

基于可视分析的网络异常检测系统

张浩城, 吴晓洁, 唐翔, 舒润萱, 丁天琛, 董笑菊

(上海交通大学电子信息与电气工程学院, 上海 200240)

摘要: 在网络安全数据的规模和复杂度不断攀升的今天, 传统可视化方法已不再适用。现有的网络安全可视化系统和方法仍具有缺陷, 它们无法对网络安全数据进行时序变化上的展示, 在信息展示的完备性和用户交互性上表现较差。针对现有方法的不足, 设计了一种多视图联动的网络安全可视化系统, 将力导向模型和阶段式动画相结合, 展示网络静态状态和动态变化, 提供协议、IP 段、端口的展示与筛选功能, 为使用者展示全面丰富的网络数据。

关键词: 信息可视化; 网络安全可视化; 可视化系统; 交互

中图分类号: TP393.08

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2018015

System detecting network anomaly with visualization techniques

ZHANG Haocheng, WU Xiaojie, TANG Xiang, SHU Runxuan, DING Tianchen, DONG Xiaojie

School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200240, China

Abstract: With the fast development of information technology and computer network, the scale and complexity of network security data grows rapidly. Traditional visualization techniques are no longer suitable. In addition, it designs interactive functions based on the feature of network security analysis, in order to assist the network security analyst. The existing approaches for network security visualization have some defects, which fail to provide a good indication of network security data in terms of timing and also fail to display information completely and realize user-friendly interaction. A multi-view network security visualization system was proposed, which provided the analysts of both the static status and dynamic changes of the network by combining the force-oriented model and the staged animation. It provides comprehensive information with display and filter of protocols, IP segment and port. The system on Shanghai Network database were evaluated.

Key words: information visualization, network security visualization, visualization system, interaction

1 引言

随着互联网的不断普及与网络通信技术的不断进步, 越来越多的网络应用应运而生。从最初的计算机网络只用于发送电子邮件或共享设备,

到如今的银行系统、行政管理系统甚至军事系统纷纷实现电子信息化管理, 人们生活世界中的各个方面正逐渐通过互联网紧密地联系在一起。然而, 互联网应用的高速发展, 不可避免地导致各种网络安全隐患和漏洞。针对这些漏洞的攻击,

收稿日期: 2017-12-26; 修回日期: 2018-01-30

通信作者: 董笑菊, xjdong@sjtu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61472238, No.61772336, No.61572318)

Foundation Item: The National Nature Science Foundation of China (No.61472238, No.61772336, No.61572318)

可能造成个人隐私、财产的损失,以及导致重要公共设施体系的瘫痪、国家级别机关的损害,后果极其严重。因此,网络安全已成为一个不可忽视的问题。

在网络安全领域中,预防与探查网络攻击的一个主要方式为分析监控设备产生的日志数据记录。在日益复杂的网络环境下,数据与信息呈爆炸式增长,对海量日志数据的分析提出了更大的挑战。首先,需要分析的日志数据来源多,构成相异,数据规模成倍增长,同时日志文件多存在记录不完整、缺乏必要字段等问题,日志的分析工作已然成为一大负担;其次,分析人员往往局限于局部异常,难以掌握宏观的网络态势;其三,网络攻击技术在不断迭代更新,出现了许多新型的网络攻击手段,攻击的复杂度也越来越高,导致传统方法不再适用。

近年来,网络安全自动化检测技术不断发展,减少了大量人力开支的同时也存在若干问题。一方面,训练数据规模带来了巨大的存储成本以及时间成本;另一方面,复杂的网络环境和攻击方式的隐蔽性导致不少漏报误报。因此,需要将人员参与分析的过程中,做出更准确的判断。

在此过程中,如何将海量繁杂的数据进行体系化、结构化表示是一个关键性问题,网络安全可视化这一交叉研究领域应运而生。运用网络安全可视化技术将抽象繁杂的数据映射为具象简单的图表,辅之以友好易用的交互功能,使网络安全分析人员可以借助可视化系统直观地分析数据中的联系与异常,从而更加高效准确地发现数据中隐含的信息。这样,便进一步提高了分析人员对繁杂日志数据的处理能力,更容易感知、分析和解决网络安全问题,从而大大促进网络安全保障工作。将网络安全维护与可视化结合,犹如将人类视觉观察与程序判断结合;人类视觉观察分析有所局限,程序判断的结果也未必可靠直观。但二者的结合弥补了各自的不足之处,开创出数据识别的新思路。

网络安全可视化还是一个非常新的领域^[1,2]。它研究的主要内容,如何通过对已有的通用型可视化方法进行调整,使之可以运用在网络安全数

据的可视化工作上。由于用户的认知能力、偏好不尽相同,无法完全预测最终用户将如何解读其设计。但是通过对用户需求和认知能力的调研,可以确定适当的展示内容和相应设计。网络安全可视化将设计过程以用户为核心,精心设计人机交互功能,尽可能直观、完整地将信息传递给用户。

在这一过程中,通常需要围绕3个对象进行:网络节点的拓扑结构、节点间的通信情况以及节点内端口的通信情况。针对这3个对象,需要解决以下4个问题:1)如何在保持用户心理印象的同时展示网络拓扑结构的动态变化;2)如何将网络结构与通信情况紧密结合又使视图不至于复杂;3)在大量节点存在的情况下端口间的收发数据;4)在直接应用图布局算法不够理想的情况下,如何结合网络拓扑数据的特性改进。

本文就IP地址之间的联系与通信状态数据,提出了一种可视化数据分析方案。从信息可视化方面着手,通过多种视图对通信数据进行多方面、多角度的表征;同时为了便于可视化的实时分析,将这些视图进行联动,从而形成一个多维度交互的数据可视化系统。系统由以下3个视图构成:1)基于阶段式动画的动态网络;2)包含IP树和协议展示圈的协议圆周图;3)源-目的IP、端口散点图。基于阶段式动画的动态网络,在静态时展示各个网络节点之间的通信情况,包括节点通信相对密度,节点流量大小、输入输出情况等;在动态时展示各个连接的变化过程,从而有效地展现拓扑结构的动态变化。协议圆周图实时展示当前网络的协议分布情况、通信时间分布等信息以展现拓扑结构上的通信变化。通信双方IP、端口散点图展示当前网络所有连接的IP、端口情况,通过交互详细表现IP的异常情况。这3个视图使用同一个时间轴对数据进行筛选,相互之间同步联动,从而通过逐层筛选可以最终锁定问题关键所在的各种信息。在主视图的布局上,笔者在众多力导向算法中选择了最合适的FR(Fruchterman-Reingold)算法,并针对网络数据的特点做了全局优化,动画上采用了精简后的阶段式动画;在2个辅视图中,协议圆周图内部的IP子树弥补了主视图中无法观察节点之间、IP段

关系的缺漏,外部的协议展示圈则提供该节点在一段时间的通信情况,给用户一些综合考虑的空间。IP 端口散点图展示了 IP 段之间、IP 之间和端口之间的通信流量情况,采用一些针对性的优化措施使整体更加易用。最后,将系统整体串联起来的多视图联动功能和时间轴筛选功能,则通过将时间序列缩略图和时间轴相结合,使整个系统可以满足从概览整体到探究细节的所有需求。

2 相关工作介绍

2.1 可视化在网络安全中的应用

2.1.1 网络监控

对主机的网络监控即对 IP 地址的监控。IP 地址有分段的特性,其中,当前使用最广的 IPv4 规定地址长度为 32 bit,通常写作 A.B.C.D,4 个用小点分开的十进制数。由于 IP 地址众多,如何进行取舍以达到比较好的可视化效果一直是学者努力的方向。IP 矩阵^[3]监控的是某一特定 AB 网段下的主机事件,通过二维坐标定位。而针对端口的监控(如 PortVis^[4])同样使用矩阵和颜色来映射所有的端口的流量情况。考虑到常用的端口并不多,可以将常用端口和非常用端口做区分,PortMatrix^[5]将端口号分为 4 类,不常用的系统动态分配的端口被缩略展示。

