, os clientes são independentes de sistema operacional. A flexibilidade premeditada permite que se tenha uma *honeynet*, ou um sistema misto de *honeypots*, dependendo da instalação escolhida para o cliente.

Porém, as *honeynets* têm suas limitações. Elas são primariamente uma ferramenta de aprendizado, as quais são usadas para a pesquisa e coleta de

informações. Elas não são a solução definitiva para todos os problemas de segurança. As *honeynets* vêm para agregar valor onde são implantadas, sendo usadas para guiar futuras políticas e implantações de segurança.

3.1.2 Honeypots

*Honeypo*t é um recurso de segurança preparado especificamente para ser sondado, atacado ou comprometido e para registrar essas atividades [18]. Já *honeynet* é uma rede projetada especificamente para ser comprometida e utilizada para observar os invasores. *Honeypots* são unidades de uma topologia de *honeynet* que, normalmente, é composta por sistemas reais e necessita de mecanismos de contenção eficientes e transparentes para evitar que seus sistemas sejam usados para ações mal intencionadas para outras redes.

3.1.3 Tipos de Honeypots

Existem dois tipos de *Honeypots*:

- *Honeypots* de baixa interação: apenas emulam serviços e sistemas operacionais, não permitindo que o atacante interaja com o sistema, normalmente usa-se o programa Honeyd [24][25], um pequeno *daemon* que cria *hosts* virtuais em uma rede [26];
- *Honeypots* de média interação: a interação entre a h*oneypot* e o atacante é maior, mas não equivale a um sistema real. Os serviços oferecidos ainda são emulados, mas estes respondem as requisições do atacante como serviços reais. Desta forma mais dados sobre o ataque são obtidos. Devido ao maior grau de interação com o sistema, os riscos também aumentam;
- *Honeypots* de alta interação: são compostos por sistemas operacionais e serviços reais e permitem que o atacante interaja com o sistema.

3.1.4 Aplicação

O valor dos *honeypots*/*honeynets* baseia-se no fato de que tudo o que é observado é suspeito e potencialmente malicioso, e sua aplicação depende do tipo de resultado que se quer alcançar.

Honeypots de baixa interatividade oferecem baixo risco de comprometimento e são indicados para redes de produção, quando não há pessoal e/ou *hardware*

disponível para manter uma *honeynet*, ou quando o risco de um *honeypot* de alta interatividade não é aceitável.

Normalmente, o uso de *honeypots* de baixa interatividade também está associado aos seguintes objetivos:

- detectar ataques internos;
- identificar varreduras e ataques automatizados;
- identificar tendências;
- manter atacantes afastados de sistemas importantes;
- coletar assinaturas de ataques;
- detectar máquinas comprometidas ou com problemas de configuração;
- coletar código malicioso (*malware*).

Já *honeypots* de alta interatividade são indicados para redes de pesquisa. Podem ser utilizados para os mesmos propósitos que os *honeypots* de baixa interatividade, mas introduzem um alto risco para instituição, e são justificáveis quando o objetivo é estudar o comportamento dos invasores, suas motivações, além de analisar detalhadamente as ferramentas utilizadas e vulnerabilidades exploradas. É importante lembrar que o uso de *honeypots* de alta interatividade demanda tempo, pessoal mais qualificado e técnicas de contenção mais eficientes.

A Tabela 1 pode auxiliar na decisão sobre que tipo de *honeypot* deve ser implementado em uma instituição.

Tabela 1. Comparativo entre os tipos de *honeypots*

Baixa Interatividade	Alta Interatividade
Emulam sistemas e serviços	Executam as versões reais
Simples. Fácil gerenciamento	Cuidados na instalação e configuração. Coleta de artefatos
Atacante não tem controle	Controle total
Ações limitadas, captura de tráfego e *malware*	Captura demais informações, incluindo ferramentas e comandos
Difíceis de iludir atacantes avançados/determinados	Difíceis de iludir atacantes avançados/determinados

O projeto brasileiro de *honeynet* [27] utiliza *honeypots* de alta interação. Eles são sistemas reais, porém com algumas modificações que permitem a captura de todos os dados, inclusive os criptografados. De certo modo, como descrito na literatura, a própria *honeynet* pode ser considerada como um único *honeypot* composto por diversos sistemas. Eles, porém, preferem referir a cada sistema individual dentro da *honeynet* como um *honeypot* individual.

3.1.5 Honeynets no Brasil

Foram criada e mantida em parceria por especialistas do Instituto Nacional de Pesquisas Espaciais (INPE) [28] e do grupo brasileiro de resposta a incidentes de segurança NBSO (NIC BR *Security Office*) [29][30].

O projeto teve início com uma palestra do especialista Lance Spitzner, um dos criadores do *Honeynet Project*. Em junho de 2000, tiveram a chance de assistir uma apresentação de Lance Spitzner sobre o conceito de *honeynets* e seu potencial. Depois disso, passaram a acompanhar o progresso do projeto. Outros contatos com ele e com o projeto surgiram, mas ainda não havia uma estrutura para implementar uma *honeynet* no Brasil.

Apesar das dificuldades, a idéia ganharia força a partir de 2001. Nessa época, o INPE e o NBSO iniciaram uma cooperação maior e, no final daquele ano, surgiu à

idéia de implementar um protótipo do projeto como um laboratório do Curso de Pós-graduação em Segurança de Sistemas de Informação do INPE.

Assim, em março de 2002 começaram as operações do *Honeynet.BR*. Três meses após o seu lançamento, o projeto receberia um importante reconhecimento. Devido à orientação do projeto à pesquisa, em junho de 2002 o Projeto *Honeynet.BR* tornou-se membro da *Honeynet Research Alliance*, que reúne diversos grupos de várias partes do mundo, todos empenhados em desenvolver a tecnologia de *honeynets*.

A idéia surgiu no Programa de Engenharia de Sistemas e Computação (PESC) da COPPE/UFRJ (Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia da Universidade Federal do Rio de Janeiro). Vendo à oportunidade de desenvolver um trabalho semelhante ao conhecido *Honeynet Project*, estudando de forma mais aprofundada os ataques que ocorrem na Internet [31].

3.1.6 Custos e profissionais habilitados

A viabilização de uma *honeynet* gera custos similares aos gastos numa rede normal. Os custos são de equivalentes aos envolvidos em manter uma rede conectada à Internet e envolvem conectividade, máquinas, espaço físico e *etc* [5].

O uso de *software* livre é uma boa alternativa para baratear o projeto. Porém, devem-se pesar os outros tipos de despesas e dificuldades associados ao *software* livre, como: capacitação e suporte especializado. Na rede administrativa, existe a possibilidade de utilização de *software* livre para todas as atividades, reduzindo o custo ao *hardware*. No caso dos *honeypots* vai depender do sistema operacional e dos aplicativos que se pretende instalar. Caso os *honeypots* sejam todos baseados em *software* livre, o custo será reduzido. Porém, se houver a intenção de utilizar sistemas como o Windows, Solaris, etc, então os custos com licença de *software* necessitam ser levados em consideração [32]. Em ambos os casos, o *hardware* utilizado não necessita ser topo de linha.

