ISSN 1673-9418 CODEN JKYTA8
Journal of Frontiers of Computer Science and Technology
1673-9418/2015/09(04)-0451-11
doi: 10.3778/j.issn.1673-9418.1405025

E-mail: fcst@vip.163.com
http://www.ceaj.org
Tel: +86-10-89056056

网络流量可视化的新方法*

高丕红+,徐明伟
清华大学 计算机科学与技术系,北京 100084

New Method of Network Flow Visualization*

GAO Pihong+, XU Mingwei
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
+ Corresponding author: E-mail: danceintheair57@gmail.com

GAO Pihong, Xu Mingwei. New method of network flow visualization. Journal of Frontiers of Computer Science and Technology, 2015, 9(4):451-461.

Abstract: Flow visualization is of great help for network researchers and supervisors to detect network anomalies, understand the characteristics of network traffic and find time trend, etc. Existing flow visualization technologies have failed to fully meet the growing needs of traffic visualization. Aiming at time, dimension and structure characteristics of flow, this paper adopts spiral visualization technology, dimension classification method and space filling technology to meet the need of different characteristics of flow. By visualizing the data from 4over6 project and compared with the methods currently used in the project, the results show that spiral visualization can excavate the time characteristics of traffic data more effectively than line mapping; dimension classification is more effective for special dimensions visualization than parallel coordinates and space filling can express node information with no loss of structure characteristics which node-link mapping can't achieve.
Key words: visualization; 4over6; traffic characteristics; derived data

摘 要:流量可视化对于网络研究和监管人员侦测网络异常,了解网络中流量的特征和趋势等有着重要的意义。现有的流量可视化技术不能满足日益广泛的流量可视化需求。针对流量数据的时间、维度、结构等特征,

* The National Natural Science Foundation of China under Grant No. 61161140454 (国家自然科学基金); the National High Technology Research and Development Program of China under Grant No. 2011AA01A101 (国家高技术研究发展计划(863计划)); the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20120002110060 (高等学校博士学科点专项科研基金).
Received 2014-05, Accepted 2014-11.
CNKI网络优先出版:2014-12-11, http://www.cnki.net/kcms/detail/11.5602.TP.20141211.1049.004.html

分别提出了螺旋可视化技术、维度分类方法和空间填充技术来满足不同特征下的可视化需求。通过对4over6项目中流量数据可视化，并与目前使用的技术进行对比分析，得知时间特征中螺旋线技术较直线映射技术更易发现数据周期特性；维度特征中，维度分类较之于平行坐标技术，更能适应特殊维度要求；结构特征中，空间填充技术较之于仅呈现结构特性的结点链路技术，能同时实现结点信息的呈现。

关键词：可视化；4over6；流量特征；派生数据

文献标志码：A 中图分类号：TP393

1 引言

随着互联网的广泛使用，分析和研究网络流量越来越重要。然而，在网络数据日益膨胀的今天，通过分析日志记录和网络数据等方式研究网络流量已经日趋困难。信息可视化技术[1]能够利用人类视觉感知系统将数据以图形化的方式展现出来，辅助研究人员了解数据特点，发现数据中内在规律。与传统方式相比，它可以有效地把网络研究人员和管理人员从复杂繁多的流量数据中解放出来。

直线映射[2]、平行坐标[3]、结点链路[4]等技术是目前流量可视化中使用最为广泛的技术。这些技术对于发现流量数据的基本特征有着不错的效果，但不能满足深入挖掘流量特征和特殊情况下（如流量的周期性）的需求。本文以在特定情况下对流量数据特征的挖掘为目标，引入了目前尚未在流量数据分析中使用的3种可视化技术（螺旋可视化技术、维度分类方法、空间映射技术）。通过对4over6（IPv4 over IPv6）项目[5]中流量数据的分析，得出新的可视化技术对于流量的特征挖掘有着更好的效果。

本文组织结构如下：第2章介绍了可视化过程和流量可视化的现状和不足，提出了3种新的改进方法；第3章详细描述了使用3种新方法对流量数据进行可视化的过程；第4章通过对比新方法和传统方法，得出新方法比传统方法更能满足特殊情况下的可视化需求；第5章是总结和展望。

2 流量可视化研究背景

2.1 可视化管线模型

可视化管线模型是根据用户的目标，选择合理的状态转换操作，把数据从源数据状态转换为视图状态的过程。该模型经过了多个可视化领域学者不断改进，基本模型如图1所示[6]。

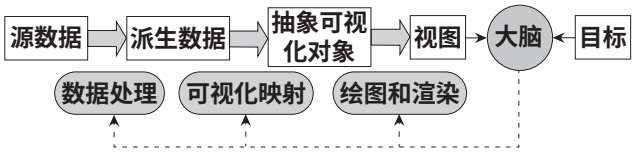


Fig.1 Visualization pipeline

图1 可视化管线

可视化管线包括4个数据状态（图1中白色方框）、3个状态转换过程（图1中灰色方框）和用户交互（图1中虚线）3个单元。

状态转换过程是连接各个数据状态的桥梁，是可视化管线重要的组成部分。其中，数据处理过程是把源数据转换为派生数据的过程，包括数据转换和过滤两部分。首先根据可视化任务对数据进行聚类、抽取，得到有效的结构化信息，再过滤得到需要的数据项。可视化映射过程是把派生数据映射到抽象可视化对象，利用点、线、颜色、位置、大小等可视化对象属性描述数据属性。可视化绘制过程是把抽象可视化对象转化为视图，通常借助于现有绘图工具实现绘制过程。视图通过视觉感知系统进入人脑，并与用户的目标相比较。若有不满，则不断调节任务，直到满足用户的目标为止。

2.2 流量可视化的不足及新方法的提出

在分析流量可视化现状之前，有必要对可视化技术的分类进行阐明。可视化技术是指在可视化映射过程中选择的技术。从图1中可以看出，可视化技术的选择与派生数据有很大的关系。根据可视化目的侧重的不同，派生数据可从时间、维度、结构3个特征进行分类，如表1所示（其中值数据类型的3种分类适用的可视化映射技术各有不同）。

