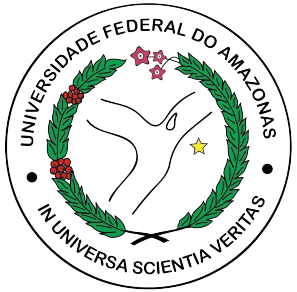
AMGenerator e AMExplorer: Geração de Metadados e Construção de Datasets Android



UFAM

Universidade Federal do Pampa

Vanderson Rocha
Joner Assolin
Hendrio Bragança
Diego Kreutz
Eduardo Feitosa

Universidade Federal do Amazonas (UFAM)
Universidade Federal do Pampa (UNIPAMPA)

Problemas em *Datasets*

Dataset	Tempo	# Maliciosos	Threshold	Metadados
MalGenome	2010 - 2012	1.264		Não
DREBIN	2013	5.560	2	Não
PiggyBacking	2016	1.136	1	Não
AMD	2010 - 2016	24.533	28	Não

Problemas em *Datasets*

Dataset	Tempo	# Maliciosos	Threshold	Metadados
MalGenome	2010 - 2012	1.264		Não
DREBIN	2013	5.560	2	Não
PiggyBacking	2016	1.136	1	Não
AMD	2010 - 2016	24.533	28	Não

Tamanho e Cobertura

Problemas em *Datasets*

Dataset	Tempo	# Maliciosos	Threshold	Metadados
MalGenome	2010 - 2012	1.264		Não
DREBIN	2013	5.560	2	Não
PiggyBacking	2016	1.136	1	Não
AMD	2010 - 2016	24.533	28	Não

Ausência de Metadados

Exempos: Descrição, Pacote, API de Compilação, etc.

Problemas em *Datasets*

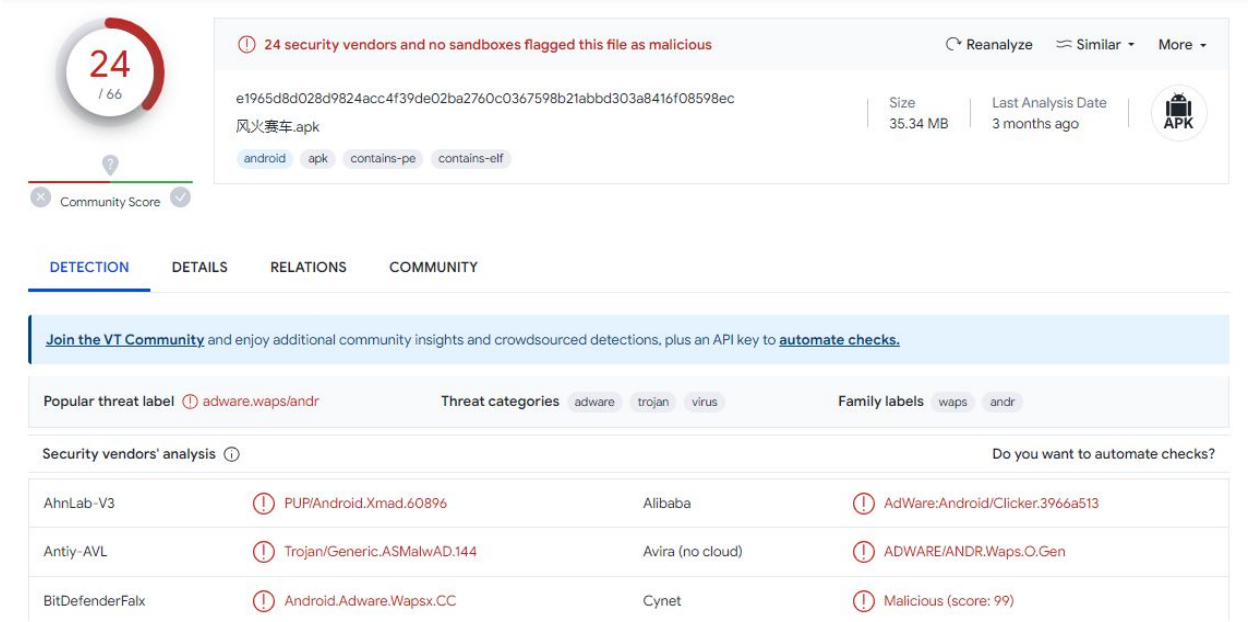
Dataset	Tempo	# Maliciosos	Threshold	Metadados
MalGenome	2010 - 2012	1.264		Não
DREBIN	2013	5.560	2	Não
PiggyBacking	2016	1.136	1	Não
AMD	2010 - 2016	24.533	28	Não

