Análise de Desempenho de Computadores de Baixo Custo em um Sistema de Detecção de Intrusão

Aluno: Lucas Seiki Oshiro,
Orientador: Daniel Macêdo Batista

20 de abril de 2019

1 Introdução

Com a obtenção de quantidades massivas de dados de tráfego e de aplicações de rede, surge a necessidade de métodos mais "inteligentes" para buscar incidentes de segurança, principalmente porque passa a ser possível encontrar informações novas por meio da correlação das informações e também porque uma análise de força-bruta levaria muito tempo para ser finalizada. Essa necessidade já vem sendo discutida há alguns anos tanto na academia [Brown et al. 2015] quanto na indústria, que já fornece diversos serviços para clientes dos mais diversos tamanhos [Splunk 2019]. Entretanto, as implementações dos métodos não seguem uma arquitetura que seja eficiente para todos os tipos de organizações. De fato, tem-se notado que cada vez mais as arquiteturas precisam ser otimizadas e personalizadas para cada tipo de usuário [Feth 2015]. Além disso, é desejável que o sistema a ser desenvolvido seja capaz de antecipar, ao máximo possível, incidentes em tempo real e que o mesmo não tenha um custo elevado tanto em termos financeiros para adquirir equipamentos, quanto em termos de consumo de energia. Uma forma de antecipar incidentes é com a utilização de mecanismos baseados em aprendizado de máquina. Uma forma de reduzir os custos é com a utilização de clusters baseados em hardware de baixo custo.

No escopo do projeto **GT-BIS – Mecanismos para Análise de Big Data em Segurança da Informação** [GT-BIS 2018], foi desenvolvido um

sistema capaz de detectar ataques a partir da análise de grandes volumes de logs de servidores web e de servidores de banco de dados, por meio de técnicas de aprendizado de máquina. No momento é necessário melhorar o sistema com a adição da detecção de novos tipos de incidentes de segurança, como por exemplo aqueles causados por ataques automatizados que usam ofuscação e que burlam mecanismos mais tradicionais de proteção.

Este trabalho de conclusão do curso é uma continuação de um trabalho de iniciação científica[1] do ano 2018, em que foram obtidos resultados parciais com a análise de diferentes ataques contra aplicações web e com a análise de desempenho de uma unidade Raspberry Pi 3 Model B em cenários de uso intensivo de CPU. Como resultado foi possível identificar algumas características dos ataques que podem ser úteis para o treinamento de um modelo de detecção de ameaças. Já os resultados da análise de desempenho da Raspberry Pi mostraram que ela apresenta boa vazão de tráfego de rede e temperatura dentro de limites seguros mesmo quando submetida a altas cargas de processamento.

2 Conceitos Básicos

Os números na figura descrevem os dados que são passados entre cada componente do sistema: **1)** logs de serviços e de sistemas de segurança. **2)** dados normalizados, filtrados e enriquecidos. **3)** fluxos de dados organizados em filas para serem processados. **4)** informações brutas e de dados já processados (p. ex. alertas); comandos de consultas. **5)** dados para serem persistidos ou recuperados. **6)** alertas de ameaças cibernéticas compartilhados pelo sistema central ou parceiros. **7)** dados da interação do administrador com a interface Web. **8)** informações disponíveis nos componentes de Processamento (alertas, dados brutos, dados processados); comandos de consultas.

`Logstash`, `Kafka`, `Spark` e `ElasticSearch`, apresentadas na Figura 1, são ferramentas escaláveis e paralelizáveis de software livre para, respectivamente, filtrar e normalizar dados de logs, implementar um sistema produtor/consumidor para acesso a filas de dados, processar dados empregando por exemplo mecanismos de aprendizado de máquina e armazenar e visualizar dados. Todas essas ferramentas podem ser replicadas em diferentes nós em *clusters* a fim de tornar o funcionamento escalável em função da carga

[1] Trabalho feito pelo mesmo aluno e orientador

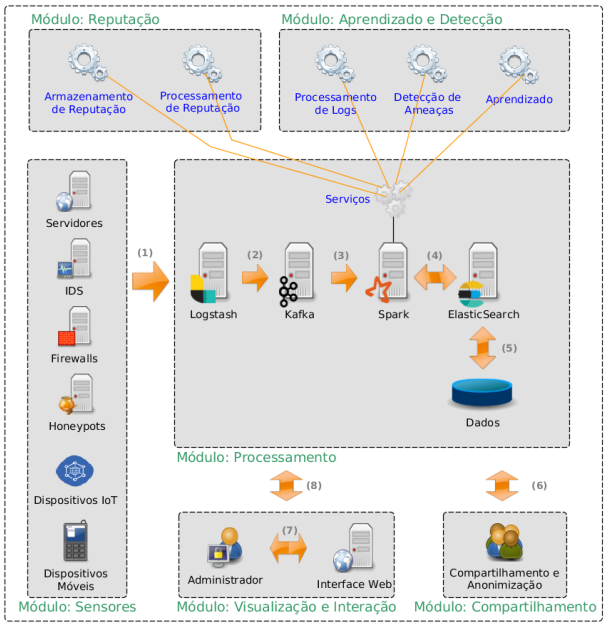

Figura 1: Arquitetura do sistema para detecção de ameaças por meio da análise de logs [GT-BIS 2018]

de dados de entrada e é nesse sentido que propomos a utilização de unidades de Raspberry Pi em cada um desses *clusters*.

