



云公司数据字段分析 (DPI)

中国电信广东公司

WiFi 运营中心

2016-5

目 录

WIFI 运营中心.....	0
1、 4G 网络数据信息采集设备总体架构.....	5
2、 移动 DPI-4G.....	6
2.1 逐个字段分析.....	6
2.1.1 imsi.....	6
2.1.2 msisdn.....	6
2.1.3 imei.....	6
2.1.4 apn.....	7
2.1.5 destinationip, destinationport.....	7
2.1.6 sourceip, sourceport.....	7
2.1.7 sgwip.....	7
2.1.8 mmeip.....	7
2.1.9 pgwip.....	7
2.1.10 sai.....	8
2.1.11 tai.....	8
2.1.12 visitedplmnid.....	8
2.1.13 ratype.....	8
2.1.14 protocolid.....	9
2.1.15 servicetype.....	10
2.1.16 starttime.....	10
2.1.17 endtime.....	10
2.1.18 duration.....	11
2.1.19 inputoctets.....	11
2.1.20 outputoctets.....	11
2.1.21 inputpacket.....	11
2.1.22 outputpacket.....	11
2.1.23 pdnconnectionid.....	11
2.1.24 bearerid.....	11
2.1.25 bearerqos.....	12
2.1.26 recordclosecause.....	12
2.1.27 useragent.....	12
2.1.28 destinationurl.....	13
2.1.29 domainname.....	13
2.1.30 host.....	13
2.1.31 contentlen, contenttype.....	13
2.1.32 iflink.....	13
2.1.33 refer.....	13
2.1.34 httpaction.....	14
2.1.35 httpstatus.....	14
2.2 分析和说明.....	17

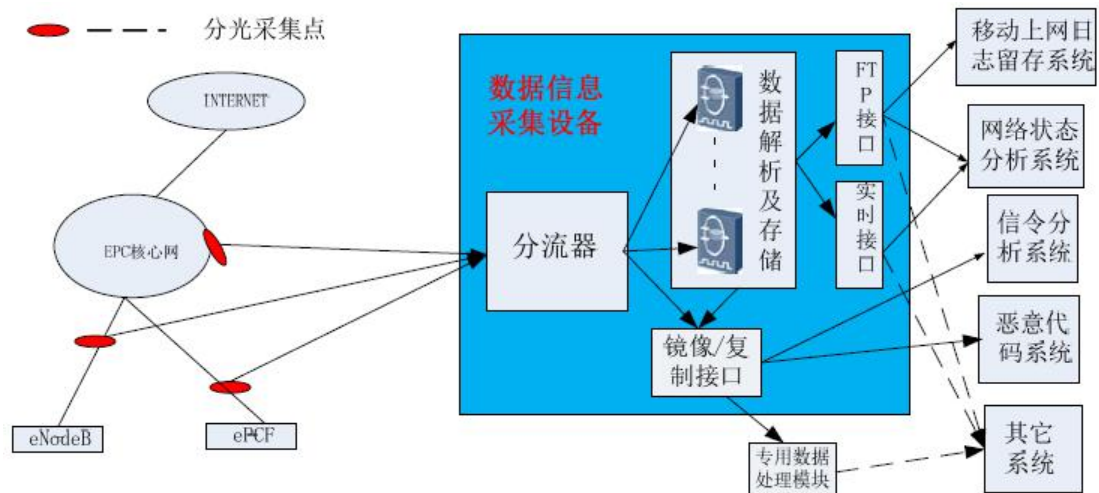
2.2.1 http 解析.....	17
2.2.2 业务类型码表.....	18
2.2.3 价值分析.....	25
3、移动 3GDPI.....	26
3.1 字段分析.....	26
3.1.1 IMSI.....	26
3.1.2 MDN.....	26
3.1.3 MEID.....	26
3.1.4 NAI.....	26
3.1.5 DestinationIP, DestinationPort.....	26
3.1.6 SourceIP, SourcePort.....	27
3.1.7 PDSNIP.....	27
3.1.8 PCFIP.....	27
3.1.9 HAIP.....	27
3.1.10 UserZoneID.....	27
3.1.11 BSID.....	28
3.1.12 Subnet.....	28
3.1.13 ServiceOption.....	28
3.1.14 ProtocolID.....	28
3.1.15 ServiceType.....	28
3.1.16 StartTime.....	28
3.1.17 EndTime.....	29
3.1.18 Duration.....	29
3.1.19 InputOctets.....	29
3.1.20 OutputOctets.....	29
3.1.21 InputPacket.....	29
3.1.22 OutputPacket.....	29
3.1.23 SessionID.....	30
3.1.24 RecordCloseCause.....	30
3.1.25 UserAgent.....	30
3.1.26 DestinationURL.....	30
3.1.27 DomainName.....	30
3.1.28 Host.....	31
3.1.29 ContentLen.....	31
3.1.30 ContentType.....	31
3.1.31 IfLink.....	31
3.1.32 Refer.....	31
3.1.33 HttpAction.....	31
3.1.34 HttpStatus.....	32
3.1.35 RespDelay.....	32
3.1.36 BehaviorTarget.....	32
3.2 分析和说明.....	32
3.2.1 价值分析.....	32

