Modelagem de um Sistema Automatizado de Análise Forense: Arquitetura Extensível e Protótipo Inicial

Marcelo Abdalla dos Reis
Instituto de Computação
Universidade Estadual de Campinas
13083-970 Campinas - SP
marcelo.reis@ic.unicamp.br

Paulo Lício de Geus
Instituto de Computação
Universidade Estadual de Campinas
13083-970 Campinas - SP
paulo@ic.unicamp.br

RESUMO

A análise forense de um sistema computacional invadido depende totalmente da experiência do investigador. Entretanto, parte desse conhecimento pode ser transferida para um sistema automatizado através de técnicas aplicadas em detecção de intrusão. Nesse sentido, os autores deste artigo propõem um modelo de sistema automatizado capaz de identificar e correlacionar evidências de intrusões, apresentando sua arquitetura e um protótipo inicial. **Palavras-chave:** *segurança, forense computacional, coleta e correlação de evidências.*

ABSTRACT

The forensic analysis of a hacked computer system totally depends on the investigator's experience. However, part of his knowledge can be transfered to an automatic system through techniques applied in intrusion detection. In this sense, the authors of this paper propose a design model for an automatic forensic system capable of identifying and correlating intrusion evidences, presenting its architecture and an early prototype. **Keywords:** *security, computer forensics, information gathering, evidence correlation.*

1 Introdução

A análise forense de um sistema computacional invadido compreende as seguintes etapas: coleta de informações para análise (*information gathering*), busca e extração de evidências nas informações coletadas e correlação das evidências encontrados.

A etapa de coleta de informações busca reunir o máximo de dados sobre o sistema invadido sem, no entanto, alterá-los. Dentre tais informações podem ser citadas: imagens dos discos do sistema, estado das conexões de rede, estado dos processos em execução e, até mesmo, conteúdo da memória. Durante a etapa de busca e extração

de evidências, as informações coletadas são minuciosamente analisadas pelo investigador, que se baseia em seu conhecimento a respeito dos indícios que podem estar relacionados com a invasão. Por fim, o investigador correlaciona as evidências encontradas, buscando formular conclusões acerca da intrusão (causas, responsabilidades e efeitos, por exemplo).

A experiência do investigador faz-se necessária durante toda a análise forense do sistema computacional. Na etapa de *information gathering* o investigador deve decidir quais informações são relevantes e qual a maneira mais adequada de coletá-las. Durante a busca de evidências, a experiência do investigador permite identificar indícios relacionados com a intrusão dentre uma quantidade grande de informações. O correlacionamento das evidências também necessita do conhecimento do investigador a respeito dos parâmetros necessários para relacionar duas ou mais evidências.

A automatização do processo de análise forense torna-se uma necessidade à medida que a quantidade de informações armazenadas nos sitemas computacionais aumenta. Muitas vezes uma investigação completa e detalhada torna-se inviável devido à quantidade massiva de informação a ser analisada. A automatização também permite a implementação de protocolos e procedimentos devidamente testados e avaliados[1], impedindo que o investigador cometa erros que comprometam a investigação (alterações nas informações originais, por exemplo).

Entretanto, para se automatizar o processo de análise forense é necessário um mecanismo que permita transferir parte do conhecimento do investigador para um sistema automatizado capaz de coletar, identificar e correlacionar evidências. Tal mecanismo pode ser encontrado nos sistemas de detecção de intrusão (IDS). A detecção de intrusão utiliza uma série de técnicas (como, por exemplo, *threshold detection*, redes neurais, sistemas expecialistas e abordagens baseadas em transição de estados) que permitem "instruir" o IDS a reconhecer situações intrusivas [2].

A forense computacional constitui uma instância *post-mortem* de detecção de intrusão, de modo que a assinatura da invasão é representada por um conjunto de evidências correlacionadas que descreve o cenário da intrusão (vulnerabilidades exploradas, ações do invasor, *timeline* dos eventos, origem e finalidade do ataque).

