大数据系列丛书

大数据挖掘及应用

王国胤 刘群 于洪 曾宪华 编著

清华大学出版社

大数据系列丛书

大数据挖掘及应用

王国胤　刘群　于洪　曾宪华　编著

清华大学出版社

北京

内 容 简 介

本书围绕大数据背景下的数据挖掘及应用问题，从大数据挖掘的基本概念入手，由浅入深、循序渐进地介绍了大数据挖掘分析过程中的数据准备和预处理方法、数据可视化技术、数据挖掘理论和经典算法、常用大数据分析计算平台的编程模型、并行化程序设计技术、统计分析 R 语言基础等内容。其中数据挖掘理论和经典算法不仅覆盖了传统的关联分析、分类和聚类，还包括深度学习理论等数据挖掘研究和发展的潮流主题。每一章内容都尽量从不同角度进行深入浅出的剖析，还配以丰富的习题和参考文献，对于读者掌握大数据挖掘及应用领域的基本知识和进一步研究都具有参考价值。本书可以作为高校本科相关专业数据分析类课程教材和面向各专业的数据科学通识教材，也可供广大 IT 从业人员参考。

图书在版编目(CIP)数据

大数据挖掘及应用/王国胤等编著. —北京：清华大学出版社，2017
(大数据系列丛书)
ISBN 978-7-302-46927-8

Ⅰ. ①大…　Ⅱ. ①王…　Ⅲ. ①数据采集—研究　Ⅳ. ①TP274

中国版本图书馆 CIP 数据核字(2017)第 074345 号

责任编辑：张　玥　战晓雷
封面设计：常雪影
责任校对：焦丽丽
责任印制：杨　艳

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　**邮　　编**：100084
社 总 机：010-62770175　　**邮　　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课 件 下 载：http://www.tup.com.cn，010-62795954
印 刷 者：北京富博印刷有限公司
装 订 者：北京市密云县京文制本装订厂
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　张**：26　　**彩　插**：8　　**字　　数**：624 千字
版　　次：2017 年 7 月第 1 版　　**印　　次**：2017 年 7 月第 1 次印刷
印　　数：1～2000
定　　价：59.50 元

