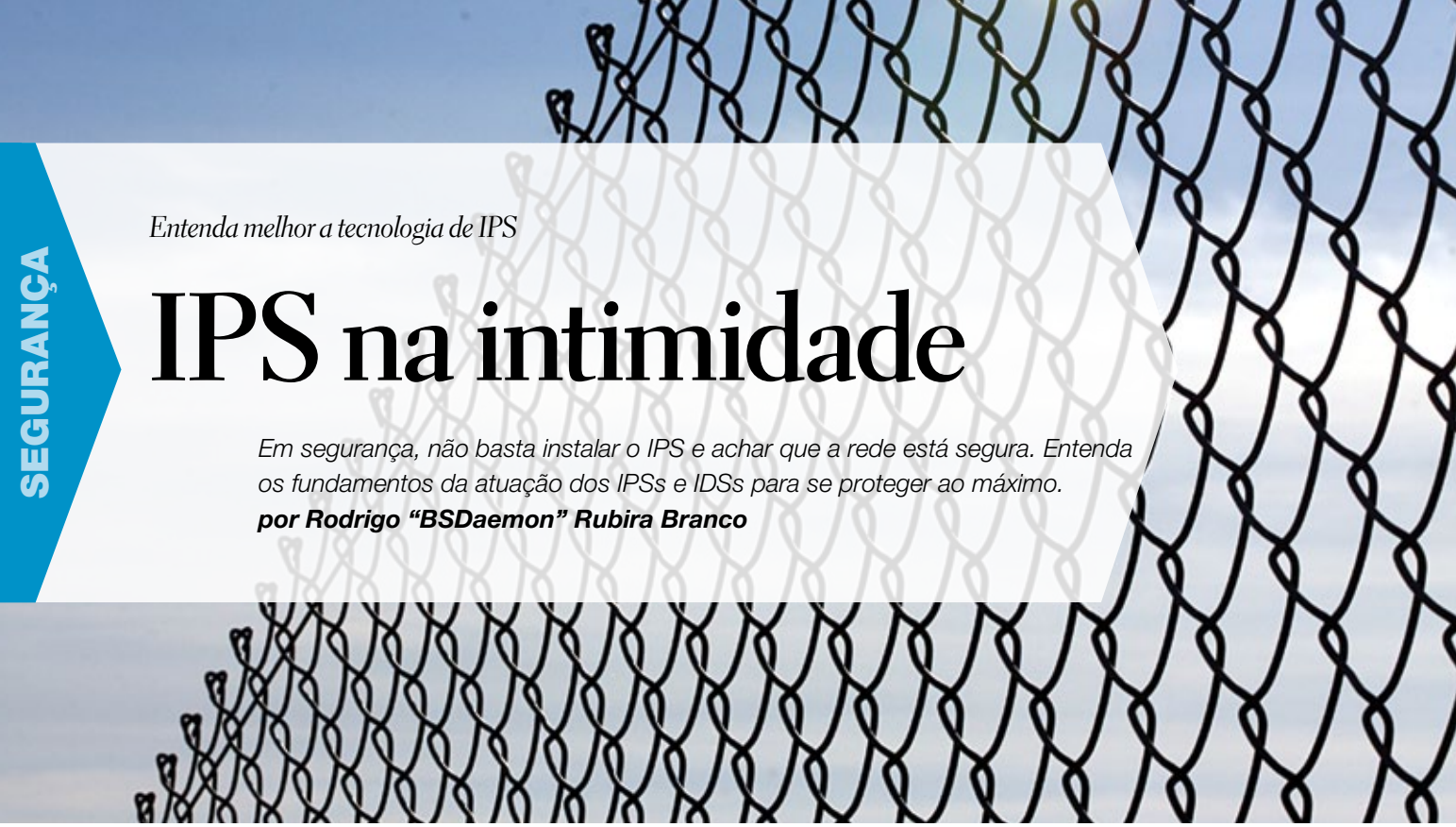

Entenda melhor a tecnologia de IPS

IPS na intimidade

Em segurança, não basta instalar o IPS e achar que a rede está segura. Entenda os fundamentos da atuação dos IPSs e IDSs para se proteger ao máximo.

por Rodrigo "BSDaemon" Rubira Branco

É fato que sistemas para a prevenção de intrusão vêm se tornando realidade para grande parte das empresas, embora exista uma panaceia de termos e tecnologias criadas, o que dificulta (e muito) o real entendimento dos objetivos e funcionamentos de tais sistemas.