4、固网 DPI	33
4.1 字段分析	33
4.1.1 UserAccount	33
4.1.2 Protocol type	33
4.1.3 SourceIP	33
4.1.4 DestinationIP	33
4.1.5 SourcePort	33
4.1.6 DestinationPort	33
4.1.7 DomainName	34
4.1.8 URL	34
URL	34
4.1.9 REFERER	34
4.1.10 UserAgent	34
4.1.11 Cookie	34
4.1.12 accessTime	34
4.2 分析和说明	35
4.2.1 价值分析	35
5、移动语音	36
5.1 字段分析	36
5.1.1 BIZ_TYPE	36
5.1.2 CALL_TYPE	36
5.1.3 ROAMTYPE	36
5.1.4 IMSI	36
5.1.5 MSISDN	36
5.1.6 OTHER_PARTY	36
5.1.7 THIRD_PARTY	36
5.1.8 START_TIME	37
5.1.9 END_TIME	37
5.1.10 CALL_DURATION	37
5.1.11 MSC	37
5.1.12 TRUNK_GROUPOUT	37
5.1.13 TRUNK_GROUPIN	37
5.1.14 CALLING_LAC	37
5.1.15 CALLED_LAC	37
5.1.16 CALLING_CELL	37
5.1.17 CALLED_CELL	37
5.1.18 HOME_AREA_CODE	38
5.1.19 VISIT_AREA_CODE	38
5.1.20 CALLED_HOME_CODE	38
5.1.21 CALLED_CODE	38
5.1.22 THIRD_HOME_CODE	38
5.1.23 THIRD_CODE	38

5.1.24 FORWARDCAUSE.....	38
5.1.25 VPN_FLAG.....	38
5.1.26 LONG_TYPE.....	39
5.1.27 CARRY_TYPE.....	39
5.1.28 CARRIER_CD.....	39
5.1.29 CFEE.....	39
5.1.30 ACCT_ITEM_TYPE_A.....	39
5.1.31 LFEE.....	39
5.1.32 ACCT_ITEM_TYPE_B.....	39
5.1.33 FEE_ADD.....	40
5.1.34 ACCT_ITEM_TYPE_C.....	40
5.1.35 SERVID.....	40
5.1.36 BILLING_MODE.....	40
5.1.37 EVENT_TYPE.....	40
5.1.38 PRODUCT_ID.....	40
5.1.39 ESN_CODE.....	40
5.1.40 IMEI.....	40
5.1.41 PROV_OFFER_ID.....	40
5.1.42 PROV_RATEDATE.....	41
5.1.43 PROV_BILLING_CYCLE_ID.....	41
5.1.44 CDR_KEY.....	41
5.1.45 RATE_TIMES.....	41
5.2 分析和说明.....	41
5.2.1 价值分析.....	41

1、4G 网络数据信息采集设备总体架构

数据信息采集设备并接在 4G 网络上，通过分光方式或镜像方式获取 IP 链路上的用户互联网访问数据，并根据本规范的要求进行数据解析、生成和存储相关记录，记录以文件接口、实时消息接口方式发送给外部需求系统，同时还应能够根据运营需要镜像/复制相关接口原始报文给外部需求系统。



2、移动 DPI-4G

2.1 逐个字段分析

2.1.1 imsi

国际移动用户识别码（IMSI：International Mobile Subscriber Identification Number）是区别移动用户的标志，储存在 SIM 卡中，可用于区别移动用户的有效信息。其总长度不超过 15 位，同样使用 0~9 的数字。其中 MCC 是移动用户所属国家代号，占 3 位数字，中国的 MCC 规定为 460；

- 我们拿到的数据被 MD5 32 位不可逆加密
- 数据样例：09320F861286E071B22104E98D3CD03F
- 价值低。

2.1.2 msisdn

MSISDN 是指主叫用户为呼叫 GSM PLMN 中的一个移动用户所需拨的号码，作用同于固定网 PSTN 号码；是在公共电话网交换网络编号计划中，唯一能识别移动用户的号码。用户手机号码。

- 我们拿到的数据被 MD5 32 位不可逆加密
- 数据样例：9D3C03CC3AF14656014330135C33FBB7
- 价值低。

2.1.3 imei

IMEI(International Mobile Equipment Identity)是国际移动设备身份码的缩写，国际移动装备标识码，是由 15 位数字组成的"电子串号"，它与每台移动电话机一一对应，而且该码是全世界唯一的。每一只移动电话机在组装完成后都将被赋予一个全球唯一的一组号码，这个号码从生产到交付使用都将被制造生产的厂商所记录。移动台设备标识

- 我们拿到的数据被 MD5 32 位不可逆加密
- 数据样例：CEADE65B68F17D7E196B8BE6018E6AA4
- 价值低。

2.1.4apn

APN 指一种网络接入技术，是通过手机上网时必须配置的一个参数，它决定了手机通过哪种接入方式来访问网络。从运营商角度看，APN 就是一个逻辑名字，APN 一般都部署在 GGSN 设备上或者逻辑连接到 GGSN 上，用户使用 GPRS 上网时，都通过 GGSN 代理出去到外部网络，因此，APN 设置、过滤、统计等，就成为一个对 GPRS 计费、GPRS 资费有重要参考价值的参数之一（因为 APN 可以区分一个业务或者外部网络）。

- 数据为用户上网的接入形式
- 数据样例： `ctlte.MNC011.MCC460.GPRS`

2.1.5destinationip, destinationport

访问目标的 IP 和端口

- 数据样例： `183.60.17.203, 80`

2.1.6sourceip, sourceport

本机（移动终端）源 IP 地址和端口

- 数据样例： `100.103.143.12, 48164`

2.1.7sgwip

Serving GateWay, 服务网关 IP

- 数据样例： `115.169.133.149`

2.1.8mmeip

MME（Mobility Management Entity）是 3GPP 协议 LTE 接入网络的关键控制节点，它负责空闲模式的 UE(User Equipment)的定位，传呼过程，包括中继，简单的说 MME 是负责信令处理部分。Mmeip 即这个关键接入节点的 ip。