O *Honeynet.BR*, iniciou sua operação contando com equipamentos doados. O projeto, inicialmente, utilizou equipamentos disponibilizados pelo INPE e por

membros do projeto. Mas também recebeu doações de equipamentos por parte do Ministério da Ciência e Tecnologia (MCT) [33].

Além dos equipamentos necessários, outro ponto importante é o perfil do profissional que estará atuando nesse tipo de projeto. Conhecimentos profundos em TCP/IP e *Firewall* são alguns dos principais requisitos exigidos.

3.2 Spam

Spam é o termo usado para referir-se aos *e-mails* não solicitados, que geralmente são enviados para um grande número de pessoas. Quando o conteúdo é exclusivamente comercial, esse tipo de mensagem é chamada de UCE (do inglês *Unsolicited Commercial E-mail*) [33].

3.2.1 Spam zombies

O recebimento de *spams* é um tanto normal nas contas de e-mails. Os *spams* têm diversas origens e conteúdos, muitos são propagandas de medicamentos, produtos, já outros com algum tipo de código malicioso. Essa última categoria, quando acessada pode comprometer seu computador, normalmente com: *worms, bots,* vírus e cavalos de tróia. Estes códigos maliciosos, uma vez instalados, permitem que *spammers* utilizem a máquina para o envio de *spam*, sem o conhecimento do usuário. Enquanto utilizam máquinas comprometidas para executar suas atividades, dificultam a identificação da origem do *spam* e dos autores também. Os *spam zombies* são muito explorados pelos *spammers*, por proporcionar o anonimato que tanto os protege [35].

3.2.2 Motivadores de envio de *spam*

A Internet causou grande impacto na vida das pessoas, tornando-se um veículo de comunicação importante, evoluindo para revolucionar a maneira de fazer negócios e buscar e disponibilizar informações. Ela viabiliza a realidade da globalização nas diversas áreas da economia e do conhecimento. Por outro lado, esse canal acabou absorvendo diversas práticas ruins.

O *spam* é uma das práticas ruins. Ele ficou famoso ao ser considerado um tormento para os usuários de *e-mail*, impactando na produtividade de funcionários e

degradando o desempenho de sistemas e redes. No entanto, poucos se lembram de que já enfrentou algo semelhante, antes de utilizar o *e-mail* como ferramenta de comunicação.

As cartas de correntes para obtenção de dinheiro fácil, encontradas nas caixas de correio, as dezenas de panfletos recebidos nas esquinas e as ligações telefônicas oferecendo produtos são os precursores do *spam*. A principal diferença, extremamente relevante, é o fato de que para enviar cartas ou panfletos e ligar para nossas casas, o remetente tinha de fazer algum investimento. Este muitas vezes inviabilizava o envio de material de propaganda em grande escala.

Com o surgimento e a popularização da Internet e, conseqüentemente, do uso do *e-mail*, remetentes de cartas de corrente ou propagandas obtiveram a oportunidade e a facilidade de atingir um número muito maior de destinatários. Tudo isso com a vantagem de investir muito pouco ou nada para alcançar os mesmos objetivos em uma escala muito maior. Por essa razão, esse é um dos maiores motivadores para o envio de *spam*.

Desde o primeiro *spam* registrado e batizado como tal, em 1994, essa prática tem evoluído, acompanhando o desenvolvimento da Internet e de novas aplicações e tecnologias. Atualmente, o *spam* está associado a ataques à segurança da Internet e do usuário, propagando vírus e golpes. Tão preocupante quanto o aumento desenfreado do volume de *spam* na rede, é a sua natureza e seus objetivos.

O *spam* ganhou popularidade, é tema tratado em vários sites e protagonista de notícias na imprensa, muitas vezes abordando mecanismos de prevenção ou defesa. O combate ao *spam* e o desenvolvimento de mecanismos de prevenção e proteção tornaram-se serviços de destaque oferecidos por provedores de acesso e empresas fabricantes de *software/hardware*.

3.3 Projeto Spampot

O CERT.br fornecem uma visão do problema do *spam* em redes brasileiras, através da análise das reclamações recebidas. Estas reclamações se dividem em

reclamações de envio de *spam*, de páginas que fazem propaganda de produtos oferecidos em *spam*, de abusos de *relays* e *proxies* abertos [35]

Nos últimos anos as reclamações de *spams* enviados viam o abuso de máquinas brasileiras com *proxies* abertos ou *proxies* instalados por códigos maliciosos, têm sido de 30% a 40% do total de reclamações. Porém, praticamente nenhum dado existe sobre a natureza, a origem ou destino desse tipo de *spam*. Sendo extremamente importante a obtenção de métricas que permitam entender melhor o perfil do abuso de *proxies* no Brasil, de modo a facilitar a proposição de formas de prevenção mais efetivas.

O objetivo do projeto é obter, através de *honeypots* de baixa interatividade, dados relativos ao abuso de máquinas conectadas via redes de banda larga para envio de *spam*. Este projeto tem seu foco nas redes ADSL e Cabo nas versões doméstica e empresarial que possuam IP roteável [36].

3.3.1 Arquitetura

O projeto utiliza *honeypots* de baixa interatividade e quando um atacante interage com um destes *honeypots*, ele não está interagindo diretamente com o sistema, mas sim com um programa que emula suas características, como sistema operacional e versões de aplicativos. A Figura 2 apresenta a arquitetura do Spampot.

Esses *honeypots* foram instalados em 5 operadoras diferentes de cabo e DSL, em conexões residenciais e comerciais.