Este artigo procura esclarecer de uma maneira mais abrangente o que significa uma solução de prevenção de intrusos, possibilitando ao leitor uma compreensão maior das tecnologias, sem foco específico em algum produto ou solução.

Uma solução de prevenção de intrusos não visa a substituir as necessidade de aplicações de *patches* de segurança em servidores da rede, mas diminuir o intervalo entre a divulgação de uma vulnerabilidade e a aplicação das correções na rede.

Cabe dizer que existem diferentes estados para uma vulnerabilidade: *não pública*, *pública sem correção* e *pública com correção*.

Dizemos que uma vulnerabilidade existe, mas não é pública, quando ela está nas mãos de poucas pessoas e ainda não existe uma correção. Tais pessoas podem ser idôneas (em geral, desenvolvedores das empresas do software vulnerável e pesquisadores de segurança que descobriram a falha) ou não (criminosos que encontram a falha e a estão explorando com um objetivo qualquer).

Uma vulnerabilidade pode ser considerada *pública sem correção* em razão de dois motivos completamente diferentes:

- Um pesquisador encontrou a vulnerabilidade e, após múltiplas tentativas de contactar o fornecedor, não obteve resposta, ou a resposta não foi satisfatória – discordância a respeito da existência da vulnerabilidade, vulnerabilidade em produto com suporte descontinuado, tempo para desenvolvimento da correção ou entendimento do problema muito prolongados –, fazendo com que o pesquisador divulgue a falha encontrada, ainda que não exista correção;
- Profissionais de segurança descobrem que alguma vulnerabilidade não pública foi explorada em seus sistemas (sejam eles *honeypots* ou não) e examinam os detalhes da exploração para descobrir qual a falha explorada. Este caso é o mais crítico, pois significa que a vulnerabilidade já está sendo explorada e as empresas fornecedoras de TI nem sequer sabem de sua existência.

Quando uma vulnerabilidade é dita *pública com correção*, muitas vezes isso significa que ela só foi divulgada quando já existia uma correção.

Independentemente do estado das vulnerabilidades, sabe-se que existe um intervalo de tempo entre a divulgação de uma correção, se ela existir, e a efetiva aplicação nos sistemas de uma rede. Mesmo em redes com diversas camadas de segurança, sistemas automatizados para instalação de correções e NAC (controle de acesso a rede), sabe-se que o tempo para a configuração efetiva de auditoria, instalação das correções e quarentena continua sendo insuficiente, dada a velocidade com que as falhas têm sido exploradas. Isso sem falar das vulnerabilidades públicas sem correção.

Esse intervalo entre a divulgação da vulnerabilidade e a aplicação de sua correção é denominado *Tempo de Exposição*: trata-se do tempo que uma empresa fica exposta a uma ameaça conhecida (sem falar das ameaças desconhecidas, abordadas mais adiante). Um sistema de IPS (no original em inglês, *Intrusion Prevention System* – Sistema de Prevenção de Intrusões) existe primariamente para reduzir ao máximo o tempo de exposição.

Existem duas formas de se prevenir contra uma ameaça:

- prevenção contra a exploração da vulnerabilidade: entendimento e validação de protocolos, assinaturas específicas para os padrões do ataque, análise comportamental, entre outras tecnologias e procedimentos que podem ser implementados com tal finalidade;
- prevenção contra uma forma de se explorar uma vulnerabilidade: também conhecido como "assinaturas para exploits", que nada mais são do que maneiras de se detectar uma forma conhecida de se explorar uma dada vulnerabilidade. Tais assinaturas geralmente são mais simples de desenvolver e costumam ser as primeiras a ser oferecidas pelos fabricantes, pois não exigem o completo entendimento das vulnerabilidades.

Tais assinaturas, em geral, podem ser contornadas quando se tem um conhecimento da tecnologia de proteção e da assinatura em si. Por esta razão, alguns fabricantes fecham suas assinaturas ao público. Outro motivo para fechar as assinaturas ao público é a existência de assinaturas para falhas ainda não divulgadas. Isso ocorre quando o fabricante compra vulnerabilidades de segurança ou possui times de pesquisa internos para detectá-las antecipadamente.

As tecnologias de prevenção de intrusão, de uma forma em geral, evoluíram para proporcionar a detecção com base em diversos fatores (tecnologias híbridas): assinaturas, comportamento, validação de protocolos e tecnologias de detecção.