- 数据样例： `115.169.133.2`

2.1.9pgwip

PGW(PDN GateWay, PDN 网关)是移动通信网络 EPC 中的重要网元。Pgwip

即 PDN 网关 IP。

- 数据样例：115.169.194.12

2.1.10sai

服务区标识

- 数据样例：46011126963969
- 数据样例：46011123937073
- 数据样例：46011141057330

2.1.11tai

跟踪区标识

- 数据样例：46011126963969
- 数据样例：46011123937073
- 数据样例：46011141057330

2.1.12visitedplmnid

拜访网络标识

- 数据样例：46011

2.1.13ratype

rat=Radio Access Technologies，无线接入技术，ratype 为无线接入类型。即用户所接入的网络类型。

- 数据样例：1004
- 数据样例：2003

0	WLAN
1	unknown
1000	UTRAN ，（UMTS Terrestrial Radio Access Network）是 UMTS 的陆地无线接入网（URAN）。它是第三代移动通信技术 UMTS 最重要的一种接入方式，并且适用范围最广。
1001	GERAN 是 GSM/EDGE 无线接入网，它采用了 EDGE 的无线传输技术，网络组成与 GPRS 相同。EDGE 的目的是为了在现有蜂窝系统中提供更高的数据速率。

1002	<p>GAN Generic access networks , a telecommunication system that extends mobile voice, data and IP Multimedia Subsystem/Session Initiation Protocol (IMS/SIP) applications over IP networks.</p> <p>(GAN =通用访问网络,一个电信系统扩展移动语音、数据和 IP 多媒体子系统/会话发起协议(IMS / SIP)应用程序通过 IP 网络。)</p>
1003	<p>HSPA, High-Speed Packet Access, WCDMA 的 R99 和 R4 系统能够提供的最高上下行速率分别为 64kbps 和 384kbps, 为了能够与 CDMA1XEV-DO 抗衡, WCDMA 在 R5 规范中引入了 HSDPA, 在 R6 规范中引入了 HSUPA, HS-DPA 和 HSUPA 合称为 HSPA</p>
1004	<p>EUTRAN, Evolved Universal Terrestrial Radio Access Network, 演进的通用陆基无线接入网</p>
2000	<p>CDMA2000 (Code Division Multiple Access 2000) 是一个 3G 移动通讯标准, 国际电信联盟 ITU 的 IMT-2000 标准认可的无线电接口, 也是 2G CDMAOne 标准的延伸。 根本的信令标准是 IS-2000。 CDMA2000 与另一个 3G 标准 WCDMA 不兼容。</p>
2001	<p>HRPD (High Rate Packet Data), 又被称为 1xEVDO, 或 HDR (High Data Rate)。 HRPD 意思是“高速分组数据”, 它是由高通发明的基于 CDMA 的高速无线数据技术, 国际标准编号为 IS-856。</p>
2002	<p>UMB 是 CDMA2000 系列标准的演进升级版本, 可升级至 20MHz 的带宽, 可在现有或新分配的频段中部署。 UMB 系统是以 OFDMA(正交频分复用接入)技术为基础、专门针对无线移动环境和实时应用优化的移动无线宽带系统</p>
2003	<p>eHRPD (Evolved High Rate Package Data, 演进的高速分组网络, eHRPD 技术是为 CDMA 至 LTE 而设计的, 因此其参考标准由 3GPP2 制定</p>

2.1.14protocolid

协议类型, 被码表化了, 需要对应码表才知道具体采用的是什协议访问网络。

● 数据样例: 1

标准协议 (适用于 3G 和固网)

编号	协议类型
1	HTTP
2	WAP (包含 wap1.x 和 wap2.0)
3	SMTP

4	POP3
5	IMAP4
6	FTP（包含动静态 FTP）
7	RTSP
8	MMS
9	ICMP
10	Radius
11	H. 323
12	SIP
13	MGCP
14	H. 248
15	OTHER

非标准协议（适用于 3G 和固网）

编号	协议类型
16	P2P 下载
17	P2P 视频
18	即时消息
19	VoIP
20	网络游戏
21	证券交易
22	在线阅读
23	在线音乐
24	应用商店
25	其它

2.1.15servicetype

业务应用，被码表化了，需要对应码表才知道具体服务类型。

- 数据样例：**E99**

2.1.16starttime

业务开始时间，格式 yyyyMMddhhmmss

- 数据样例：**20160419120442**

2.1.17endtime

业务结束时间，格式 yyyyMMddhhmmss

- 数据样例：20160419120547

2.1.18duration

持续时长单位毫秒

- 数据样例：64917

2.1.19inputoctets

发给用户的业务字节数。

- 数据样例：448

2.1.20outputoctets

用户发出的业务字节数。

- 数据样例：510

2.1.21inputpacket

发给用户的业务数据包数量。

- 数据样例：1

2.1.22outputpacket

用户发出的业务数据包数量。

- 数据样例：2

2.1.23pdnconnectionid

PDN 连接的 ID 标识号，采集设备自动生成，以 session 开始的时间戳为 ID 号，同一个用户中的该 ID 相同表明是同一次 PDN 连接中的不同记录，格式为 yyyymmddhhmmss（24 小时制）

- 数据样例：20160418205112

2.1.24bearerid

承载标识号。

- 数据样例：5

2.1.25bearerqos

承载 QoS 从 GTP-C 消息中获取 BearerID 对应的 QoS。

- 数据样例：

00000411FF000010000028FF11FFFF0000000410FF000010000028FF11FF
FF00

2.1.26recordclosecause

记录关闭原因，被码表化了，需要对应码表才知道具体原因。

- 数据样例：2
- 数据样例：0

0	session 释放
1	业务应用结束；
2	业务应用等待超时
3	用户位置变化
4	用户接入方式变化

2.1.27useragent

User Agent 信息，标注浏览器信息，通常指各个手机开发包的名称及版本信息。

- 数据样例 1：Dalvik/1.6.0 (Linux; U; Android 4.4.4; HUAWEI ALE-CL00 Build/HuaweiALE-CL00)。可以分析出用户用的数华为的手机，**Android** 的操作系统。
- 数据样例 2：Mozilla/5.0 (iPhone; CPU iPhone OS 9_2 like Mac OS X; zh-CN) AppleWebKit/537.51.1 (KHTML, like Gecko) Mobile/13C75 UCBrowser/10.7.11.672 Mobile。可以分析用户用的是苹果手机，**OS9** 操作系统，火狐浏览器。