Nesse sentido, um conjunto de possíveis evidências (vestígios mais prováveis de serem encontrados, como, por exemplo, alterações em arquivos sensíveis), bem como relações entre elas, podem ser definidas antecipadamente segundo experiências anteriores do investigador. Essas informações podem ser armazenadas em uma base de dados utilizada por um sistema automatizado capaz de executar as etapas de *information gathering*, busca e correlação de evidências.

Este artigo propõe uma arquitetura extensível para um sistema automatizado de análise forense capaz de coletar informações, identificar e correlacionar as evidências de uma intrusão. Um protótipo inicial, denominado AFA (Automated Forensic Analyser), foi desenvolvido implementando o *framework* da arquitetura e as etapas de *information gathering* e busca de evidências. A arquitetura do sistema é descrita na seção 2 e o protótipo inicial é apresentado na seção 3.

[1] Os autores deste artigo desenvolvem paralelamente um trabalho a respeito da padronização de procedimentos e protocolos para a forense computacional [5].

2 Uma Arquitetura Extensível

O desenvolvimento de um sistema automatizado de análise forense requer uma arquitetura que permita ao investigador configurar, de acordo com suas experiências de investigações anteriores, todo o processo de análise. Deve ser possível determinar quais informações devem ser coletadas para análise e qual a maneira mais adequada de realizar a coleta, quais as técnicas mais adequadas para a busca de cada evidência e como as evidências podem estar relacionadas. Além disso, tal arquitetura deve ser concistente com princípios básicos da forense, como, por exemplo, a não alteração do sistema analisado, a realização da imagem do sistema e análise sobre a cópia, e a utilização de ferramentas confiáveis [5].

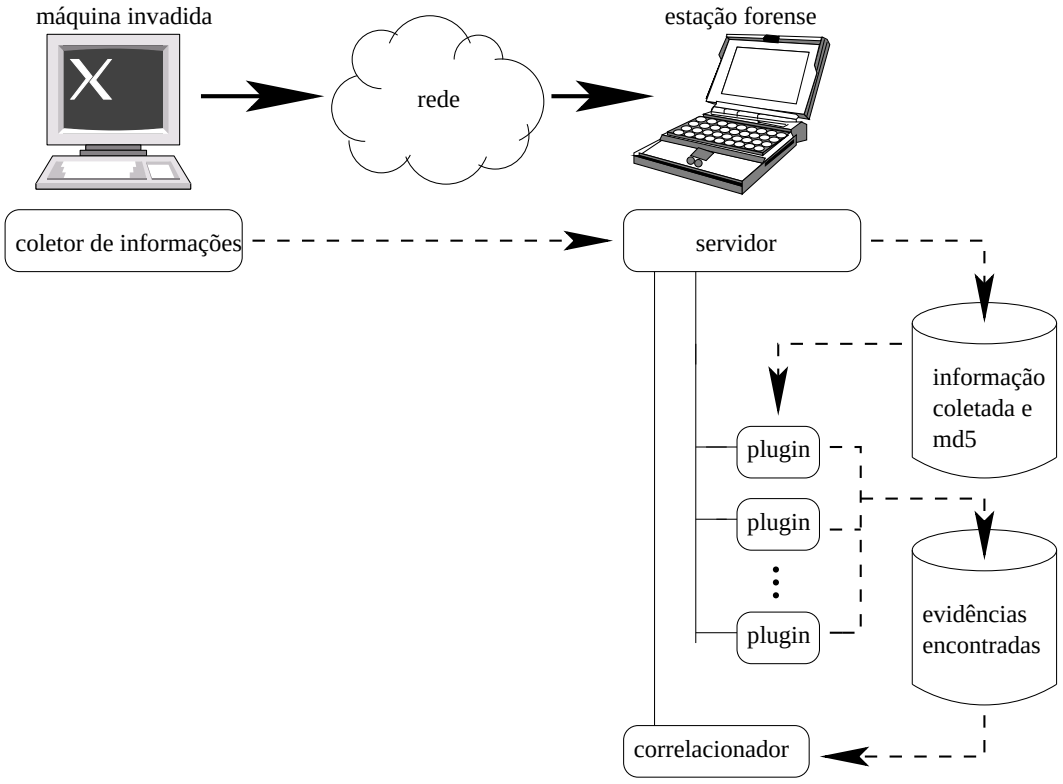

Figura 1: Arquitetura extensível para um sistema automatizado de análise forense.