Assinaturas

Muitos entendem assinaturas como simples busca de padrões (também conhecido como *pattern matching*), ou busca de strings. Embora em algumas tecnologias isso muitas vezes seja verdade – por exemplo, *Snort*, quando não tratamos as regras dinâmicas –, isso não se aplica a todas as soluções. Uma assinatura pode ser um conjunto elaborado de condições e regras para validação dos dados que passam pelo sistema. Por isso é importante entender quais os protocolos analisados e entendidos pelo sensor a ser avaliado. Por exemplo, uma simples busca de strings poderia identificar um `1=1` como sendo um ataque e emitir um alerta (a propósito, `1=1` é um padrão comumente utilizado em injeção de comandos SQL, uma classe de vulnerabilidades existente primariamente em sistemas baseados na Web). Mas tal sistema não seria capaz de diferenciar se essa string está sendo enviada como parte de uma mensagem inofensiva de email ou no corpo de uma página, ou ainda em uma URL. Vale lembrar também das possíveis variações para essa assinatura, pois `10=10` possui o mesmo efeito do ponto de vista do atacante, mas não geraria um alerta.

Comportamento

A detecção baseada em comportamento envolve a criação de um *baseline*, um padrão normal do comportamento da rede. Mudanças nesse padrão são indicativas de um ataque. Algumas tecnologias possuem recursos mais avançados para isso, enquanto outras dependem de elementos externos. O importante é entendermos que, embora seja um recurso precioso para detecção de ameaças como *worms*, que sabidamente mudam o comportamento da rede, não são eficientes para todos os casos. Ataques visando um alvo específico normalmente não geram anomalias perceptíveis pela rede. Por essa razão, os sensores modernos usam tecnologias híbridas.

Validação de protocolos

Todas as soluções modernas oferecem algum tipo de validação de protocolos, independentemente das assinaturas existentes. Deveriam realizar, por exemplo, a remontagem de sessões (identificar uma sessão, ou seja, sequência de dados pertencentes a uma mesma transação por parte de um protocolo específico), a remontagem de pacotes (chamada de *desfragmentação virtual*, semelhante à remontagem de sessões, porém ocorrida na camada de rede, mais baixa que a de sessão), entre outras.

Tecnologias de detecção

De uma forma geral, todas as soluções – proprietárias ou não – possuem outras formas de detecção de ataques. Um exemplo é a verificação de instruções de máquina passando pela rede e a emulação das instruções para evitar falsos positivos. Este método serve para a detecção genérica de explorações com inserção de código arbitrário (*shellcodes*).

Habilitação de regras (assinaturas e configurações)

É importante mencionar que a gerência de sensores em uma rede é um dos principais fatores para o su-

cesso de uma solução de prevenção de intrusos. As assinaturas não farão diferença se não estiverem habilitadas.

Algumas assinaturas são associadas ao equipamento alvo, ou seja, ao sistema operacional e seus softwares. Por questões de otimização de desempenho dos equipamentos, é importante que vulnerabilidades que não possam afetar a rede – por já estarem corrigidas ou porque os sistemas alvos não existem – tenham suas respectivas assinaturas desabilitadas, a menos que a solução ofereça degradação desprezível de desempenho mesmo com milhares de assinaturas habilitadas.

Diversas soluções já possuem formas automatizadas de reconhecimento de mudanças da rede para habilitação e desabilitação de assinaturas, evitando exposições indevidas – por exemplo, quando um novo servidor não autorizado é instalado na rede. Tal reconhecimento pode ser realizado pelo próprio sensor, sendo este o responsável final pela implantação da segurança, de forma passiva, ou por outro elemento na rede, como um analisador de vulnerabilidades reativo que precisa ser executado na rede periodicamente, ou ainda um analisador de rede integrado ao sensor.

O poder de customização de assinaturas é fundamental para este tipo de tecnologia. Softwares desenvolvidos exclusivamente pela empresa, soluções menores de tecnologia e outras necessidades especiais dificilmente serão atendidos pelos fabricantes. Customização é a palavra-chave nestes casos. Mesmo que a equipe responsável não tenha conhecimento suficiente para realizá-la, é possível a contratação externa de um especialista para seu desenvolvimento. Isto diminui a ocorrência de falsos positivos ou falsos negativos e melhora o desempenho da solução utilizada, com consequente ganho de produtividade por parte da equipe que gerencia a solução.