Threshold de Detecção

Varia Arbitrariamente

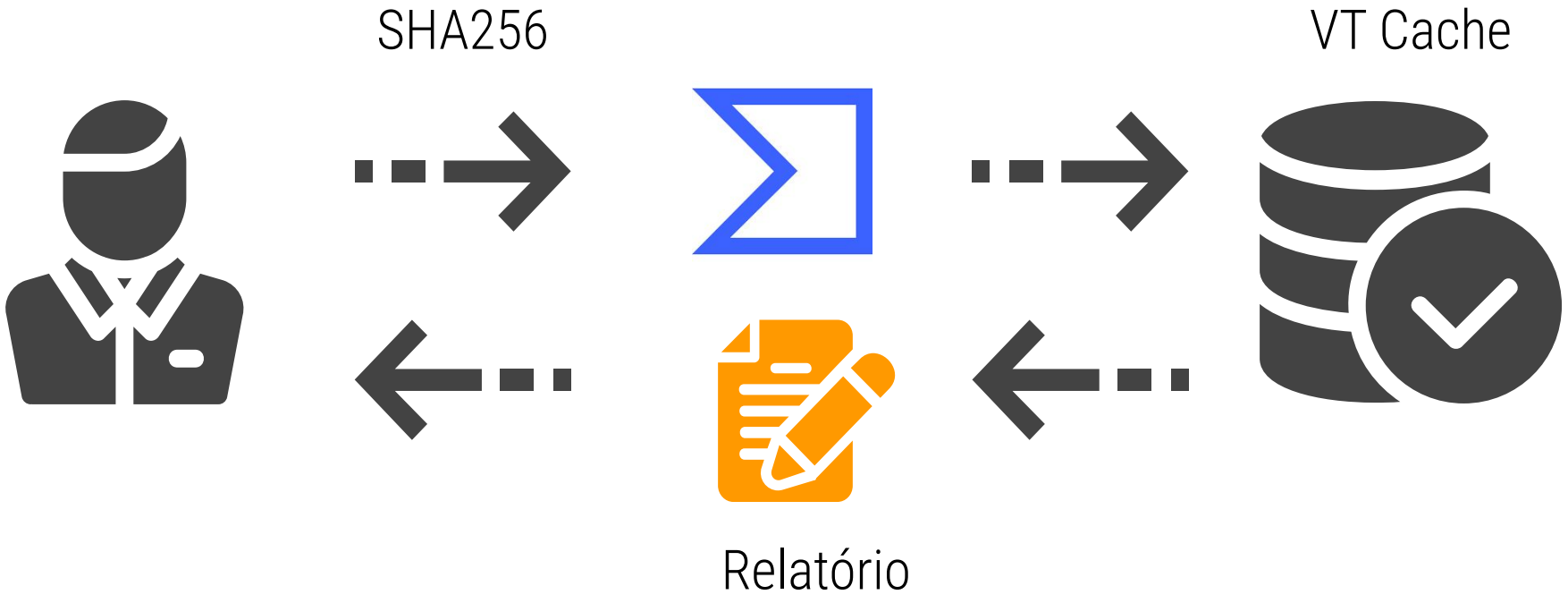
Rotulação Com VirusTotal

VIRUSTOTAL

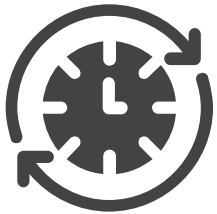


Getting Started

Rotulação Com VirusTotal



Rotulação Com VirusTotal



Dados Defasados

Aplicativo: **Corrida de Fogo de Vento**

E1965D8D028D9824ACC4F39DE02BA2760C0367598B21ABBD303A8416F08598EC

- 27 de Maio de 2014: 1 *Scanner*
- 09 de Junho de 2023: 11 *Scanners*
- 14 de Setembro de 2023: 24 *Scanners*

Rotulação Com VirusTotal

Interpretação do Resultados

- Não Rotula os Aplicativos
- ~ 60 Scanners
- Rotulação Depende da Interpretação do Usuário