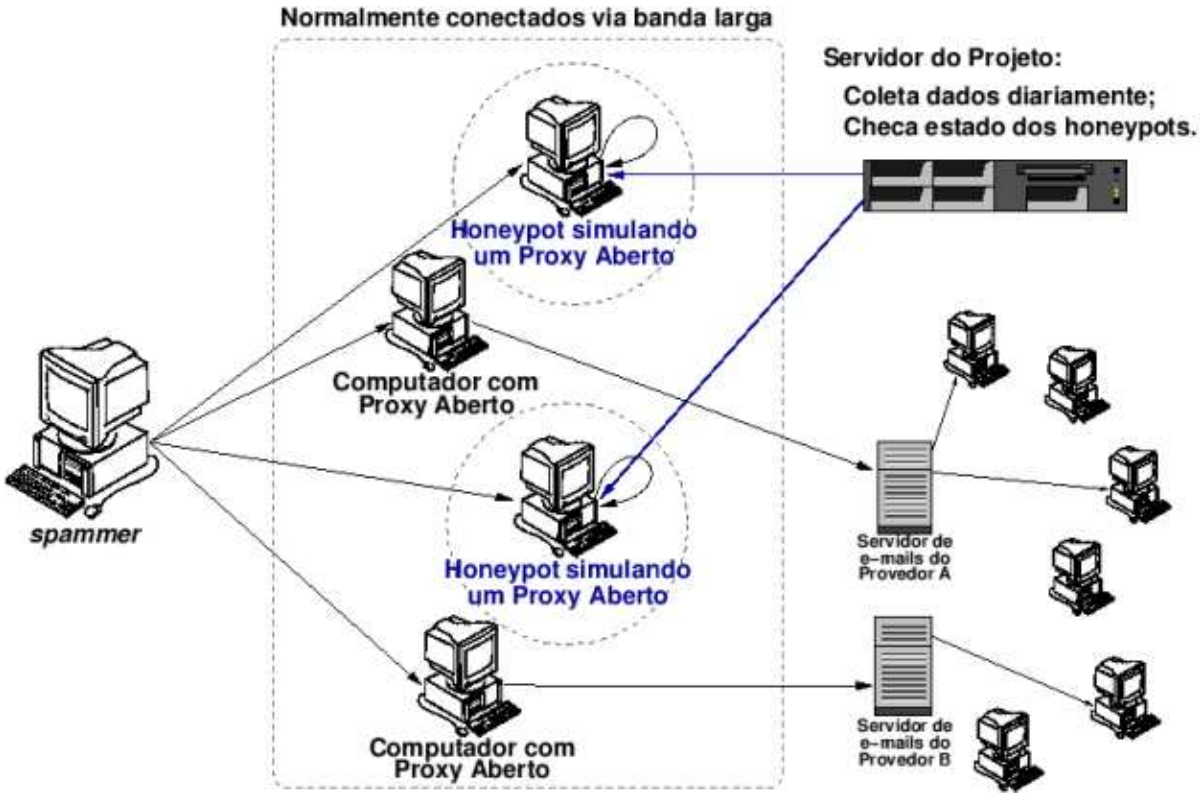

Figura 2. Arquitetura de um *Spampot* [7]

Um *spammer* que tentar abusar de um desses *honeypots* para o envio de *spam*, estará interagindo com programas projetados para fazê-lo acreditar que está conseguindo enviar seus *e-mails*. Deste modo, nenhum *spam* foi realmente enviado aos destinatários, mas apenas coletado para análise por um servidor central.

4 Experimento

Neste capítulo, apresentamos a concepção e desenvolvimento de uma ferramenta para o estudo dos *spams*. Essa ameaça que vem normalmente de servidores infectados por algum tipo de *malware,* onde são controlados remotamente para enviar *spam* ou por falhas de segurança exploradas pelos *spammers*.

4.1 Objetivo

Spam é uma preocupação mundial, pois cada dia mais recebemos lixo eletrônico em nossas caixas postais. Por mais medidas que se tomem e por mais segurança que se aplique para tentar derrubar, ainda assim recebemos spam. O objetivo do aplicativo desenvolvido, intitulado Mail-pot, é deixar uma máquina se passando por um servidor de correio. O Mail-pot fará os *spammers* pensar que estão conseguindo enviar *spams* utilizando uma falha de configuração no "servidor". O servidor será programado para aceitar os *spams*, mas não repassá-los, apenas armazená-los em banco de dados.

Após um período de coleta dos *spams*, ocorrerá à fase da análise no banco de dados. Nela será verificado, por exemplo, qual o país de origem que mais tentou enviar *spam*, qual o país que seria mais atingido, que tipo de *spam* está sendo enviado (*spam, phishing, scam, vírus*), qual o tipo de propaganda que mais tentam difundir (venda de medicamentos *on-line*, pornografia infantil, utensílios domésticos).

4.2 Emulador do Servidor SMTP

O Mail-pot utiliza o protocolo SMTP que é normalmente associado à porta 25/TCP, onde seu principal objetivo é o transporte de e-mails de uma maneira confiável e eficiente [8]. Este tipo de servidor é utilizado abusivamente para enviar *spam*, uma vez que dificulta a detecção da origem real. Também é utilizado para burlar listas de bloqueio de e-mail.

Os *spammers* scaneiam a Internet a procura de falhas em servidores. Uma vez localizadas, eles iniciam a exploração e o envio dos *spams.*

O servidor SMTP emulado foi implementado em JAVA e utiliza banco de dados MySQL. Ele pode rodar em qualquer sistema operacional que possua um JRE instalado e uma conexão com a internet.

O servidor está configurado na porta de correio padrão (25) e aberto para qualquer exploração, não foi feito nenhuma política de segurança, nem tão pouco facilitação para os *spammers* agirem.

Os dois principais elementos do projeto são a captura de dados e análise. A captura de dados foi realizada no período de 30 dias, onde foi filtrado todo e-mail "enviado" pelo Mail-pot. O elemento crítico é controlar o número de envios por destinatário (*spammers*) para evitar que apenas uma fonte tome conta, por tempo indeterminado do Mail-pot, mascarando as análises dos dados.

O servidor recebe o tráfego da porta 25/TCP (*Transmission Control Protocol*) e simula as repostas para o remetente de um servidor SMTP comum. As mensagens são aceitas e armazenadas localmente. O *spammer* ao receber a confirmação de envio do "servidor" acredita que o seu e-mail foi enviado, porém eles nunca são entregues aos destinatários, mas sim, armazenadas no banco de dados.

Esse módulo teve apenas algumas funções implementadas de um servidor SMTP real:

- Helo ou Ehlo;
- Mail From;
- Rcpt To;
- Data.

Ao iniciar uma sessão com o emulador a função "helo" retornará uma mensagem com o código 220, indicando que se estabeleceu a conexão:

```
220 mail.abcdario.com.br
```

Para iniciar uma sessão de e-mail, o *spammers* envia um comando MAIL informando o remetente::

```
Mail from: email@dominio.terminacao
250 2.1.0 user OK
```

O servidor sempre retornará “ok”, já que ele não tenta validar o e-mail, caso a construção esteja de acordo. Em seguida, é identificado o destinatário usando o comando RCPT:

```
Rcpt to: email@dominio.terminacao
250 2.1.0 user OK
```

Logo após, deve-se enviar o conteúdo da mensagem por meio do comando DATA, porém dessa vez aparece uma resposta com código 354 que confirma o campo para inserção da mensagem e informa como terminá-la.

```
Data
354 Enter mail, end with \".\" on a line by itself
```

No campo data, pode ser inserido o subject. Depois do subject e da mensagem inserida, incluir um “.” na última linha para confirmar o envio. Pronto, o emulador lhe retornará uma mensagem informando que o e-mail foi enviado, mas como o projeto propõe, o *spam* não é enviado, e sim armazenado em banco de dados para a realização de estudo sobre eles.

```
250 2.0.0 k3C6UDVc028967 Message accepted for delivery
```

A conexão é pertinente até o envio do comando QUIT, onde retorna uma mensagem encerrando a conexão com o Mail-pot:

```
221 2.0.0 mail.abcdario.com.br closing connection
```

Depois do encerrado uma conexão o emulador salva o e-mail de origem e destino destrinchando em login, domínio e terminação. Armazena em outras tabelas o assunto, hora e a mensagem, assim como o IP de origem e portas utilizadas.

