



基于信息汇聚平台的快数据应用设计与实现

赵彦杰

(中国电信股份有限公司上海分公司, 上海 200120)

摘要: 用户行为信息中蕴含着巨大的价值, 如何采集并加工用户行为信息, 使之成为新的业务增长点, 是运营商面临的重要课题。通过对信息汇聚、采集、加工等关键技术的研究, 给出了基于信息汇聚平台的快数据应用的实现方案, 能够通过用户开关机、位置变化、语音通话等行为触发用户关怀、企业名片、信息推送等业务, 该方案已经广泛用于电信运营商的新兴业务领域, 取得了显著的经济效益。

关键词: 移动通信; 大数据; 快数据

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016027

Design and implementation of fast data application based on information aggregation platform

ZHAO Yanjie

Shanghai Branch of China Telecom Co., Ltd., Shanghai 200120, China

Abstract: User behavior information contains great value. Telecom operators must collect and deal with this information to create new business growth point. Through the research of information collection and processing, the implementation scheme of fast data application was given. Services such as user care and information push were provided based on the behavior of position change and voice calls. The application was used in the new business field of China Telecom and achieved remarkable results.

Key words: mobile communication, big data, fast data

1 引言

工业和信息化部发布的通信运营统计公报显示, 2014 年中国手机用户数已经达到 13 亿户, 并且仍在不断高速增长。随着移动通信行业的巨大发展, 运营商越来越意识到, 用户行为中蕴含着巨大的价值, 除了对用户信息进行大数据挖掘, 分析用户特点、偏好之外, 如何围绕移动用户特有的实时行为数据, 提供对广大用户具有吸引力的快数据业务来提升服务体验, 是一个需要深入研究的课题。

快数据是基于用户行为的近实时数据流, 与普通大

数据处理动辄数十分钟至数小时的跨度不同, 其从产生、采集、加工到业务提供, 往往在数百毫秒至数秒之间, 需要在事件发生的同时立即做出反应与操作, 运营商目前的信息采集及分析系统并不具备这种实时事件的反应能力。此外, 运营商目前信令系统原始数据的来源主要是类 A 口信令, 因此针对移动用户信息分析局限于用户的开关机及忙闲状态, 存在信息源单一、数据挖掘不充分、接口不灵活等问题, 不能充分满足上层业务的需求。比如本地用户国际漫游回归欢迎词业务, 由于早期系统无法采集用户的漫游状态, 导致该业务需求不能快速实现。

为了更好地发挥信息提供阀门作用,必须设计一套新的信息汇聚系统,对更多的移动通信网络用户行为原始数据进行汇聚和采集,同时对采集的数据进行分析、过滤、加工和封装,最后通过标准的接口将数据传送至业务平台,为业务平台执行上层业务逻辑为用户提供个性化快数据服务打下坚实的基础,满足不断发展的业务需求。

2 信息汇聚平台功能架构设计

信息汇聚平台是通信网络用户行为消息汇聚、采集和加工的核心网元,其系统架构设计自下而上可分为 3 层:数据采集层、数据处理层和应用接口层。信息汇聚平台的系统架构如图 1 所示。

数据采集层负责从各类网元收集并汇聚如类 A 口事件(基站到移动交换机的信令)、深度分组检测信息、MAP (mobile application part, 移动应用部分)消息等源数据,并将所得到的数据存入原始数据库。消息解码内容包括开关机、主被叫、收发短信、周期性位置更新等。数据处理层负责对原始数据进行预处理、加工和关联,将所有信息梳理成用户维度的实时与非实时状态表,并将用户行为细节信息存入用户数据库。应用接口层负责按照信息汇聚平台定义的标准接口对外提供数据,包括 Web services 接口、

socket 接口、FTP 接口以及其他因为特定用户的特殊要求而实现的专用接口。接口层同时还具备各种数据访问逻辑以及安全控制,通过内部中间件向数据处理层发起业务请求,提供上层业务平台所需的信息查询、定制或主动推送服务,如单个用户实时状态查询、基于用户的实时状态变化通知、批量的基站维度查询以及支持号码、位置和状态 3 个维度的用户历史轨迹查询等。信息汇聚平台的对外接口见表 1。

3 业务网络数据采集及过滤

在传统电信网络中,信令监测只是一种故障定位和运维保障的手段,而信息汇聚平台的出现使得这些包含用户信息的信令可以为各种上层业务需求提供原始数据,因此,信令的采集和过滤技术显得尤为重要。

类 A 口信令是信息汇聚平台采集的重要数据,包含用户的开关机事件、起呼事件、短信事件、登记事件以及呼叫详细记录,其采集的原理是在基站到移动交换机的数据链路上跨接探头,以获取相关数据流量,跨接点一般选择在汇聚基站的数据设备上。

