分析攻击者的攻击目的和攻击能力

目录

一．分析和涉及思路概述

以第一题和第二题的数据集基础上，进行数据分析和处理。

1. 攻击者的攻击目的：通过分析攻击者的攻击目标和网络攻击手法，可得到攻击者的攻击社会行业的倾向，以及通过哪些网络攻击达到控制目标主机或者摧毁目标主机的目的。
   a) 攻击目标分析：以攻击者 ip 以及其社区的 ip 列表为分析主体，获取攻击社区解析过的域名、所属域名行业类别、解析次数。
   b) 网络攻击手法分析：以攻击者 ip 以及其社区的 ip 列表为分析主体，获取攻击社区流量日志中，攻击种类，给攻击种类的攻击次数。

2. 攻击者攻击能力，通过社区分析和社群发现，得到攻击者的社区关系，得到攻击者的控制主机的范围和能力（详见题目二设计文档），在攻击目标和网络攻击手法中的分析也可以达到攻击者的攻击次数和解析次数，得到攻击者的攻击力度。

二．收集的数据集

1. 基础数据集：官方给定的数据

   IP 日志，部分终端恶意样本（样本名和终端指纹），IP 绑定过域名（域名），IP 解析过的域名，360 威胁情报

2. 题目一和题目二中处理得到的数据集(位于 ip_file 中)

   攻击 ip 的关联 ip 列表 ip_data_friends.csv，

   流量日志中检测出的攻击手法和得分（logs_ip_score.csv）（按标准 WAF 规则计算）。

注释：攻击 ip 的关联 ip 列表数据由于脚本运行时间过长，考虑到现场复现的时长，此处直接给出处理后得到的数据文件。流量日志中检测出的攻击手法和得分，属于第一题的处理过程中得到的数据，考虑到复现时长，并且不属于答题范围，也直接给出结果数据文件。

三．处理数据

1．攻击者解析过的域名记录，以及域名行业类别信息

基础数据：ip 解析过的域名 dns.csv，域名和该域名的行业分类信息 domain_category.json。

处理思路：对每条 ip 解析域名记录做域名和与该域名的行业分类信息做关联，得到 ip，domain_name，domain_category 对应数据。处理详情见源码。

结果文件：ip_file\\ ip_log_dns_cate.csv

2．攻击者 ip 以及其社区 ip 的 360 威胁情报。

基础数据：360 威胁情报数据.json

处理思路：将 json 文件处理为对应的 dataframe 数据格式，将每个 ip 的属性字段提取出来。

结果文件：ip_file\\ ip_file\\ioc_data.csv

四．可视化分析

以题目二中分析得到的最具攻击能力的 ip：120.200.167.93 以及其社区 ip 列表进行分析。

网络攻击手法分析

1. 从 logs_ip_score.csv 中获取 ip：120.200.167.93 相关日志攻击数据，发现只有一条 get 记录，而且几乎没有可用数据。由此可见，该攻击者具有强烈的隐秘性，预计他主要通过控制其他主机发动网络攻击。
2. 获取社区 ip 列表的日志攻击数据，果然发现大量攻击数据。（score 表示日志中该攻击手法的攻击次数）

149	101.95.175.54	MSG	0
256	103.244.89.19	MSG	0
354	103.67.23.140	MSG	0
766	110.87.188.33	:'Detects MySQL UDF injection and other data/s...	198
767	110.87.188.33	:'Detects blind sqli tests using sleep() or be...	256
768	110.87.188.33	:'PHP Injection Attack: Low-Value PHP Function...	5660
769	110.87.188.33	:'PHP Injection Attack: PHP Open Tag Found',	84
770	110.87.188.33	:'Path Traversal Attack (/../)',	4874
771	110.87.188.33	:'Remote Command Execution: Unix Shell Express...	2331
772	110.87.188.33	:'Remote Command Execution: Wildcard bypass te...	35
773	110.87.188.33	:'SQL Comment Sequence Detected.',	1047
774	110.87.188.33	:'SQL Injection Attack',	5920

统计攻击类别：发现攻击者极度偏向注入攻击，sql 和 php 注入。猜测攻击者的攻击目的在获取数据或者想要长期控制目标主机。但涉及的攻击方法多种，远程命令执行，xss 攻击，路径遍历，也会多种数据库的攻击，可见攻击者知识面很广，能力很高。

:'PHP Injection Attack: PHP Open Tag Found',	186	0.000405
:'XSS Filter - Category 3: Attribute Vector',	191	0.000416
:'Detects blind sqli tests using sleep() or benchmark().',	256	0.000558
:'Possible XSS Attack Detected - HTML Tag Handler',	280	0.000610
:'XSS Filter - Category 1: Script Tag Vector',	292	0.000636
:'Detects MySQL UDF injection and other data/structure manipulation attempts',	586	0.001277
:'Remote Command Execution: Unix Shell Expression Found',	2455	0.005351
:'Path Traversal Attack (/../)',	5288	0.011526
:'SQL Comment Sequence Detected.',	5471	0.011925
:'Looking for basic sql injection. Common attack string for mysql, oracle and others.',	21677	0.047249
:'PHP Injection Attack: Variables Found',	94679	0.206371
:'PHP Injection Attack: High-Risk PHP Function Call Found',	98490	0.214678
:'PHP Injection Attack: Low-Value PHP Function Call Found',	101503	0.221245
:'SQL Injection Attack',	127076	0.276986

