Extract Me If You Can:

Abusing PDF Parsers in Malware Detectors

Curtis Carmony

Syracuse University

curtis.carmony@gmail.com

## 中文标题：

滥用恶意软件检测器中的PDF解析器

## Questions:

## 一句话概述：

## 中文摘要：

由于PDF格式的普及和Adobe Reader的不断发展，恶意PDF的检测仍然是一个问题。所有现有的检测技术在一定程度上依赖于PDF解析器，而PDF格式的复杂性为解析器混淆留下了充足的空间。为了量化这些解析器与Adobe Reader之间的差异，我们通过在大多数自动二进制分析技术识别的位置直接使用Adobe Reader创建参考JavaScript提取器。通过将此参考提取器的输出与从VirusTotal获取的大型数据集上的几个开源JavaScript提取器的输出进行比较，我们能够识别数百个样本，其中现有的提取器无法从中提取JavaScript。通过分析这些样本，我们能够确定这些提取器中的每一个的几个弱点。基于这些经验教训，我们对恶意PDF样本进行了几次混淆处理，可以成功回避所有经过测试的恶意软件检测器。我们称这种逃避技术为a*PDF解析器混淆攻击*。最后，我们演示了参考JavaScript提取器提高了现有基于JavaScript的分类器的准确性，以及如何使用它来缓解实际设置中的这些分析器限制。

## 亮点解读

       提出了一种大多数自动化的平台无关点识别技术，用于正确识别与Adobe Reader中用于开发参考JavaScript提取器的JavaScript解析和执行相关的抽头点。

•          使用我们的参考提取器，我们系统地评估现有JavaScript抽取工具的缺点。我们已经识别出数百个PDF样本（无论是良性的还是恶意的），而现有的提取器无法从中提取JavaScript。我们手动调查其中的很多，并确定其根源。

•          我们通过结合我们分析中确定的几个混淆因素构建了几个PDF解析器混淆攻击。这些逃避已被证明有效地成功避开了我们测试过的所有恶意软件检测器。

•          我们讨论几种缓解技术。特别是，我们证明，使用我们的参考JavaScript提取器，现有分类器的检测率在我们的样本集中从68％显着增加到96％，并为参考提取器提供了可能的部署方案。

## 论文总结

在本文中，我们对一种称为*PDF解析器混淆攻击*的新规避技术进行了系统研究，该技术旨在混淆恶意软件检测器中的PDF解析器以逃避检测。为了进行系统的研究，我们开发了一个*参考JavaScript抽取器*，直接使用Adobe Reader并提供了一种用于开发它的大部分自动化技术。通过研究参考提取器和现有提取器之间的这些差异，我们发现了几个新的混淆，并进一步量化了它们在JavaScript提取器和恶意软件检测器的解析器混淆攻击中的影响。通过结合其中几个混淆，我们制作了一个可以成功回避*所有问题*的恶意PDF评估恶意软件检测器，包括基于签名的，基于结构/元数据的和基于JavaScript的检测器。为解决解析器混淆攻击，我们讨论了几种缓解技术。特别是，我们证明，使用我们的参考JavaScript提取器，现有分类器的检测率已经从我们的样本集的68％显着增加到96％，并为参考提取器提供了可能的部署方案。