A figura 1 ilustra a arquitetura proposta para o desenvolvimento de um sistema automatizado de análise forense. Tal arquitetura é composta de um servidor, executado na estação forense (com sistema e ferramentas confiáveis), e um cliente (coletor de informações) presente na máquina analisada. O servidor é o componente principal, encarregado de receber as informações provenientes da máquina analisada, organizá-las devidamente na estação forense e buscar as evidências e as correlações. A parte cliente é responsável pela coleta de informações na máquina analisada e envio delas para o servidor.

A arquitetura proposta é descrita, segundo as etapas da análise forense, como segue.

Coleta de Informações

A coleta de informações na máquina analisada é feita pelo *coletor de informações*, segundo as instruções do investigador. Esta etapa consiste na obtenção de dados para análise dentro das seguintes fontes possíveis de evidências:

- registradores e cache (difícil coletar dados sem alterações no sistema);
- memória;
- estado do sistema (conexões de rede, processos em execução, usuários logados, tabelas de rotas, estado das interfaces, data e hora do sistema, entre outras informações);
- tráfego de rede;
- dispositivos de armazenagem secundária (espaços não alocados, *file slacks* e sistema de arquivos).

O investigador deve determinar ao *coletor de informações* quais dados devem ser coletados e como eles devem ser obtidos (qual ferramenta deve ser utilizada). A imagem do disco (ou discos) da máquina analisada também é realizada nesta etapa e pode ser feita de duas maneiras. Caso a máquina analisada esteja ligada, o *coletor de informações* pode enviar a imagem pela rede ao *servidor*, caso contrário, os discos a serem copiados podem ser conectados na estação forense e duplicados através do *servidor*.

É importante observar que toda informação coletada é enviada, juntamente com seu hash criptográfico, pela rede para o servidor (utilizando opcionalmente um canal cifrado ou *crossover*), de modo que nada é escrito na máquina analisada (um dos princípios da forense [5]). Outra observação é a forma como a parte cliente é executada na máquina analisada: toda ferramenta (*coletor de informações* e programas auxiliares) e configurações necessárias para a coleta de informações devem ser acessadas a partir de uma mídia removível produzida pelo investigador. Além disso, tais ferramentas devem ser compiladas estaticamente e forensicamente testadas [5].

Busca e Extração de Evidências

A etapa de busca e extração de evidências representa uma instância *post-mortem* de detecção de intrusão. A busca de evidências de uma intrusão em um sistema invadido é equivalente à detecção de intrusão em *batch mode* e, portanto, pode utilizar as diversas técnicas de análise empregadas na detecção de intrusão [2].

Esta etapa é automatizada através da especificação de um conjunto de *plugins* que buscam evidências em locais específicos do sistema, utilizando as técnicas que melhor se adequam para a detecção dessas evidências. A cada *plugin* é especificado um conjunto de possíveis evidências que devem ser procuradas no sistema. Algumas possíveis evidências são ilustradas na figura 2 de maneira genérica.

Diante da inexistência de um método único de detecção que possa ser aplicado com total eficácia [1] e considerando a existência de várias técnicas [2], cada qual com suas vantagens, a arquitetura proposta apresenta características extensíveis, permitindo incorporar as diversas soluções existentes e suportar sem grandes esforços a incorporação de novos métodos.

