Detecção de Anomalias em Contêiner: Uma Avaliação Considerando o Nível de Observação das Aplicações no Contêiner

Anderson Aparecido do Carmo Frasão
Departamento de Informática
Universidade Federal do Paraná
Curitiba, PR, Brasil

Sumário

Introdução

- Os contêineres representam uma revolução na implementação de aplicativos.
- Algumas vantagens dos contêineres em comparação com máquinas virtuais:
  - Maior desempenho e escalabilidade
  - Uso eficiente de recursos computacionais
  - Menos sobrecarga e maior número de instâncias
- Riscos (Imagem, Contêiner, Sistema Operacional do Host, ...)
- Mitigações (Atualizações regulares, Monitoramento constante, Escolha de imagens confiáveis, ...)

Introdução

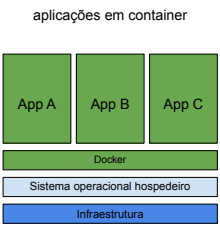
- Algumas soluções para Detecção de Intrusão em Contêineres:
  - Integridade de Instâncias de Contêiner (Amazon)
  - Análise de Comportamento (docker)
  - Sistemas de Prevenção de Intrusão (IPS)
- Esse estudo se concentra em:
  - Gerar um sistema de detecção de intrusão.
  - Comparar diferentes pontos de coletas de dados.

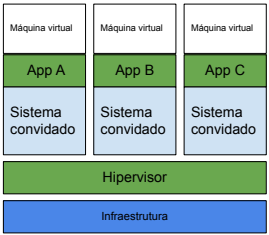
Sistema de detecção de intrusão

- Ferramentas para detectar comportamentos intrusivos
- Análise de dados de auditoria
- Prevenir futuros ataques
- Categorias de Detecção de Intrusão
  - Detecção baseada em assinatura (SD)
    - Comparação com padrões conhecidos
    - Identificação de ameaças e comportamentos suspeitos
  - Detecção baseada em anomalia (AD)
    - Comparação de comportamentos com perfis normais
    - Identificação de comportamentos não usuais ou suspeitos
- Sistemas de Detecção de Intrusão de Rede (NIDS) e Host (HIDS)

Virtualização Baseada em Contêiner

- Aproveita recursos do kernel para criar ambientes isolados para processos.
- Diferença estrutural em relação à virtualização baseada em hipervisor.





Trabalhos Relacionados

Trabalho	Estratégia	Objeto alvo
(Abed et al., 2015b)	BoSC com janela deslizante em epochs	Contêiner
(Castanhel et al., 2021)	Strace	Contêiner Docker
(Du et al., 2018)	cAdvisor, Heapster, InfluxDB e Grafana	Contêiner Kubernetes
(Rocha et al., 2022)	Sysdig	Contêiner Kubernetes

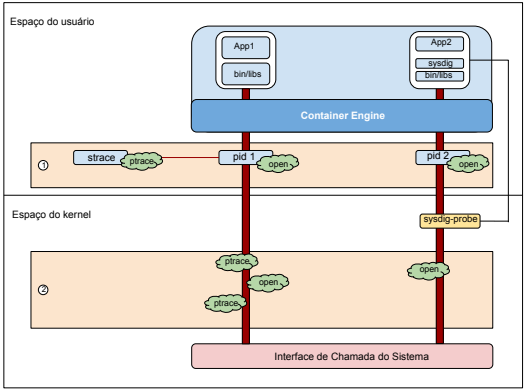
Tabela: Trabalhos Relacionados

Proposta

- Qual a viabilidade de considerar o isolamento do contêiner e diferentes níveis de observação do ambiente?
- É possível criar uma nova estratégia para a detecção de anomalias em um ambiente de contêiner?
- Requisitos:
  - Se faz necessária uma base representativa de dados para detecção e comparação

Proposta

- O Sysdig foi a ferramenta escolhida para comparação com o Strace
- A estruturas dessas ferramentas são descritas a seguir:
  1. Dados gerados antes da coleta de dados.
  2. Resultados da coleta de dados após a interação com a interface de chamada do sistema.
- Destacando a sobrecarga causada pela troca de contexto do strace durante a coleta de dados.



Proposta

- Este estudo foca em uma aplicação executando em um contêiner
  - Aplicação alvo: Wordpress
- A parte de anomalias do nosso conjunto de dados sysdig apresenta as seguintes vulnerabilidades:
  - Injeção arbitrária de código
  - Falha na validação de extensões de ficheiros, permitindo o upload e execução de arquivos PHP.
  - Injeção de SQL.
  - Download remoto de arquivos...
- Dataset formado por uma sequências de system calls

Proposta

- Conjunto de dados reunido com strace:
  - Coleta estruturada, 10 comportamentos, 5 execuções cada
  - Totalizando 50 arquivos de log
- Conjunto de dados reunido com sysdig:
  - Coleta estruturada, 20 comportamentos, 10 execuções cada
  - Totalizando 200 arquivos de log

Proposta

- Algoritmos de Machine Learning utilizados:
  - Ada Boost (AB)
  - Decision Tree (DT)
  - Multilayer Perceptron (MLP)
  - Nu-Support Vector (NuSV)
  - Random Forest (RF)
  - Stochastic Gradient Descent (SGD)
  - XGBoost (XGB)
- Dois experimentos foram definidos:
  - Analise do impacto que as diferentes perspectivas de observação têm nos dados recolhidos por duas soluções heterogêneas de rastreio de syscall (strace e sysdig).
  - Solução de detecção de anomalias possibilitada pelo sysdig.