2.1.2 异常检测

网络异常的范围很广,包括流量异常、越权访问、设备失效等,其中,流量是异常检测的主要监控信息。对于流量异常的检测方法,2.1.1 节中基于 IP 地址和端口的监控方案就可以胜任。此外也有许多其他的流量监控方案,如文献[5]采用基于热力图的流量监控可视化系统。而 Abdullah^[6]采用一个基于时序的端口流量分组聚合分析的堆叠图。

除了流量异常之外,主机、设备的突然失效也值得注意。节点连接图是监控网络连接情况的常用可视化方案,具体实现时又分为主机位置固定和动态布局 2 种策略。

2.2 动态网络中的时间编码

对于动态网络的表示方法,最常见的就是网格图。任何静态网络可视化系统都可以用于

在任意时间点生成当前网络的快照,这些快照可以使用表格进行显示^[7~9]。

2 个或者更多网络之间的差异图可以直接比较网络的时间步长,但单独一个差异图不足以探索长时间步长的序列^[10~13]。颜色也被用于展示长期的变化,如文献[14]中用颜色表示节点和连接在一个对整个网络的简单聚合图中的首次出现。然而,这种方法很难对其他时间度量进行编码。文献[15,16]认为值通常是聚合的,而目标是提供一个单一的度量来描述整个动态变化,因此,省略关于各个时间点的信息。然而,单个图像对时间变化显示编码的增强同时也增加了网络表示的视觉复杂度^[17]。

针对子图在特定时间步长内的连通性以及高维属性显示的问题,本系统提供了一致的视觉界面,结合了网格图表的优势,提供一个总览,帮助用户将时间维度中的表示方式与差异图像的显式编码以及动画转换的灵活性相结合。

2.3 动画与时间导航

动画作为一种向用户展示界面变化的手段已经在心理学^[18,19]、人机交互^[20]和信息可视化^[21]中得到了广泛的研究,且已被证明在统计信息图像之间和多变量的散点图之间^[22]的切换以及突出显示文本档的历史修改记录^[23]中具有很好的效果。Heer 等^[22]也偏向于过程更长的动画交互效果。Chevalier 等^[20]放弃分阶段转换从而缩短转换时间,更说明动画能够有效利用短期记忆。

分阶段的动画也被用于可视化动态树^[24,25],如在 SpaceTree 中扩展子树时和探索树随时间的变化之时^[26]。DOITrees^[27]也因此使用动画转换,不同的是它并行运行不同类型的动画(子树扩展、布局适配)。

Eades 等^[28]第一次提出了动画可以在浏览集群图像时提高对变化的理解程度。Friedrich 等^[29,30]提出了用于可视化动态网络中变化的若干转化步骤。Brandes^[31]则提出三阶段转换,但在实际转换过程中很难跟踪具体的变化。

综上所述,动画的许多重要方面尚未得到充分研究:步长、分阶段、阶段的排序,转换图形渲染方式等。需要更多涉及更高级任务、涵盖更大数据集的实验来更好地进行探索并找到有效支

持动态网络探索的交互技术。

2.4 动态网络布局算法

动态网络可视化的另一个关键是如何在每个时间步骤中布置网络。Gephi 采用的迭代布局求解器在用户更改显示的时间范围时可以提供很好的连续性^[32]。虽然这种方法在步骤之间提供了一些连续性,但是不够稳定,重新访问之前某个时间段内网络状态的任务很难完成。为了避免这些变化,TempoVis 根据上一个时间步长的布局、步骤间节点颜色和位置的线性插值创建一个新的布局^[33]。

虽然这种方法可以随着时间推移对布局进行更大的改变,但 Eades 等认为它更利于全局布局的稳定性,从而更好地保护用户的心理图^[28]。Purchase 和 Samra 的研究^[34]表明,全局稳定或局部优化应该是首选,而不是中间解决方案。Archambault 等^[35]比较了网格图和动画在保存心理图上的优劣,表明稳定的布局不会提高性能。然而,在后来的研究中,Archambault 和 Purchase^[36]表明,稳定的布局能够更好地支持网络探索,这一发现得到了 Ghani 等另一项研究的证实。

这些实验结果表明,具体的布局策略应该根据用户的具体情况进行权衡。在所有时间步骤中最稳定的布局可能不适用于这些步骤中的任一步。针对每个时间步长优化之后的布局在步骤之间可能非常不稳定。关于布局算法的决策以及要使用的动画类型高度依赖于用户执行的实际任务。虽然布局与本系统中的其他功能是正交的,但是所选择的布局会强烈地影响其适当的使用,详见 3.2.1 节。

3 可视化系统设计与实现

3.1 系统总览

3.1.1 数据特征

本系统针对的数据集为含有源、目的 IP、源、目的端口,传输协议等维度,以传输时间为轴的网络流量数据。

3.1.2 系统简介

由于数据量较大,本系统针对引言中提出的 4 个问题采用 3 个视图解决,以展示数据集各个维度的信息,如图 1 所示(见彩插页 9)。

为有效展现拓扑结构的动态变化,主视图采用基于阶段式动画的力导向动态网络。静态时展示了各个网络节点之间的通信情况,包括节点通信相对密度,节点流量大小、输入输出情况等,当视图出现变化时,使用阶段式动画高亮新增的和消失的连接。其中,对于拓扑结构的布局,本系统以 Fruchterman-Reingold 算法为基础,针对网络数据的特点做了全局优化,使整体布局更加分散、易读。

为将拓扑结构以及通信变化联系起来,图 1(b)采用协议圆周图,具体展示了当前网络的协议分布情况。内圈各个节点按照网段分布,并有线段指引到外圈。外圈切向为协议类别,径向为协议时间分布情况,

通过节点与内圈的连线可以展示该节点的通信协议类别、通信时间分布等信息。

因普通 IP 矩阵在点过多时无法清晰有效地显示端口数据而传达的信息量不够,图 1(c)设计采用了源-目的 IP 散点图和一个源-目的端口散

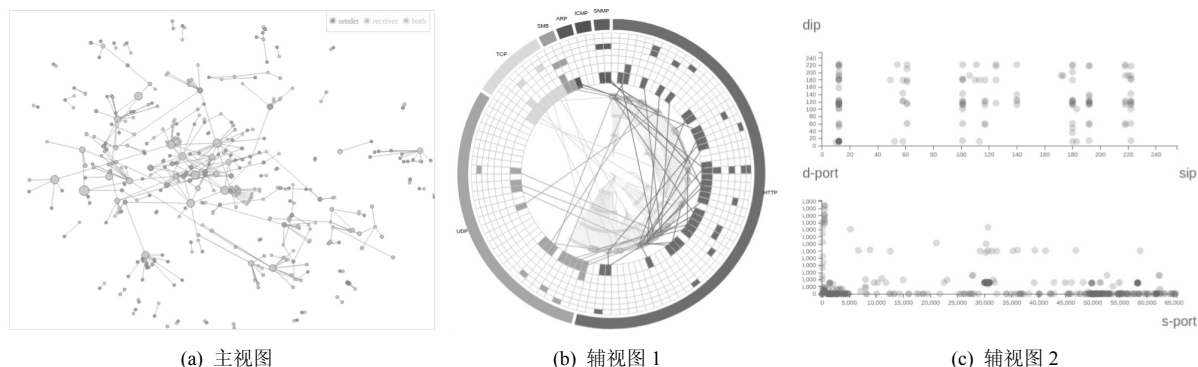


图1 系统总览

点图,展示了当前网络所有连接的 IP、端口情况,并且使用颜色的深浅表示流量大小,颜色深的点流量较大。

3 个视图均使用一个时间轴对数据进行筛选,时间轴滑块可以调整区间大小从而控制展示信息量,时间轴具有播放和暂停功能,方便浏览网络动态变化。

同时,各个视图均有交互功能且互相之间都有联动,通过一步步筛选可以最终锁定问题关键所在的各种信息。

3.2 基于阶段式动画的力导向动态网络

3.2.1 动态网络布局算法

1) 需求

对于一个网络的监控,即包括主机、端口、流量和协议的监控,将对端口和协议的监控置于辅视图中展示,因此,动态网络这个视图中需要展示的是主机和通信流量的情况,2.1.1 节中已经提到了几种基于主机监控的可视化系统,然而不仅要体现主机的流量,还要展示出主机之间的通信流量大小,并且不能将主机简单地分成诸如内外网等几类。因此,笔者决定采用节点连接图进行展示,考虑到网络可能分为几个集群,最终采用了主机位置动态布局的方法,使用了力导向模型。考虑到大型网络中难以对主机进行简单的拓扑压缩,笔者并未对网络进行简化,而是直接展示实际情况,这样可以避免信息的遗漏。