2.1.28destinationurl

用户访问的目标网站的 URL

- 数据样例: <http://183.60.17.203/cgi-bin/httpconn?mType=Other>。可以分析用户在上什么网站。

2.1.29domainname

外部网站的域名

- 数据样例: 183.60.17.203。
- 数据样例: kwurl.ucweb.com。
- 数据样例: pingma.qq.com。

2.1.30host

访问目标主机

- 数据样例: 183.60.17.203。
- 数据样例: kwurl.ucweb.com 他并没有把域名解析为主机 IP(14.152.65.28)
- 数据样例: pingma.qq.com。他并没把域名解析为主机 IP (14.17.42.58)

2.1.31contentlen, contenttype

内容大小, 内容类型

- 数据样例: 283, text/octet。文本字节, 长度 283
- 数据样例: 1426, application/octet-stream。应用程序, 流字节, 1426
- 数据样例: 39275, image/jpeg。图片, 39275

2.1.32iflink

是否是链接访问 (0 否; 1 是)

- 数据样例: 0
- 数据样例: 1

2.1.33refer

链接源信息, 根据 GET 请求包中 Referer 字段提取

- 数据样例：
- 数据样例：

yd.baidu.com/view/4987019c6294dd88d0d26bce?token=5mvo4S5sb%252BVAp4hlawDWyBe%2F21YoeEulp0LRBWIA0%252B1%252B9sw0RiUTT8VbxlpuZ4Z5uk4PyJulQM0%3D&novel_id=4305071941&token=5mvo4S5sb%252BVAp4hlawDWyBe%2F21YoeEulp0LRBWIA0%252B1%252B9sw0RiUTT8VbxlpuZ4Z5uk4PyJulQM0%3D&novel_id=4305071941&token=5mvo4S5sb%252BVAp4hlawDWyBe%2F21YoeEulp0LRBWIA0%252B1%252B9sw0RiUTT8VbxlpuZ4Z5uk4PyJulQM0%3D&novel_id=4305071941&token=5mvo4S5sb%252BVAp4hlawDWyBe%2F21YoeEulp0LRBWIA0%252B1%252B9sw0RiUTT8VbxlpuZ4Z5uk4PyJulQM0%3D&novel_id=430

2.1.34httpaction

HTTP 操作类型（5: Post,6:Get）

- 数据样例： 5
- 数据样例： 6

2.1.35httpstatus

http statusCode(状态码)

201-206	都表示服务器成功处理了请求的状态代码，说明网页可以正常访问。
200（成功）	服务器已成功处理了请求。通常，这表示服务器提供了请求的网页。
201（已创建）	请求成功且服务器已创建了新的资源。
202（已接受）	服务器已接受了请求，但尚未对其进行处理。
203（非授权信息）	服务器已成功处理了请求，但返回了可能来自另一来源的信息。
204（无内容）	服务器成功处理了请求，但未返回任何内容。
205（重置内容）	服务器成功处理了请求，但未返回任何内容。与 204 响应不同，此响应要求请求者重置文档视图（例如清除表单内容以输入新内容）。

206（部分内容）	服务器成功处理了部分 GET 请求。
300-3007	表示的意思是：要完成请求，您需要进一步进行操作。通常，这些状态代码是永远重定向的。
300（多种选择）	服务器根据请求可执行多种操作。服务器可根据请求者 来选择一项操作，或提供操作列表供其选择。
301（永久移动）	请求的网页已被永久移动到新位置。服务器返回此响应时，会自动将请求者转到新位置。您应使用此代码通知搜索引擎蜘蛛网页或网站已被永久移动到新位置。
302（临时移动）	服务器目前正从不同位置的网页响应请求，但请求者应继续使用原有位置来进行以后的请求。会自动将请求者转到不同的位置。但由于搜索引擎会继续抓取原有位置并将其编入索引，因此您不应使用此代码来告诉搜索引擎页面或网站已被移动。
303（查看其他位置）	当请求者应对不同的位置进行单独的 GET 请求以检索响应时，服务器会返回此代码。对于除 HEAD 请求之外的所有请求，服务器会自动转到其他位置。
304（未修改）	自从上次请求后，请求的网页未被修改过。服务器返回此响应时，不会返回网页内容。
305（使用代理）	请求者只能使用代理访问请求的网页。如果服务器返回此响应，那么，服务器还会指明请求者应当使用的代理。
307（临时重定向）	服务器目前正从不同位置的网页响应请求，但请求者应继续使用原有位置来进行以后的请求。会自动将请求者转到不同的位置。但由于搜索引擎会继续抓取原有位置并将其编入索引，因此您不应使用此代码来告诉搜索引擎某个页面或网站已被移动。
4XXHTTP 状态码表示请求可能出错，会妨碍服务器的处理。	
400（错误请求）	服务器不理解请求的语法。
401（身份验证错误）	此页要求授权。您可能不希望将此网页纳入索引。
403（禁止）	服务器拒绝请求。