Analise do impacto de diferentes perspectivas de observação

- Diferença de Chamadas de Sistema de interações normais entre Sysdig (esquerda) e Strace (direita)

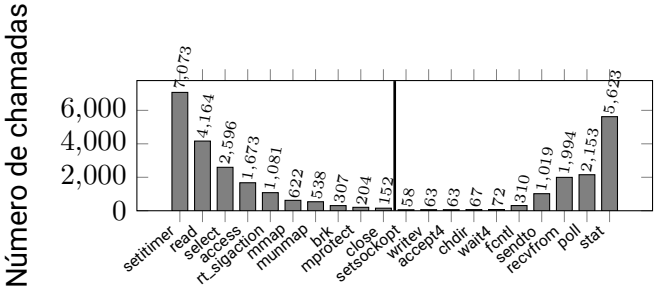


Figura: Sysdig | Strace

Analise do impacto de diferentes perspectivas de observação

- Diferença de Chamadas de Sistema de interações anormais entre Sysdig (esquerda) e Strace (direita)

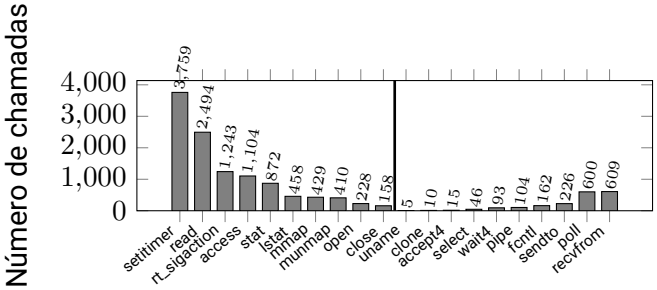
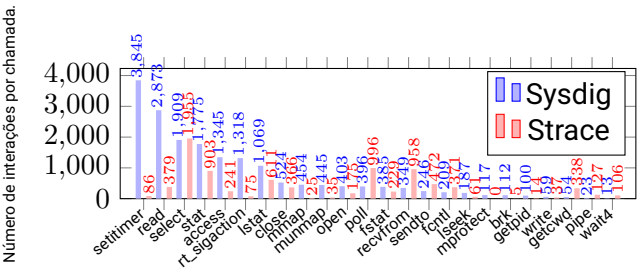


Figura: Sysdig | Strace

Analise do impacto de diferentes perspectivas de observação

- Chamadas de sistema com destaque em interações mal-intencionadas.
- Variação nos conjuntos específicos de chamadas entre as duas perspectivas de observação.
- Complicação no processo de identificação de anomalias.



Solução de detecção de anomalias possibilitada pelo sysdig

- Treinamento com 50% dos dados disponíveis e testes com os 50% restantes.
- A curva Receiver Operating Characteristic (ROC) indica que três dos classificadores alcançam resultados satisfatórios

Classificador	***Precision***	***Recall***	**F1**	**ROC**
AB	83.67%	83.67%	83.67%	95.84%
MLP	93.33%	85.71%	89.36%	95.88%
NuSV	86.00%	87.76%	86.87%	94.14%
RF	80.39%	83.63%	82.00%	90.84%

Solução de detecção de anomalias possibilitada pelo sysdig

- Em relação ao F1 Score, o MLP foi o melhor classificador e apenas um modelo ficou abaixo do limiar de 80%.
- Apesar das margens de erro nos modelos, os resultados indicam o potencial da proposta para detecção de anomalias, com possíveis aplicações em contextos específicos.

Conclusão

- Nesse trabalho levantamos os seguintes pontos:
  - Qual a viabilidade de considerar o isolamento do contêiner e diferentes níveis de observação do ambiente?
  - É possível criar uma nova estratégia para a detecção de anomalias em um ambiente de contêiner?

Conclusão

- O sysdig destaca-se por um aumento significativo em operações temporais, leitura, gestão de processos e manipulação de dispositivos
- Os dados obtidos do strace revelam maior ênfase em chamadas de sistema relacionadas à comunicação e manipulação do sistema operativo
- É viável monitorar o comportamento de contêineres por meio de análise de system calls utilizando o Sysdig.
- O classificador Multi-Layer Perceptron se destacou para possível implementação.

Obrigado