Table 1 Classification of derived data

表1 派生数据的分类

数据分类特征	具备此特征的数据类型	不具备此特征的数据类型
时间特征	时间序列	非时间序列
维度特征	多维	低维
结构特征	树，图	值（可分类，可比较，可量化）[7]

因此流量可视化技术也可分为基于时间、流量和结构特征3类。在流量可视化中，时间特征是指对流量数据随时间变化情况的可视化；维度特征是指对流量数据的多个属性的可视化；结构特征是指对不同结点间流量的分布和流动情况的可视化。可视化不同特征决定了派生数据的划分类型，从而也决定了可视化技术选择的范围。3种特征通常不会同时呈现。如以反映数据时间特征为目的，则需降低数据维度并只能针对值类型数据进行可视化。即反映数据的时间特征时，在数据处理时生成的派生数据只具有时间特征。本文后续涉及到的可视化技术实现均是基于某一特征，不涉及多种特征的结合。

2.2.1 流量可视化不足

目前互联网中流量可视化应用较广泛，大量的网络监管工具中集成了可视化的功能。下面从时间特征、维度特征和结构特征3方面来介绍目前流量可视化的现状和不足之处。

在流量时间特征可视化中，大量的流量可视化工具均采用直线映射技术。文献[8]总结了大量的网络流量监控和分析工具，这些工具对于时间特征的可视化均是把时间数据映射到直线上，方便发现流量数据的变化趋势。然而，直线映射技术对流量的周期特征没有进行有效的探索和挖掘。

在流量维度特征的可视化中，使用较多的是平行坐标技术。文献[9]以观察局域网内部与外部其他主机数据传输所产生的流量分布为目标，使用平行坐标技术对流量进行可视化，反映流量数据基于源、目的地址的聚类情况。该技术把维度数据映射到平行的坐标轴上，然而这种技术对所有维度都采用同样的映射过程，在某些情况下不利于数据呈现。如地理信息由于经度纬度共同决定位置这一性质，在使用平行坐标时无法发现地理位置之间的临近关系，故平行坐标技术不适用于这种需求。

在流量结构特征的可视化中，目前使用的技术均为结点链路技术。文献[10]采用Flow-inspector工具中实现的结点链路映射技术来反映流量在不同IP之间的流动情况。该方法突出了链路的特性，对结点信息通过标签等交互式的方式呈现。然而，标签方式在某一时刻只能查看一个结点的信息，不能对结点信息有整体直观的认识，即结构和结点信息无法同时呈现。

综上所述，目前流量可视化技术对于一般情况的可视化需求有着不错的效果；然而对于某些特殊的需求，如流量的周期特性、特殊维度的要求以及结点信息的呈现仍待改进。

2.2.2 流量可视化的新方法

针对流量可视化目前的使用技术难以满足特殊需求的问题，本文提出了网络流量可视化的新技术。第一，在流量的时间特征中，针对目前可视化技术较难发现流量周期特性这一缺点，提出了使用螺旋线技术对流量的时间特征进行可视化。通过使用EISD[11]（enhanced interactive spiral display）工具实现螺旋线技术，可以快速地发现周期特性。此外，在数据量大的情况下，该方法较之于直线映射可以极大地减少可视化视图的空间占用。第二，在维度特征可视化中，针对在平行坐标中无法有效映射特殊维度的特点，采用了对维度进行分类的方法解决该问题。即维度分别对待，不同分类采用不同的映射方式。本文针对地理信息的例子，通过使用JMP（http://www.jmp.com/software/jmp/）工具，对地理信息维度特征进行分类的可视化，地理维度映射到直角坐标轴，其他维度映射到颜色。该方法有效解决了平行坐标对于地理信息特殊维度可视化的局限性。第三，在流量的结构特征中，针对结点链路技术聚焦于链路特性可视化而忽略结点信息的现状，提出了空间填充技术对结点信息进行可视化的方法。本文采用JMP工具中空间填充技术，实现了结点信息的可视化。通过使用这些新技术对流量数据进行可视化，并与目前的可视化技术进行对比，说明了新技术在特定场景下的优势。

3 流量可视化新方法的实现

本文针对4over6项目[5]中的流量数据进行了可视化，来分析和阐述新方法对流量的时间、维度、结构等特征的可视化效果。从2.1节可视化管线模型中可以看出，可视化过程就是对数据进行状态转换的过程。其中数据处理分为数据转换和数据过滤。数据转换是针对源数据的结构化；数据过滤是根据目标过滤无关数据项，与数据的目标紧密联系的，在下文的任务分析中详细介绍。可视化映射技术和绘制渲染过程在2.2.2节中已经确定。故可视化实现面临的主要问题包括数据源的采集，数据转换的实现以及可视化工具使用。此外，对于初步接触的用户来说，使用新技术得到的视图理解起来有一定难度，本文将简要说明视图含义。下面从3个问题出发，介绍可视化的实现过程。

3.1 数据来源

本文使用的数据源包括百所高校流量数据和百所高校地理信息两部分。下面分别介绍两种数据的获得方式和数据源的格式。

流量数据来源于百所高校部署4over6设备实现过渡技术的项目。该项目采用清华大学下一代互联网实验室开发的4over6过渡技术在全国一百所高校进行4over6PE（IPv4 over IPv6 provider edge）设备部署，通过CERNET2（China education and research network II）网络连接互联网，实现IPv6网传输IPv4数据包的过程。目前数据的采集和处理由4over6过渡网管系统实现[12]。它是基于Cacti（流量监控软件）系统平台建立，后台使用SNMP（simple network management protocal）[13]网络管理协议和MySQL数据库，采用RRDTool（round robin database tool）完成图形绘制和处理，并使用Weathermap（Cacti常用绘图插件）绘制拓扑。每隔5 min进行流量数据采集和更新。RRD数据库具有存储空间固定和空间重复使用的特点。随着时间的变化，之前的数据会被覆盖。本文使用该网管系统服务器中存储的全国百所高校的IPv4和IPv6流量数据以及核心结点的流量数据作为源数据。