产品编号：074001-01

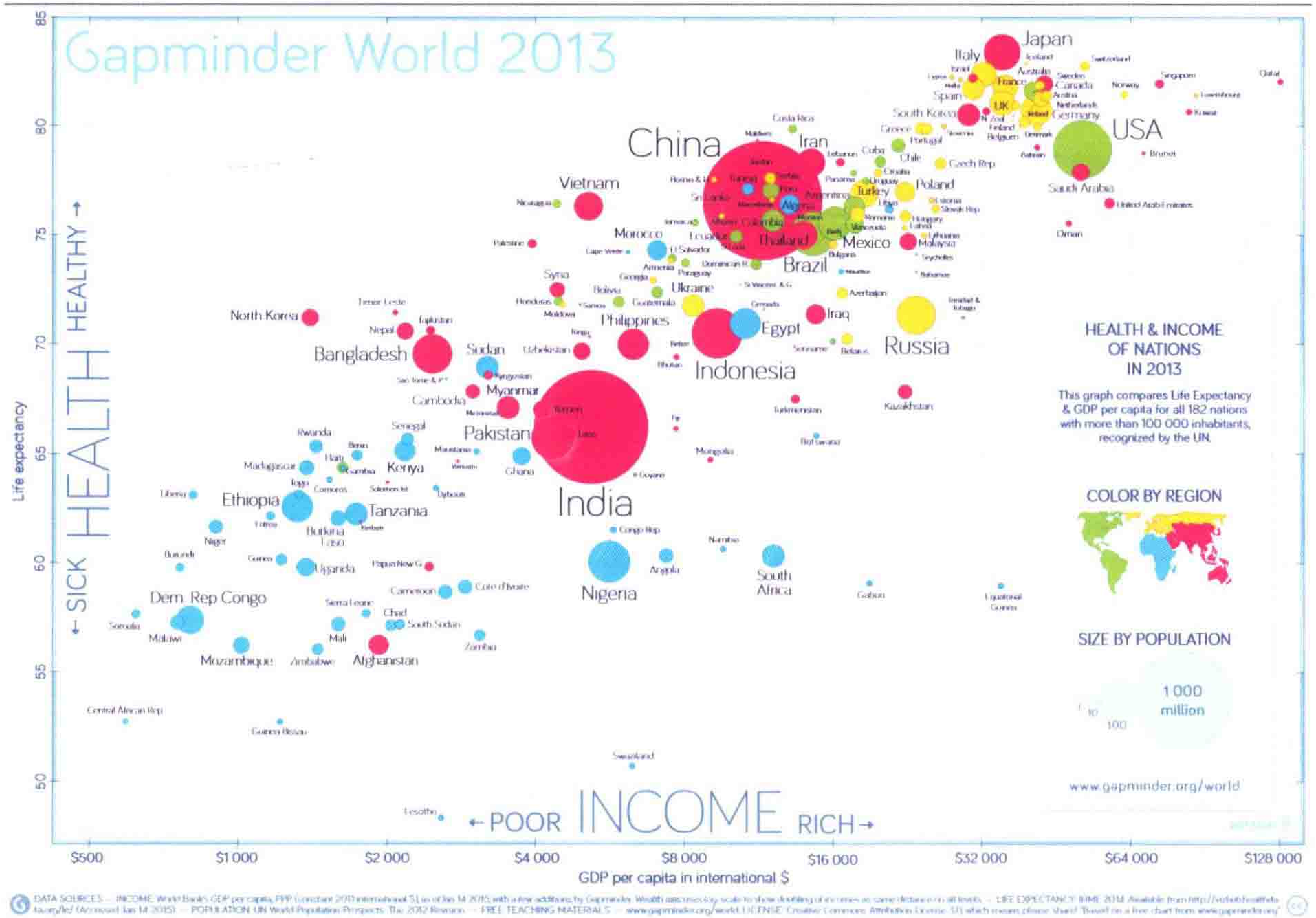

图 3-6　2013 年世界各国预期寿命与人均国内生产总值的散点图可视化

图片来源：http：//www.gapminder.org

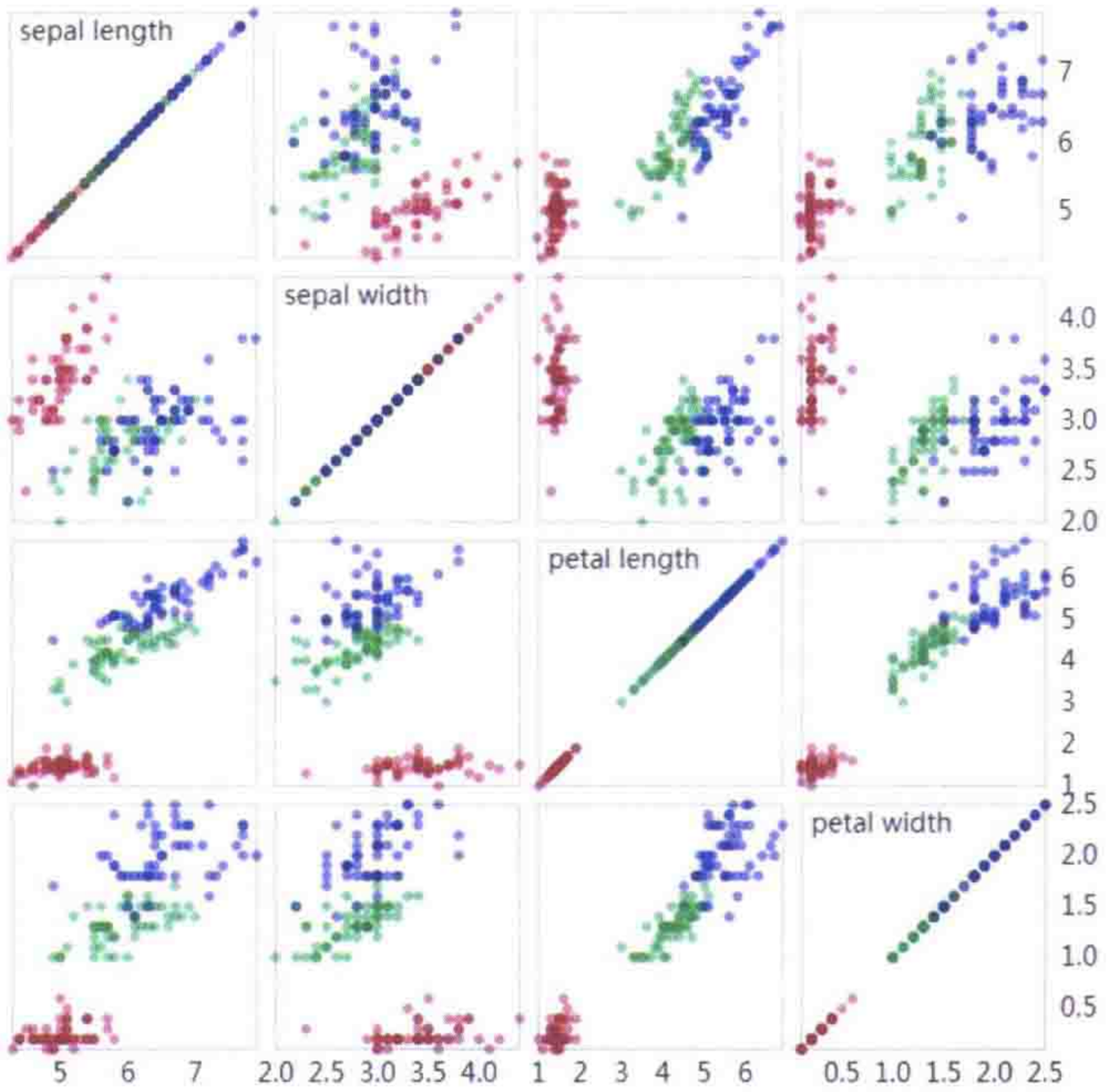

图 3-7　鸢尾花散点图矩阵

图片来源：https://github.com/d3/d3/wiki/Gallery

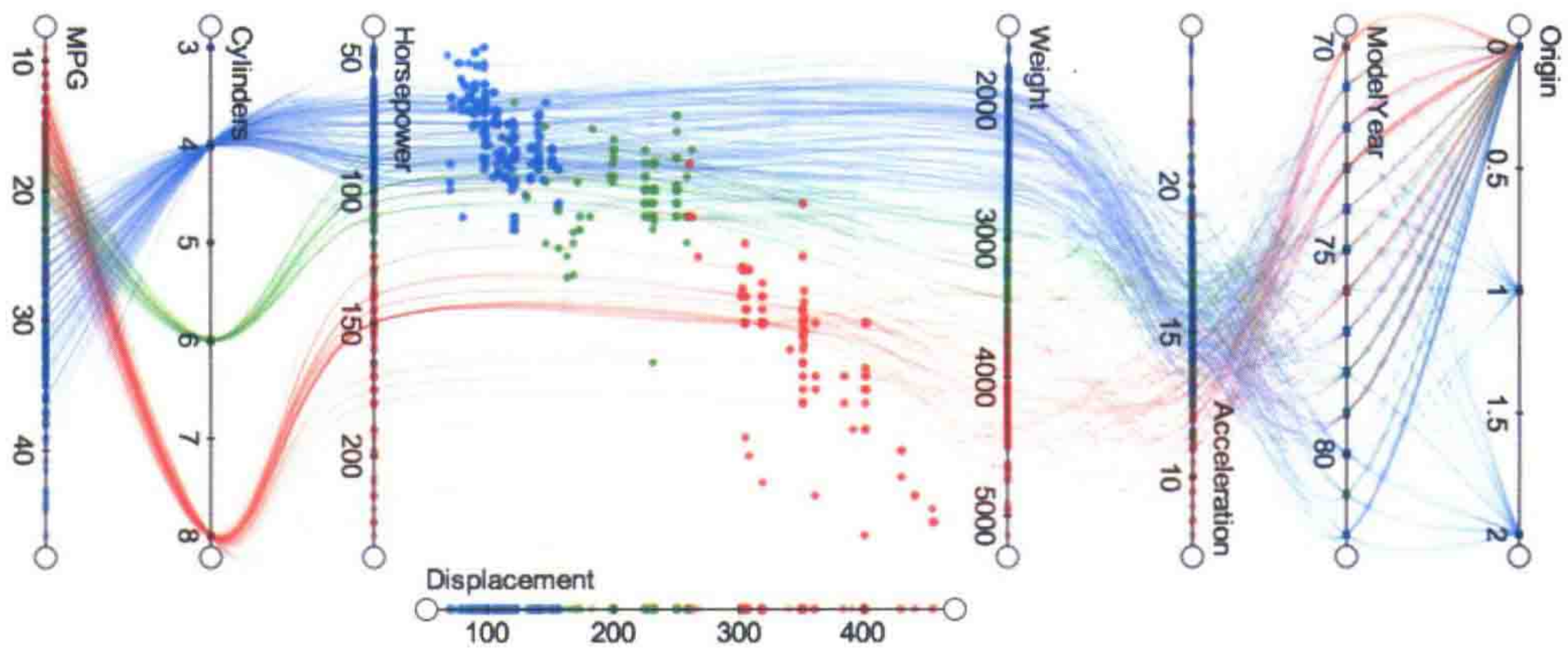

图 3-9　平行坐标结合散点图

图片来源：http://vis.pku.edu.cn/mddv/val/gallery

