black hat ARSENAL

DECEMBER 9-12, 2024

EXCEL LONDON / UNITED KINGDOM

NotPacked++

Evading Static Packing Detection

By Alexandre D’Hondt, Jaber Ramhani, Charles-Henry Bertrand Van Ouytsel and Axel Legay

Outline

1. Introduction
2. Background
3. Adversarial Tool
4. Experiments & Results
5. Conclusion

Outline

1. Introduction
2. Background
3. Adversarial Tool
4. Experiments & Results
5. Conclusion

- Problem statement
- Objectives

Problem statement (1)

Packing =

- Set of transformations
- On binary file
- That preserves the original working at runtime

→ Large coverage in scientific literature, yet an open issue

→ Often employed with malware

→ Static detection increasingly relying on Machine Learning

Problem statement (2)

- Dedicated experimental toolkit
- Solves experiments repeatability
- Includes adversarial and unsupervised learning capabilities

Packing Box: Playing with Executable Packing (BHEU22)
Packing-Box: Breaking Detectors & Visualizing Packing (BHEU23)

Static detection challenges (con't) :

- Design efficient or refine existing attacks against common techniques and state-of-the-art features
- Static features robustness evaluation

- Few focus on problem-space adversarial learning yet
- No operational adversarial tool for evading packing static detection

Objectives

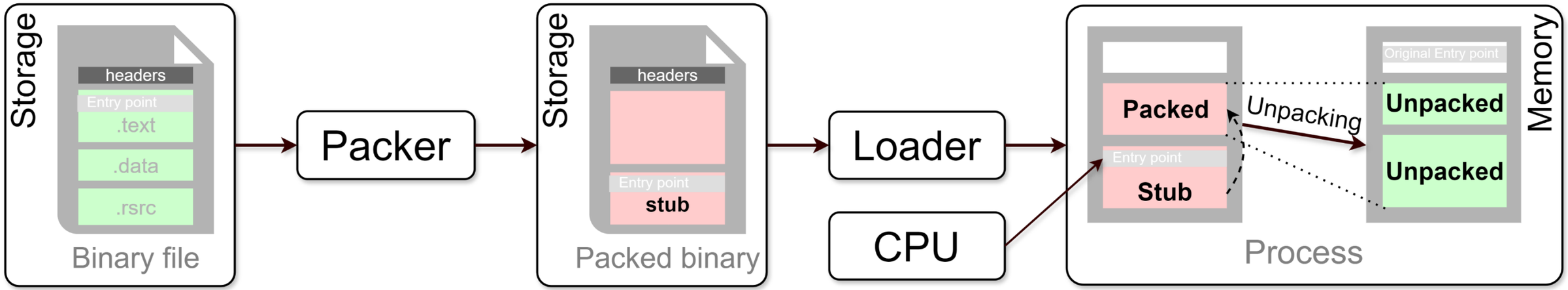
1. Review Packing Box's alterations (fix & improve)
2. Design an easy-to-deploy adversarial tool
3. Test this new tool on some data

Outline

1. Introduction
2. Background
3. Adversarial Tool
4. Experiments & Results
5. Conclusion

- Packing / unpacking
- Static detection & features
- Learning Pipeline
- Adversarial Learning

Packing / unpacking



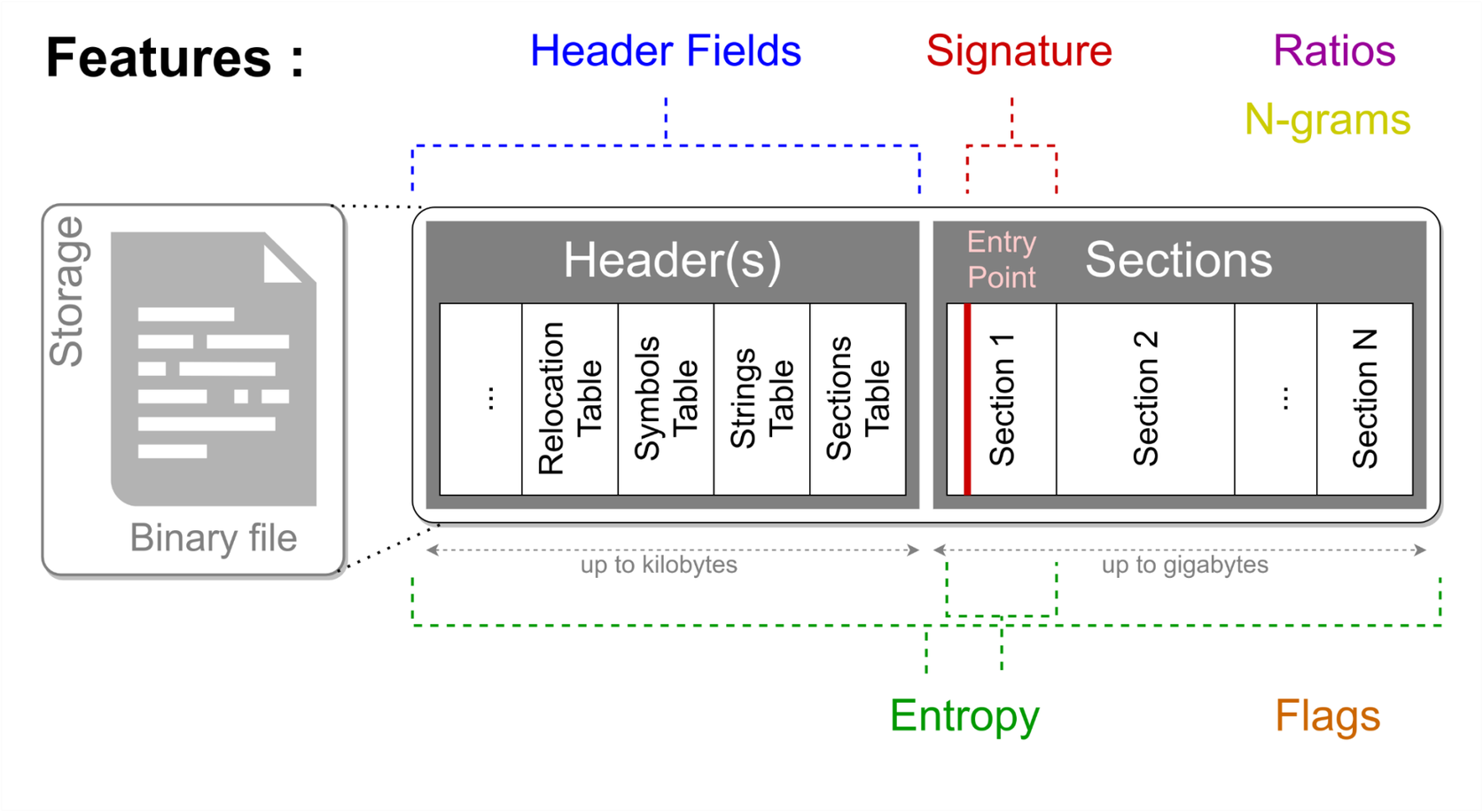
Transformations :

- Compression
- Encryption
- Encoding
- Protection
- Bundling
- Mutation
- Virtualization

Common usage :