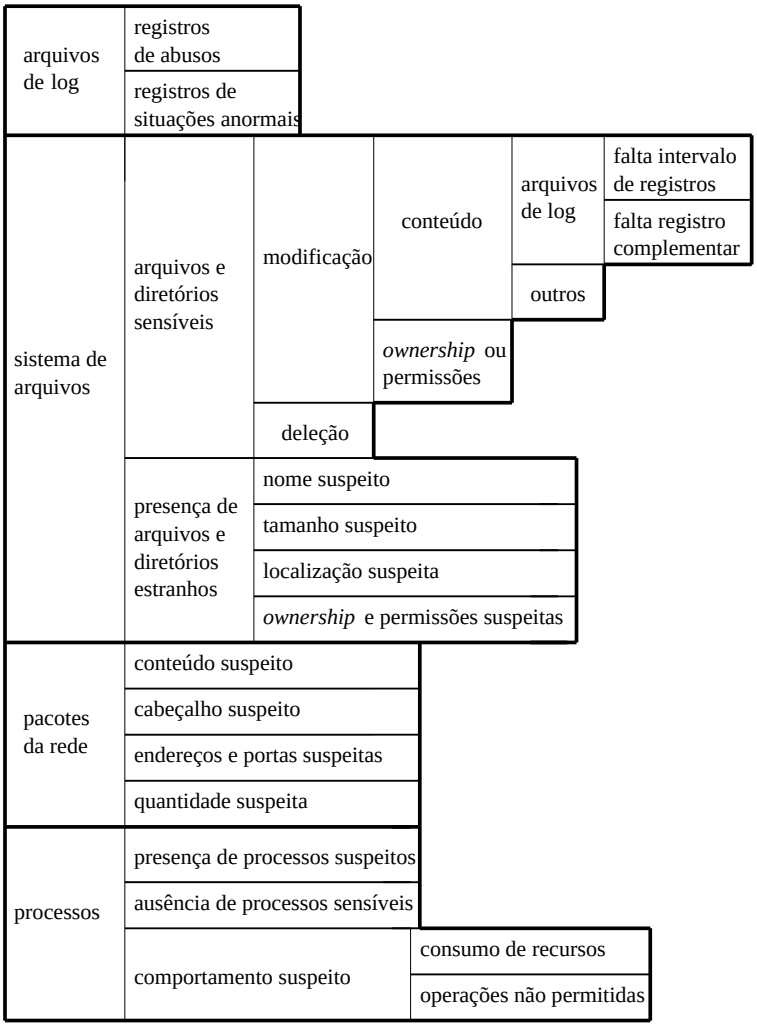

Figura 2: Conjunto de possíveis evidências a serem procuradas.

Correlacionamento de Evidências

As evidências encontradas na etapa anterior são armazenadas e descritas de modo a conter toda informação necessária para um possível correlacionamento (data e hora de um registro de log, *owner* e hora de inicialização de um processo, origem de uma conexão, por exemplo).

Com base em alguns parâmetros, como, por exemplo, usuário, endereço origem ou intervalo de tempo, as evidências encontradas podem ser correlacionadas. Dentre as técnicas que podem ser empregadas no correlacionamento de evidências podem ser citadas: *Bayesian reasoning* [4] e sistemas expecialistas (if-then-else) [2].

3 Um Protótipo Inicial

Foi desenvolvido um protótipo inicial que implementa o *framework* geral da arquitetura proposta e as etapas de *information gathering* e busca de evidências. O protótipo, denominado AFA (*Automated Forensic Analyser*), foi implementado em *Perl* e limita-se à análise do ambiente *Linux*.

A etapa de *information gathering* encontra-se totalmente implementada e a etapa de busca de evidências conta apenas com um *plugin* desenvolvido, responsável pela análise dos processos em execução no sistema analisado. Apesar de seu estágio embrionário, o AFA permite colocar em prática a arquitetura proposta neste trabalho, de modo a consolidar os conceitos e entender a fundo questões que só aparecem durante a implementação, como detalhes de configuração, interface e funcionamento.