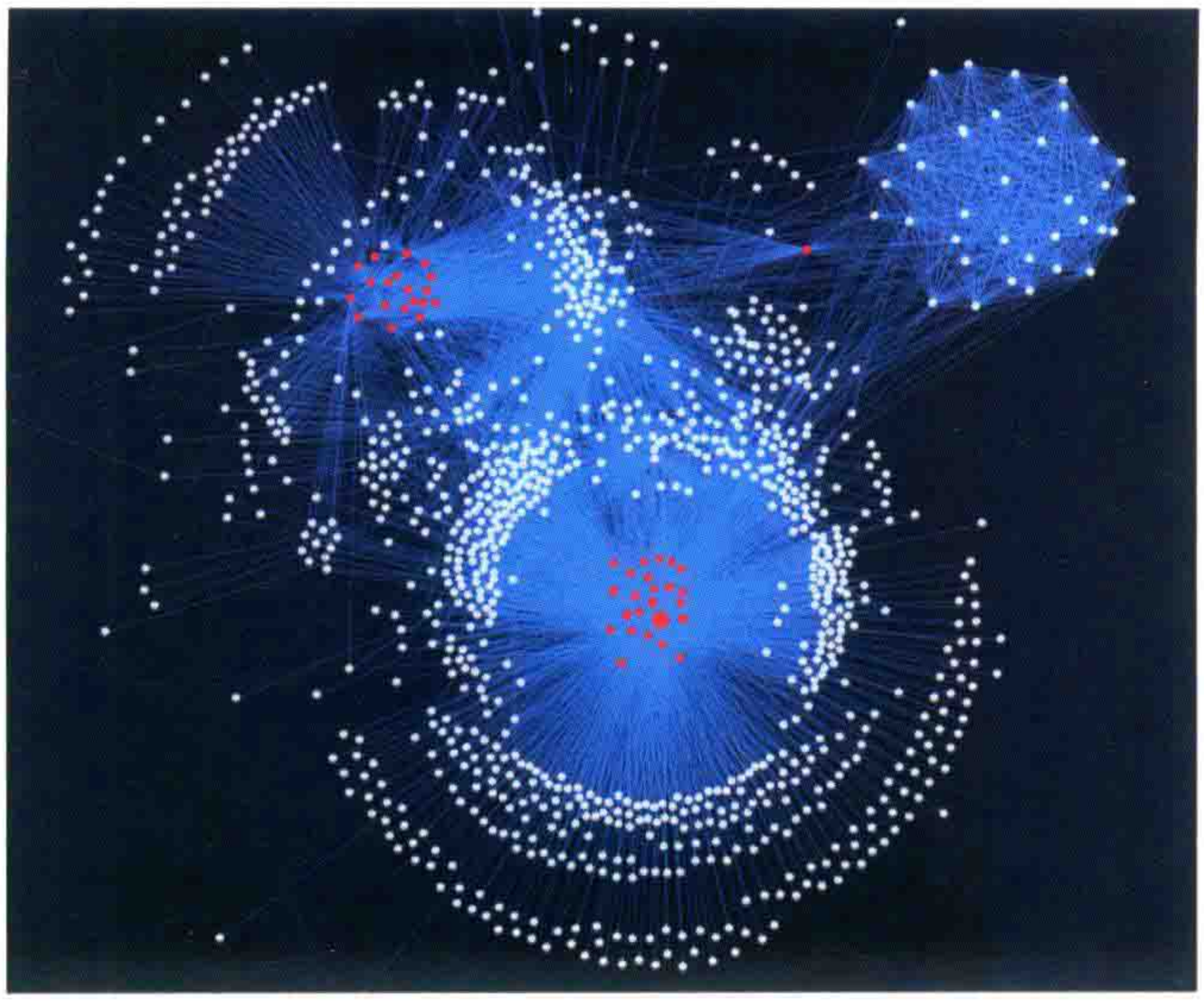

图 3-20　内网主机与服务器之间通信关系的力导向图

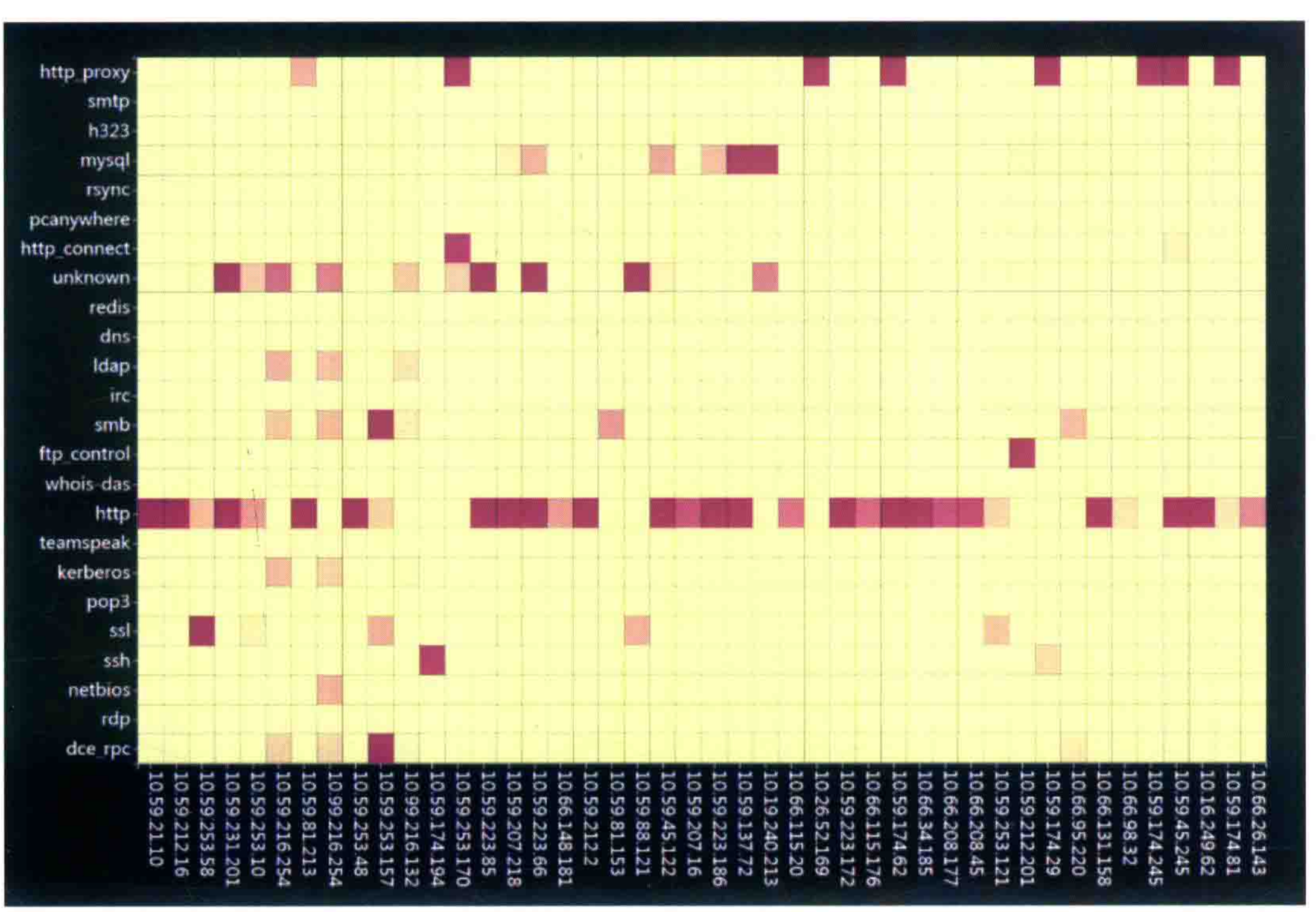

图 3-22　服务器类型的热点图

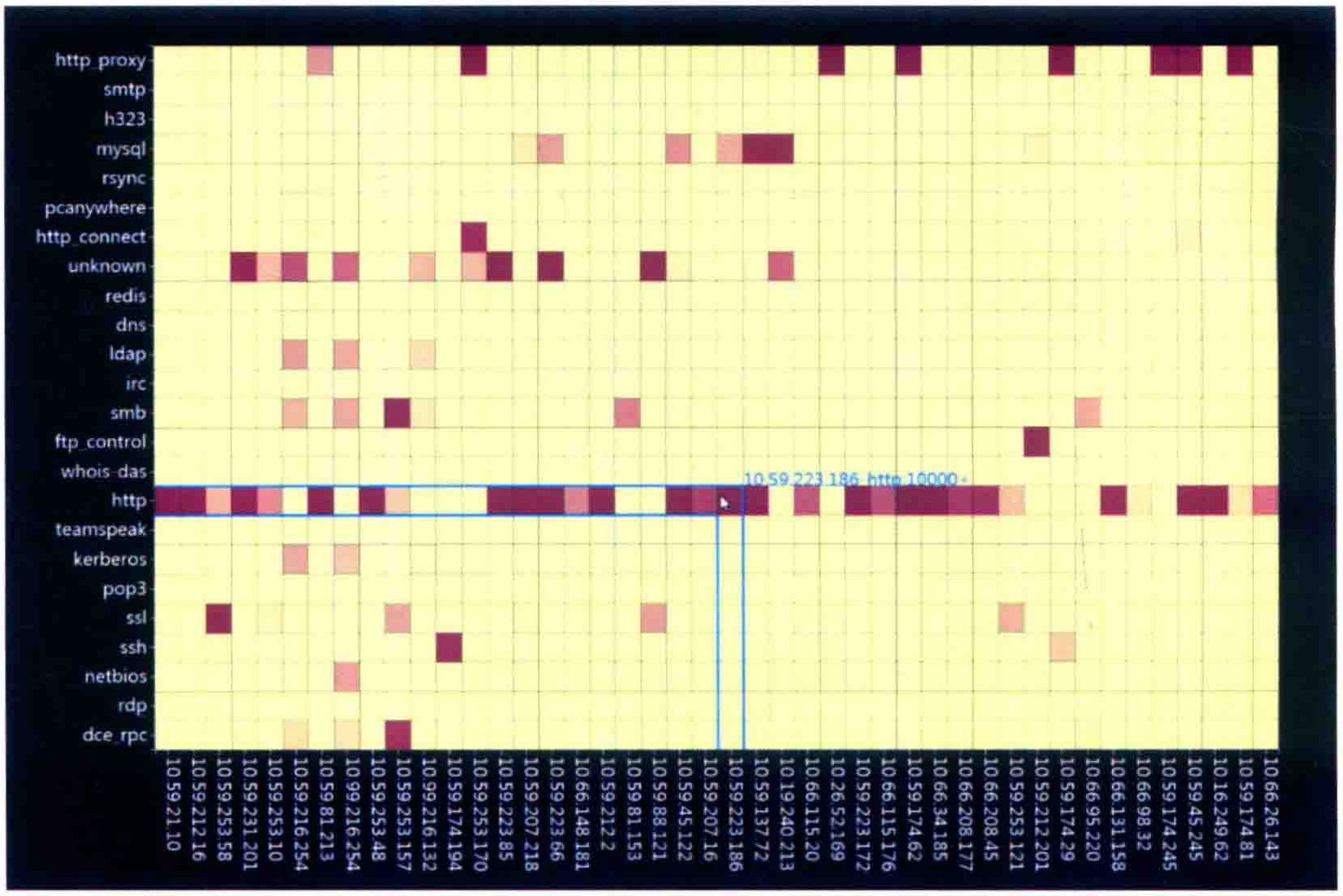

图 3-23　服务器与客户端通信的热点图

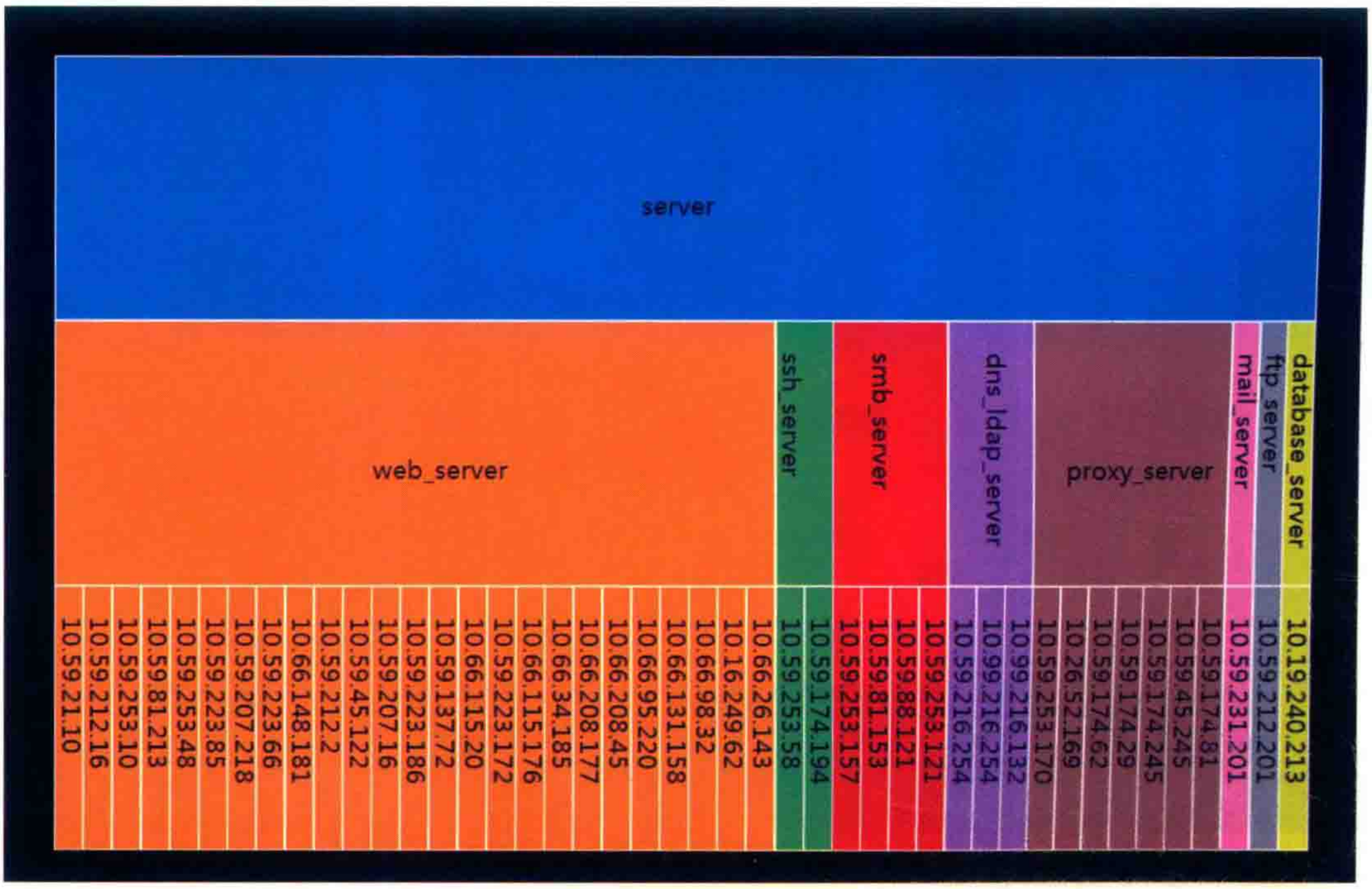

图 3-24　服务器分类结果的树图

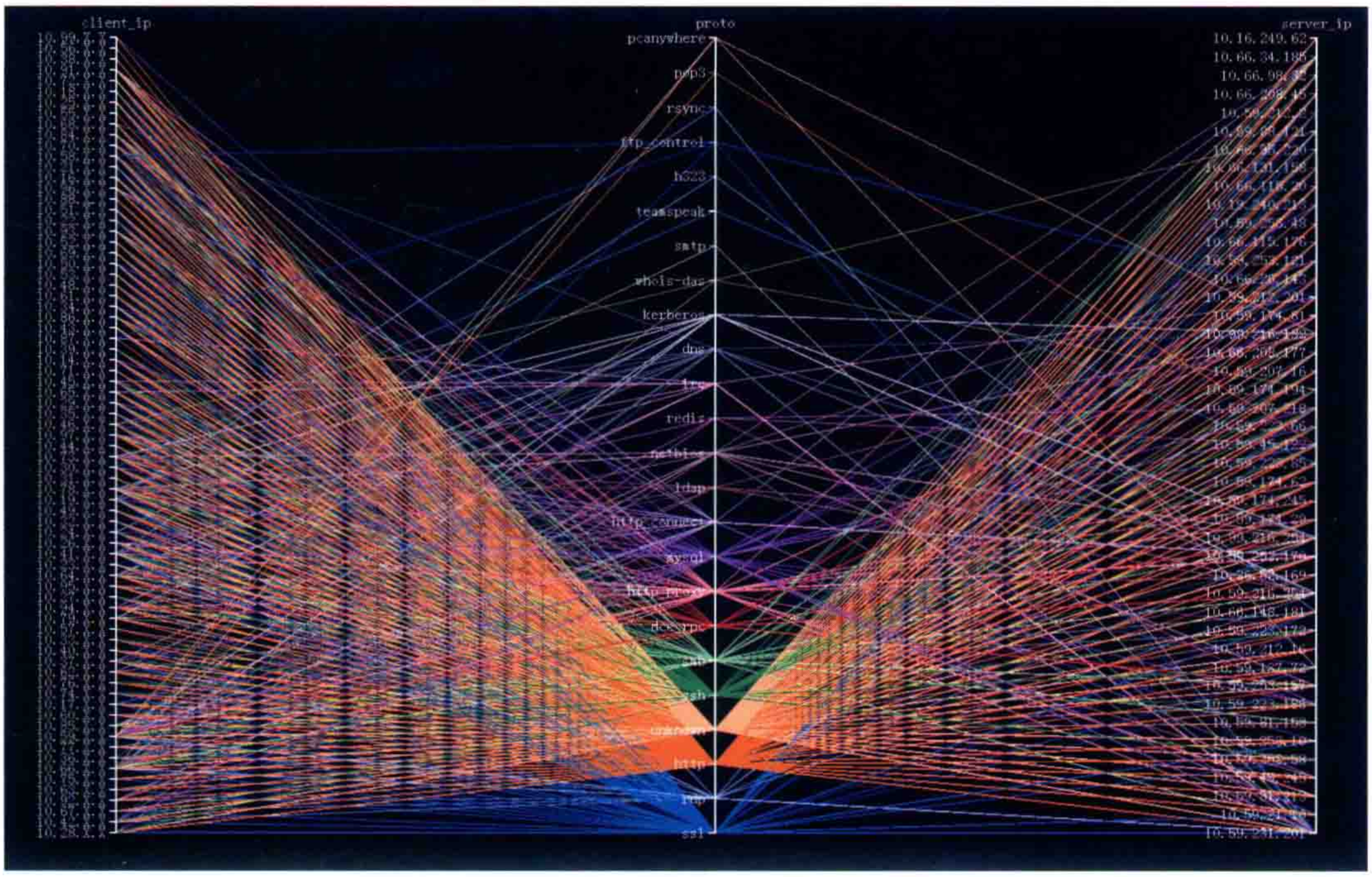

图 3-25　内部通信的平行坐标系

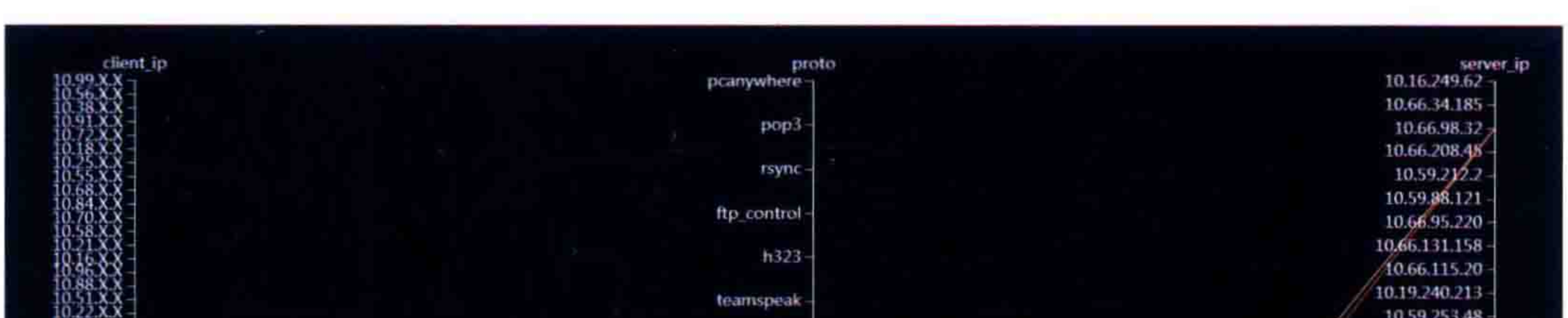

