Packer Classifier Based on PE Header Information

Qiao Jin, Jiayi Duan, Shobha Vasudevan, Michael Bailey
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

Motivation and Approach

- Goal: Establish an efficient binary packer classifier with PE-header data mining, unknown packer detection
- Technique:
  - PE-header feature extraction
  - Build up binary program clustering pattern parallel to their packer type
  - Generate an efficient binary classifier which is able to correctly tell whether a binary is packed and which packer is used

Background

- Malware: programs designed to control operating systems sources without the explicit permission of the user
- Binary packer:
  - originally used to compress binary files for transmission over slow modem links
  - widely used in malware manufacturing
  - often destroys the ability of antivirus to find the string of bytes representing the signature within the new, compressed binary, leaving the effectiveness of antivirus severely compromised
- An efficient packer classifier can help malware detection
- More knowledge about packers' syngen & working principle

Previous Work

- Unpacking Research
  - Omniunpack: uses dynamic analysis to create packed versions of the packed binaries
  - Signature-based packer classifier: Sigbuster, Peid, etc.
- There are already many evading methods: Disguised entry point, Vmprotect, etc.

PE-header Info Extraction

- PE-header describes many aspects of binaries.
- Different packers leave different PE-header styles
- PE-header patterns can probably be used on a binary file to detect whether a packer is used and which packer it is

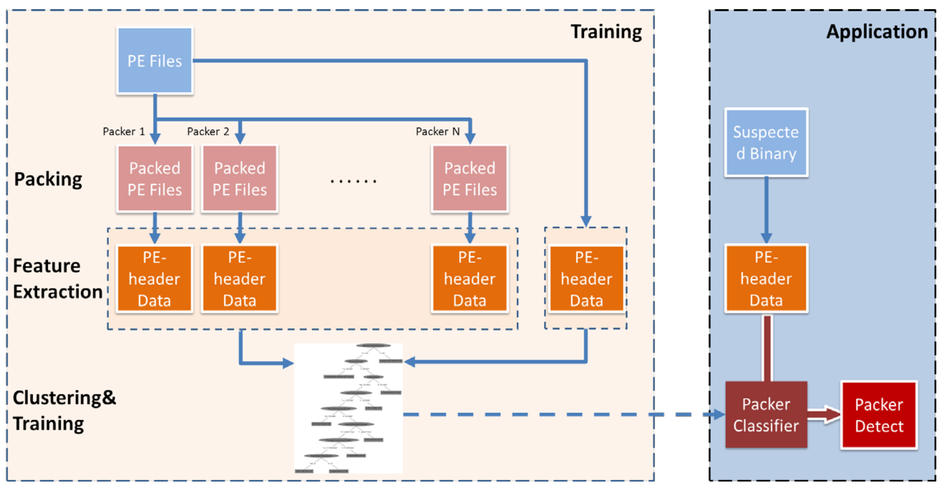


Figure 1: Binary Packer Classifier Model

Clustering Approach

- We hypothesize that programs generated from the same packer share common properties
- Clustering PE-header attributes of programs can help find these common properties
- Previously unseen clusters or binaries that do not fit into existing clusters potentially represent new packers

Packer Classifier

- A packer classifier can be built using supervised learning with labeled training data.
- Training dataset: prepare a set of executable files, pack them with different types of packers, then label them with the names of packers
- Train a classifier with training dataset
- With this classifier, a new program can be identified as packed with a specified packer or not packed

Experiments

- Experiment data: 15602 PE-header entries across 3 kinds of binaries
- The accuracy seems to be satisfying
- Detection rules are simpler
  ---->better efficiency

Table 1:Experiment Result

Class	Precision	Recall	F-Measure
Unpacked	0.985	0.998	0.992
UPX	0.997	0.985	0.991
Upack	0.997	0.977	0.987
Overall	0.991	0.991	0.991

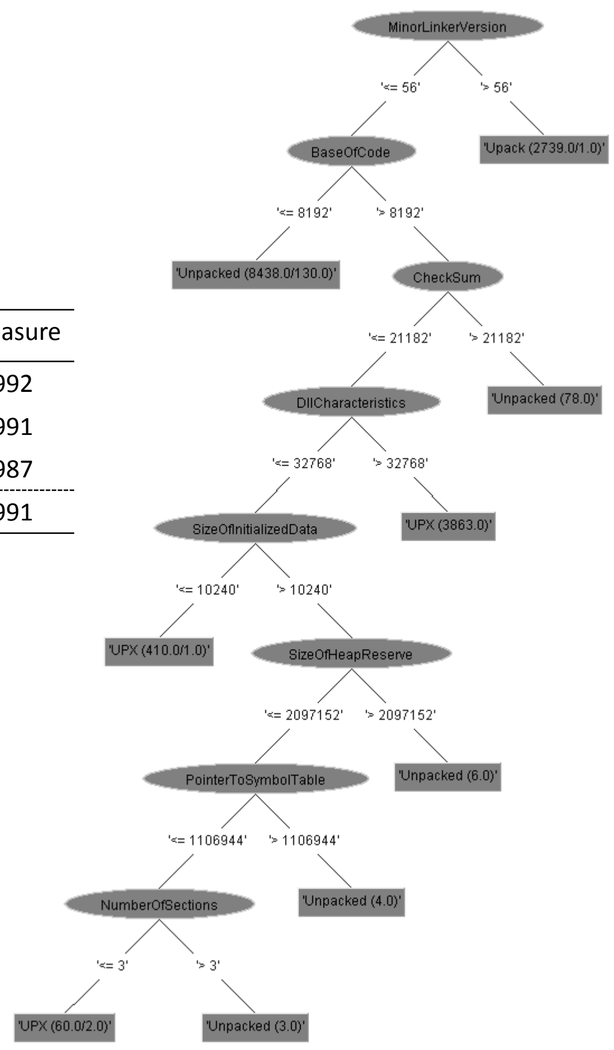


Figure 2: A Binary Packer Classifier Instance

http://hot-sos.org/

The Science of Security initiative is funded by the National Security Agency.