Nenhum teste quantitativo foi realizado, no sentido de avaliar a eficiência e precisão do sistema, devido ao estágio inicial em que se encontra o protótipo. Entretanto, as características principais da arquitetura, quais sejam, *framework*, capacidade de extensão e configuração, e concistência com os princípios da forense, puderam ser avaliados qualitativamente de maneira satisfatória.

4 Conclusões

A experiência do investigador sempre se fará necessária em algum momento da análise forense. No entanto, o processo de identificação e correlacionamento de evidências pode ser automatizado, através do armazenamento do conhecimento adquirido pelo investigador na base de dados do sistema proposto.

A arquitetura proposta neste artigo baseia-se no conhecimento prévio de um conjunto de informações que podem representar evidências de uma intrusão, bem como dos possíveis relacionamentos existentes entre tais informações. Desse modo, o investigador deve especificar quais informações devem ser coletadas na máquina analisada e qual a maneira correta de fazer a coleta; quais dados podem representar indícios da intrusão e qual a melhor técnica para detectá-los; e como duas ou mais evidências podem ser relacionadas (segundo determinados parâmetros).

A automatização do processo de análise forense permite uma maior rapidez na formulação de conclusões acerca de uma intrusão e viabiliza análises onde a quantidade de informação a ser examinada é muito grande. Além disso, impede que o investigador cometa erros que comprometam a investigação, através da implementação de protocolos e procedimentos devidamente testados e avaliados.

A arquitetura proposta apresenta características importantes para a automatização do processo de análise forense: é facilmente extensível e configurável, permitindo a incorporação de novas técnicas para a detecção de evidências; e implementa os princípios básicos da forense.

5 Trabalhos Futuros

Como continuidade ao trabalho apresentado neste artigo, os autores pretendem:

- Refinar a arquitetura proposta neste artigo;
- Derivar uma classificação de evidências, permitindo definir o escopo de atuação dos *plugins*;
- Implementar outros *plugins* para análise, por exemplo, de arquivos de log;
- Definir uma meta-linguagem de representação das evidências encontradas;

- Especificar os parâmetros para correlacionamento das evidências;
- Definir a abordagem a ser utilizada para a correlação de evidências;
- Efetuar testes com exemplos de análises de sistemas invadidos (como, por exemplo, a análise do *Honeynet Project's Forensic Challenge* [3])
- Desenvolver um módulo adicional para converter as informações resultantes da análise forense em uma assinatura do ataque. Tal assinatura pode ser fornecida a um sistema de detecção de intrusão por mau uso. A arquitetura proposta neste artigo, em conjunto com o referido módulo adicional, pode ser aplicado na geração automática de assinaturas de ataques, utilizada no sistema de detecção de intrusão baseado no sistema imunológico proposto em [6];

Referências

[1] Ross Anderson e Abida Khattak. The use of information retrieval techniques for intrusion detection. Em *First Workshop on the Recent Advances in Intrusion Detection*, Louvain-la-Neuve, Belgium, Setembro de 1998.

[2] Rebecca G. Bace. *Intrusion Detection*. Macmillan Technical Publishing, Indianapolis, IN, 2000.

[3] The honeynet forensic challenge, 2000. Disponível *online* em agosto de 2001 na URL http://project.honeynet.org/challenge/.

[4] T. Goan. A cop on the beat: Collecting and appraising intrusion evidence. *Communications of ACM*, Julho de 1999.

[5] Marcelo A. Reis e Paulo L. Geus. Forense computacional: Procedimentos e padrões. Em *Anais do SSI'2001: 3º Simpósio Segurança em Informática*, São José dos Campos, SP, 2001.

[6] Marcelo A. Reis, Fabrício S. Paula, Diego M. Fernandes, e Paulo L. Geus. A hybrid ids architecture based on the immune system. Em *A ser publicado no Wseg2002: Workshop em Segurança de Sistemas Computacionais*, Búzios, RJ, 2002.