数据源格式是按学校命名的多个XML（extensible markup language）文件，图2显示了某高校的流量数据。采集到的初始数据分别按照5 min、0.5 h、2 h以及1 d的数据间隔存储于XML文件中。每个学校包含3个文档IPv4.xml、IPv6.xml、USER.xml。

bit_traffic_in.xml

```
<rra><cf> AVERAGE </cf>
    <pdp_per_row> 1 </pdp_per_row> <!-- 300 seconds -->
        <database>
        <!-- 2012-12-04 09:20:00 CST / 1354584000 -->
        <row><v> 4.4535810668e+01 </v><v> 1.1106170395e+01 </v></row> … …
        </database>
</rra>
<rra><cf> AVERAGE </cf>
    <pdp_per_row> 6 </pdp_per_row> <!-- 1800 seconds --> <database>… … </database>
</rra>
<rra><cf> AVERAGE </cf>
    <pdp_per_row> 24 </pdp_per_row> <!-- 7200 seconds --> <database>… … </database>
</rra>
<rra><cf> AVERAGE </cf>
    <pdp_per_row> 288 </pdp_per_row> <!-- 86400 seconds --> <database>… … </database>
</rra>
```

Fig.2 IPv4_traffic data file of some university

图2 某高校的IPv4流量数据文件

为了分析流量数据的时间、维度、结构特征，深入了解流量信息在不同地理位置的分布情况，本文采集了百所高校所在的省份、经纬度和区域信息，如图3所示。其中高校名称缩写隐含于XML文件的文件名中；区域信息包含在网管系统文件分类中；省份和经纬度信息不包含在原始数据中。本文根据百所高校名称，google定位其地理位置，获取了百所高校的经纬度和所在省份的数据，最终把这些数据项存储于school_info.txt中。

school_info.txt

高校名称缩写	省份	经度	纬度	区域
Bit	北京	39.9581577	116.302782	区域1
Bjfu	北京	39.9999341	116.339271	区域1
Bjut	北京	39.8704232	116.473285	区域1
		… …		
Zzu	河南	34.743333	113.636945	区域1

Fig.3 Geographic information file of universities

图3 百所高校地理信息文件

3.2 数据转换和实现

为了结合流量信息和地理信息生成有效的结构化数据以及方便后续处理，本文通过程序把包含百所高校的IPv4（IPv6）流量数据的多个XML文件和高校地理信息表汇聚于同一数据表中，生成包含采集时间、高校名称缩写、省份、经度、纬度、区域、入流量和出流量等数据项组成的4个数据表。4个数据表是

根据数据采集时间间隔划分的，分为time_interval_1（5 min）、time_interval_6（0.5 h）、time_interval_24（2 h）及time_interval_288（1 d）。图4呈现了数据转换后生成的时间间隔为5 min的流量地理信息汇总表。数据转换的过程如图5所示。

time_interval_1.txt

时间	高校名称缩写	省份	经度	纬度	区域	入流量	出流量
2012-12-04 09:20:00	bit	北京	116.302782	39.9581577	区域1	4.4535810668e+01	1.1106170395e+01
2012-12-04 09:25:00	bit	北京	116.302782	39.9581577	区域1	5.2539600443e+01	2.0010851827e+01
			… …				
2012-12-06 11:15:00	zzu	河南	113.636945	34.743333	区域1	5.1804694166e+00	3.6955366481e+01

Fig.4 File of traffic and geographic information

图4 流量和地理信息汇总表

数据转换过程使用Visual studio工具C#语言编程实现。该代码可以实现任意多个流量数据文件的数据转化过程。数据转换过程如下：

⑴ 建立高校类，包括高校名称缩写、省份、经度、维度区域5项属性。读取高校地理数据，并记录在高校类数组中。

⑵ 建立4个文件result1、result6、result24、result288，用于存储不同时间间隔的流量数据。

⑶ 搜索源文件夹，若有文件，读取一个文件，取文件名中下划线之前数据为高校缩写Name_u，并打开文件，继续步骤⑷；若无文件则跳转到⑹。

⑷ 搜索下一个< >或者> <中内容并写入text，若text为AVERAGE（数据块开头标示），则跳转到⑷，搜索每行数据；若text为pdp_per_row，即本数据块有未读取的行数，则继续下一步来读取数据；若text为空，即本文件内数据块读取完，则根据Name_u搜索高校类数组，把高校缩写相应的高校地理数据写入result1（result6/result24/result288）对应的属性中，并跳转到⑶，读取下一个文件。

⑸ 读取pdp_per_row后数据写入i（表示数据采集时间），根据时间间隔i选择需要写的result(i)文件。读取database与/database之间时间间隔、入流量、出流量数据并写入result(i)中，跳转到⑷。

⑹ 保存result1、result6、result24、result288文件并退出。

3.3 可视化工具实现

本文使用了EISD和JMP两种可视化工具，其中EISD是针对流量时间特征的可视化，而JMP是针对流量维度和结构特征的可视化。下面简要介绍两种工具实现可视化映射和渲染的过程以及所产生的视图含义。

3.3.1 EISD工具

EISD工具是Tominski[11]结合了螺旋线和双色编码[14]的优点，使用Java语言开发的交互式螺旋线绘图工具。该工具提供了聚焦和标签（focus and context）、

