Uso de Workflows Científicos para Apoiar a Elaboração de Técnicas de Predição de Invasão de Sistemas

Valeria Farias Alves
valeria.farias@gmail.com
CEFET RJ

Cristina Gomes de Souza
crisgsouza@gmail.com
CEFET RJ

Raphael Machado
raphael@clavis.com.br
CLAVIS RJ

Eduardo Ogasawara
eogasawara@@cefet-rj.br
CEFET RJ

Eduardo Bezerra
ebezerra@@cefet-rj.br
CEFET RJ

Resumo:O avanço tecnológico proporciona oportunidades de maior conectividade entre sistemas e pessoas. Como consequência, observamos um ecossistema global com bilhões de dispositivos conectados a uma miríade de serviços que consomem e produzem uma grande quantidade de dados. A existência desse ecossistema inevitavelmente suscita surgimento dos ataques cibernéticos. Esses ataques compreendem as inúmeras formas de violação da segurança da informação, que incluem acessos não autorizados, roubos de informação e paralização de serviços. De fato, os ataques cibernéticos vêm crescendo em ocorrência e estão se tornando nocivos para uma sociedade cada vez mais dependentes de transações digitais. Estes ataques podem ocorrer a partir de motivações social, política ou econômica. As principais técnicas de prevenção são predominantemente baseadas em análise individual de pacotes de rede. Este trabalho propõe o uso de workflows científicos como uma forma de apoiar a elaboração de novas técnicas de predição de invasão de sistema, atuando na análise do fluxo de acessos e pacotes de dados, de forma a analisar e detectar comportamentos de tráfego suspeitos.

Palavras Chave: Segurança - Ataque Cibernético - Workflow Científico - Invasão -

1. INTRODUÇÃO

O contínuo avanço tecnológico e o crescimento da internet alavancou a interconexão entre inúmeros dispositivos. Em paralelo a este crescimento observou-se, também, uma evolução dos chamados ataques cibernéticos, procedimentos que visam comprometer a segurança da informação e de sistemas computacionais. Com o aumento da dependência da sociedade em relação a sistemas de software, os impactos de um ataque cibernético tornam-se cada vez mais relevantes.

O aumento da quantidade e da relevância dos ataques cibernéticos traz a necessidade do desenvolvimento de sistemas de defesa contra ataques cibernéticos. No núcleo de tais sistemas de defesa estão os chamados "sistemas de detecção de ataque", cujo propósito é detectar, por meio da análise de uma imensa quantidade de dados em trânsito, a presença de tráfego potencialmente malicioso. A maioria das técnicas de defesa baseia-se em análise individual de cada pacote trafegado, o que torna cada vez mais difícil escalar tal abordagem, principalmente nos cenários de negação de serviço, onde é mais importante analisar o comportamento dos acessos, conhecidos como fluxos de acesso, do que os pacotes individualizados (HELLEMONS et al., 2012).

Apesar do grande número de ataques, poucos esforços foram realizados no sentido de consolidar bases de dados com as ocorrências de ataques para que pesquisadores possam aumentar ou melhorar os mecanismos de defesa (PFAHRINGER, 2000; SPEROTTO et al., 2009). Iniciativas como a de Sperotto et al. (2009) procuram deixar disponível para a comunidade uma base de dados para o desenvolvimento de técnicas de mineração de dados que possibilitem identificar comportamento de invasões. Entretanto, a ausência de uma sistematização deste procedimento pode tornar as bases obsoletas frente ao surgimento de novos ataques cibernéticos.

Um fenômeno similar aconteceu com o conjunto de dados disponibilizado pela KDD CUP 99 (PFAHRINGER, 2000), que hoje é considerado obsoleto (VASUDEVAN et al., 2011). Como esses esforços são espúrios, hoje não temos uma base atualizada disponíveis para a realização de pesquisas e testes. As bases que sofrem constante atualizações, são as bases relacionadas a assinaturas de vírus, presentes em softwares de antivírus, de forma a gerar a proteção que se espera deste tipo de ferramenta.