- Size reduction
- SW piracy prevention / License management
- Malware

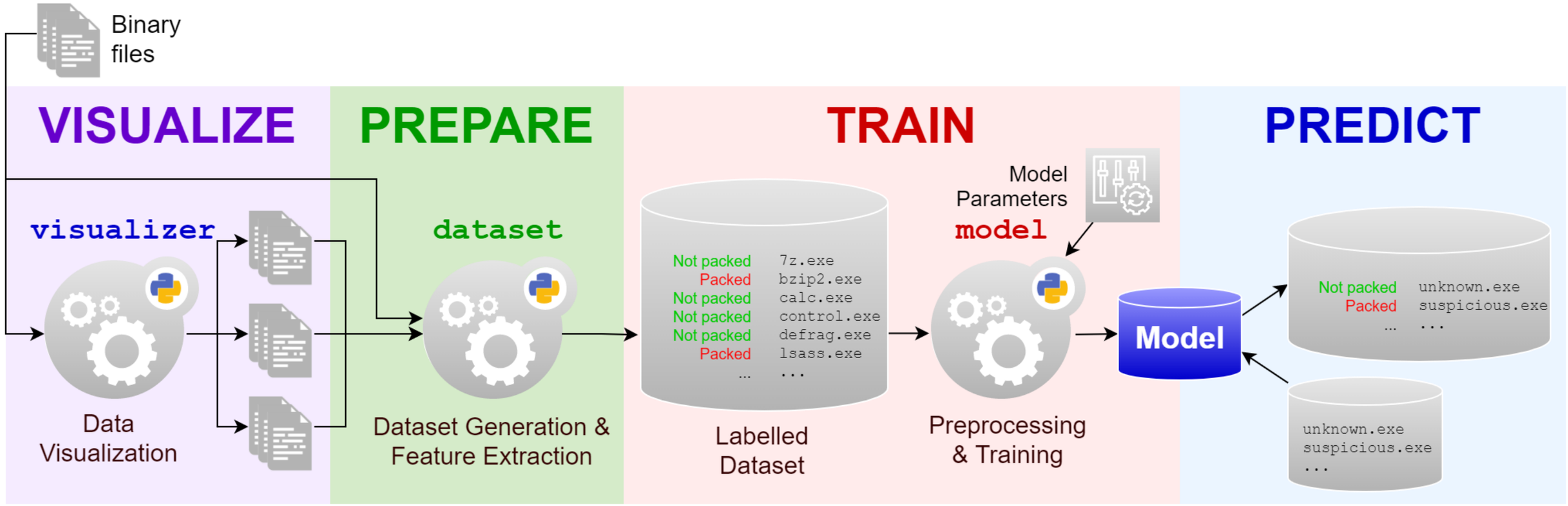
Static detection & features



Static (no execution) :

- Entropy threshold
- Pattern matching
- Signatures
- Heuristics
- Disassembly
- Control-Flow Graphs
- …

Learning pipeline



Adversarial Learning (1)

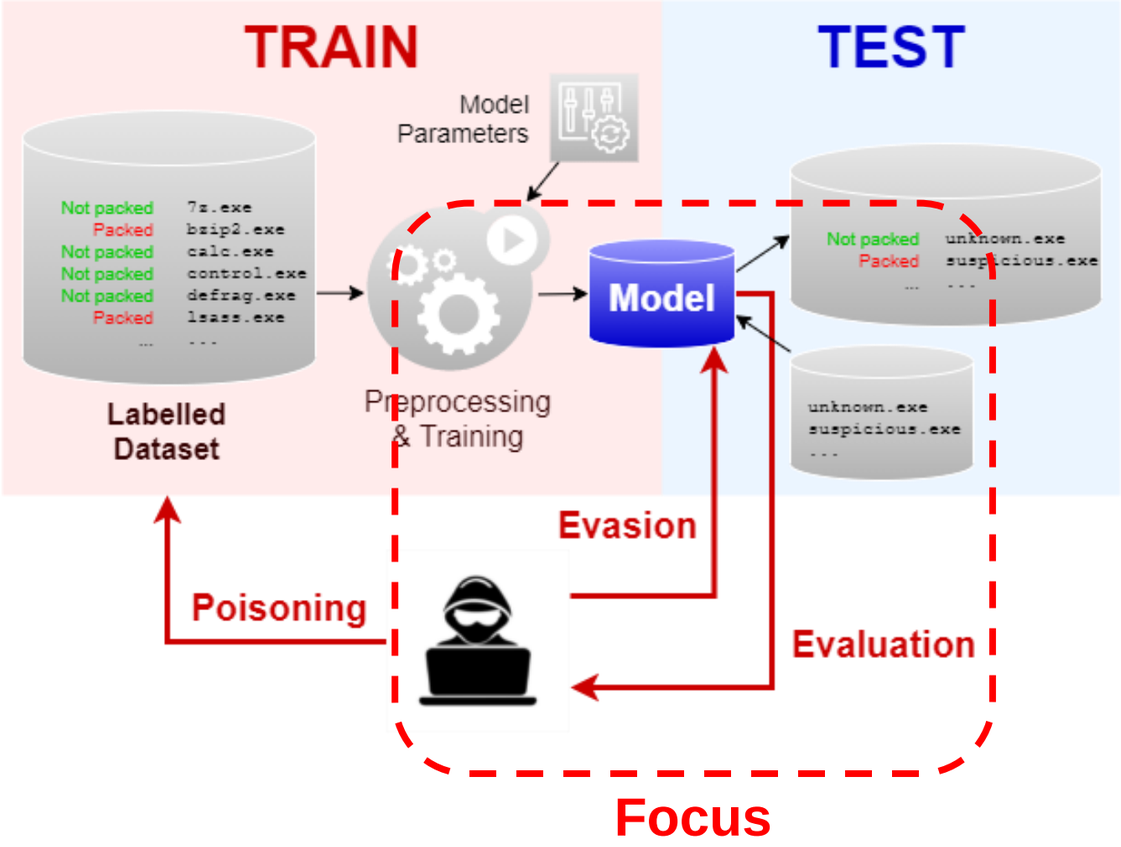
Threat model

- Attack Surface :

  Train (poisoning) VS Test (evasion) phase

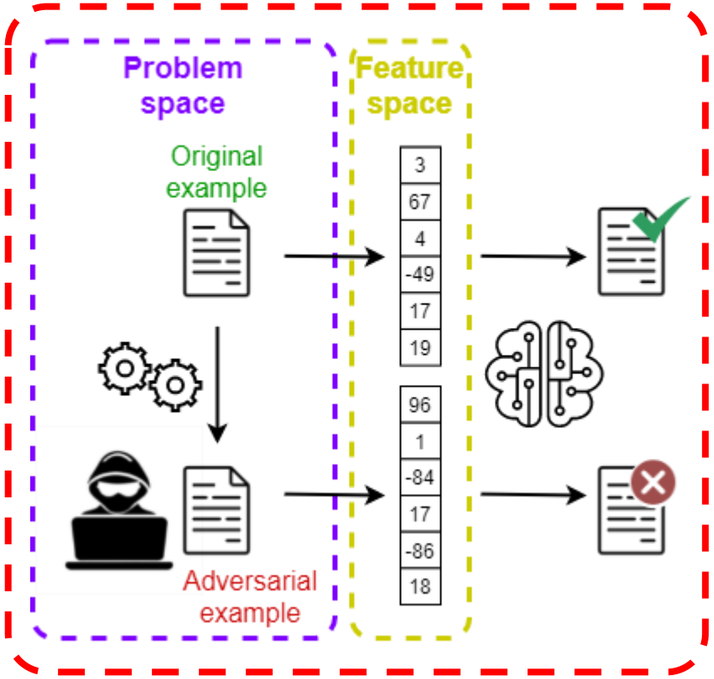
- Adversary :
  - ➢ Goal : Untargeted VS targeted
  - ➢ Capatibilities : ability to modify samples

    (tied to executable formats)
  - ➢ Knowledge : white-box VS black-box



Adversarial Learning (2)

