# Fengjiao研究方向规划

目录

[Fengjiao研究方向规划 1](#_Toc513753292)

[人工智能赋能文档检测 1](#_Toc513753293)

[工作概述 1](#_Toc513753294)

[目标O ： 1](#_Toc513753295)

[时间范围： 1](#_Toc513753296)

[Q2 季度KR： 1](#_Toc513753297)

# 人工智能赋能文档检测

摘要：

恶意文档（如pdf）的检测是工业界和学术界的难题，并且四大安全会议还是把重心放在pdf上，原因之一在于其**结构与形式的多样**，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，攻击者手段之丰富（如通过**电子邮件的附件传播**，**钓鱼网站加密混淆等**）（如：CVE的漏洞）。其次网络攻击者正在**转向基于文件的恶意软件**，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。

此研究项目的目标旨在为工业界（如蓝盾）和学术界提供一个基于AI的文档（PDF，word等）分类器，此分类器具有 准确度极高，误报率极低和抗逃逸能力强三大特点。通过对文档领域特征的挖掘，并结合数据科学中的复杂建模，模型成功被构建。此模型可广泛应用于蓝盾多条产品线上，如在**终端安全产品**上加入此模组，又或者**在邮件服务器**上加入此模组，均是非常有意义的应用场景，AI文档分类器不仅限于pdf，如可扩展到word，execl，ppt等有结构的文档。

## 工作概述

● 收集 现有 数据集，使 其总数据集 规模 达 **20万 级别**

● 添加 新近的，自己生成的**，变异 PDF恶意样本**约**7000个**（级别）

● 利用人工智能生成 基于文档 的检测模型 **准确率 >= 99.6%** ，误报率 <= 0.020%（**时间：单机版预测 200个 样本需要 10s ，Spark 上运行 2千多个 预测需要1.4s**）

● 阅读并收集 **10份 顶会论文**，并有读书笔记上次

● 基于人工智能的恶意软件检测引擎 的 **技术文稿** 发表于蓝盾公众号

## 目标O ：

● 实现基于AI的文档分类引擎和对抗模型，使其广泛应用于蓝盾产品线上

## 时间范围：

● 2018 Q2-Q4

## Q2 季度KR：

* 1）（基础）**加固** 分类器 的方法（参考Feature Squeezing一文）
* 2）（基础）阅读并收集 **至少 5份 顶会论文**，每周 1-2 个读书笔记分享
* 3）（中级）根据现有技术，对已经选取的**特征进行计算合并**，并且在新的模型中 优化现有特征
* 4）（中级）在5月 **完成中文论文的修稿工作**，以及图表制作，数据整理，模型训练等
* 5）（中级）研究现有攻击逃逸的技术，并针对现有的逃逸技术，训练一个 **具有抗逃逸** 的分类器 一个
* 6）（高级 ）完成 **一篇 论文** 并 **提交** 到 合适的会议 中