Em uma visão mais ampla, dado o grande volume de dados associado à variedade e à complexidade dos métodos de ataque atualmente à disposição de usuários maliciosos, torna-se fundamental que sistemas de detecção de intrusão sejam desenvolvidos em conformidade com preceitos científicos, com vistas a garantir a reprodutibilidade dos eventos detectados. Essa conformidade também inclui dominar o processo de produção de fluxos de acesso, de produção de bases curadas (i.e., com marcação de ataques validadas por especialistas) e de avaliação.

Visando, em especial, todos aqueles que lidam com a gestão de Tecnologia da Informação (TI) nas mais diversas organizações, o trabalho apresenta uma proposta de sistema de detecção de intrusão de tempo real fazendo uso de três workflows de apoio. Esses workflows, que consistem na definição de uma sequência de processos e tarefas a serem desenvolvidos em determinada base de dados, visam detectar ataques cibernéticos apoiando a análise e a predição de incidentes relacionados à segurança da informação.

Esse trabalho encontra-se organizado em seis seções. Na seção 2 são descritos diversos tipos de ataques cibernéticos. A seção 3 traz as formas de análise desses ataques. A seção 4 aborda o conceito de workflow e algumas considerações sobre sua aplicabilidade na predição das invasões. A seção 5 apresenta a proposta do sistema de detecção de intrusão de

tempo real. As considerações finais sobre a utilização dos workflows propostos constam na seção 6.

2. ATAQUES CIBERNÉTICOS

A temática relacionada à ataque cibernético está tomando força a cada dia, devido a dependência cada vez maior da tecnologia para a realização das mais diversas atividades, sejam elas profissionais ou pessoais, no mundo da internet. Os crimes cibernéticos não se limitam apenas aos danos diretos, como a interrupção de determinado serviço, que poderia provocar danos a determinada marca, mas a danos indiretos a toda a sociedade, como é o caso das espionagens e outros atos antiéticos utilizados por empresas privadas e governos de alguns países, de forma a se beneficiarem de informações privilegiadas.

Com objetivos que podem ter cunho financeiro, político ou até mesmo pessoal, por meio de programas maliciosos, informações podem ser obtidas indevidamente, computadores podem ser manipulados silenciosamente, paralisações de serviços podem ser realizadas, entre os mais diversos malefícios que programas que trabalham de forma sigilosa podem provocar. Estes programas maliciosos podem chegar até os sistemas de várias formas, podendo gerar danos irreparáveis, interrupções de serviços por várias horas, provocando assim grandes perdas financeiras, seja por roubo de informações ou por danos causados a uma marca. A seguir, são apresentadas algumas formas de ataques aos sistemas e as suas principais características:

- Ataques de Negação de Serviço (DOS) - Os ataques DOS (do inglês *Denial of Service)*, caracterizam-se pelo uso de ferramentas automáticas, que enviam comando de solicitações a um determinado servidor, sobrecarregando-o, a ponto de torná-lo bem lento ou indisponível. Conforme (JORGE e WENDT, 2013), o ataque de DOS é tipificado no Código Penal Brasileiro, nos artigo 166 e 265, por perturbação e interrupção de serviços, quando provocado a indisponibilidade de serviços de utilidade pública, cabendo assim punição legal para o autor.
- Vírus - É um programa malicioso desenvolvido para causar ações danosas e ilícitas em um sistema, sendo transmitido de computador para computador, a partir de downloads ou cópias de arquivos, sendo instalados apenas após a sua execução.
- *Worm* - São programas semelhantes ao vírus, com a diferença de que são capazes de se propagarem automaticamente entre máquinas conectadas a rede ou através de pen drives, não dependendo de interferência humana para se multiplicar, enviando copias de si mesmo para outras máquinas. Sua propagação se dá por meio de falhas ou vulnerabilidades de sistemas. Após a infecção por *worms*, as máquinas podem ser facilmente acessadas e manipuladas remotamente por outras pessoas, de forma silenciosa.
- *Botnets* - Se caracteriza por uma rede de computadores infectados por um aplicativo capaz de se comunicar com os invasores, podendo ser programado para realizar tarefas específicas dentro do computador do usuário afetado.
- Cavalo de Tróia - Também conhecido como Trojan, do inglês *Trojan Horse*, é um código malicioso que se esconde sob a forma de algo útil ou inofensivo. Não se reproduz e não se espalha automaticamente, necessitando que um arquivo seja executado para que ele seja instalado. Normalmente, os Cavalos de Tróia são anexos de e-mails ou links de páginas falsas, que quando são abertos instalam o programa mal intencionado.
- *Keylogger* - Termo oriundo da língua inglesa, que significa registrador de teclado. São programas espiões que, uma vez instalados no sistema do usuário, realizam o

