项目编号：T030PRP28062



**本科生研究计划（PRP）研究论文**

(第28期)

**论文题目： 阶段式动画动态网络在网络安全分析中的应用**

**项目负责人： 董笑菊 学院（系）：电子信息与电气工程学院**

**指导教师： 董笑菊 学院（系）：电子信息与电气工程学院**

**参与学生：\_\_\_\_\_项一男、舒润萱、丁天琛\_\_\_\_\_\_\_\_\_**

**项目执行时间：\_2015\_年\_\_9\_\_月 至 \_2016\_年\_\_9\_\_月**

阶段式动画动态网络在网络安全分析中的应用

电子信息与电气工程学院 F1403001 舒润萱 5140309010

指导教师 电子信息与电气工程学院 董笑菊

摘 要

现如今，互联网变得越来越普及，网络安全也成为了一个重要的课题。在海量的网络数据中，可视化是从中提取信息的一种直观、快速的方式。本文将提出一种可视化方案，多维度地分析上海市城域网的网络监控数据。本文主要介绍基于阶段式动画与力导向算法的动态网络在分析网络监控数据中的应用，以及可视化系统中各视图的联动。

关键词：网络安全，可视化，阶段式动画，动态网络，联动

ABSTRACT

In present days, Internet becomes increasingly common, which makes network security an important issue. Data visualization is an intuitive and fast way to extract the information from the data. In this paper, a plan of visualization will be presented. It can be used to analyze the monitoring data of metropolitan area network of Shanghai multi-dimensionally. This paper will focus on the application of staged-animation-based and force-directed dynamic network on analyzing network monitoring data. The interaction of views in the visualization system will also be mentioned.

**Keywords:** network security, data visualization, staged animation, dynamic network, interaction

1. 简介

动态网络是数据可视化中最常用的方案之一，它可以表达包括社会网络、商业网络、人际关系网络与互联网网络等各种网络。在这之中，由于动态网络的特征，使得它与互联网网络有着极高的契合度：动态网络可以简单、快速地表达各IP地址之间的通信状态，若稍加修改，则可以包含更多的信息。

然而，互联网通信的数据量是极大的，即使是某个大型局域网，每一时刻都可能有上千条连接的状态发生改变，这对可视化的动态表达来说是个巨大的挑战。合适的动态表达可以让用户对大型网络状态变化的认识更加简单高效，从而从中提取更多的信息。本文提出了一种基于阶段式动画的动态网络可视化方案，目标在于有效地减少用户的认知成本。

本项目提出了一种通过多个视图对网络连接数据进行多角度分析的方案。本文将重点关注系统中的动态网络部分、以及各视图的联动效果。

1. 相关工作

数据可视化早已是一个成熟的课题，在这个课题之下有许多的可视化方案被应用于各个领域。在网络安全领域，也有非常多的方案[1]。早期的工作如Girardin采用了一种“自组织图”（self-organizing map）[2]的方式进行恶意网络行为的分析；Krasser则采用了一种动态平行坐标轴的方式[3]进行分析。

在这些研究中，使用动态网络的方式进行分析的并不多，然而动态网络与互联网通信天生有着极高的契合度：动态网络可以直观地显示出互联网的拓扑结构及其变化。基于这一点，本文采取了动态网络这一方式对互联网数据进行可视化。而为了改进动态网络的动画效果，本文基于Bach的研究[4]的阶段式动画，并对其进行改进，最终应用于互联网网络安全数据分析。

1. 系统设计与实现
   1. 系统总览

本系统的目标是多角度地对网络监测数据进行分析。因此，我们使用了三个视图对同一组数据进行分析。系统的总览如图3.1所示。

(a)

(d)

(c)

(b)

