1. **简介**

随着时间的推移，PDF规格和样式变得更加丰富。新版本增加脚本的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，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攻击，并且能更灵活地隐藏恶意有效载荷，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版本进行更新，最终使这些攻击取得成功。

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恶意文件检测方法有基于Shellcode的检测[15]、基于签名的检测[13]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，最开始使用机器学习的方法是2011年Nedim Srndic等人，主要对javascript进行提取分析，之后很多研究者基于内容和结构提取文件的静态特征[8],或是基于元数据与结构提取文件特征[24]使用SVM和决策树对文件进行分类，经过AI算法调优后，可以达到很好的效果。我们的方法结合他们所提到的一些特征进行提取，包含有结构，内容，javascript,元数据信息等。并通过研究发现，我们使用随机森林的算法比使用SVM 算法的准确率要高。同时我们也考虑到了AI模型的安全问题，Nedim Srndic等人在后来的研究中关注的不止是模型的准确率，而是AI模型的抗逃逸与鲁棒性，他们在IEEE会议上[4]针对AI模型的逃逸提出了几种假设，最后成功逃逸分类器，在我们的实验当中同时也使用了其中的4种方法来验证我们的模型是否具有这种抗逃逸的能力，我们通过对模型的一些特征选取和算法调优，发现可以使之前逃逸的一部分样本被检测出来，说明我们的模型有一定的抗逃逸能力。

本文的主要贡献如下：

* 一个PDF数据集，总样本数201368 个，其中恶意样本173036个，正常样本28332个；
* 精心选取了一套静态特征集（133个）以用于刻画PDF恶意文档形象，以用于区分恶意与良性样本；
* 模型准确率高达99.82%，误报率0.01%，单个文件检测时间维持在毫秒水平；
* 成功使用自己生成的变种病毒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。