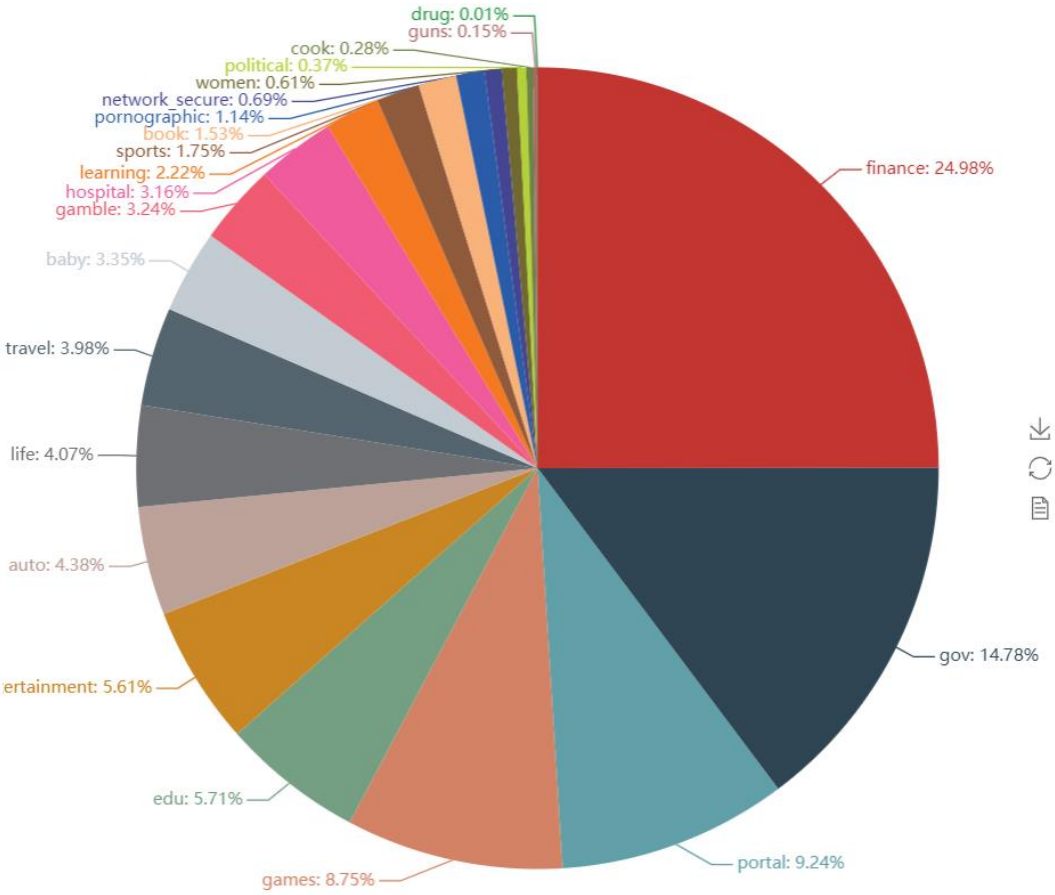
攻击目标分析

1. 从题目中所得数据 ip 解析域名以及域名类型文件 ip_log_dns_cate.csv。获得社区 ip 列表的数据。

	ip	date	domain_name	resolve_number	domain_category
0	1.180.18.194	2018.11.21	0008013e9645e0dc4213abea560dfc26.b99219cf9436a...	1	unknown
1	1.180.18.194	2018.11.21	005b20e3ad6e90c6ebecdf4954862304.a852d3a7c2f51...	2	unknown
2	1.180.18.194	2018.11.21	00d552795cf142a0c85587459c972a90.com	1	unknown
3	1.180.18.194	2018.11.21	011c660a4f29c5d72329566158dfe3b2.com	3	unknown
4	1.180.18.194	2018.11.21	011f4ab0082265dba74c8a4f0d8daf55.cu	1	unknown

2. 统计该攻击社区解析的域名记录中域名种类统计：

很直观的发现，该攻击者攻击目标主要在，经济（finance），教育（gov），门户网站（portal）



攻击者社区的情报分析

将ip和该社区ip列表与威胁情报ioc_data.csv以及第二题中得到的ip恶意样本列表，

投放恶意样本记录数，解析域名列表合并为攻击者的综合数据。

从这份数据中具备很多指标，ip 投放过的恶意样本种类，ip 解析过的域名种类，ip 关联的社区列表，这都与攻击者的攻击能力成正向关系。

包括社区中 ip 的恶意标签种类：

```
In [11]: most_friends_info.malicious_label.value_counts()

Out[11]: [DDOS, WEB_ATTACKER]                    7
         []                                      5
         [DDOS, SCANNER, WEB_ATTACKER]           4
         [DDOS]                                  3
         [SCANNER, WEB_ATTACKER]                 2
         [DDOS, SCANNER]                         1
         [DDOS, SPAM, WEB_ATTACKER]              1
         [DDOS, SCANNER, SPAM, WEB_ATTACKER]     1
         [WEB_ATTACKER]                          1
         [SCANNER, SPAM]                         1
         Name: malicious_label, dtype: int64
```

五．源码复现说明

由于大多数是进行数据分析，所有代码都在 jupyter notebook 下编写。

环境：anaconda 3（64bit）python 3

第三方库：

from pyecharts import Graph

from pyecharts import Bar

源码说明：

1. 处理数据源码位于 get_attacker_data.ipynb，所有代码已进行函数归类，每个函数都有相应注释。需要将所用基础数据文件路径进行更换。
2. 代码生成的所有文件都将存放于相对路径 ip_file 文件夹下，

3. ip_file 存放有第一题的过程文件 logs_ip_score.csv)。
4. 第二题的结果文件的部分数据 ip_data_friends.csv。由于提交文件有限制，无法提交完整第二题的结果数据。最好是将第二题中复现的文件 ip_data_friends.csv 替换此处提交的同名文件。
5. 可视化分析位于 visi_analysis.ipynb