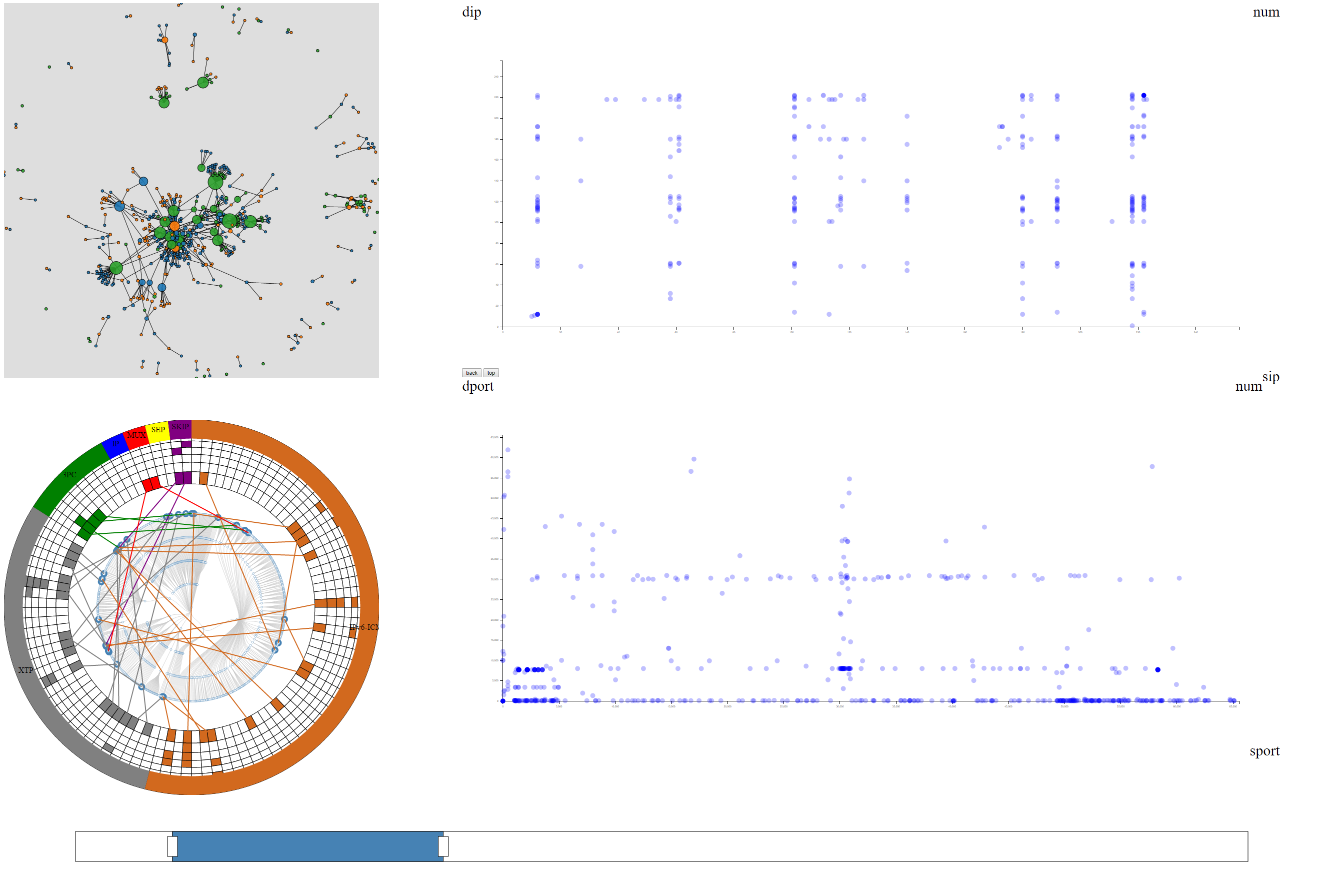


图3.1 系统总览图：(a) 基于阶段式动画的动态网络；(b) 由通信双方的IP及端口组成的散点图；(c) 由IP及通信协议组成的可视化方案；(d) 进行数据筛选所用的时间轴。

图3.1(b)上半部分显示了源IP、目的IP组成的通信分布情况，由IP段进行分类。下半部分则显示了源端口、目的端口组成的通信分布情况。点击其中的某一个点可以对数据进行筛选。通过该视图，可以观察到通信IP及端口的分布情况，通过点的分布（横纵规律、密集程度等）以及深浅（通信量大小、频繁程度）来找出可疑的通信。

