**基于机器学习的PDF恶意软件检测与分类器对抗性学习研究**

凤娇，江纬和杨育斌A1，3，作者B2

（1.单位名1,城市名1 邮政编码1）

（2.单位名2,城市名2 邮政编码2）

**英文简介还没有看-** Nowadays, with the highly rapid development of information, it is becoming more and more important to detection based on malicious documents (such as PDF). A attack for document type is targeted and more vulnerable . Because of the diversity of its structure and form, attackers have various accesses and its technique is hard to be retrieved. Therefore, this research project aims to provide document (PDF) classifier based on AI for industry and academia. The study has collected 200,000 samples data and analyzed training process. This results show that the accuracy for training model highly rated 99.82% while the false positives only rated 0.01%; and through the study of antagonism, the model has certain ability to resist attacks and good robustness. This model can be widely used in the Terminal Security Products or mail servers when a file is detected at a millisecond level. It is a feasible model for AI is not limited to use in PDF document classifier, also can be extended the structure of documents such as word, excel etc.

**Key Words：**AI Security; Machine Learning; Maldoc Detection; Adversarial ML

**摘 要：**在当今信息安全领域，基于恶意文档（如pdf）的恶意软件检测越发重要，基于文档的攻击通常具有针对性（targeted attack），因此更容易攻击成功。加上其文件结构的多样性，攻击手段变得丰富且容易隐藏。此研究项目的目标旨在为工业界和学术界提供一个基于AI的文档（PDF）恶意软件分类器原型。直到目前为止，我们收集了20万个文档，对其进行解析（parsing），并训练生成模型。我们模型的准确率达到99.82%，而误报率却只有0.01%，单个文件的检测时间为毫秒。进一步地，我们通过使用对抗性学习方法，用实验数据证明模型具有很好的抗攻击（样本逃逸）能力和鲁棒性。此模型可广泛部署在终端安全产品，邮件服务器等，这些均是非常有意义的应用场景。

**关键词：** 人工智能之于信息安全；机器学习；恶意文件检测；对抗性学习

作者简介：作者1姓名（19XX－ ），性别，籍贯，职称，最后学位（或在读学历），主要研究方向，通信地址和邮编，手机，电话，Email;作者2姓名（19XX－ ），性别，籍贯，职称，最后学位（或在读学历），主要研究方向，通信地址和邮编，手机，电话，Email。项目基金：名称、编号

1. **简介**

随着时间的推移，PDF规格和样式都在变得丰富。增加的脚本功能使文档与可执行文件几乎能以相同的方式工作，如连接到Internet的能力，运行进程以及与其他文件/程序进行交互等。这种复杂性的增长为攻击者提供了更多的武器来发动黑客攻击，并且能更灵活地隐藏恶意有效载荷（例如加密，隐藏为图像，字体或Flash内容等）并逃逸检测。由于企业和个人普遍对此类安全漏洞反应迟缓，安全意识不足，导致大量的用户系统未采用最新的浏览器插件（例如，Adobe的 Flash Player 和 Acrobat Reader）修补程序进行更新，导致这些攻击取得成功。

在2012年，开创性的漏洞攻击工具Blackhole专门针对Java，PDF和Flash文件。2013年，通过网络传输的文档类型的文件，主要针对PDF和Flash文件。在2014年发现的24个0-day中，有16个是针对Adobe Reader和Flash Player的。在通用漏洞与披露（CVE）也可以明显观测到，从2015年开始关于Adobe Reader发现的漏洞呈现高增长态势，这给基于PDF的文档攻击敞开了大门。

针对于近几年遇到的多种基于文档的攻击，传统的PDF恶意软件检测方法有基于Shellcode的检测[58]、基于签名的检测[31]等。这些方法均存在识别率不高、无法及时更新恶意代码等普遍问题。基于机器学习的PDF恶意软件检测为此提供了崭新的方向，通过把PDF文档内容、结构、附加信息等作为特征，综合利用随机森林、SVM、决策树等AI算法构建PDF恶意软件分类器，此方向的研究有了突破性进展。

本文的主要贡献如下：

* 一个全新的PDF数据集，总样本数达到约二十万，其中恶意样本173036个，正常样本28332个。无论在数量还是质量上，均达到业界领先水平；
* 精心选取了一套静态特征集（133个）以用于刻画PDF恶意软件形象（Profile），使其能快速有效的区分恶意与良性样本；
* 模型准确度高达99.82%，误报率0.01%，单个文件检测时间维持在毫秒水平，领先业界；
* 使用逃逸样本生成算法生成7000个恶意软件的变种以用于对抗性学习研究；
* 成功使用变异样本对分类器发动逃逸攻击，分类器根据攻击进行自我修复，重新训练得出一个鲁棒性强，抗逃逸能力强的模型