monitoramento de suas atividades e enviam as informações coletadas para terceiros, por meio da internet.

3. FORMAS DE ANÁLISES DE ATAQUES

Ataques cibernéticos são realizados por meio de técnicas que exploram falhas nos protocolos da internet, sistemas operacionais e aplicativos de software. Várias medidas de segurança têm sido propostas visando a identificar atividades de invasão. Em particular, ferramentas automáticas de detecção vêm sendo amplamente adotadas e difundidas. Neste contexto, os *sistemas de detecção de invasão* (IDS, do inglês *Intrusion Detection Systems*) procuram monitorar os eventos que ocorrem nos computadores ou na rede a procura de sinais de invasão (OLUSOLA et al., 2010).

Conforme Nakamura e Geus (2007), um IDS tem como objetivo detectar atividades suspeitas, impróprias, incorretas ou anômalas, sendo capaz de detectar ataques trabalhando como uma câmera, monitorando todo tipo de tráfego e podendo realizar a detecção com base em algum tipo de conhecimento tais como perfil de um ataque ou desvio de comportamento, gerando logs dos tráfegos para análise e atuação. Dentre a funções de um IDS podemos destacar a coleta, análise, armazenamento e, em alguns casos, resposta a atividades suspeitas. A Figura 1 ilustra, de forma esquemática, as funções do IDS.

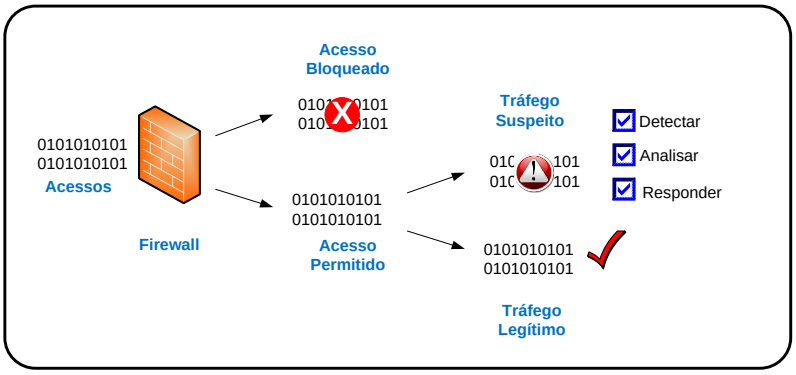

Figura 1: Funções do IDS.

Com as informações registradas nos logs do IDS, é gerado um banco de dados com o fluxo das informações de uma determinada comunicação ou comunicações na internet. Por meio de técnicas de análise de ataques cibernéticos, a análise das informações registradas nos logs podem ser processadas com o objetivo de extrair informações para apoio a tomada de decisões.

Existem dois tipos de técnicas para detecção: as técnicas *baseadas em conhecimento* e aquelas *baseadas em comportamento*. As técnicas baseadas em conhecimento buscam pelas denominadas "assinaturas de ataque", padrões conhecidos de ataques cadastrados na base de dados do IDS. As técnicas baseadas em comportamento procuram ataques pelas detecções de mudanças de padrões de utilização ou mudanças de padrões comportamentais no sistema (PFAHRINGER, 2000).