随着业务需求的扩充,信息汇聚平台还需要采集移动用户登记消息,用于配合上层消息推送平台进行漫游用

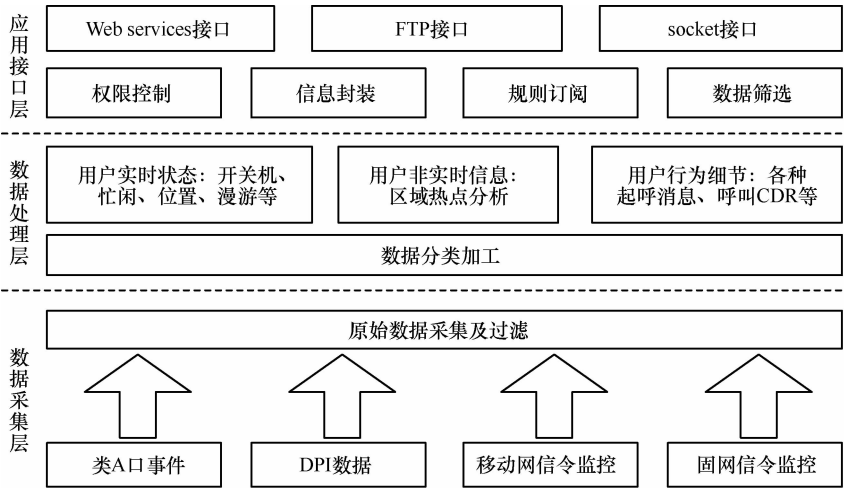


图 1 信息汇聚平台功能模块设计

表 1 信息汇聚平台应用接口

接口	内容	应用
查询单个/多个用户的信息	用户号码、移动用户识别码、基站号、扇区号、小区基站经纬度、忙闲状态、开关机状态	通过用户号码或识别码获取用户位置(扇区号、小区基站经纬度、小区基站经度)
查询某个基站/区域下的用户信息	用户号码集	结合用户位置信息,通过号码集进行信息推送或业务推广
当用户状态发生变化时通知	开关机状态、忙闲状态、位置信息、国际漫游	根据用户状态变化,触发用户关怀、业务提醒等服务
实时信令数据过滤转发接口	用户号码(语音通话、收发短信等)	根据用户行为,通过主被叫号码进行信息推送或业务推广



户的欢迎词服务,此类消息一般可以从运营商已有的移动网信令监测系统获得。移动网信令监测系统从原始信令中过滤出登记通知消息,进行解码、分析、合成,根据一定的规则生成相关的记录(包括移动用户识别码、用户号码、时间等信息),通过北向 socket 接口实时发送给信令汇聚平台。信息汇聚平台通过采集层收到移动网信令监测系统的数据后,经过统一封装模块,将数据发送给上层业务平台。

数据过滤是数据采集技术中的重要环节,比如根据主被叫号码或指定区域进行过滤,根据不同的需求,采用不同的处理模式。与其他信息采集系统采用先存储后处理的方式不同,信息汇聚平台采用基于实时事件的数据流处理方式,真正满足了快数据业务实现的高速性。在信息汇聚平台基于基站列表的过滤功能中,可以采用混合过滤模式,即各类业务统一对原始数据进行扫描,按照不同的过滤要求同时查找各类规则,最后生成针对不同业务的输出。这种方式下,系统扫描数据时需要对全部规则进行匹配,实时性稍差,但适合以文件方式传输的准实时业务需求,如用户关怀、信息推送等业务。混合过滤方案的实现模式如图 2 所示。

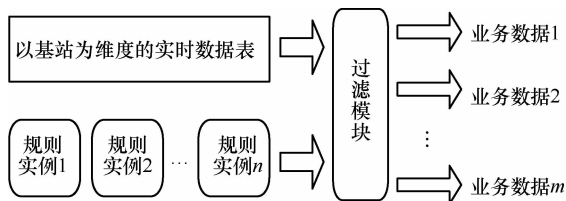


图 2 混合过滤方案的实现模式

为了保证数据的实时输出,企业名片、挂机短信等高实时性业务建议采用以业务为单位的链式过滤,对每个业务配置独立的过滤器。这种面向业务的独立处理方式针对各个业务独立进行处理,实时性高,但是每次有新的业务需求时,都需要新建过滤器,适合于数据源单一、数据加工简单的业务。