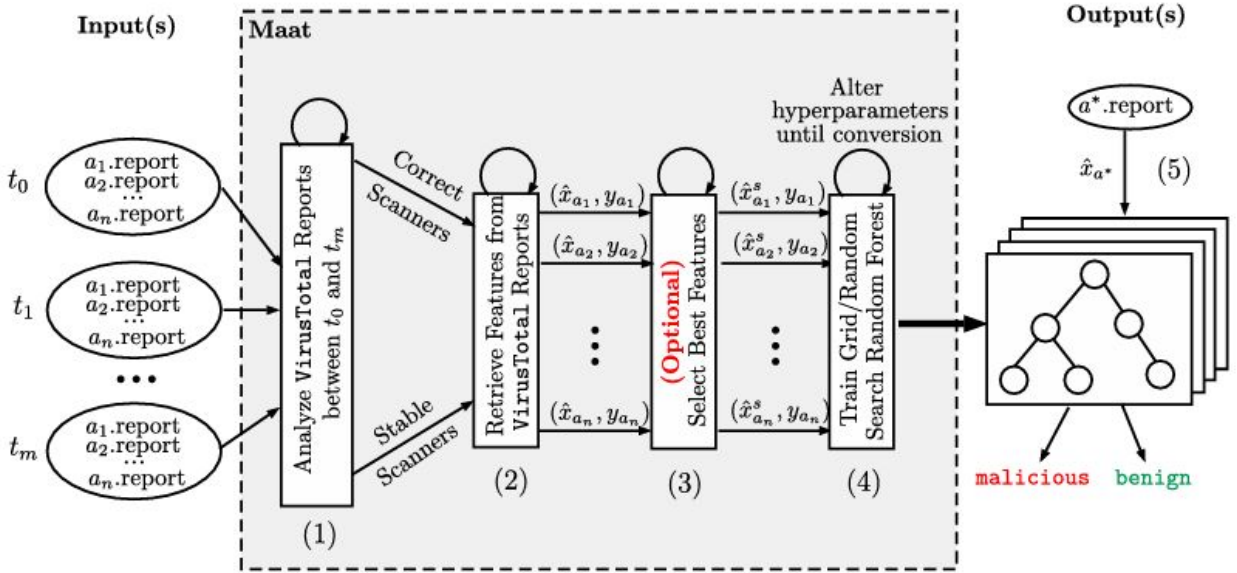
Rotulação Com VirusTotal

Interpretação do Resultados

Paper					
Threshold	1	1	2	4	5

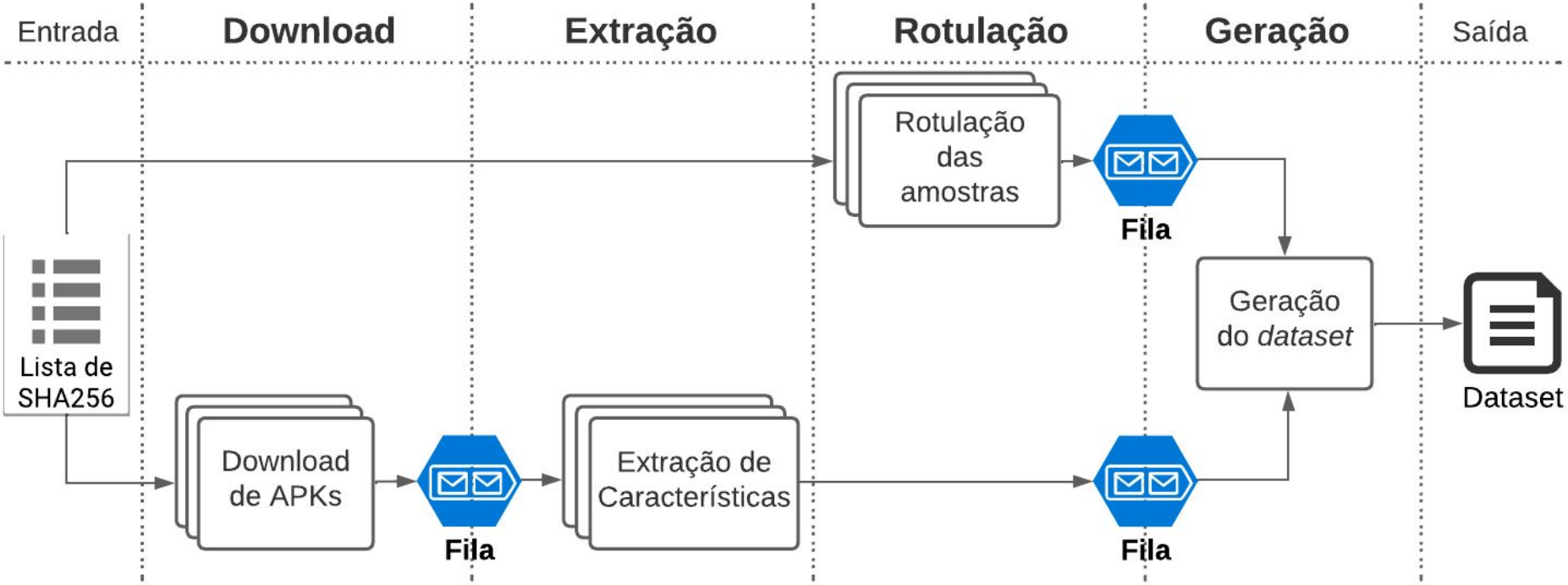
Rotulação Com VirusTotal

Interpretação do Resultados

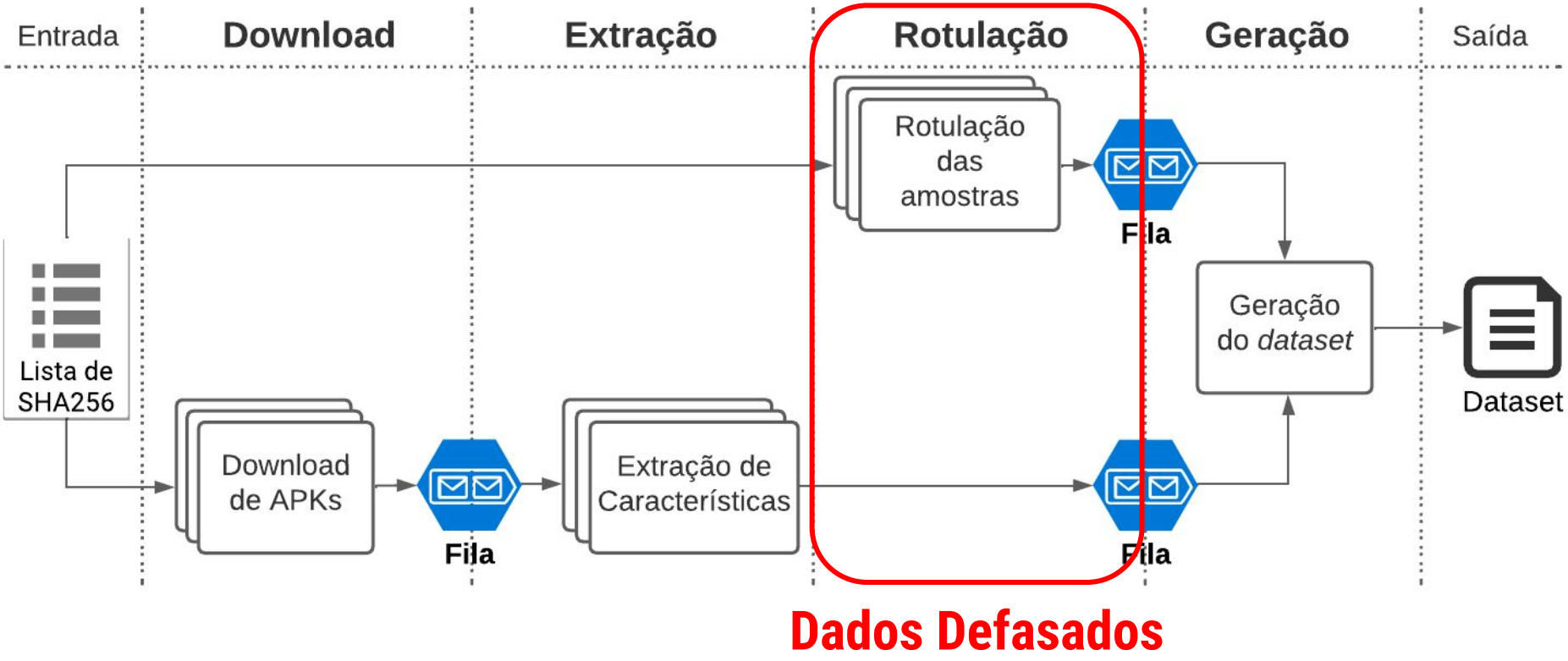


Maat

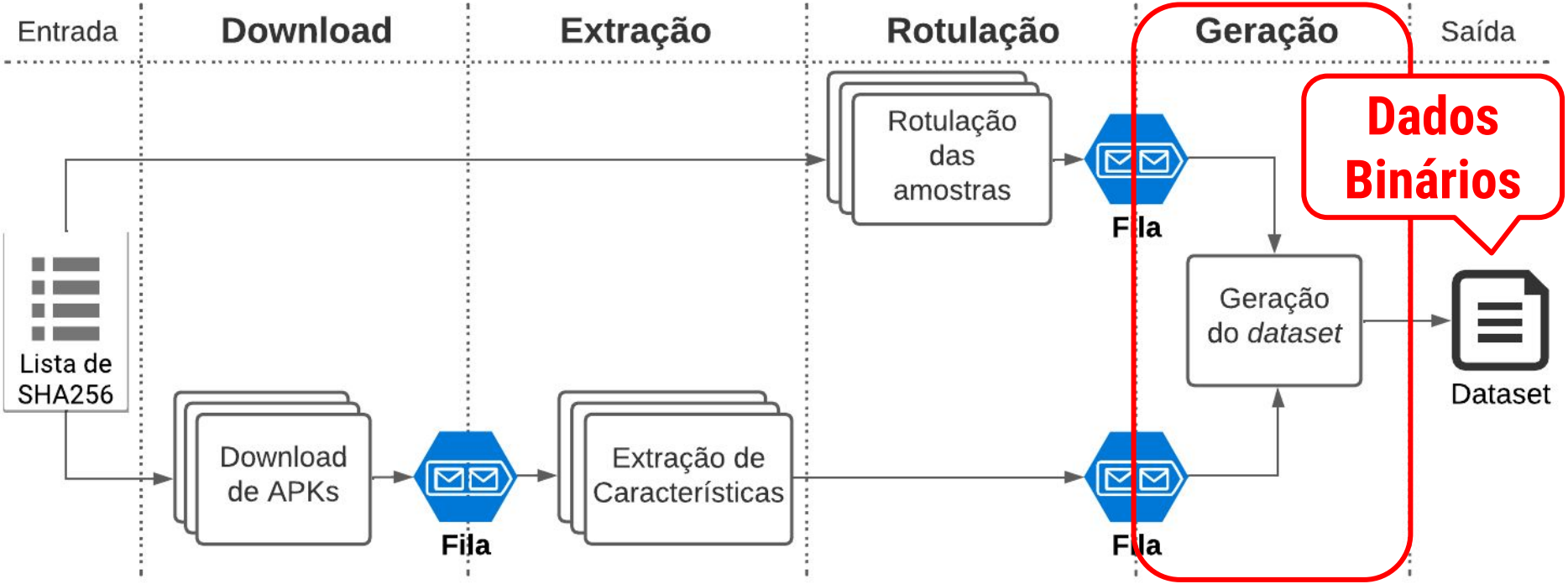
ADBuilder


Entrada
Download
Extração
Rotulação
Geração
Saída
Lista de SHA256
Download de APKs
Fila
Extração de Características
Rotulação das amostras
Fila
Fila
Geração do dataset
