



大数据背景下运营商建设统一 DPI 系统的思考

侯慧芳, 潘洁

(中国移动通信集团设计院有限公司, 北京 100080)

摘要: 移动数据流量业务逐渐增大, 运营商的经营模式正在从“话务量经营”向“流量经营”转变, “大数据经营”将成为运营商的主要发展方向。运营商积极探索 DPI 技术对于大数据运营的价值和应用方向, 制定标准的统一 DPI 系统规范, 为大数据运营提供全面、标准的流量数据。系统分析了传统 DPI 部署方式存在的弊端, 提出建设统一 DPI 系统对推动运营商大数据发展提供的重要基础数据保障作用。

关键词: 统一 DPI; 大数据; 电信运营商

中图分类号: TN919.5

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017107

Discussion of constructing unified DPI system by operators in the background of big data

HOU Huifang, PAN Jie

China Mobile Group Design Institute Co., Ltd., Beijing 100080, China

Abstract: Mobile data traffic increases gradually. Operator's business model is transferring from “telephone traffic management” to “traffic management”. Big data management will become the main development direction of operators. For big data management, operators explore the value and application of DPI technology and make the unified DPI system standards to offer comprehensive and standard traffic data. The insufficiency of traditional DPI deployment mode was analyzed, and the unified DPI system to offer basic data security for accelerating development of big data was presented.

Key words : unified DPI, big data, telecom operator

1 引言

随着移动互联网业务的快速发展, 数据业务取代了语音和短信业务成为运营商的收入支柱, 运营商积极探索经营新模式, 从“话务量经营”向“流量经营”转变, “大数据经营”将成为运营商的主要发展方向。国内互联网公司已涌现出

一批先行者(例如阿里巴巴、百度、腾讯等)积极开展大数据的研究和开发, 电信运营商比互联网公司拥有更多的天然优势, 因为电信运营商本身掌握了丰富的基础数据资源。

对用户话单和网络信令中的信息进行提取分析是大数据运营的基础, 而运营商传统的 DPI (deep packet inspection) 部署方式以系统为单位进



行建设,存在重复建设、重复存储、重复分析以及数据格式不标准的弊端,分散的信息形成一个个信息“孤岛”,不能合成全面完整的用户信息^[1]。统一 DPI 系统对传统 DPI 系统存在的问题加以改进和规范,着眼于部署位置合理化、设备功能规范化和数据格式标准化,为大数据运营提供准确丰富的基础数据,推动大数据应用的建设发展。

2 传统 DPI 系统存在的问题

DPI 是一种基于数据分组的深度检测技术,针对不同的网络应用层载荷(例如 HTTP、DNS 等)进行深度检测,通过对报文的有效载荷检测决定其合法性。DPI 设备通过对网络关键点的流量和报文内容进行检测分析,可以根据事先定义的策略对检测流量进行过滤控制,能完成所在链路的业务精细化识别、业务流量流向分析、业务流量占比统计、业务占比整形、应用层拒绝服务攻击、对病毒/木马进行过滤和控制 P2P 的滥用等功能。

传统 DPI 设备以业务系统为单位进行建设,每个业务系统在同一采集点均建设了自己的 DPI 设备,并且不同采集点、不同业务系统和不同厂商对 DPI 设备采集的数据格式并未进行统一规范,造成传统 DPI 系统存在重复采集、重复存储、重复分析和数据格式不标准等弊端。

2.1 重复采集

每个业务系统都建设了从数据采集层到应用层的一整套设备,因此全网 DPI 系统的数量就等于全网业务系统的类别数量,而在数量庞大的 DPI 设备中,有约 70% 存在数据重复采集的现象。DPI 系统的烟囱式部署,导致了多点建设 DPI 位置重复、同一位置建设多套 DPI 功能重复的问题,也给运营商造成了巨大的投资浪费。

2.2 重复存储

传统的 DPI 系统部署时存在同一采集点部署

多套 DPI 设备的情况,当同一采集点的不同应用系统均需要进行数据存储时,数据被分别存储于各个应用系统,就会产生数据重复存储的现象。而全网的每个采集点至少都部署了监测系统和管理系统,再加上各类个性化应用,运营商重复存储的数据量会非常庞大。

2.3 重复分析

传统的 DPI 系统在建设时只考虑各应用系统自身的数据采集需求,不同的应用系统可能对数据进行了相同方式的采集和分析,只是应用系统将分析数据进行了不同层面的展示,造成了数据重复分析、应用重复开发的问题,给运营商带来较大的投资浪费。

