腾讯反诈大脑数据调研

本文提及所需原始数据均不出网，腾讯互联网侧黑产库数据单向同步到部署的位置与原始数据碰撞进行离线分析。仅腾讯黑产数据和模型更新结果返回已确认是否正确更新，检测疑似“诈骗结果”（非原始全量数据）数据需送往腾讯做交叉验证（如非需要用到腾讯侧数据，也可以不传送，覆盖率和准确性将降低），并将腾讯侧验证的结果返回系统所部署网络。数据做双向对称不可逆加密，为手机号/账号+时间戳+疑似诈骗场景+作案嫌疑人/受害人。不含任何个人信息。

1. 数据需求

得到其所涉及的相关数据：如果不使用话单和短信亦可只用网络访问日志，但将缺失电话和短信诈骗的撒网覆盖。

渠道	数据
电话	通话话单（实时话单）
短信	短信话单（实时话单+内容）
网络流量	网址日志（上行分光流量移网+固网）

1.1. 通话话单

注：（1、2、3、4）G 话单，漫游话单全量推送

提供语音话单文件，话单命名方式：yd_cdr_YYYYmmddHHMMSS.txt，话单要求以小文件的形式准实时（延时 10 分钟以内）推送，单个话单文件大小建议不要超过 50MB，话单文件需包含以下字段，如果有更多的字段，欢迎提供；

字段	字段含义	是否必须	能否使用
主叫号码	主叫端的号码	是	
被叫号码	被叫端的号码	是	
话单类型	标识本条话单是来自主叫端还是被叫端	是	
通话时间	该通通话的开始时间，时间日期格式：年月日时分秒，类似 20160420161900	是	
通话时长	该通通话的通话时长，整数，单位秒，未接通时为 0	是	
被计费移动用户 IMSI	计费用户 国际移动用户识别码	最好提供	
被计费移动用	计费用户 国际移动设备身份码	最好提供	

户 IMEI			
话单来源 LAC（位置区识别码）	该通通话来自哪个基站的 LAC 标识，LAC 交换机位置区，主叫流程为主叫用户的值，被叫流程为被叫用户的值	是	
话单来源 Cell_id（小区号码）	该通通话来自哪个基站的 Cell_id 标识，Cell_id 交换机小区号，主叫流程为主叫用户的值，被叫流程为被叫用户的值	是	
话单来源经度	该通通话来自基站的经度	最好提供	
话单来源纬度	该通通话来自基站的纬度	最好提供	
话单来源城市	该通通话来自哪个城市	最好提供	

1.2. 短信话单

提供短信话单文件，话单命名方式：sms_cdr_YYYYmmddHHMMSS.txt，话单要求以小文件的形式准实时（延时 10 分钟以内）推送。单个话单文件大小建议不要超过 50MB，话单文件需包含以下字段，如果有更多的字段，欢迎提供；

字段	字段含义	是否必须	能否使用
发送号码	发送端号码	是	
接收号码	接收端号码	是	
发送类型	该条短信记录来自发送端还是接收端	是	
短信内容	短信的内容	是	
短信长度	短信的长度	是	
发送时间	发送短信的时间	是	
被计费用户 IMSI	计费用户 国际移动用户识别码	最好提供	
被计费用户 IMEI	计费用户 国际移动设备身份码	最好提供	
发送来源 LAC	该条短信来自哪个基站的 LAC 标识，LAC 交换机位置区	是	
发送来源 Cell_id	该条短信来自哪个基站的 Cell_id 标识，Cell_id 交换机小区号	是	
话单来源经度	该通通话来自基站的经度	非必须，最好提供	
话单来源纬度	该通通话来自基站的纬度	非必须，最好提供	

发送来源城市	该条短信来自哪个城市	最好提供	

1.3. 网址日志

字段	备注	能否使用
手机号码	不包含字冠如+86,0086，86	
位置区编码	LAC	
Cell_Id	当有网络切换时，选择第一个 Cell_Id	
IMEI	IMEI 码	
IMSI	IMSI 码	
访问时间	YYYY-MM-DD HH:MM:SS	
访问网址	https://www.baidu.com/	
终端 IP	用户每次请求和应答的 IP 地址	

1.4. 网络登录日志

目前需要 qq、微信的登录日志

字段	备注
手机号码	不包含字冠如+86,0086，86
位置区编码	LAC
Cell_Id	当有网络切换时，选择第一个 Cell_Id
IMEI	IMEI 码
IMSI	IMSI 码
特征信息	QQ、微信账号……
登录时间	YYYY-MM-DD HH:MM:SS
结束时间	YYYY-MM-DD HH:MM:SS
持续时长	以秒为单位
上行流量	以 bytes 为单位
下行流量	以 bytes 为单位
总流量	以 bytes 为单位
终端 IP	用户每次请求和应答的 IP 地址

2.数据量调研

硬件数据和配置与具体的数据量有直接关系。

项目	数量
每日通话话单总量	
每日短信话单总量	
每日网址日志总量（Http get 的总量）	

批注 [T1]: 请在规模处填写总量，
话单以日记总条数精确到百万级即可，具体数据格式要求见后文描述。
短信以条数记日均总量精确到百万级即可
分光以实际日均流量计，同时提供带宽量，分移动网和固网