Dataset

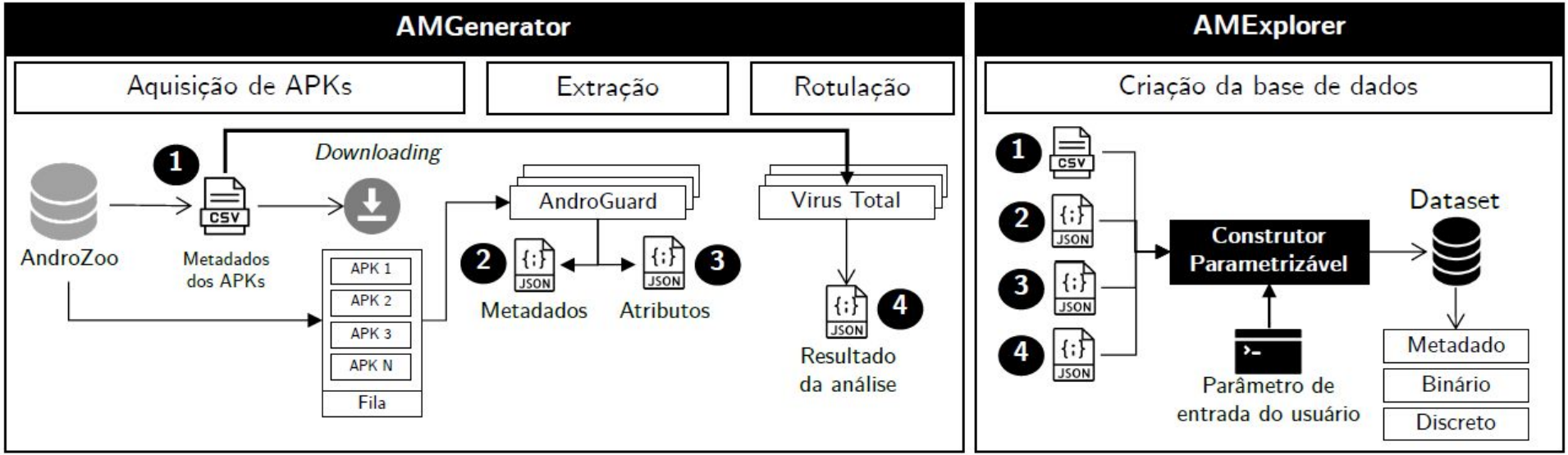
ADBuilder


Entrada
Download
Extração
Rotulação
Geração
Saída
Lista de SHA256
Download de APKs
Fila
Extração de Características
Rotulação das amostras
Fila
Fila
Geração do dataset
Dataset
Dados Defasados

ADBuilder


Entrada
Download
Extração
Rotulação
Geração
Saída
Lista de SHA256
Download de APKs
Fila
Extração de Características
Rotulação das amostras
Fila
Fila
Geração do dataset
Dataset
Dados Binários
Ausência de Metadados