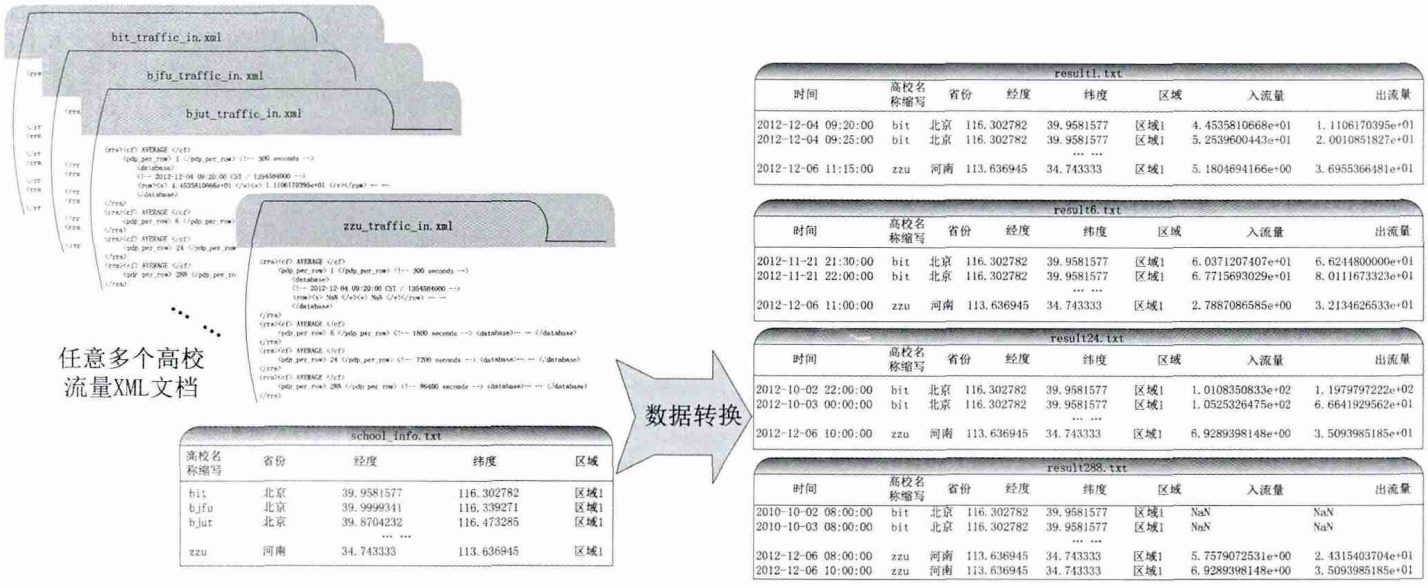


Fig.5 Process of data conversion

图5 数据转换过程

全局和细节观察(zoom and pan)等交互式功能。通过调整每圈数据项的数量,可以探索数据中的周期特征。其中双色编码是螺旋线中精确读取时间轴数值的关键技术,有必要介绍双色编码技术的基本原理。

在颜色编码中,传统的编码方式有离散编码、连续编码。图6针对最上方的曲线图,分别使用离散(DISCRETE)、连续(CONTINUOUS)以及双色编码(TWO-TONE)进行可视化。从图6中可知,离散编码由于对同一区间内数据采样同种颜色编码,得到的数据值不够精确;连续编码虽然精确地对数据值进行编码,然而相邻数据值采用相近的颜色编码来反映,很难从颜色中精确得到数据值的大小;双色编码使用某种颜色反映特定数据值的大小。对于其他的数据值采用两种相邻数据值对应的颜色进行编码,实现了编码的精确性以及视觉上的可读性。其原理是首先从两种颜色读取数据值的取值范围,再通过两种颜色的比例分析具体值的大小。例如对于图6下图红色框中数据,首先由绿色和蓝色两种颜色组成数据值,可以读出数据值的范围在30~40之间。其次根据蓝色与绿色的比例1/3×10得出偏差值3,因此该处双色编码的值为33。

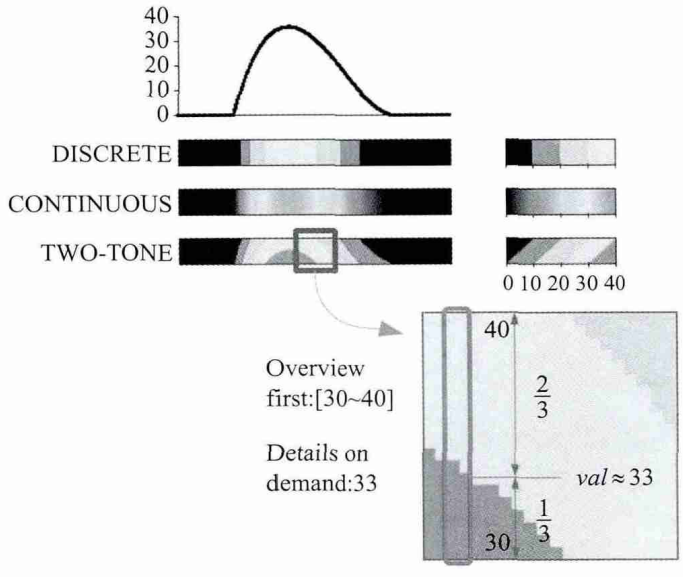


Fig.6 Conventional and two-tone pseudo coloring

图6 传统编码和双色编码

图7是EISD对4over6核心结点IPv4流量数据时间特征可视化得到的视图(刻度盘是为方便说明添加的而非EISD生成)。通过在交互面板中设置每环分段数来发现数据的周期特性。图中时间轴方向为从中心到边缘;双色编码采用温度7色和线性的编码方式,即从蓝色到红色流量数据按线性增长;采集间隔为5 min,故288份×5 min为一天。颜色盘(图右)表示了不同颜色代表的数据值大小。流量大小通过5 min内数据包的多少来描述。从图7可知,流量数据具有明显以天为周期的规律。早晨8点到12点为流量高峰期;12点到14点流量明显下降;从22点到次日8点夜间流量非常小。且流量数据在短期内没有明显随时间变化的趋势。此外,可以清晰地发现流量数据在最后两个时间段为蓝色(即数据为零),说明该技术还可以用来识别流量数据异常。

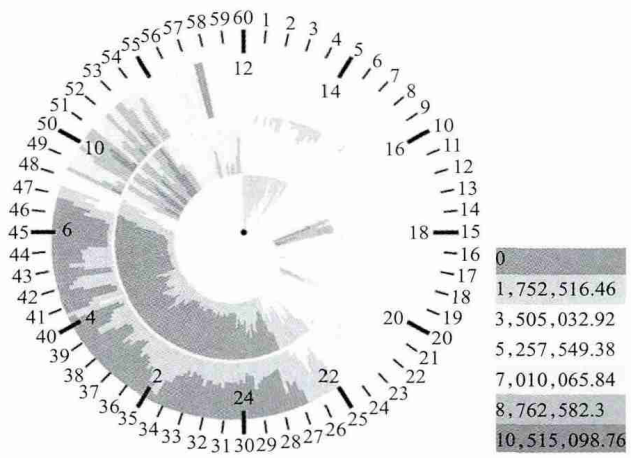


Fig.7 IPv4 traffic visualization of 4over6 core node by EISD tool

图7 EISD工具对4over6核心结点IPv4流量可视化

3.3.2 JMP工具

JMP11(http://www.jmp.com/software/jmp/)是SAS(statistical analysis system)最新推出的交互式可视化统计发现软件。该工具支持交互式地进行数据分析、挖掘和可视化等功能。此外,JMP内置了多个国家的地理地图、轮廓图、街区地图,这对于分析流量数据的地理分布非常有效。JMP支持csv、txt、xlsx等多种数据输入格式。JMP工具操作简单,功能强大,本文维度特征和结构特征的新方法均采用该工具实现。

