项目编号：T030PRP28062



**本科生研究计划（PRP）研究论文**

(第28期)

**论文题目： 信息可视化在网络安全数据分析中的应用**

**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**

信息可视化在网络安全数据分析中的应用

电子信息与电气工程学院 F1403004 丁天琛 5140309081

指导教师 电子信息与电气工程学院 董笑菊

摘 要

随着现代社会科技的发展，网络已成为人们生活中不可缺少的一环。伴随着网络的发展和普及，网络安全的重要性及隐患也逐渐显露了出来。本文为研究和分析网络安全数据提供一些信息可视化的方法，以助于识别和预警网络中存在的攻击和安全隐患。本文主要介绍IP/端口散点图和协议圆周图两种可视化方法，它们各自从不同的方面解读网络通信的数据，并和其他视图联动起来达到相乘的效果。

关键词：网络安全，可视化，网络通信，散点图，圆周图

ABSTRACT

With the development of science and technology in modern society, Internet has become one of the necessary parts in people’s life. The importance and hidden danger of network security are also appearing due to the development and spread of Internet. This paper will provide some ways of data visualization for researching and analyzing the data of network security, in order to recognize and alert the attacks and security bugs in network. This paper will focus on scatter plot of IP addresses and ports, and circle plot of protocols in. These two ways read and analyze the network communication data from different aspects, and will be combined with other views so that their effects will be improved.

Keywords: network security, data visualization, network communication, scatter plot, circle plot

1. 绪论

随着现代社会科技的发展，网络已成为人们生活中不可缺少的一环。伴随着网络的发展和普及，网络安全的重要性及隐患也逐渐显露了出来。特别是在最近，网络安全问题层出不穷，如针对用户的木马传播、XSS和CSRF攻击等，针对服务器的SQL注入、DDOS、文件上传提权等已是屡见不鲜。对于负责维护和管理网络的安全人员来说，及时找到漏洞加以弥补或是防患于未然便成了一件非常重要的事。那么，如何及时从数量庞大的网络通信数据中识别出网络攻击和有攻击倾向的数据呢？对于海量的数据，肉眼显然不能分辨出结果，而通过程序识别出的结论也未必完全准确和可靠，将肉眼与程序结合便成为了一种新的思路。因此，本文从信息可视化方面着手，提出通过多种视图对通信数据的多种方面来进行可视化分析的方案。本文将着重介绍IP/端口散点图及通信协议圆周图。

1. IP/端口散点图

2.1 初期构想

在思考如何将大量通信数据以一种较为直观的形式展现出来时，首先能够想到的便是将通信的源IP地址与目的IP地址进行可视化处理。为展现这一种二维数据的对应关系，我们考虑使用散点图或是矩阵实现。然而，由于IP地址数量过多，使用矩阵会使得整张图的方格密密麻麻，导致用户很难看清高亮方格，起到了反效果。而散点图则较为灵活多变，而且在空旷的背景下更能强调出它自身的存在以吸引用户的注意。由于源端口和目的端口的对应关系也类似与IP地址，所以我们决定将端口的对应关系也做成散点图，并使这两者联动。

接下来是对散点图的具体实现。我们初步先决定将所有的IP地址和端口数据全加入坐标轴观察效果，结果如图2.1所示（横坐标为源IP地址和端口，纵坐标为目的IP地址和端口，IP地址根据其二进制形式比较大小排列）。