云公司数据字段分析

404（未找到）	服务器找不到请求的网页。例如，对于服务器上不存在的网页经常会返回此代码。
405（方法禁用）	禁用请求中指定的方法。
406（不接受）	无法使用请求的内容特性响应请求的网页。
407（需要代理授权）	此状态码与 401 类似，但指定请求者必须授权使用代理。如果服务器返回此响应，还表示请求者应当使用代理。
408（请求超时）	服务器等候请求时发生超时。
409（冲突）	服务器在完成请求时发生冲突。服务器必须在响应中包含有关冲突的信息。服务器在响应与前一个请求相冲突的 PUT 请求时可能会返回此代码，以及两个请求的差异列表。
410（已删除）	请求的资源永久删除后，服务器返回此响应。该代码与 404（未找到）代码相似，但在资源以前存在而现在不存在的情况下，有时会用来替代 404 代码。如果资源已永久删除，您应当使用 301 指定资源的新位置。
411（需要有效长度）	服务器不接受不含有有效内容长度标头字段的请求。
412（未满足前提条件）	服务器未满足请求者在请求中设置的其中一个前提条件。
413（请求实体过大）	服务器无法处理请求，因为请求实体过大，超出服务器的处理能力。
414（请求的 URI 过长）	请求的 URI（通常为网址）过长，服务器无法处理。
415（不支持的媒体类型）	请求的格式不受请求页面的支持。
416（请求范围不符合要求）	如果页面无法提供请求的范围，则服务器会返回此状态码。
417（未满足期望值）	服务器未满足"期望"请求标头字段的要求。
500 至 505 表示的意思是：服务器在尝试处理请求时发生内部错误。这些错误可能是服务器本身的错误，而不是请求出错。	

500（服务器内部错误）	500（服务器内部错误）
501（尚未实施）	服务器不具备完成请求的功能。例如，当服务器无法识别请求方法时，服务器可能会返回此代码。
502（错误网关）	服务器作为网关或代理，从上游服务器收到了无效的响应。
503（服务不可用）	目前无法使用服务器（由于超载或进行停机维护）。通常，这只是一种暂时的状态。
504（网关超时）	服务器作为网关或代理，未及时从上游服务器接收请求。
505（HTTP 版本不受支持）	服务器不支持请求中所使用的 HTTP 协议版本。

- 数据样例：200。

2.2分析和说明

2.2.1http 解析

数据信息采集设备能够准确识别标准端口和非标准端口上的 HTTP 流，并通过 HTTP 协议直接获取 UserAgent、DestinationURL、DomainName、Host、Contentlen、ContentType、Referer、HttpAction 和 HTTPSTATUS 等字段，也能够根据业务流综合分析获取 RESPDELAY 和 Behavior Target，同时能够对解析、统计、分析的结果进行记录。

HTTP 浏览信息解析规则：业务流创建后，生成一条记录。要求以 get/post 请求为单位上报。

对于 DestinationURL 字段信息遵循以下统一的规则：

DestinationURL 字段获取长度可配置，单位为 Byte，默认值为空，表示记录完整 URL。DestinationURL 字段获取文件类型可配置，支持正向选择文件类型，或者反向选择文件类型。比如经过配置各信息采集点只记录扩展名为 html、htm、jsp、asp 的 URL，或者经过配置各信息采集点不记录扩展名为 css、js 的 URL。默认为空，表示全部记录。如果 DestinationURL 字段信息包含“|”、回车、换行字符，那么需要转义后记录到数据文件中。DestinationURL 字段为必须解析项。

2.2.2业务类型码表

业务类别	业务应用编码	业务应用名称
P2P 下载	101	迅雷
	102	eMule
	103	BT
	104	网际快车
	105	QQ 旋风
	106	VaGaa
	107	Maze
	108	汉魅
	109	酷我音乐盒
	110	酷狗音乐盒
	111	115 优蛋
	112	RaySource
	113	飞速土豆
	114	纳米盘
	115	360 软件管家
	116	Shareaza
	117	Poco
	118	百纳 Biget
	199	Other P2P
信息即时交互 服务	201	QQ
	202	MSN
	203	旺旺
	204	飞信
	205	微信
	206	米聊
	207	ECP（电信自营业务）
	208	飞聊
	209	Gtalk
	216	陌陌
	211	go 聊
	212	百度 Hi
	213	新浪 UC
	214	有你短信
	215	网易泡泡
	299	Other IM
HTTP 应用	301	www.vnet.cn
	302	新浪微博
	303	腾讯微博
	304	搜狐微博

云公司数据字段分析

305	开心网
306	土豆网
307	大众点评
308	人人网
309	网易微博
310	移动 139 微博
311	百度
312	淘宝
313	天涯社区
314	新浪论坛
315	京东
316	QQ 空间
317	腾讯网
318	亚马逊
319	新浪新闻娱乐
320	1 号店
321	当当
322	凡客
323	猫扑
324	人脉库
325	世纪佳缘
326	MiniWeb
328	腾讯搜索服务
329	都市在线
330	太平洋汽车网
331	汽车之友
332	汽车之家
333	汽车在线
334	汽车点评网
335	六度健康网
336	99 健康网
337	39 健康网
338	中国足彩网
339	中国体彩网
340	中国竞彩网
341	中国福彩网
342	新浪彩票
343	腾讯彩票
344	搜狐彩票
345	网易彩票
346	趣天下
347	去哪儿网团
348	清团网

云公司数据字段分析

	349	天猫商城
	350	国美电器
	351	淘宝商城
	352	苏宁易购
	353	腾讯 qq 商城
	354	腾讯 qq 网购
	355	腾讯拍拍
	356	腾讯微卖场
	357	网易商城
	358	亲亲网
	359	千品网
	360	高朋团购
	361	糯米网
	362	米奇网
	363	猛买网
	364	美团网
	365	苏泊尔
	366	三星
	367	美菱
	368	美的
	369	康佳
	370	九阳
	371	海信
	372	海尔
	399	普通网页浏览
VPN 类应用	401	VPN
电子邮件服务	501	Email
WAP 类应用	601	彩信
	699	其他 WAP 浏览
VoIP	701	Skype
	702	QQ 语音
	703	新浪 SHOW
	704	新浪 UTGAME
	705	六间房
	706	歪歪语音
	707	ET 语音
	708	Doshow 视频聊天
	709	IsPeak
	710	Lava-Lava
	711	9158 多人视频
流媒体应用	799	Other VoIP
	801	MediaPlay
	802	RealPlay

