Bifocal Agent: Identificando automaticamente funções maliciosas para aumentar o foco do analista de malware

Leonardo Gonçalves Chahud,
Rafael Rocha,
Prof. Dr. Lourenço Pereira Junior,
Prof. Dr. Idilio Drago

Instituto Tecnológico de Aeronáutica
Università di Torino

Agenda

1. Motivação
2. Objetivos
3. Escopo de Trabalho
4. Trabalhos Relacionados
5. Bifocal Agent
6. Resultados
7. Conclusão
8. Limitações
9. Trabalhos Futuros

Motivação

Motivação

NEWS 28 NOV 2023

Ukraine Police Dismantle Major Ransomware Group

Malware-as-a-Service Now the Top Threat to Organizations ...

Infosecurity Magazine › news › malware-service-top-threat

NEWS 27 NOV 2023

SysJoker Malware: Hamas-Related Threat Expands With Rust Variant

CPR said the malware now uses OneDrive instead of Google Drive for storing dynamic C2 server URLs

Objetivos

Objetivos

Auxiliar o **analista de malware**

Classificar **funcionalidades**

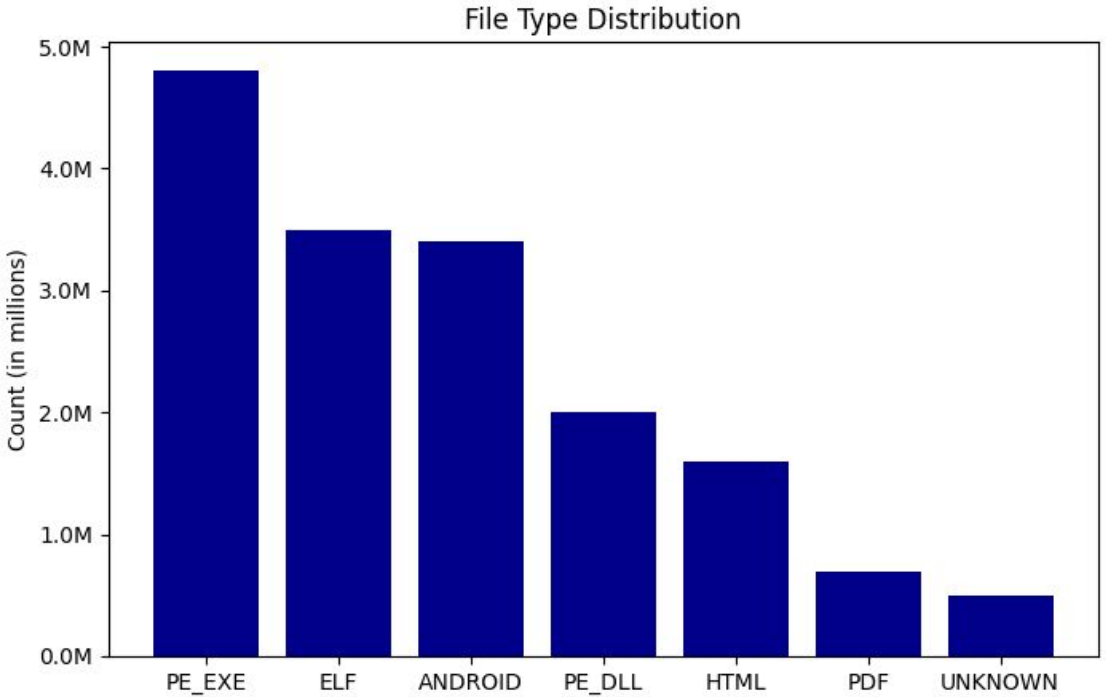
Análise por **decomposição**

Escopo de Trabalho

Escopo de Trabalho

Por que Windows?

Por que formato PE?