Problem-space VS Feature-space attacks



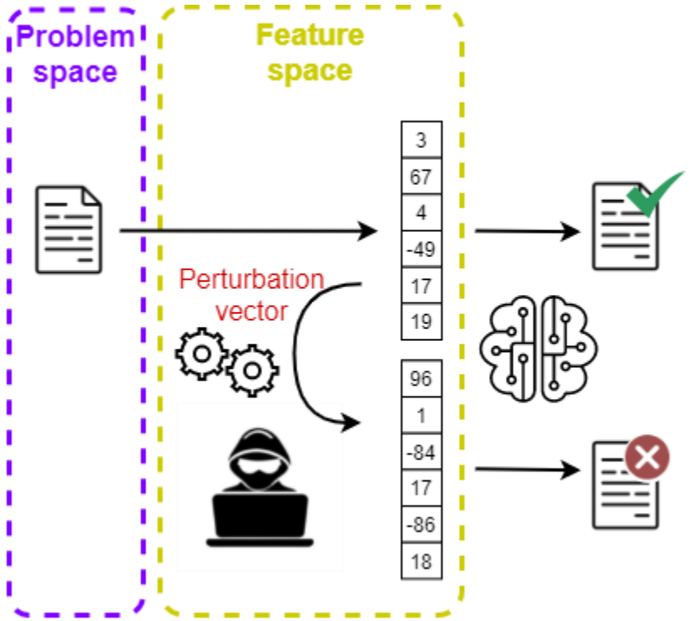
Focus

Problem-space : data transformation

- Can check validity of data
- No direct control on features

Feature-space : features perturbation

- Requires to feed features to the model
- Feature-to-problem mapping
- Easier



Outline

1. Introduction
2. Background
3. Adversarial Tool
4. Experiments & Results
5. Conclusion

- Design & Methodology
- Architecture
- Alterations
- Capabilities
- Getting Started

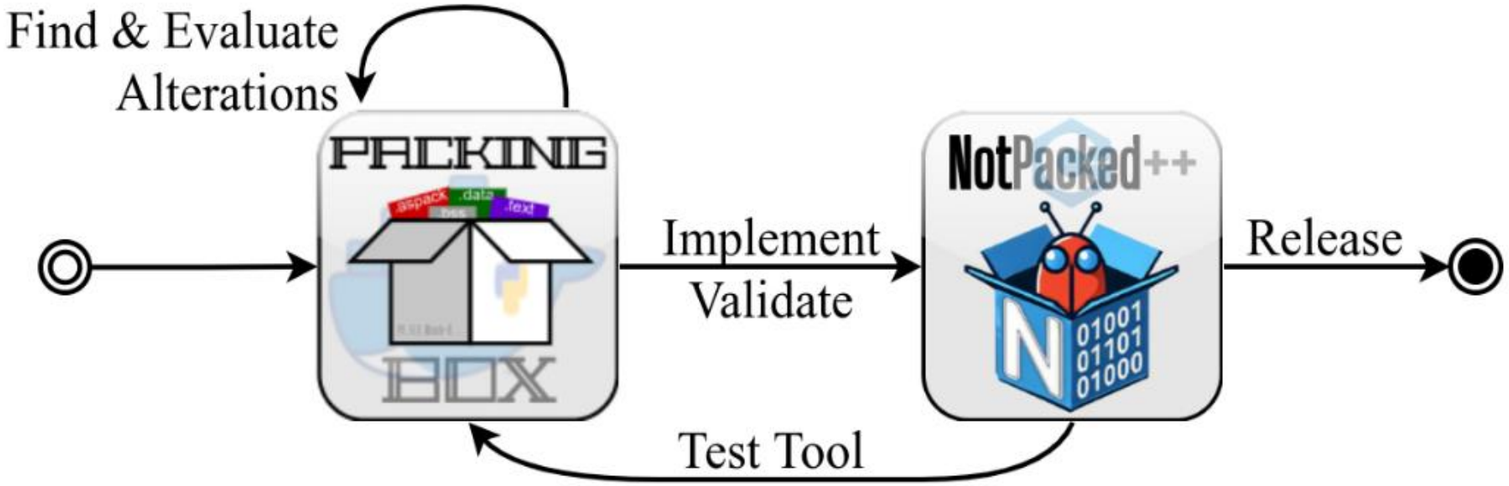
3. Adversarial Tool

Design & Methodology

- Self-contained tool
- PE format only (at this time)
- Written in C++
- Transforms packed executables to appear as not packed

Therefore aptly named **NotPacked++**

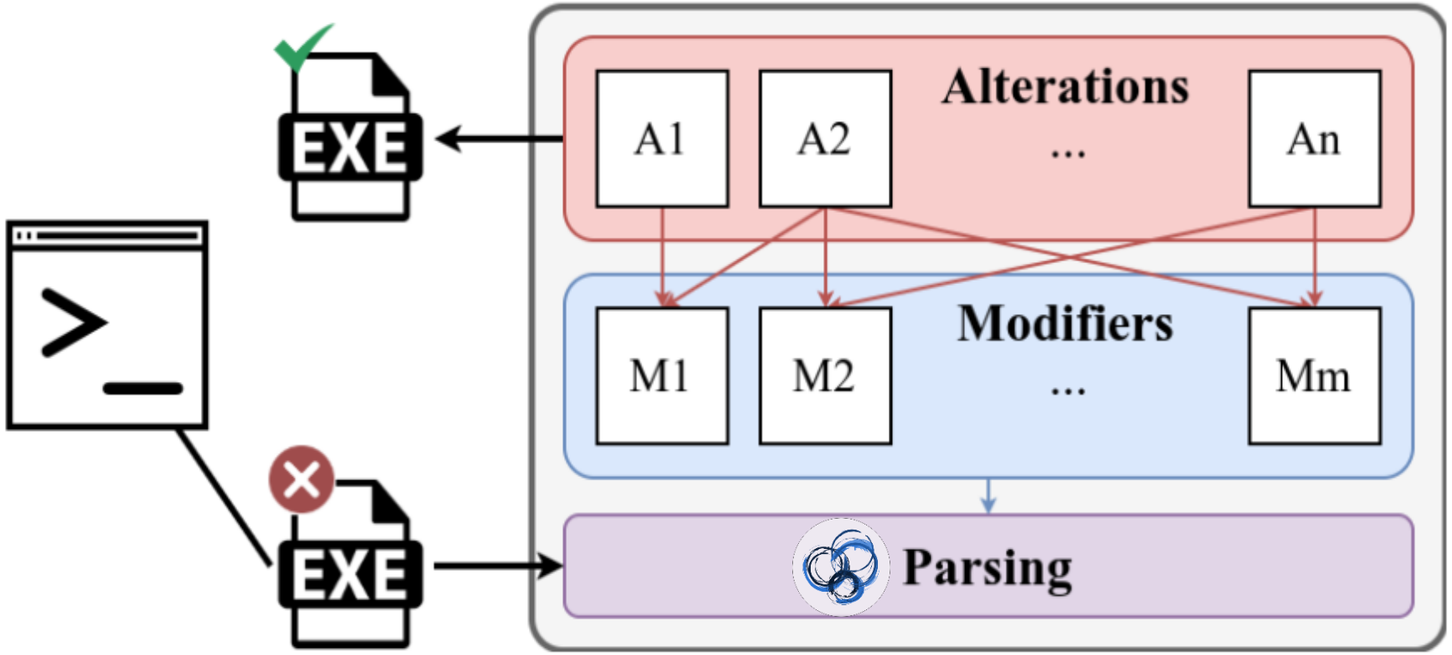
In reference to Notepad++ text editor



Architecture

3 components

1. Parsing library (i.e. LIEF)
   Abstraction of the input binary
2. Modifiers
   Specific transformations applied to input binary executable
3. Alterations
   Combination of alterations and modifiers



Alterations (1)

Starting point : **`alterations.yml`** = D'Hondt (2022) + Jennes (2023)

Caveats regarding current *Move EP to new section*

- LIEF fails to patch the IAT (creates a custom section "*./1*")
- Trampoline code broken, only valid in non-ASLR environment
- Fixed pattern, vulnerable to signature-based detection

Alterations (2)

Fixing *Move EP to new section*

Improved trampoline code

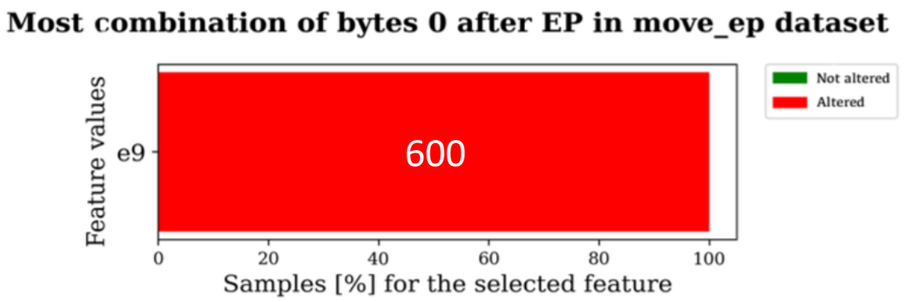
- Jump to offset, now valid in ASLR environment

