Identificação e análise de tráfego malicioso através do uso de honeypots

Vinícius Batistela[1]
Marco Antônio Sandini Trentin[2]

Resumo: O conceito de honeypot representa uma nova abordagem na área de segurança de redes de computadores, que envolve, por exemplo, a utilização de computadores atuando como sensores, a fim de identificar atividades maliciosas na rede. Assim, o presente trabalho utilizou esta nova abordagem e, com a utilização de um honeypot atuando como sensor na internet, buscou identificar a presença de atividades que possam ser consideradas maliciosas. A partir disso, o tráfego malicioso detectado pelo honeypot foi analisado e, após, relatos a seu respeito foram feitos, bem como apresentadas informações, tais como sobre a quais serviços disponíveis da rede se refere, quais foram suas origens e, no caso da identificação da presença de algum malware, como pode ter entrado.

Palavras-chave: *Honeypot. Honeynet. Segurança de Redes.*

Abstract: *The honeypot concept represents a new approach on computer network security area, that involve for example the use of computers working as sensors, to identify malicious activities on the network. So, the present paper used this new approach and, using a honeypot working as a sensor on the Internet, tried to identify the presence of activities that could be consider malicious. From that, the detected malicious traffic was analyzed and then reports was done showing information like the network service involves with the malicious traffic, from where was it from and, in the case of the identification of the presence of some malware, how did it come in.*

Keywords: *Honeypot. Honeynet. Network Security.*

1 Introdução

A abordagem tradicional de segurança, puramente defensiva, com a utilização de firewalls, sistemas de detecção de intrusão, criptografia e outros mecanismos, não é suficiente para garantir a segurança das redes de computadores atualmente. O problema desta abordagem é que é o atacante quem toma a iniciativa, estando sempre um passo à frente das tecnologias de segurança existentes. Honeypots e, posteriormente, honeynets, surgiram com o objetivo de mudar esta abordagem, permitindo a descoberta de novas ferramentas de intrusão e o estudo da metodologias de ataque e das motivações dos atacantes, com o intuito de contribuir para o desenvolvimento de boas práticas de segurança.

Assim, o presente trabalho buscou utilizar esta nova abordagem de segurança e, com a utilização de honeypots atuando como sensores no tráfego na internet, identificar a presença de atividades que possam ser consideradas maliciosas. A partir disso, essas atividades foram relatadas, juntamente com informações a seu

[1]Graduado do curso de Ciência da Computação da UPF
{vinicius@batistela.org}
[2]Curso de Ciência da Computação, UPF, Campus 1 – BR 285 – Passo Fundo (RS) – Brasil
{trentin@upf.br}

doi: 10.5335/rbca.2009.002

respeito, como a quais serviços disponíveis da rede se referem, quais foram suas origens e, no caso da identificação da presença de algum malware, qual vulnerabilidade foi explorada.

2 Honeypots

Em agosto de 1986, os administradores da rede de computadores do Lawrence Berkley Laboratory, um famoso centro de pesquisas dos Estados Unidos, localizado em São Francisco, na Califórnia, perceberam que alguém estava atacando a rede e obtendo sucesso nas tentativas de invasão. Entretanto, ao invés de tomar providências para interromper os ataques e manter o invasor longe dos sistemas, os administradores resolveram deixá-lo agir e monitorar as suas ações. Esse incidente, levado a público por Clifford Stoll [10], é tido como o primeiro relato de implantação de mecanismos para acompanhar as atividades de um invasor.

As máquinas que foram comprometidas no incidente relatado por Clifford Stoll não eram especialmente preparadas para este fim; eram máquinas utilizadas no Lawrence Berkley Laboratory, contendo arquivos e serviços reais. Em 1991,Bill Cheswick [4], ao perceber uma tentativa de intrusão a um computador da AT&T Bell Laboratories, resolveu seguir o mesmo caminho de Clifford Stoll, com uma diferença: preparou uma máquina especialmente para ser invadida, desenvolvendo um ambiente para enjaular o invasor e restringir as suas ações.

Em 1998, uma ferramenta desenvolvida por Fred Cohen [5], denominada The Deception Toolkit, foi distribuída gratuitamente na Internet. Tal ferramenta permite simular, no sistema em que for instalada, a existência de um grande número de softwares com vulnerabilidades conhecidas; o seu funcionamento consiste em gerar respostas específicas, diante de tentativas de intrusão dos atacantes, fazendo-os acreditar que estão explorando as vulnerabilidades e obtendo acesso ao sistema. Desse modo, surgia a primeira ferramenta cujo objetivo era colher informações a respeito dos métodos utilizados para invadir sistemas.

Apesar de os conceitos envolvendo honeypots terem sido introduzidos por Clifford Stoll e Bill Cheswick, apenas em 2002 surgiu uma definição clara a respeito do que é um honeypot. Naquele ano, o termo honeypot foi definido como sendo um recurso de segurança cujo valor está na sua sondagem, ataque ou comprometimento [8]. O fato de essa definição ser tão genérica deve-se a que não são apenas computadores que podem ser utilizados como honeypots, mas qualquer equipamento capaz de ser preparado para ser comprometido, como, por exemplo, modens e roteadores. O valor de um honeypot está diretamente ligado ao fato de que tal recurso não será utilizado por usuários ou para prover serviços para outros sistemas, ou seja, nenhuma atividade envolvendo o honeypot será esperada, de modo que qualquer atividade que existir pode ser considerada suspeita [8].

Honeypots podem ser usados para dois propósitos: pesquisa ou produção. Honeypots de pesquisa são utilizados para capturar informações. A partir disto, novos malwares, novas ferramentas e novas táticas utilizadas pelos atacantes podem ser identificadas rapidamente, possibilitando a divulgação de alertas e, também, dar aos administradores mais segurança aos seus sistemas. Honeypots de produção tem como propósito fornecer segurança às organizações, podendo agir de três formas:

- prevenção: honeypots podem prevenir ataques, detectando scans na rede e interagindo com os mesmos, fazendo com que o atacante receba respostas mais lentamente e, até mesmo, interrompendo o ataque. Desse modo, o atacante terá dificuldades em detectar vulnerabilidades no sistema. Honeypots também realizam a prevenção, intimidando e confundindo o atacante. Se o atacante sabe que a organização utiliza esse recurso de segurança, mas não sabe quais sistemas são honeypots e quais são legítimos, pode desistir de realizar o ataque. Por outro lado, se ele não souber da existência dos honeypots, poderá perder tempo interagindo com eles, deixando os sistemas legítimos livres dos ataques;
- detecção: quanto mais rapidamente uma organização detectar um ataque, mais rapidamente poderá responder a este, interrompendo-o ou minimizando os possíveis danos. Honeypots são ferramentas poderosas para este propósito, pelo fato de qualquer atividade que os envolva ser considerada suspeita;

- resposta: para responder a um invasor, uma organização necessita ter um profundo conhecimento a respeito do que ele fez, de como invadiu e das ferramentas que utilizou. Obter essas informações a partir de um sistema de produção, como um servidor de e-mail, é muito difícil em razão da quantidade de atividade envolvida (usuários efetuando login no servidor, e-mails sendo enviados, dentre outros). Além de ser difícil separar as informações referentes a atividades legítimas das referentes ao ataque, um sistema de produção não pode ser desligado, de modo que a análise tem de ser feita com o sistema funcionando e gerando ainda mais informações. Honeypots capturam somente atividade maliciosa e podem ser desligados a qualquer momento, tornando-se muito mais fácil extrair informações a partir deles [8].