Com o crescente aumento no volume e na vazão de dados na internet, é impraticável analisar cada pacote de dado recebido, principalmente no momento de um ataque cibernético (SPEROTTO e MEENT, 2007). Em vez disso, são necessárias técnicas de amostragem e consolidação de dados. Em particular, abre-se espaço para análise não mais apenas por pacotes ou características de pacotes, mas sim pelo comportamento dos acessos. Os acessos

de uma mesma origem em uma determinada janela de tempo determinam um *fluxo de conexões* (HELLEMONS et al., 2012).

4. WORKFLOWS

A definição de workflow surgiu incialmente para detalhar processos administrativos realizados em um escritório, definindo as etapas de execução de determinada tarefa. Em 1996, a WfMC (*Workflow Management Coalition*), organização global formada por analistas, consultores, universitários e pesquisadores de gerenciamento dos processos de negócios, definiu workflow como 'a automação de um processo de negócio, total ou parcial, na qual documentos, informações ou tarefas são transferidas entre participantes de acordo com um conjunto de regras'.

Workflows podem ser implementados, portanto, para definição de processos de negócios e também para a estruturação de dados, com o objetivo de se obter resultados. Um workflow envolve a combinação de dados e processos, em uma sequência estruturada de passos, visando à solução de um problema.

O workflow é composto pelo sequenciamento de tarefas, pelas parametrizações destas tarefas, pela definição dos dados que serão inseridos e pelo controle dos fluxos necessários, obtendo-se um resultado, a partir do qual será gerada a análise. Para a análise do resultado, o especialista deverá consultar a execução das atividades, ou seja, a execução do workflow e identificar se ocorreu algum erro na execução de algum processo ou se existe algum fator que mereça destaque na avaliação. É de grande importância que os testes e seus resultados possam ser reproduzidos.

O uso de workflows pode ser aplicado a várias áreas do conhecimento como, por exemplo, a biologia. Através da bioinformática – que é a área responsável pela criação, desenvolvimento e operação de banco de dados e outras ferramentas computacionais para coletar, organizar e interpretar dados biológicos derivados de diversos experimentos – pode-se obter resultados quantitativos e qualitativos.

Nesse trabalho buscou-se aplicar os workflows para apoiar a predição de invasão. É importante observar que os sistemas de detecção de intrusão são potencialmente mais eficazes quanto mais completos forem os dados utilizados em sua tarefa de classificação. Assim sendo, é possível construir sistemas de detecção com elevadas taxas de acerto, uma vez que se tenha acesso aos dados adequados, por exemplo, dados coletados em diversos pontos de uma rede e em diversos formatos.

Um sistema que tenha acesso aos pacotes que trafegam por diversos locais da topologia de uma rede, e que cruze tais dados com logs de *hosts* e equipamentos de rede, pode ser capaz de uma taxa de acerto bastante próxima dos 100% (SPEROTTO et al., 2009). No entanto, dificilmente um sistema como o descrito é capaz de ser utilizado para a detecção em tempo real de ataques e tentativas de intrusão (SPEROTTO e MEENT, 2007).

Denominamos *detectores avançados* aos sistemas de detecção de intrusão que buscam, a partir do cruzamento de dados de diversas fontes, atingir patamares superiores de taxa de acerto. Em oposição aos detectores avançados que são eficazes, porém lentos, temos os *sistemas de detecção de tempo real*. Tais sistemas são responsáveis por indicar, a cada instante de tempo, a presença de tráfego malicioso. Sistemas de detecção de tempo real operam sobre dados consolidados (*flows*) e, muitas vezes, incompletos (*i.e.*, obtidos por meio de amostragem).

5. PROPOSTA DE WORKFLOWS PARA APOIAR A PREDIÇÃO DE INVASÃO

Nessa seção é apresentada uma proposta de sistema de detecção de intrusão de tempo real cuja característica fundamental é a utilização de três workflows de apoio, conforme ilustrado na Figura 2. A estratégia consiste em usar detectores avançados com o objetivo de monitorar a taxa de acerto do sistema. Outro aspecto a ser considerado é que os detectores avançados permitem que se gerem conjuntos de dados de ataques sempre atualizados, os quais podem ser utilizados no treinamento de novos sistemas de detecção de tempo real.

