DataCon 大数据安全分析比赛

DNS 方向赛题描述与提交说明

背景

假如你是某网络的管理员，近日接到告警称，网络中存在 DNS 攻击行为，希望你进行调查。现捕获到网络中的 DNS 流量，请对其中的攻击行为进行分析。

要点

DNS 是互联网中重要的基础设施之一，对网络的稳定运行有至关重要的作用。然而，由于设计缺陷，DNS 存在诸多脆弱点，因此可被利用于诸多攻击。本题对常见的 DNS 安全问题进行考查。

题目一：DNS 攻击流量识别

设计说明

模拟网络管理员的攻击分析过程。给定的流量中，包含五种 DNS 攻击流量。选手需要准确判断出五种 DNS 攻击，并说明 pcap 文件中哪些数据包是攻击流量。

数据说明

提供的流量为 pcap 格式数据包，大小为 2.6GB。

持续时间

本题的答题时间为 2019 年 4 月 1 日至 2019 年 4 月 8 日。

提交形式和文件格式

选手提交 2 个 csv 文件、Writeup、解题代码（如有），统一使用 zip 格式打包提交。注意，压缩文件请勿加密，并确保可以使用 unzip 命令无参数解压。请勿压缩文件夹。

提交文件的命名规则和内容如下：

- traffic.csv：**列名：packetid,attackid**
  - 每一行包含两列：数据包的顺序 ID **packetid**、数据包对应攻击类别代码 **attackid**。
  - 数据包中只包含五类攻击，攻击类别代码请使用数字 1、2、3、4、5，评分时忽略其他的攻击类别代码；提交的答案多于五类将只取使用 1-5 编号的类别，少于五类将默认剩余类别为空；评分与选手攻击编号顺序无关。
  - 同一个数据包不会同时涉及两种攻击；若提交答案中同一个数据包同时出现在多个攻击类别，将以第一次出现的记录为准。
  - 仅按格式提交你认为的所有攻击**请求数据包（DNS query）**，请不要提交响应数据包。
  - 选手应使用半角逗号作为列间分隔符，下同。
  - 选手的 csv 文件请使用 UTF-8 编码，并使用\n 作为换行符，下同。
- attack.csv：攻击类别描述，**列名：attackid,description**
  - 在此文件中，将简要攻击描述（例如 xx attack，2-5 词即可）和 traffic.csv 中的攻击类别代码进行对应。
  - 为避免编码问题，攻击描述请勿包含非 ASCII 字符。
- Writeup 和代码
  - 在 2019 年 4 月 8 日答题截止前最后一次提交时，请确保包含解题 Writeup，如有代码请一并提交相关代码 zip 文件。Writeup 和代码将由组委会审核并作为成绩评定依据，应详细包含解题思路、数据分析步骤、代码功能、代码运行环境和说明、每种攻击的详细信息等内容。
  - Writeup 请使用 PDF 格式。代码请打包为 zip 格式文件。
  - 未提交 Writeup 和解题代码，将影响题目最终得分。
- 示例（见 example_dns1.zip）

在如下的数据包中，判断 67、73、76、82 号数据包是攻击流量，70、79 号数据包是正常流量。并且，67、73 号数据包是同一种攻击 attack1，76、82 数据包是另一种攻击 attack2。

No.	Time	Source	Destination	Length	Protocol	Info
67	0.355902	192.168.2.193	192.168.2.1	86	DNS	Standard query 0xcf72
70	0.409498	192.168.2.193	192.168.2.1	86	DNS	Standard query 0xf9cf
73	0.442261	192.168.2.193	192.168.2.1	86	DNS	Standard query 0x9f56
76	0.462612	192.168.2.193	192.168.2.1	86	DNS	Standard query 0x1064
79	0.523110	192.168.2.193	192.168.2.1	86	DNS	Standard query 0x186c
82	0.552305	192.168.2.193	192.168.2.1	86	DNS	Standard query 0x3959

我提交的 traffic.csv 内容应为：

packetid	attackid
67	1
73	1
76	2
82	2

我提交的 attack.csv 内容应为：

attackid	description
1	dnsattack1
2	dnsattack2

我最后一次提交的 zip 文件，应该包含上面两个 csv 文件、一个 Writeup PDF 文件和代码 zip 文件（如有）。之前用于阶段评分的提交，可以只包含两个答案 csv 文件。

提交规则

选手在本题的持续时间内，提交次数不限。

主办方将不定期进行阶段评分，将队伍本题的得分以及当前得分排名范围反馈给选手。在答题截止前，选手可以根据阶段反馈情况调整答案。

本题最终得分为选手在 2019 年 4 月 8 日 24 时答题截止前**最后一次提交结果**的得分。

评分规则

该题得分由每一类得分累加而成，每一类攻击占比 20%。

选手提交的 5 个类别，将与标准答案中的 5 个类别进行比较，并对每一类分别进行指标计算和累加。选手提交的类i和标准答案中的类j，评分指标计算如下：

$$Score_{i,j} = max\left(\frac{1}{N}\cdot w\cdot\sum_{k=1}^{m}([P_k \in FTrue_j]),0\right)$$

各项符号的含义：

$FTrue_j$：标准答案的第j类

N：$FTrue_j$的长度

m：选手提交的第i类的长度

P_k：选手提交的第i类中的第k个数据包

[]：其中为判断语句。语句成立时，值为 1；不成立时，值为-1

$max()$：取最大值函数

w：积分权重

选手提交的类别，将和标准答案的类别进行全排列比较并累加。取累加平均后得分最高的一种计算结果，作为选手该题的得分。