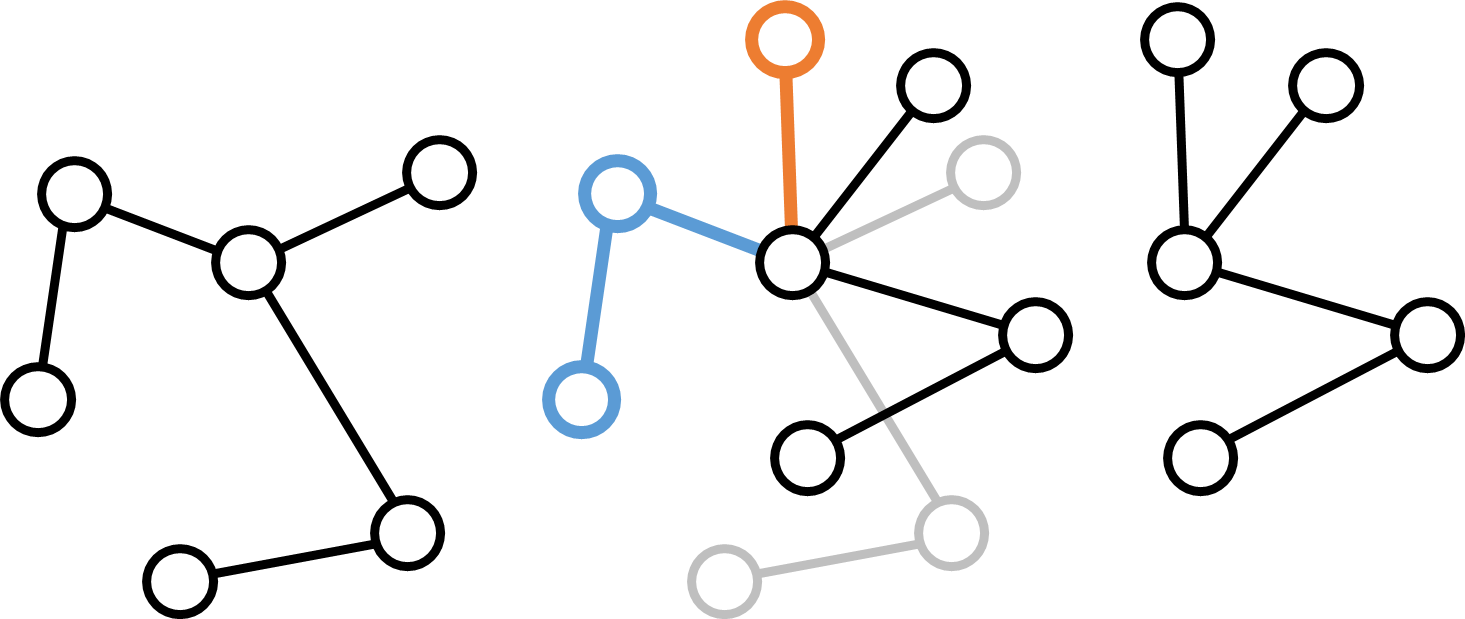
图3.1(c)由中心的IP树及外层的协议类型组成。连接IP和协议的线代表了该IP通信时使用了该协议。通过该视图，可以观察到各IP某时刻所使用的网络协议的状态，可以反应网络异常的可能攻击类型。

图3.1(d)的时间轴可以对所有三个视图进行时间上的筛选。通过调整时间可以观察到网络通信情况的改变。

本文将着重于图3.1(a)的动态网络与三个视图的联动效果。

* 1. 基于阶段式动画的力导向动态网络
     1. 阶段式动画

动态网络的一大优点就是可以展示网络结构的变化。为此，合适的动画效果是十分重要的。对于数据量较小、结构较为简单的动态网络，一般的做法是使用平行式动画，如图3.2所示。

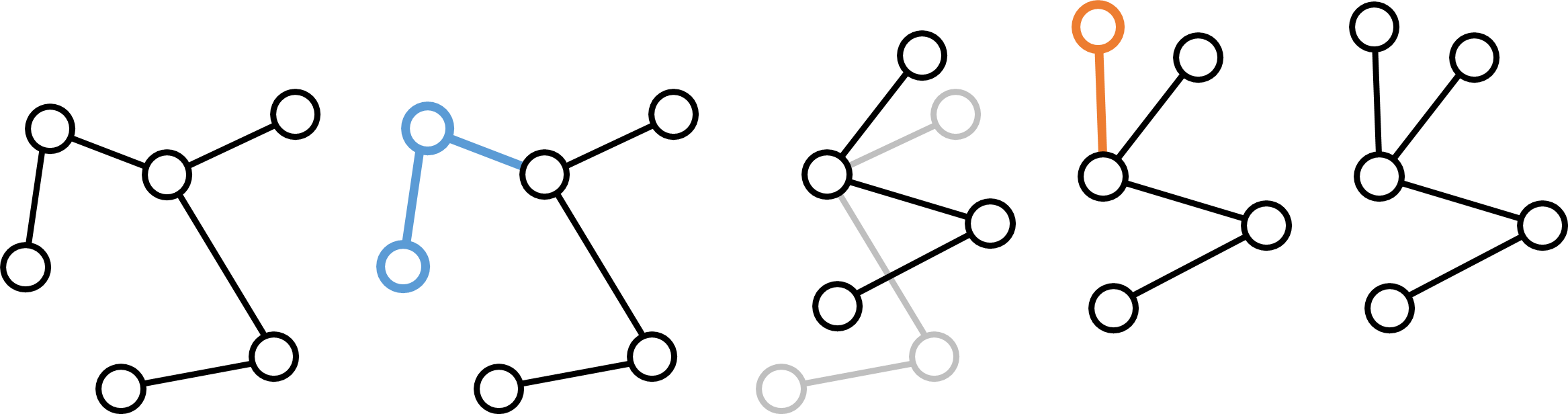


1. (b) (c)