2) 力导向模型

力导向模型的基本思想是,将图形用一个物理系统模拟,每个节点都受到其他节点的拉力和斥力,在这种相互作用力之下整个系统达到平衡,节点不再震荡保持静止时,就获得了合理的布局。

具体拉力和斥力的选择,Eades 在前人的基础上提出了第一个力导向模型,他将网络中的节点视作刚性的环,链接视作圆环间的弹簧,整个网络就是一个弹簧系统。但是在实现上 Eades 并没有遵照胡克定律,而是采用了自己的弹簧力公式。他的另一个不同于物理现实的地方在于力的作用方式:所有节点两两之间都有斥力,而拉力只存在于相邻节点之间。这样做降低了算法的复杂度,使计算拉力的复杂度降到了 $O(E)$,虽然计算斥力的复杂度依旧是 $O(N^2)$ 。

Eades 没有采用胡克定律是因为它将节点之间的距离设置太大,Kamada 和 Kawai^[37]则是改进了胡克定律,提出了 KK 模型。他们同样将图形视作弹簧系统,但是加入了一个理想距离的概念,2 个节点的理想距离与它们之间的最短路径长度相关。Kamada 和 Kawai 将图形的绘制视作一个降低整个弹簧系统总能量的过程,通过最小化系统能量,节点之间的距离会最接近与它的理想值。他们将系统能量定义为

$$E = \sum \sum \frac{1}{2} k_{ij} (|p_i - p_j| - l_{ij})^2$$

其中, p_i 为节点 v_i 对应刚性圆环的位置, k_{ij} 为 p_i 和 p_j 之间的弹力系数, l_{ij} 为节点 v_i 和 v_j 之间的理想距离。

$$l_{ij} = \frac{L_0}{\max_{i < j} (d_{ij})} d_{ij}$$

系统通过解偏微分方程获取节点的新位置,并且仅通过移动使系统能量减少的节点来减少整体能量。这一过程重复进行直到系统能量达到预设阈值。KK 模型跟 Eades 模型很大的一个不同在于一次只移动一个节点,这使每次只要计算一个节点对系统能量的贡献,将算法的复杂度降到了 $O(N)$ 。

3) Fruchterman-Reingold 算法

Fruchterman 和 Reingold 对 Eades 的模型进行了改进。他们受到自然中电子斥力的启发,将模型中节点之间的斥力用电子斥力进行模拟,定义 2 个节点之间的电子斥力为

$$F = \frac{G^2}{|p_i - p_j|}$$

其中, G 为电子力常数,可以看出 2 个节点的距离越小则相互之间斥力越大,这样就解决了节点重叠的问题。当然运动并未被完全模拟,力转化的是速度而不是加速度,这是因为转化加速度达到的是一个动态的平衡,而实际追求的是一个画面上静态的平衡。

每一个循环内都有 3 步,计算每个节点受到的引力,计算每个节点受到的斥力,然后通过“温度”限制总的移动距离。这个“温度”是受到 Davidson 和 Harel^[38]的启发引入冷却函数模拟退

火过程的控制变量。每一次迭代都将减小节点的最大移动距离，这样可以使系统能量快速减小，从而减少布局时间。

将 k 设置为节点之间的理想距离，即

$$k = C \sqrt{\left(\frac{\text{area}}{\text{number of vertices}} \right)} \quad (1)$$

其中，number of vertices 为节点数量，area 为面积， C 需要通过经验来验证，式(1)基本的思想为希望所有的点在整个区域内广泛分布，那么 k 就是理想节点周围的空旷区域的半径。如果用 f_a 表示引力， f_r 表示斥力， d 表示 2 个节点之间的距离，那么

$$f_a(d) = \frac{d^2}{k}$$

$$f_r(d) = -\frac{k^2}{d}$$

2 个力随距离的变化如图 2 所示^[39]，可以看到当距离为理想距离时，二者正好互相抵消，引力和斥力函数是经过一些实验得来的，有一些其他的选择，但是很容易囿于局部最小值。

另外，冷却函数 $\text{cool}(t) = \alpha t$ ，其中， α 为冷却系数，经验上设置为 0.95 可以达到比较好的效果。

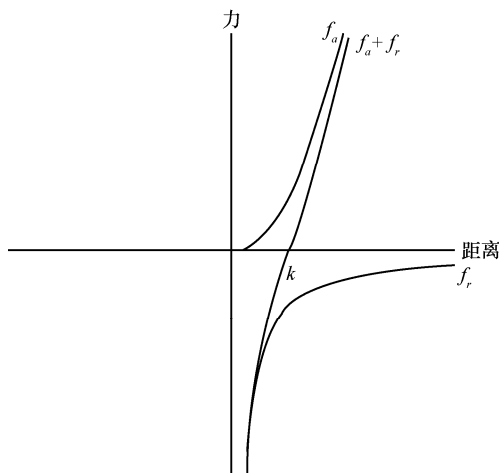


图2 引力斥力和 2 个节点距离函数

4) 针对网络安全数据特性的改进

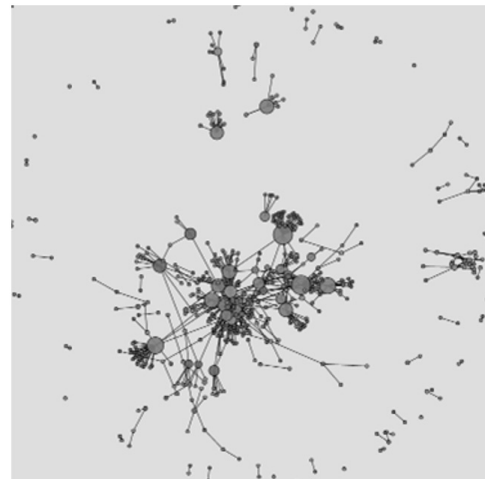
在使用 FR 算法布局大型网络时发现了一些其他的问题，因为数据上显示出网络通信的一些特点：几个流量特别大的节点周围存在很多小节点，而流量特别大的节点之间通信也很频繁，还有一些节点是脱离这些大节点而独自存在的小点

对。这就导致这些小点被远远地排斥在外只留下中心节点，在展示区域有限的情况下将导致节点分布不均。为了解决这个问题，笔者给节点附加一个拉力权重 W_a 和一个斥力权重 W_r 。

$$W_a = k_a \left(1 - \frac{D_i}{D_{\max}} \right)$$

$$W_r = k_r \left(1 - \frac{D_i}{D_{\max}} \right)$$

其中， k_a 和 k_r 分别为常数， D_i 为节点 p_i 的总连接数， $D_{\max}$ 为所有节点的最大连接数。通过减少大流量节点之间的引力，加大它们之间的斥力，使大节点之间较普通节点更为分散。同时，减小孤立点对共同受到的中心节点的斥力，缩小中空区域，使整体布局更加分散、易读。图 3（见彩插页 10）是改进前后的布局变化。