É muito importante diferenciar também o que são assinaturas de ataques com destino a servidores e assinaturas de vulnerabilidades em softwares clientes. Isto porque se um sistema de prevenção de intrusos for instalado para proteger uma DMZ, quase seguramente poderemos desabilitar todas as assinaturas de clientes e melhorar consideravelmente o desempenho geral, pois permitimos assim a inspeção de tráfego unidirecional. O inverso também é valido se o sistema de prevenção servir para a proteção de uma rede que consiste apenas de sistemas de usuários.

Desempenho

Desempenho é um fator fundamental para o sucesso de uma solução de prevenção de intrusão.

Os fabricantes em geral não divulgam as avaliações de desempenho de seus sistemas ou equipamentos com todas as assinaturas de prevenção ativadas, e um parâmetro de desempenho sem especificação de quais proteções estão ativas pode não refletir o comportamento da solução no dia a dia. Lembre-se sempre de definir quais as proteções mínimas necessárias para seu ambiente. Preferencialmente, antes de escolher qualquer configuração, o ideal é um teste utilizando as definições de vulnerabilidades a ser protegidas e as formas de proteção exigidas.

Topologias e disponibilidade

Diretamente relacionada ao desempenho vem a questão topológica dos equipamentos. Isto porque um IPS trabalha de forma *inline* (em linha, ou seja, todas as conexões inspecionadas devem necessariamente passar por ele) e pode ou não possuir balanceamento de tráfego e alta disponibilidade.

Em geral, um sensor não possui IP e atua em modo *bridge*. Este é um dos grandes diferenciais entre um sensor embarcado – comumente visto em caixas UTM – e um sensor dedicado. Sensores embarcados em geral estão em equipamentos com IP, pois atuam também como firewalls na rede. A questão de a profundidade e os recursos de inspeção presentes em tais sensores serem inferiores a sistemas dedicados não pode ser tratada como verdadeira em todos os casos. Deve-se sempre validar se existem formas de priorização de serviços para evitar que inspeções do IPS afetem o desempenho do equipamento como firewall.

As soluções de IPS hoje possuem também a possibilidade de instalação como IDS – *Intrusion Detection System*, Sistemas de Detecção de Intrusão –, sendo estes sistemas não-inline (não-em-linha, ou seja, mesmo se o sensor estiver inativo, o tráfego continuará fluindo, pois o sensor recebe apenas uma cópia do que passa pela rede). Existem ainda recursos de aprendizado por assinatura – nele, todo o sensor está em modo de aprendizado e não efetua bloqueios, apenas gera alertas para os ataques – ou ainda modos em que cada assinatura é configurada separadamente, podendo estar em modo de aprendizado, com algumas assinaturas específicas em modo de bloqueio. Vale notar que os sistemas instalados unicamente como IDS estão em franco declínio.

Outra situação de topologia comum é a implementação de um mesmo sensor para múltiplos segmentos de rede. Embora o desempenho mereça ser considerado para que ataques em um segmento não prejudiquem os demais, esse requisito é muito comum e pode ser fornecido de diversas formas. Uma delas é a existência de diferentes políticas por segmento – isto é, por par de interfaces físicas ou *tags* de VLAN

(rede virtual) em um sensor. Esta é a forma mais comum, pois permite a visualização clara de qual segmento possui quais regras, embora dificulte a gestão do sensor em si, que funciona como se existissem múltiplos sensores na prática. Outra abordagem é o uso de exceções por regra ou por sensor, em que diversos parâmetros podem ser utilizados para criar exceções às regras. Por exemplo, determinado conjunto de IPs pode ser desconsiderado na regra de vulnerabilidades em servidores Apache. Essa abordagem facilita o gerenciamento das assinaturas e proporciona mais segurança, mas dificulta a compreensão do que está ativo para cada segmento.

Por estarem inline no tráfego de rede, as soluções de prevenção de intrusão devem possuir algum tipo de redundância.

Interfaces de *bypass* são placas de rede especiais que permitem ao tráfego fluir mesmo com o equipamento desligado. Essa é uma das formas mais usadas e de melhor custo, podendo ser internas ou externas. Quando se deseja alta disponibilidade, podem-se usar dois sensores inline com interfaces de bypass interligadas. Na prática, todo o tráfego é inspecionado duas vezes (duplicando a latência) e, no caso da falha de um sensor, a interface manterá o tráfego passando até o outro. O problema, nesse caso, é evidente: em caso de falha na interface de bypass, a rede para.