图 3-26　内部通信的平行坐标系

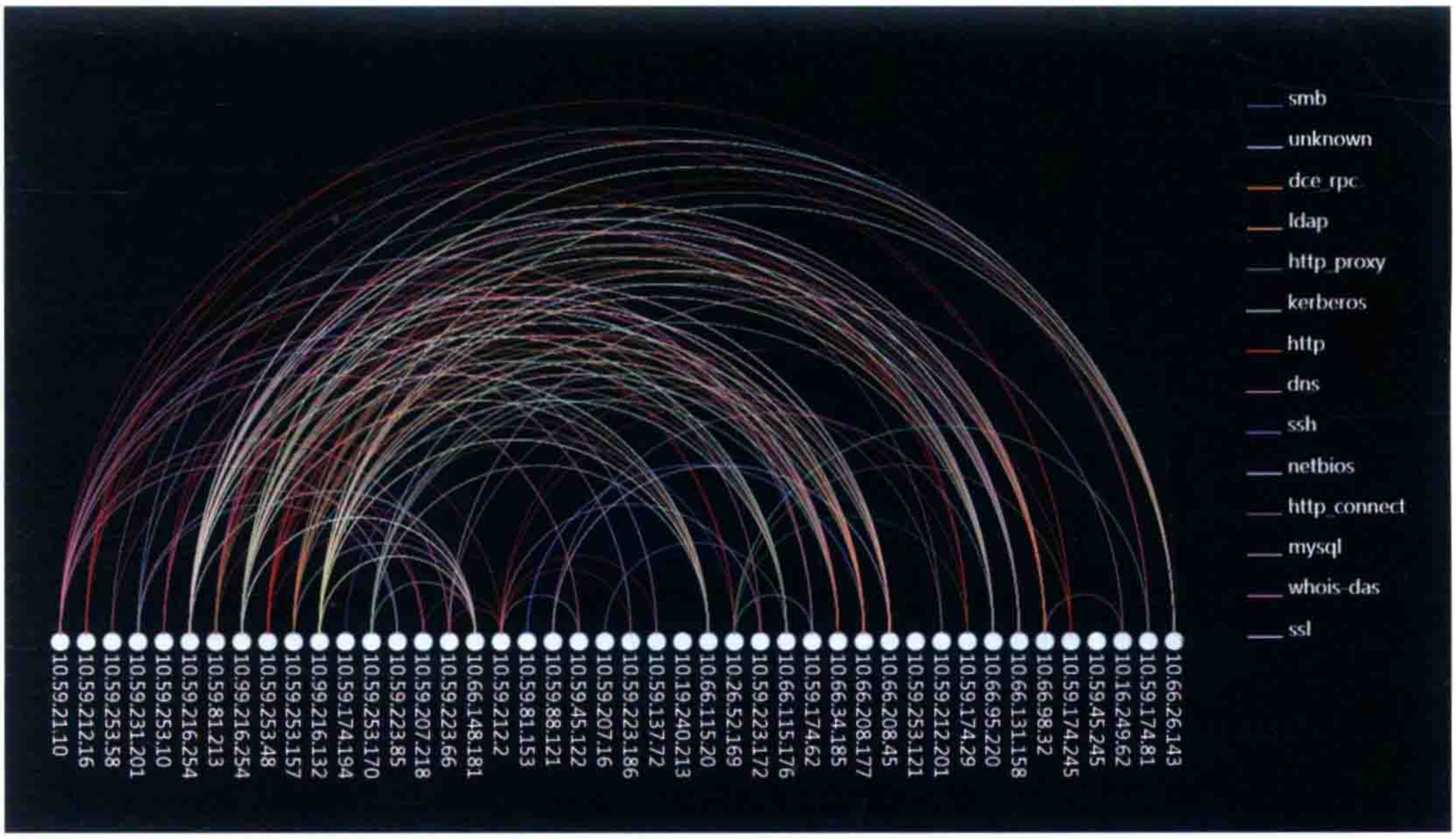

图 3-27　服务器与服务器通信的弧长链接图

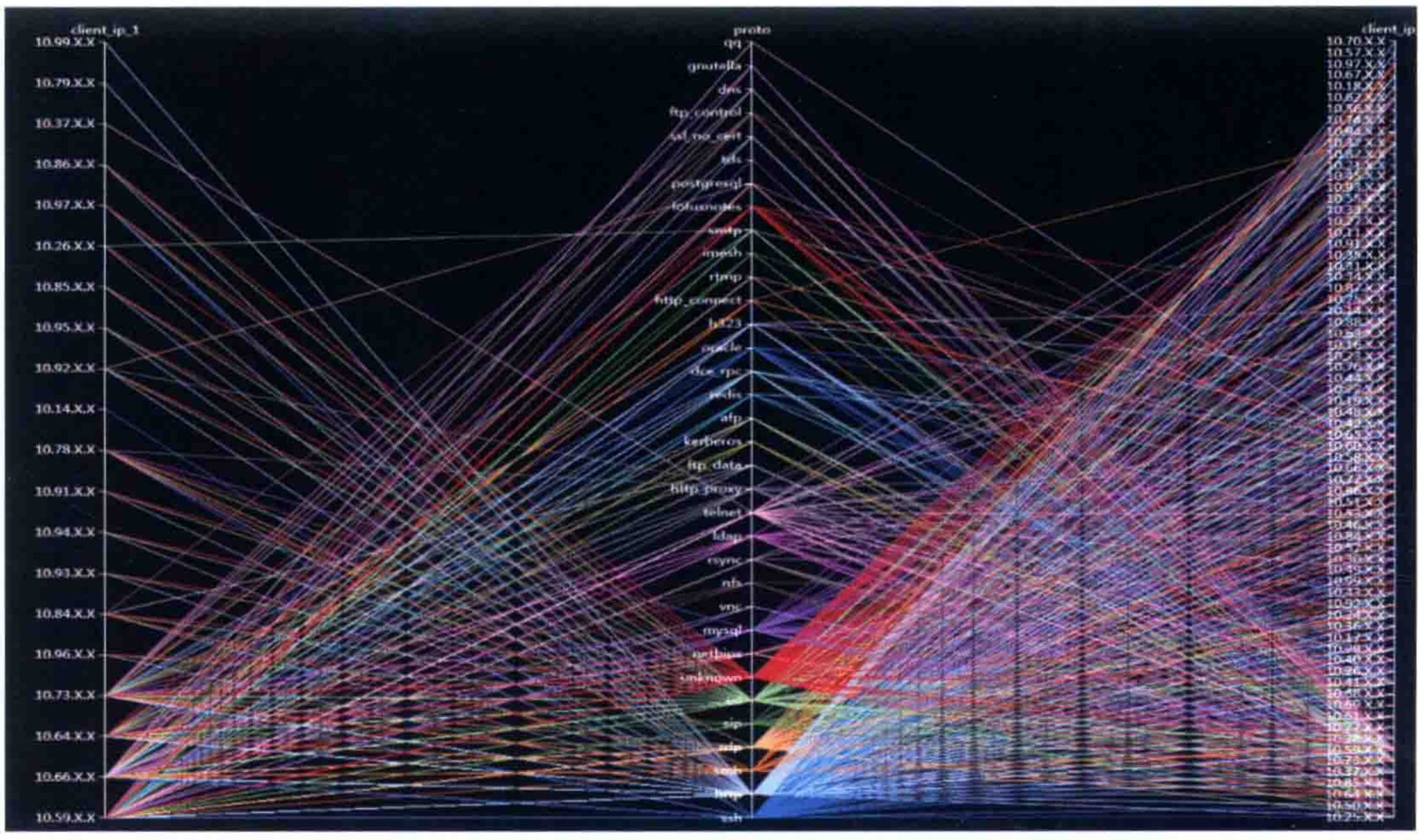

图 3-28　客户端与客户端通信的平行坐标系

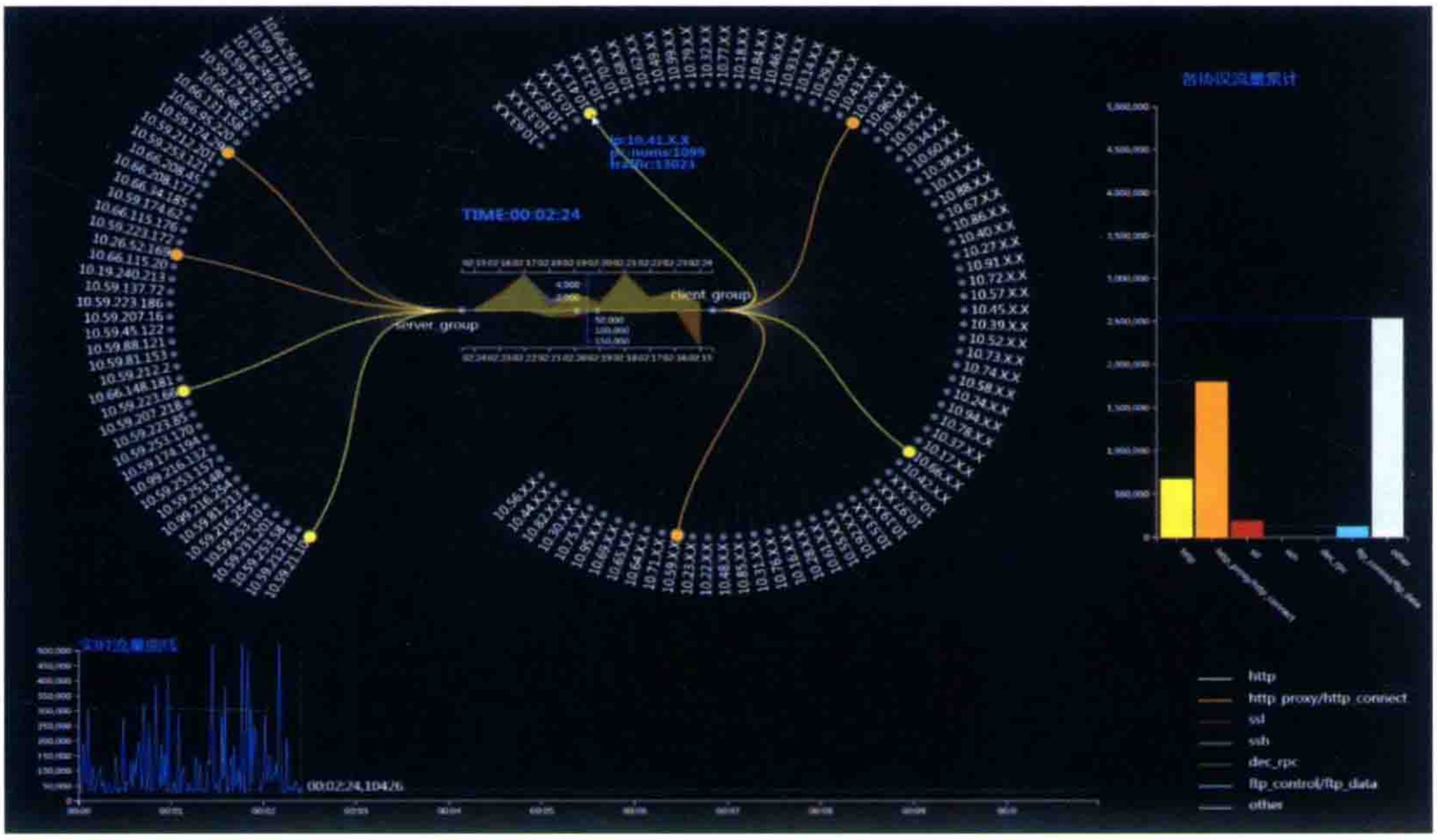

图 3-29　模拟内部网络的实时通信流量

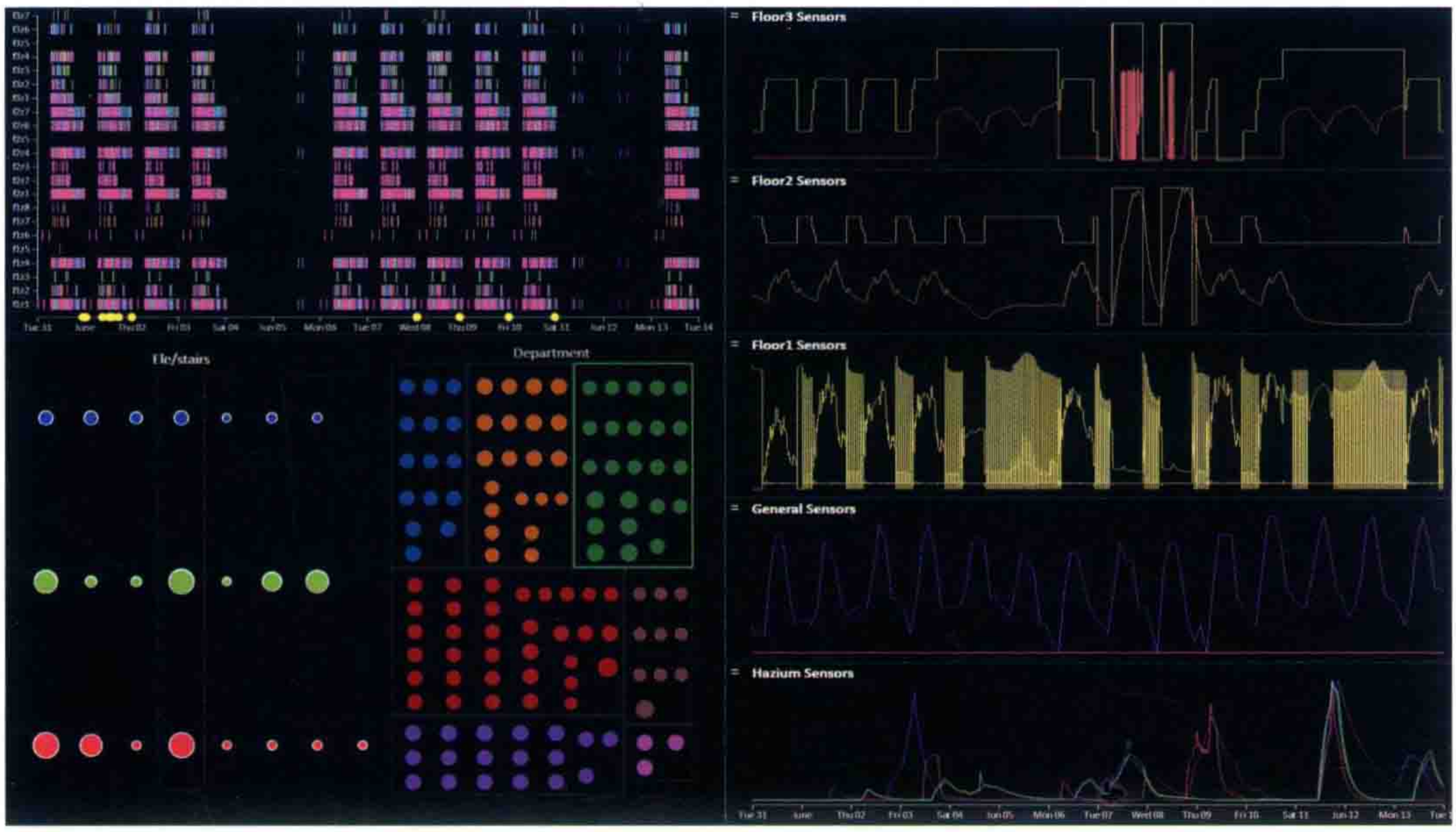

图 3-30　整体设计图

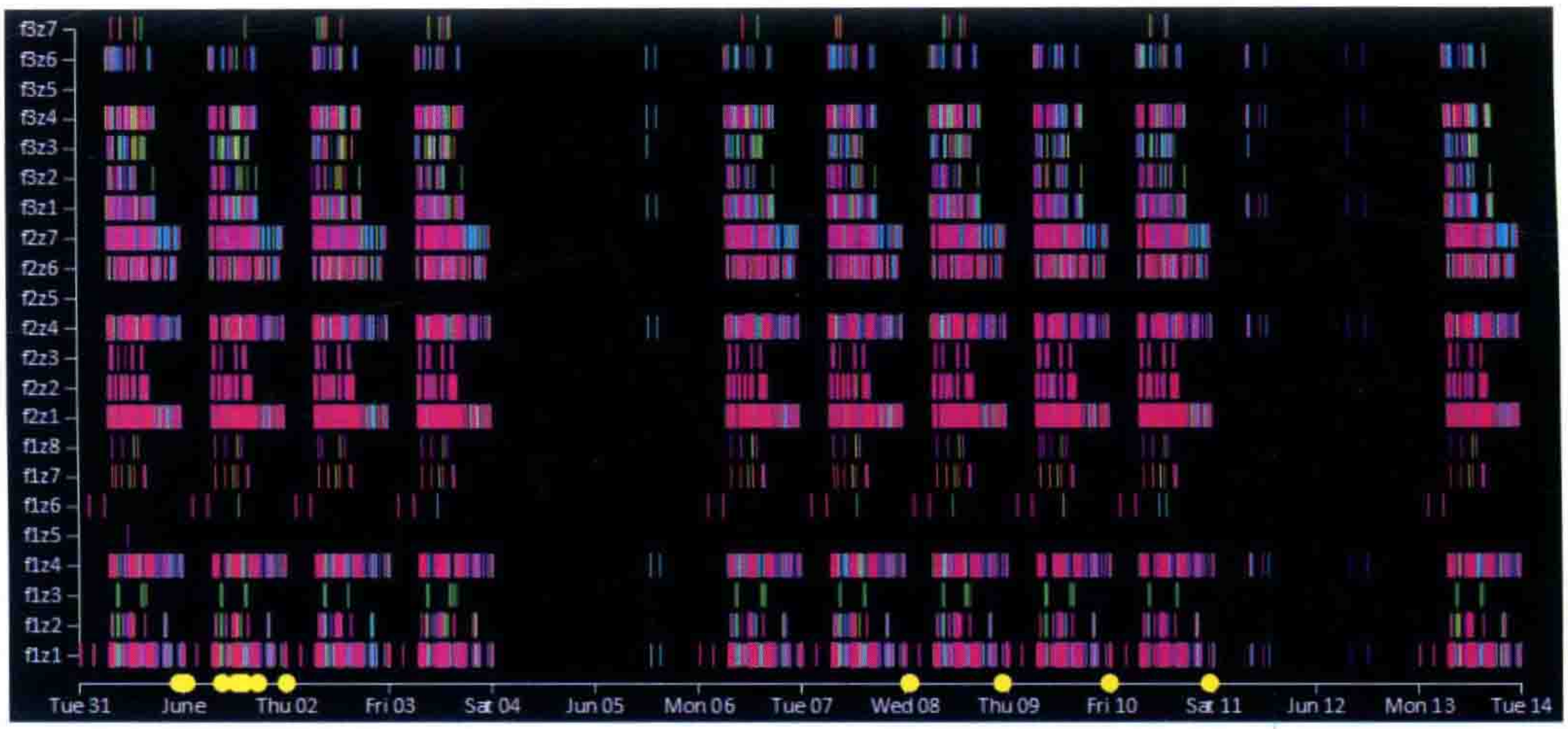

图 3-31　prox 卡数据的散点图

图 3-32　分析雇员轨迹(一)