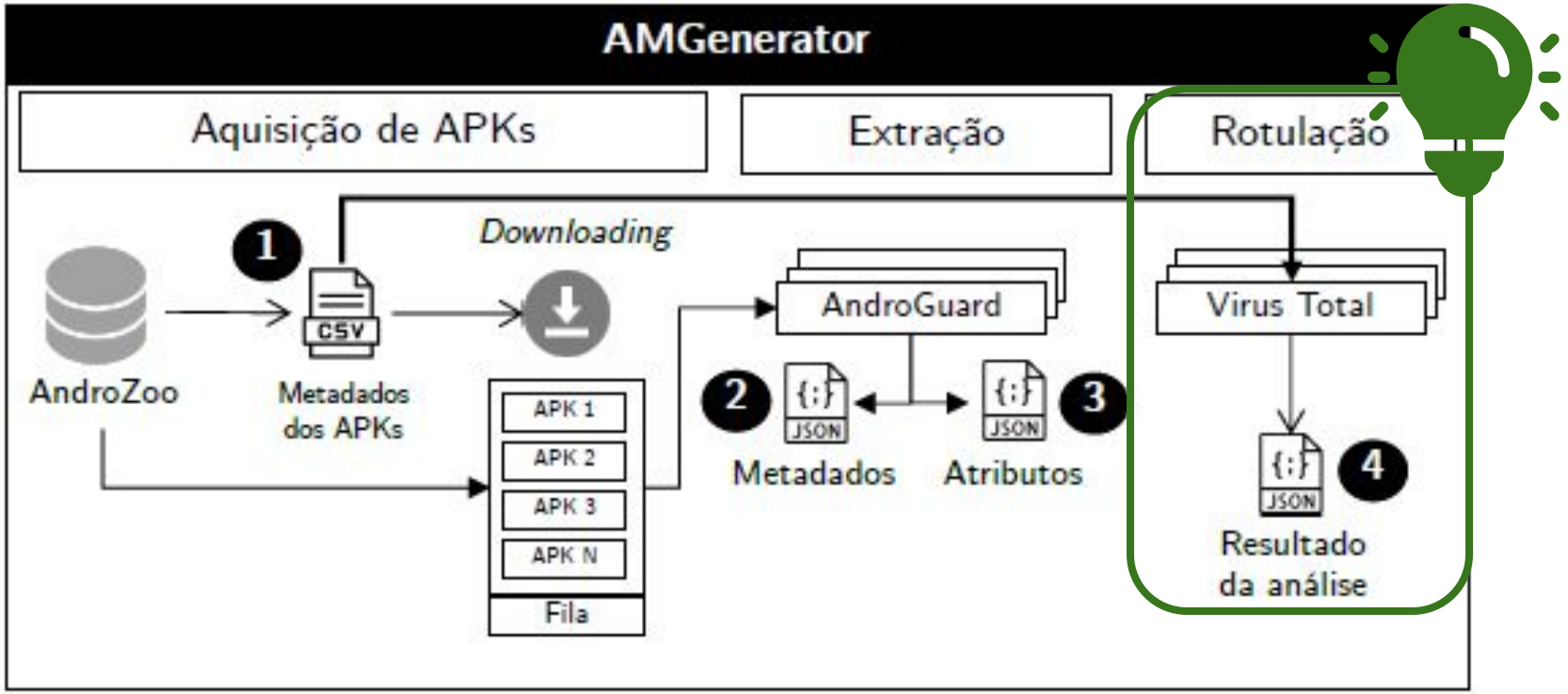
AMGenerator e AMExplorer



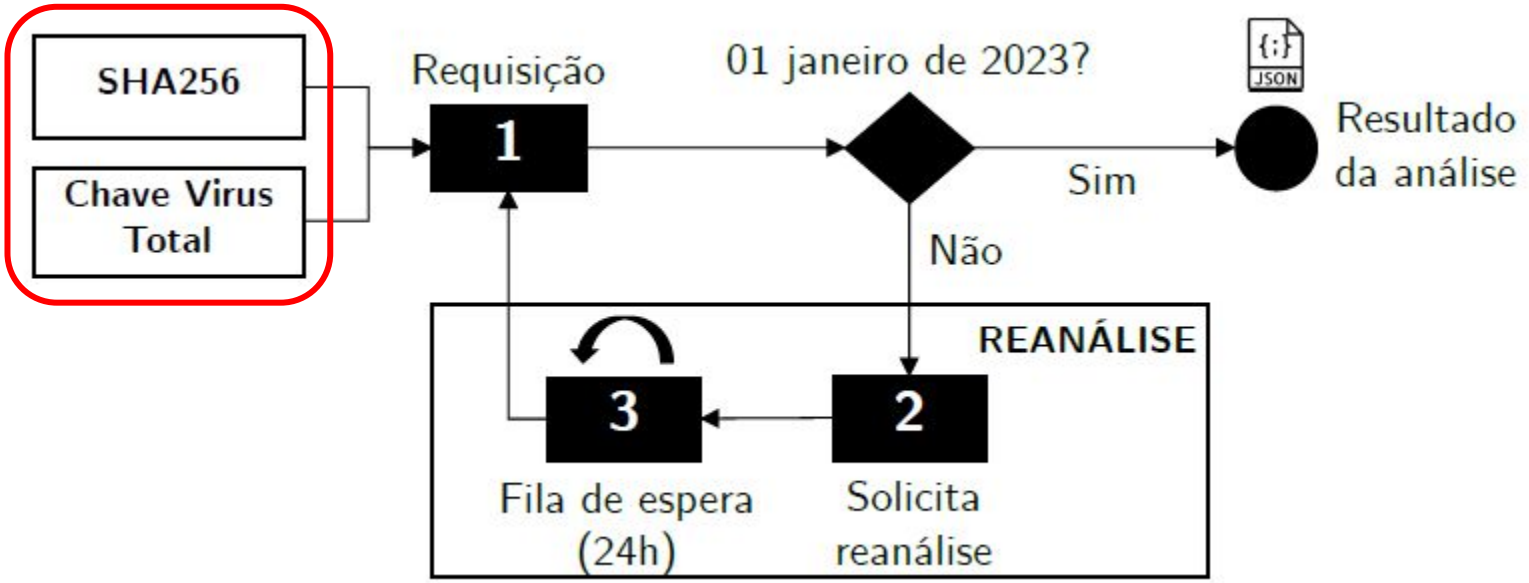
Dados Atualizados

Datasets Expressivos

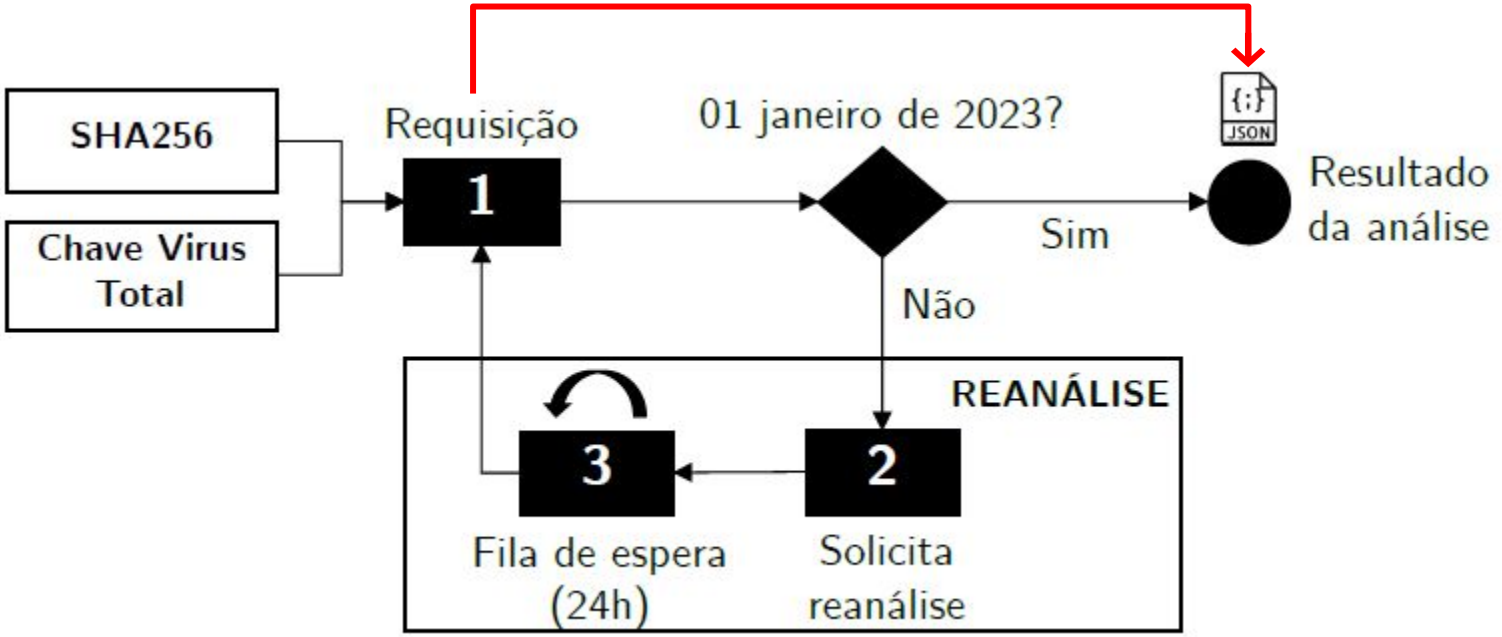
AMGenerator



AMGenerator (Rotulação)