5 Resultados

Os dados recolhidos permitiram uma melhor análise do problema *spam* e também um melhor entendimento de como o *spam* está sendo enviado. Neste capítulo abordaremos as fases de **captura e análise dos dados.**

5.1 Captura dos dados

O Mail-pot foi instalado em um servidor Linux, distribuição Fedora Core release 4, de configurações simples. O sistema operacional não tinha nenhuma camada de segurança a mais que os padrões desse tipo de distribuição. O Mail-pot funciona de maneira similar a um servidor SMTP, porém sem as funções de repassar o tráfego recebido.

Foi verificado que os *spammers* não interagiram com emulador por muito tempo, desde o envio da primeira mensagem. Como não foi feito nenhum cadastro em sites que poderiam enviar *spams* através das contas de *e-mail* cadastradas, o Mail-pot ficou dias sem nenhuma atividade. Decidiu-se então repassar a mensagem para o destino especificado no *e-mail*.

Horas após a primeira etapa do experimento o emulador se encontrava fora do ar. O Mail-pot não suportou a quantidade de solicitações simultâneas e encerrou seu funcionamento. Constatou-se que a única mensagem que continha era apenas uma isca para verificar a veracidade do “servidor”. A partir de então, o servidor começou a ser “visitado” por diferentes *spammers.*

A captura das informações foi realizada num período de 30 dias, onde as conexões com o Mail-pot aumentavam gradativamente. Todo tráfego que flui na ferramenta passa apenas por uma *interface*. Isso aumenta o risco, uma vez que o armazenamento das informações capturas fica contido num banco de dados na mesma máquina, situação que poderia ser facilmente detectada, caso tivesse uma intervenção humana anteriormente a interação dos *spammers.*

Por existir apenas uma camada, o cuidado e a manutenção do Mail-pot exigiu uma atenção constante. Diariamente eram feitas cópias de segurança para evitar qualquer invasão à máquina e que os dados fossem formatados. Contudo, em caso

de comprometimento do sistema, não foi elaborado nenhum procedimento de reação.

5.2 Análise dos dados

Nesse aplicativo não foi necessária uma filtragem dos *e-mails* que são armazenados em banco de dados, pois diferentemente dos aplicativos padrões, onde pode levar semanas para identificar *spams* que trafegam na rede, no Mail-pot todo o tráfego é suspeito. Assim sendo, todo o tráfego no servidor é potencialmente informação útil.

O armazenamento das atividades foi feito em um banco de dados MySQL, dividido em 8 tabelas:

- Email;
- Dominio;
- Final;
- Emaildominio;
- Msgs;
- Msgsdata;
- Msgsip;
- Srcs.

A tabela Email armazena os *e-mails* de origem, normalmente falsos, e os *e-mails* de destino. O domínio e a terminação (*top-level*) *de* cada *e-mail* são armazenados nas tabelas Dominio e Final, respectivamente. A tabela Emaildominio é utilizada para fazer a contagem dos *e-mails,* domínios e terminações que aparecem nos *spams*. A tabela Msgs tem a data de interação dos *spammers* com o Mail-pot e o assunto de cada mensagem, assim como as referências para identificar os *e-mails* de origem e destino. A tabela Msgsdata trás o conteúdo de todos os *spams*. Por último, as tabelas Msgsip trás as portas de origem dos *spammers* e a tabela Srcs trás os IPs de origem. Toda a lógica do banco é feita para não escrever nenhuma entrada repetida, utilizando as tabelas Emaildominio, Msgsdata e Msgsip para fazer a contagem das interações.

Tendo, portanto, as tabelas seus dados inseridos, utilizou o pacote *Microsoft Office,* especificamente o programa Excel, para gerar as tabelas e gráficos. Cada gráfico apresenta apenas as maiores ocorrências, deixando, portanto, as menores representadas pela denominação “outros”, indicando ocasiões distintas e que pouco se repetem entre os dados analisados.

O estudo proposto começa com a localização dos *spammers.* Verificou-se uma maior incidência do Brasil no envio, porém uma predominância regional asiática. Países como China, Turquia, Coréia, Índia e Tailândia apareceram em massa, porém foi freqüente a presença de outros países do sudeste asiático como: Filipinas, Vietnã, Singapura e Indonésia. Nas Américas foi verificada forte presença dos Estados Unidos e algumas ocorrências da Colômbia. Outro forte disseminador de *spams* verificado foi o continente Europeu, a Rússia e países como Itália, Polônia e França foram identificados nas pesquisas.

A Figura 3 apresenta um gráfico mais detalhado dos países que mais enviaram *spam* no período de coleta.

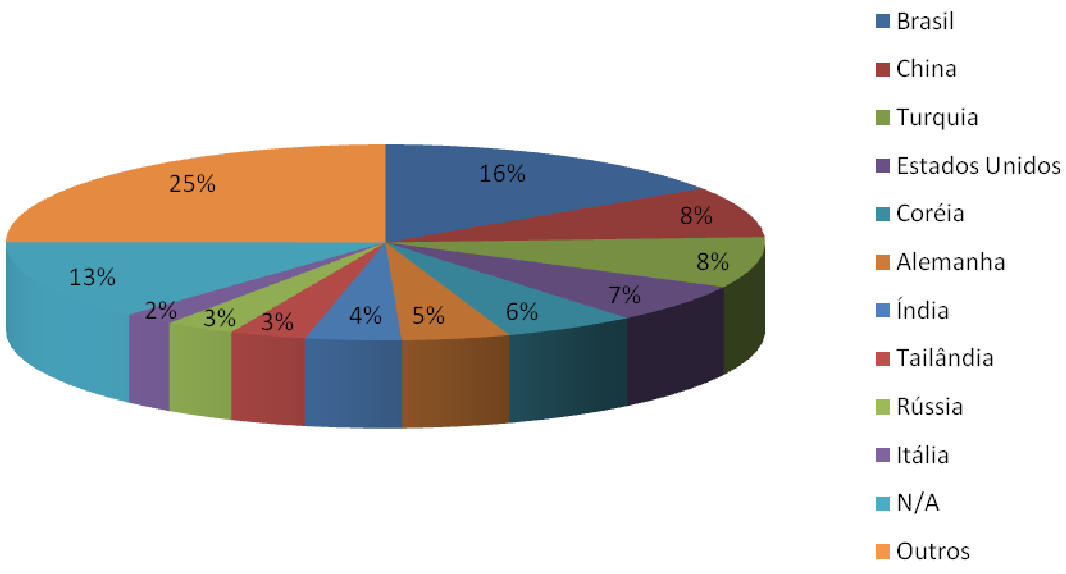

Figura 3. Países que mais enviaram spams

As ocorrências N/A significam que não foi identificado o endereço encontrado como um endereço válido, possivelmente trata-se de informações falsas para evitar a identificação das fontes emissoras de *spams.* A seguir, a Tabela 2 faz a amostragem quantitativa dos *spams* por região.