2.4 数据格式不标准

由于不同通信设备厂商 DPI 设备实现方式不同,导致数据呈现未遵循统一格式;以业务划分方式为例,不同厂商对业务划分颗粒度和归属关系的定义不一致,导致 DPI 设备识别的业务命名、定义和数量均不同。

不同采集位置数据格式 TLV(类型(type)、长度(length)、值(value))缺乏统一规划,难以关联,无法端到端分析;以移动用户使用微信业务为例,传统烟囱式 DPI 部署方式下,若 Gn 口和省网出口两个位置得到的用户信息并不能完全对应,就无法完成用户数据端到端的分析。

不同业务系统定义的业务分类体系不同,使得不同业务系统 DPI 设备识别的基础数据无法达成一致。因此,无论从运营业务系统本身对业务数据分类的定义,还是通信设备厂商的实现方式,都没有做到统一规范,造成各个系统输出信息的字段和格式不完全相同,各字段的内容描述和具体取值不完全相同,导致数据格式不一致,各说各话。

由于 DPI 系统烟囱式部署、无序规划、重复建设,导致在网络安全、运行维护、建设投资和有效利用等方面存在一系列问题,同时数据

格式的不一致导致无法端到端呈现全局数据, 难以以为大数据平台提供基础数据。没有统一规划的 DPI 发展模式, 在未来大数据时代将导致大量的投资浪费、能力不足等问题, 需从全网角度对 DPI 设备的部署进行统筹规划。

3 统一 DPI 系统建设的必要性

为实现 DPI 部署位置合理化、设备功能规范化、数据格式标准化, 运营商对全网 DPI 系统进行整合, 建设统一 DPI 系统。统一 DPI 是指一套 DPI 设备对链路上的流量进行采集与识别, 并将满足其他系统所需的流量和分析统计数据分发给各第三方应用系统服务器。网络中各类流量管理相关的应用系统, 主要由 DPI 设备和应用服务器组成。其中, DPI 设备负责数据采集、流量分析统计、日志合成, 应用服务器主要完成对数据的进一步分析处理, 合理组织和存储数据, 并进行呈现。统一 DPI 整合了多个第三方应用系统的 DPI 设备, 通过 DPI 复用为多个系统提供 DPI 能力。

3.1 统一位置

运营商网内流量, 从用户出发, 按照接入网、城域网、省网和骨干网的顺序在网内传输, 由此, 在全网 5 个点进行数据采集, 部署 DPI 设备可捕捉网内流量全貌, 如图 1 所示。

同一个采集点上采用一套统一 DPI 设备为多套应用系统提供数据, 实现 DPI 复用。通过减少采集点和 DPI 设备数量, 实现多系统间同一采集点的 DPI 设备共享, 力争通过最少的采集点获取最全的数据。

3.2 统一格式

通过制定相关规范实现数据标准化, 并确保运营商各部门、各省公司采用相同规范制定数据流量的标准日志格式和内容, 要求各厂商执行相同的数据标准, 对于 DPI 七层识别的全部结果, 实现 TLV 全字段的统一和对齐, 如图 2 所示。

除此之外, 制定相关规范, 规定统一的业务分类, 实现最细粒度的统一业务划分规则, 在各采集点、各厂商、各分析系统之间形成通用语言, 奠定大数据端到端分析的基础。

3.3 统一功能

按照应用系统所需数据内容和格式的不同, 统一 DPI 系统可提供 4 种数据复用方式。对于系统特征库特殊、不易整合的应用系统, 可通过链路分光器或统一 DPI 设备的分光功能, 将光路信号完整地复制分发给应用系统; 对于明确需要某种报文的系统, 统一 DPI 设备将全部原始报文或者经过特定条件过滤后的部分原始报文复制分发给应用系统; 对于以分析流量日志为主的应用系