图8是采用JMP工具实现维度分类的例子。该图是JMP对50所高校流量数据的维度特征进行可视化的视图。其中地理信息使用直角坐标系映射(横

轴表示纬度，纵轴表示经度）；各省流量总和使用颜色的深浅表示，颜色越深表示流量越大。白色区域表示4over6项目没有覆盖到的省份。从图8中可以一目了然地观察各省份4over6流量的分布情况。其中山东省产生的流量总和最大（颜色最深）。这是维度分类对特殊维度（地理信息）可视化的例子。

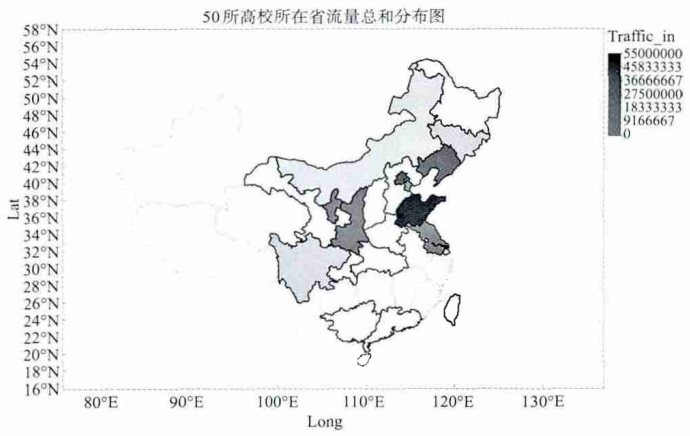


Fig.8 Provincial distribution map of 50 universities traffic

图8 50所高校流量的省分布图

4 互联网流量可视化实验

为了分析新方法优点，本文针对4over6流量数据进行可视化，通过与目前可视化方法进行对比，分析了新方法在时间、维度、结构3方面各自的优势。从时间特征出发，使用螺旋线映射技术对流量时间数据可视化，说明螺旋线映射技术在周期发现中的优势；从维度特征出发，采用维度分类方法实现地理信息和流量的分类可视化，说明维度分类解决特殊维度的需求的优势；从结构特征出发，使用空间填充技术对流量结构特征进行可视化，说明空间填充在简单结构下对结点信息呈现的优势。

4.1 流量时间特征可视化对比

目前在4over6项目中，采用了RRDtool工具实现流量数据时间特征的直线映射技术。本文通过使用EISD工具实现螺旋线映射技术，并与RRDtool工具中直线映射技术进行对比，说明螺旋可视化技术在发现周期方面的优势。

在此采用4over6核心结点IPv4流量按照0.5 h时间间隔得到数据进行可视化实验。在时间特征中，仅关心时间、入流量和出流量，故针对3.2节中数据转换后的数据result6进行数据过滤，得到仅包含这3种属性的数据。RRDtool与EISD对数据可视化的结果分别如图9与图10所示。

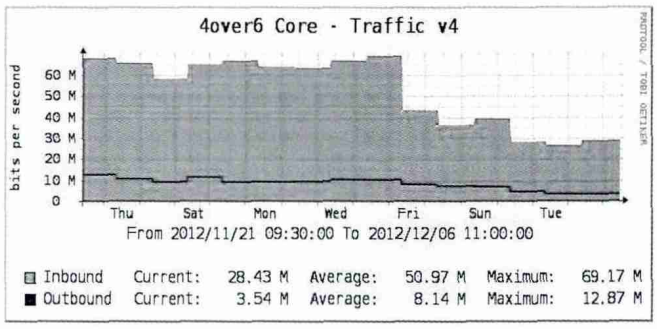
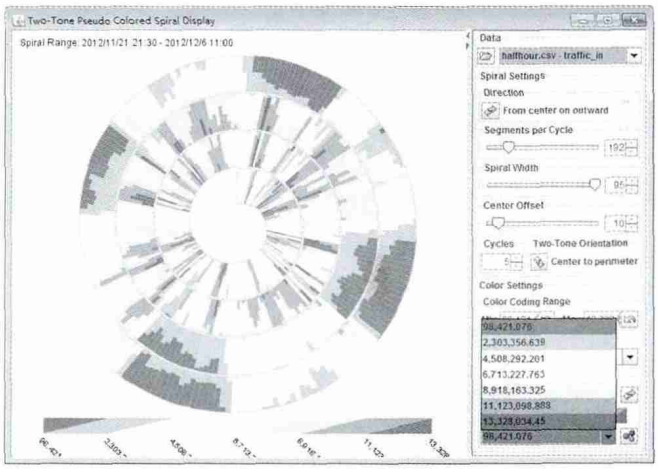
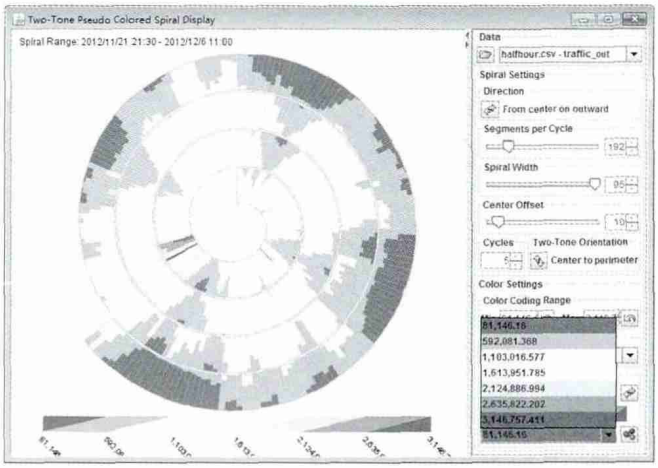