Embora o conceito envolvendo honeypots seja simples, é esta simplicidade que proporciona as vantagens e desvantagens características destes recursos. Como vantagens apresentam-se as seguintes:

- captura de uma pequena quantidade de dados: analisar os logs de um sistema de produção à procura de algum tipo de atividade maliciosa é um trabalho árduo e demorado, em razão da grande quantidade de logs existentes. Desse modo, como toda atividade relacionada a um honeypot pode ser considerada suspeita, embora menos dados sejam coletados, estes dados são muito mais valiosos, pois são todos relacionados a tentativas de ataque ou intrusão ao sistema. Assim, analisar os dados e retirar informações torna-se muito mais fácil;
- captura de novas ferramentas e táticas: honeypots conseguem identificar novas ferramentas e táticas utilizadas pelos atacantes;
- necessidade de poucos recursos: um honeypot apenas registra as atividades maliciosas, de modo que um velho Pentium com 128MB de RAM e processador de 100MHz pode ser utilizado tranquilamente como um honeypot;
- captura de tráfego encriptado: enquanto ferramentas de IDS podem não detectar atividades envolvendo dados criptografados, em um honeypot, toda e qualquer atividade poderá ser capturada, armazenada e ser analisada posteriormente;
- informação: honeypots podem coletar informações que poucas, ou nenhuma outra técnica é capaz de coletar, como, por exemplo, novas ferramentas que estejam sendo utilizadas por atacantes;
- simplicidade: a utilização de honeypots é relativamente simples, pois basta configurar um ambiente e monitorá-lo, não envolvendo o desenvolvimento de algorítimos complexos, por exemplo [8].

Como toda tecnologia, os honeypots também possuem alguns pontos fracos. Podemos citar as seguintes desvantagens relacionadas aos honeypots:

- visão limitada: somente podem capturar atividades diretamente relacionadas a ele; não capturam ataques contra outros sistemas, a não ser que ocorra interação com o honeypot;
- riscos: todo recurso de segurança possui riscos. Um firewall corre o risco de ser burlado, uma criptografia corre o risco de ser quebrada e um IDS corre o risco de falhar em detectar um ataque. Honeypots não são diferentes. O maior risco que se corre ao utilizar um honeypot é que seja usado pelo invasor para lançar ataques a outros sistemas [8].

Para ajudar a compreender o funcionamento dos variados tipos, honeypots são classificados de acordo com o nível de interatividade que proporcionam ao atacante, ou seja, de acordo com o que o atacante será capaz de fazer no honeypot. Com base no nível de interatividade define-se a quantidade de informação que poderá ser capturada, a dificuldade que se terá em instalar, configurar e manter o honeypot e os riscos que proporcionará ao ambiente em que estiver instalado. Utilizando esta abordagem, existem três classificações:

- baixa interatividade: em termos de instalação, configuração e manutenção, honeypots de baixa interatividade são os mais fáceis de implementar. Esse tipo de honeypot é capaz de simular serviços básicos, como Telnet e FTP, limitando o atacante a interagir apenas com esses serviços pré-configurados. Pelo fato de não permitir a entrada do atacante, honeypots de baixa interatividade apresentam a vantagem de não correrem o risco de serem utilizados como base para outros ataques, sendo os mais indicados para pessoas ou organizações que querem começar a trabalhar com este tipo de recurso de segurança;
- média interatividade: este tipo de honeypot permite uma maior interação por parte do atacante, mas ainda não permite uma interação com um sistema operacional real. Um exemplo de honeypot de

média interatividade é a criação de um ambiente protegido utilizando as funcionalidades providas pelo chroot, presente nos sistemas operacionais baseados em Unix. Desse modo, o administrador pode criar um sistema operacional virtual dentro do sistema operacional real. O objetivo da criação de um ambiente protegido é fazer com que o atacante, ao obter acesso, caia nesse ambiente, que pode ser facilmente monitorado e analisado;

- alta interatividade: honeypots de alta interatividade oferecem aos atacantes sistemas operacionais e aplicações reais para serem atacados. Os serviços não são simulados, não são utilizados ambientes especiais e nada é restringido. Este tipo de honeypot é o mais complexo de ser implementado, é o que oferece mais riscos, mas é o que possibilita uma maior coleta de informações [8].

Quando vários honeypots são colocados em uma mesma rede, todos atrás de um gateway responsável por controlar o fluxo de dados que entram e saem dos mesmos, tem-se o que é chamado de honeynet. Honeynets são essencialmente um conjunto de honeypots de pesquisa, ou seja, é uma rede projetada especialmente para ser comprometida. Uma vez comprometida, é utilizada para observar o comportamento dos invasores, possibilitando a descoberta de novas ferramentas e novas vulnerabilidades que estejam sendo exploradas [6].

Diante disso, esse projeto de pesquisa objetivou conhecer com um maior número de detalhes a implantação e o funcionamento de um honeypot, analisando os resultados obtidos a fim de estudar a presença de malwares na internet. Organizações como o CERT.br (Centro de Estudos, Resposta e Tratamentos de Incidentes de Segurança do Brasil), grupo de resposta a incidentes de segurança para a internet brasileira, mantido pelo Comitê Gestor da Internet no Brasil (CGI.br), também utilizam honeypots para analisar atividades maliciosas na internet, como pode ser visto em [3]. Outro projeto que se utiliza de honeypots para capturar informações é o Consórcio Brasileiro de Honeypots[3], que se trata de uma rede de honeypots distribuídos por várias organizações, como universidades, empresas de telecomunicações e órgãos do governo federal.