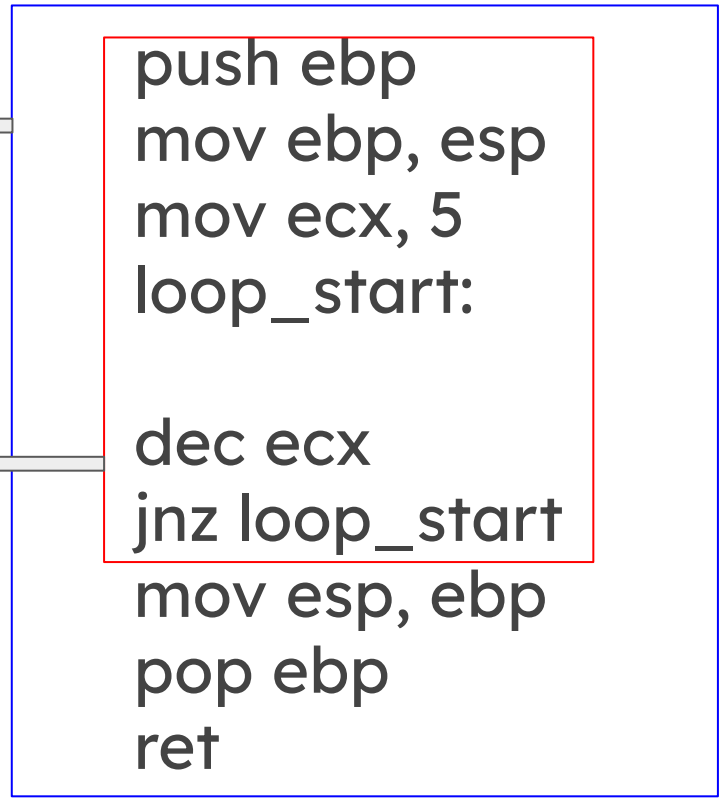


Escopo de Trabalho

Funções

Blocos básicos

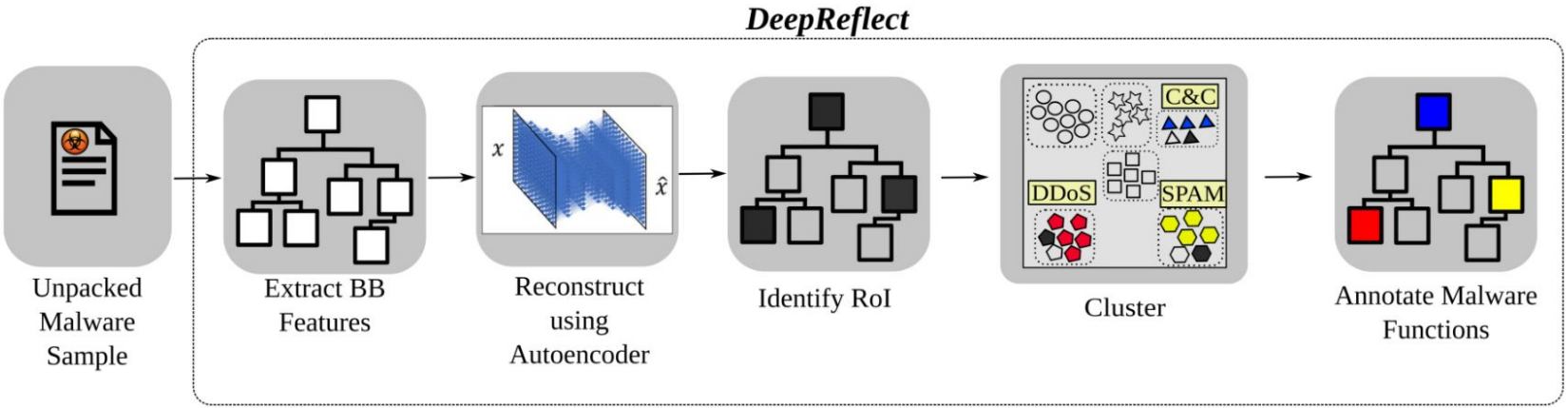
- Incluem instruções **call**



Trabalhos Relacionados

Trabalhos Relacionados

Arquitetura do DeepReflect



Trabalhos Relacionados

Atributos do DeepReflect - Blocos Básicos

offspring
betweenness
arith_bit_shift
arith_basic_math
arith_logic_ops
trans_stack
trans_reg
trans_port

api_dll
api_file
api_network
api_object
api_process
api_registry
api_service
api_sync
api_sysinfo
api_time

Trabalhos Relacionados

Amostras Ground-Truth

- Rbot (2004)
- Pegasus (2016)
- Carbanak (2014)

Ferramentas baseline

- VGG19 model + SHAP (deep learning comparison)
- CAPA (FireEye)
- FunctionSimSearch (Google Project Zero)

Trabalhos Relacionados - Limitações

Tabela 1. Tabela de funcionalidades presentes em cada solução.