(a) 算法改进前



(b) 算法改进后

图3 FR 算法对布局的改进

3.2.2 阶段式动画

1) 需求

网络安全数据是一个具有时序性的数据，2.3 节中比较了网格图和动画在展示探索随时间变化图像的差异时的表现，虽然网格图有准确、容易绘制等优点，但考虑到本系统是运行在一个大小受限的页面之上而复杂网络的静态图像本来就已经很大，若采用网格图来展示变化，将会丢失很多数据上的细节，并且交互起来会非常不方便。因此本文决定用动画展示。使用动画展示图形的差异也有很多设计方案可以选择，涉及节点和边的变化顺序和变化方式等。最简单的就是平行变换，将前后 2 张图的所有差异进行一次性的转化，这种变换方式的优点在于非常节省播放动画所需要的时间，从而达到快速浏览的目的。然而这在本系统中是不适用的，首先，复杂网络由于本身含有的元素众多，其产生的各种变化也非常多，过快地展示所有的变化给用户带来认知负担，反而降低了可视化的效果。其次，对网络的监控不仅仅包括对当前网络主机通信情况的掌握，还包括网络之中产生的诸多变化，如突然增多的流量、突然失效的设备等。因此，对网络变化的展示与对当前网络状态的展示同样重要，比较之下，阶段式动画是一个很好的选择，通过合理地分配图像转化阶段，不仅可以使网络状态的变化跃然眼前，还可以维持用户的心理图，降低用户的认知负担。

2) 实现与改进

将动画分为 3 个阶段：

- ① 失效节点和边的渐出与高亮；
- ② 已有节点和边的换位；
- ③ 新增节点和边的渐进与高亮。

笔者认为这样分配动画较为合理，尤其对于动态网络来说非常适用。新增节点和边可以展示出突然增高的流量，或者新出现的扫描行为；已有节点和边的换位展示的是已有链接的流量变化；失效节点和边可以展示出突然失效的设备。这样分开的展示可以使用户不遗漏任何信息。

考虑到系统的实际情况，笔者最终没有采用局部优化的布局方式，理由为：局部优化下的布局不具有整体性，对于存疑部分的网络状况有反复查看的需要，而局部优化的算法布局会受到历史操作记录的影响。因此，最终选用了全局优化的布局，前后视图节点的位置并不会发生改变，这就不再需要动画的第二步，并且也能够给用户提供良好的前后连贯性，保持心理图。

最终的动画效果如图 4 所示（见彩插页 11），分为两步，失效的节点和边采用蓝色短暂持续的高亮边缘渐出，短暂的间隔之后新增的节点和边采用橙色短暂持续的高亮边缘渐进。

3.3 协议圆周图

3.3.1 简述

本系统中使用的圆周图是对文献[40]中提到的雷达图的实现与改进。笔者借用了其内部展现



(a) 蓝色高亮显示渐出



(b) 橙色高亮显示渐进

图4 阶段式动画

网络结构，外部展现通信情况的思想，绘制了一个监控网络协议的圆周图。整个图分为2个部分，内部为IP树，外部为通信协议圈，如图5所示（见彩插页12）。

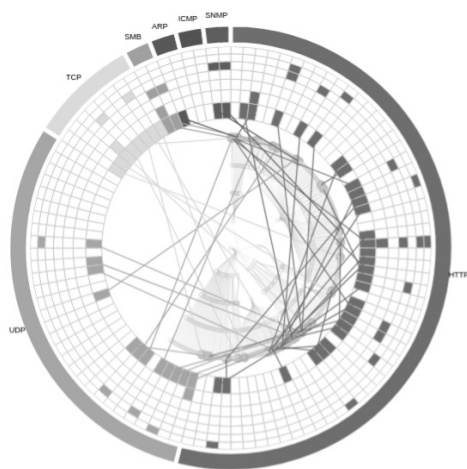


图5 协议圆周图

3.3.2 IP树

圈的内部是由IP网段组成的树，每个叶子节点都是一个IP地址，除了根节点外，其他节点都代表了某个网段，数据结构如图6所示。

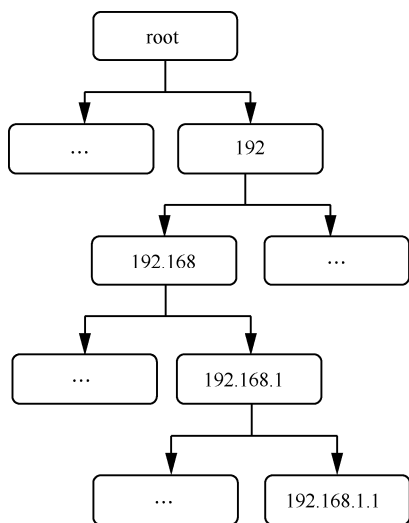


图6 IP树数据结构

这样分配IP地址的优点在于，各个主机是否处在同一网段一目了然。由于主视图采用了节点连接图，并且连接展示的是节点之间的通信情况，没办法展示它们之间IP地址的关联，所以在这里进行补充。将IP组织成这样的网段树，拥有相同父节点的IP都处于同一网段内，即使数据量扩

大，也能够很快识别出各个IP之间的位置关系，为用户的判断提供帮助。

3.3.3 协议展示图

外圈的协议展示圈是一个100×9的环形网格图，用于展示当前网络通信的协议以及时间信息。圆环的切向为x轴，表示的是协议的类型，最外层是具体协议类型示意，径向为y轴，表示的是通信的时间，越是靠内的方格表示通信发生的时间越接近当前选择时间范围的终止时间，即最靠近内圈的方格表示该通信发生的时间为所选时间轴的末尾；越是靠外的方格表示通信发生的时间越接近当前选择时间范围的起始时间，即最靠近最外圈的方格表示该通信发生的时间为所选时间轴的起始。通过展示通信的时间，可以发现某个IP的历史行为而不用通过切换已选择的时间区间长度进行观察，从而帮助分析人员更快找到异常行为。

内外圈通过与协议提示颜色相同的连线将IP树中的节点与外圈方格相连，这样，颜色的一致性也方便观察某些节点通信协议的规律。

需要注意的是，由于展示的空间有限，当数据规模过大时可能无法容纳，为了解决这个问题将展示的数据进行了预处理。受到外圈规模限制，传入的数据较多时预处理程序将根据通信量、发生时间、该协议的占用情况等自动筛选出一部分数据进行展示，因此一些数据可能会被遗漏。若是不想程序进行筛选，可以将时间筛选长度减小，从而减少需要展示的数据。

3.3.4 交互

笔者通过给线段和方格赋予不同的颜色来区分通信的协议，这么做有一个缺陷，即数据量增多时，画面中会出现大量的连线和方格，降低可视效果。同时，多视图联动的意义在于可以对数据进行筛选，因此本视图中的交互也是必不可少的。

首先，为了让使用者能够在信息繁多的情况下定位到自己需要的IP、通信情况等数据，本文在对线段、节点、方块的交互上添加了以下功能。

1) 当用户的鼠标移经某个IP地址的节点时，将该节点以及与该节点相连的通信信息高亮，并

同时淡化其他所有数据。

2) 当用户的鼠标移经某个连线或方格时, 仅高亮该条通信的信息而淡化其他所有数据。

3) 考虑到用户可能希望保留若干条自己已经看过的数据, 保持它们的高亮, 也增加了对于鼠标单击事件的响应: 当用户鼠标单击图上的数据时, 锁定该数据的高亮, 鼠标移开也不会被淡化, 同样按住 **ctrl** 键单击为多选, **shift** 单击为取消选择。当用户选中需要的数据后, 其他用户不关心的“无用”数据都将淡化融入背景中, 只显示需要的数据。