3 Estrutura utilizada

O curso de Ciência da Computação da UPF – Universidade de Passo Fundo possui, desde 2006, um Grupo de Pesquisa em Segurança de Redes, o GSEG, visando à utilização de honeypots para realizar diversos estudos e pesquisas relacionados à segurança. A estrutura montada por esse grupo, e que foi utilizada na realização do estudo envolvendo o honeypot atuando na internet, consiste em uma honeynet formada por um gateway e cinco máquinas atuando como honeypots, como pode ser observado na figura abaixo. Além disso, há uma máquina responsável por armazenar os logs dos honeypots e uma outra rodando um NMS (Network Management System), responsável por monitorar as demais.

[3] http://www.honeypots-alliance.org.br

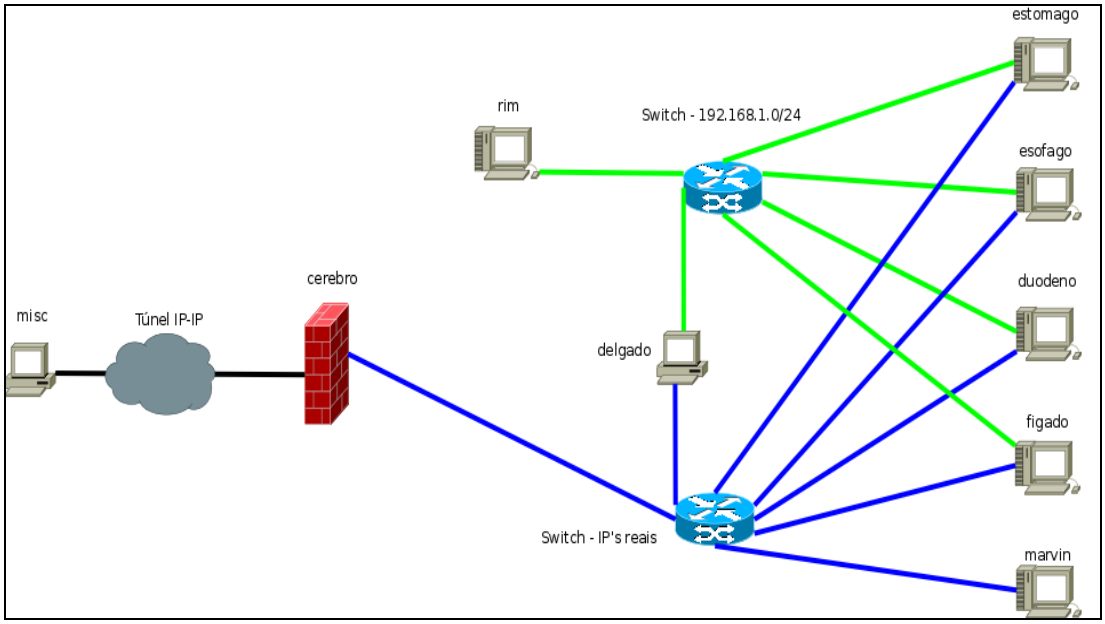


Figura 1. Diagrama da estrutura da honeynet do GSEG

Para a construção da rede do GSEG e para possibilitar as suas pesquisas, sem que houvesse qualquer restrição de acesso à internet, a UPF cedeu 13 endereços IP reais para serem utilizados. Além disso, todo o tráfego destinado a esses IP's é roteado de forma a ser entregue ao gateway da honeynet que se encarrega de entregá-lo aos honeypots.

Dos cinco honeypots pertencentes à rede do GSEG, apenas um foi utilizado para coletar os dados pretendidos para essa pesquisa, o qual foi configurado e colocado em funcionamento no dia 05/09/08, às 21 horas. A configuração utilizada é extremamente simples e pode ser facilmente reproduzida. Como sistema operacional, foi escolhido o Linux Ubuntu 8.04 na sua versão para servidor, pelo fato de ser um sistema estável e de a UPF possuir repositórios internos para o mesmo, facilitando a instalação e atualização. A primeira configuração feita nos honeypots foi a criação de uma única regra para o Iptables, com o objetivo de armazenar em logs todo o tráfego com destino ao honeypot. Com isso, pretendeu-se coletar informações tais como as portas mais acessadas e a origem das tentativas de acesso.

O objetivo maior deste trabalho consistiu em analisar a presença de malwares na internet. Para atingir esse objetivo, procurou-se buscar softwares que atuassem como honeypots e que se mostrassem adequados para a realização desta tarefa. Assim, dois honeypots pesquisados mostraram-se os mais completos, no sentido de suas funcionalidades, e também os mais utilizados por renomados órgãos de pesquisa, tais como o CERT.br: o Nepenthes[4] e o Honeyd[5].

Tanto o Honeyd quanto o Nepenthes são honeypots de baixa interatividade, ou seja, apenas emulam serviços. O Honeyd possui fácil configuração, permitindo a criação de redes virtuais, com rotas virtuais entre essas redes e a simulação de *links* congestionados, manipulando o tempo das respostas efetuadas pelo honeypot. Permite também que para cada porta emulada seja associado um script que simule a troca de mensagens realizada pelos protocolos, tais como SMTP e SSH, de modo a tentar iludir o invasor de que ele está interagindo em um sistema real. Outra funcionalidade interessante do Honeyd é a capacidade de identificar de modo passivo o sistema operacional utilizado pelo atacante.

Dentre todas as funcionalidades do Honeyd apresentadas até aqui, o Nepenthes possui apenas a de emular a troca de mensagens entre os protocolos. Entretanto, o Nepenthes possui uma característica única, que o fez ser escolhido o mais adequado para a utilização neste trabalho, que é a capacidade de realizar o download dos

[4] http://nepenthes.mwcollect.org
[5] http://www.honeyd.org

binários dos malwares que chegam até ele. O modo de funcionamento do Nepenthes será mais detalhado na seção a seguir.

3.1 Nepenthes

O Nepenthes é um honeypot de baixa interatividade que foi desenvolvido com o objetivo específico de capturar malwares. Atua emulando serviços do sistema operacional Windows com o objetivo de enganar ataques automatizados; está organizado em módulos, cada um com uma função específica [1].

Os módulos de vulnerabilidade são responsáveis por emular serviços vulneráveis e fazer os malwares acreditarem que podem explorar esses serviços. A partir disso se consegue capturar o payload do ataque, o qual é passado então ao próximo módulo, chamado shellcode parsing module, responsável por analisar este payload em busca da URL onde será possível fazer o download do malware. Caso esta URL seja encontrada, os módulos chamados fetch modules encarregam-se de fazer o download do malware. Existem sete fetch modules distintos, que possuem a capacidade de realizar downloads via TFTP, HTTP, FTP e csend/creceive (uma implementação do protocolo IRC), assim como outros protocolos próprios utilizados pelos malwares. Por fim, após realizar o download do malware, entram em ação os submissions modules, responsáveis por dar um destino ao que foi obtido. O Nepenthes suporta basicamente três tipos de tratamento para os malwares coletados: armazená-los localmente ou enviá-los para uma base de dados centralizada, ou, ainda, enviá-los diretamente ao Norman SandBox Information Center [1].

