Unacceptable Behavior: Robust PDF Malware Detection Using Abstract Interpretation

Alexander Jordan[1], François Gauthier[1], Behnaz Hassanshahi[1], David Zhao[2]

[1]{alexander.jordan, francois.gauthier, behnaz.hassahshahi}@oracle.com, [2]d-z@outlook.com
[1]Oracle Labs, [2]University of Sydney

1. Abstract

The popularity of the PDF format and the rich JavaScript environment that PDF viewers offer make PDF documents an attractive attack vector for malware developers. Because machine learning-based approaches are subject to adversarial attacks that mimic the structure of benign documents, we propose to detect malicious code inside a PDF by statically reasoning about its *possible behaviours* using abstract interpretation. A comparison with state-of-the-art PDF malware detection tools shows that our conservative abstract interpretation approach achieves similar accuracy, is more resilient to evasion attacks, and provides explainable reports.

2. Motivating example

```
function urpl(sc) {
    var keyu = "%u";
    var re = /XY/g;
    sc = sc.replace(re, keyu);
    return sc;
}
var unes = unescape
var pGvRIJZpqdN
for (i = 0; i < 18000; i++)
    pGvRIJZpqdN = pGvRIJZpqdN + 0x77;
var s = "XY104CXY106FXY1072XY1065XY106D" +
        "XY1020XY1061XY1064XY1069XY10…";
pGvRIJZpqdN = unes(urpl(s));
```

Attackers can obfuscate their payloads in countless ways and embed it in benign code to evade ML-based detectors, an attack known as *reverse mimicry*.

3. Evaluation

Comparison to state-of-the-art

Tool	FP Rate	Recall	Accuracy
Slayer [3]	2.99%	99.23%	97.89%
Hidost [4]	1.53%	99.67%	98.95%
SAFE-PDF [2]	2.70%	99.93%	98.34%

SAFE-PDF achieves comparable accuracy to state-of-the-art malware PDF detectors.

Resilience to evasive PDFs

Obfuscation	Slayer	Hidost	SAFE-PDF
O1, O2	✓	✓	✓
O1, O3	✗	✓	✓
O1, O2, O3	✗	✓	✓
O1, O4	✗	✗	✓
O1, O2, O4	✗	✗	✓
O1, O2, O4, O5	✗	✗	✓
O6	✓	✓	✓
O1, O2, O6	✓	✓	✓
O1, O2, O3, O5, O6	✗	✓	✓
Reverse mimicry + O1-O6	✗	✗	✓

Carmony et al. [1] crafted malicious PDF documents that use obfuscation (O1-O6) and reverse mimicry to evade malware detectors. *SAFE-PDF* detects them all.

Explainability of *SAFE-PDF* reports

Report Cause	Count	Percentage
Malicious behavior	8655	88.92%
Unexpected behavior	709	7.28%
Other	369	3.80%

96.2% of *SAFE-PDF*'s detection reports highlight malicious or unexpected code behaviours.

Conclusion

The goal of any malware is to execute a specific set of malicious operations on its host. Because abstract interpretation reasons about *semantics*, it can detect, report, and explain such operations despite obfuscations. This is the first study to demonstrate the applicability of abstract interpretation for PDF malware detection and we believe that it could be used alongside other detectors to capture advanced evasive malware.

References

[1] Curtis Carmony, Xunchao Hu, Heng Yin, Abhishek Vasisht Bhaskar, and Mu Zhang. Extract Me If You Can: Abusing PDF Parsers in Malware Detectors. In *NDSS'16*, 2016.

[2] Alexander Jordan, François Gauthier, Behnaz Hassanshahi, and David Zhao. Unacceptable Behavior: Robust PDF Malware Detection Using Abstract Interpretation. In *PLAS'19*, pages 19–30, 2019.

[3] Davide Maiorca and Giorgio Giacinto. Looking at the Bag is not Enough to Find the Bomb: An Evasion of Structural Methods for Malicious PDF Files Detection. *ASIA CCS'13*, pages 119–129, 2013.

[4] N Srndic and Pavel Laskov. Detection of Malicious PDF Files Based on Hierarchical Document Structure. *NDSS'13*, 2013.