图3.2 平行式动画：(a) 网络初始状态；(b) 正在发生变化的网络：消失（蓝色）、出现（橙色）、位置变化（灰色）；(c) 网络最终状态。

平行式动画最大的优点在于，在网络结构不复杂时，平行式动画可以有效减少动画长度，让用户专注于可视化传达的信息本身。但是当信息来源主要是动画，或者网络结构复杂时，平行式动画的劣势也显而易见。

在分析网络监控数据时，网络结构复杂，且其变化也提供了相当的信息量。为了解决平行式动画在此方面的劣势，通常的做法是将动画效果减慢。然而，这样的做法会大大降低用户认知的连贯性[4]。为此，Bach[4]提出了一种阶段式动画，它的效果如图3.3。

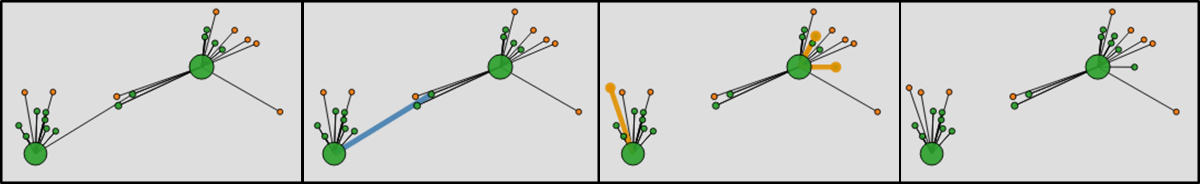


(a) (b) (c) (d) (e)

图3.3 阶段式动画：(a) 网络初始状态；(b) 消失阶段（蓝色）；(c) 位置变化阶段（灰色）；(d) 出现阶段（橙色）；(e) 网络最终状态。

在阶段式动画中，图3.2(b)被拆分成图3.3(b)(c)(d)三个部分。整体上看，动画时间被延长了两倍，但是相较同样长度的平行式动画，每一部分的动画时长都被有效地控制。在每一阶段，用户对网络变化状态的认知是不同的。因此，阶段式动画可以大大提高用户认知的连贯性。

在实际应用阶段式动画时，本系统对此进行了一些改动。图3.4所示为本系统的局部效果展示。



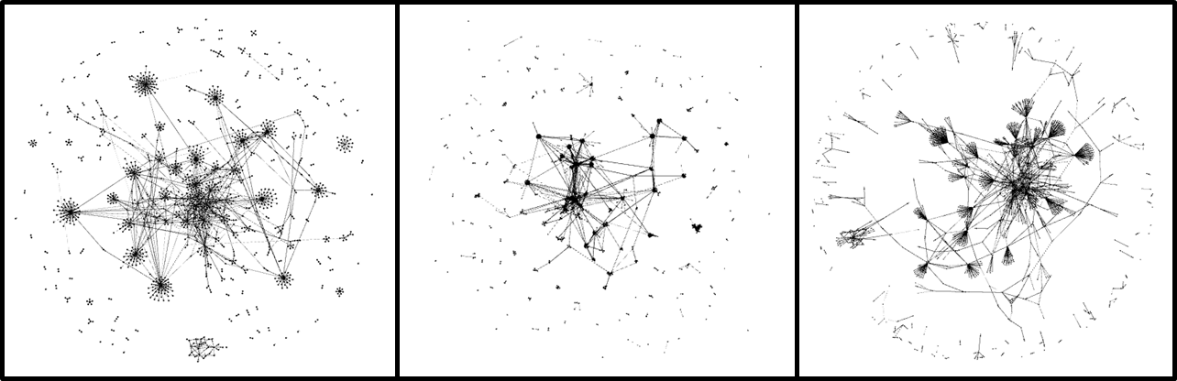
1. (b) (c) (d)