O Norman SandBox Information Center é um site mantido por uma empresa fabricante de sistemas antivírus, no qual é possível submeter um arquivo considerado suspeito de estar infectado. Este arquivo é então analisado e informações, tais como o nome do malware encontrado, os arquivos que ele cria e apaga ao se instalar no sistema e alterações que realiza no registro do Windows são retornadas por e-mail. Entretanto, neste trabalho optou-se por armazenar os malwares localmente e analisá-los posteriormente através do site http://www.virustotal.com, que analisa o arquivo utilizando diferentes sistemas antivírus.

No sistema operacional Ubuntu, a instalação do Nepenthes pode ser feita utilizando-se do sistema de gerenciamento de pacotes provido pelo mesmo, por meio do comando apt-get install nepenthes. O Nepenthes foi utilizado na sua configuração padrão, que consiste em todos os módulos de vulnerabilidade ativos, emulando serviços em vinte portas[6] TCP diferentes, realizando o armazenamento dos malwares capturados localmente no diretório /var/lib/nepenthes/binaries, sendo estes os arquivos que serão posteriormente analisados. Os logs referentes aos downloads, contendo data, hora, endereço IP e método de download, assim como o MD5 do arquivo obtido, encontram-se no arquivo /var/log/nepenthes/logged_submissions.

O Nepenthes é um sistema que possibilita escalabilidade, no sentido de poder emular mais de um host, respondendo por vários endereços IP. O Nepenthes é capaz de responder em uma única máquina por um número aproximado de dezesseis mil endereços IP [1]. Essa máquina permaneceu ativa e estável por mais de cinco meses. Na realização do presente trabalho, entretanto, o Nepenthes atuou respondendo por um endereço IP apenas.

4 Análise dos logs

Analisar um arquivo de log contendo milhares de linhas sem o auxílio de alguma ferramenta é algo trabalhoso e pouco produtivo. Por exemplo, o arquivo de logs do Iptables obtido pelo honeypot colocado na internet, por um período de dez dias teve mais de 110 mil linhas, cada uma correspondendo a um pacote que teve como destino o honeypot. Por esse motivo, foram buscadas ferramentas automatizadas para a realização desta tarefa, principalmente para se analisar os logs do Iptables, pois o arquivo de log gerado pelo Nepenthes pode ser facilmente interpretado e analisado manualmente. Na figura abaixo pode-se observar um exemplo de arquivo de log gerado pelo Nepenthes, contendo informações a respeito dos malwares capturados.

[6] 21, 42, 80, 110, 135, 139, 143, 220, 443, 445, 465, 993, 995, 1023, 1025, 2105, 3372, 5000, 10000 e 17300.

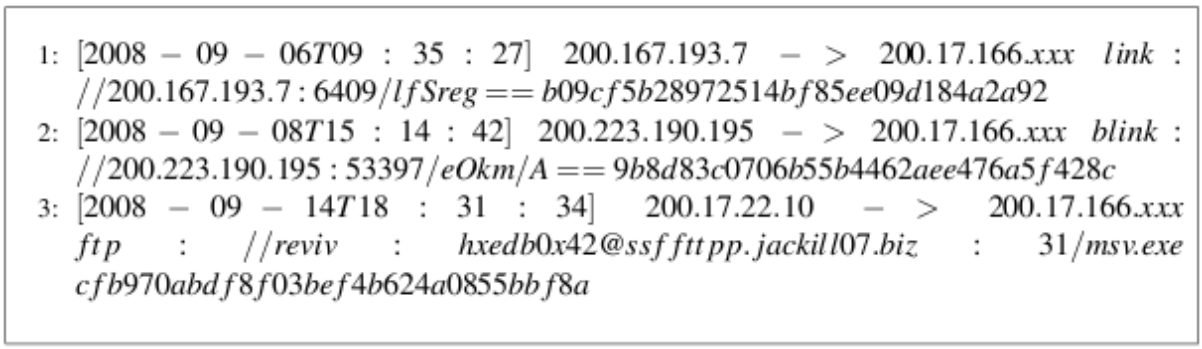


Figura 2. Exemplo de arquivo de log gerado pelo Nepenthes

As informações presentes no *log* apresentado na Figura 2 são, respectivamente, a data e a hora de download do malware, o IP de origem e o IP de destino da conexão (200.17.166.xxx, o honeypot), o link pelo qual foi feito o download e o hash MD5 do binário obtido. O IP utilizado no honeypot foi camuflado a fim de que seu conhecimento público não atrapalhe futuras pesquisas.

Na Figura 2 pode-se observar a presença de links começando com "link://" e "blink://". Estas linhas tratam de tipos de transferência de dados próprios de alguns tipos de malwares. O funcionamento desses dois tipos de transferência se dá da seguinte maneira: primeiramente, o malware envia um shellcode[7] que armazena um cookie[8], com um tamanho de 4 bytes e codificado em base64 na máquina alvo, além de deixar uma porta aberta. Posteriormente, uma segunda conexão é estabelecida e um segundo shellcode é enviado à máquina alvo. Este segundo shellcode é responsável por enviar o cookie para a máquina atacante e fazer o download do binário do malware para a máquina alvo. O cookie funciona, então, como uma espécie de chave, sem o qual o binário do malware não pode ser obtido. Dessa forma, o significado deste tipo de link é: link://ip_do_host_hospedeiro:porta_do_host_hospedeiro/base64(cookie). A diferença entre os links "link://" e "blink://" ocorre no segundo estágio da conexão; no primeiro, é o atacante que se conecta à vítima para requisitar o cookie e, no segundo, é a vítima que se conecta ao atacante e o envia.

Para a análise dos logs gerados pelo Iptables foi utilizada a ferramenta wflogs[9], disponível nos repositórios do Ubuntu, podendo ser instalada por meio do comando apt-get. Esta ferramenta analisa logs gerados por diversas aplicações, como Iptables, Ipchains, Ipfilter e Snort, permitindo a conversão dos logs para XML, HTML, um formato de texto puro mais legível que o *log* original, ou, ainda, converter do log de uma aplicação para o de outra.

