

题目描述

【背景】

各电信运营商、银行、电力等各行各业出于经营需要，逐渐积累了大量个人信息数据与企业运营相关数据；随着社会发展，数据的价值越来越大，网络威胁日益复杂严峻，网络攻击多样化，内部人员泄漏行为频出，网民个人信息的泄露与财产损失的不不断增加。

【场景描述】

某电力公司安全部门在监测数据流量时，监测到一段异常的网络流量信息，安全部门负责人怀疑可能有人恶意获取数据。请您协助分析截获的数据包中是否存在潜在的敏感数据传输及数据泄露风险。

【题干】

请参赛选手从给定的名为“transfer_info.pcap”的 PCAP 包中，解出通过 http 协议传输的文件名为“test.xlsx”的 excel 文件。excel 文件中第一个 sheet 页中有若干的列数据，其中包含身份证号码、驾照编号、银行卡号、MAC 地址、邮箱等五种敏感信息（还包括若干不满足特征的干扰数据）。请参赛选手识别出包含上述五种敏感信息且对应正确敏感信息占比大于 50% 的列，然后根据给定的《数据分类分级规则.xlsx》中对数据分类分级的定义，确定每种敏感信息的数据分类、数据级别，并作为所在列的数据分类和数据级别。然后将识别出的列的分类分级结果按列名数字编号从小到大的顺序依次拼接，拼接格式如下：“列 M 的列名_列 M 的数据分类_列 M 的数据级别;列 N 的列名_列 N 的数据类别_列 N 的数据级别;”（拼接结果以英文分号“;”结尾），然后计算拼接结果的 MD5 值（MD5 值英文字符全小写，长度 32 个字符），并将 MD5 值提交至平台。

excel 文件内容说明：excel 文件中每个列中最多包含一种正确敏感信息，对应列中也可能存在干扰数据。身份证号码、驾照编号、银行卡号、MAC 地址、邮箱的具体规则见“附录：标签说明”。

答案拼接格式说明：同一列的列名、数据类别、数据级别之间用英文下划线（“_”）分隔，两个列的分类分级结果之间用英文分号（“;”）分隔，不同列的分类分级结果按照列名数字编号从小到大的顺序依次拼接，并且最后以英文分号（“;”）结尾。数据类别只取数字编号部分（如：1-1-2-1），数字间使用英文中划线“-”。数据级别转换为数字，1 级数据为 1，2 级数据为 2，3 级数据为 3，4 级数据为 4。

提交说明

提交结果为答案的 MD5 值，只有 MD5 值完全匹配，方得分。

答案示例：c1_1-1-1-1_3;c3_1-3-1-1_2;c4_1-1-1-2_2;c5_1-1-1-1_3;

提交结果示例：8abce13b2c7ca091f475809740d90ce0

附录：标签说明

1. 身份证号码

本题目中的身份证号码是指 18 位二代身份证号码，例如：430101194507137672

具体规则如下：

- 1) 1-6 位 —— 430101，行政地区代码(不同年份有修订)
- 2) 7-14 位 —— 19450713，出生年月日，出生年月日不早于 1930 年 1 月 1 日，不晚于当前日期
- 3) 15-17 位 —— 767，顺序码，任意 3 位数字，第 17 位奇数为男性，偶数为女性
- 4) 18 位 —— 2，校验码，取值范围为 0-9 或者 X
- 5) 身份证号码校验算法参照国标 GB11643-1999《公民身份证号码》

本题的结构化文件《test.xlsx》中正确的身份证号码前 6 位范围为：131123、130522

2. 驾照编号

本题目中的驾照编号是指 12 位纯数字的驾照编号，例如 410324362842

具体规则如下：

- 1) 1-4 位 —— 4103，省市编号
- 2) 5-12 位 —— 24362842，驾驶证的顺序编号

本题的结构化文件《test.xlsx》中正确的驾照编号前 4 位范围为：1101、1305

3. 银行卡号

本题目中的银行卡号是指 19 位纯数字的银行卡号，例如 6200262341334968741

具体规则如下：

- 1) 银行卡号由发卡银行标识 BIN(Bank Identification Number)码、账户标识、校验位组成
- 2) BIN 码 —— 620026，开头 6 至 7 位
- 3) 账户标识 —— 234133496874，通常包含 6-12 位数字
- 4) 校验位 —— 1，最后 1 位，用于校验卡号是否合法，校验算法采用 Luhn 算法

本题的结构化文件《test.xlsx》中正确的银行卡号的发卡银行标识范围为：622481、622513、622250

4. MAC 地址

本题目中的 MAC 地址是指以 12 个大写十六进制数对表示的网络设备物理地址，每对十六进制数之间用英文冒号(:)分隔，例如 00:1A:2B:3C:4D:5E

具体规则如下：

- 1) 前 6 个十六进制数对 —— 00:1A:2B，设备的厂商标识码
- 2) 后 6 个十六进制数对 —— 3C:4D:5E，设备的唯一标识符，由厂商自己分配

本题的结构化文件《test.xlsx》中正确的 MAC 地址的前 6 个十六进制数对范围为：AA:AD:7C、00:2D:4B、00:1A:5F

5. 邮箱

本题目中的邮箱有用户名@域名组成，不限制长度，例如 1234567@qq.com

具体规则如下：

- 1) 用户名：由字母、数字、下划线构成，不可以下划线开始
- 2) 域名：至少包含一个".", 以最后一个"."分割为两部分，前半部分由字母、数字或"."构成，后半部分由字母、数字构成

本题的结构化文件《test.xlsx》中正确的邮箱的域名范围为：163.com、qq.com、hotmail.com