图3.4 阶段式动画的实际应用效果：(a) 网络初始状态；(b) 消失阶段（蓝色）； (c) 出现阶段（橙色）；(d) 网络最终状态。

在将阶段式动画实际应用与网络监控数据分析时，本系统去掉了图3.3(c)的位置变化阶段。在数据的所有时间点上，某一节点的位置不会随时间的变化而变化。做出此改动出于以下几点考虑：首先，在互联网网络中，节点个数极多，即使在本系统所使用的上海市城域网数据，每一时刻（跨度为1毫秒）即有30~50条数据。在如此庞大的网络中，其拓扑结构的变化不会十分显著，一个明显的表现是，无论何时，服务器节点依然是服务器，其行为模式（高吞吐量、连接节点数量等）并不会发生显著变化；其次，若在全局对节点进行合理的布局，局部区域的连接情况变化对全图的布局将不会产生影响（布局策略将在下一部分阐述）；最后，由于本系统使用JavaScript实现，受限于语言本身特性和硬件设备，实时改变布局对设备的算力是个巨大的挑战。

* + 1. 动态网络布局策略

在对互联网网络数据进行动态网络布局的时候，特别是在要揭示其变化的过程的情况下，要考虑到其全局布局的美观以及局部布局的合理性。为此，在实现动态网络前，我们使用了开源软件Gephi[5]进行布局的尝试与研究。在Gephi中，流行的布局算法包括ForceAtlas2[6]、OpenOrd[7]与Yifan Hu[8]。这些算法的效果如图3.5所示。



1. (b) (c)

图3.5 Gephi中各算法对网络监控数据的布局效果：(a) ForceAtlas2算法；(b) OpenOrd算法；(c) YiFan Hu算法。

这三种算法各有优劣。ForceAtlas2算法适合中、小型网络，虽然布局较为合理，集群清晰，但对于网络监控数据这一类较大的网络，其运算时间较长。OpenOrd算法性能很高，但它适合更加大型的网络，其集群过于集中，不是很适合本系统。YiFan Hu算法运算速度较快，布局合理，且对于边权值的体现较为出色。其对于小型集群的扇形布局也可以减少空间压力。综合考虑，最终决定采用Yifan Hu算法进行动态网络的布局。

* + 1. 节点配色、大小与边权值

一个优秀的网络可视化要求其节点、边及它们的各项属性有着直观、合理的意义。好的网络可以让用户短时间提取出有效信息。在本方案中，也对这方面进行了研究。视图的效果如图3.6所示。

* + - 1. 节点

在本视图中，将节点的颜色按其行为模式分为三种：蓝色、橙色以及绿色。蓝色节点代表相应IP仅接收数据，橙色节点代表相应IP仅发送数据，绿色节点代表相应IP既接收又发送数据。同时，将节点的大小按其度（连接的节点数量）进行分类。节点越大，代表其连接的节点越多。

在这种分类方式下，IP的行为模式可以被非常轻松地分类：在图3.6中，可以看到很多非常显眼的绿色较大的节点。可以推测这些点很可能是上海市及其他省市的网络出入口服务器。

* + - 1. 边

边的长度与粗细在网络可视化中也是一个重要的部分。在本方案中，由于节点数量较多，改变边的粗细会导致视觉效果混乱，所以在本方案中没有采用边的粗细这一属性。

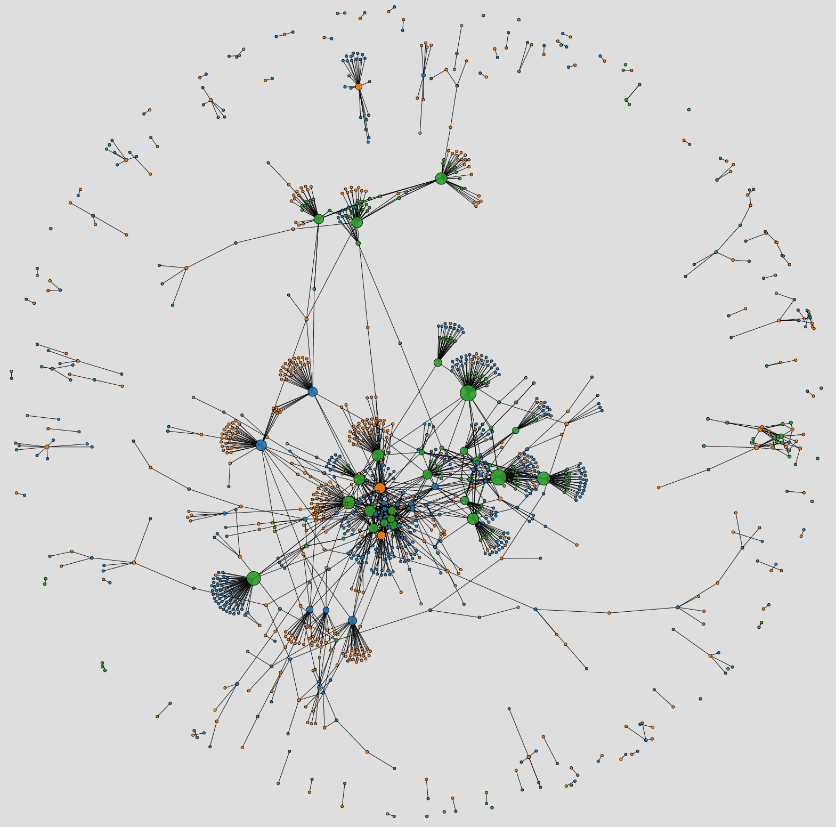
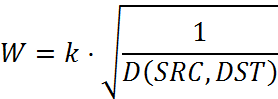


