Detecção Hierárquica Confiável de Malware de Android Baseado em Arquiteturas CNN

Jhonatan Geremias, Eduardo Kugler Viegas, Altair Olivo Santin, **Pedro Horchulhack**, Alceu de Souza Britto Jr

*Programa de Pós-Graduação em Informática – Pontifícia Universidade Católica do Paraná, Brasil

{jgeremias, santin, pedro.horchulhack, alceu}@ppgia.pucpr.br

SBSeg 2024

Simpósio Brasileiro em Segurança da Informação (SBSeg)

PUCPR

SecPLab

Disclosing your future needs for Security & Privacy

Agenda

- ***Introdução***
- Fundamentação Teórica
- Declaração do Problema
- Proposta
- Avaliação
- Conclusão

PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

Introdução

Contextualização

- Sistema operacional móvel Android
  - 2.5 bilhões de dispositivos ativos
  - 70% do total de mercado de smartphones
- Aplicativos Android maliciosos
  - O número de aplicativos maliciosos está aumentando
  - Afetando 24% de todos os usuários do Android
  - Mais de 67% de todos os aplicativos de malware foram originados de mercados de aplicativos oficiais.

Introdução

Contextualização

- Técnicas baseadas em análise dinâmica
  - Avaliação do comportamento do aplicativo Android realizado em tempo de execução
  - O aplicativo é executado em ambiente de *sandbox*
  - Malwares modernos são capazes de detectar quando estão sendo monitorados por uma *sandbox*, permanecendo indetectáveis
- Técnicas baseadas em análise estática
  - Avaliação das características do aplicativo de maneira *offline*
  - Não requer a execução do aplicativo Android analisado.

Introdução

Contextualização

- Técnicas de *Deep learning* para classificação de aplicativos maliciosos
  - Os arquivos fonte compilados em Java (*dex*) são representados como uma imagem para a tarefa de classificação.
    - O arquivo binário .dex é traduzido para um formato de imagem
    - Cada byte do arquivo .dex é representado na imagem como pixel
  - A imagem é classificada por uma rede neural convolucional (*CNN*)
    - O dataset composto por amostras de malware e aplicativos benignos
    - O modelo construído é utilizado em produção para a identificação de novos aplicativos Android maliciosos.

PUCPR

SecPLab

Disclosing your future needs for Security & Privacy

Introdução

Limitação da abordagens tradicionais

- Abordagens atuais do estado da arte
  - Os esquemas propostos muitas vezes não são confiáveis para configurações do mundo real
  - Os modelos construídos avaliam a classificação entre diversas famílias de malware
    - Ignoram a identificação inicial da amostra do aplicativo ser maliciosa
    - Não existe garantia de que as taxas de reportadas serão alcançadas em ambientes reais.
  - Os datasets utilizados na avaliação em geral são irrealistas
    - Coletados de uma única fonte (problema de generalização)

PUCPR

SecPLab

Disclosing your future needs for Security & Privacy

Introdução

Objetivo

- Desenvolver um modelo de CNN hierárquico e confiável para a classificação de malwares em Android
  - Uma CNN baseada em imagens em um ambiente de classificação local hierárquico
    - Classificados como amostras benignas ou maliciosas em um nó pai
    - Família dos aplicativos maliciosos classificados no nível anterior
  - Aumentar a confiabilidade da classificação
    - Abordagem de Classificação com rejeição
  - Apenas amostras de malware com alta confiança devem ter sua família identificada, mantendo a acurácia do sistema.

Introdução

Contribuições

- Um novo dataset de malware Android
  - Composto por mais de 29 mil amostras de aplicativos benignos e maliciosos
  - Dividido em 29 famílias
- Avaliação das abordagens CNN para tarefa de classificação de malware Android
  - Demostrar a falta de confiabilidade para fornecer altas taxas de acurácia quando uma dataset mais desafiador é utilizado
- Um novo modelo CNN baseado em imagem hierárquico e confiável para detecção de malware Android
  - Um modelo capaz de melhorar a precisão da detecção quando comparado a trabalhos relacionados.