云公司数据字段分析

	803	QuickTime
	899	Other RTSP
P2P 视频	901	PPLive (PPTV)
	902	PPStream
	903	迅雷看看
	904	风行
	905	腾讯视频
	906	优酷
	907	土豆
	908	天翼视讯（电信自营业务）
	909	手机电视
	910	UUSEE
	911	快播
	912	酷 6
	913	暴风影音
	914	皮皮影视
	915	M1905 电影网
	916	VG0 网络电视
	917	QQ Live
	918	搜狐视频
	919	新浪视频
	920	网易视频
	921	A67 手机电影网
	922	80s 手机电影网
	923	7060 手机电影
	924	激动网
	925	乐视网
	926	56 视频
	927	芒果 TV
	928	奇艺影音
	929	CNTV
	930	手机电视 Dopool
	931	优米网
	932	东方宽频
	933	凤凰视频
	934	3G 生活网
	935	163TV
	936	517dv
	937	金鹰网
	938	UC 影音
	999	其他基于 P2P 的流媒体应用
网络游戏	A01	爱游戏（电信自营业务）
	A02	三国杀

云公司数据字段分析

	A03	QQ 游戏大厅
	A04	浩方游戏平台
	A05	魔兽世界
	A06	地下城与勇士
	A07	联众世界
	A08	浩方对战
	A09	热血江湖
	A10	跑跑卡丁车
	A11	劲舞团
	A12	街头篮球
	A13	梦幻西游
	A14	巨人征途
	A15	天堂 II
	A16	剑侠世界
	A17	剑侠情缘 3
	A18	穿越火线
	A19	天龙八部
	A20	天道
	A21	QQ 飞车
	A22	新浪游戏
	A23	太平洋游戏
	A24	游侠网
	A25	17173
	A26	QQ 音速
	A27	中国游戏中心
	A28	征服
	A99	其他网络游戏
证券交易	B01	同花顺
	B02	大智慧
	B03	商旅在线
	B04	Elong 商旅
	B05	证券之星
	B06	财华终端
	B07	投资堂
	B08	益盟操盘手
	B09	华泰联合
	B10	和讯
在线阅读	B99	其他证券应用
	C01	百阅
	C02	天翼阅读（电信自营业务）
	C03	爱动漫（电信自营业务）
	C04	熊猫看书
	C05	起点手机阅读

云公司数据字段分析

C06	起点中文网
C07	掌上书院
C08	空中书城
C09	小说阅读网
C10	纵横中文网
C11	掌阅
C12	天涯阅读
C13	新浪读书
C14	快眼看书
C15	潇湘书院
C16	红袖添香
C17	QQ 书城
C18	百阅
C19	中国论文联盟
C20	中国留学网
C21	外国教育网
C22	中华军事
C23	中国战略网
C24	中国军网
C25	战争类小游戏
C26	战略网
C27	新军事网
C28	新华网军事
C29	西陆军事
C30	人民网军事
C31	热门军事新闻
C32	枪炮世界
C33	千龙军事
C34	专利之家
C35	中华网文史
C36	中国知网
C37	中国诗歌流派
C38	中国期刊网
C39	中国国家地理网
C40	语文资源网
C41	有道网
C42	邮政报刊订阅网
C43	环球网历史
C44	环球科学
C45	互动百科
C46	果壳网
C47	国学堂
C48	国学论坛

云公司数据字段分析

	C49	国学经典
	C50	国学讲座
	C51	国家精品课程资源网
	C52	公务员考试网
	C53	中国文艺网
	C54	中国诗人论坛
	C55	徐志摩诗网
	C56	中国作家网
	C57	作家在线
	C58	作家网
	C59	卓芒文学网
	C60	斩龙小说网
	C61	重要小说网
	C62	追书网
	C63	着笔中文网
	C64	左书屋
	C65	智耐小说网
	C66	盗墓笔记小说网
	C67	娱乐秀小说网
	C68	歪歪小说
	C69	阅书网
	C70	烟雨中文网
	C71	岳麓小说网
	C72	银星社区
	C73	言情小说库
	C74	暴走漫画
	C99	其他在线阅读
在线音乐	D01	酷狗
	D02	酷我音乐盒
	D03	QQ 音乐
	D04	爱音乐（电信自营业务）
	D05	天天动听
	D06	搜狗音乐
	D07	虾米音乐
	D08	一听音乐
	D09	百度音乐
	D10	九天音乐
	D11	K 歌达人
	D12	豆瓣电台
	D13	懒人听书
应用商店	D99	其他在线音乐
	E01	天翼空间（电信自营业务）
	E02	安卓市场

	E03	安智市场
	E04	AppStore
	E05	91 手机助手
	E06	掌上应用汇
	E07	智汇云
	E08	google play
	E09	豌豆荚
	E99	其他应用市场
其他网络应用	F01	FTP
	F02	SSL

2.2.3 价值分析

Imsi 和 **msisdn** 字段，为客户的 imsi 号以及手机号，很重要的字段，可惜我们拿到手的是经过 MD5（32 位）不可逆加密了的字符串，价值大打折扣。我们需要加密前的数据。

Useragent 字段，能够分析用户使用的手机型号，在客户消费潜力模型里面能够提供依据。

ServiceType、**DestinationIP**、**DestinationURL** 这三个字段对分析用户网上行为偏好有作用。

3、移动 3GDPI

3.1 字段分析

3.1.1 IMSI

国际移动用户识别码（IMSI：International Mobile Subscriber Identification Number）是区别移动用户的标志，我们拿到的数据被 MD5 32 位不可逆加密。