本视图主要展示了通信数据中的协议信息, 同时具有对通信数据的协议筛选功能。通过单击最外圈的协议标注圆环, 可以通过单击选中某个协议, 同样也是按住 **ctrl** 单击进行多选, 按住 **shift** 单击取消选择, 如图 7 所示 (见彩插页 12)。

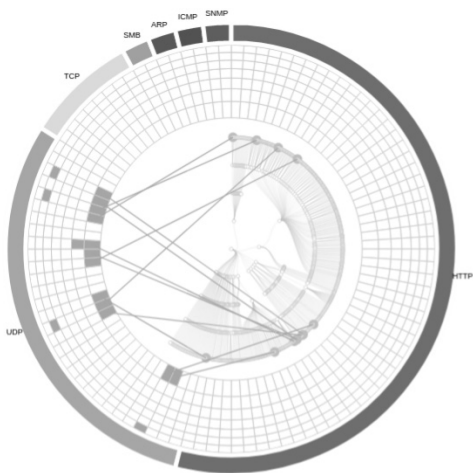


图7 协议筛选

通过筛选出用户需要的协议再加上与其他视图的联动可以帮助发现某些特定的网络异常, 如判断洪泛的类型等。通过过滤噪声, 也能够更好地分析数据特征。

3.4 IP、端口散点图

3.4.1 源-目的 IP 散点图

用户往往并不关心全局的通信情况, 他们大多只想知道特定网段的情况, 而其他网段的 IP 地址及其通信情况与他们无关。因此笔者决定分网段展示散点图, 这样横纵坐标轴的范围就能缩减到 0~255。IP 地址写作 a.b.c.d 这 4 段的话, 散点图的初始界面展示的是 a 段的通信情况, 在 (x, y)

位置有蓝点则说明在以 x 为首段的 IP 地址和以 y 为首段的 IP 地址之间存在通信。用户要是想查看特定 a 网段之间的通信数据, 在单击之后可以看到 b 网段的对应情况, 以此类推, 最终可以看到任意 IP 网段或地址之间的通信情况。

仅仅这样处理只是解决了点过多的问题, 但是整张图的信息量太少了, 只是知道此处有通信, 流量多少却毫无概念。为了完善这一视图, 使之达到更好的效果, 将每个点的流量情况用透明度进行反馈, 展示的是 IP 网段之间的通信时, 颜色深浅是统计好的数据。这样, 用户能够轻松在散点图上获知网段流量, 而且流量热点之处非常醒目, 如图 8 所示 (见彩插页 12)。

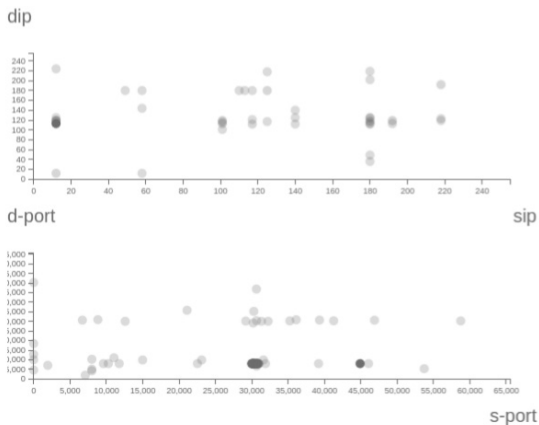


图8 IP 散点图

3.4.2 源-目的端口散点图

计算机网络共有 65 535 个端口, 因此将其一次性展示也不会过于麻烦, 笔者依旧对端口流量进行了统计, 使用颜色的深浅表示流量的大小。不过端口需要注意的一点是, 平时经常使用的端口在 10000 以下, 比较有监听意义的大多在 1000 以下, 因此需要更多地关注较小的端口号, 而那些超过 10000 的端口被认为是系统自动分配的一些随机端口, 可以少一点关注。

3.5 多视图联动

3.5.1 各视图交互方式

主视图提供的交互方式包括: 用户自定义筛选节点, 缩放、拖动平移画面。其中, 筛选节点会对另外 2 个视图产生影响。

协议圆周图提供的交互方式包括用户自定义筛选某一则通信数据和筛选某些协议的通信数

据,均会对另外2个视图产生影响。

IP/端口散点图提供的交互方式包括用户自定义筛选IP网段内通信和用户自定义筛选对应端口通信,同样均会对另外2个视图产生影响。

3.5.2 多视图的筛选与联动

3个视图同时既有信息展示、传达的功能,又有筛选、探索的功能,不管对哪一个视图进行操作,都将获得统一一个整体的展示,这使信息的获取没有延迟,不需要额外的操作确认信息是否一致,更加便于分析网络状态,检测网络异常。

3.5.3 时间轴

对于一份时序的数据来说,时间轴是一个非常重要的部分,2.2节探讨了一些动态网络中的时间编码问题,时间序列图依旧有非常大的局限性,但是可以将它补充完整。本文最终使用了带有缩略图预览的时间轴,如图9所示(见彩插页12)。

图9中上部为时间轴,具有一个可以调节时间间隔的播放滑块,当前展现的网络状态即为该滑块所确定的起止时间下的网络状态。下部为由缩略图组成的时间序列图,这些给定时间段内的布局很好地展示了网络的变化过程,笔者用橙蓝两色分别表示消失和新增的节点与边,使前后时间片段的网络差异更加明显。当然,时间序列图支持交互,它与时间轴是完全同步的,高亮的那一帧即为距离当前所选最近的一帧。也可以通过单击时间序列图进行跳转。缩略图显示的内容也可以选择,用户可以只看新增的部分或只看消失的部分。

这样的设计主要有2个考虑:一是完全使用时间序列图并不能满足探索细节的需求,而且在数据量大时要显示所有时间序列图较为困难;二是只有时间轴也是不够的,通过播放动画来了解整体的变化有时并不是用户需要的,在任务简单易行需要快速完成的时候,一个缩略图就可以达到很好的效果。综上所述,二者相辅相成,组合在一起提供了非常友好的体验。

4 案例分析

4.1 数据集

本次案例分析使用的数据集为上海市城域网运行安全实时监测元数据样本,数据截取时间为2016年5月,数据经过了范式化处理后记录的数据。

数据总共含有3 000条,每一条包含源、目的IP、源、目的端口,协议和接收时间等信息。协议大体上分为IMAP、SMB、ARP、ICMP、TCP、UDP、HTTP这7类,一些只出现了一两次的协议并未被展示出来。

4.2 网络监控

本系统的第一个应用场景即为对网络状态的实时监控,通过将各种网络监控设备生成的日志文件导入可以将当前设备所捕捉到的网络通信进行可视化处理,对当前网络的运行状况、流量情况、通信协议情况都能很好地掌握。