Outra forma de se implementar a alta disponibilidade é por meio de um equipamento externo que, em caso de falhas, desvia o tráfego para outro sensor. A grande maioria das soluções com essa topologia tem problemas de continuidade em virtude do roteamento assimétrico: a volta dos pacotes muitas vezes ocorre por um sensor diferente daquele pelo qual chegaram originalmente. Independentemente do bom funcionamento desta topologia, há perda de segurança, já que as sessões não são completamente remontadas para detecção dos ataques, a menos que a questão do roteamento assimétrico seja resolvida ou haja sincronismo de informações entre os sensores.

A forma mais vantajosa de implementar a alta disponibilidade é com balanceamento de carga (*load sharing*). Com ele, em caso de falha de um sensor, o outro equipamento mantém a inspeção. Tais soluções em geral encarecem o projeto por três motivos:

- exigem mais de um dos equipamentos, e cada um dos equipamentos precisa ser capaz de sustentar sozinho a vazão (*throughput*) total da rede, pois, na falha de um, o outro equipamento fica encarregado de toda a rede;
- não atendem aos requisitos de escalabilidade tão facilmente: por exemplo, em caso de adição de mais equipamentos para melhor desempenho;
- diminuem o desempenho dos equipamentos. São necessários equipamentos de maior desempenho para atender à mesma vazão, já que a sincronização dos estados internos do IPS exige muita performance e degrada o equipamento.

Uma opção interessante disponível no mercado é a implementação de topologia de balanceamento de carga **sem** a sincronização dos estados de IPS. Nela, os IPSs dividem o trabalho de inspeção, pois, na prática, não precisam se comunicar para isso – e as sessões são sempre tratadas pelo mesmo equipamento.

Esta solução permite a escolha entre segurança e conectividade. Caso se opte pela segurança, a queda de um equipamento faz com que as **sessões** ativas sejam bloqueadas pelo outro equipamento (*out-of-state*). Já se for feita a opção pela conectividade, as sessões ativas serão permitidas, por um período de tempo, sem inspeção. Novas conexões serão normalmente inspecionadas.

Appliance x software

A escolha entre *appliance* e software depende fortemente da integração com parceiros e fornecedores, pois algumas empresas possuem mais facilidades no atendimento, suporte e preços com determinadas marcas, bem como das diretivas de negócio de cada empresa. Muitas, por exemplo, optam pelo uso de appliances por facilitarem a manutenção e o suporte em geral.

Cada fabricante possui sua própria estratégia nesse quesito, mas a escolha do que é melhor cabe ao cliente, não sendo uma vantagem possuir uma opção híbrida no caso de um cliente que opte apenas por appliances.

O mais importante é conseguir os números corretos de desempenho, baseados em critérios reais e não na famosa "vazão de pacotes UDP de 1500 bytes" sem qualquer assinatura ativada.

Falsos positivos e negativos

Um dos principais motivos para o insucesso na adoção de soluções de prevenção de intrusos está na alta taxa de falsos positivos – eventos normais e inofensivos detectados como ataques –, o que gera bloqueios indevidos. Há também os falsos negativos – excesso de ataques que conseguem de alguma forma passar pelo sensor sem ser detectados.

Evitar totalmente falsos negativos é fácil. Basta detectar todo o tráfego como ataque. Isso nos dá 0% de falsos negativos, mas ao mesmo tempo algo próximo de 100% de falsos positivos. O oposto também é verdadeiro, ou seja, se não alertarmos para nada, não teremos falsos positivos, mas teremos 100% de falsos negativos para

os ataques. O grande desafio está no equilíbrio entre ambos.

Como já mencionado, o aprendizado da rede melhora o desempenho do equipamento, desabilitando assinaturas desnecessárias, mas também pode ajudar na diminuição de falsos positivos. A customização de assinaturas também auxilia na melhora, pois ajuda a diminuir falsos positivos e a aumentar a detecção de ataques de sistemas cujas vulnerabilidades não possuam assinaturas.

Cada assinatura possui uma probabilidade de ser um ataque real. Os fabricantes em geral definem essa taxa com base em critérios próprios, em sua maioria assertivos, mas nem sempre válidos para todos os casos. A customização de assinaturas deve permitir a modificação da probabilidade de ataques reais, bem como a definição do grau de risco aceitável para a organização.