A grande ajuda proporcionada pelo wflogs são os filtros que oferece. A partir disso é possível gerar um XML contendo apenas as portas TCP ou UDP, por exemplo, e utilizar algum script para ler este XML e gerar algumas estatísticas, como as portas mais acessadas. Esta foi uma das formas utilizadas neste trabalho para a análise dos logs, tendo sido desenvolvido um pequeno script em Python para a leitura do XML.

Além da análise dos logs com a ajuda do wflogs e de um script desenvolvido para ler seus XML gerados, foram desenvolvidos dois outros scripts também em Python, para ler os logs gerados diretamente pelo Iptables. Isso se mostrou necessário pois o wflogs, ao análisar os logs do Iptables contendo os acesso às portas TCP, considera apenas os pacotes com a flag SYN ligada, que caracterizam uma tentativa de início de conexão. Entretanto, em uma análise manual do arquivo de log foi possível verificar que não era apenas este tipo de pacote que chegava ao honeypot. Assim, o script analisa os logs do Iptables, gerando como resultado uma lista de portas de destino distintas e o número de acessos em cada uma. De outra forma, o script extrai uma lista de IPs de origem distintos, assim como o número de pacotes onde cada um aparece e sua localição geográfica, através da API disponibilizada pelo site http://www.hostip.info/use.html.

[7] Pequeno pedaço de código utilizado com payload de um exploit para alguma vulnerabilidade em um software.
[8] Grupo de dados trocados entre um cliente e um servidor, colocado num arquivo de texto criado no computador do cliente.
[9] http://www.wallfire.org/wflogs/

5 Resultados

O honeypot ficou ativo por um período de dez dias, de 05/09/08 a 16/09/08. Então, cabe ressaltar aqui que todas as estatísticas apresentadas nessa seção se referem a esse período. Durante este período, tanto o Nepenthes quanto o Iptables atuaram na coleta de informações, sendo a partir do Iptables registrada a entrada de 119.463 pacotes. Desses, 116.861 (97,82%) foram pacotes TCP; 2.272 (1,9%) pacotes UDP e 330 (0,27%), pacotes ICMP. Como já referido anteriormente, o honeypot entrou em funcionamento às 21h, e a primeira tentativa de acesso data deste mesmo dia foi por volta das 22h, ou seja, menos de uma hora após a máquina ter entrado em funcionamento. Como esse IP nunca havia sido utilizado pela UPF em nenhum servidor, conclui-se que o tráfego dirigido a este só pode ser indesejado, o que será comprovado adiante.

A Figura 3 mostra as portas UDP que mais foram acessadas e o número de acessos a cada uma. Aproximadamente 72% dos acessos foram nas portas 1026 e 1027, as quais são tipicamente utilizadas pelo serviço de mensagens do Windows. Este serviço consiste no envio de mensagens pop-up para usuários a fim de prover alguma notificação, tendo sido criado como uma ferramenta para os administradores de rede enviarem avisos para os usuários. Entretanto, malwares se utilizam desse serviço para o envio de SPAM, o que justifica a procura por essas portas. A partir do Windows XP SP2 este serviço passou a vir desabilitado. Percebe-se, então, que ainda há procura por vulnerabilidades antigas, indício que leva a crer que muitos usuários não têm o costume de atualizar seus sistemas operacionais ou que ainda utilizam sistemas operacionais antigos.

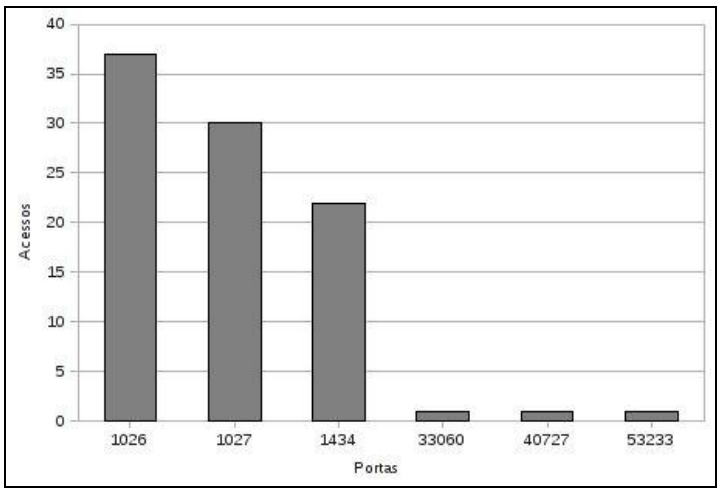


Figura 3. Portas UDP mais acessadas entre 05/09 e 16/09

Observando ainda a Figura 3, pode-se aferir que em terceiro lugar no número total de acessos se encontra a porta UDP 1434, a qual, juntamente com a porta TCP 1434 (uma das portas emuladas pelo *Nepenthes*), é utilizada por padrão pelo serviço de monitoramento oferecido pelo Microsoft SQL Server. Comumente, scans relacionados a esta porta são provenientes de máquinas infectadas com o worm SQL Slammer. Este worm surgiu em janeiro de 2003 e em menos de dez minutos infectou 90% dos hosts vulneráveis, o que ocorreu mesmo já tendo sido lançado em 2002, por parte da Microsoft, um patch para a correção da vulnerabilidade explorada pelo mesmo no SQL Server. O Slammer não possui um payload malicioso, ou seja, não causa danos à máquina infectada; seu objetivo é apenas se propagar, realizando scans em faixas randômicas de endereços IP, a fim de inundar o enlace que esteja utilizando e causar um DoS. Embora seja uma das portas emuladas pelo Nepenthes e tenha sido acessada, não foi capturado nenhum binário proveniente de acesso a esta porta, o que pode ser justificado pelo fato de o Slammer não possuir código para gravar a si mesmo no disco, atuando somente em memória, o que o torna também fácil de ser removido [7].

As demais portas UDP que aparecem na Figura 3 são oriundas de conexões abertas pelo próprio honeypot, utilizadas para a resolução de nomes por meio do DNS e para a obtenção de malwares através do protocolo TFTP. Esses dois protocolos utilizam o UDP como protocolo de transporte. Desta forma, tais portas não representam atividade maliciosa.