Tabela 2. Quantidade de *spams* enviados por país

PAÍS	Spams (Unidades)
Brasil	58
China	31
Turquia	31
Estados Unidos	24
Coréia	21
Alemanha	17
Índia	15
Tailândia	12
Rússia	11
Itália	8
N/A	49
Outros	92

A Figura 4 denota os países onde foram identificados *spammers.* Os pontos em vermelho indicam uma ocorrência no local especificado. O mapa ajuda a visualizar melhor todos os países e concentrações de *spammers* por localidade. Nele verificamos que a Oceania não teve nenhuma ocorrência, assim como a América Central. No continente africano foi constato apenas três ocorrências no Marrocos e apenas uma ocorrência em Benin. Verifica-se também a forte concentração dos disseminadores de *spams* no Brasil, continente europeu e sudeste asiático.

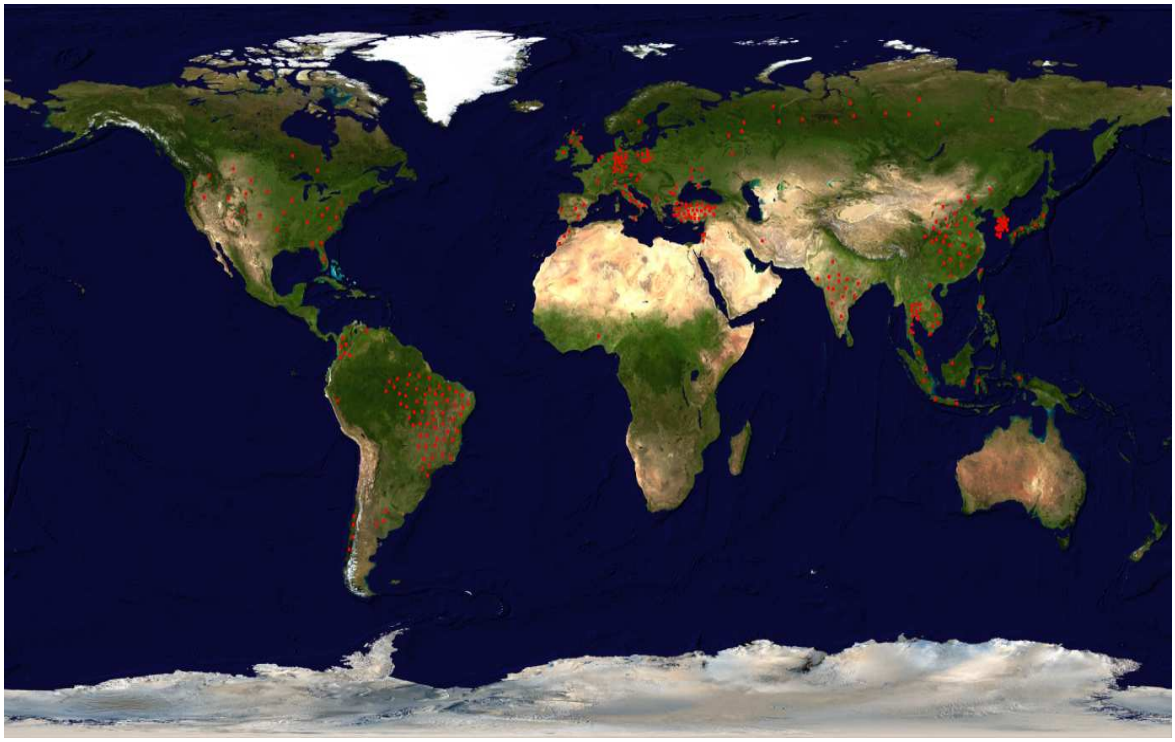

Figura 4. Mapa com todas as localidades dos *spammers* identificados.

A Figura 5 apresenta as maiores ocorrências de domínios com suas respectivas porcentagens. Verificou-se que, tanto para envio, tanto para recebimento dos *spams* houve uma concentração para a terminação ".com.br". Foi identificado fontes de várias localidades como apresentado na tabela acima, porém o destino de todos os *spams* foram o Brasil.

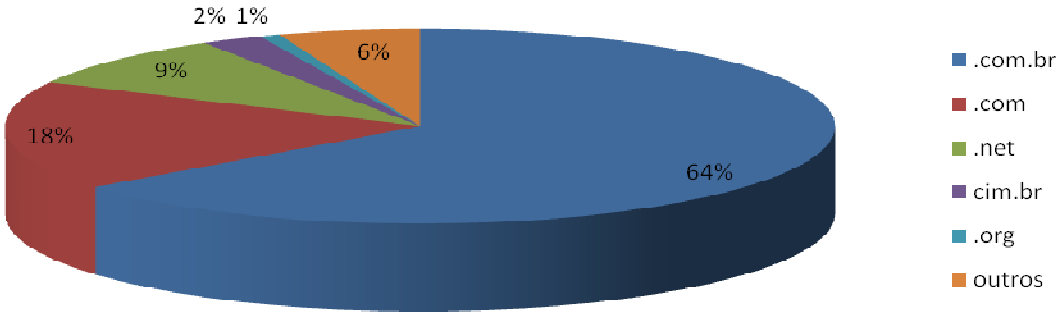

Figura 5. Relação das terminações de *e-mails* mais utilizadas para envio e recebimento de *spams*

Na Tabela 3 é apresentada a soma de todas as aparições de terminações (*top-level*). É denotada uma quantidade acima do dobro de mensagens, que seria a quantidade de *e-mails* do remetente e *e-mails* do destinatário, porém como grande parte dessas informações foi mascarada, o aplicativo auxiliar desenvolvido apenas

para a fase de análise retorna todas as ocorrências de *e-mails* no campo da mensagem, por isso a quantidade excessiva de terminações.

Tabela 3. Quantidade apresenta por cada terminação

Terminação	Quantidade (Unidade)
.com.br	966
.com	273
.net	133
.cim.br	36
.org	12
Outros	83

Outro parâmetro utilizado nos estudos foi à porta de origem, porta utilizada pelo *spammer* para iniciar a conexão com o Mail-pot. Foi verificada uma grande semelhança nas portas de acesso, onde, mesmo sendo constatado que mais de um *spammer* utilizava o Mail-pot, as interações usavam possíveis incrementos de portas para acessar o emulador. A Figura 6 apresenta esse comportamento.