3.诈骗网址实时拦截方案投入评估需求表

批注 [T2]: 网址拦截评估需要提供的信息量较多，请提供以下表格中所需内容，方可评估投入成本。

如果做诈骗网址实时自动拦截则需要提供以下具体信息，并根据要求做好资料准备。

序号	内容	备注
1	网络拓扑	详细的网络拓扑用于组网设计，后续根据拓扑确认分光的位置和回注的位置。
2	接入流量（实时）	A：固网 1. http get 上行流量大小（需明确流量是否能统一汇聚后接入反诈骗系统，还是必须分机房实现业务）、链路类型及数量； 2. radius 流量大小 （如无实时流量是否能提供实时解析好的对应关系）、链路类型及数量。 B：移网 1. 移网 S1-U 接口、S11 接口用户面 http get 上行流量大小（需明确流量是否能统一汇聚后接入反诈骗系统，还是必须分机房实现业务）、链路类型及数量； 2. 移动网信令面的 S10、S11、S5/S8、S1-MME、S6a 流量大小（信令接口类型根据 S1-MME 数据是否加密，需要进一步确认）、链路类型及数量。 注：1.客户如无法直接提供 http get 流量信息，可提供上行 http 流量信息； 2.移网如无法在核心网中直接进行部署，需明确如何在城域网中进行用户信息与 ip 的关联 。

<table>
<tr><td>3</td><td>用户数信息</td><td>项目覆盖用户数，日活用户数。</td></tr>
<tr><td>4</td><td>分流器策略（http）</td><td>分流器配置策略做源 IP 转发。（如项目通过部署分光器直接进行流量采集可忽略该条内容，由绿网进行后续的流量分流策略配置）。</td></tr>
<tr><td>5</td><td>账号位置信息</td><td>需要获取用户账号对应的地理位置信息（用于绘制欺诈地理热力图）
1：家庭宽带（固网）：通常采用静态文件标记账号物理位置 （也可定制接口获取）。
2：移动网（手机、cpe）：lac、cell-id 相关信息。</td></tr>
<tr><td>6</td><td>公网 IP 地址申请</td><td>1.公网地址用途：
提供拦截提示页服务 （上网用户被重定向到拦截提示页）；
同步更新反欺诈库 （ftp 方式）。
2.拦截提示页公网地址说明
如不做 lvs，需要 1 个公网地址；
如做 LVS，需要 3 个公网地址（建议规模在 100 万用户以上局点使用）。</td></tr>
<tr><td>7</td><td>端口申请</td><td>
<table>
<tr><td>AAA 服务器</td><td>Huas</td><td>内网</td><td>开放端口：65123</td></tr>
<tr><td>库更新服务器</td><td>FTP</td><td>公网</td><td>开放端口：2101</td></tr>
<tr><td>反欺诈平台-缓存服务器</td><td>redis</td><td>内网</td><td>开放端口：6379/6380/6381/6382</td></tr>
<tr><td rowspan="2">反欺诈平台-数据服务器</td><td>mysql</td><td>内网</td><td>开放端口：3306</td></tr>
<tr><td>rsync</td><td>内网</td><td>开放端口：873</td></tr>
<tr><td rowspan="2">反欺诈平台-落地页服务器</td><td>接口服务</td><td>公网</td><td>开放端口：80</td></tr>
<tr><td>rsync</td><td>内网</td><td>开放端口：873</td></tr>
<tr><td rowspan="2">反欺诈平台-管理服务器</td><td>管理平台</td><td>公网</td><td>开放端口：8082</td></tr>
<tr><td>数据统计平台</td><td>公网</td><td>开放端口：8083</td></tr>
<tr><td rowspan="2">负载均衡</td><td>zabbix-web</td><td>公网</td><td>开放端口：8084</td></tr>
<tr><td>zabbix</td><td>内网</td><td>开放端口：10050/10051</td></tr>
<tr><td>所有服务器</td><td>管理平台</td><td>公网/内网</td><td>开放端口：6802</td></tr>
</table>
</td></tr>
<tr><td>8</td><td>拨测资源</td><td>1. 家宽、固网：用于拨测的宽带账号、pc 机，且测试账号的网络访问流量被反欺诈网关覆盖；
2. 移动网：用于拨测的手机号或上网卡、测试终端，且测试账号的网络访问流量被反欺诈网关覆盖。</td></tr>
</table>

批注 [T3]: 1、2、3 属于需要回答的数据信息请在表格对应项中填写，拓扑图以附件形式提供。

批注 [T4]: 4、5、6、7 部分均属于需求，无需回答，但需要在实现时提供。

批注 [T5]: 测试需要的硬件资源含手机、对应运营商的手机卡和宽带账号需要甲方提供。

9	网络连通	1.账号解析服务器需要和同欺诈网关服务器互通； 2.反欺诈库更新服务器需同反欺诈网关服务器互通（反欺诈库更新服务器和反欺诈网关服务器可合设）； 3.反欺诈平台服务器之间（提示页服务器、管理服务器）内网互通； 4.反欺诈网关服务器到反欺诈平台服务器之间内网互通： 平台下发策略到拦截服务器； 拦截服务器上传拦截日志到反欺诈平台。
10	网络链路	1. 回注链路：拦截服务器的回注口可发送报文到用户端和服务端； 拦截服务器到 cr 间的链路上，如果有防火墙，需要防火墙关闭状态检测功能； 回注接口连接的 CR 上如果配置 urpf，需要关闭该接口 urpf 安全策略。 2.回注模式采用三层回注，需要提供回注口的下一跳的三层设备的接口 mac 地址，回注模式采用三层回注，需要提供默认路由的下一跳 IP 地址。 3.移动网回注如需要在回注报文中添加 vlan-ID，需提供 vlan 号和 SGW 地址的对应关系。

批注 [T6]: 9、10 均属于需求，无需答复。