Dynamic trampoline code with dead code

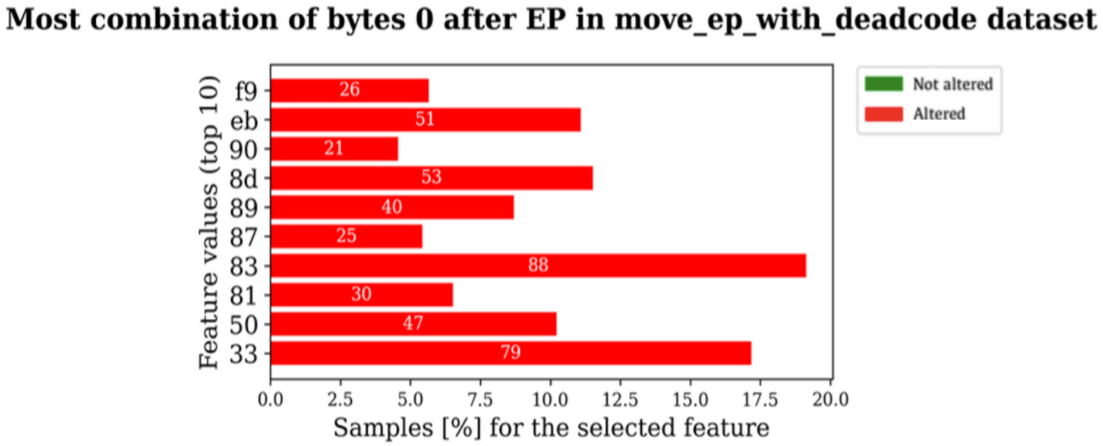
- Variable pattern, breaks signature-based detection

Sample Code

```
0x90                # NOP
0x33, 0xDB          # XOR EBX, EBX
0x83, 0xC0, 0x00    # ADD EAX, 0
0x87, 0xC0          # XCHG EAX, EAX
0x89, 0xC2          # MOV EDX, EDX
<<snipped>>
```



(a) Before



(b) After improvement with dead code injection

Alterations (3)

Status of `alterations.yml`

Alteration	Functionality
add_20_common_api_imports	Broken
add_low_entropy_text_section	Functional
fill_sections_with_zeros	Functional
move_ep_to_new_low_entropy_section	Broken
rename_packer_sections	Functional

Fixed

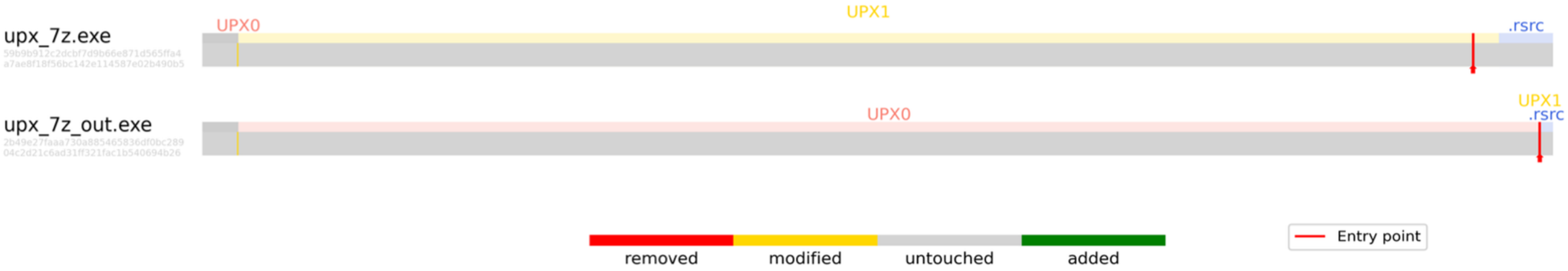
Alterations (4)

Finding new ideas

Edit SizeOfRawData

- Edit section's SizeOfRawData header, in Section Header only

- Rationale :
  Fill sections with zeros increases file size **+150%**
  Edit raw size targets same features



Alterations (5)

Finding new ideas

Change section permissions

- Targeting feature Number of (r)wx sections from Bertrand Van Ouytsel *et al.* (set of 75 most significant features among 119 from Biondi *et al.*)
- Common that not packed samples have 0

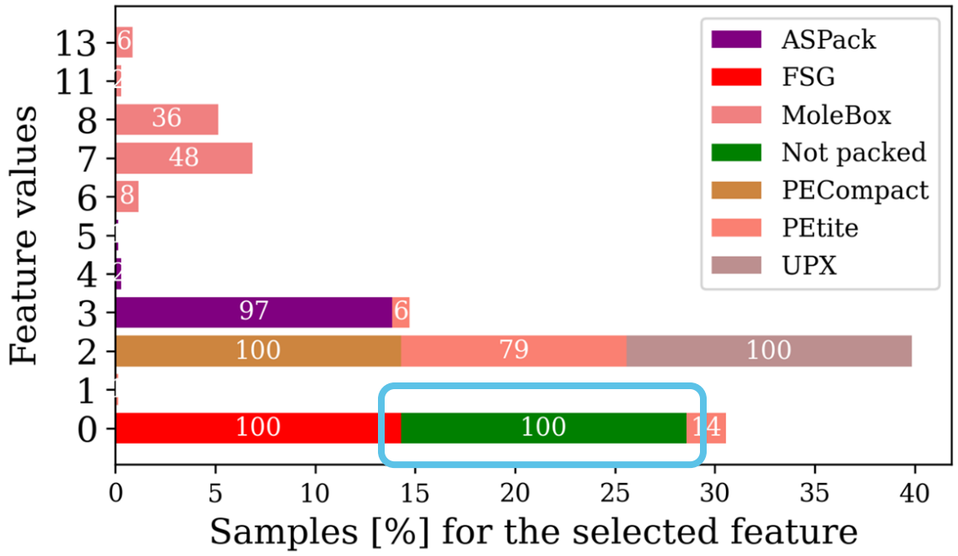
Address	Size	Info	Protection
00400000	00001000	upx_7z_out.exe	-R---
00401000	00010000	".data"	-RWC-
00411000	00009000	".rdata"	-R---
0041A000	00001000	".rsrc"	-R---
0041B000	00002000	".text"	ER---

(a) Before (State at entrypoint)

Address	Size	Info	Protection
00400000	00001000	upx_7z_out.exe	-R---
00401000	00010000	".data"	ERWC-
00411000	00009000	".rdata"	ERWC-
0041A000	00001000	".rsrc"	-RWC-
0041B000	00002000	".text"	ER---

(b) After (State after trampoline)

Num. of writable & executable sections for baseline_np

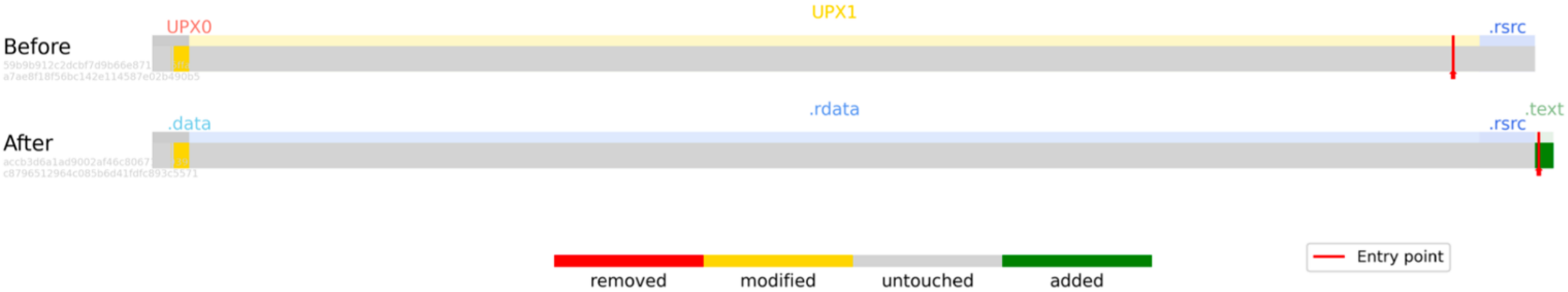


Alterations (6)