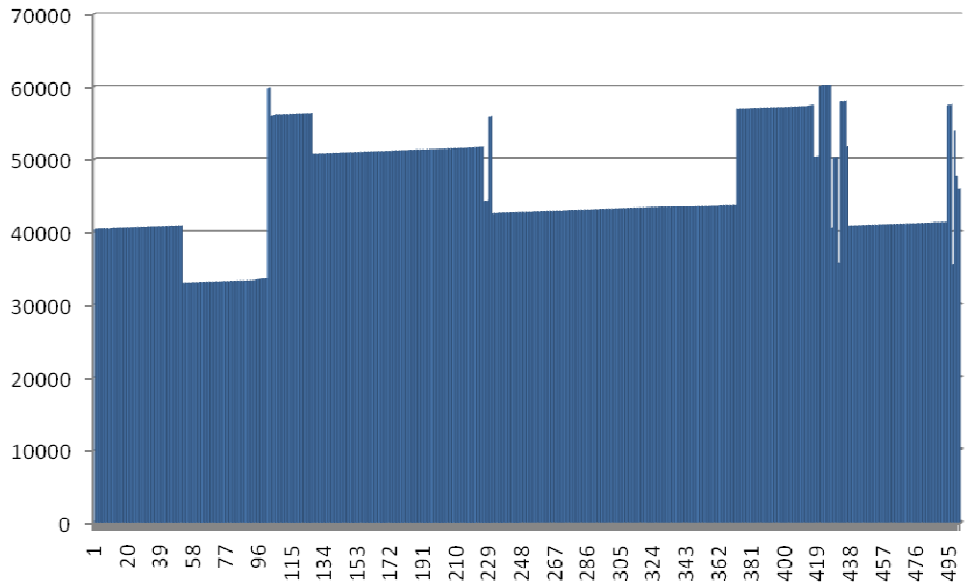

Figura 6. Relação dos *spams* (eixo das abscissas) com as portas que originaram as conexões (eixo das coordenadas).

A Figura 7(a) e a Figura 7(b) exemplificam melhor as diferenças encontradas. Enquanto na primeira figura evidenciamos um comportamento com certa tendência, onde as portas de acesso são apenas incrementadas em pequenas unidades, na segunda figura o aplicativo registrou um processo mais aleatório de acesso ao Mail-pot.

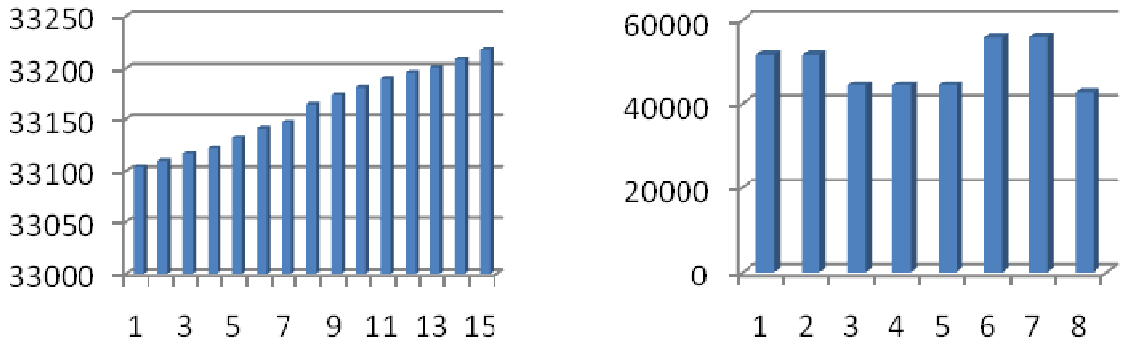

Figura 7. (a) Intervalo que possivelmente apenas um *spammers* usou o aplicativo. (b) Intervalo que houve um revezamento entre os *spammers*.

A Figura 8 apresenta um gráfico onde se buscou relacionar o acesso ao Mail-pot pela unidade de milhão da porta do *spammer*. Nele identificamos que mais da metade dos acessos foram nas portas 4XXXX e quase 40% nas portas 5XXXX. Não se constatou nenhuma tentativa de acesso de portas conhecidas, portas entre 1 e 1024. Todo o restante das portas utilizadas pelos *spammers* ficou nas portas 3XXXX e 6XXXX.

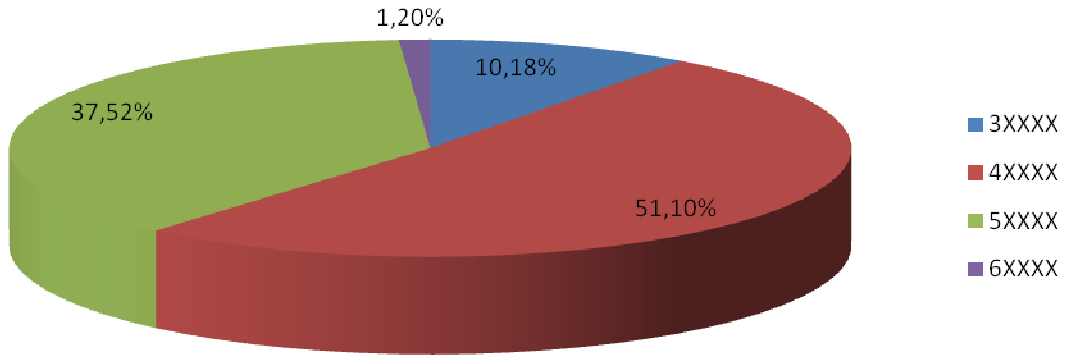

Figura 8. Incidências na unidade de milhão das portas de acesso.

A Tabela 4 mostra a quantidade por porta de acesso, seguindo a ordem de abordagem. Essas informações não são consideradas perigosas pelos *spammers*

para identificá-los, logo nenhuma porta foi mascarada nas inserções das tabelas no banco de dados. A geração dessa tabela não precisou de um aplicativo auxiliar para varrer o conteúdo do spam para encontrar essas informações.

Tabela 4. Quantidade de porta por unidade de milhão

Portas	**Quantidade**
3XXXX	51
4XXXX	256
5XXXX	188
6XXXX	6
Total	501

A Figura 9 apresenta o levantamento do conteúdo disseminado nos *spams.* O processo para quantificar cada ocorrência foi completamente manual, evitando uma predisposição da análise dos resultados.