笔者用上海市城域网运行安全实时监测元数据样本进行模拟分析。数据载入时系统显示的是最初始一段时长间隔的网络状态,可以从历史记

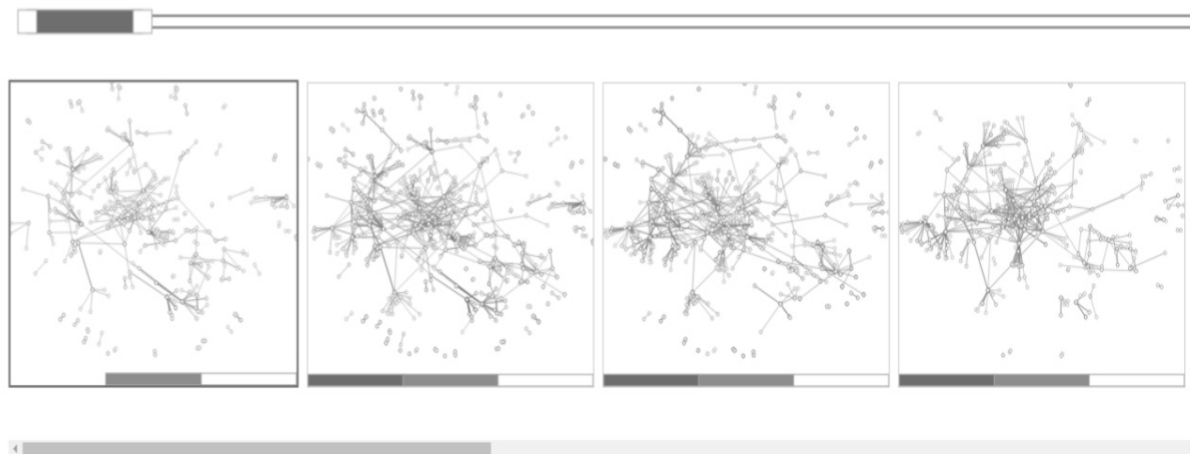


图9 时间轴

录框中知道这是2016年5月3日09时14分44秒到53秒的通信数据。在这数秒之中,从图10(见彩插页13)中可以观察到这个网络的几个特征:首先,图中有十余个半径较大的绿色圆点,这些流量较大的原点很可能是某些服务器或者正在扫描网络的可疑主机;其次,绿色圆点互相之间也有通信,部分甚至在进行了特殊处理之后也相隔很近,这些节点可能是某些邮件服务器,也可能是一些聊天软件的缓存服务器,如果需要进一步地判断,可以单击之后观察分析通信协议。余下的一些两两之间的通信散布在画面的边缘,这些流量较小的节点可以不用太在意,通常是一些正常连接。

除了主视图外,对于辅视图的观察也能补充一些信息,通过协议圆周图可以发现此时网络中的协议分布情况。这一时间段中一共出现了6种主要协议,其中大部分是HTTP。这也符合人们日常使用的情况,绝大部分网页访问都是通过HTTP达成的。如果需要观察某协议的网络通信情况,可以在协议圆周图上进行筛选,这样就可以在主视图中进行进一步的观察。

在IP、端口散点图中,也能获取一些关于网络状态的信息。IP散点图中,源IP的规律非常明显,散点构成的几条竖线在首部为10、58、100、120、180、200、220附近聚集,横向的坐标受空间限制

不太容易观察,按照推测也应该在这些值附近聚集。由此可知,这份数据包含的a网段主要就是这些。端口散点图中有大量深色点聚集在坐标轴附近,将鼠标悬停在点上可以获得具体端口号。简单查看之后发现聚集在坐标轴上的点其源端口或目的端口多为80、135、445等值。同时也能发现聚集在x轴靠右的点较聚集在y轴靠上的点颜色深,这意味着以30000以后的端口号作源端口号的较以之为目的端口号的更多,这也符合大于30000的端口号多是系统自动生成的猜想。大量系统自动生成的端口号向某一固定端口发送报文,这很像是某服务器的大量访问产生的模式特征。

接下来,可以开始观察网络状态随时间变化的情况,首先可以通过时间轴下的缩略图进行概览,有些非常明显的变化有时可以直接从缩略图中获得。如图10中的时间轴缩略图表所示,蓝色的点线为与上一缩略图相比失效的节点和连接,大量失效的连接伴随着的失效节点可以很容易地被发现,直接单击缩略图就可以直接查看当时的网络变化情况。图11(见彩插页14)为图10的缩略图列表中的第三张,其中出现了与同一节点相连的大量连接,而在下一张缩略图中随即消失,单击之后就能切换到该时间片段,如果需要返回当前访问,可以通过右上角的历史记录框返回之前的时间片段。

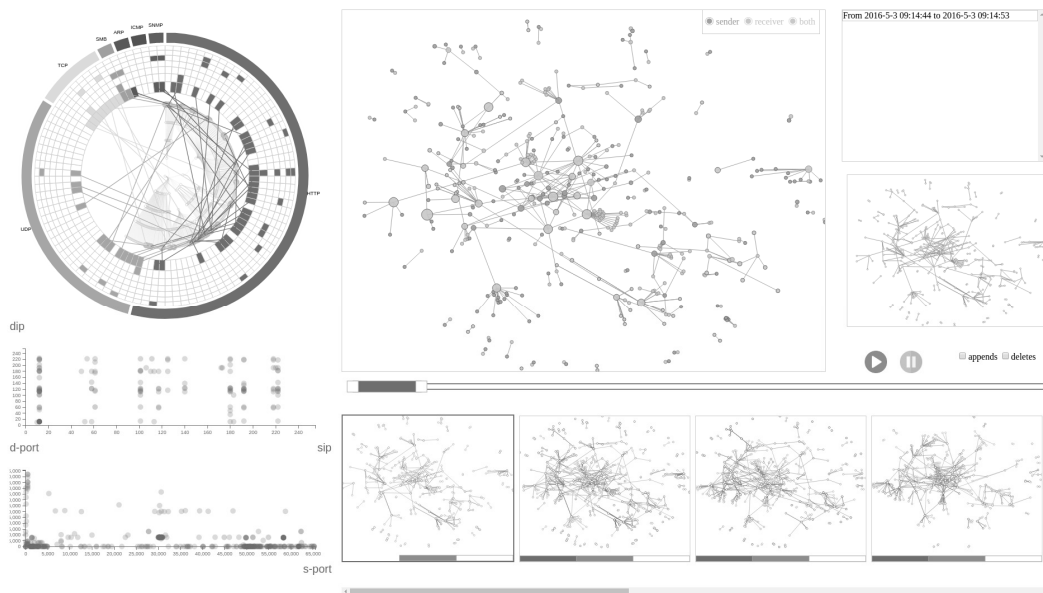


图10 系统初始界面

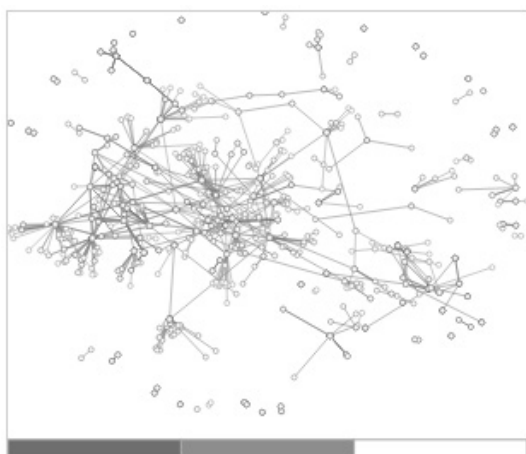


图 11 可疑节点

如果需要系统地浏览网络变化，可以按下播放键，时间轴上的滑块会自动向右移动，一步一步展示网络状态的变化。通过阶段式动画，失效的连接和节点与新增的连接和节点被分开展示，失效主机和需要重视的大量新增连接都得到突出显示。在发现感兴趣的部分时，单击暂停键，再进行进一步的分析。

4.3 异常检测

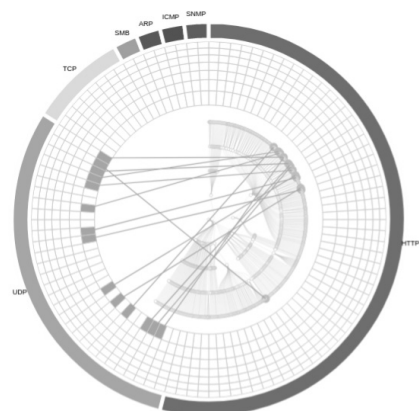
图 10 中有一个绿色节点周围有非常紧密的仅接收信息的主机，由此可以单击该点进行进一步的观察，如图 12(a)所示（见彩插页 14）。单击该绿色节点之后出现了一个非常有意思的现象，通过图 12(b)所示的协议圆周图发现该主机发出的通信全部都是 UDP 的，这对一般的主机并不常见，因为正常的网络访问多是使用 TCP/HTTP，同时从 IP 树中发现大多数通信目的 IP 都非常靠近，这意味着它们 IP 地址的 a、b、c 段都完全相同，即它们很可能处在同一个子网中。