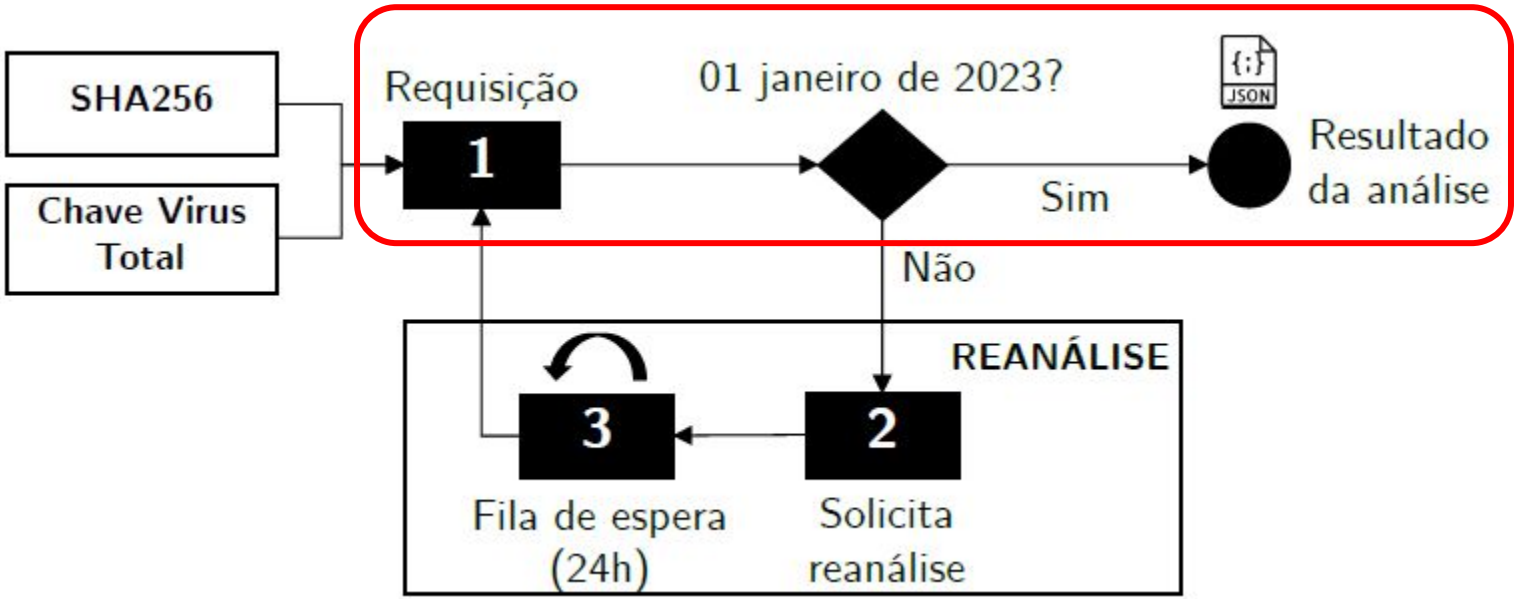
AMGenerator (Rotulação)



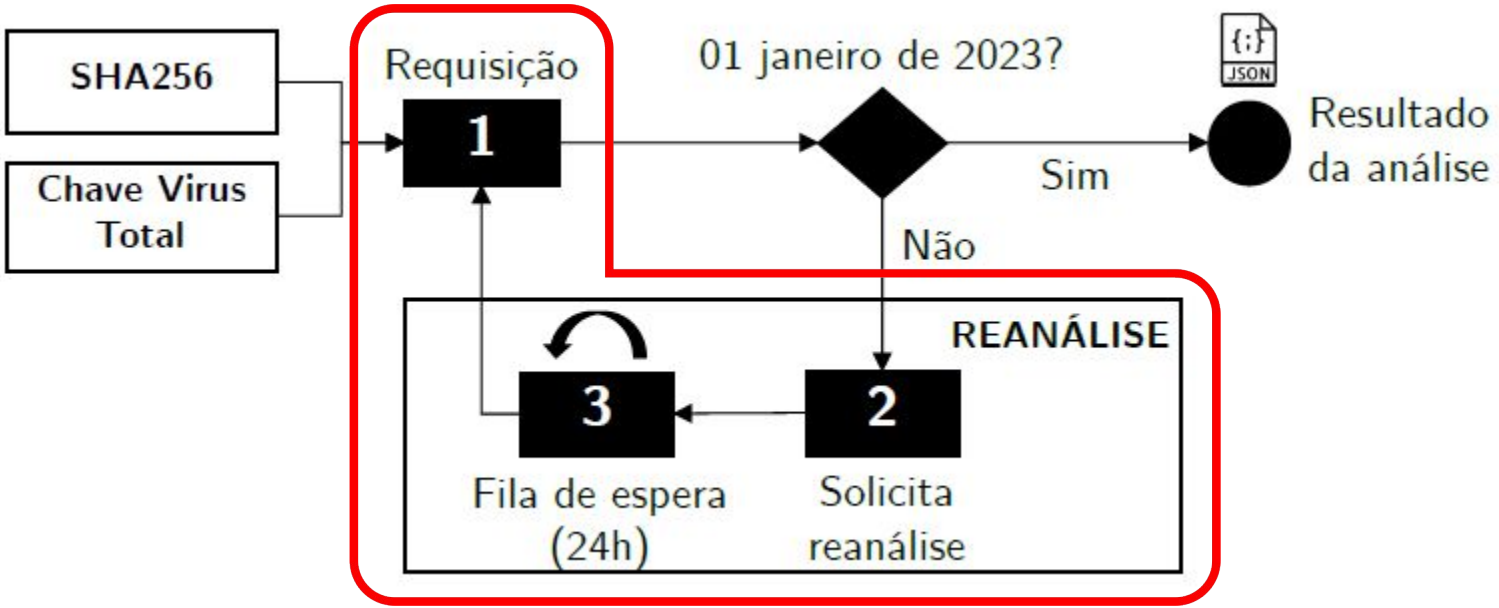
AMGenerator (Rotulação)


```
"last_analysis_results": {
    "Ikarus": {
        "category": "malicious",
        "engine_version": "6.1.14.0",
        "result": "PUA.AndroidOS.Nativead",
        "engine_update": "20230809"
    }
},
"last_analysis_stats": {
    "type-unsupported": 11,
    "malicious": 2,
    "undetected": 62
},
"last_analysis_date": 1691604279
```