图 3-33　分析雇员轨迹(二)

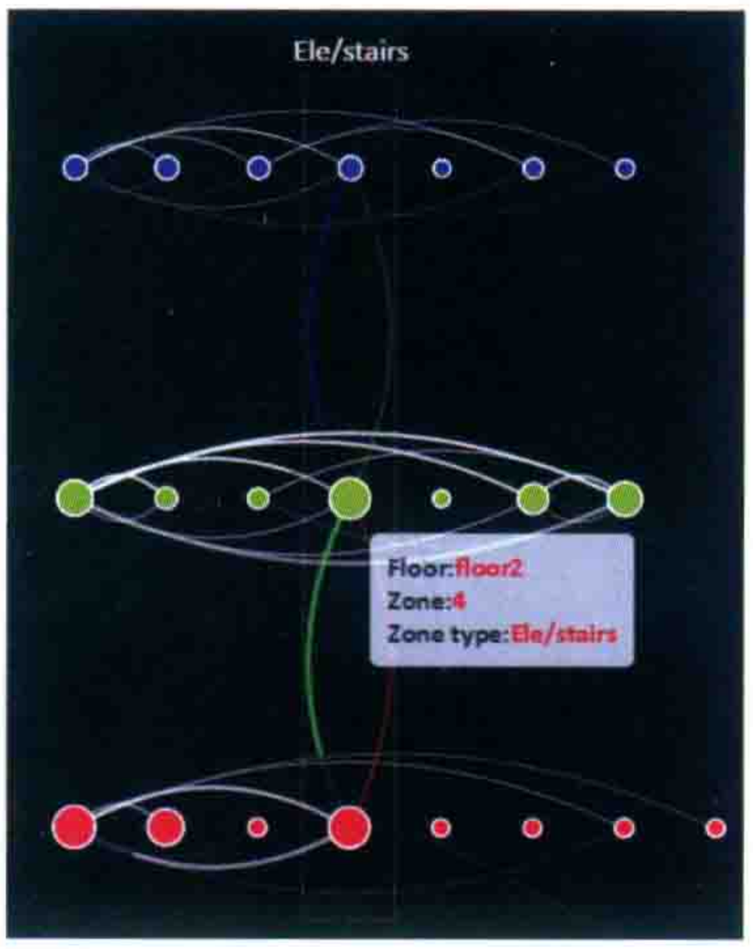

图 3-34　雇员在不同 prox 区域的轨迹图

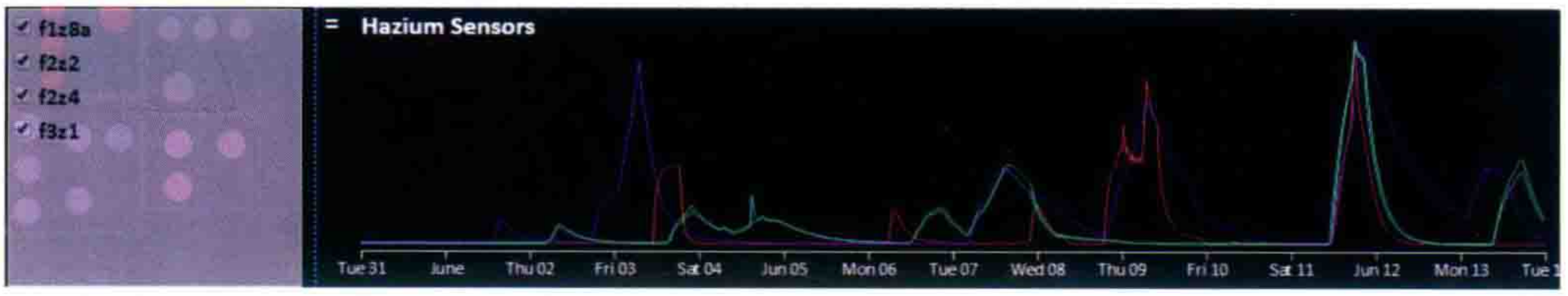

图 3-35　Hazium 传感器数据读数的折线图

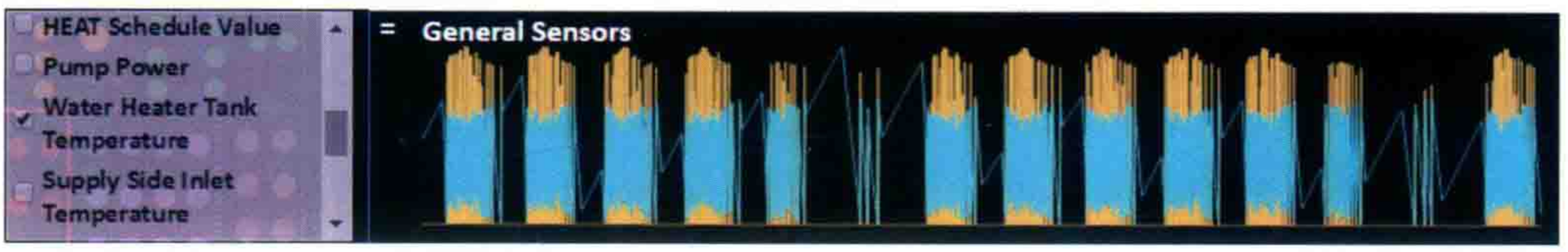

图 3-36　热水器燃气消耗和热水器水箱温度变换规律的折线图

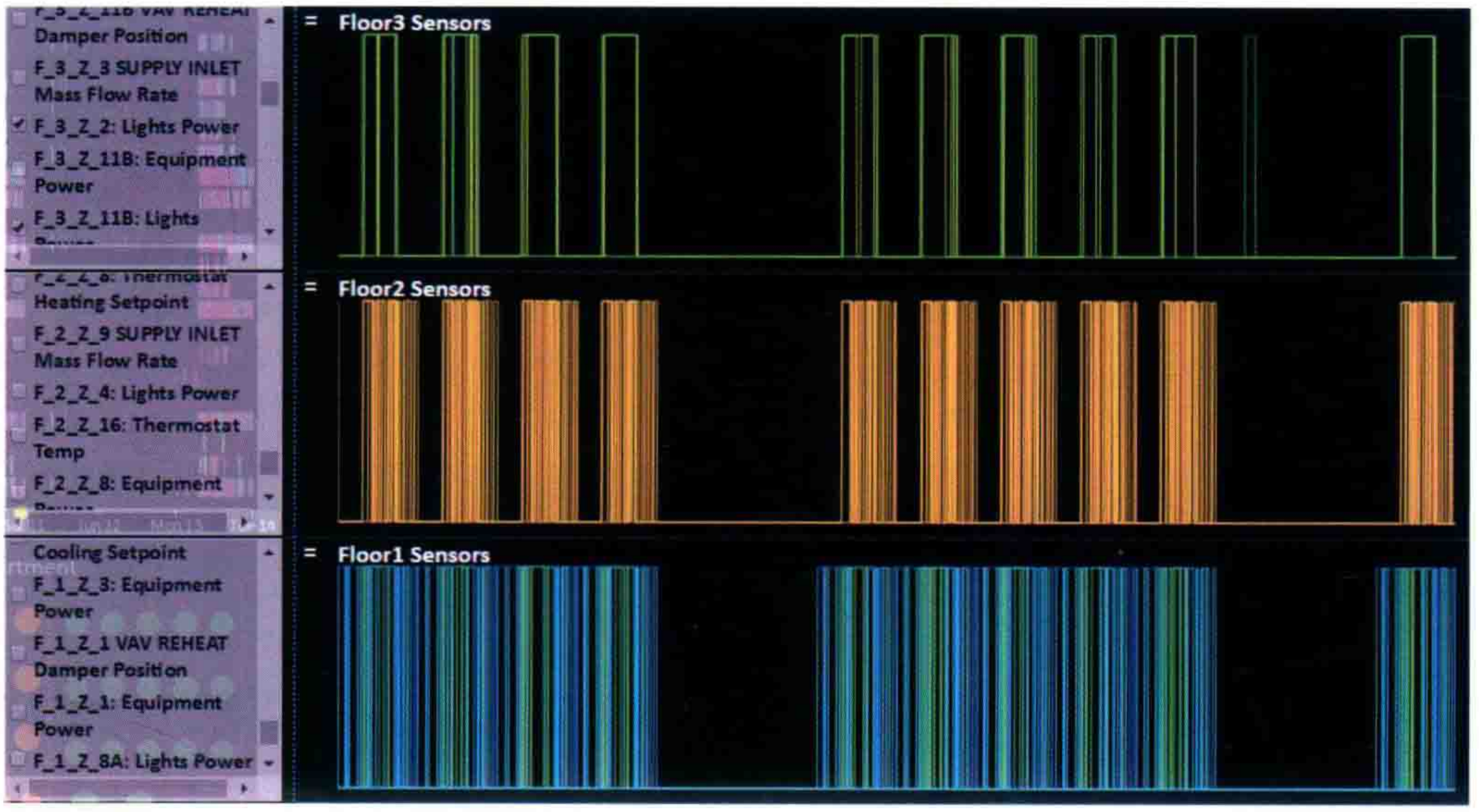

图 3-37　建筑物内所有区域的照明电路开关图