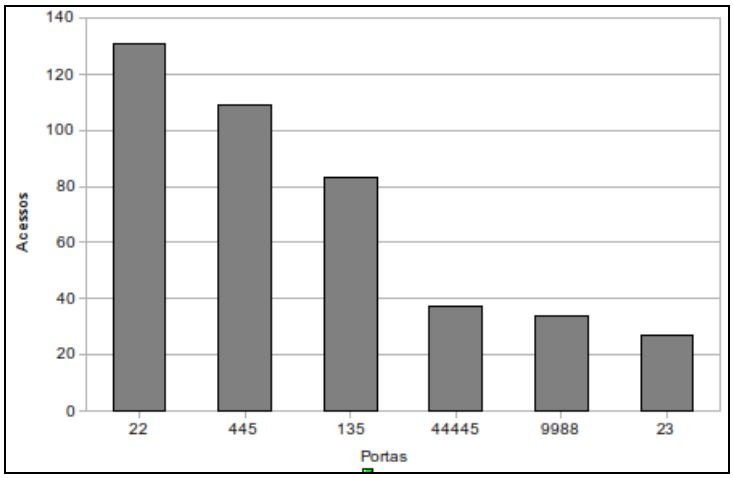


Figura 4. Portas TCP mais acessadas entre 05/09 e 16/09 (Tentativas de inicio de conexão)

Tratando-se do protocolo TCP e considerando apenas as tentativas de início de conexão, as portas 22 e 445 mostraram-se as mais procuradas, como pode ser observado no gráfico da Figura 4. A porta 22 é historicamente muito procurada pelo fato de rodar tipicamente o serviço SSH, onde qualquer tentativa bem sucedida garante acesso direto a uma shell no sistema comprometido. Já a porta 445 é a porta padrão utilizada pelo Windows para o compartilhamento de arquivos. Malwares que tentam acesso a essa porta geralmente estão à procura de compartilhamentos abertos pelos quais eles poderão se propagar e infectar a máquina encontrada.

Em terceiro lugar dentre as portas mais acessadas apresentadas na Figura 4 está a TCP 135. Esta é uma das portas onde o Nepenthes emula vulnerabilidades do Windows, mais especificamente, do serviço RPC (Remote Procedure Call), onde diversos worms tentam acesso para infectar os seus hosts alvos. As portas TCP 44445 e 9988 não são portas conhecidas, ou seja, não são portas que fornecem serviços padrão, tais como FTP ou HTTP. Apesar disso, podem ter sido acessadas por alguma ferramenta ou malware que realize port scan que estivesse buscando identificar portas abertas no honeypot. A porta 9988 é, também, procurada por um worm denominado W32.Rahack[10]. Este malware busca por hosts que contenham versões vulneráveis do software Radmin, que disponibiliza controle e acesso remoto para computadores rodando o sistema operacional Windows. A última porta listada no gráfico (23) representa as tentativas de acesso à porta padrão do serviço Telnet, que, assim como o SSH, oferece uma shell de acesso a máquina, com a diferença de a conexão não ser criptografada.

Ao serem levados em consideração todos os pacotes TCP que chegaram ao honeypot, a porta 21 foi a mais procurada, com um total de 87.831 pacotes a ela destinados, como pode ser observado na Figura 5. A sumarização das portas contando todos os pacotes é importante, pois existem tipos de ataques de reconhecimento, como port scans, que se utilizam de pacotes contendo, por exemplo, a flag *FIN* ativada (não caracterizando início de conexão), como uma técnica de port scan utilizada para burlar sistemas de firewall. Além da porta 21, duas outras portas, que não aparecem no gráfico da Figura 4, estão presentes na Figura 5: a TCP 139 e a 2222. Destas, a primeira é utilizada por sistemas Windows para o compartilhamento de arquivos através do protocolo NetBIOS[11], sendo, assim, como a TCP 445, procurada por malwares que estejam em busca de compartilhamentos abertos por onde possam se propagar. A segunda (2222) foi utilizada para instalar um servidor SSH legítimo, a fim de permitir acesso remoto ao honeypot para manutenção e coleta de informações, não caracterizando então, atividade maliciosa.

[10] http://www.symantec.com/security_response/writeup.jsp?docid=2007-011509-2103-99&tabid=2

[11] Arquitetura que utiliza, ao invés de endereços IP, nomes, para efetuar comunicação. Esta arquitetura é utilizada para comunicação entre sistemas Windows através do nome da máquina, que deve ser único para cada uma.

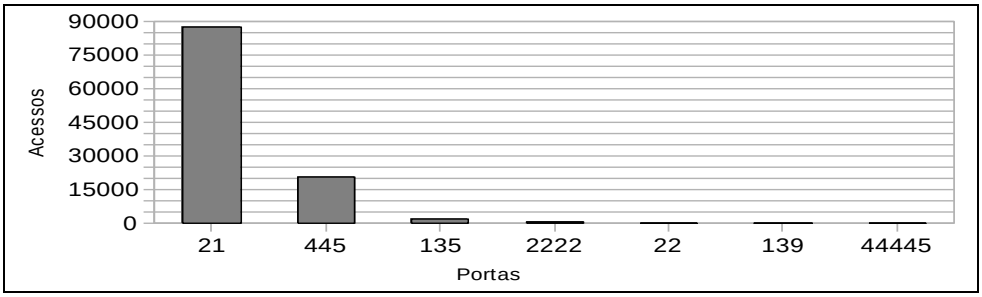


Figura 5. Portas TCP mais acessadas entre 05/09 e 16/09

Dos 87.831 pacotes destinados à porta TCP 21 que foram registrados pelo honeypot, 47.814 vieram de apenas um endereço IP localizado em Taiwan, representando aproximadamente 54% dos acessos a esta porta. Por este motivo, esse endereço IP acabou ficando em primeiro lugar na lista dos cinco IP's que mais interagiram com o honeypot (levando em consideração o número de pacotes enviados), apresentada na Figura 4. O grande número de acessos à porta TCP 21 deve-se ao fato de o Nepenthes emular um servidor FTP a fim de registrar tentativas de ataques, como, por exemplo, ataques de força bruta ou de dicionário. O Nepenthes não registra em log os comandos passados ao servidor de FTP emulado, mas registra as conexões realizadas e, pelo cruzamento dessas informações com os logs do Iptables, foi possível perceber a existência de conexões que duraram até 18 horas, o que sugere um ataque de força bruta ou de dicionário, com o objetivo de descobrir um usuário e respectiva senha válidos para se conectar ao FTP.

