# **fast-flux域名检测**

# **一、背景介绍**

“Fast flux”技术是指不断改变域名和IP地址映射关系的一种技术，即在短时间内查询使用Fast flux技术部署的域名，会得到不同的结果。尽管这种技术有一些合法用途，但它也成为了网络犯罪分子非常喜欢的发起协作攻击的技术。在Honeynet首次观察到这种攻击活动后，fast-flux已被广泛用于网络钓鱼、恶意软件传播、垃圾邮件和其他与犯罪组织有关的恶意活动。

Fast flux的“fast”是指通过将DNS的Time-To-Live（TTL)设置为一个较小的值，造成DNS响应不会被长时间缓存。它可以快速地在多个设备之间快速变化，每个设备只在很短时间内使用，使得基于IP的黑名单屏蔽和删除工作变得困难。

Fast-flux的检测工作中，面临以下难点：

**1）数据严重不平衡**，真实场景中，fast-flux仅占恶意域名中的少部分，在未知域名中的比例会更低；

2）一些**正常业务**也具有类似fast-flux域名的行为，增加了检测的难度。

# **二、数据描述**

本题目提供训练集和测试集，为了方便参赛选手训练模型，train目录下分为数据集文件pdns.csv，和fastfluxt标签文件fastflux\_tag.csv，flux域名标签为**1**，非fastflux域名标签为**0**；test目录下为初赛使用的pdns数据集文件、

pdns文件为域名在pdns数据库中的查询结果，具体提供的数据及说明见下表：

A）count：解析次数

B）time\_first：该解析记录首次观察到时间

C）time\_last：该解析记录最后一次观察到时间

D）rrname：解析的域名

E）rrtype：解析的类型，具体查询DNS解析类型（如A，NS，MX，SOA...）。

F）rdata：解析结果，根据不同解析类型响应产生不同结果，比如，A记录对应的是域名解析到IP的结果。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| count | time\_first | time\_last | rrname | rrtype | rdata |
| 875 | 1398943391 | 1439353872 | advection.net. | A | ['23.23.125.143'] |
| 44 | 1502359032 | 1502770467 | advection.net. | A | ['38.68.11.57'] |
| 1638 | 1496133298 | 1502352740 | advection.net. | A | ['38.68.11.157'] |

对于域名解析的IP信息，如需要可向评委申请关于IP信息查询的云端接口。如地理位置信息的查询。

# **三、提交格式**

通过训练集训练AI模型，对测试集数据进行预测，输出csv标准格式文件，其中fastflux域名标签为**1**，非fastflux标签域名为**0**，样例如下（注：不含表头行）：

nofastflux1.com.,0

fastflux.com.,1

nofastflux2.com.,0

# **四、评分方法**

|  |  |
| --- | --- |
| **指标** | **说明** |
| TP | 表示fastflux域名被预测为fastflux域名的数目 |
| FP | 表示非fastflux域名被预测为fastflux域名的数目（误报） |
| TN | 表示非fastflux域名被预测为非fastflux域名的数目 |
| FN | 表示fastflux域名被预测为非fastflux域名的数目（漏报） |
| **最终结果评分方法** | |
| **计算公式** | precision = TP / (TP + FP)  recall = TP / (TP + FN)  score = 2 \* precision \* recall / (precision + recall) |
| **说明** | 当前不考虑准确率和召回率在实际应用中的差异，认为准确率和召回率同等重要。 |

# **五、训练集、测试集说明**

# **六、注意事项**

# **七、出题人**