4 基于通话行为的快数据应用设计

通过信息汇聚平台对外开放的信令数据,可以获取用户的实时通话行为,并以此设计挂机短信与来电名片业务,这也是电信运营商面向中小企业用户推广的重点新兴业务。

挂机短信的场景是当移动用户拨打签约此项业务的企业号码后,企业可识别拨打行为并向主叫发送预设的

短信,作为企业维系客户的手段。要实现此项业务,信息汇聚平台必须采集移动网原始信令,对 MAP 消息进行分拣和合成,获取所有移动用户的呼叫情况,同时结合类 A 口的被叫侧起呼和挂机事件,补充相关呼叫的状态,最后根据企业清单过滤出主叫号码发送给上层业务平台进行预设短信的发送。对于用户起呼、终呼在不同移动交换机的场景,则需要信息汇聚平台将各个交换机的详单及所有的切换消息进行全局的匹配与合成,从而形成正确的跨域挂机事件。

来电名片业务的申请者为固话用户,当申请该业务的用户拨打其他用户时,将同时给被叫发送闪信/短信,短信内容包含该用户预设的名片信息。来电名片的数据来源主要是固网到关口局的 IAM(initial address message,起始地址消息),对于已完成软交换改造的电信网络,需调整为采集 SIP(session initiation protocol,会话初始协议)消息。信息汇聚平台根据 IP 地址实时采集各个软交换汇接局到关口局的起呼消息,根据主被叫进行解码,获取签约号码的呼叫行为清单后发送给业务平台。挂机短信及来电名片业务的实现流程如图 3 所示。

其具体实现流程如下:

- (1)主叫号码 A 呼叫被叫号码 B;
- (2)信息汇聚平台通过数据采集层的探针获取呼叫发起和振铃消息事件;
- (3)信息汇聚平台数据处理层对采集到的原始消息事件进行预处理;
- (4)数据处理层将消息关联成用户号码维度的实时状态表,根据消息内容补充被叫状态;
- (5)数据处理层根据主被叫号码的业务订购关系过滤相关事件;
- (6)信息汇聚平台应用接口层根据事先定义的标准接口将数据发送至上层业务平台;
- (7)来电名片平台将主叫事先定义的名片信息通过闪信发送给被叫号码;
- (8)B 应答后,A、B 双方通话,通话结束后双方挂机;
- (9)信息汇聚平台通过数据采集层的探针获取呼叫应答和挂机消息事件;
- (10)信息汇聚平台数据处理层对原始数据进行加工、关联和过滤;
- (11)信息汇聚平台应用接口层通过标准接口将数据发送至上层业务平台;

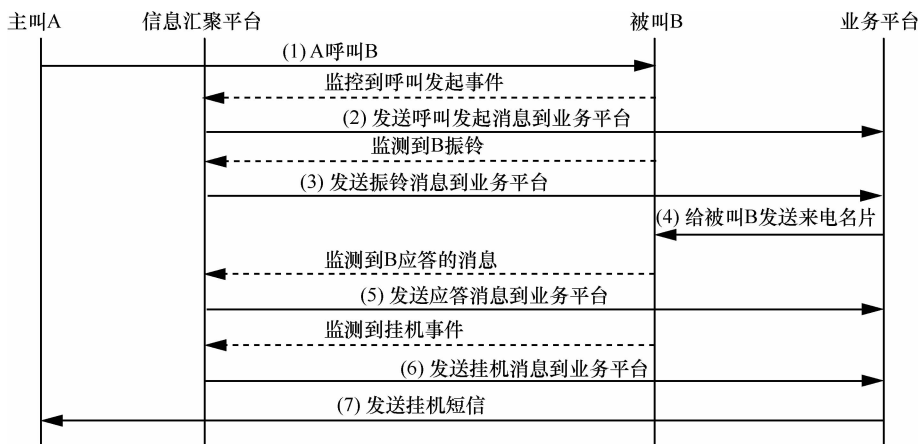


图3 通话快数据业务实现流程

(12) 挂机短信平台将被叫事先定义的推广信息通过短信发送给主叫号码。

为了维护用户行为隐私,对外提供的数据都需要按照实际申请业务的用户进行发送。用户清单的获取方式有两种:一是上层挂机短信或来电名片平台定期(每天)提供全量或新增的用户清单,信息汇聚平台根据新的用户清单对数据进行过滤;二是信息汇聚平台通过客户关系管理系统获取新开通业务的用户清单定时更新。信息汇聚平台与上层业务平台采用 socket 方式进行通信,在正常情况(网络带宽充足、数据接收服务器未堵塞数据等)下,信息汇聚系统解码过滤分析处理后向业务平台标准接口发送数据的时延可以控制在收到起呼或终呼消息后的 2 s 之内,相当于用户呼叫或挂机后马上就能收到消息推送,完全可以满足快数据应用的需要。