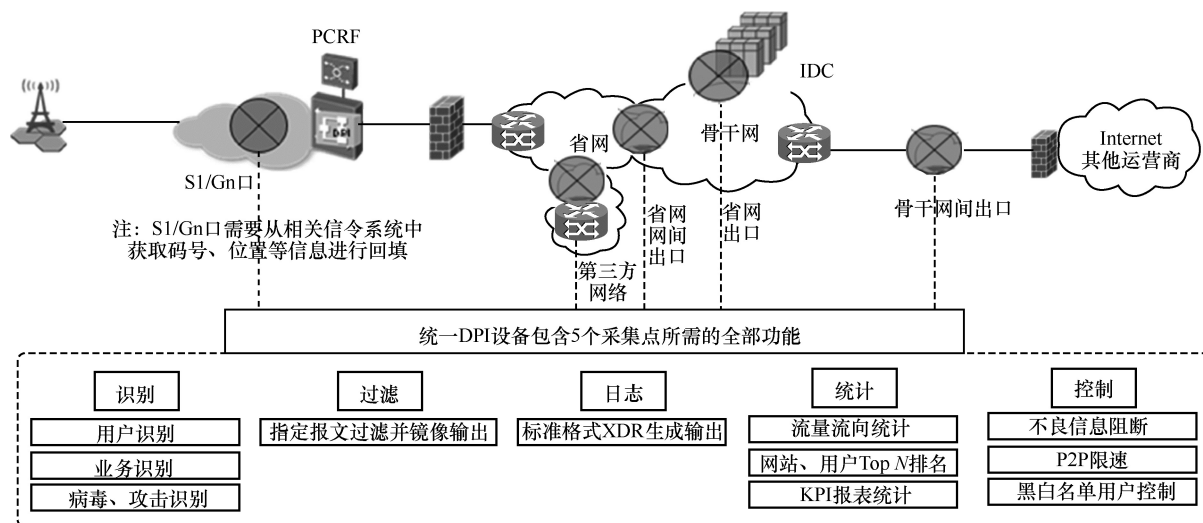


图 1 统一 DPI 系统部署位置

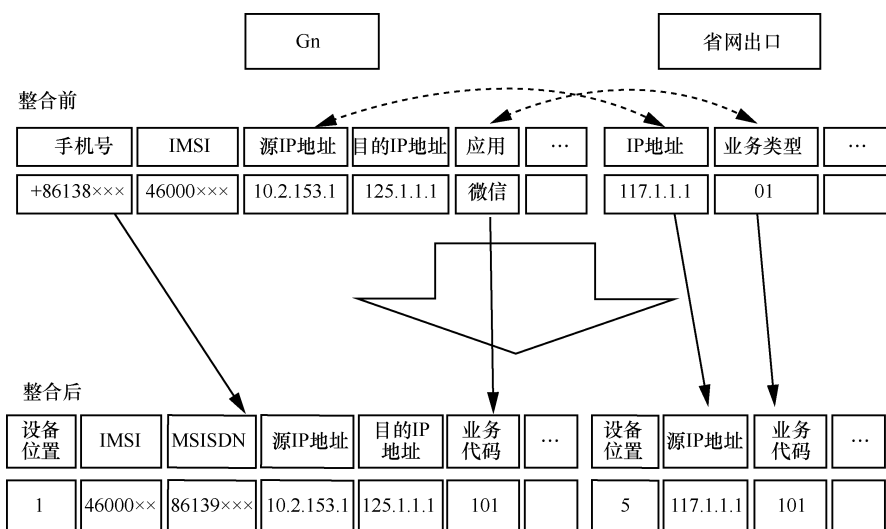


图2 统一 DPI 系统数据格式规范化示例

统，统一 DPI 设备分析记录流量的关键信息形成 XDR（X data decoding）话单，并将会话数据按指定格式发送给应用系统；对于需要简单明确的统计指标的应用系统，DPI 设备统计流量的各项指标，并将统计结果发送给应用系统。统一 DPI 系统数据复用方式如图 3 所示。

统一 DPI 系统定位于为上层应用提供全面的基础数据，除了满足各类应用系统所需数据的镜像需求，也能完成一些简单的数据处理操作，而个性化、复杂的数据处理则由上层应用系统和大数据平台实现。统一 DPI 系统对流量采集和数据解析功能进行了统一规范，打破 DPI 设备和上层应用系统间的耦合，形成开放、标准的 DPI 设备体系。

4 统一 DPI 系统为大数据应用提供重要数据保障

当今信息产业呈现出前所未有的繁荣，新的互联网技术不断涌现。从传统互联网的 PC 终端，到移动互联网的智能手机，再到物联网传感器，技术革新使数据生产能力呈指数级提升，而运营商是用户访问互联网的管道建设者，拥有多维度、丰富、持久的互联网用户数据，在洞察用户方面拥有得天独厚的优势，管道中海量的用户上网行为数据是运营商可以开采的下一个金矿，而统一 DPI 系统正是开启这座“金矿”的钥匙^[2]，如图 4 所示。