PUCPR

SecPLab

Disclosing your future needs for Security & Privacy

Agenda

- Introdução
- ***Fundamentação Teórica***
- Declaração do Problema
- Proposta
- Avaliação
- Conclusão

PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

Fundamentação Teórica

Deteção de Malware Android

- A tarefa de detecção baseado em análise estática pode ser represetada por quatro módulos sequenciais:
  - Aquisição de dados: extrai o arquivo .dex do arquivo .apk do aplicativo Android analisado
  - Construtor de Imagem: representa o conteúdo do arquivo .dex como uma imagem, representando cada *byte* como um pixel de imagem
  - Classificação: recebe as imagens construídas e aplica um modelo CNN previamente treinado
  - Alerta: sinaliza as amostras classificadas como malware.

Fundamentação teórica

Detecção de malware Android com Deep Learning

- Na literatura vários trabalhos propõem o uso de CNNs para a classificação de malware em Android
  - Reportam o desempenho com altas taxas de acurácia
  - As abordagens classificam amostras de aplicativos analisadas de acordo com sua família de malware
  - Negligenciam a identificação inicial das amostras de malware
- Consequentemente, mesmo que consigam fornecer esquemas baseados em CNN altamente precisos
  - Limita-se apenas na classificação de amostras de aplicativos que já são conhecidas como malware
  - Deixando o desempenho de detecção entre malware e benigno ainda por ser conhecido.

Agenda

- Introdução
- Fundamentação Teórica
- ***Declaração do Problema***
- Proposta
- Avaliação
- Conclusão

PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

Declaração do Problema

Dataset de malware Android de imagem realista

- Dataset para classificação de amostras de aplicativos Android baseada em imagens:
  - Os datasets utilizados na literatura não refletem características realistas de ambientes do mundo real
- Problema inerente à sua fonte de dados:
  - Os dataset devem ser compostos por um número representativo de malware e amostras benignas
  - Coletado de uma variedade de fontes ao longo do tempo

Declaração do Problema

Detecção de malware em Android por classificação de imagens

- Detecção de aplicativos Android maliciosos por meio de CNNs baseadas em imagens ainda está em estágios iniciais.
- As técnicas tradicionais baseadas em CNN não conseguem identificar de forma confiável amostras de aplicativos Android maliciosos
- A classificação de famílias de malware Android com CNNs baseadas em imagens não oferece confiabilidade

Agenda

- Introdução
- Fundamentação Teórica
- Declaração do Problema
- ***Proposta***
- Avaliação
- Conclusão

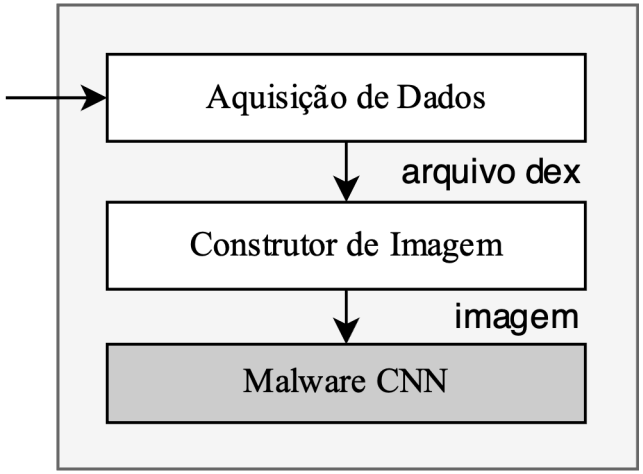
PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

Proposta

Construtor de Imagem

- O processo de classificação inicia com um arquivo apk Android como entrada para o módulo Aquisição de Dados
  - O arquivo dex é extraído do arquivo apk
  - No módulo Construtor de Imagem o arquivo dex é representado em formato de imagem
- A imagem construída é classificada como benigno ou malware por um módulo Malware CNN.