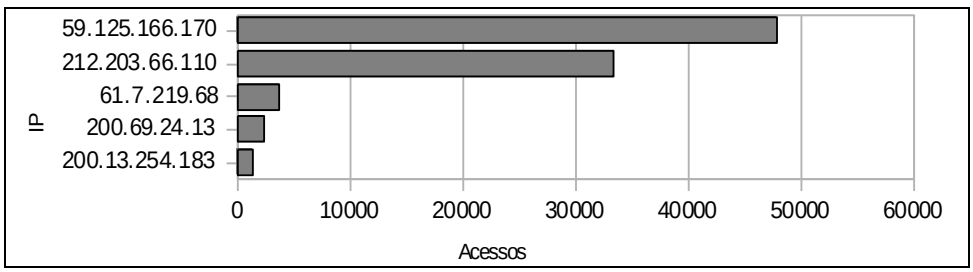


Figura 6. Cinco endereços IP que mais enviaram pacotes ao *honeypot*

Observando ainda o gráfico apresentado na Figura 6, o segundo, terceiro e quarto IPs da lista (provenientes, respectivamente, de China, Tailândia e Argentina), assim como o primeiro (proveniente de Taiwan), também estabeleceram conexão com o servidor de FTP emulado pelo Nepenthes, sendo o segundo responsável por aproximadamente 37% dos acessos a este serviço. Somados, então, os pacotes enviados pelo primeiro e segundo IPs apresentados na Figura 4 representam 91% do total de pacotes, tendo como destino a porta TCP 21 do honeypot. O host 61.7.219.68, antes de realizar conexão na porta 21, tentou, sem sucesso, conexão com a porta 22, que não possuía serviço real ou emulado esperando conexão. Por fim, o quinto IP que mais enviou pacotes para o honeypot, proveniente da Colômbia, foi o responsável pela origem de cinco dos 11 tipos diferentes de malwares identificados por meio do Nepenthes.

Dos endereços IP que foram registrados pelo honeypot, 112 vieram da China, país que liderou o número de acessos, seguido por Estados Unidos, com 66 IPs, e Brasil, com 48, como pode ser observado na Figura 7.

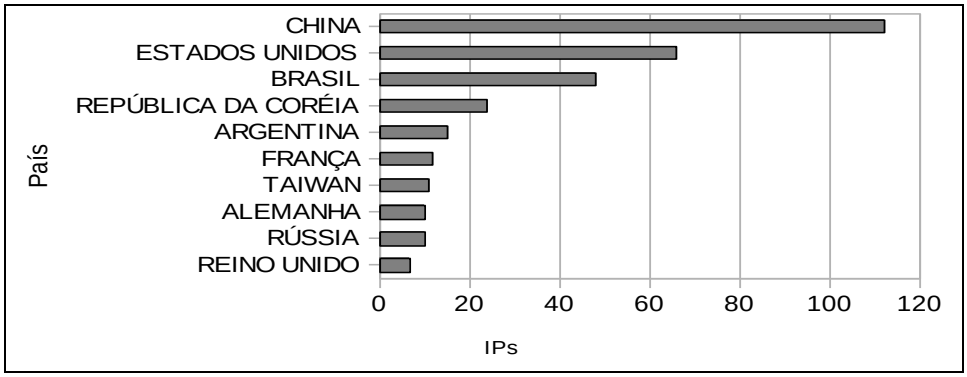
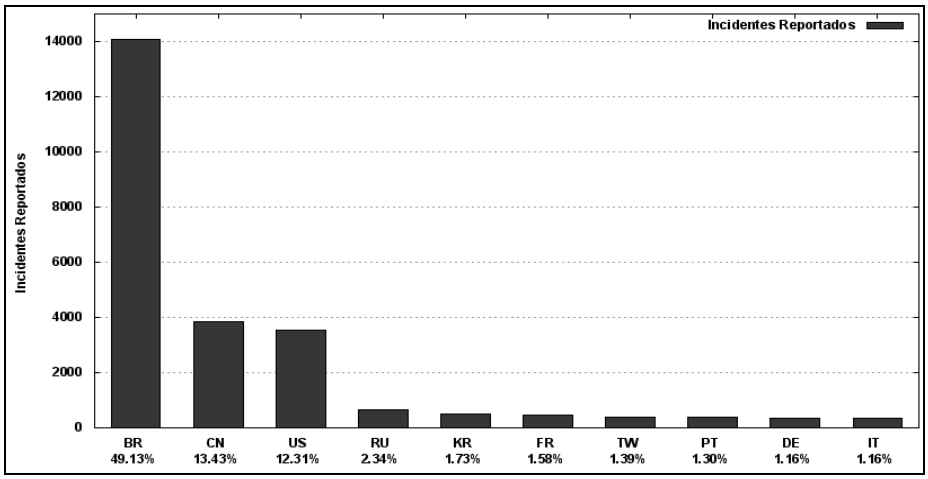


Figura 7. Países com maior número de IPs distintos que acessaram o honeypot

Comparando a origem das tentativas de acesso ao honeypot com estatísticas obtidas por outras organizações que realizam pesquisa semelhante no Brasil, foi possível perceber que os três primeiros colocados, geralmente, são China, Estados Unidos e Brasil, não necessariamente nesta ordem. Como exemplo se pode observar a Figura 6, que mostra um gráfico contendo os dez países de onde mais originaram incidentes de segurança reportados ao CERT.Br no período de julho a setembro de 2008. Pode-se observar ainda que República da Coreia, França, Taiwan, Alemanha e Rússia são países que aparecem em ambos os gráficos apresentados nas Figuras 7 e 8. Outra fonte de dados passíveis de comparação é o Projeto Brasileiro de Honeypots Distribuídos, que fornece a partir do endereço http://www.honeypots-alliance.org.br/stats/ diversas estatísticas referentes aos dados obtidos pelos honeypots, entre elas os países com maior taxa de pacotes por segundo destinados aos honeypots. Desse modo, foi possível comprovar novamente a presença de Brasil, China e Estados Unidos entre os líderes na origem de atividade maliciosa na internet no Brasil.



Fonte: CERT.BR [2]

Figura 8. Países que mais originaram incidentes de segurança reportados ao CERT.BR de julho a setembro de 2008

O primeiro malware coletado pelo Nepenthes foi obtido no dia 06/09/08, um dia após o honeypot ter sido colocado em funcionamento. Ao todo, baseado nos diferentes MD5 gerados, foram capturados no período 11 tipos diferentes de malwares, os quais foram, então, submetidos à avaliação no site http://www.virustotal.com, onde foi possível identificar os nomes de cada um. O sistema empregado por tal site analisa o arquivo submetido utilizando 37 diferentes antivírus. Destes, foi escolhido o AVG como fonte de nomes para os malwares, pelo fato de ser um dos poucos a identificar todos os binários submetidos e por ser um antivírus popular, bastante utilizado por usuários do sistema operacional Windows.

A Tabela 1 mostra os malwares obtidos, ordenados pelo número de ocorrências. Como se pode observar, existem alguns binários que foram identificados pelo AVG com o mesmo nome, do que se pode concluir que se tratam de variações do mesmo malware.

Tabela 1. Dados referentes aos malwares obtidos no período de 05/09/08 a 16/09/08.