运营商智能管道中大数据应用的目标是实

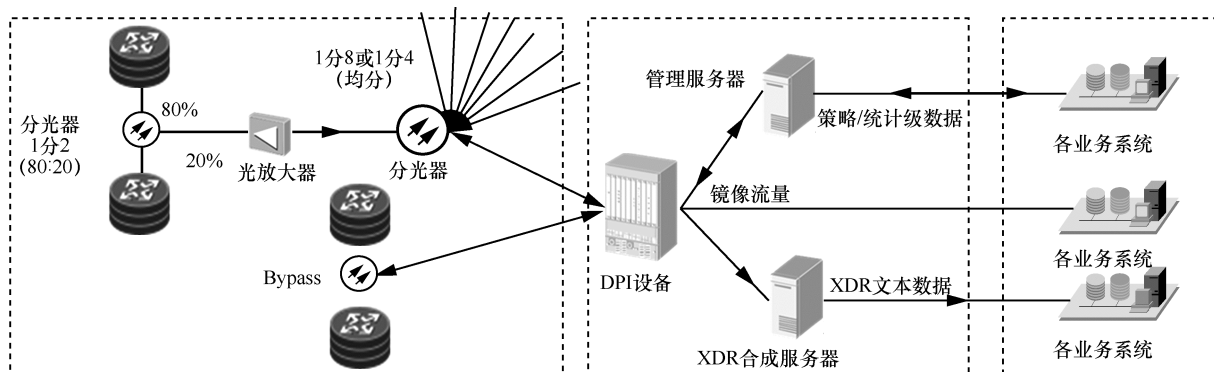


图3 统一 DPI 系统数据复用方式



图4 统一 DPI 系统作为用户和大数据应用联系的纽带

现复合关联、泛在挖掘、快速实时，把分散在网络中各个环节的数据汇聚拼接出用户的全貌，把握用户的真正需求，挖掘潜在用户群的行为特征，利用大数据洞察人的数字生活，并及时地反映到对用户的服务上^[3]。而统一 DPI 系统实现全网数据的采集识别，处于网络中的采集层，提供全景化的用户数据，是实现大数据应用的第一步。大数据应用从采集层到应用层的整体架构如图5所示。

统一 DPI 系统的建设为大数据应用的发展奠定了强大的数据基础，使运营商摆脱沦为单纯“管道”的尴尬局面，其提供的全面化、规范化、系统化的用户信息，为大数据应用提供了精准的数据来源，提升用户体验，进一步促进了大数据更好地发展。

4.1 全面采集，精准分析

统一 DPI 系统通过对链路数据一次采集多次复用，为第三方大数据咨询业务提供数据来源，

运营商可以开展一些基于网络流量的新业务，统一 DPI 系统提供灵活的镜像方式，将符合特定要求的用户、协议、应用的流量镜像给大数据中心和第三方系统，以支撑新的业务的开展，如缓存系统、数据挖掘系统、精准广告系统、防病毒系统及其他第三方系统，如图6所示。

统一 DPI 系统通过对全网流量的采集，掌握流量整体趋势，了解网内、网间流量流向，重点业务/网站流量占比等指标，通过流量流向可视实现热点应用分析；通过业务应用/协议的流量分布及发展趋势，ICP、域名、URL 排名实现业务热点洞察；通过对用户访问日志的采集分析挖掘用户特征，了解用户网络使用情况，分析用户流量组成和流向、用户兴趣和热点资源，实现用户行为可视化，为个性化差异服务提供数据支撑；通过对网络时延等网络质量的监控，改善用户体验；通过检测垃圾邮件、僵尸、DDoS 网络、手机病毒、恶意网站和不良内容，了解安全威胁，实现安全