Finding new ideas

Chaining alterations

1. (Fixed & improved) Move EP to new section
2. Rename packer section names
3. Add a low-entropy section
4. Change section permissions



Capabilities

1. Fixed and enhanced alteration : *Move EP to new section*

   Issue fixed with new EP jump (ASLR, x64), trampoline code made dynamic

2. New alterations : *Raw size edit*, *Change section permissions*

   Modification of related header without increasing the actual raw size

   Touching section permissions to make them appear as in not packed samples

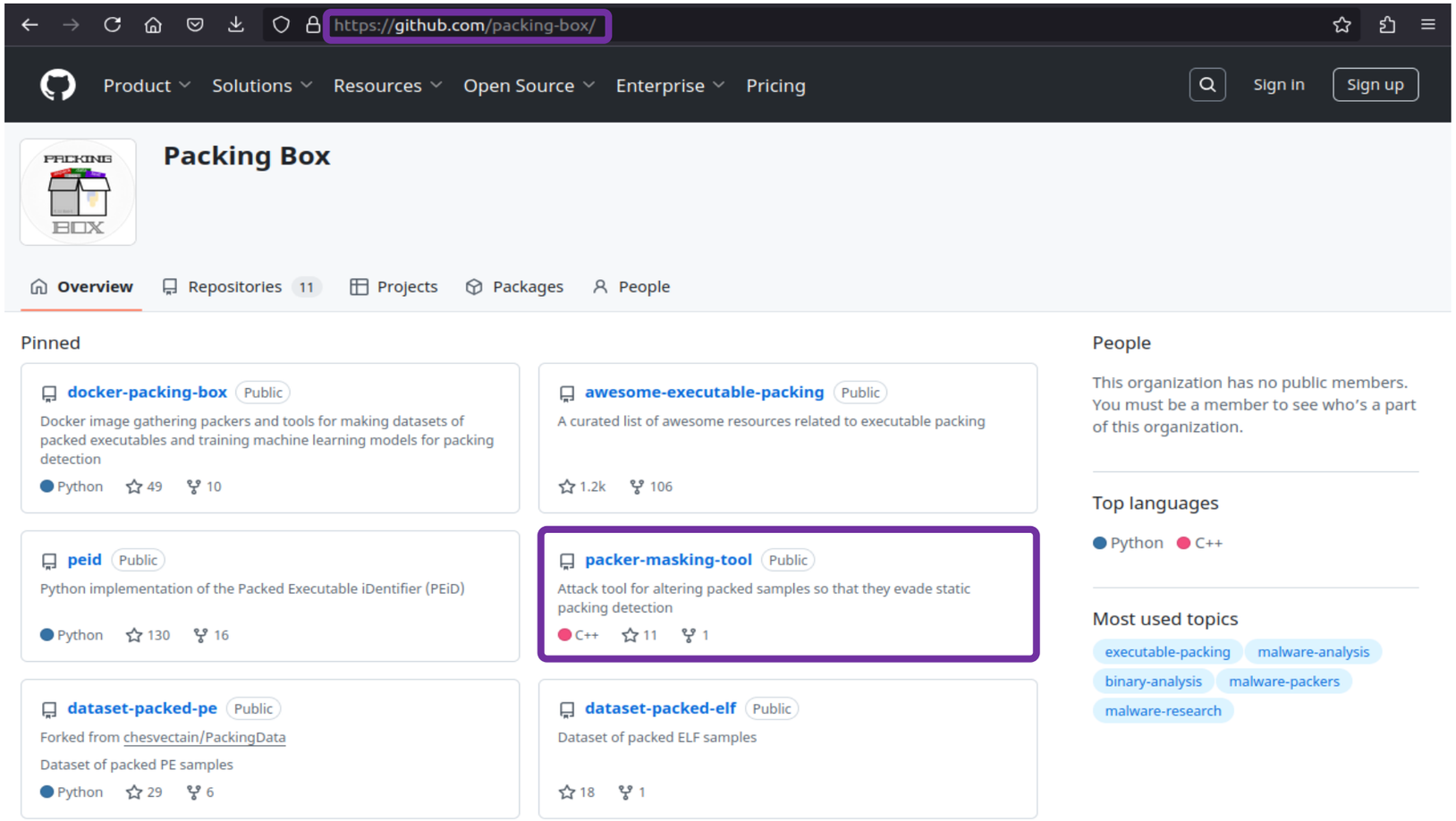
3. Super-alteration

   Chaining of *Move EP to new section* > *Rename sections* > *Add a low entropy section* > *Change section permissions*

Getting started (1)

Starting point :

1. Open terminal
2. Clone the repo



Getting started (2)

Build

Try It Now !

```
# ./install_lief.sh
# make
g++ -Wall -std=c++17  -c src/main.cpp
<<snipped>>
rm -f *.o
```

```
# docker build -t notpackedpp .
[+] Building 52.9s (13/13) FINISHED
<<snipped>>
# docker run -it -h notpackedpp -v `pwd`:/mnt/share notpackedpp
user@notpackedpp:/mnt/share$ make
g++ -Wall -std=c++17  -c src/main.cpp
<<snipped>>
rm -f *.o
user@notpackedpp:/mnt/share$ exit
```

Use

Try It Now !

```
$ ./notpacked++ test-upx-packed.exe
$ ./notpacked++ test-upx-packed.exe --permissions --fill-sections -o test-not-packed.exe
```

Outline

1. Introduction
2. Background
3. Adversarial Tool
4. Experiments & Results
5. Conclusion

- Setup
- Impact on features
- Impact on detection

4. Experiments & Results

Setup

Ingesting dataset-packed-pe

- aspack
- ...
- petite
- ...
- upx

```
┌──[user@packing-box]──[~]──────────
$ git clone https://github.com/packing-box/dataset-packed-pe
┌──[user@packing-box]──[~]──────────
$ dataset ingest dataset-packed-pe --labels dataset-packed-pe/labels.json
    --exclude outliers
<<snipped>>
```

Creating our **baseline** dataset

```
┌──[user@packing-box]──[~]──────────
$ dataset select upx baseline -n 100
100% ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 100/100 samples • 0:00:17 • 0:00:00 • baseline
┌──[user@packing-box]──[~]──────────
$ dataset select aspack baseline -n 100
100% ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 100/100 samples • 0:00:17 • 0:00:00 • baseline
┌──[user@packing-box]──[~]──────────
$ dataset select petite baseline -n 100
100% ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 100/100 samples • 0:00:17 • 0:00:00 • baseline
<<snipped>>
```

4. Experiments & Results

Visualization with PE-Bear

Try It Now !

```
$ ./notpacked++ test-upx-packed.exe -o test-upx-packed_masked.exe
```

Before

Disasm: UPX1 | General | Strings | DOS Hdr | Rich Hdr | File Hdr | Optional Hdr | Section Hdrs

Name	Raw Addr.	Raw size	Virtual Addr.	Virtual Size	Characteristics	Ptr to Reloc.	Num. o
UPX0	400	0	1000	10000	E0000080	0	0
UPX1	400	8C00	11000	9000	E0000040	0	0
.rsrc	9000	600	1A000	1000	C0000040	0	0

After

Disasm: .text | General | Strings | DOS Hdr | Rich Hdr | File Hdr | Optional Hdr | Section Hdrs

Name	Raw Addr.	Raw size	Virtual Addr.	Virtual Size	Characteristics	Ptr to Reloc.	Num.
.data	400	10000	1000	10000	C0000040	0	0
.rdata	10400	9000	11000	9000	40000040	0	0
.rsrc	19000	1000	1A000	1000	40000040	0	0
.text	1A000	2A00	1B000	281C	60000020	0	0

Impact on features (1)