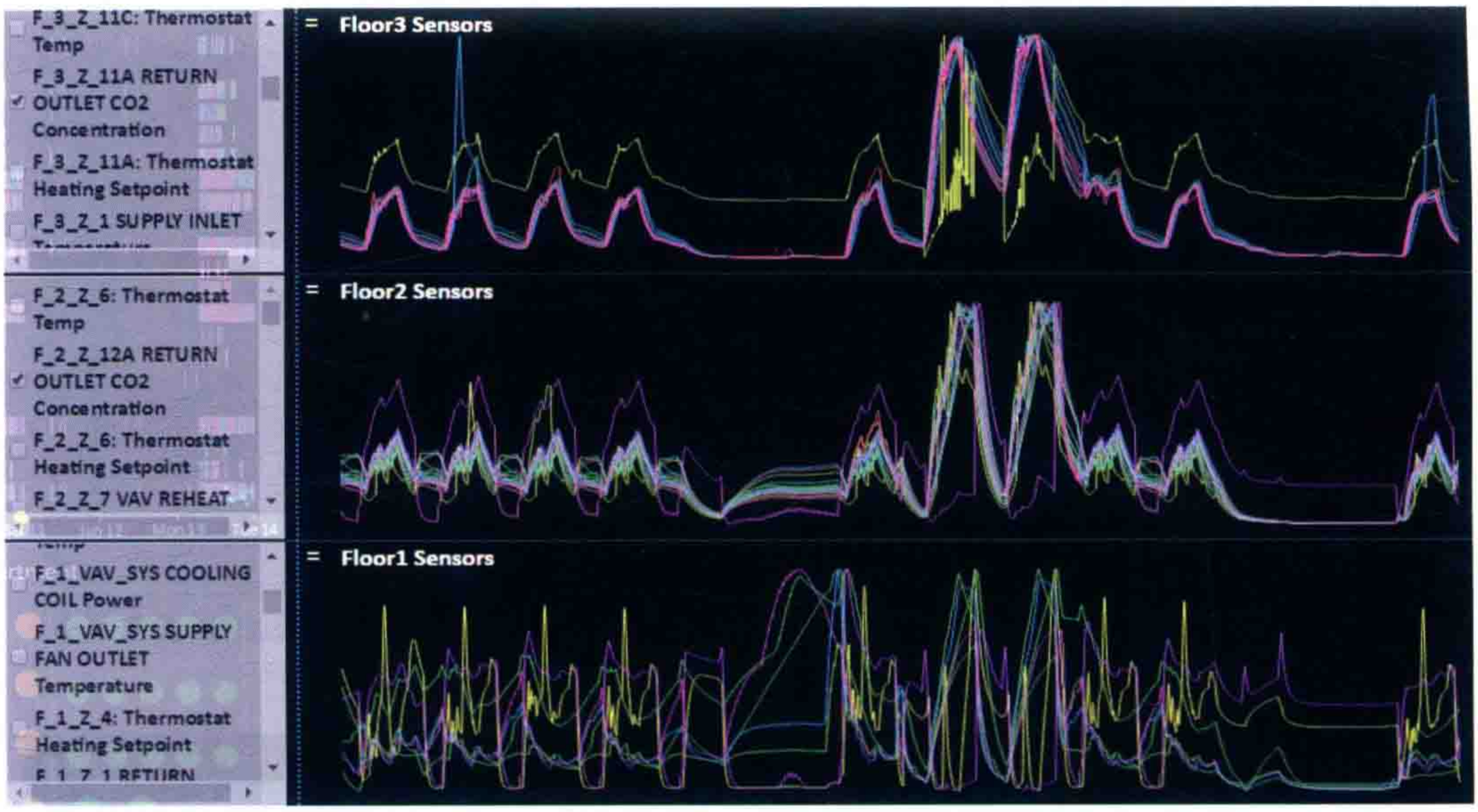

图 3-38　通风口处二氧化碳浓度传感器读数变化的折线图

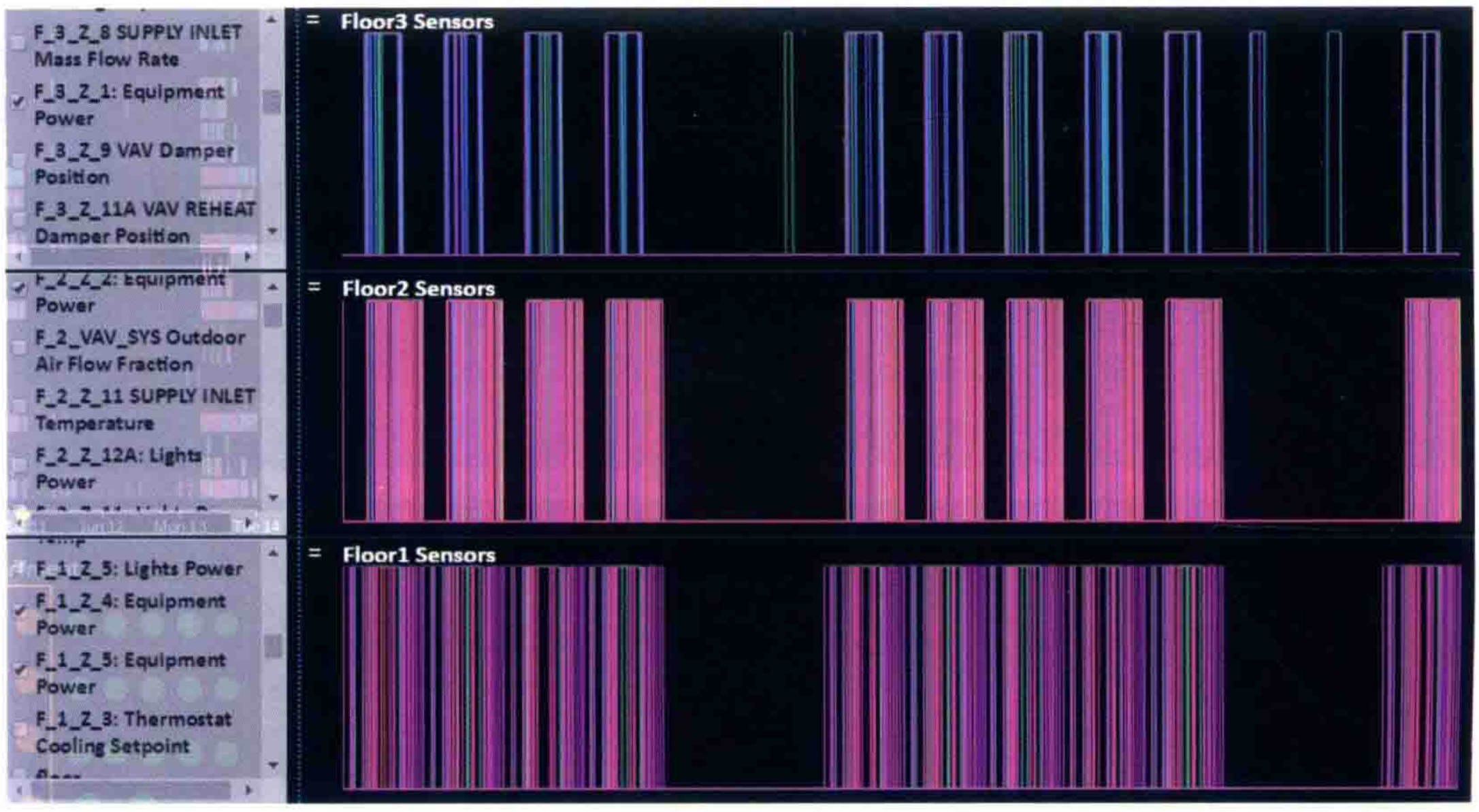

图 3-39　建筑物设备用电量变化规律图

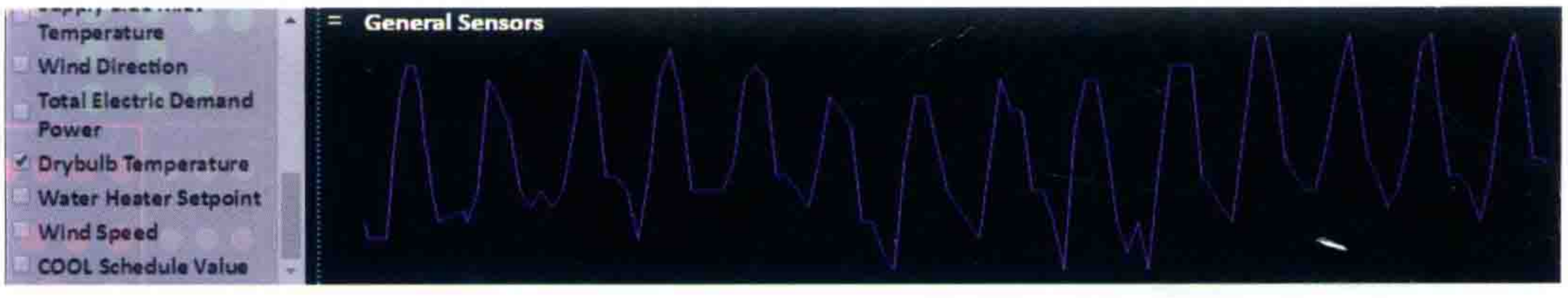

图 3-40　干球温度传感器的读数变化图

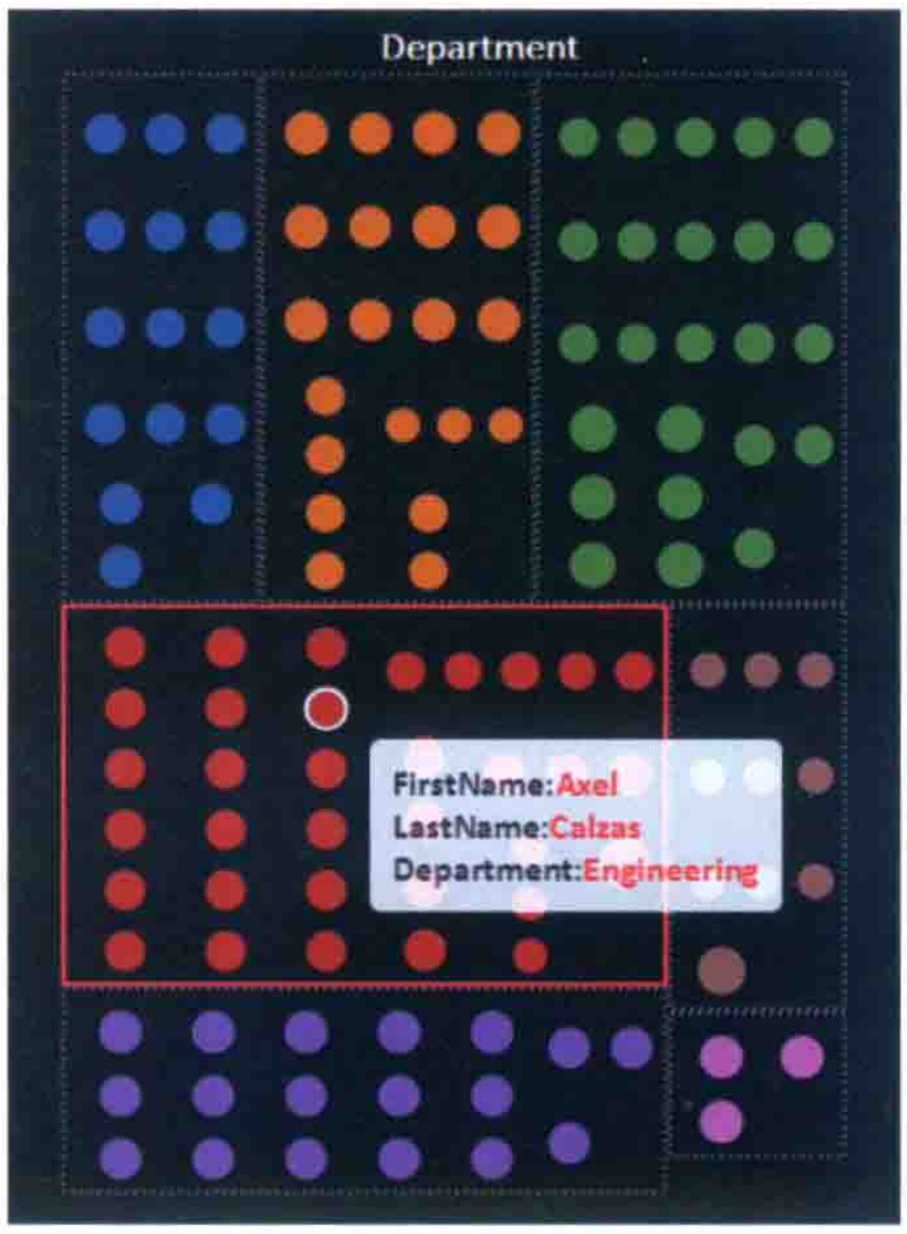

图 3-41　不同部门的雇员分布

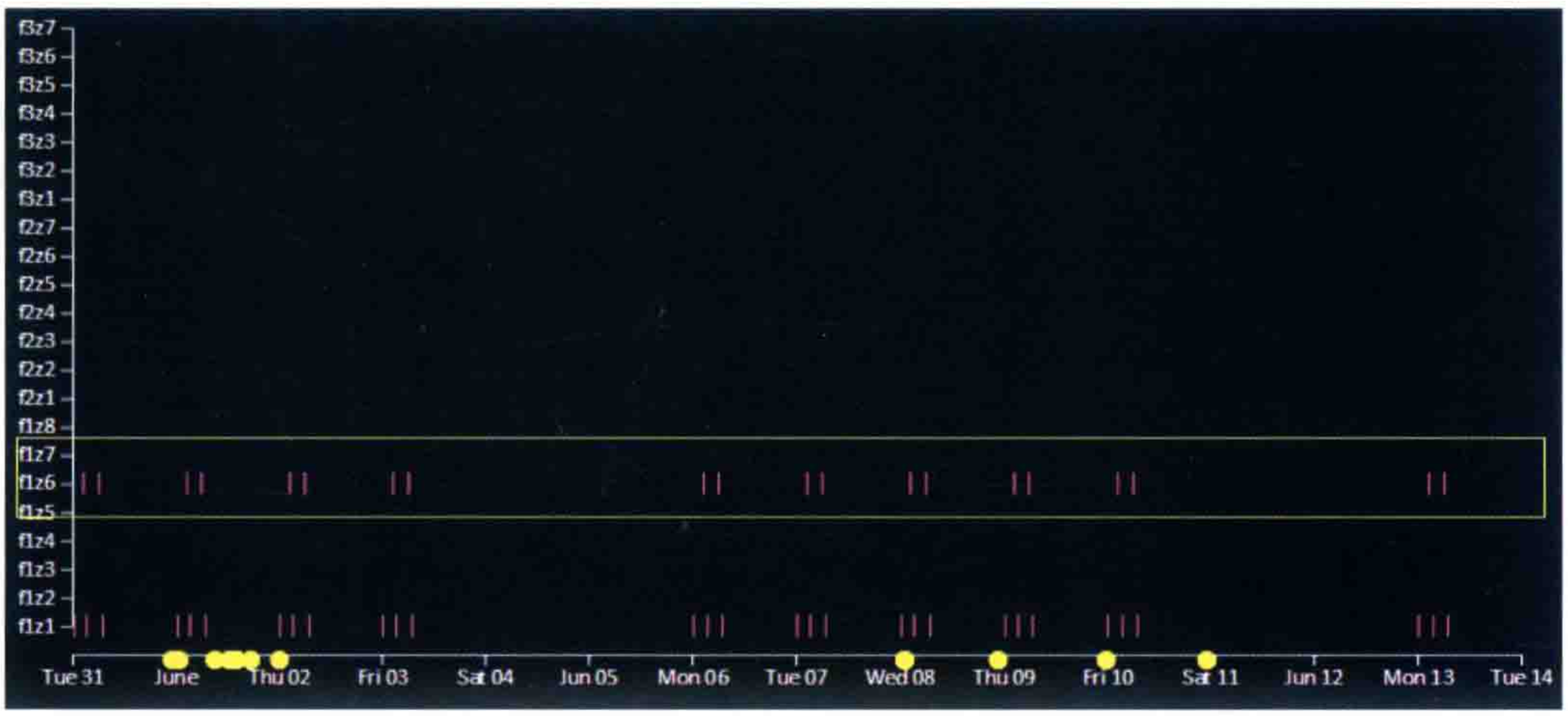

图 3-42　对一层区域 6 的分析图