AMGenerator (Rotulação)



AMGenerator (Rotulação)



AMGenerator (Parâmetros)

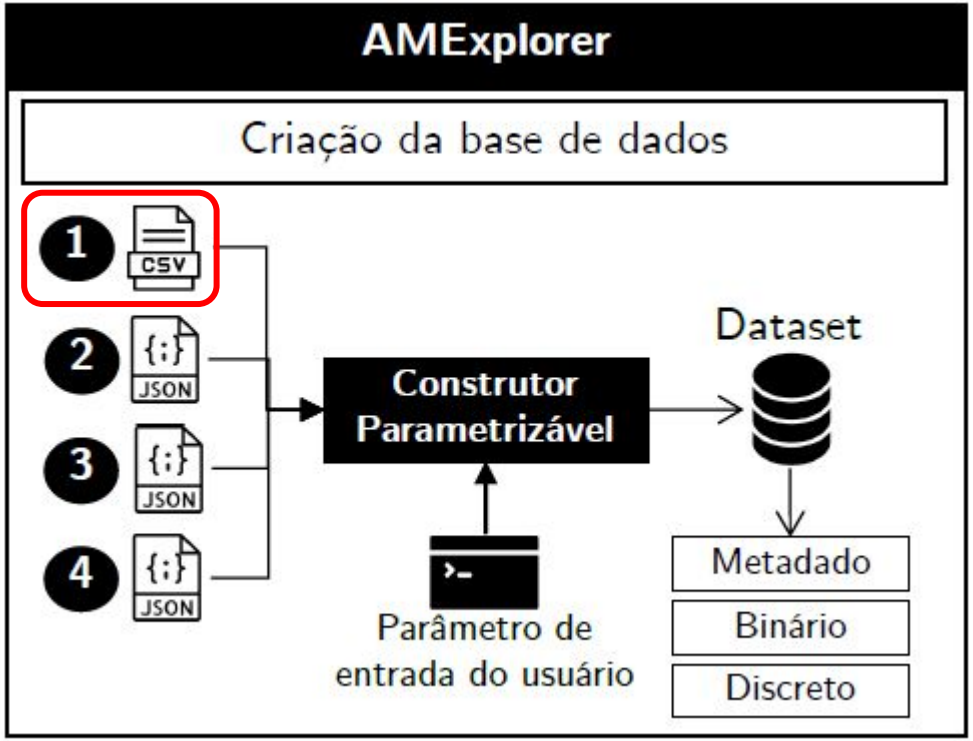
```
Show Help:
  -h, --help            show this help message and exit

AMGenerator parameters:
  --file FILE           File With a List of APKs SHA256 (One Per Line)
  --download            Download APK files
  --download-dir PATH   Directory to/from Downloads
  --androzoo-key KEY, -azk KEY
                        Androzoo API Key
  --num-parallel-download INT, -npd INT
                        Number of Parallel Downloads
  --extraction          APK Metadata and Features Extraction
  --num-parallel-extraction INT, -npe INT
                        Number of Parallel Process for Feature Extraction
  --label               VirusTotal Labelling
  --label-deadline INT, -ld INT
                        Deadline for Analysis to be Considered Current (in Epoch Format)
  --vt-key KEY, -vtk KEY
                        VirusTotal's API Key
  --reanalyze-time INT, -rt INT
                        Time to Wait for Reanalysis (in Minutes)
  --output-data PATH    Data Output Directory
```