Datasets

- Reference : not packed samples
- Super-alterations (AllFill, AllRaw, Permissions)
- Individual alteration (MoveEP, ..., Rename)

Move EP to new section
> Rename sections
> Add a low entropy section
> Change section permissions

Plotting information gain from our dataset


```
$ dataset plot infogain-compare not-packed --max-features 20 \
    --datasets AllFill AllRaw Permissions MoveEP RawSize Fill_0 AddApi Rename
```

4. Experiments & Results

Impact on features (2)

1 reference dataset
3 datasets related to super-alterations
5 datasets ; 1 for each alteration

20 most relevant features

Infogain comparison (reference = Not-packed)

	AllFill	AllRaw	Permissions	MoveEP	RawSize	Fill_0	AddApi	Rename
ratio_standard_sections	0.21	0.21	0.15	1.00	0.77	0.77	0.88	0.51
number_wx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
number_rwx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
byte_0_after_ep	0.23	0.23	0.23	0.23	1.00	0.98	1.00	1.00
number_non_standard_sections	0.18	0.18	0.12	0.88	0.71	0.72	1.00	0.63
ep_ratio	0.33	0.45	0.46	0.27	0.99	0.65	1.00	1.00
byte_X_after_ep	0.25	0.24	0.25	0.24	0.98	0.90	1.00	1.00
number_rx_sections	0.31	0.31	0.31	0.86	0.99	1.00	0.86	1.00
number_x_sections	0.33	0.33	0.33	0.86	0.99	1.00	0.86	1.00
is_x_section_not_code	0.01	0.01	0.01	1.00	0.99	1.00	0.98	1.00
entropy_section_with_ep	0.74	0.73	0.75	0.51	1.00	1.00	0.99	1.00
is_code_section_not_present	0.00	0.00	0.00	0.00	0.99	1.00	1.00	0.80
max_ratio_rdata	0.74	0.68	0.57	0.47	1.00	0.92	0.83	1.00
number_sections_vsize>dsize	0.52	1.00	1.00	1.00	1.00	0.52	1.00	1.00
min_ratio_rdata	0.69	1.00	0.46	0.46	0.95	0.69	0.46	0.46
is_ep_not_in_code_section	0.01	0.01	0.01	0.68	0.98	1.00	1.00	0.87
is_ep_not_in_standard_section	0.00	0.00	0.00	0.65	0.98	1.00	1.00	0.75
size_of_code	0.42	0.40	0.42	1.00	0.84	0.84	0.84	0.84
ratio_malicious_api_imported	0.71	0.72	1.00	1.00	0.99	0.94	0.71	1.00
number_malicious_api_imported	0.61	0.62	0.94	0.94	0.93	1.00	0.61	0.94

4. Experiments & Results

Impact on features (3)

Lighter color : more impact on detection
Darker color : less impact on detection

Infogain comparison (reference = Not-packed)

	AllFill	AllRaw	Permissions	MoveEP	RawSize	Fill_0	AddApi	Rename
ratio_standard_sections	0.21	0.21	0.15	1.00	0.77	0.77	0.88	0.51
number_wx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
number_rwx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
byte_0_after_ep	0.23	0.23	0.23	0.23	1.00	0.98	1.00	1.00
number_non_standard_sections	0.18	0.18	0.12	0.88	0.71	0.72	1.00	0.63
ep_ratio	0.33	0.45	0.46	0.27	0.99	0.65	1.00	1.00
byte_X_after_ep	0.25	0.24	0.25	0.24	0.98	0.90	1.00	1.00
number_rx_sections	0.31	0.31	0.31	0.86	0.99	1.00	0.86	1.00
number_x_sections	0.33	0.33	0.33	0.86	0.99	1.00	0.86	1.00
is_x_section_not_code	0.01	0.01	0.01	1.00	0.99	1.00	0.98	1.00
entropy_section_with_ep	0.74	0.73	0.75	0.51	1.00	1.00	0.99	1.00
is_code_section_not_present	0.00	0.00	0.00	0.00	0.99	1.00	1.00	0.80
max_ratio_rdata	0.74	0.68	0.57	0.47	1.00	0.92	0.83	1.00
number_sections_vsize>dsize	0.52	1.00	1.00	1.00	1.00	0.52	1.00	1.00
min_ratio_rdata	0.69	1.00	0.46	0.46	0.95	0.69	0.46	0.46
is_ep_not_in_code_section	0.01	0.01	0.01	0.68	0.98	1.00	1.00	0.87
is_ep_not_in_standard_section	0.00	0.00	0.00	0.65	0.98	1.00	1.00	0.75
size_of_code	0.42	0.40	0.42	1.00	0.84	0.84	0.84	0.84
ratio_malicious_api_imported	0.71	0.72	1.00	1.00	0.99	0.94	0.71	1.00
number_malicious_api_imported	0.61	0.62	0.94	0.94	0.93	1.00	0.61	0.94

Useful for detection

Useless for detection

4. Experiments & Results

Impact on features (4)

High impact of super-alterations on detection
6 commonly used features highly impacted

Infogain comparison (reference = Not-packed)

	AllFill	AllRaw	Permissions	MoveEP	RawSize	Fill_0	AddApi	Rename
ratio_standard_sections	0.21	0.21	0.15	1.00	0.77	0.77	0.88	0.51
number_wx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
number_rwx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
byte_0_after_ep	0.23	0.23	0.23	0.23	1.00	0.98	1.00	1.00
number_non_standard_sections	0.18	0.18	0.12	0.88	0.71	0.72	1.00	0.63
ep_ratio	0.33	0.45	0.46	0.27	0.99	0.65	1.00	1.00
byte_X_after_ep	0.25	0.24	0.25	0.24	0.98	0.90	1.00	1.00
number_rx_sections	0.31	0.31	0.31	0.86	0.99	1.00	0.86	1.00
number_x_sections	0.33	0.33	0.33	0.86	0.99	1.00	0.86	1.00
is_x_section_not_code	0.01	0.01	0.01	1.00	0.99	1.00	0.98	1.00
entropy_section_with_ep	0.74	0.73	0.75	0.51	1.00	1.00	0.99	1.00
is_code_section_not_present	0.00	0.00	0.00	0.00	0.99	1.00	1.00	0.80
max_ratio_rdata	0.74	0.68	0.57	0.47	1.00	0.92	0.83	1.00
number_sections_vsize>dsize	0.52	1.00	1.00	1.00	1.00	0.52	1.00	1.00
min_ratio_rdata	0.69	1.00	0.46	0.46	0.95	0.69	0.46	0.46
is_ep_not_in_code_section	0.01	0.01	0.01	0.68	0.98	1.00	1.00	0.87
is_ep_not_in_standard_section	0.00	0.00	0.00	0.65	0.98	1.00	1.00	0.75
size_of_code	0.42	0.40	0.42	1.00	0.84	0.84	0.84	0.84
ratio_malicious_api_imported	0.71	0.72	1.00	1.00	0.99	0.94	0.71	1.00
number_malicious_api_imported	0.61	0.62	0.94	0.94	0.93	1.00	0.61	0.94

Useful for detection

Useless for detection

4. Experiments & Results

Impact on features (5)

Feature “number of writable and executable sections” :

Fewly impacted by single alterations

Fully impacted with super-alterations

Infogain comparison (reference = Not-packed)