把目光移向另一个辅视图（如图 12(c)所示），IP、端口散点图，在这里看到相关 IP 的分布。散点图中有 2 个点，它们目的 IP 的 a 段完全相同，只是源 IP 的 a 段有差异，可以看到左边的点颜色较深，统计下一阶段的流量较大，而右边的点颜色较浅，流量较小，再注意到图中不仅有一个数据收发方，可以猜测右边的点是由除选中点之外的 2 个绿色节点发出的。通过进一步单击左边这个点，可以一步步探索具体通信发生的源、目的 IP 的分布。锁定这一网络异常涉及的网段和 IP。再观察源、目

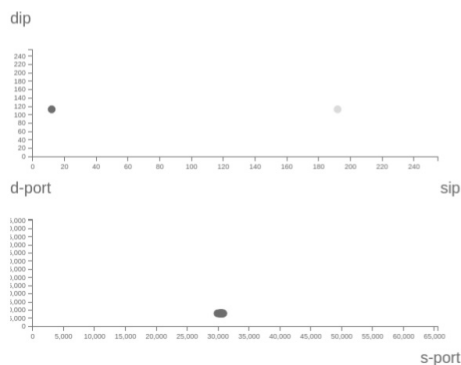
的端口图，发现源端口在 30000 附近密集分布，而目的端口均为 445。



(a) 主视图选中后



(b) 协议全部为 UDP



(c) 端口全部集中在 445

图 12 大量 UDP 445 端口连接

445 端口是一个特殊的端口，Windows 开启这个端口的本意是在局域网中提供共享本地文件、操作打印机等正常功能。然而由于空会话的存在，445 成为了一个经典的病毒入侵的端口，大量发往 445 端口的通信请求很可能是蠕虫病毒正在传播。因此这个端口值得关注。

锁定 UDP 下的这些异常后再确认一下这个

节点有没有“同伙”，即同样在传播蠕虫病毒的主机，这时候取消当前的节点选择，改为只选择 UDP。如图 13 所示（见彩插页 15），再通过查找端口散点图中目的端口为 445 的点就可以确定图中高亮的节点是否是异常点。

5 结束语

在信息技术和计算机网络不断发展的今天，网络安全数据的规模和复杂度不断攀升，传统的可视化系统和方法已不再适用于网络安全大数据。在此背景下，网络安全可视化成为了一个新兴的研究领域，它对现有方法进行改造和重组，使之能

适应网络安全领域，并针对网络安全分析的特点设计交互功能，辅助网络安全分析人员的研究分析工作。现有的网络安全可视化系统和方法仍具有缺陷，它们无法对网络安全数据进行时序变化上的展示，在信息展示的完备性和用户交互性上表现较差。针对现有方法的不足，本文设计了一种多视图联动的网络安全可视化系统，将力导向模型和阶段式动画相结合，如图 14 所示（见彩插页 16），展示网络静态状态和动态变化，提供协议、IP 段、端口的展示与筛选的功能，为用户展示全面丰富的网络数据，实现用户友好的交互功能。本文使用上海市城域网数据集进行实验，通过案

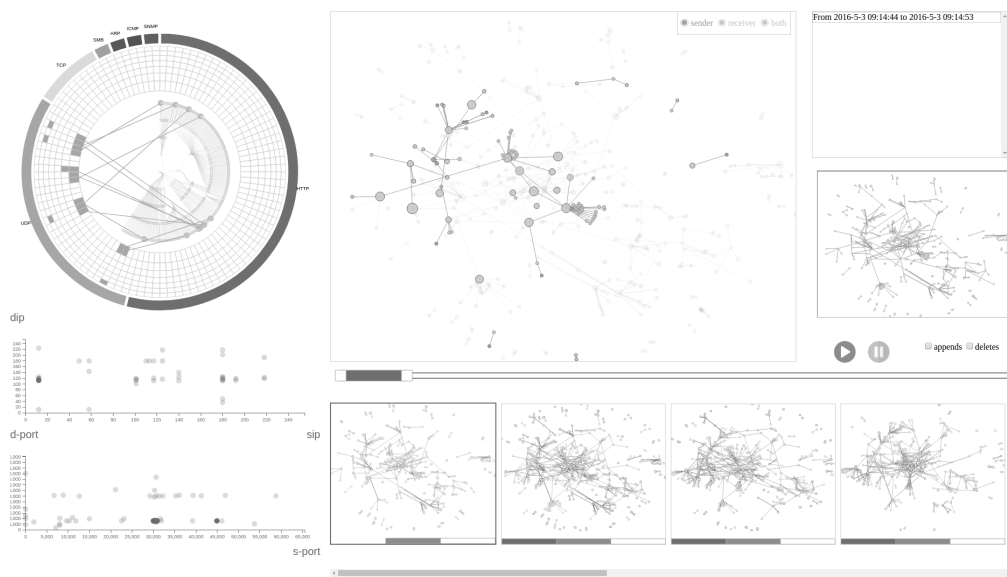


图 13 筛选 UDP

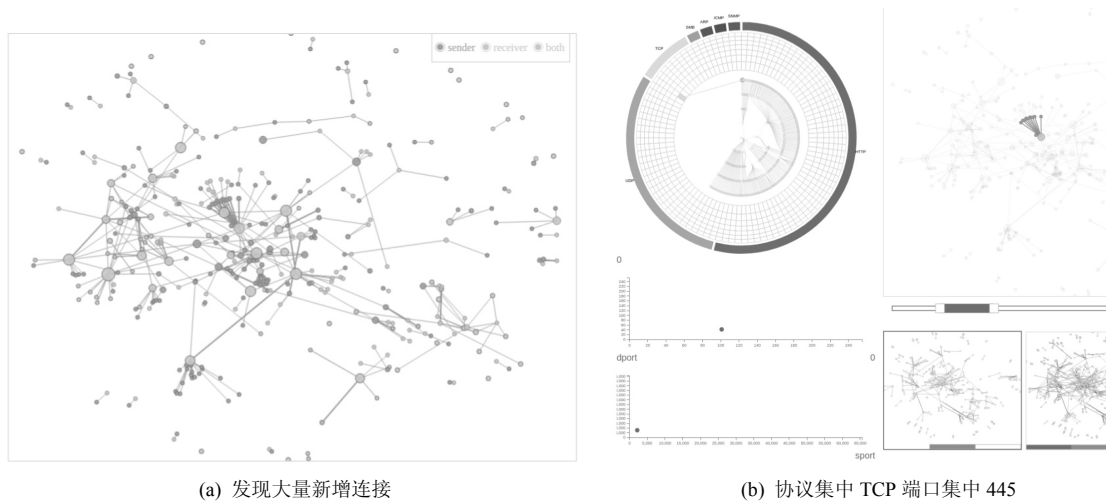


图 14 网络动态变化的异常检测

例分析展示系统功能。

本文工作虽然已经可以很好地满足网络分析的需求,但是还有很多方向可以进行优化。在未来的研究工作中,将从以下几个方面着手。

1) 优化布局算法,降低算法复杂度,以满足实时生成的流式数据规模。

2) 实现复现功能。全局统一的力导向布局可能不适用于某些场景,需要引入新的带有复现能力的局部优化算法。为了完善优化系统在更大数据量的数据集下的表现,可以综合运用一些简化图像的办法,如实现通过交互可以束拢展开的图压缩节点连接图。

3) 用户体验优化。将时间序列图隐藏在时间轴之中,当鼠标移到时间轴上的呈现“放大镜”效果。

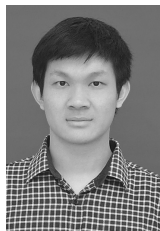
参考文献:

- [1] CONTI G. Security data visualization[M]. No Starch Press, 2007.
- [2] MARTY R. Applied security visualization[M]. Addison-Wesley Professional, 2009.
- [3] KOIKE H, OHNO K, KOIZUMI K. Visualizing cyber attacks using IP Matrix[C]// IEEE Visualization for Computer Security. 2005: 91-98.
- [4] MCPHERSON J, MA K L, KRYSTOSK P, et al. PortVis: a tool for port-based detection of security events[C]//ACM Workshop on Visualization and Data Mining for Computer Security. 2004:73-81.
- [5] ZHAO Y. MVSec: a novel multi-view visualization system for network security[D] Changsha: Central South University, 2013.
- [6] ABDULLAH K, LEE C P, CONTI G, et al. Visualizing network data for intrusion detection[C]//The Sixth IEEE SMC Information Assurance Workshop. 2005:100-108.
- [7] POHL M, REITZ F, BIRKE P. As time goes by: integrated visualization and analysis of dynamic networks[C]//ACM Working Conference on Advanced Visual Interfaces. 2008:372-375.
- [8] FEDERICO P, AIGNER W, MIKSCHE S, et al. A visual analytics approach to dynamic social networks[C]//The International Conference on Knowledge Management and Knowledge Technologies. 2011:1-8.
- [9] FARRUGIA M, HURLEY N, QUIGLEY A. Exploring temporal ego networks using small multiples and treering layouts[C]//The Fourth International Conference on Advances in Computer-Human Interactions. 2011:79-88.
- [10] ANDREWS K, WOHLFAHRT M, WURZINGER G. Visual graph comparison[C]//IEEE International Conference on Information Visualisation. 2009:62-67.
- [11] ALPER B, BACH B, RICHEL N H, et al. Weighted graph comparison techniques for brain connectivity analysis[C]//Sigchi Conference on Human Factors in Computing Systems. 2013:483-492.
- [12] DRAGICEVIC P. Interactive graph matching and visual comparison of graphs and clustered graphs[C]//The International Working Conference on Advanced Visual Interfaces. 2012:522-529.
- [13] COLLBERG C, KOBROUOV S, NAGRA J, et al. A system for graph-based visualization of the evolution of software[C]//ACM Symposium on Software Visualization. ACM, 2003:77-ff.
- [14] LERMAN K, GHOSH R, KANG J H. Centrality metric for dynamic networks[C]//KDD Workshop on Mining and Learning with Graphs. 2010:70-77.
- [15] FEDERICO P, PFEFFER J, AIGNER W, et al. Visual analysis of dynamic networks using change centrality[C]//The International Conference on Advances in Social Networks Analysis and Mining. 2012:179-183.
- [16] DWYER T, EADES P. Visualising a fund manager flow graph with columns and worms[C]//The International Conference on Information Visualisation. 2002:147-152.
- [17] BRANDES U, CORMAN S R. Visual unrolling of network evolution and the analysis of dynamic discourse[C]// IEEE Symposium on Information Visualization. 2002:145-151.
- [18] TVERSKY B, MORRISON J B, BETRANCOURT M. Animation: can it facilitate?[J]. International Journal of Human-Computer Studies, 2002, 57(4):247-262.
- [19] SHANMUGASUNDARAM M, IRANI P. The effect of animated transitions in zooming interfaces[C]// ACM Working Conference on Advanced Visual Interfaces. 2008:396-399.
- [20] CHEVALIER F, DRAGICEVIC P, BEZERIANOS A, et al. Using text animated transitions to support navigation in document histories[C]//The International Conference on Human Factors in Computing Systems. 2010:683-692.
- [21] HEER J, ROBERTSON G G. Animated transitions in statistical data graphics[J]. IEEE Transactions on Visualization & Computer Graphics, 2007, 13(6):1240-1247.
- [22] BEZERIANOS A, CHEVALIER F, DRAGICEVIC P, et al. GraphDice: a system for exploring multivariate social networks[J]. Computer Graphics Forum, 2010, 29(3):863-872.
- [23] HEER, JEFFREY, STUART K, et al. Prefuse: a toolkit for interactive information visualization[J]. Proc Chi, 2015, 37(4):421-430.
- [24] YEE, PING K. Animated exploration of dynamic graphs with radial layout[C]//IEEE Symposium on Information Visualization. 2002: 43-50.
- [25] GUILMAINE D, VIAU C, MCGUFFIN M J. Hierarchically animated transitions in visualizations of tree structures[C]//The 2012 International Conference on Advanced Visual Interfaces. 2012: 514-521.
- [26] CARD S K, SUH B, PENDLETON B A, et al. Time tree: exploring time changing hierarchies[C]//2006 IEEE Symposium On Visual Analytics Science And Technology. 2006:3-10.
- [27] HEER J, CARD S K. DOTrees revisited:scalable, space-constrained visualization of hierarchical data[C]//Working Conference on Advanced Visual Interfaces(AVI 2004). 2004: 421-424.
- [28] EADES P, HUANG M L. Navigating clustered graphs using force-directed methods[M]// Graph Algorithms And Applications 2. 2000:191-215.
- [29] FRIEDRICH C, EADES P. The marey graph animation tool demo[J]. Lecture Notes in Computer Science, 2001, 1984(1984): 396-406.
- [30] FRIEDRICH C, HOULE M E. Graph drawing in motion[C]//The International Symposium on Graph Drawing. 2001:220-231.
- [31] BRANDES U, WAGNER D. Analysis and visualization of social networks[M]// Drawing Software, Mathematics and Visualization. Berlin: Springer. 2004: 321-340.
- [32] BASTIAN M, HEYMANN S, JACOMY M. Gephi: an open source software for exploring and manipulating networks[C]//The Third In-

ternational Aaai Conference On Weblogs And Social Media. 2009.

- [33] AHN J W, TAIEB-MAIMON M, SOPAN A, et al. Temporal visualization of social network dynamics: prototypes for nation of neighbors[C]//The International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction. 2008:309-316.
- [34] PURCHASE H C, SAMRA A. Extremes are better: investigating mental map preservation in dynamic graphs[C]//The International Conference on Theory and Application of Diagrams. 2008:60-73.
- [35] ARCHAMBAULT D, MUNZNER T, AUBER D. TopoLayout: multilevel graph layout by topological features[J]. IEEE Transactions on Visualization & Computer Graphics, 2007, 13(2):305.
- [36] ARCHAMBAULT D, PURCHASE H, PINAUD B. Animation, small multiples, and the effect of mental map preservation in dynamic graphs[J]. IEEE Transactions on Visualization & Computer Graphics, 2011, 17(4):539-552.
- [37] KAMADA T, KAWAI S. An algorithm for drawing general undirected graphs[J]. Information Processing Letters, 1989, 31(1):7-15.
- [38] DAVIDSON R, HAREL D. Drawing graphs nicely using simulated an-nealing[J]. ACM Trans Graphics, 1996, 15(4):301-331.
- [39] FRUCHTERMAN T M J, REINGOLD E M. Graph drawing by force-directed placement[J]. Software Practice & Experience, 1991, 21(11):1129-1164.
- [40] FORESTI S, AGUTTER J, LIVNAT Y, et al. Visual correlation of network alerts[J]. IEEE Computer Graphics & Applications, 2006, 26(2):48-59.

[作者简介]



张浩城(1993-),男,上海人,上海交通大学硕士生,主要研究方向为可视化与可视分析。



吴晓洁(1995-),女,江西信丰人,上海交通大学本科生,主要研究方向为可视化与可视分析。



唐翔(1992-),男,江苏扬州人,上海交通大学硕士生,主要研究方向为可视化与可视分析。



舒润萱(1997-),男,吉林吉林人,上海交通大学本科生,主要研究方向为可视化与可视分析。



丁天琛(1996-),男,上海人,上海交通大学本科生,主要研究方向为可视化与可视分析。

董笑菊(1975-),女,吉林公主岭人,博士,上海交通大学副教授,主要研究方向为可视化与可视分析、形式化方法。