AMExplorer

Tipo: Metadados

Origem: Androzoo

Módulo AMGenerator: Aquisição

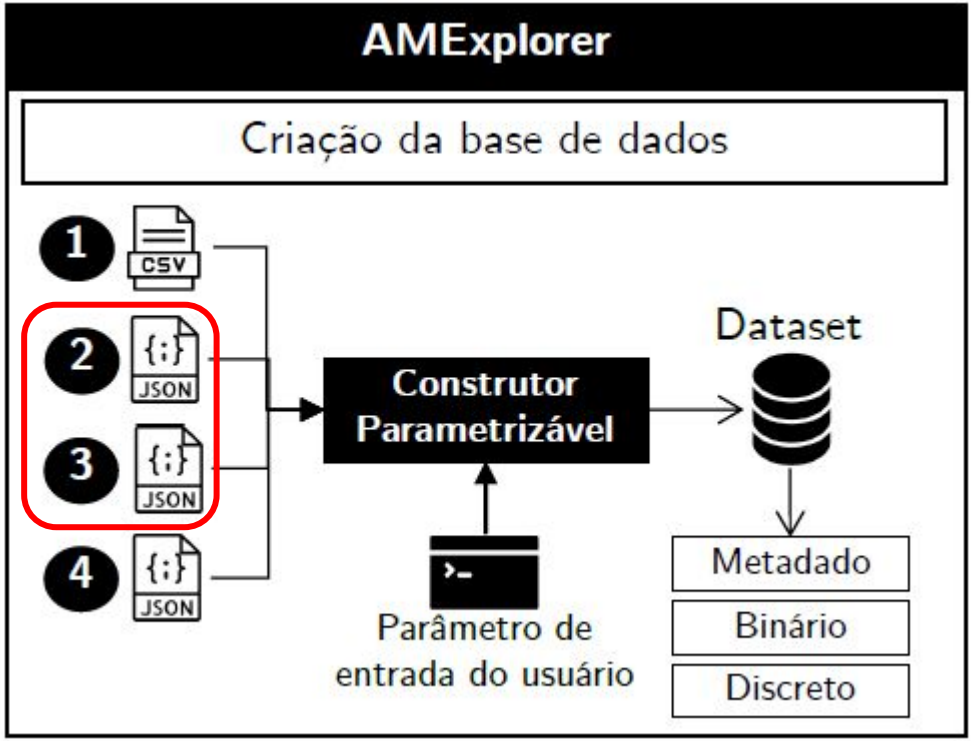


AMExplorer

Tipo: Metadados e Atributos

Origem: AndroGuard

Módulo AMGenerator: Extração

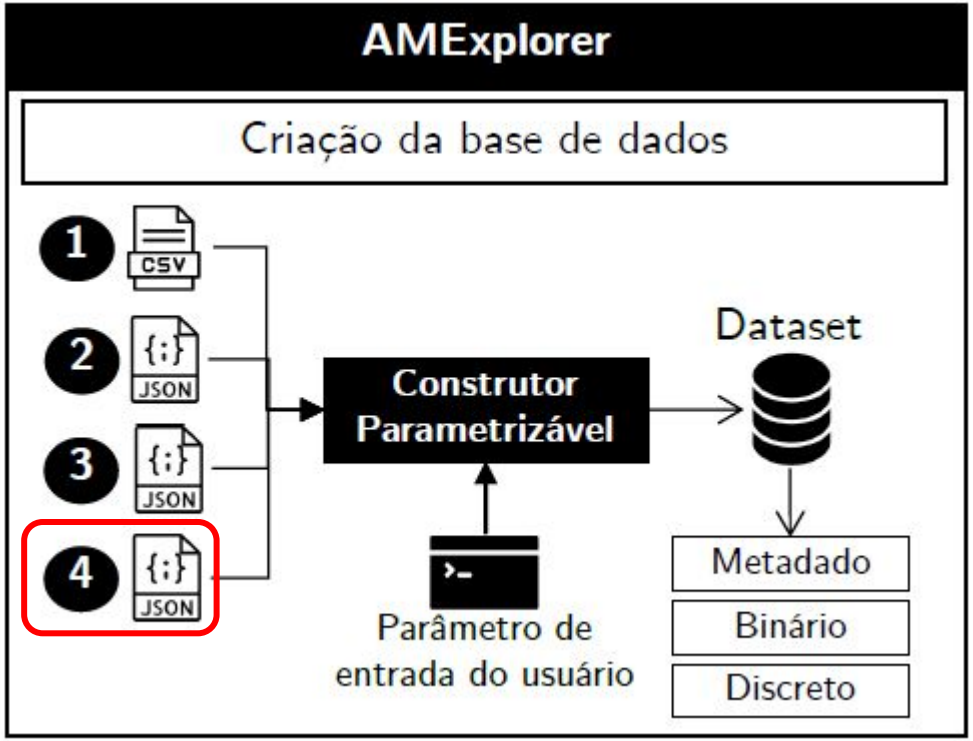


AMExplorer

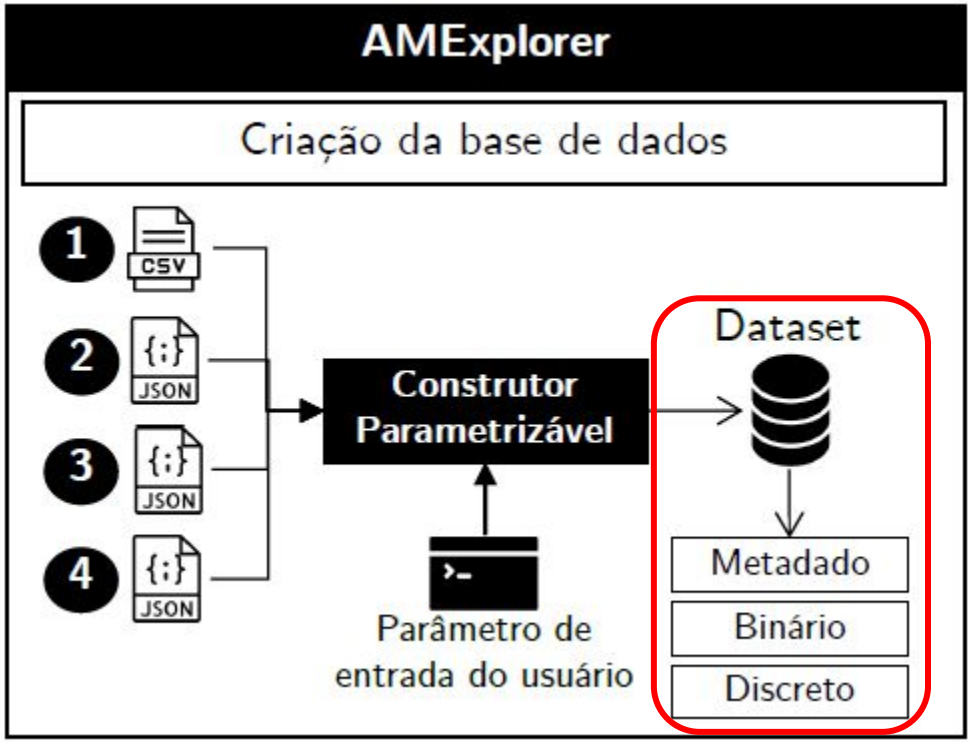
Tipo: Relátorios

Origem: VirusTotal

Módulo AMGenerator: Rotulação



AMExplorer


AMExplorer
Criação da base de dados
1
CSV
2
JSON
3
JSON
4
JSON
Construtor Parametrizável
Parâmetro de entrada do usuário
Dataset
Metadado
Binário
Discreto

AMExplorer (Parâmetros)

```
Show Help:
  -h, --help            show this help message and exit

ADBuilder Parameters:
  --dataset-type TYPE [TYPE ...]
                        Type of Dataset To Be Generated. Choices: ['metadata', 'binary', 'discrete']
  --dataset-type-all    Generate All Dataset Types
  --output-dir PATH     Dataset Output Directory (Default: 'output_amexplorer')
  --prefix PREFIX       Prefix To Be Used in Output Dataset

Metadata Dataset:
  --androguard-features PATH, -agf PATH
                        Directory Path of AndroGuard Features JSON Files
  --androguard-metadata PATH, -agm PATH
                        Directory Path of AndroGuard Metadata JSON Files
  --virustotal-metadata PATH, -vtm PATH
                        Directory Path of VirusTotal JSON Files
  --androzoo-metadata FILE_PATH, -azm FILE_PATH
                        AndroZoo CSV File Path

Binary or Discrete Dataset:
  --metadata-dataset FILE_PATH, -md FILE_PATH
                        Metadata Dataset CSV File Path
  --threshold INT, -th INT
                        Number of VirusTotal Scanners to Define Malware (Default: 4)
```

Repositórios

AMGenerator

AMExplorer

- Exemplos de Uso
- Scripts de Demonstração
- Uso em Docker
- Argumentos Disponíveis

OBRIGADO

Perguntas?

vanderson@ufam.edu.br
ppgi.ufam.edu.br