- 数据样例：5088FE5552C3864F17F8F7C5D9E6E3DD

3.1.2 MDN

用户手机号码，我们拿到的数据被 MD5 32 位不可逆加密

- 数据样例：6F5208215F691CA4FD8951DBB95B805D

3.1.3 MEID

移动台设备标识 ID，我们拿到的数据被 MD5 32 位不可逆加密

- 数据样例：68705C73917777E8BAE07F5F3AD9CE17

3.1.4 NAI

网络接入标识，用户的帐号，标识用户名和归属网络，格式为：user@realm

- 数据样例：ctnet@mycdma.cn

3.1.5 DestinationIP, DestinationPort

用户访问目标的 IP 地址和端口号

- 数据样例：211.155.94.202, 80

3.1.6 SourceIP, SourcePort

用户访问外部网络时使用的 IP 地址和端口号

- 数据样例: 10.91.21.78, 2626

3.1.7 PDSNIP

PDSN (Packet Data Serving Node) 中文名为分组数据服务节点、分组业务数据节点 IP,采用点分十进制表示法, NAS-IP-Address, PDSN 作为 RADIUS 客户端的地址。

- 数据样例: 115.168.82.161

3.1.8 PCFIP

PCF (Packet Control Function) IP,采用点分十进制表示法, 3GPP2_PCF IP_Addr, 服务 RAN 中的 PCF。

- 数据样例: 0.0.0.0

3.1.9 HAIP

HA 的 IPv4 或 IPv6 地址, 3GPP2_HA_IP_Addr, 用户在 MIP 会话时使用的 HA IP 地址.如[RFC 2002]定义。

- 数据样例: 11.115.159.0

3.1.10 UserZoneID

用户区 ID, 3GPP2_User_ID, 最低有效的 16 位容纳用户区 ID(UZ_ID),下 15 有效位容纳用户区系统 ID(UZ_SID),最高有效位置为 0

- 数据样例: 87
- 数据样例: 86

3.1.11BSID

基站标识，3GPP2_BSID，每项使用十六进制的 ASCII 字符编码。

- 数据样例：360CFFFF139C

3.1.12Subnet

子网号，DO 系统的子网和扇区标识，3GPP2_Subnet，HRPD 系统的子网。

- 数据样例：
01136800A8C0006F000000810B110926000000021200A8C0006F0000008
10B11

3.1.13ServiceOption

用户所接入的 cdma 网络类型，3GPP2_SO，接入网络类型

- 数据样例：59

3.1.14ProtocolID

协议类型。十六进制整形表示

- 数据样例：1

3.1.15ServiceType

应用类型。

- 数据样例：E99
- 数据样例：302

3.1.16StartTime

业务流开始时间，格式为 yyyymmddhhmmss（24 小时制），如果开启中间记录模式，每条记录都填写相同的开始时间。

- 数据样例：20160419110209

3.1.17EndTime

业务流结束时间，格式为 `yyyymmddhhmmss`（24 小时制），如果开启中间记录模式，只在最后一条记录填写结束时间。

- 数据样例：20160419110650

3.1.18Duration

持续时间，单位毫秒

- 数据样例：281005

3.1.19InputOctets

发给用户的业务字节数

- 数据样例：0

3.1.20OutputOctets

用户发出的业务字节数

- 数据样例：668

3.1.21InputPacket

发给用户的业务数据包数量

- 数据样例：0

3.1.22OutputPacket

用户发出的业务数据包数量

- 数据样例：5

3.1.23SessionID

session 的 ID 标识号，采集设备自动生成，以 session 开始的时间戳为 ID 号，同一个用户中的该 ID 相同表明是同一次 PPP 连接中的不同记录，格式为 yyyymmddhhmmss（24 小时制）

- 数据样例：20221018173113

3.1.24RecordCloseCause

记录关闭原因，采集设备自动生成，数字编号（0：session 释放；1：业务应用结束；2：业务应用等待超时；3：用户位置变化；4：用户接入方式变化）

- 数据样例：2

3.1.25UserAgent

User Agent 信息，User-Agent，标注浏览器信息，通常指各个手机开发包的名称及版本信息

- 数据样例：Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0)

3.1.26DestinationURL

用户访问的目标网站的 URL。

- 数据样例：
<http://craw.adpro.cn/crawler.js?url=http%3A%2F%2Fwww.maogoule.com%2F&body=4422&hash=G:WwoyDShz0KPOy4AkEzSwJuuryLYlADGldwIo:syDSE0KPK04SqLYlz9dwIo>

3.1.27DomainName

外部网站的域名

- 数据样例：craw.adpro.cn

3.1.28Host

外部网站主机名

- 数据样例: crawl.adpro.cn

3.1.29ContentLen

发给用户的业务数据包数量

- 数据样例: 0

3.1.30ContentType

内容大小, Content-Length, 协议标准字段。

- 数据样例: 0

3.1.31IfLink

是否是链接访问, 0=否, 1=是, 根据 GET 请求包中是否包含 Referer 来判断。

- 数据样例: 1

3.1.32Refer

链接源信息, 根据 GET 请求包中 Referer 字段提取。

- 数据样例: www.maogoule.com/

3.1.33HttpAction

HTTP 操作类型 (5: Post, 6:Get)

- 数据样例: 6

3.1.34HttpStatus

HTTP 状态码（4GDPI 里面有相同说明）。

- 数据样例：0

3.1.35RespDelay

响应时延，单位毫秒。

- 数据样例：0

3.1.36BehaviorTarget

目标行为，0：会话是用户主动点击的页面；1：站点目标而产生的页面。

- 数据样例：1

3.2分析和说明

3.2.1价值分析

Imsi 和 **mdn** 字段，为客户的 imsi 号以及手机号，很重要的字段，可惜我们拿到手的是经过 MD5（32 位）不可逆加密了的字符串，价值大打折扣。我们需要加密前的数据。