5 基于位置信息的快数据应用设计

位置信息是移动互联网应用的核心要素,通过信息汇聚平台类 A 口信令中包含的基站和扇区信息,能够准确地获取用户所在的小区位置。一般在城市市区或郊区城镇区域,一个基站覆盖的范围为半径 500 m 的区域,而在比较空旷的农村区域,一个基站的覆盖范围在 1.5~3 km。类 A 口信令可以给出用户所在小区对应基站的 IP 地址,通过查询基站数据库的映射关系,就可以找到对应的基站名称、地址及经纬度信息。

除了对单个用户进行定位外,信息汇聚平台还可以结合时间轴给出用户的动态轨迹,或者过滤出处于某个基站或某片区域的用户集,开展信息推送类的快数据应用。运营商可以根据合作方提供的签约用户号码,查询用户所在的

商圈,同时从大数据系统获取用户的各类标签,结合用户位置和标签进一步筛选和分类用户,针对不同用户推送其感兴趣的信息。基于位置信息的实时信息推送应用过程设计如下:

(1) 业务发布人员在上层管理平台创建任务、定制业务触发规则;

(2) 管理平台批量导入指定区域的基站列表,自动同步到信息汇聚平台;

(3) 信息汇聚平台生成指定区域采集自动机,采集区域内的用户位置变化信息;

(4) 用户进入指定区域,向移动交换机发起位置登记及状态更新消息;

(5) 信息汇聚平台通过数据采集层的探针获取到用户登记消息;

(6) 信息汇聚平台数据处理层对采集到的原始消息事件进行预处理;

(7) 数据处理层根据预设的业务触发规则及基站列表过滤相关用户事件;

(8) 信息汇聚平台通过大数据平台进行特征匹配,打上用户喜好标签;

(9) 信息汇聚平台应用接口层通过标准接口将号码集发送至小区短信或信息推送平台;

(10) 上层业务平台根据信息汇聚平台上报的号码集进行信息推送。

对于指定区域信息推送的业务,要特别注意对该区域的常驻用户进行过滤,以免对用户形成骚扰。常驻用户一般有两种类型:该区域居住用户(驻留时间为夜晚)和该区域上班用户(驻留时间为白天)。信息汇聚平台需要对常驻用户



进行定义,如连续 N 天 A 时刻- B 时刻出现在指定基站或者扇区下的用户, N 、 A 、 B 是可配置参数,一般居住用户设置成连续5天凌晨0:00-6:00、上班用户设置成连续5天9:00-16:00。数据挖掘模块定期分析指定区域的常驻用户,生成文件放在指定目录下,服务提供层根据业务需求,对需要过滤常驻用户的数据推送,会自动获取对应的用户清单,实现常驻用户过滤的功能。

6 结束语

随着移动网络建设的不断完善,通信网络覆盖和用户数都在大幅增长。在这种形势下,电信运营商除了提供用户基础的语音、短信、上网业务之外,还需要进一步提供更加丰富和全面的增值服务来吸引用户,提高自身的竞争力,创造新的价值增长点。信息汇聚平台具有强大的网络数据采集能力和丰富的对外接口,通过信息采集和汇聚可以实现小区短信、漫游用户欢迎词、公益信息发布、企业来电名片、挂机短信等多种实时的快数据应用,为运营商有效利用用户行为数据开展快数据业务起到了示范作用,取得了良好的经济效益和社会效益。如何对采集的信息进一步分析挖掘,实现移动通信网络行为的大数据分析,与快数据应用结合开展业务,是下一步研究的重点课题。

参考文献:

- [1] 郑凤. 移动互联网技术架构及其发展 [M]. 北京: 人民邮电出版社, 2013.
ZHENG F. Mobile Internet Technical Architecture [M]. Beijing: Posts & Telecom Press, 2013.
- [2] 尹立强, 张海燕. 移动通信[M]. 西安: 西北工业大学出版社, 2013.
YIN L Q, ZHANG H Y. Mobile Telecommunication [M]. Xi'an: Northwestern Polytechnical University Press, 2013.
- [3] 张传福. 移动互联网技术及业务[M]. 北京: 电子工业出版社, 2012.
ZHANG C F. Mobile Internet Technology and Business [M]. Beijing: Publishing House of Electronics Industry, 2012.
- [4] 涂子沛. 数据之巅: 大数据革命, 历史、现实与未来[M]. 北京: 中信出版社, 2014.
TU Z P. Big Data History, Reality and Future [M]. Beijing: China Citic Press, 2014.

[作者简介]



赵彦杰(1982-),男,中国电信股份有限公司上海分公司工程师、交换网络处副处长,主要研究方向为业务平台的维护管理及产品支撑。