**Group Planning PDF Malware detection**

**前沿研究方向—ML Malware detection研究规划 1**

**机器学习运用-基于恶意文件检测 1**

1. **背景**

网络攻击者正在转向基于文件的恶意软件，因为用户聪明起来恶意电子邮件附件和网络链接，许多防病毒（AV）供应商。用户通常会被浏览器，电子邮件代理或AV产品更多地警告可执行文件的危险，而像PDF这样的文档由于受到静态文件的影响而受到的关注和审查要少得多，并且几乎没有什么危害。

但是，随着时间的推移，PDF规格已经改变。增加的脚本功能可以使文档以与可执行文件几乎相同的方式工作，包括连接到Internet的能力，运行进程以及与其他文件/程序进行交互。内容复杂性的增长为攻击者提供了更多的武器来发动强大的攻击，并且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）并逃避检测。

在2015年发布了137个CVE，在2016年仅有Adobe Acrobat Reader（AAR）。AAR的普及及其大型攻F击面使其成为攻击者的首要攻击目标，其次是浏览器和操作系统内核。在引入了类似Chrome的沙盒机制后，在pwn2own竞赛中，单个漏洞的价值可能高达7万美元。收集的恶意软件样本显示，许多Adobe组件已被利用，包括元素解析器和解码器[ 37 ]，字体管理器以及JavaScript引擎。系统范围的依赖性，如图形库[ 23 ]也在攻击者的雷达

2016年，Cymantec观察到大量通过恶意Office宏（W97M.Downloader和变体）和JavaScript下载程序文件（JS.Downloader和变体）分发勒索软件和网上银行威胁的电子邮件活动。 在他们之间，他们在2016年终端上的检测数量达到了700万，并且主宰了网络犯罪威胁的形势，特别是在当年下半年

1. **文件检测类型**

**PDF文件检测**