Características	DeepReflect	Jarv1s	CodeAnalyzer	BifocalAgent
Aprendizado de Máquina	X		X	X
Blocos básicos	X			X
Funções		X	X	X
Atributos de API	X		X	X
Densidade API				X
Densidade de Leitura				X
Densidade de Escrita				X

Bifocal Agent

Bifocal Agent

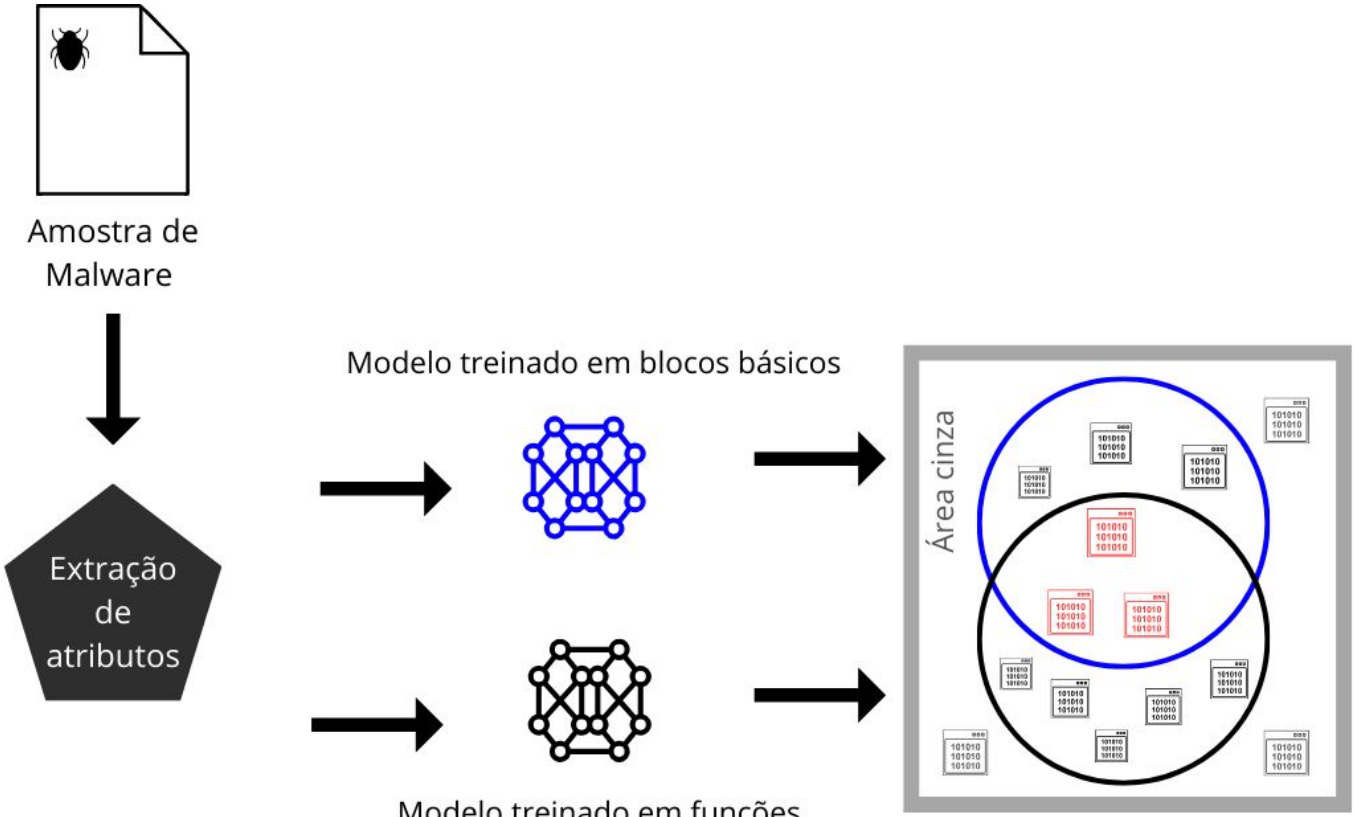
Considera tanto blocos básicos como funções