图3.6 动态网络应用效果

在本方案中，边的权值由以下公式决定：



其中，为边的权值，为一常数，对整张图的大小有一定的影响。代表与间的通信次数。边的权值与其两端节点的电斥力成线性关系[8]。因此，最终结果中边的长度与其两端IP的通信次数大致成反线性关系。通信次数越多，边长越短，反映在图上效果为两点距离更近。



* + 1. 交互

交互作为用户体验中最重要的一部分，是可视化方案中最表层、最直接的部分。在本方案中，交互包括鼠标单击节点进行筛选、鼠标拖动平移、缩放以及调整时间轴。其效果如图3.7所示。

当鼠标单击一个节点时，这个节点和与它直接相连的节点和线会被高亮，其它节点和线会变暗。这一交互可以帮助用户快速查看某一局域网或集群。筛选可以选中多个节点。单击空白区域可以取消筛选。

当滚动鼠标滚轮时，会以鼠标停留的位置为中心进行缩放操作。这一交互可以方便用户观察网络的局部特征。缩放后点的大小和线的粗细不会变化，同时阶段式动画的出现、消失高亮的效果也不会变化，使得视图的视觉效果不会特别挤。

(a)

(c)

(b)

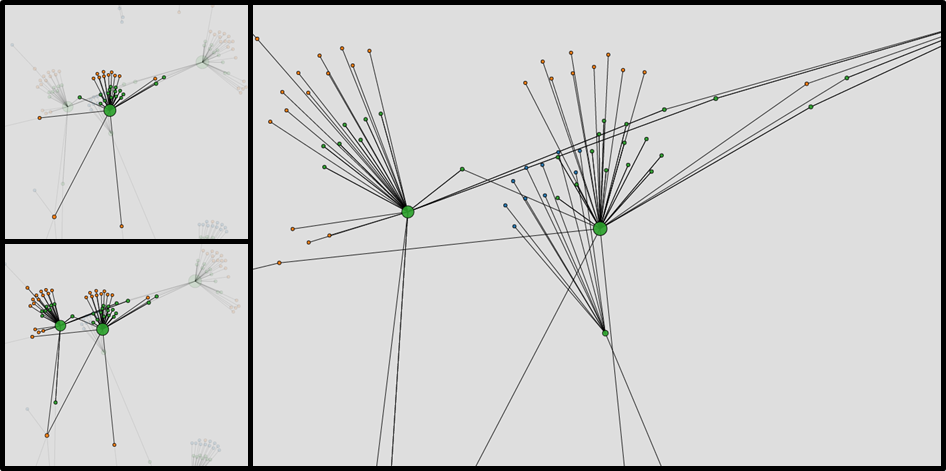


图3.7 动态网络中的交互：(a) 筛选一个节点；(b) 筛选两个节点；(c) 局部放大。

1. 可视化系统多视图联动
   1. 各视图交互

视图1（图3.1(a)）提供的交互方式包括：用户自定义筛选节点、缩放、拖动平移。其中，筛选节点会对另外两个视图产生影响。

视图2（图3.1(b)）提供的交互方式包括：用户自定义筛选IP网段内通信、用户自定义筛选同一端口内通信。这两种交互均会对另外两个视图产生影响。

视图3（图3.1(c)）提供的交互方式包括：根据通信协议筛选通信数据。该交互均会对另外两个视图产生影响。

视图2和视图3交互的效果可以参考本项目的另一篇论文，由丁天琛同学完成。

* 1. 多视图联动

在这三个视图中，用户均可以筛选部分数据，来集中观察自己所需要的数据。在一张视图上进行筛选时，其它视图也会产生相应的改变。

* + 1. 时间轴

本系统三个视图均可以通过时间筛选数据，它们共用同一个时间轴。时间轴可以改变起始时间、结束时间，以及同步修改起始结束时间。其效果如图4.1所示。

* + 1. 数据筛选

本系统的三个视图均可以通过某些规则筛选数据。这些规则包括：筛选某节点及与其直接连接的节点（视图1提供）、筛选某网段的节点或使用某端口的节点（视图2提供）和筛选使用某通信协议的节点（视图3提供）。当用户在某一视图上进行筛选时，其它的视图也会作出反应，如图4.2所示。