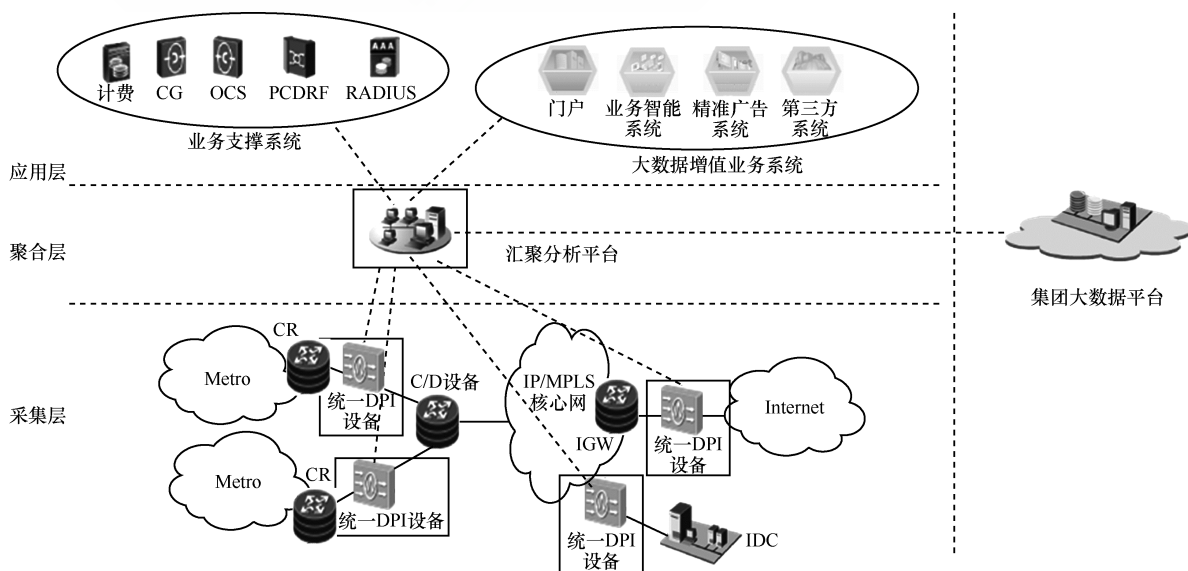


图5 大数据应用功能实现架构

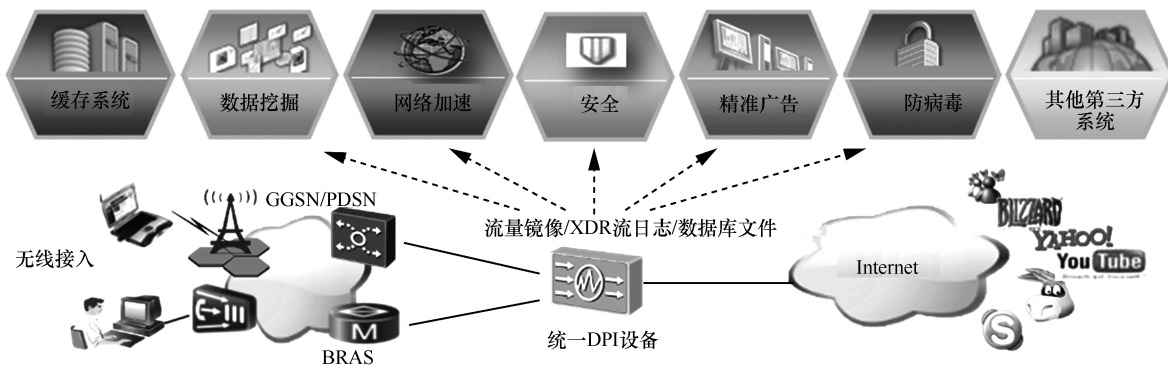


图6 统一 DPI 系统为大数据应用提供数据支撑

威胁可视化，为网络安全防护提供支撑。

4.2 建立网络应用模型

统一 DPI 系统为大数据中心提供海量用户数据，大数据应用通过基于用户的上网行为数据提取自然人特征，并为其建立包含各种属性的关系链，基于特征标识库，对海量数据进行行为分类；挖掘用户使用习惯，创建常规应用模型，实现用户画像标签，如图 7 所示。

以广告推送为例，如今大数据已应用于支撑精准广告的投放，统一 DPI 系统实现网站分析统计及搜索关键字抓取，提供用户上网行为数据，大数据中心结合强大的数据挖掘技术与行业知识库，对 DPI 网元覆盖区域的终端用户进行用户行为分析，建立精准的用户画像，为信息推送（广

告）服务平台的信息推送策略提供数据依据，进而结合用户需求，在正确的时间窗口触达用户，实现针对用户的精准营销^[1]。

大数据中心的构建与用户上网行为的挖掘分析，除了支撑精准广告的推送，未来在市场调研、金融征信等领域也有很广阔的应用前景。

4.3 基于统计数据的经营建议

运营商基于统一 DPI 系统提供的统计数据，为运营商提供多样化经营建议，在开源节流方面提供数据支撑，促进业务模式创新，实现业务的增值运营，为管道注智，为流量增值。