Utilização de área cinza

Vetores de atributos modificados

Pré-processamento dos dados

Bifocal Agent - Arquitetura



Bifocal Agent - Novos Atributos

+instruction_count = contagem **total** de instruções

+api_calls = contagem **total** de chamadas

+api_read = contagem de chamadas de **leitura**

+api_write = contagem de chamadas de **escrita**

+api_density = chamadas / instruções

Bifocal Agent - Novos Atributos

+api_read_density = chamadas de leitura / chamadas

+api_write_density = chamadas de escrita / chamadas

Como separar chamadas de leitura e escrita?

Bifocal Agent - Novos Atributos

Utilização de expressão regular

RegQueryValueExA

ReadFile

GetVolumeInformationA

SetFileAttributesW

WriteFileEx

UpdateResource

Bifocal Agent - Dataset

Amostras de Treino

- 201,549 amostras benignas
- 20,000 amostras selecionadas aleatoriamente
- 13,868 amostras benignas filtradas após pré-processamento

Amostras Ground-Truth

- Rbot (2004)
- Pegasus (2016)
- Carbanak (2014)

Bifocal Agent - Pré Processamento

Granularidade a nível de funções

Exclusão de vetores praticamente vazios

Seleção de vetores até o percentil 95 para remover amostras discrepantes (instruction-count)

Normalização dos atributos pelo valor máximo (DeepReflect)

Bifocal Agent - Pré Processamento

Granularidade a nível de blocos básicos

Exclusão de vetores praticamente vazios

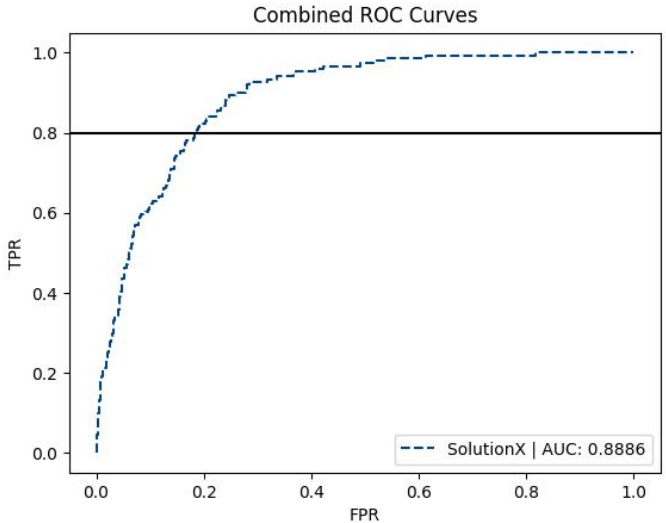
Assimetria positiva -> Limites de Tukey sob contagem de instruções

Limites de Tukey + Cálculo de entropia -> Seleção de vetores preenchidos

Normalização pelo valor máximo (DeepReflect)

Bifocal Agent - Thresholds

- Mesmo processo do DeepReflect
- Curva ROC + TPR = 80% para o RBOT



Resultados

Bifocal Agent - Resultados

Áreas sobre as curvas ROC

Solução denota o conjunto de atributos

Solução	Dataset	Granularidade	Rbot	Pegasus	Carbanak	Combinado
BifocalAgent	BifocalAgent	Função	0.7242	0.7677	0.8423	0.8886
BifocalAgent	BifocalAgent	Bloco básico	0.8385	0.6109	0.6756	0.7989
DeepReflect	BifocalAgent	Função	0.6359	0.6459	0.8027	0.7120
DeepReflect	BifocalAgent	Bloco Básico	0.6818	0.7349	0.8314	0.7612
DeepReflect	DeepReflect	Bloco Básico	0.8429	0.7926	0.7634	0.8319

Bifocal Agent - Novos Atributos

Melhoras de 24,8% e 5% na área combinada através da engenharia de atributos.