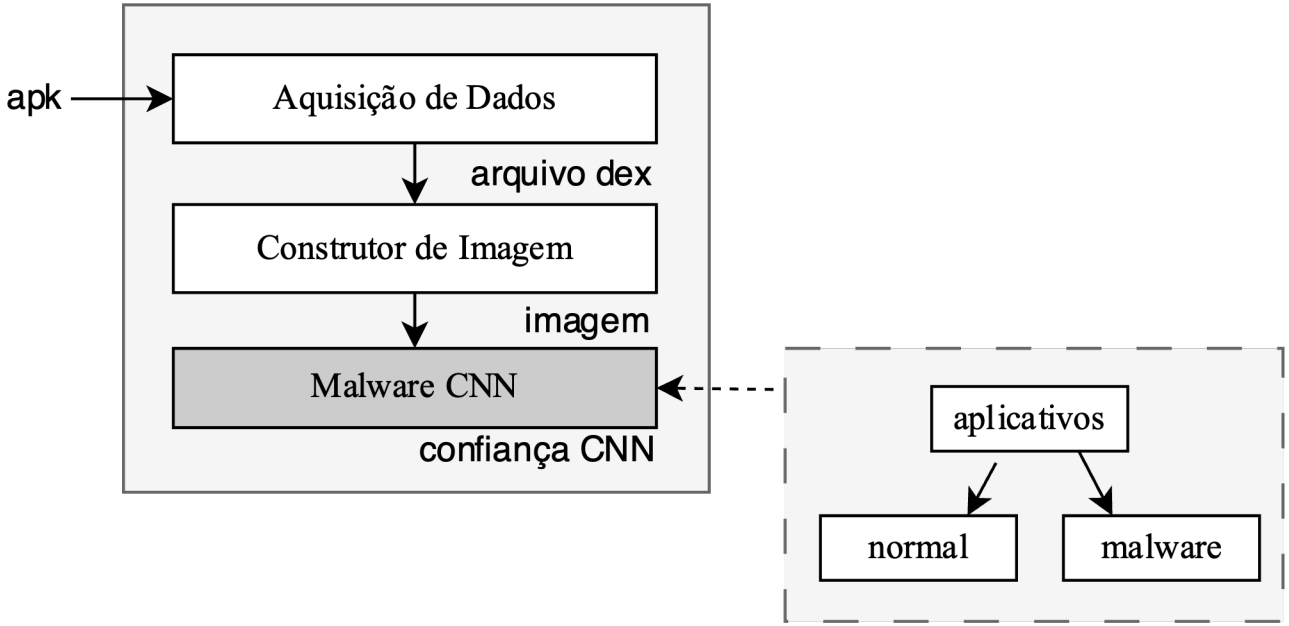
PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

Proposta

Classificação Aplicativo Android

- O modelo proposto realiza a classificação de aplicativos Android considerando uma estrutura de classificação local hierárquica.



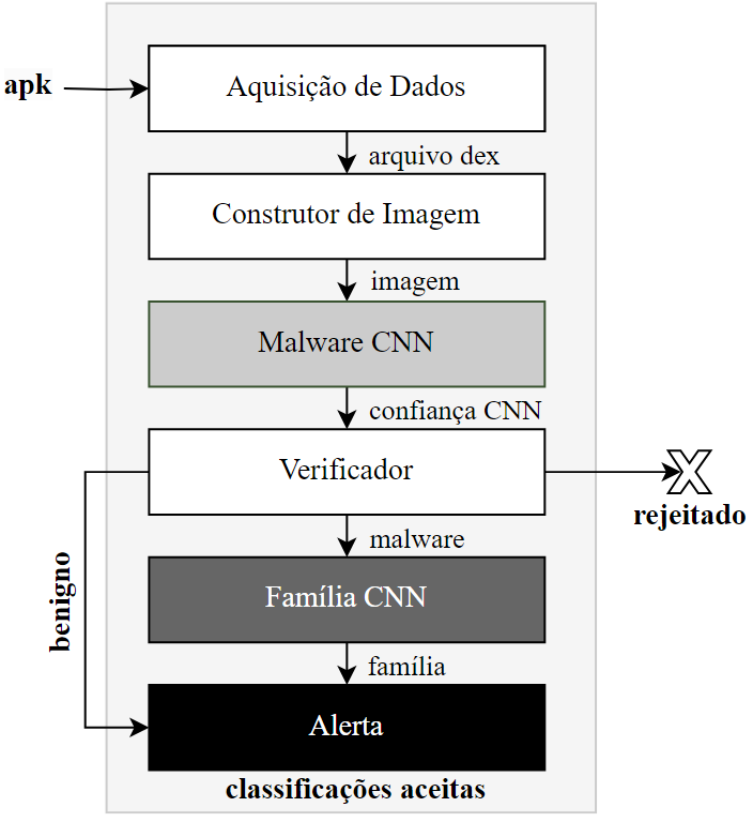
PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