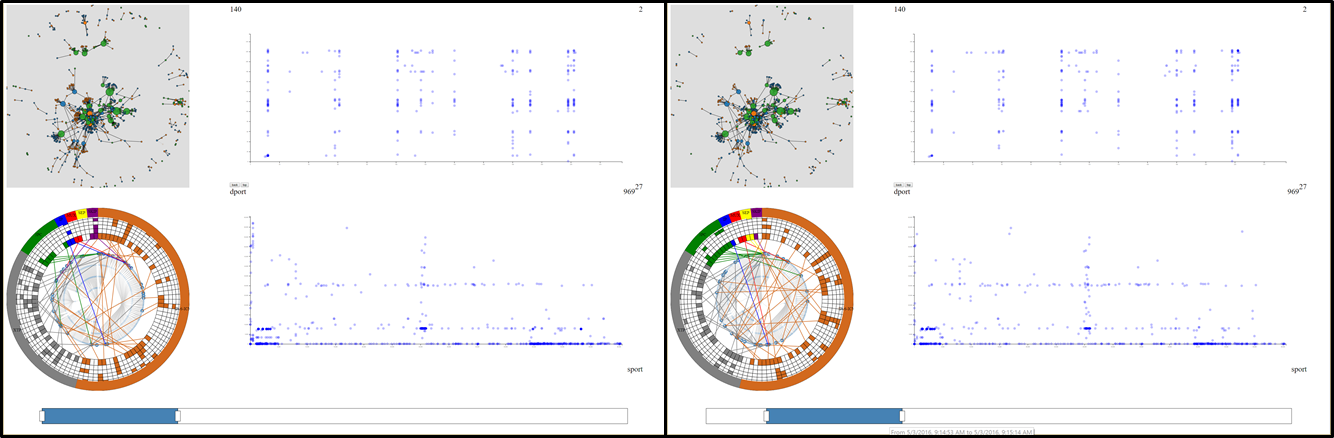


图4.1 对所有视图进行时间筛选

(c)

(b)

(a)

(d)