Solução	Dataset	Granularidade	Rbot	Pegasus	Carbanak	Combinado
BifocalAgent	BifocalAgent	Função	0.7242	0.7677	0.8423	0.8886
BifocalAgent	BifocalAgent	Bloco básico	0.8385	0.6109	0.6756	0.7989
DeepReflect	BifocalAgent	Função	0.6359	0.6459	0.8027	0.7120
DeepReflect	BifocalAgent	Bloco Básico	0.6818	0.7349	0.8314	0.7612
DeepReflect	DeepReflect	Bloco Básico	0.8429	0.7926	0.7634	0.8319

Bifocal Agent - Granularidade

Melhora de 11,2% para o **novo** vetor de atributos
Piora de 7,5% com o vetor de atributos **antigo**

Solução	Dataset	Granularidade	Rbot	Pegasus	Carbanak	Combinado
BifocalAgent	BifocalAgent	Função	0.7242	0.7677	0.8423	0.8886
BifocalAgent	BifocalAgent	Bloco básico	0.8385	0.6109	0.6756	0.7989
DeepReflect	BifocalAgent	Função	0.6359	0.6459	0.8027	0.7120
DeepReflect	BifocalAgent	Bloco Básico	0.6818	0.7349	0.8314	0.7612
DeepReflect	DeepReflect	Bloco Básico	0.8429	0.7926	0.7634	0.8319

Bifocal Agent - Atributos + Granularidade

Melhora de 17% na área combinada

Solução	Dataset	Granularidade	Rbot	Pegasus	Carbanak	Combinado
BifocalAgent	BifocalAgent	Função	0.7242	0.7677	0.8423	0.8886
BifocalAgent	BifocalAgent	Bloco básico	0.8385	0.6109	0.6756	0.7989
DeepReflect	BifocalAgent	Função	0.6359	0.6459	0.8027	0.7120
DeepReflect	BifocalAgent	Bloco Básico	0.6818	0.7349	0.8314	0.7612
DeepReflect	DeepReflect	Bloco Básico	0.8429	0.7926	0.7634	0.8319

Bifocal Agent - Precisão

Melhora de 5,0% apenas mudando de granularidade

Granularidade	Funções	Blocos Básicos	Granularidade Múltipla
Precision	0.926	0.882	0.968
Recall	0.815	0.815	0.663
F1-Score	0.867	0.847	0.787
Acertos	89	85	67
Erros	23	27	5
Amostras Cinzas	0	0	40

Bifocal Agent - Precisão

Melhoras de 4,5% e 9,7% na precisão

Granularidade	Funções	Blocos Básicos	Granularidade Múltipla
Precision	0.926	0.882	0.968
Recall	0.815	0.815	0.663
F1-Score	0.867	0.847	0.787
Acertos	89	85	67
Erros	23	27	5
Amostras Cinzas	0	0	40

Conclusão

Conclusão

Trabalhar a nível de funções gera melhor desempenho

Novos atributos representam melhor as funcionalidades

Granularidade múltipla é preferível para maior precisão

Limitações

Bifocal Agent - Limitações

Amostras que passaram por **packing**

Ataques adversariais

Dataset de amostras Ground-Truth **pequeno**

Desbalanceamento do dataset em relação às chamadas de API

Trabalhos Futuros

Trabalhos futuros

- Estudar performance em amostras com packing
- Expandir dataset de avaliação
- Testar thresholds diferentes
- Testar heurísticas diferentes com os dois modelos

Obrigado!

Autores

- Leonardo Gonçalves Chahud
- Rafael Rocha
- Prof. Dr. Lourenço Alves Pereira Junior
- Prof. Dr. Idilio Drago

Contato

- leonardo.chahud.res@gmail.com