MD5	Nome	Ocorrências	IPs Distintos
b09cf5b28972514bf85ee09d184a2a92	BackDoor.Generic_c.CER	29	8
24d9bb993fda99e73be788ed9e724662	BackDoor.VB.LJ	3	1
08627e41d99990bedddab6ca99e6d4f6	Worm/Agobot.CYA	2	2
cfb970abdf8f03bef4b624a0855bbf8a	Dropper.Agent.JXC	2	1
4691ee7642634b6dc45e16bb179991c3	BackDoor.RBot.KB	1	1
474312616dce04c03f13a20a6244b2bf	BackDoor.RBot.KB	1	1
7eeee76d65f7b9417331de4d13bce703	BackDoor.RBot.KB	1	1
3cd1361df4a9b8399802401bcd03d4f6	BackDoor.RBot.KB	1	1
9b8d83c0706b55b4462aee476a5f428c	BackDoor.RBot.KB	1	1
64686bda68f229c82cd54479c8c887bc	Win32/Themida	1	1
8676210e6246948201aa014db471de90	Worm/Lovsan.A	1	1

Como é possível observar na Tabela 1, a predominância dentre os malwares encontrados é de backdoors, os quais são geralmente provenientes de trojans e instalam-se no sistema quando o trojan é executado, fornecendo uma porta de acesso para invasores. Todos os backdoors capturados foram obtidos por meio de tentativas de conexão na porta TCP 135, uma das portas onde o Nepenthes emula serviços. Nessa porta, no sistema operacional Windows, roda por padrão o serviço RPC. Vulnerabilidades nesse serviço são bastante exploradas por diversos malwares. Entre os demais malwares presentes na Tabela 1, apenas o de nome Dropper.Agent.JXC foi obtido a partir de uma porta diferente da 135, a TCP 445.

6 Considerações finais

Com base nas informações apresentadas, foi possível perceber que a maior parte do tráfego gerado a partir da internet foi destinada a ataques contra o serviço FTP, provavelmente de força bruta ou de dicionário. Isso justifica e ressalta a preocupação que universidades como a de Passo Fundo têm e/ou devem ter a respeito das senhas dos usuários de sua rede, principalmente quando o serviço FTP é oferecido a todos e a maioria dos usuários utiliza números sequenciais, como no caso das matrículas, para entrar no sistema. Tais nomes de usuários são muito fáceis de descobrir para serem utilizados em ataques ao FTP.

Em relação aos malwares capturados, foi possível constatar que o principal método de propagação épor meio de vulnerabilidades do RPC e de compartilhamentos abertos. Por isso, ressalta-se a importância de que usuários do sistema operacional Windows se utilizem de métodos de proteção como firewalls, que bloqueiam acessos indevidos a esses serviços, diminuindo as chances de se adquirir um malware através dos mesmos. Isso é válido principalmente para usuários do Windows que estão expostos diretamente a internet, através de um endereço IP real (roteável).

Por fim, ressalta-se a importância da constante monitoração das atividades maliciosas presentes nas redes de computadores através da utilização de honeypots, realizadas tanto em meio acadêmico quanto por instituições

como o CERT.br, em nível nacional, e como o SANS Internet Storm Center, em nível mundial. Dessa forma, garante-se a rápida descoberta de novas ameaças, diminuindo consideravelmente as chances de que alguma delas venha a interferir, ou até mesmo interromper, nas comunicações através da internet ou nas redes locais das organizações.

Como trabalhos futuros propõem-se a investigação e análise de outros tipos de vulnerabilidades e atividades maliciosas, como, por exemplo, as que envolvem o envio de SPAM por meio de servidores de e-mail com relay aberto, podendo, assim, identificar as fontes de envio de SPAM e os seus destinos com a instalação de um servidor de e-mail em um honeypot. Assim, essas fontes podem ser comunicadas às autoridades competentes e podem ser adicionadas a black lists, listas que podem ser consultadas por servidores de e-mail que contêm domínios não confiáveis, cujo e-mail enviado pode ser considerado SPAM e descartado. Outro trabalho possível é a utilização de honeypots de alta interatividade, que proporcionam uma maior interação para permitir a entrada de um atacante e estudar o modo como ele invadiu o sistema, identificando possíveis novas ferramentas e vulnerabilidades que estejam sendo exploradas.

Referências

[1] BAECHER, P. et al. **The nepenthes platform:** an efficient approach to collect malware. In: Recent Advances in Intrusion Detection. [S.l.]: Springer Berlin/Heidelberg, 2006. p. 165-184. Disponível em: <http://www.springerlink.com.w10001.dotlib.com.br/content/660141007r57tk41/fulltext.pdf>. Acesso em: 8 nov. 2008.

[2] CERT.BR. **Stats: julho a setembro de 2008**. 2008a. Disponível em: <http://www.cert.br/stats/incidentes/2008-jul-sep/top-atacantescc.html>. Acesso em: 8 nov. 2008.

[3] CERT.BR. **Stats: julho a setembro de 2008**. 2008b. Disponível em: <http://www.cert.br/stats/incidentes/2008-jul-sep/tipos-ataque.html>. Acesso em: 23 abr. 2008.

[4] CHESWICK, Bill. **An evening with Berferd, in which a cracker is lured, endured, and studied**. 1991. Disponível em: <http://www.tracking-hackers.com/papers/berferd.pdf>. Acesso em: 14 maio 2008.

[5] COHEN, Fred. **The deception toolkit**. Risks Digest, v. 19.62, mar. 1998. Disponível em: <http://catless.ncl.ac.uk/Risks/19.62.html#subj11>. Acesso em: 12 maio 2008.

[6] JABOUR, E. C. M. G.; DUARTE, Otto C. M. B. **Honeynets*:*** invasores, ferramentas, técnicas e táticas. 2005. Disponível em: <http://www.gta.ufrj.br/seminarios/CPE825/tutoriais/eugenia/honeynets.PDF>. Acesso em: 5 maio 2008.

[7] MOORE, David; et al. **Inside the slammer worm**. Security & Privacy, IEEE}, v.1, n.4, p. 33--39, July-Aug. 2003. Disponível em: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1219056&isnumber=273%99>. Acesso em: 6 nov. 2008.

[8] SPITZNER, Lance. **Honeypots***:* tracking hackers. Boston: Addison Wesley, 2002.

[9] SPITZNER, Lance. **Honeypots:** definitions and value of honeypots. 2003. Disponível em: <http://www.tracking-hackers.com/papers/honeypots.html>. Acesso em: 13 maio 2008.

[10] STOLL, Clifford. **Stalking the wily hacker.** 1988. Disponível em: <http://pdf.textfiles.com/academics/wilyhacker.pdf>. Acesso em: 14 maio 2008.