Fig.9 Visualization of traffic time feature (straight line mapping)

图9 流量时间特征可视化（直线映射）



(a)Inbound



(b)Outbound

Fig.10 Visualization of traffic time feature (spiral curve mapping)

图10 流量时间特征可视化（螺旋线映射）

图9中横轴表示时间,纵轴表示流量数据;入流量采用柱状图表示,出流量采用曲线表示。图10中,时间轴沿着螺旋线向外延伸,双色编码表示流量大小;图(a)为入流量,图(b)为出流量。设置每环分段数为192(每份为0.5 h,192份为4 d),颜色编码采用温度7色,模式为线性。

通过对比图9与图10可知,双色编码螺旋可视化映射能够方便研究人员更加容易地发现时间周期特性。而传统直线映射的流量可视化方法需要观察者花费更多的时间和精力去发现数据规律。此外,对于特定时间点的流量值,RRDTool可以方便地读出数据的范围。对于EISD用户也可以通过颜色快速地读取数据范围甚至数据值大小。

然而由于直线映射使用非常广泛,而螺旋线映射只有极少应用,螺旋线映射技术的变化趋势对于首次接触观察者可能会不够直观,用户需要一定的了解和适应才能达到快速理解的效果。在实际使用过程中,可以结合两种可视化映射的方式,从数据值和周期特性等方面对时间特征数据进行挖掘。

4.2 流量维度特征可视化对比

目前4over6网管系统中未对维度特征进行可视化。为了与目前广泛使用的平行坐标技术进行对比,来说明维度特征分类方法的优势,本文实现了流量数据的维度特征的两种可视化方法。通过两者的比较,说明维度分类可视化技术对特殊维度的优势。

经过3.2节中的数据转换,流量数据维度信息包括采集时间、高校名称缩写、省份、经度、纬度、区域、入流量和出流量这些数据项。为了简化实验,突出维度特征优势,本文针对result6中数据进行过滤,得到经度、维度、区域、省份、入流量等维度信息。其中区域信息是4over6网管系统把百所高校分为4个区,而得到高校区域信息。为了方便观察,按省份对入流量累加,采用平行坐标技术和维度分类方法的可视化分别如图11和图12所示。

在图11中,每个维度映射到一条坐标轴,可视化效果与平行坐标中坐标轴的顺序有关系,且在平行坐标技术中,经度(long)、纬度(lat)作为两个独立的坐标轴不利于发现数据的地理信息。在图12中,按照经度、纬度、区域、省份等维度信息确定数据的坐标位置,区域信息进行视图分割,各省份的入流量使用颜色进行编码。

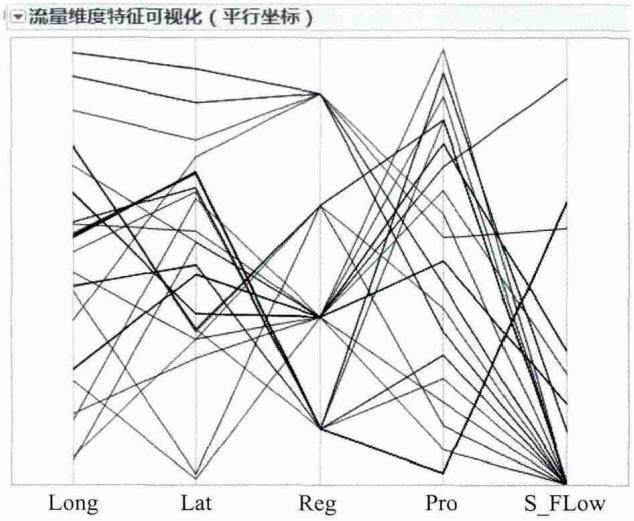


Fig.11 Visualization of traffic dimension feature (parallel coordinates)

图11 流量维度特征可视化(平行坐标)

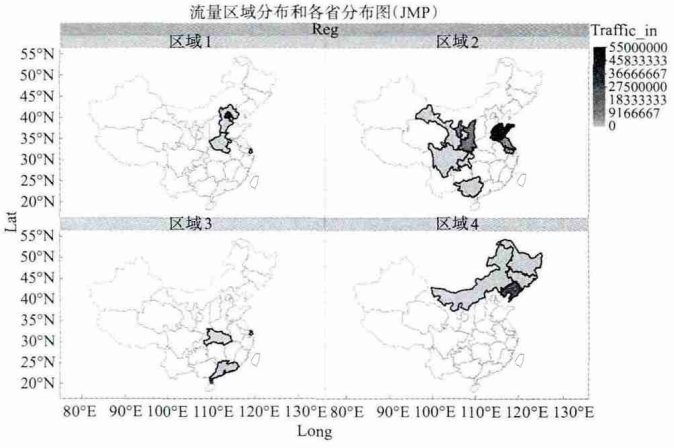


Fig.12 Visualization of traffic dimension feature (dimension classification)

图12 流量维度特征可视化(维度分类)

通过对比图11与图12可知,维度分类的方法更方便用户发现流量数据在各省份和各区域的分布情况。通过结合颜色编码和直角坐标系映射的方式克服了平行坐标的不足。维度分类的方法能够针对每个维度各自的物理意义和值类型选择合适的可视化映射方法,得到每种维度信息可视化的最大化。而平行坐标技术不论维度数据是何种类型的值数据

(可分组、可比较或者可量化) 都采用坐标轴映射方式，而且具有可视化效果与各条坐标轴之间的顺序有关联的弊端。

在使用过程中，如果对数据的维度映射没有特别的可视化需求，平行坐标有着维度数量任意多的优点，因此是可视化高维度数据的首选。对于有着特殊需求和特定物理意义的维度，则采用维度分类的方法，针对不同维度采用不同的可视化映射技术达到更佳的可视化效果。

4.3 流量结构特征可视化对比

在4over6项目中，目前使用Weathermap绘图插件对流量的拓扑结构进行可视化，该插件采用了结点链路技术。本文通过使用JMP工具，并采用空间链路技术对拓扑(结构)信息进行可视化，说明空间填充技术在结点信息呈现方面的优势。