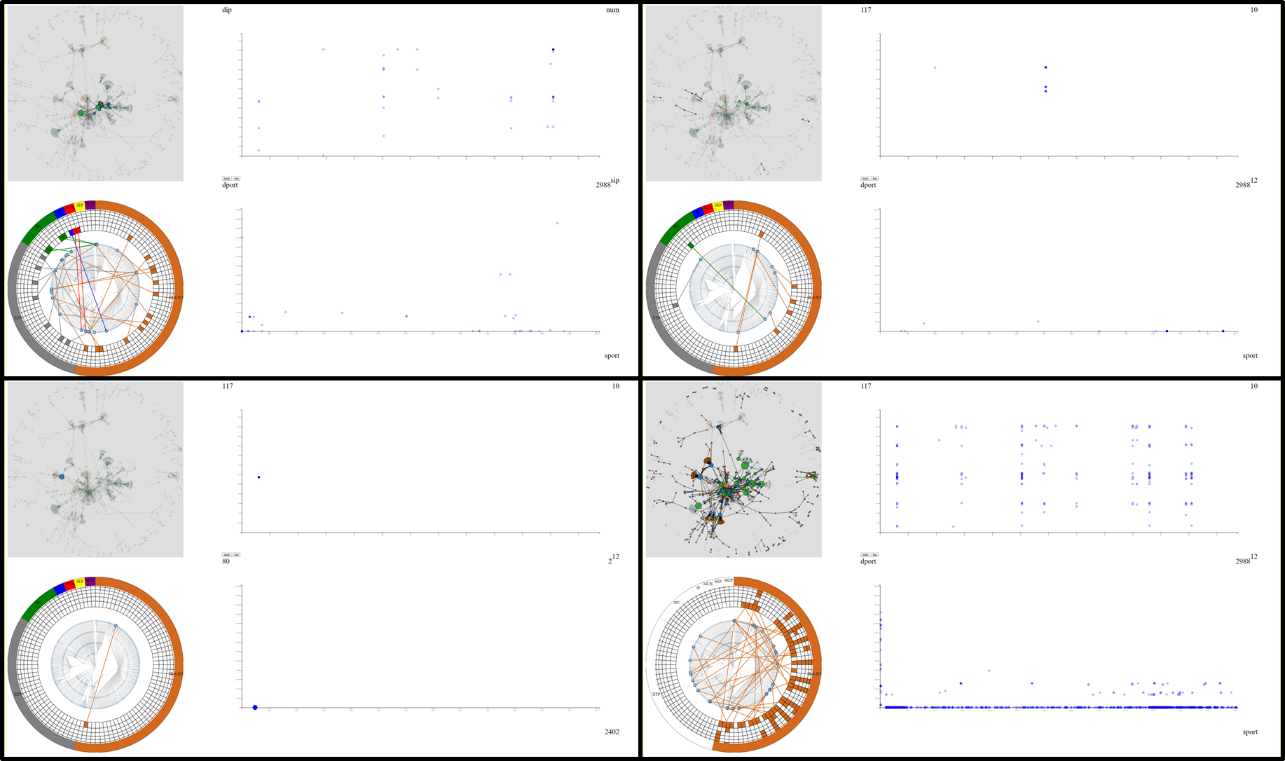


图4.2 在不同的视图进行条件筛选时的联动效果：(a) 筛选某节点及其直接连接的节点；(b) 筛选某网段的节点；(c) 筛选使用特定端口进行通信的节点；(d) 筛选使用特定通信协议的节点。

1. 结论与项目后续

本系统最初目标在于通过多视图共同工作，多角度、多维度地分析网络监控数据，并期望可以做到发现甚至预警可能的网络恶意攻击。由于我们所使用的上海市城域网数据的时间跨度从2016年5月3日上午9点14分44秒到同日9点16分16秒，总共3000条数据，在短短一分半内没有发生网络异常，所以，在这一方面的成效需要后期继续研究。另一方面，通过这些数据，我们已经可以十分清楚简单地看出网络中的IP行为模式、端口分布与协议使用情况，理论上可以快速发现网络的异常状态。

本系统使用JavaScript实现，使用了d3js数据可视化库。由于JavaScript语言性能不及其它编译型底层语言，所以对于更大的数据量可能会有性能瓶颈。当然，JavaScript语言的特性可以让我们在任何浏览器上运行这一系统，这是优势之一。后续若数据量持续增大，可能会将系统用C++与OpenGL等图形库重新实现这个系统。这样也便于我们继续改进系统并增加新功能，如实时对动态网络进行布局等。高性能的实现也有助于实时分析数据，及时预警可能的网络攻击。

参考文献

[1] Khanh Dang T, Tri Dang T. A survey on security visualization techniques for web information systems[J]. International Journal of Web Information Systems, 2013, 9(1): 6-31.

[2] Girardin L. An Eye on Network Intruder-Administrator Shootouts[C]//Workshop on Intrusion Detection and Network Monitoring. 1999: 19-28.

[3] Krasser S, Conti G, Grizzard J, et al. Real-time and forensic network data analysis using animated and coordinated visualization[C]//Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop. IEEE, 2005: 42-49.

[4] Bach B, Pietriga E, Fekete J D. GraphDiaries: animated transitions andtemporal navigation for dynamic networks[J]. IEEE transactions on visualization and computer graphics, 2014, 20(5): 740-754.

[5] Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks[J]. ICWSM, 2009, 8: 361-362.

[6] Jacomy M, Venturini T, Heymann S, et al. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software[J]. PloS one, 2014, 9(6): e98679.

[7] Martin S, Brown W M, Klavans R, et al. OpenOrd: an open-source toolbox for large graph layout[C]//IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2011: 786806-786806-11.

[8] Hu Y. Efficient, high-quality force-directed graph drawing[J]. Mathematica Journal, 2005, 10(1): 37-71.

谢辞

首先感谢我的指导教师董笑菊老师，在项目中给我提供了包括理论、技术方面的帮助。董笑菊老师也帮助我们获取了一份真实的宝贵数据。感谢与我共同完成此系统的丁天琛同学，在项目中与我一起讨论，给我提出了许多宝贵的意见，并且与我一同修复系统中的错误。丁天琛同学也在我遇到技术困难时耐心地提供给我许多帮助，让我对JavaScript这门语言的理解与应用更加深入。同时感谢上海市网络与信息安全应急管理事务中心提供的上海市城域网网络安全监测数据，这份数据详细且真实，提供了很多有效信息，对项目的进展有着莫大的帮助。最后感谢其他所有在项目中帮助过我的同学们。