2.1 Ataques Contra Aplicações Web

Ataques contra redes de computadores ocorrem por conta de vulnerabilidades em diversas camadas da arquitetura Internet. Especificamente no caso de aplicações web, aquelas projetadas para serem acessadas principalmente por meio de um navegador web tendo o protocolo HTTP como base, falhas podem levar à negação do serviço, ao acesso indevido a informações ou mesmo à execução de comandos arbitrários do lado do servidor.

Pelo fato do protocolo HTTP ser um protocolo sem estado, boa parte das aplicações web requer a utilização de algum Sistema Gerenciador de Banco de Dados (SGBD). Tais sistemas costumam ser alvos de atacantes em busca de falhas que possam afetar a infraestrutura de uma aplicação web. Outro alvos para a exploração de falhas é a construção de URLs maliciosas.Alguns exemplos de ataques desses tipos são: **Injeção de SQL** – inserção de dados

que, se não tratados, realizam ações indevidas no banco de dados; **XSS Refletido** – inserção de código HTML e JavaScript malicioso em uma URL, que é executado quando um site é aberto por meio daquela URL; **XSS Persistido** – envio de código HTML e JavaScript malicioso, que fica armazenado e é executado no navegador de todos os usuários que acessarem o site.

A distinção de ataques como os listados acima no meio de acessos legítimos pode ser dificultada caso o atacante utilize técnicas de ofuscação facilitadas por ferramentas que automatizam a exploração desses ataques.

2.2 Plataformas de Computação de Baixo Custo

Muitas aplicações para Internet das Coisas [Singh and Kapoor 2017] dependem de placas e sistemas embarcados que funcionem tanto como sensores, obtendo informações do mundo real, quanto como processadores, analisando os dados capturados e enviando comandos de atuação. A necessidade de implantação desses elementos em larga escala justifica a utilização de hardware de baixo custo. Dentre as várias opções de hardware para esse objetivo destacam-se o Arduino [Arduino 2019] e a Raspberry Pi [Raspberry Pi Foundation 2018]. O Arduino é uma plataforma de código aberto para computação física baseada em entradas e saídas simples [Banzi 2011], enquanto que a Raspberry Pi, embora também seja um dispositivo em placa única, é um computador e possui maior poder de processamento que o Arduino. O grande sucesso na utilização desses dispositivos em aplicações de Internet das Coisas tem levado à sua utilização em outros domínios de aplicação. No caso da Raspberry Pi, vários projetos tem utilizado diversas unidades conectadas via rede criando assim um *cluster* de baixo custo [Pahl et al. 2016]. No Brasil, uma unidade Arduino UNO pode ser adquirida por cerca de R$50,00, enquanto uma unidade Raspberry Pi 3 Model B pode ser adquirida por cerca de R$200,00.

3 Cronograma e principais atividades

- Abril
  - Montar o ambiente de medição de consumo de energia;
  - Preparação da apresentação do artigo que fala dos avanços obtidos até o momento (A ser apresentado no Workshop de Trabalhos de

Iniciação Científica e de Graduação do SBRC)

- Maio
  - Participação do Workshop de Trabalhos de Iniciação Científica e de Graduação do SBRC;
  - Realização de experimentos para medição do consumo de energia com as aplicações sintéticas (vcgenmod/iperf);
  - Realização de experimentos para análise de desempenho com aplicações reais (logstash);
- Junho
  - Realização de experimentos para análise de desempenho com aplicações reais (Kafka, Spark e ElasticSearch);
- Julho
  - Integração da unidade Raspberry Pi com todo o sistema de detecção de intrusão;
- Agosto
  - Realização de experimentos para análise de precisão na detecção de ataques;
- Setembro
  - Realização de experimentos para análise de precisão na detecção de ataques;
- Outubro
  - Escrita da monografia;
- Novembro
  - Preparação do poster para exibição e dos slides para apresentação oral;

Referências

[Arduino 2019] Arduino (2019). Arduino - Home. `https://www.arduino.cc/`. Último acesso em 22 de Março de 2019.

[Banzi 2011] Banzi, M. (2011). *Getting Started with Arduino*. Make: projects. O'Reilly Media.

[Brown et al. 2015] Brown, S., Gommers, J., and Serrano, O. (2015). From Cyber Security Information Sharing to Threat Management. In *Proceedings of the 2Nd ACM WISCS'15*, pages 43–49.

[Feth 2015] Feth, D. (2015). User-centric Security: Optimization of the Security-usability Trade-off. In *Proceedings of the 10th ESEC/FSE 2015*, pages 1034–1037.

[GT-BIS 2018] GT-BIS (2018). GT-BIS – Mecanismos para Análise de Big Data em Segurança da Informação. `http://gtbis.ime.usp.br/`. Último acesso em 22 de Março de 2019.

[Pahl et al. 2016] Pahl, C., Helmer, S., Miori, L., Sanin, J., and Lee, B. (2016). A Container-Based Edge Cloud PaaS Architecture Based on Raspberry Pi Clusters. In *4th IEEE FiCloudW*, pages 117–124.

[Raspberry Pi Foundation 2018] Raspberry Pi Foundation (2018). Raspberry Pi – Teach, Learn, and Make with Raspberry Pi. `https://www.raspberrypi.org/`. Último acesso em 22 de Março de 2019.

[Singh and Kapoor 2017] Singh, K. J. and Kapoor, D. S. (2017). Create Your Own Internet of Things: A survey of IoT Platforms. *IEEE Consumer Electronics Magazine*, 6(2):57–68.

[Splunk 2019] Splunk (2019). SIEM, AIOps, Application Management, Log Management, Machine Learning, and Compliance — Splunk. `https://www.splunk.com/en_us`. Último acesso em 22 de Março de 2019.