Proposta

Classificação Confiável de Família de Malware

- O módulo verificador avalia o valor de confiança da classificação
  - Garantir que o nível desejado de limite de confiança de classificação seja atendido



PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

Agenda

- Introdução
- Fundamentação Teórica
- Declaração do Problema
- Proposta
- ***Avaliação***
- Conclusão

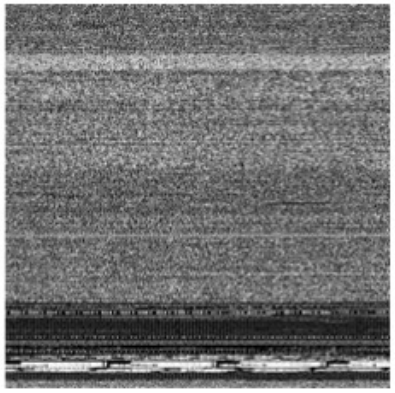
PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

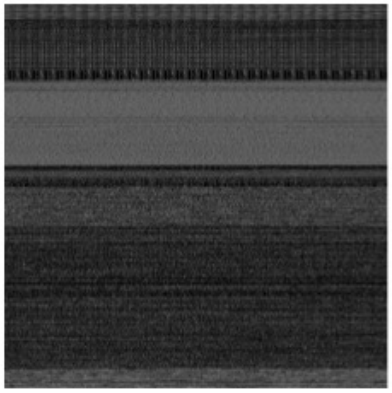
Avaliação

Image-based Android Malware Dataset (IAMD)

- Para nossa avaliação construímos o dataset IAMD
  - Composto por 26.799 amostras de aplicativos de Android
    - Divididas em 6.815 amostras benignas e 19,984 maliciosas
    - Distribuidas em 29 famílias
  - Os arquivos apk Android coletados são representados como uma imagem em nosso dataset
    - O arquivo dex é extraído do arquivo do aplicativo apk
    - Convertido em um arquivo de imagem em tons de cinza
    - O dataset é composto por mais de 13,4 GB de imagens



normal



malware

PUCPR

SecPLab

Disclosing your future needs for Security & Privacy

Avaliação

Construção do Modelos

- O modelo proposto foi implementado fazendo uso da arquitetura Resnet50
  - A arquitetura foi implementada em Keras e TensorFlow
- O dataset foi dividido em três partes: treinamento, validação e teste (40%, 30% e 30%, respectivamente)
  - As imagens foram geradas por meio da API *Python Image Library* (PIL)
- As dimensões das imagens geradas variam de acordo com o tamanho do arquivo dex
  - Redimensionadas para 224x224 antes de serem utilizadas nas arquiteturas de CNN.

Avaliação

Questões de Pesquisa

- Como as arquiteturas CNN tradicionais classificam amostras de malware e benignos entre aplicativos Android?