4over6过渡项目把百所高校分为4个区域，Weathermap实现了对4个区域的流量结构特征的可视化。由于Weathermap仅显示当前的流量分布，故没有历史数据的分布图进行对比实验。通过两种可视化技术对不同时刻数据的可视化呈现进行对比，分析两种方法的特点。首先通过数据过滤去除无关信息，得到高校名称缩写、区域入流量数据项。Weathermap和JMP两种工具可视化实现的结果分别如图13和图14所示。

其中Weathermap实现了4over6中100所高校流量分布情况的可视化。图中中间结点表示连接到CERNET2网络的区域路由器，颜色表示高校4over6设备与区域路由器间流量的流动情况。该图形可以方便用户观察系统的整体运行情况，通过颜色识别及时发现数据中的异常。JMP针对4over6中50所高

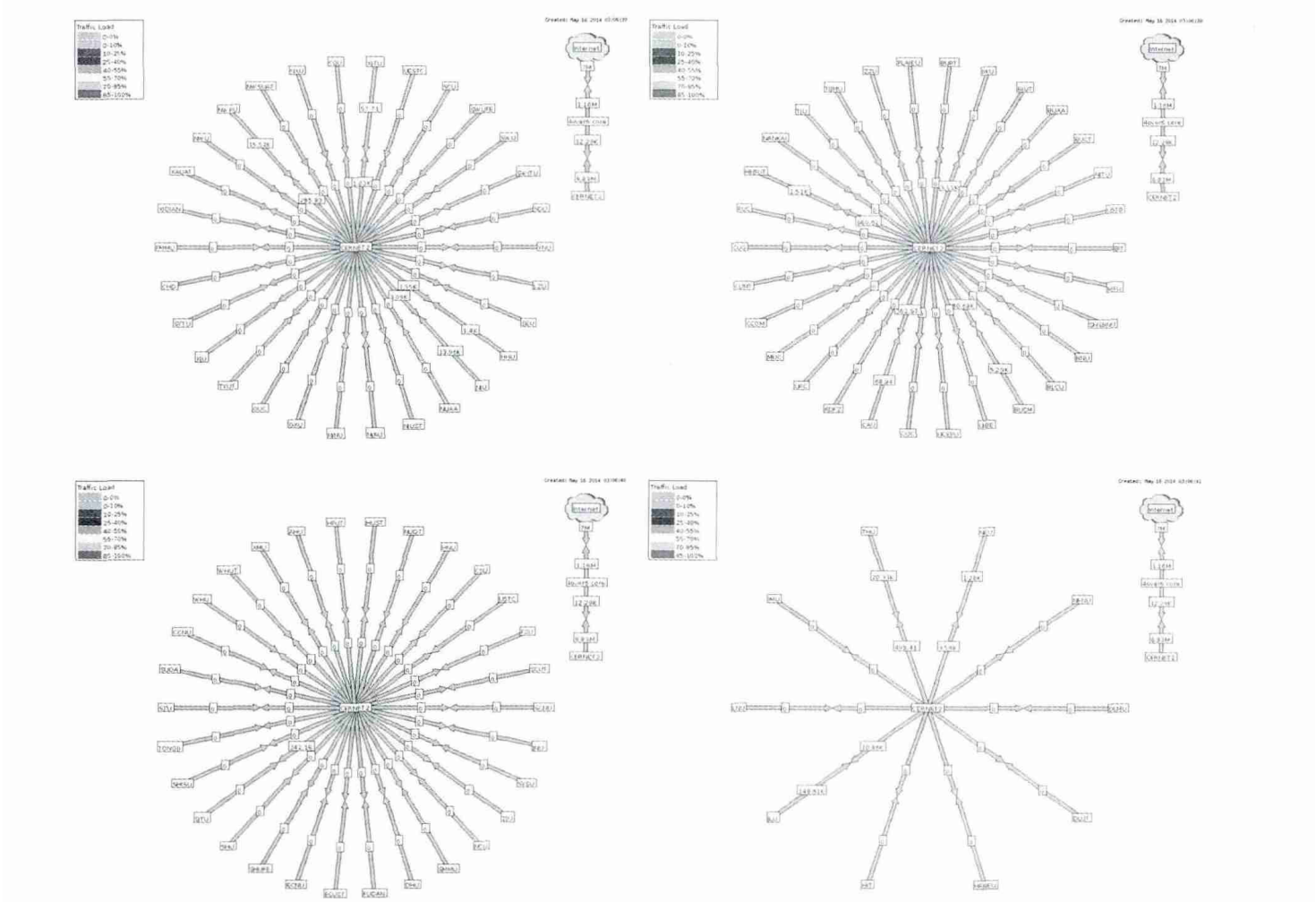


Fig.13 Visualization of traffic structure feature (node-link mapping)

图13 流量结构特征可视化(结点链路映射)

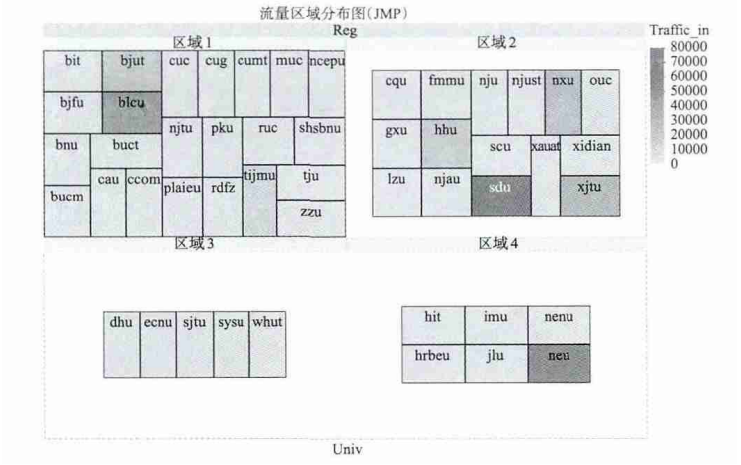


Fig.14 Visualization of traffic structure feature (space-filling mapping)

图14 流量结构特征可视化(空间填充映射)