Os três workflows utilizados nessa proposta de sistema de detecção de intrusão de tempo real são descritos a seguir:

- Workflow 1 – **Construção de conjuntos de dados de ataques**: Este workflow trabalha tanto com os pacotes coletados como também na transformação dos pacotes em fluxos, visando identificar grupos. Os grupos apresentam comportamentos similares que servem para produzir amostras estratificadas para detectores avançados. Todos os agrupamentos identificados são rotulados produzindo bases marcadas.
- Workflow 2 – **Treinamento de sistema de detecção de tempo real**: Este workflow recebe um conjuntos de dados de fluxos rotulados (preferencialmente curados) por ataque e o utiliza no processo de treinamento de um novo sistema de detecção.
- Workflow 3 – **Monitoração de sistema de detecção de tempo real**: Este workflow recebe a classificação decidida pelo sistema de detecção e compara com a classificação decidida pelo detector avançado, determinando, assim, uma "taxa de acerto" do sistema de detecção. Enquanto a taxa de acerto estiver adequada, mensagens de relatórios são enviadas aos administradores. No caso de uma degeneração, mensagens de alerta passam a ser enviadas.

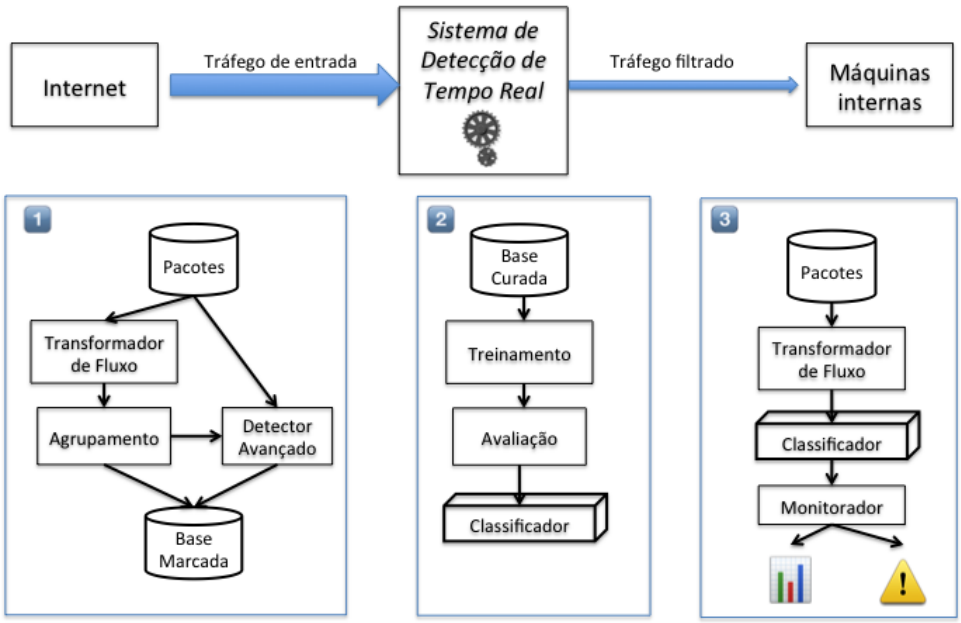

Figura 2: Estratégia de detecção juntamente com os workflows.

Nos três workflows propostos, os dados foram coletados e, em cada um deles, foi realizada um sequenciamento de atividades, ou seja, manipulação de dados, de forma a se obter resultados com grau de exatidão consideráveis.

No primeiro workflow os dados coletados foram transformados em pacotes de fluxos, criando assim grupos de comportamento similares. Desta forma, a análise é realizada em função de comportamento similares. Quando algum comportamento diverge do padrão, um alarme pode ser gerando, de modo a identificar o ataque.

O segundo workflow trabalha com dados rotulados, onde os ataques estão definidos nestes rótulos. Comparativos são realizados sinalizando os dados que se adequam ao perfil rotulado.