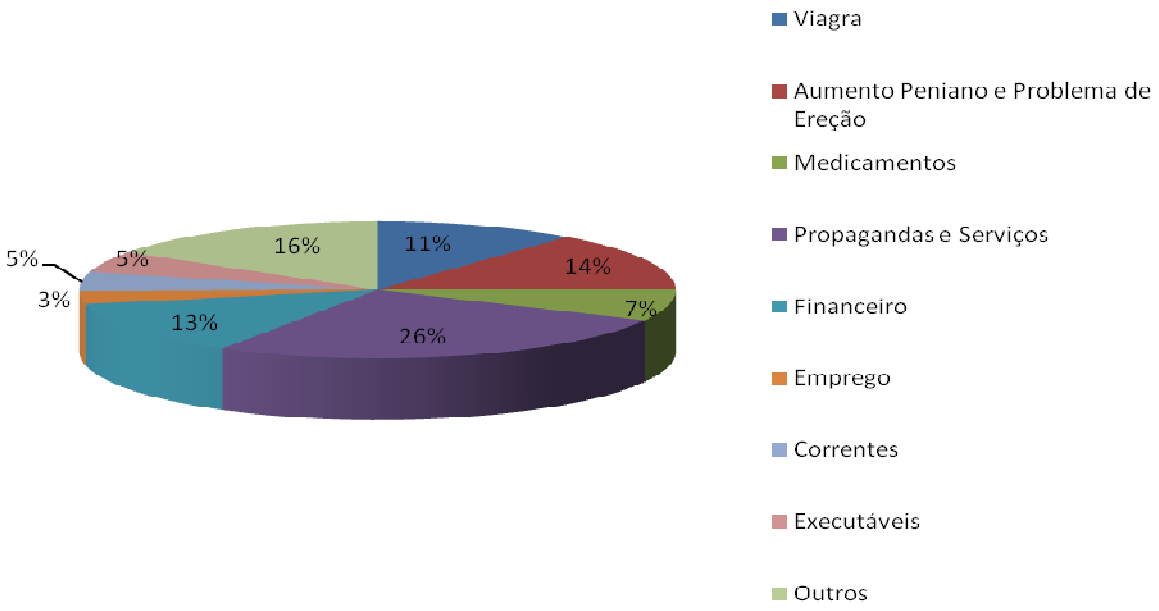

Figura 9. Conteúdo disseminado nos *spams* estudados

Foi verificado que os *spammers* tentam burlar filtros de *spams* e *blacklists*, modificando as formas de escrita do seu produto. Fato observado principalmente no

medicamento Viagra, muitas vezes escrito como V!agra, \/iagra, V1agr@ e variâncias. Apesar de ser um medicamento, ele se encontra separado na análise, pela quantidade exorbitante de ocorrências e por conta das particularidades que os *spammers* adotaram para divulgar esse produto.

Os resultados mostraram uma predominância de assuntos relacionados ao desempenho sexual, mas conforme outros estudos como o do *Spam Filter Review* [37], onde aponta que 25% dos *spams* disseminados estão relacionados a produtos e 20% relacionado a Finanças, o Mail-pot apresentou a mesma tendência, tendo 26% de propaganda de produtos ou serviços de empresas e 13% relacionado ao Financeiro.

A Tabela 5 apresenta a quantidade de *spams* por conteúdo. A Tabela reflete o resultado do gráfico acima, quantizando cada tipo de conteúdo, num total de 501 mensagens estudadas.

Tabela 5. Quantidade apresenta de cada conteúdo

Conteúdo	**Quantidade**
Viagra	55
Aumento peniano e problema de ereção	70
Medicamentos	37
Propagandas e Serviços	132
Financeiro	65
Emprego	15
Correntes	23
Executáveis	25
Outros	79
Total	501

Foi observada também uma quantidade expressiva de mensagens com conteúdo malicioso, contendo executáveis ou direcionamentos para sites de mesmo propósito. Conteúdos em outros idiomas foram descartados da análise, a ocorrência “outros” teve sua coluna incrementada nesse caso.

6 Conclusão

A preocupação com a segurança da informação no âmbito organizacional e pessoal tem motivado várias pesquisas nessa área. A grande dificuldade de fazer políticas e implantações de segurança da informação desencadeou o desenvolvimento desse projeto, que tem como intuito obter informações de como os *spammers* agem, sua localização e o que costumam disseminar em suas mensagens.

Como 40% de todos os *e-mails* enviados no mundo são *spams* [37], a ferramenta proposta vem para ajudar a evitar uma grande transmissão e armazenamento de volume de dados desnecessários. Essa carga excessiva de *spams* em servidores de correio pode causar diversos problemas, como:

- Não recebimento de *e-mails*, já que boa parte dos provedores de Internet limita o tamanho da caixa postal do usuário;
- Gasto desnecessário de tempo, pois cada usuário demanda tempo para identificar o *e-mail* como *spam* e removê-lo;
- Perda de produtividade, principalmente para quem utiliza o *e-mail* como ferramenta de trabalho;
- Além de prejuízos financeiros causados por fraude, onde o *spam* tem sido amplamente utilizado como veículo para disseminar esquemas fraudulentos, que tentam induzir o usuário a acessar páginas clonadas de instituições financeiras ou a instalar programas maliciosos projetados para furtar dados pessoais e financeiros.

O Mail-pot é uma ferramenta para aprender. Ela foi criada parar coletar informações e fazer a análise delas. O aplicativo consiste em um servidor SMTP emulado que não repassa o tráfego que recebe. Ao invés disso ele armazena todo o conteúdo das mensagens em banco de dados. Posteriormente faz-se uma análise do conteúdo.

Como o Mail-pot foi criado com intuito de ser comprometido, de modo que todo tráfego que entra é suspeito por natureza. Isso torna muito mais fácil a coleta e

sua posterior análise dos dados. Essa análise serve como um guia para auxiliar na confecção de políticas e implantações de segurança.

O aplicativo proposto teve o intuito de ficar com a porta 25/TCP aberta e capturar todas as mensagens que passam pela única interface de rede. Após iniciada a conexão, o *spammer* tem todas as confirmações que um servidor SMTP real faria, porém é apenas uma impressão já que a mensagem não é repassada, e sim armazenada em banco de dados. Porém, problemas não previstos ocorreram como mensagens sem *e-mail* de destinatário ou remetente, e existiu a necessidade de gerar novos aplicativos para varrer o conteúdo de todas as mensagens e capturar os dados antes mascarados pelos *spammers.* Informações como *e-mails* de origem e destino, IPs de origem não foram detectados inicialmente, porém o aplicativo auxiliar obteve essas informações e inseriu-as corretamente no banco de dados.

Apesar do pequeno número de *spams* coletados e o pouco tempo para análise, o projeto aqui apresentado, seguiu todas as funcionalidades propostas. Os resultados esperados foram constatados, mas algumas fontes, localidade dos *spammers*, não foram identificadas. Os resultados mostraram uma predominância de países asiáticos como disseminadores de *spams,* assim como o Brasil. A terminação (*top-level*) mais afetado foi o ".com.br" e foi encontrado diferentes comportamentos em relação às portas que originaram as conexões, porém não foi possível comprovar a existência de apenas uma fonte ou mais utilizando o aplicativo ao mesmo tempo. Contudo, é possível concluir que o Mail-pot proposto é um aplicativo que provê informações que podem auxiliar no desenvolvimento de políticas e implantações de segurança de servidores SMTP, e que é tão eficiente quanto o Spampot proposto pelo CERT.br para identificar fontes emissoras de *spams*.