校流量分布情况的可视化(由于百所高校信息采集时间不一致,本文通过过滤,得到50所采集时间相近的高校,而Weathermap是实时绘制无法修改)。JMP使用空间填充技术对各个学校的流量数据分区域进行可视化。其中流量的大小使用图中高校对应的矩形框的颜色来表示。该方法同样可以发现高校整体运行情况和数据的异常。

虽然两种方法针对的数据集有所不同,但是可以对两者进行定性的比较。对比图13与图14可以发现,JMP能够在实现Weathermap对于流量数据分布和异常发现功能的同时,更加直观地呈现结点的信息,包括高校名称和流量大小。作为补偿,空间填充技术中结构信息隐含在结点矩形框排列方式中。而结点链路技术使用大部分空间来呈现结构信息,故对于结点信息没有呈现。在4over6的拓扑可视化中,结构信息非常简单,因而通过空间填充技术能在兼顾结构信息的同时,更好地呈现高校名称和流量大小。

空间填充技术由于结构信息隐含在矩形框中,且日常使用较少,需要熟悉才能更好地理解。在使用过程中,当结构很复杂(图结构)时,选择结点链路技术来呈现结构信息;当结构比较简单(树结构)时,可以采用空间填充技术来呈现结构信息和结点信息,这样可利用树结构特征因简单而浪费的空间,呈现更多的结点信息。

5 结束语

本文针对流量数据的时间、维度及结构特征可视化在涉及到特殊可视化、传统可视化技术不能达到特殊可视化目的的情况下,提出了3种新的可视化方法,即螺旋线可视化技术、维度分类方法和空间填充技术。通过与广泛使用的直线映射、平行坐标和结点链路技术进行比较,说明了新方法在特殊情况下的优势。本文基于4over6项目流量数据进行可视化实现,通过对比得出:

(1) 在流量时间特征中,螺旋线技术对流量周期规律的发现有着显著的优势。

(2) 在流量维度特征中,维度分类方法对于特殊维度信息可视化较平行坐标技术有着更好的可读性。

(3) 在流量结构特征中,空间填充技术对于结构简单的树类型可以在显示结构特征的同时,呈现结点的更多信息。

References:

[1] Card S K, Mackinlay J D, Shneiderman B. Readings in information visualization: using vision to think[M]. San Francisco, USA: Morgan Kaufmann Publishers, 1999.

[2] Kriglstein S, Pohl M, Smuc M. Pep up your time machine: recommendations for the design of information visualizations of time-dependent data[M]//Handbook of Human Centric Visualization. New York: Springer, 2014: 203-225.

[3] de Oliveira M C F, Levkowitz H. From visual data exploration to visual data mining: a survey[J]. IEEE Transactions on Visualization and Computer Graphics, 2003, 9(3): 378-394.

[4] Cui Weiwei, Qu Huamin. PhD qualifying exam (PQE) report: a survey on graph visualization[R]. Kowloon, Hong Kong: Computer Science Department, Hong Kong University of Science and Technology, 2007.

[5] Wu Jianping, Li Xing, Cui Yong, et al. 4over6: IPv4 network interconnection over IPv6 backbone without explicit tunneling[J]. Chinese Journal of Electronics, 2006, 34(3): 454-458.

[6] Haber R B, McNabb D A. Visualization idioms: a conceptual model for scientific visualization systems[J]. Visualization in Scientific Computing, 1990, 74: 93.

[7] Stevens S S. On the theory of scales of measurement[J]. Sci-

ence, 1946, 103(2684): 677-680.

[8] So-In, a survey of network traffic monitoring and analysis tools[EB/OL]. [2014-03-20]. http://www.cs.wustl.edu/~jain/cse567-06/ftp/net_traffic_monitors3.pdf.

[9] Yin Xiaoxin, Yurcik W, Treaster M, et al. VisFlowConnect: netflow visualizations of link relationships for security situational awareness[C]//Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security, Washington, USA, Oct 29, 2004. New York, NY, USA: ACM, 2004: 26-34.

[10] Braun L, Volke M, Schlamp J, et al. Flow-inspector: a framework for visualizing network flow data using current Web technologies[J]. Computing, 2014, 96(1): 15-26.

[11] Tominski C, Schumann H. Enhanced interactive spiral display[C]//Proceedings of the Annual SIGRAD Conference, Special Theme: Interactivity, Stockholm, Sweden, 2008: 53-56.

[12] Wang Hao. Research and implementation of 4over6 transition network management system for next generation Internet[D]. Beijing: Computer Science and Technology, Tsinghua University, 2012.

[13] Stallings W. SNMP, SNMPv2, and CMIP: the practical guide to network management[M]. Boston, MA, USA: Addison-Wesley Longman Publishing Co, Inc, 1993.

[14] Saito T, Miyamura H N, Yamamoto M, et al. Two-tone pseudo coloring: compact visualization for one-dimensional data[C]//Proceedings of the 2005 IEEE Symposium on Information Visualization. Piscataway, NJ, USA: IEEE, 2005: 173-180.

附中文参考文献：

[5] **吴建平**, **李星**, **崔勇**, **等**. 4over6: **基于非显式隧道的** IPv4 **跨越** IPv6 **互联机制**[J]. **电子学报**, 2006, 34(3): 454-458.

GAO Pihong was born in 1990. She is a master candidate at Department of Computer Science and Technology, Tsinghua University. Her research interests include Internet measurement and visualization, etc.

高丕红（1990—），女，山西吕梁人，清华大学计算机科学与技术系硕士研究生，主要研究领域为网络测量，可视化等。

XU Mingwei was born in 1971. He received the Ph.D. degree from Department of Computer Science and Technology, Tsinghua University in 1998. Now he is a professor and Ph.D. supervisor at Tsinghua University, and the senior member of CCF. His research interests include future Internet architecture, Internet routing, high-performance router, internet measurement and visualization, etc.

徐明伟（1971—），男，辽宁朝阳人，1998 年于清华大学计算机科学与技术系获得博士学位，现为清华大学教授、博士生导师，CCF 高级会员，主要研究领域为未来互联网体系，互联网路由，高性能路由器，网络测量，可视化等。