	AllFill	AllRaw	Permissions	MoveEP	RawSize	Fill_0	AddApi	Rename
ratio_standard_sections	0.21	0.21	0.15	1.00	0.77	0.77	0.88	0.51
number_wx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
number_rwx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
byte_0_after_ep	0.23	0.23	0.23	0.23	1.00	0.98	1.00	1.00
number_non_standard_sections	0.18	0.18	0.12	0.88	0.71	0.72	1.00	0.63
ep_ratio	0.33	0.45	0.46	0.27	0.99	0.65	1.00	1.00
byte_X_after_ep	0.25	0.24	0.25	0.24	0.98	0.90	1.00	1.00
number_rx_sections	0.31	0.31	0.31	0.86	0.99	1.00	0.86	1.00
number_x_sections	0.33	0.33	0.33	0.86	0.99	1.00	0.86	1.00
is_x_section_not_code	0.01	0.01	0.01	1.00	0.99	1.00	0.98	1.00
entropy_section_with_ep	0.74	0.73	0.75	0.51	1.00	1.00	0.99	1.00
is_code_section_not_present	0.00	0.00	0.00	0.00	0.99	1.00	1.00	0.80
max_ratio_rdata	0.74	0.68	0.57	0.47	1.00	0.92	0.83	1.00
number_sections_vsize>dsize	0.52	1.00	1.00	1.00	1.00	0.52	1.00	1.00
min_ratio_rdata	0.69	1.00	0.46	0.46	0.95	0.69	0.46	0.46
is_ep_not_in_code_section	0.01	0.01	0.01	0.68	0.98	1.00	1.00	0.87
is_ep_not_in_standard_section	0.00	0.00	0.00	0.65	0.98	1.00	1.00	0.75
size_of_code	0.42	0.40	0.42	1.00	0.84	0.84	0.84	0.84
ratio_malicious_api_imported	0.71	0.72	1.00	1.00	0.99	0.94	0.71	1.00
number_malicious_api_imported	0.61	0.62	0.94	0.94	0.93	1.00	0.61	0.94

Useful for detection

Useless for detection

Impact on features (6)

Alteration “move entry point to new section” (now fixed and improved) has a high impact on features “byte X after the EP”

Infogain comparison (reference = Not-packed)

	AllFill	AllRaw	Permissions	MoveEP	RawSize	Fill_0	AddApi	Rename
ratio_standard_sections	0.21	0.21	0.15	1.00	0.77	0.77	0.88	0.51
number_wx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
number_rwx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
byte_0_after_ep	0.23	0.23	0.23	0.23	1.00	0.98	1.00	1.00
number_non_standard_sections	0.18	0.18	0.12	0.88	0.71	0.72	1.00	0.63
ep_ratio	0.33	0.45	0.46	0.27	0.99	0.65	1.00	1.00
byte_X_after_ep	0.25	0.24	0.25	0.24	0.98	0.90	1.00	1.00
number_rx_sections	0.31	0.31	0.31	0.86	0.99	1.00	0.86	1.00
number_x_sections	0.33	0.33	0.33	0.86	0.99	1.00	0.86	1.00
is_x_section_not_code	0.01	0.01	0.01	1.00	0.99	1.00	0.98	1.00
entropy_section_with_ep	0.74	0.73	0.75	0.51	1.00	1.00	0.99	1.00
is_code_section_not_present	0.00	0.00	0.00	0.00	0.99	1.00	1.00	0.80
max_ratio_rdata	0.74	0.68	0.57	0.47	1.00	0.92	0.83	1.00
number_sections_vsize>dsize	0.52	1.00	1.00	1.00	1.00	0.52	1.00	1.00
min_ratio_rdata	0.69	1.00	0.46	0.46	0.95	0.69	0.46	0.46
is_ep_not_in_code_section	0.01	0.01	0.01	0.68	0.98	1.00	1.00	0.87
is_ep_not_in_standard_section	0.00	0.00	0.00	0.65	0.98	1.00	1.00	0.75
size_of_code	0.42	0.40	0.42	1.00	0.84	0.84	0.84	0.84
ratio_malicious_api_imported	0.71	0.72	1.00	1.00	0.99	0.94	0.71	1.00
number_malicious_api_imported	0.61	0.62	0.94	0.94	0.93	1.00	0.61	0.94

Useful for detection

Useless for detection

Impact on features (7)

New “edit raw size” alteration seems less effective than “fill sections with zeros”

Infogain comparison (reference = Not-packed)

	AllFill	AllRaw	Permissions	MoveEP	RawSize	Fill_0	AddApi	Rename
ratio_standard_sections	0.21	0.21	0.15	1.00	0.77	0.77	0.88	0.51
number_wx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
number_rwx_sections	0.00	0.00	0.00	0.80	0.79	0.80	1.00	0.80
byte_0_after_ep	0.23	0.23	0.23	0.23	1.00	0.98	1.00	1.00
number_non_standard_sections	0.18	0.18	0.12	0.88	0.71	0.72	1.00	0.63
ep_ratio	0.33	0.45	0.46	0.27	0.99	0.65	1.00	1.00
byte_X_after_ep	0.25	0.24	0.25	0.24	0.98	0.90	1.00	1.00
number_rx_sections	0.31	0.31	0.31	0.86	0.99	1.00	0.86	1.00
number_x_sections	0.33	0.33	0.33	0.86	0.99	1.00	0.86	1.00
is_x_section_not_code	0.01	0.01	0.01	1.00	0.99	1.00	0.98	1.00
entropy_section_with_ep	0.74	0.73	0.75	0.51	1.00	1.00	0.99	1.00
is_code_section_not_present	0.00	0.00	0.00	0.00	0.99	1.00	1.00	0.80
max_ratio_rdata	0.74	0.68	0.57	0.47	1.00	0.92	0.83	1.00
number_sections_vsize>dsize	0.52	1.00	1.00	1.00	1.00	0.52	1.00	1.00
min_ratio_rdata	0.69	1.00	0.46	0.46	0.95	0.69	0.46	0.46
is_ep_not_in_code_section	0.01	0.01	0.01	0.68	0.98	1.00	1.00	0.87
is_ep_not_in_standard_section	0.00	0.00	0.00	0.65	0.98	1.00	1.00	0.75
size_of_code	0.42	0.40	0.42	1.00	0.84	0.84	0.84	0.84
ratio_malicious_api_imported	0.71	0.72	1.00	1.00	0.99	0.94	0.71	1.00
number_malicious_api_imported	0.61	0.62	0.94	0.94	0.93	1.00	0.61	0.94

Useful for detection

Useless for detection

Impact on detection (1)

Datasets

- One per alteration
- Baseline
- All alterations combined

Applying detection on a given dataset with a given detector


```
$ detector baseline_upx --binary -d die -m
100.00%,100.00%,100.00%,100.00%
```

4. Experiments & Results

Impact on detection (2)

8 common static detectors

Dataset	Bintropy	DIE	Manalyze	PEiD	PePack	PyPackerDetect	Reminder	RetDec
not-packed	100.0%	100.0%	100.0%	99.5%	98.6%	83.8%	100.0%	99.5%
baseline	69.5%	100.0%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
add_api	66.3%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
rename_sections	70.0%	99.8%	11.0%	100.0%	100.0%	97.2%	50.5%	100.0%
add_section	0.0%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
fill_sections	56.5%	81.6%	47.7%	82.2%	82.8%	100.0%	18.2%	83.50%
edit_raw_size	70.4%	97.7%	47.7%	97.8%	100.0%	100.0%	50.3%	97.8%
move_ep_original	70.0%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	65.5%
move_ep_deadcode	68.3%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	50.0%
change_permissions ☆	11.2%	0.3%	0.0%	0.2%	68.8%	49.0%	0.0%	19.2%
perm_fill_api (All)	8.2%	0.2%	0.0%	0.0%	38.3%	0.0%	0.0%	19.5%

Similar datasets than for the previous experiment

Baseline

Single alterations

Super alteration

ALL alterations

- Baseline
- No impact or negative
- Very little impact
- Significant impact
- High impact (< 10%)

Comparison of **Accuracy** for 8 different static detectors

Impact on detection (3)

Not-packed dataset has decent detection rates, yet some detectors have false positives

Dataset	Bintropy	DIE	Manalyze	PEiD	PePack	PyPackerDetect	Reminder	RetDec
not-packed	100.0%	100.0%	100.0%	99.5%	98.6%	83.8%	100.0%	99.5%
baseline	69.5%	100.0%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
add_api	66.3%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
rename_sections	70.0%	99.8%	11.0%	100.0%	100.0%	97.2%	50.5%	100.0%
add_section	0.0%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
fill_sections	56.5%	81.6%	47.7%	82.2%	82.8%	100.0%	18.2%	83.50%
edit_raw_size	70.4%	97.7%	47.7%	97.8%	100.0%	100.0%	50.3%	97.8%
move_ep_original	70.0%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	65.5%
move_ep_deadcode	68.3%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	50.0%
change_permissions ☆	11.2%	0.3%	0.0%	0.2%	68.8%	49.0%	0.0%	19.2%
perm_fill_api (All)	8.2%	0.2%	0.0%	0.0%	38.3%	0.0%	0.0%	19.5%