运营商针对用户的上网行为，分析用户可能的离网倾向，制定营销策略，推送优惠信息，提前防止用户流失；通过展现手机视频、手机阅读、

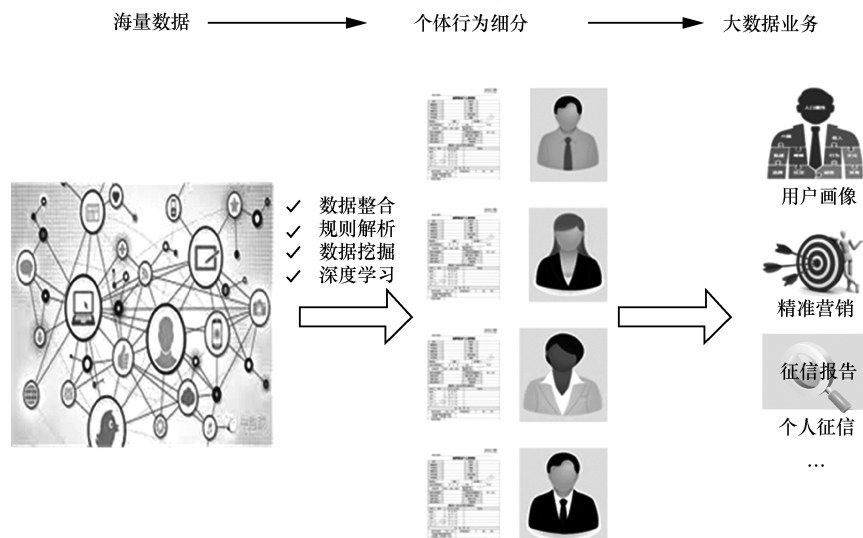


图7 基于海量数据构建用户模型

应用商店等各类应用的趋势、占比和排名,更好地了解网络流量构成,为精细化运营提供数据支撑;通过分析运营商传统语音、短信业务和数据流量业务占比,为运营商业务管控策略提供数据依据;通过精准分析,实现终端类型可视化,了解终端组成情况,分析手机终端类型、操作系统类型、浏览器类型和对应的流量和用户数,为业务营销提供数据支撑^[3]。

除内部服务外,运营商还可以向行业用户、外部合作伙伴提供数据出售、数据咨询、数据能力和数据解决方案4种业务形态的数据服务。通过对外合作的模式,可融合合作伙伴的数据内容,根据合作伙伴和客户的需求,将电信的数据资产进行包装,向合作伙伴提供数据服务,开展大数据运营。

目前运营商在部分省份建设大数据中心,但仍存在跨省、跨域数据分析难的问题存在,整合跨省、跨域数据,建设集团层面的统一的大数据中心,支撑业务流程端到端分析和面向移动互联网的大数据分析仍是未来大数据中心建设的目标。

5 结束语

本文通过分析运营商传统 DPI 部署方式存在的问题,提出在大数据应用快速发展的大背景下,运营商整合全网的 DPI 资源,建设统一 DPI 系统的必要性,同时对统一 DPI 系统在部署位置合理化、数据格式规范化和实现功能统一化方面做出的改进进行了阐述,提出统一 DPI 系统作为实现大数据应用第一步的重要作用,其提供的全面化、规范化、系统化的用户信息,保证了运营商实现

用户数据的全面采集、精准分析,帮助运营商建立网络应用模型,有效的用户统计数据更为运营商提供了合理的经营建议。因此,统一 DPI 系统为大数据的应用发展奠定了强大的数据基础,是推动运营商大数据发展的重要保障。

参考文献:

- [1] 毋毅. 面向通信智能管道运维体系的统一 DPI 网络流量大数据分析平台研究[J]. 信息与电脑, 2016(18): 159-160.
WU Y. Research on the big data analysis platform of unified DPI traffic for communication intelligent pipeline operation and maintenance system[J]. China Computer&Communication, 2016(18): 159-160.
- [2] 冉萌, 韩玉辉. DPI 技术在移动大数据中的应用[J]. 邮电设计技术, 2016(8): 33-36.
RAN M, HAN Y H. Application of DPI technology in mobile in big data[J]. Designing Techniques of Posts and Telecommunications, 2016(8): 33-36.
- [3] 殷鹏. 智能管道运营中大数据的应用分析[J]. 电信技术, 2013, 1(4): 47-49.
YIN P. Analysis of the application of big data in intelligent pipeline operation[J]. Telecommunications Technology, 2013, 1(4): 47-49.

[作者简介]



侯慧芳(1986-), 女, 中国移动通信集团设计院有限公司工程师、咨询设计师, 主要研究方向为数据网络。



潘洁(1978-), 女, 中国移动通信集团设计院有限公司工程师、高级咨询设计师, 主要研究方向为数据网络。