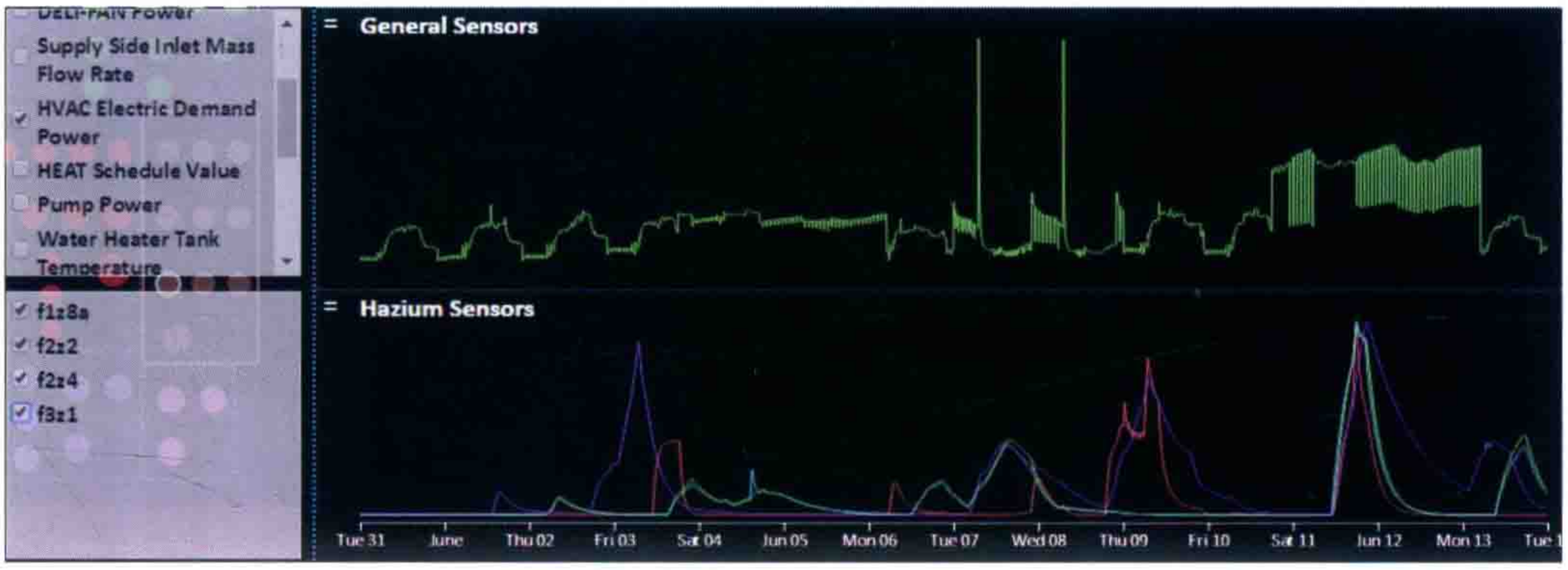

图 3-43　HVAC 区域的用电量和 Hazium 浓度之间的关系图

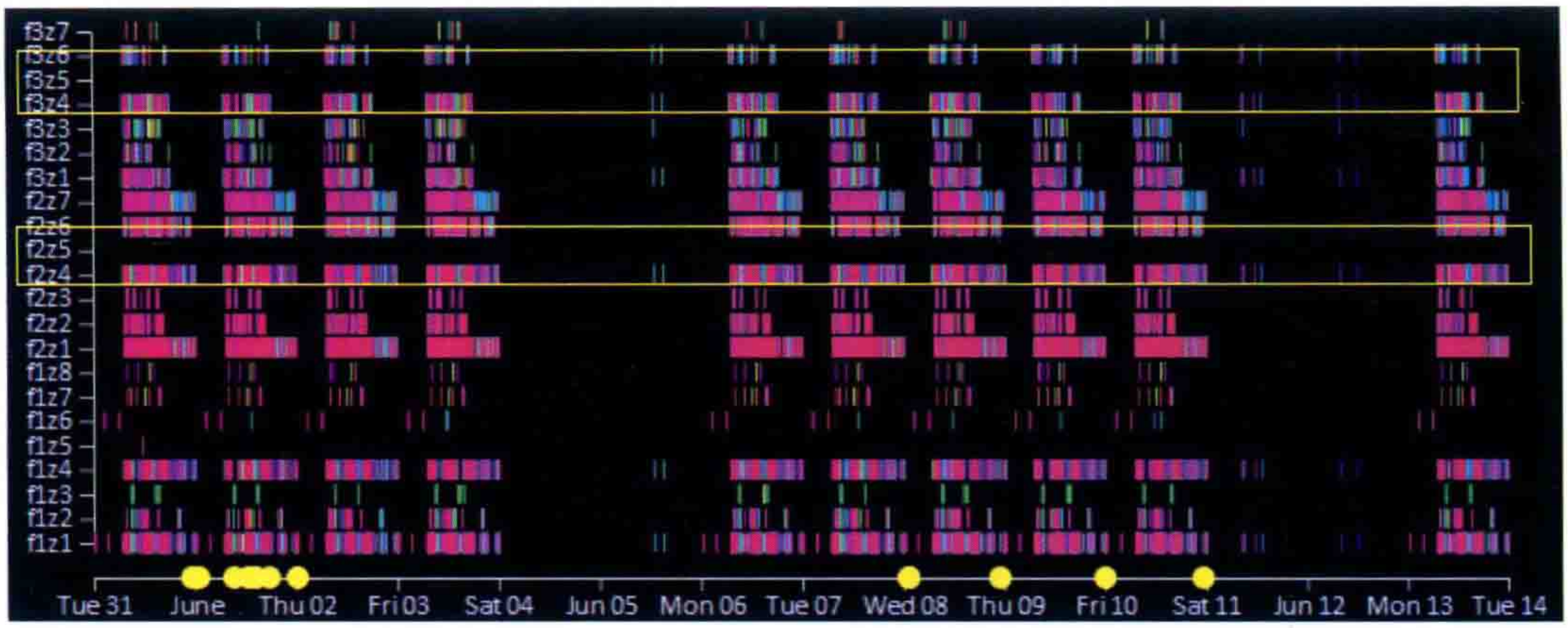

图 3-44　prox 卡散点图

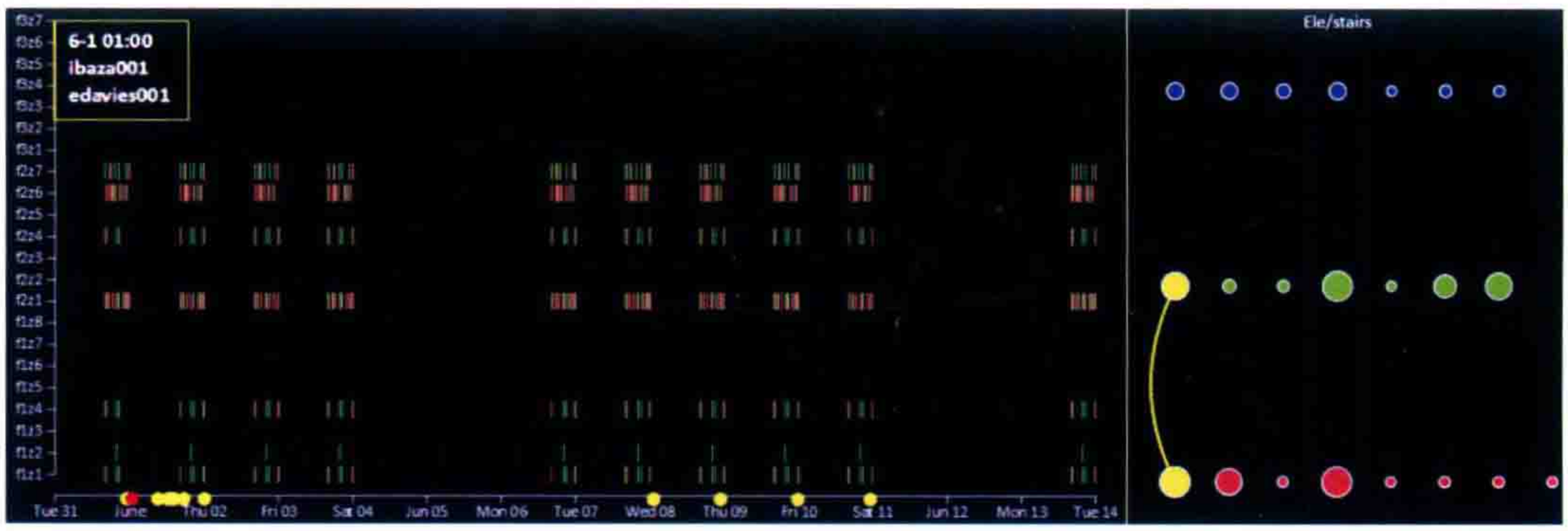

图 3-45　对 ibaza001 和 edavies001 活动异常的分析

图 3-46　SUPPLY INLET Temperature 传感器读数异常

图 3-47　一层的区域 2 设备功率传感器异常

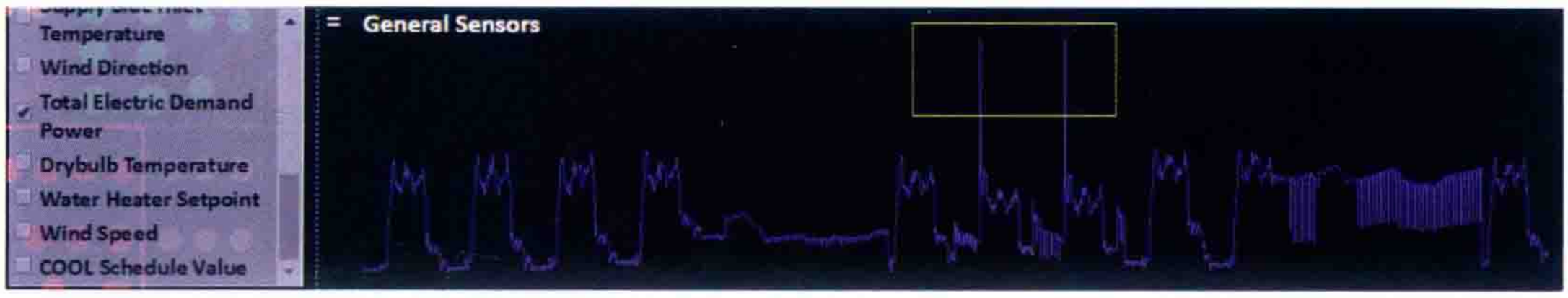