参考文献:

1. Davide Balzarotti, Marco Cova, Christoph Karlberger, Christopher Kruegel, Engin Kirda, and Giovann Vigna. Efficient Detection of Split Personalities in Malware. In Proceedings of the 17th Annual Network and Distributed System Security Symposium (NDSS), San Diego, CA, February–March 2010
2. Curtis Carmony, Mu Zhang, Xunchao Hu, Abhishek Vasisht Bhaskar, and Heng Yin. Extract Me If You Can: Abusing PDF Parsers in Malware Detectors. In Proceedings of the 2016 Annual Network and Distributed System Security Symposium (NDSS), San Diego, CA, February 2016
3. Nedim ˇ Srndic and Pavel Laskov. Detection of Malicious Pdf Files Based on Hierarchical Document Structure. In 20th Network and Distributed System Security Symposium (NDSS), 2013
4. Nedim Srndic and Pavel Laskov. Practical Evasion of a Learning- Based Classifier: A Case Study. In Proceedings of the 35th IEEE Symposium on Security and Privacy (Oakland), San Jose, CA, May 2014
5. Nedim ˇ Srndic and Pavel Laskov. Mimicus: A Library for Adversarial Classifier Evasion. <https://github.com/srndic/mimicus>.
6. Nedim Šrndic and Pavel Laskov . Hidost: a static machine-learning-based detector of malicious files, Šrndi′c and Laskov EURASIP Journal on Information Security (2016) 2016
7. Weilin Xu, Yanjun Qi, and David Evans. Automatically Evading Classifiers: A Case Study on PDF Malware Classifiers. In Proceedings of the 2016 Annual Network and Distributed System Security Symposium (NDSS), San Diego, CA, February 2016. http://evademl.org/
8. VirusTotal. Free Online Virus, Malware and URL Scanner.https://www.virustotal.com/.
9. Stephan Chenette. Malicious Documents Archive for Signature Testing and Research - Contagio Malware Dump. http://contagiodump.blogspot.de/2010/08/ malicious-documents-archive-for.html.
10. D. Maiorca, D. Ariu, I. Corona, and G. Giacinto, “A Structural and Content-Based Approach for a Precise and Robust Detection of Malicious PDF Files”, in Proceedings of the 1st International Conference on Information Systems Security and Privacy. ScitePress Digital Library,2015, pp. 27-36.
11. Charles Smutz and Angelos Stavrou. Malicious PDF Detection Using Metadata and Structural Features. Technical report, 2012.
12. C. Cortes and V. Vapnik. Support vector networks. Machine Learning, 20:273–297, 1995.
13. <https://www.symantec.com/content/dam/symantec/docs/reports/istr-22-2017-en.pdf>
14. Igino Corona, Davide Maiorca, Davide Ariu, and Giorgio Giacinto. Lux0R: Detection of Malicious PDF-embedded JavaScript Code through Discriminant Analysis of API References. In Proceedings of the Artificial Intelligent and Security Workshop (AISec), 2014.PDFrate
15. <https://github.com/srndic/mimicus>

[34] DavideMaiorca, DavideAriu, IginoCorona, andGiorgioGiacinto. An Evasion Resilient Approach to the Detection of Malicious PDF Files. In Proceedings of the International Conference on Information Systems Security and Privacy (ICISSP), 2016.

[35] DavideMaiorca,IginoCorona,andGiorgioGiacinto.Lookingat the Bag is not Enough to Find the Bomb: An Evasion of Structural Methods for Malicious PDF Files Detection. In Proceedings of the 8th ACM Symposium on Information, Computer and Commu- nications Security (ASIACCS), Hangzhou, China, March 2013.

[53] Nedim Srndic and PavelLaskov. Practical Evasion of a Learning- Based Classifier: A Case Study. In Proceedings of the 35th IEEE Symposium on Security and Privacy (Oakland), San Jose, CA, May 2014.

[27] Nedim ˇ Srndic and Pavel Laskov. Mimicus: A Library for Adversarial Classifier Evasion. https://github.com/srndic/mimicus.

[29] Nedim ˇ Srndic and Pavel Laskov. Practical Evasion of a Learning-Based Classifier: A Case Study. In 35th IEEE Symposium on Security and Privacy (Oakland), 2014.

[30] Meng Xu and Taesoo Kim, Georgia Institute of Technology:PlatPal: Detecting Malicious Documents with Platform Diversity . 26th USENIX Security Symposium 2017