UserAgent 字段，能够分析用户使用的手机型号，在客户消费潜力模型里面能够提供依据。

ServiceType、**DestinationIP**、**DestinationURL** 这三个字段对分析用户网上行为偏好有作用。

4、固网 DPI

4.1 字段分析

4.1.1 UserAccount

用户账号。我们拿到的样例数据是某种可逆加密后的密文数据

- 数据样例: WSt1oYFrFk2G1i0xHQtl2Q==
- 数据样例: ze4Y8ibaF2hqrJJsAbYMbA==

4.1.2 Protocol type

协议类型。十六进制整形表示

- 数据样例: 6

4.1.3 SourceIP

源 IP

- 数据样例: 220.180.34.25

4.1.4 DestinationIP

目的 IP

- 数据样例: 180.97.178.210

4.1.5 SourcePort

源端口，十六进制整形表示

- 数据样例: 121e

4.1.6 DestinationPort

目的端口，十六进制整形表示

- 数据样例: 50

4.1.7DomainName

域名

- 数据样例: roomapicdn.plu.cn

4.1.8URL

URL

- 数据样例:
http://roomapicdn.plu.cn/room/roomstatus?pagetype=live&profileField=1&roomid=215644&callback=_c_536165

4.1.9REFERER

链接源信息

- 数据样例: <http://star.longzhu.com/105232>

4.1.10UserAgent

UserAgent 信息

- 数据样例: Mozilla/4.0 (compatible; MSIE 9.0; Windows NT 6.1)

4.1.11Cookie

Cookie 信息

- 数据样例: PLULOGINSESSID=61b4f042089d24dcec6bead92341339c;

4.1.12accessTime

访问时间, 采用 UTC 格式

- 数据样例：1460906399

4.2分析和说明

4.2.1价值分析

URL 以及 Cookie 字段对分析用户网上行为偏好有作用。

Useragent 字段，能够分析用户使用的浏览器型号，进而推断出用户用的是普通 PC 电脑还是苹果电脑以及型号，在客户消费潜力模型里面能够提供依据。

5、移动语音

5.1字段分析

5.1.1BIZ_TYPE

业务类型 01：国内，02：国际。

5.1.2CALL_TYPE

呼叫类型 01：主叫，02：被叫，03：呼转

5.1.3ROAMTYPE

漫游类型 0:非漫游, 1:省内漫游, 4:省际出访漫游, 5: 国际漫游

5.1.4IMSI

IMSI 号

增加错误编码：

F040:IMSI 为空；

F041:长度非 15 位；

F042:非全数字；

F043:查表无码段配置；

F044:非中国电信的码段。

5.1.5MSISDN

计费号码，手机号码不带前缀 86 或者 0086

5.1.6OTHER_PARTY

对端号码，手机号码不带前缀 86 或者 0086

5.1.7THIRD_PARTY

第三方号码，手机号码不带前缀 86 或者 0086

5.1.8START_TIME

通话开始时间，YYYYMMDDHH24MISS

5.1.9END_TIME

通话结束时间，YYYYMMDDHH24MISS

5.1.10CALL_DURATION

通话时长(秒)

5.1.11MSC

交换机号

5.1.12TRUNK_GROUPOUT

出中继

5.1.13TRUNK_GROUPIN

入中继

5.1.14CALLING_LAC

主叫 pa 区

5.1.15CALLED_LAC

被叫 pa 区

5.1.16CALLING_CELL

主叫基站

5.1.17CALLED_CELL

被叫基站

5.1.18HOME_AREA_CODE

计费号码归属地区号

5.1.19VISIT_AREA_CODE

计费号码到访地区号

5.1.20CALLED_HOME_CODE

对方归属地区号

5.1.21CALLED_CODE

对方到访地区号

5.1.22THIRD_HOME_CODE

第三方归属地区号

5.1.23THIRD_CODE

第三方到访地区号

5.1.24FORWARDCAUSE

前转原因

1: 所有呼叫前转

0: 不是互转话单

V: 无条件呼叫前转

W: 有条件呼叫前转

X: 遇忙前转

Y: 无应答呼叫前转

Z: 不可及呼叫前转

5.1.25VPN_FLAG

虚拟网标志

- 0: 非智能网用户
- 1: 智能网 PPC 用户
- 2: 智能网 WVPN 用户
- 3: 智能网 IVPN 用户
- 9: 其他智能网用户

5.1.26LONG_TYPE

长途类型

5.1.27CARRY_TYPE

承载类型

5.1.28CARRIER_CD

出访运营商代码

计费号码的运营商代码：国内 1:电信,国际填实际漫游地的运营商代码；非国际漫游建议固定填"CHN01"

5.1.29CFEE

基本通话费，单位为：分

5.1.30ACCT_ITEM_TYPE_A

费用 1 帐目类型

5.1.31LFEE

长途通话费，单位为：分

5.1.32ACCT_ITEM_TYPE_B

费用 2 帐目类型

5.1.33FEE_ADD

其他通话费，单位为：分

5.1.34ACCT_ITEM_TYPE_C

费用 3 帐目类型

5.1.35SERVID

用户标识

5.1.36BILLING_MODE

付费模式

付费模式（DOMAIN_BILLING_MODE）：18B：预付费；18A：后付费；18C
准实时预付费；18D：其他；（必须经全国中心批准才能用）

5.1.37EVENT_TYPE

事件类型

5.1.38PRODUCT_ID

商品标识

5.1.39ESN_CODE

电子序列号

5.1.40IMEI

移动设备识别码

5.1.41PROV_OFFER_ID

省份批价套餐

5.1.42PROV_RATEDATE

省份批价日期

5.1.43PROV_BILLING_CYCLE_ID

省份批价帐期

省份批价的帐期标识 YYYYMM

5.1.44CDR_KEY

话单主键

省份+业务标识+流水

5.1.45RATE_TIMES

全国中心批价次数

5.2分析和说明

5.2.1价值分析

IMSI 号可以关联到手机号以及客户自然人信息，前提是大数据中心和 ODS 数据对我们开放。

由于移动话单的产生为离散的，并不能把用户完整的时空轨迹匹配到地图上，而且基站数据字段也确实具体扇区信息，本表仅仅能做用户通话对象分析，关系网所在地分析等等。长途通话的时长能够对该用户的消费能力建模和职业画像有所帮助。