O terceiro workflow trabalha de forma comparativa, porém com uma classificação pré-definida pelo sistema de detecção, onde os dados podem se adequar ao perfil classificado como de risco e, assim, serem detectados.

6. CONSIDERAÇÕES FINAIS

Os sistemas de detecção de intrusão são ferramentas importantes para realizar a análise e obter respostas, em tempo real, aos ataques. O investimento na evolução dos IDS é fundamental, uma vez que qualquer demora na atuação pode comprometer todo um sistema, podendo provocar danos imensuráveis. A possibilidade de adaptação do IDS à necessidade de cada ambiente é de extrema importância, pois cada negócio possui a sua particularidade como, por exemplo, instituições financeiras que necessitam implementar sistemas robustos e de alta confiabilidade, exigindo grande poder de processamento e equipamentos de características técnicas mais avançadas.

Nesse trabalho foram propostos três workflows como estratégia para detecção de intrusão de tempo real. Considerando a avalição em tempo real, um IDS que execute os workflows propostos pode gerar resultados que apoiem a tomada de decisão e a proteção de informações, uma vez que as avaliações serão realizadas de forma imediata, de acordo com ambiente. Nesse caso, ações de prevenção poderão ser iniciadas de forma automática e imediata, mitigando os riscos de ataque a um determinado ambiente.

A partir da arquitetura proposta, acredita-se que os três workflows podem contribuir de forma positiva para a detecção de ataques cibernético, sendo a sua aplicabilidade definida de acordo com o ambiente e quantitativo de dados a ser analisado em tempo real. No entanto, deve-se considerar que, quanto maior o fluxo de dados, maior é a necessidade de equipamentos robustos e de grande capacidade de processamento para analisar e dar a resposta em um tempo mínimo necessário para impedir o ataque.

7. REFERÊNCIAS

JORGE, H. V. N.; WENDT, E. Crimes Cibernéticos - Ameaças e Procedimentos de Investigação. 2ª Ed. 2013 ed. [s.l.] Brasport, 2013.

HELLEMONS, L. et al. SSHCure: A Flow-Based SSH Intrusion Detection System. In: SADRE, R. et al. (Eds.). Dependable Networks and Services. Lecture Notes in Computer Science. [s.l.] Springer Berlin Heidelberg, 2012. p. 86–97.

NAKAMURA, E. T.; GEUS, P. L. DE. Segurança de Redes em Ambientes Cooperativos. São Paulo: Novatec, 2007.

OLUSOLA, A. A.; OLADELE, A. S.; ABOSEDE, D. O. Analysis of KDD ‘99 Intrusion Detection Dataset for Selection of Relevance FeaturesWorld Congress on Engineering and Computer Science. Anais...2010

PFAHRINGER, B. Winning the KDD99 Classification Cup: Bagged Boosting. SIGKDD Explor. Newsl., v. 1, n. 2, p. 65–66, jan. 2000.

SPEROTTO, A. et al. A Labeled Data Set for Flow-Based Intrusion DetectionProceedings of the 9th IEEE International Workshop on IP Operations and Management. Anais...: IPOM '09.Berlin, Heidelberg: Springer-Verlag, 2009Disponível em: <http://dx.doi.org/10.1007/978-3-642-04968-2_4>. Acesso em: 11 abr. 2014

SPEROTTO, A.; MEENT, R. VAN DE. A Survey of the High-Speed Self-learning Intrusion Detection Research Area. In: BANDARA, A. K.; BURGESS, M. (Eds.). Inter-Domain Management. Lecture Notes in Computer Science. [s.l.] Springer Berlin Heidelberg, 2007. p. 196–199.

VASUDEVAN, A. R.; HARSHINI, E.; SELVAKUMAR, S. SSENet-2011: A Network Intrusion Detection System dataset and its comparison with KDD CUP 99 dataset2011 Second Asian Himalayas International Conference on Internet (AH-ICI). Anais... In: 2011 SECOND ASIAN HIMALAYAS INTERNATIONAL CONFERENCE ON INTERNET (AH-ICI). nov. 2011