6.1 Trabalhos Futuros

Como trabalho futuro, pode ser feito o aperfeiçoamento do emulador SMTP, Mail-pot, repassando mensagens de confirmações e integração com os outros aplicativos desenvolvidos para o estudo dos dados armazenados, automatizando toda varredura do conteúdo da mensagem nos casos das tentativas de camuflar os dados pelos *spammers*.

Desenvolvimento de uma ferramenta de análise dos dados para a automatização da geração dos gráficos e relatórios, facilitando o uso do aplicativo. Aumentar a quantidade de gráficos e relatórios, mais detalhados para melhor guiar o desenvolvimento das políticas e implantações de segurança pelos responsáveis pela administração dos servidores de correio.

A instalação do Mail-pot em diferentes localizações para identificar tendências com mais facilidade e rapidez. Podendo correlacionar os dados de diversas redes, confirmando, assim, tendências comuns.

Bibliografia

[1] **DARPA – *Defense Advanced Research Projects Agency***; disponível em: <http://www.darpa.mil/> Acessado em Novembro de 2008.

[2] Merkle, E. R., & Richardson, R. A. (2000). **Digital dating and virtual relating: Conceptualizing computer mediated romantic relationships**. Family Relations: Interdisciplinary Journal of Applied Family Studies, 49 (2), 187-192.

[3] Internet Systems Consortium. **Internet Domain Survey Host Count**. Disponível em: < http://www.isc.org/ops/ds/ >. Acesso em: 09 de set. 2008.

[4] The Honeynet Project. **Conheça o seu inimigo. O Projeto Honeynet. Revelando, as ferramentas de segurança, táticas e motivos da comunidade hacker**. 1° Edição. São Paulo: Pearson Education do Brasil Ltda., 2002.

[5] ***Honeynet Project***; disponível em: <http://www.honeynet.org> Acessado em Agosto de 2008.

[6] **Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil**; diponível em: <http://www.cert.br/> Acessado em Julho de 2008.

[7] CERT.br, **Resultados Preliminares do Projeto SpamPots: Uso de *Honeypots* de Baixa Interatividade na Obtenção de Métricas sobre o Abuso de Redes de Banda Larga para o Envio de *Spam*;** disponível em: <http://www.cert.br/docs/whitepapers/spampots/> Acessado em Agosto de 2008.

[8] KLESIN, J. **RFC 2821: Simple Mail Transfer Protocol**. <http://www.ietf.org/rfc/rfc2821.txt> Acessado em Setembro de 2008.

[9] **MySQL**; disponível em: <http://www.mysql.com/> Acessado em Agosto de 2008.

[10] **Java Runtime Environment (JRE)**; disponível em <http://java.sun.com/j2se/desktopjava/jre/> Acessado em Agosto de 2008.

[11] SÊMOLA, M. **Gestão da Segurança da Informação: uma visão executiva.** 1ª Edição. Rio de Janeiro: Ed. Campus, 2003.

[12] Módulo Security Solutions S.A. **Gestão de Riscos em TI utilizando a ferramenta Módulo Risk Manager.** Rio de Janeiro: jul. 2007. 189 slides. Acompanha texto.

[13] FERREIRA, F. N. F.; ARAÚJO, M. T. **Política de Segurança da Informação (guia prático para elaboração e implementação)**. Rio de Janeiro: Ed. Ciência Moderna, 2006.

[14] STOLL. CLIFF. ***The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer* Espionage**, Nova York: Pocket Books, 1990.

[15] CHESWICH, BILL. ***An Evening with Berferd In Which a Cracker is Lured, Endured, and Studied****, disponível em:* <http://www.securityfocus.com/data/library/berferd.ps> Acessado em Agosto de 2008.

[16] COHEN, F. **The Deception Toolkit**, disponível em: <http://www.all.net/dtk> Acessado em Agosto de 2008.

[17] SPITZNER, L. ***Honeypots: Tracking Hackers***. Addison Wesleyu, 2002.

[18] **Honeypots Tracking Hackers**. USA: Addison Wesley, 2003.

[19] **IPTables**; disponível em: <http://www.netfilter.org> Acessado em Agosto de 2008.

[20] **Syslog**; disponível em: <http://www.ietf.org/html.charters/syslog-charter.html> Acessado em Agosto de 2008.

[21] **Snort**; disponível em: <http://www.snort.org> Acessado em agosto de 2008.

[22] **Tcpdump**; disponível em: <http://www.tcpdump.org> Acessado em agosto de 2008.

[23] **Snort-Inline**; disponível em: <http://snort-inline.sf.net> Acessado em agosto de 2008.

[24] NIELS PROVOS, N. ***Developments of the Honeyd Virtual Honeypot****, disponível em: <http://www.honeyd.org>* Acessado em agosto de 2008.

[25] http://www.honeynet.org/papers/honeynet/index.html Acessado em Agosto de 2008.

[26] NIELS PROVOS, T. H. **Virtual Honeypots: From Botnet Tracking to Intrusion Detection**. Addison-Wesley Professional, 2007.

[27] **Projeto *HoneynetBR***; disponível em: <http://www.honeynet.org.br> Acessado em Agosto de 2008.

[28] **INPE**; disponível em: <http://www.inpe.br> Acessado em Agosto de 2008.

[29] **NBSO – NIC BR Security Office**. http://www.nbso.nic.br/stats/incidentes/.

[30] **NBSO - NIC BR Security Office**. http://www.nbso.nic.br/stats/spam/.

[31] Andrucioli, Alexandre Pinaffi. **Proposta e avaliação de um modelo alternativo baseado em honeynet para identificação de ataques e classificação de atacantes na internet**. Tese de Mestrado, COPPE/UFRJ, abril de 2005. http://www.ravel.ufrj.br/arquivosPublicacoes/tese_pinaffi.pdf.

[32] **Microsoft**; disponível em: <http://www.microsoft.com> Acessado em Agosto de 2008.

[33] **Mistério da Ciência e Tecnologia**; disponível em: <http://www.mct.gov.br/> Acessado em Agosto de 2008.

[34] **SPAM - Unsolicited Commercial E-Mail**; Disponível em: <http://epic.org/privacy/junk_mail/spam/> Acessado em Agosto de 2008

[35] **Antispam.BR**; disponível em: <http://www.antispam.br/conceito/> Acessado em Julho de 2008

[36] **Projeto e Desenvolvimento de um Sistema de Controle e Acompanhamento de Notificações de Spam, Apresentado no V Simpósio Segurança em Informática** (SSI'2003), (São José dos Campos, SP), Novembro, 2003. http://www.cert.br/docs/papers/spamctl-ssi2003.pdf

[37] Evett, Don. *Spam Filter Review,* **Spam Statistic 2006**, disponível em: <http://www.spam-filter-review.toptenreviews.com/spam-statistics.html> Acessado em Outubro de 2008.