图 3-48　电源功率异常

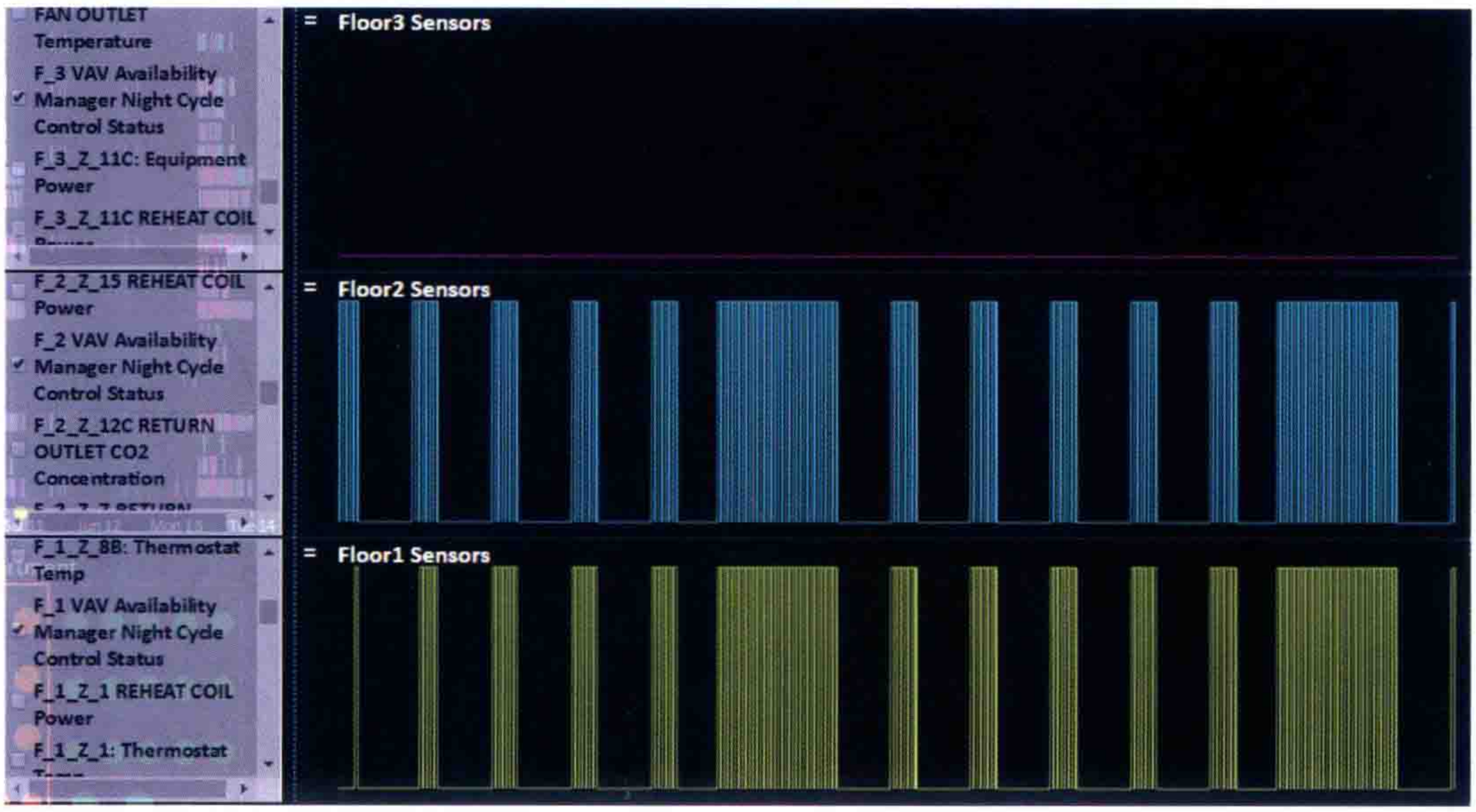

图 3-49　Availability Manager Night Cycle Control Status 的读数异常

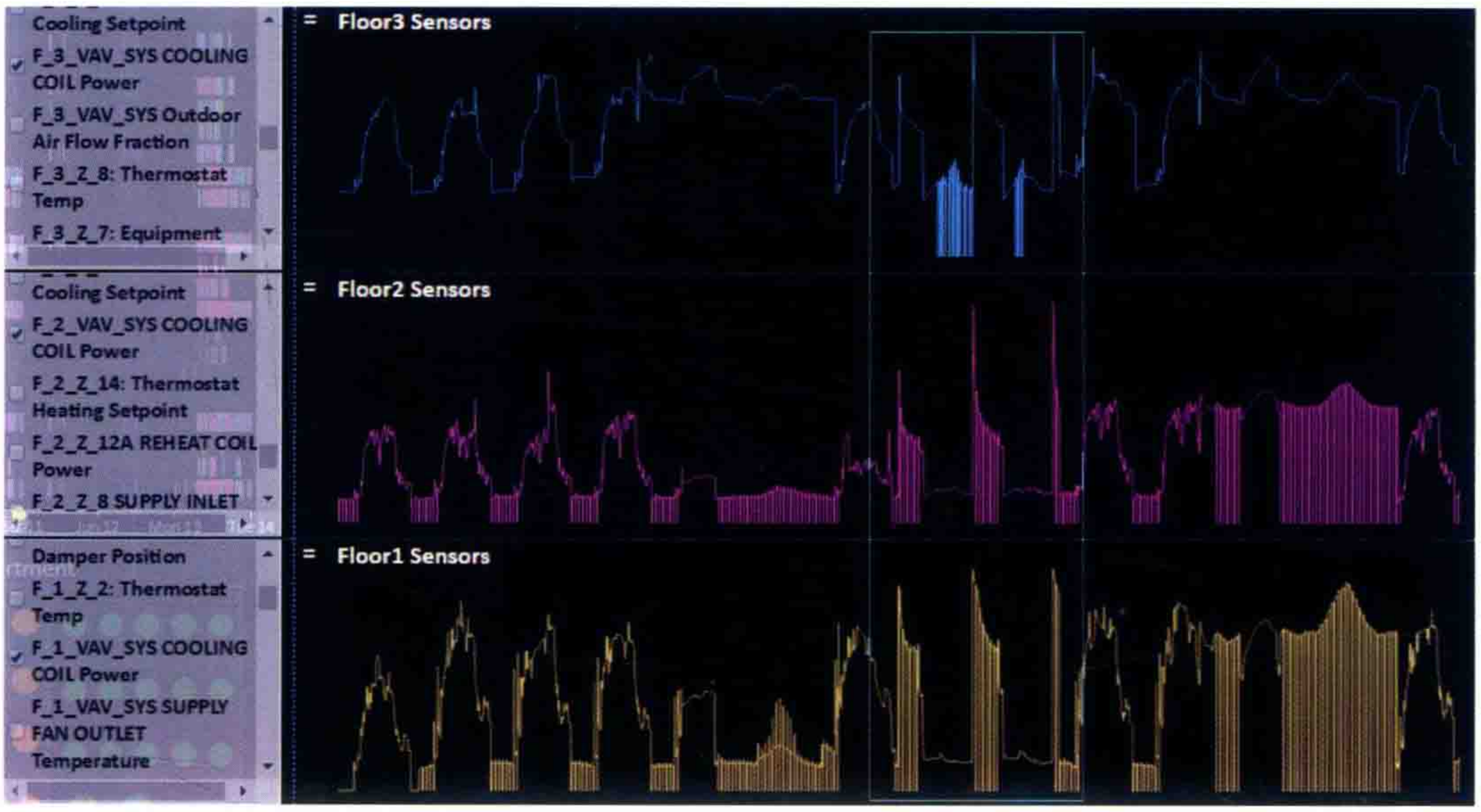

图 3-50　冷却系统的功率异常

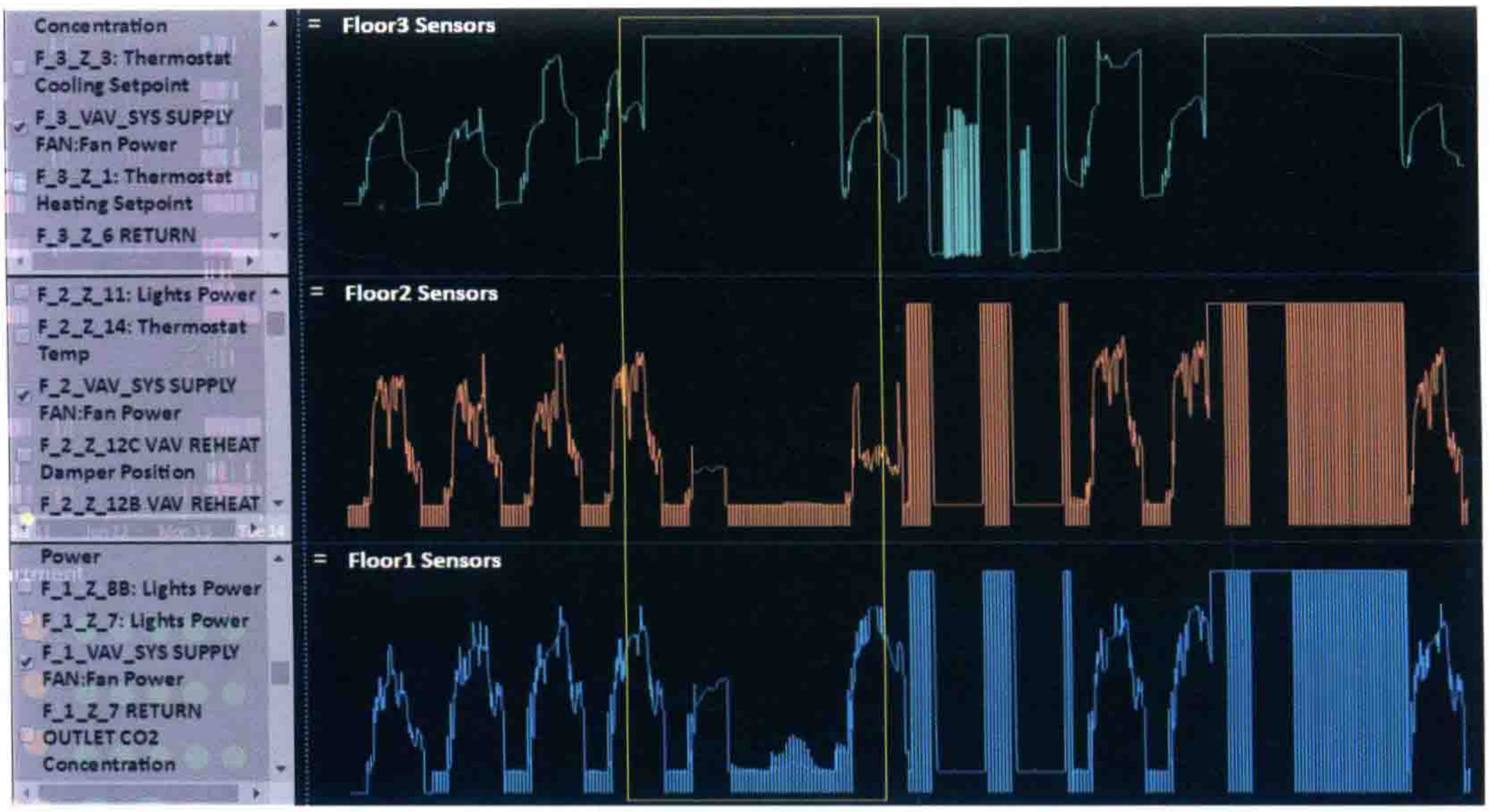

图 3-51　3 个楼层的风扇用电功率异常

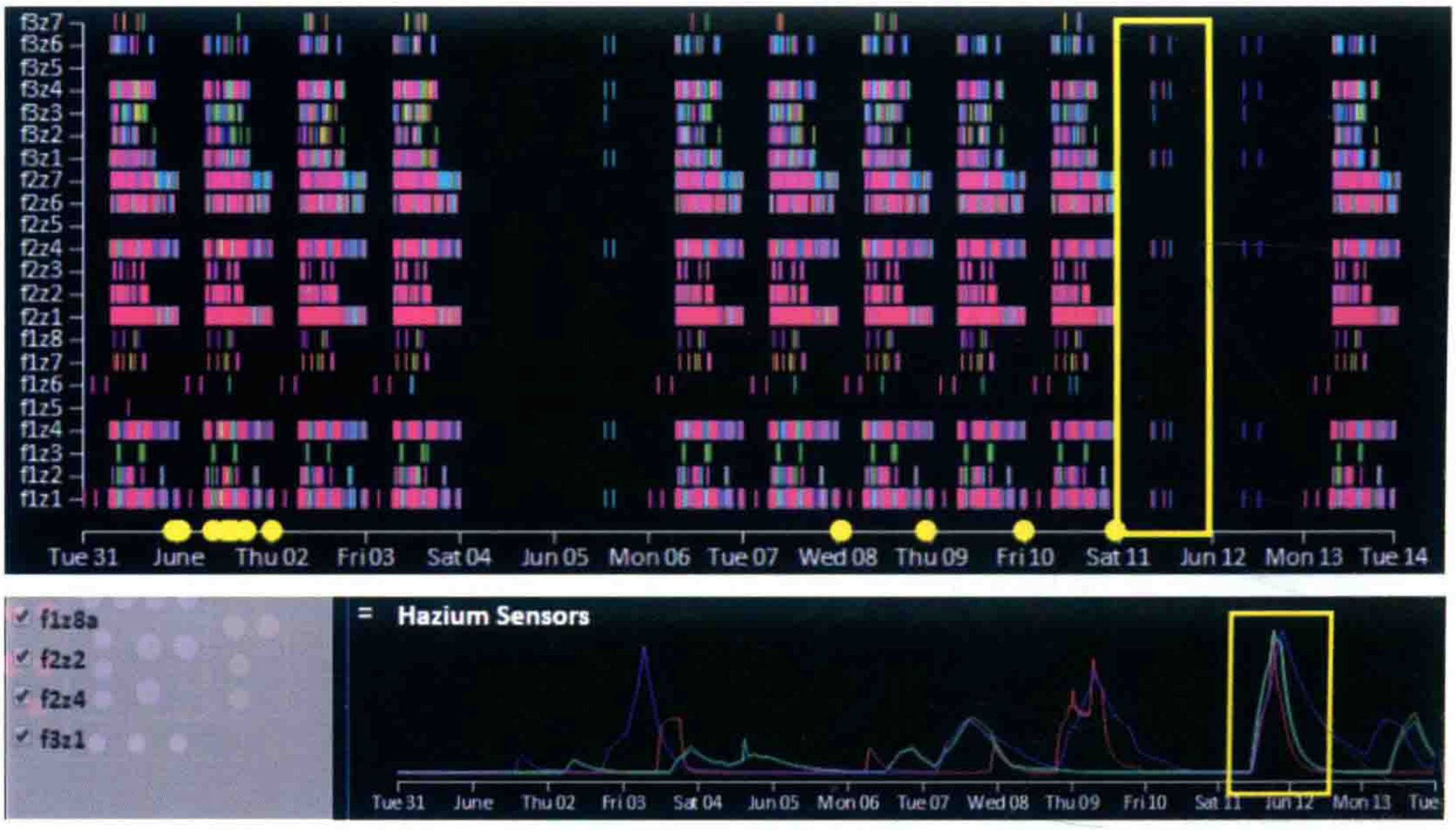

图 3-52　人员活动散点图