Baseline
Single alterations
Super alteration
ALL alterations

- Baseline
- No impact or negative
- Very little impact
- Significant impact
- High impact (< 10%)

Baseline dataset (packed samples only) shows poor detection rates for Bintropy, Manalyze and REMINDer)

Comparison of **Accuracy** for 8 different static detectors

Impact on detection (4)

Alteration “fill sections with zeros” has a great impact on REMINDer (heuristic relies on the entropy of the EP section)

Dataset	Bintropy	DIE	Manalyze	PEiD	PePack	PyPackerDetect	Reminder	RetDec
not-packed	100.0%	100.0%	100.0%	99.5%	98.6%	83.8%	100.0%	99.5%
baseline	69.5%	100.0%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
add_api	66.3%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
rename_sections	70.0%	99.8%	11.0%	100.0%	100.0%	97.2%	50.5%	100.0%
add_section	0.0%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
fill_sections	56.5%	81.6%	47.7%	82.2%	82.8%	100.0%	18.2%	83.50%
edit_raw_size	70.4%	97.7%	47.7%	97.8%	100.0%	100.0%	50.3%	97.8%
move_ep_original	70.0%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	65.5%
move_ep_deadcode	68.3%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	50.0%
change_permissions ☆	11.2%	0.3%	0.0%	0.2%	68.8%	49.0%	0.0%	19.2%
perm_fill_api (All)	8.2%	0.2%	0.0%	0.0%	38.3%	0.0%	0.0%	19.5%

Baseline

Single alterations

Super alteration

ALL alterations

- Baseline
- No impact or negative
- Very little impact
- Significant impact
- High impact (< 10%)

Comparison of **Accuracy** for 8 different static detectors

4. Experiments & Results

Impact on detection (5)

Even though new "move EP to new section" gives slightly better results, now it does not break executables anymore

Dataset	Bintropy	DIE	Manalyze	PEiD	PePack	PyPackerDetect	Reminder	RetDec
not-packed	100.0%	100.0%	100.0%	99.5%	98.6%	83.8%	100.0%	99.5%
baseline	69.5%	100.0%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
add_api	66.3%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
rename_sections	70.0%	99.8%	11.0%	100.0%	100.0%	97.2%	50.5%	100.0%
add_section	0.0%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
fill_sections	56.5%	81.6%	47.7%	82.2%	82.8%	100.0%	18.2%	83.50%
edit_raw_size	70.4%	97.7%	47.7%	97.8%	100.0%	100.0%	50.3%	97.8%
move_ep_original	70.0%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	65.5%
move_ep_deadcode	68.3%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	50.0%
change_permissions ☆	11.2%	0.3%	0.0%	0.2%	68.8%	49.0%	0.0%	19.2%
perm_fill_api (All)	8.2%	0.2%	0.0%	0.0%	38.3%	0.0%	0.0%	19.5%

Baseline

Single alterations

Super alteration

ALL alterations

- Baseline
- No impact or negative
- Very little impact
- Significant impact
- High impact (< 10%)

Comparison of **Accuracy** for 8 different static detectors

Impact on detection (6)

Super-alteration “change permissions” gives the best results while not breaking executable samples

Dataset	Bintropy	DIE	Manalyze	PEiD	PePack	PyPackerDetect	Reminder	RetDec
not-packed	100.0%	100.0%	100.0%	99.5%	98.6%	83.8%	100.0%	99.5%
baseline	69.5%	100.0%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
add_api	66.3%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
rename_sections	70.0%	99.8%	11.0%	100.0%	100.0%	97.2%	50.5%	100.0%
add_section	0.0%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
fill_sections	56.5%	81.6%	47.7%	82.2%	82.8%	100.0%	18.2%	83.50%
edit_raw_size	70.4%	97.7%	47.7%	97.8%	100.0%	100.0%	50.3%	97.8%
move_ep_original	70.0%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	65.5%
move_ep_deadcode	68.3%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	50.0%
change_permissions ☆	11.2%	0.3%	0.0%	0.2%	68.8%	49.0%	0.0%	19.2%
perm_fill_api (All)	8.2%	0.2%	0.0%	0.0%	38.3%	0.0%	0.0%	19.5%

Baseline

Single alterations

Super alteration

ALL alterations

- Baseline
- No impact or negative
- Very little impact
- Significant impact
- High impact (< 10%)

Comparison of **Accuracy** for 8 different static detectors

4. Experiments & Results

Impact on detection (7)

Super-alteration “change permissions” combined with “fill sections with zeros” and “add API imports” gives even better results but breaks executable samples

Dataset	Bintropy	DIE	Manalyze	PEiD	PePack	PyPackerDetect	Reminder	RetDec
not-packed	100.0%	100.0%	100.0%	99.5%	98.6%	83.8%	100.0%	99.5%
baseline	69.5%	100.0%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
add_api	66.3%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
rename_sections	70.0%	99.8%	11.0%	100.0%	100.0%	97.2%	50.5%	100.0%
add_section	0.0%	99.8%	47.7%	100.0%	100.0%	100.0%	50.5%	100.0%
fill_sections	56.5%	81.6%	47.7%	82.2%	82.8%	100.0%	18.2%	83.50%
edit_raw_size	70.4%	97.7%	47.7%	97.8%	100.0%	100.0%	50.3%	97.8%
move_ep_original	70.0%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	65.5%
move_ep_deadcode	68.3%	17.0%	47.7%	0.0%	28.8%	99.5%	0.0%	50.0%
change_permissions ☆	11.2%	0.3%	0.0%	0.2%	68.8%	49.0%	0.0%	19.2%
perm_fill_api (All)	8.2%	0.2%	0.0%	0.0%	38.3%	0.0%	0.0%	19.5%

Baseline
Single alterations
Super alteration
ALL alterations

- Baseline
- No impact or negative
- Very little impact
- Significant impact
- High impact (< 10%)

Comparison of **Accuracy** for 8 different static detectors

Outline

1. Introduction
2. Background
3. Adversarial Tool
4. Experiments & Results
5. Conclusion

- Contribution
- Future work

Contribution

Packing Box

- ✓ Reviewed alterations for functionality preservation
- ✓ 2 fixed, 1 improved, 2 new alterations

NotPacked++

- ✓ New open source adversarial tool
- ✓ Selection of functional and relevant alterations
- ✓ Tested in a realistic adversarial setting – can be used in design phase of security tools

Future work

- Parametric study on dead code injection
- New approaches to moving the EP (i.e. using the relocation table)
- Obfuscation of the unpacking stubs
- Impact on malware analysis (e.g. using VirtualProtect)
- Extension to other executable formats (ELF, MachO)

NotPacked++: Evading Static Packing Detection

black hat ARSENAL

DECEMBER 9-12, 2024

EXCEL LONDON / UNITED KINGDOM

awesome Excutable Packing

Awesome list gathering our whole bibliography and many other references to documentation, tools, etc.

PACKING BOX

Dataset of Packed ELF

Ready-to-use dataset of packed and not-packed ELF files

Dataset of Packed PE

Ready-to-use dataset of packed and not-packed PE files (enriched version of Choi's dataset)

bintropy

Entropy-based tool inspired from the study of Lyda et al. in 2007

REsponse tool for Malware INDication

Heuristic-based tool inspired from the study of Han et al. in 2009

PY PACKER DETECT

Operationalized fork of https://github.com/cylance/PyPackerDetect

pe peid

Python fork of the popular tool, PEiD

NotPacked++

Attack tool for altering packed samples so that they evade static packing detection

DSFF

Custom exchange format for datasets (supports conversion to ARFF, CSV, Packing-Box dataset)

Library for getting samples from multiple malware databases