Análise detalhada

Os ataques bloqueados devem ser devidamente analisados, pois podem:

- ser falsos positivos: neste caso, o tráfego normal da rede é bloqueado. Se a assinatura estiver em modo de aprendizado, o tráfego é permitido, mas um alerta é emitido;
- merecer o bloqueio em vez de simples alertas: neste caso, é preciso tomar ações, pois é fato que ao menos um ataque passou pelo sistema. Quais os impactos? Em que consiste o ataque? Tudo isso deve ser provido facilmente pelo sistema de visualização de alertas;
- tratar-se de um ataque disparado em massa, comum em situações como worms procurando automaticamente por sistemas vulneráveis e ataques a faixas de IP, entre outros. Em geral, é seguro ignorar tais alertas se o ataque for devidamente bloqueado;
- tratar-se de um ataque direcionado: esse tipo de ataque necessita de tratamento diferenciado, por ser uma tentativa específica de exploração contra a organização. Isso significa que o agressor possui algum motivo para infligir danos de alguma espécie à empresa, e que provavelmente procurará outras formas de contornar os sistemas de segurança.

As atividades desenvolvidas para os alertas gerados devem ser registradas para posterior consulta, geração de relatórios ou até mesmo interação entre múltiplos analistas. A filtragem para rápida visualização dos alertas, categorias e outros critérios deve ser facilitada, o que quase sempre exclui interfaces web, principalmente no caso de sistemas que recebam alertas de múltiplos sensores, pois a quantidade muitas vezes inviabiliza o uso de tais interfaces.

Principalmente as redes com múltiplos sensores devem ter projetos que considerem o formato de visualização dos alertas destes sensores em um ponto central, controlem o uso de disco de tais sistemas (RAID é essencial para o desempenho) e, obviamente, *correlação*. Não importa se a correlação for feita pela própria tecnologia ou por softwares terceiros, o importante é que ela seja feita. Não é necessário correlacionar ataques apenas de sensores diferentes, mas também de um mesmo sensor. Por exemplo, podemos aumentar a prioridade se múltiplos alertas forem gerados a partir de uma mesma origem ou visando um mesmo destino.

Fica aqui o alerta contra reações automatizadas, como listas negras (*blacklists*) baseadas em elementos que podem ser facilmente forjados. Imagine a situação em que automaticamente se coloca em uma lista negra o IP de um atacante que tente explorar uma falha em um servidor DNS, protocolo que utiliza normalmente a porta UDP 53. Neste caso, um atacante poderia forjar um IP e causar o bloqueio de outro equipamento, pois o protocolo UDP não possui uma sessão.

Conclusão

Este artigo tentou esclarecer um pouco dos sistemas e das tecnologias de prevenção de intrusos e, com isso, gerar diversas novas questões sobre o tema para os leitores.

Como última consideração, levantamos a importante questão da instalação desses sistemas, que, além de considerar todos os itens anteriormente mencionados, deve levar em conta os seguintes critérios:

- definição da topologia a ser utilizada;
- definição da tecnologia a ser utilizada;
- instalação do equipamento em modo de aprendizado, para evitar paradas na rede;
- visualização dos logs pelo período adequado a cada organização;
- habilitação gradual do modo de bloqueio. ■

Sobre o autor

Rodrigo "BSDaemon" Rubira Branco (rbranco@la.checkpoint.com) atua como Security Expert na empresa Check Point Software Technologies. Membro do comitê de pesquisas da Conviso e consultor sênior de vulnerabilidades para a COSEINC, também atuou como Principal Security Researcher na empresa Scanit (maior fornecedor de segurança dos Emirados Árabes) e desenvolvedor Linux no Advanced Linux Response Team da IBM. Rodrigo é mantenedor de diversos projetos de código aberto e palestrante nas mais importantes conferências de pesquisa em segurança no mundo. Como membro do grupo RISE Security (www.risesecurity.org), divulgou diversas vulnerabilidades de segurança. Também é instrutor dos treinamentos SANS e membro do comitê da certificação GIAC em Engenharia Reversa.

DECISÕES CERTAS PODEM MUDAR O RUMO DE SUA CARREIRA

Inclua em seu currículo a principal certificação Linux no mundo – LPI.

Em tempos de crise, soluções de código aberto – como o Linux – se destacam na adoção por empresas de todos os tamanhos, como solução ideal para aumentar eficiência nos negócios e reduzir custos. Atualmente há no mercado uma carência por profissionais certificados para atender a essa demanda crescente. Aproveite essa oportunidade e inclua em seu currículo a principal certificação Linux no mundo.

As datas de realização das provas são:

18/07/2009 – Vitória/ES
01/08/2009 – Fortaleza/CE
01/08/2009 – São Paulo/SP
03/10/2009 – São Paulo/SP

Inscrições e mais informações:

treinamentos@vectory.com.br
Tel (11) 4082-1305

Linux Professional Institute

Vectory
CAPACITANDO PARA O FUTURO