Arquitetura CNN	TP (%)	TN (%)	F1	AUC
Resnet50	92,95	92,77	0,929	0,97

- As abordagens avaliadas não conseguiram fornecer precisões de classificação significativamente altas
- O ResNet apresentou uma AUC de 0,97, com uma taxa de verdadeiros positivos (TP) de apenas 92,95% e uma taxa de verdadeiros negativos (TN) de apenas 92,77%
- Experimentos com abordagens de detecção amplamente utilizadas mostraram que as técnicas atuais não alcançam altas taxas em termos de acurácia

PUCPR

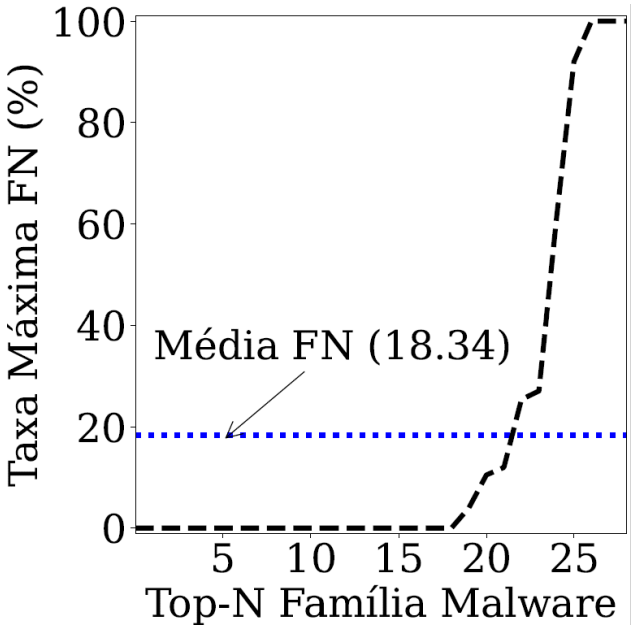
SecPLab
Disclosing your future needs for Security & Privacy

Avaliação

Questões de Pesquisa

- Como as arquiteturas CNN tradicionais realizam a classificação famílias de malware?

- A arquitetura selecionada não consegue fornecer alta precisão para a identificação da família de malware
- Em média o Resnet50 apresentou uma taxa de FN de apenas 18.34%
- As abordagens de identificação de famílias de malware falham em alcançar a confiabilidade desejada, afetando significativamente sua acurácia



PUCPR

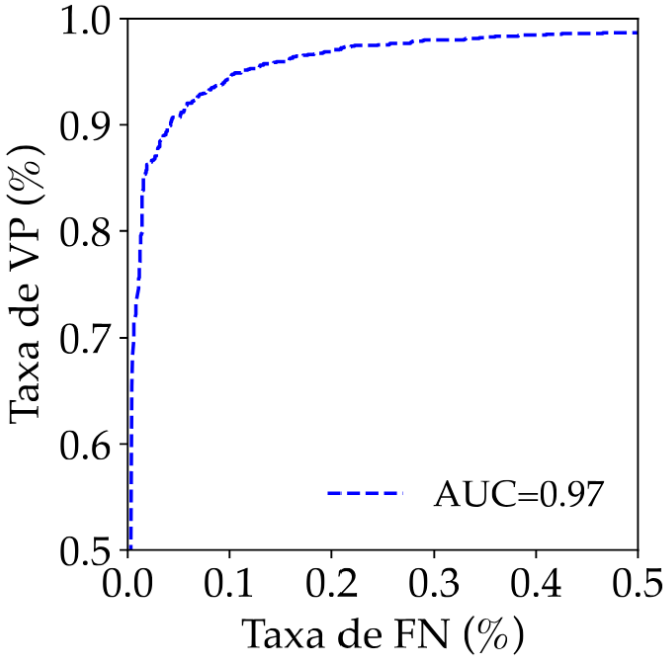
SecPLab
Disclosing your future needs for Security & Privacy

Avaliação

Questões de Pesquisa

- Como a abordagem da avaliação na classificação melhora o modelo de classificação de malware?

- A figura mostra a curva ROC da arquitetura Resnet50
- Avaliação do desempenho do modelo proposto utilizando o módulo Verificador para avaliação das classificações realizadas em ambiente de duas classes



PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

Avaliação

Questões de Pesquisa

- Como a abordagem da avaliação na classificação melhora o modelo de classificação de malware?

CNN	TP (%)	TN (%)	F1	Rejection (%)
Resnet50	92.95	92.77	0.929	0.00
	96.25	98.27	0.972	5.00
	96.54	98.30	0.970	10.00
	97.00	98.33	0.977	15.00

- Rejeitando apenas 10,0% das instâncias, nosso modelo melhora a taxa de TP em 3,59% para Resnet50
- O módulo verificador é capaz de melhorar a confiabilidade do sistema na detecção de malware

PUCPR

SecPLab

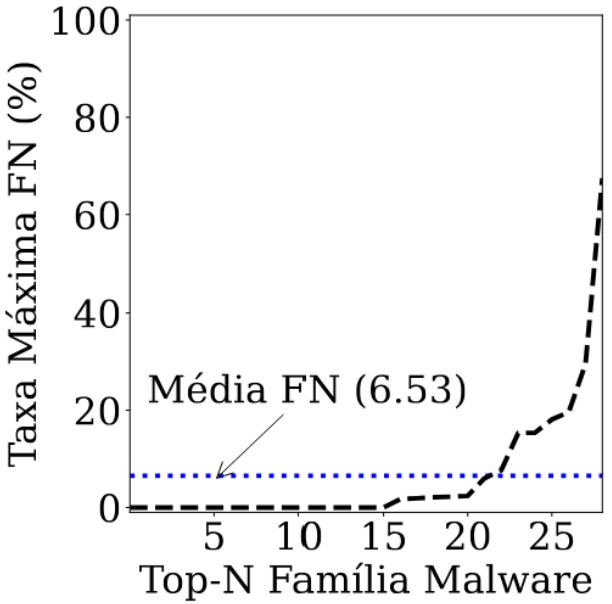
Disclosing your future needs for Security & Privacy

Avaliação

Questões de Pesquisa

- Como o modelo proposto se comporta na classificação de famílias de malware?

- O esquema proposto melhora significativamente a classificação das famílias de malware aceitas
- Rejeitando apenas 10% dos aplicativos avaliados, a taxa média de TP na identificação de famílias de malware melhora em 12,75% para Resnet50 com uma taxa máxima de 6,53% de FN



PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

Agenda

- Introdução
- Fundamentação Teórica
- Declaração do Problema
- Proposta
- Avaliação
- ***Conclusão***

PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

Conclusão

- A detecção de malware em Android usando CNNs para classificação de imagens está ainda em seus estágios iniciais
- Limitações das abordagens atuais do estado da arte
  - Os datasets utilizados na literatura são incapazes de fornecer características realistas de ambientes do mundo real
  - Não há garantia de que as taxas de acurácia relatadas serão alcançadas em ambientes reais
- Propusemos uma nova abordagem de detecção de malware em Android, confiável e hierárquica, usando CNNs para identificar aplicativos maliciosos e suas famílias
  - Capaz de melhorar a detecção de aplicativos com malware e rejeitar apenas um pequeno subconjunto de amostras de aplicativos
  - Melhora a taxa média de detecção das famílias de malware aceitas quando comparadas às técnicas tradicionais

PUCPR

SecPLab
Disclosing your future needs for Security & Privacy

Detecção Hierárquica Confiável de Malware de Android Baseado em Arquiteturas CNN

Perguntas?

Jhonatan Geremias, Eduardo Kugler Viegas, Altair Olivo Santin, **Pedro Horchulhack**, Alceu de Souza Britto Jr

*Programa de Pós-Graduação em Informática – Pontifícia Universidade Católica do Paraná, Brasil

{jgeremias, santin, pedro.horchulhack, alceu}@ppgia.pucpr.br

SBSeg 2024

Simpósio Brasileiro em Segurança da Informação (SBSeg)

PUCPR

SecPLab

Disclosing your future needs for Security & Privacy