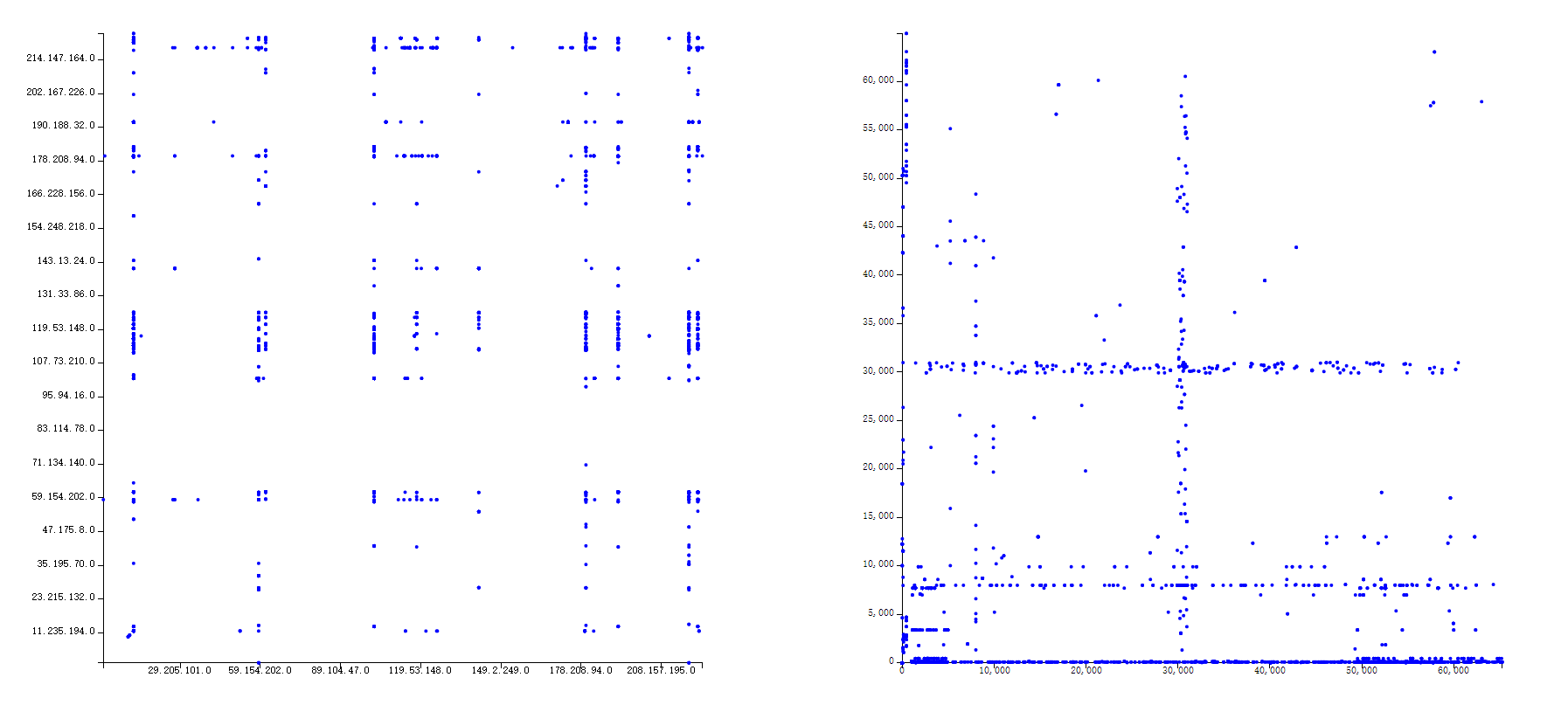


图2.1 散点图初期构想

从图2.1中可以看出，这些通信数据的IP与端口之间的对应关系确实存在着一些规律，也可以通过点的密集程度来判断通信的频繁程度。然而，我们对左侧这张IP地址的散点图却不甚满意。现在的IP散点图坐标轴上，容纳的IP数量有，即4,294,967,296个，而我们所使用的样例数据仅有3000条。试想，若是要处理上万条甚至百万条级的通信数据，用户所看到的的一定会是密密麻麻的数不清的点，这不仅会让用户感到读图困难，也会加重处理器的运算负担，从而导致整个页面响应缓慢甚至直接卡死。因此，有必要对该视图进行一定程度的改良。

2.2 最终设计

在思考改良方法时，我们想到，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范围、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，我们设计将IP地址散点图的坐标轴范围缩减为0至255。在现如今最为普遍的IPv4技术下，IP地址的形式为a.b.c.d四个IP段。在最初，坐标轴仅显示IP地址的a段，在(x,y)位置有高亮的蓝点即说明存在IP首段为x的地址向IP首段为y的地址发送的通信。如若用户对这个网段的数据感兴趣，则点击该蓝点，将会“进入”该网段，即通信双方的IP首段固定，整个散点图改为展现IP地址b段的分布情况。以此类推，用户可以逐渐“深入”地查看、获悉通信的IP分布情况。

然而这样的处理方法还存在一些缺陷，用户在看首段IP散点图时，只能知道网段的分布情况，而其他情况则无法获知。为了辅助用户查看视图，达到更直观的视觉效果，我们对于每个点做了统计处理操作。将每个点所代表的通信网段的总通信数量统计出，并将该数据以透明度的形式展现在散点图上。如此一来，用户便能在散点图上通过深浅得知网段的通信量大小，通信量较大的网段会十分醒目地被识别出。IP地址散点图的效果如图2.2所示。

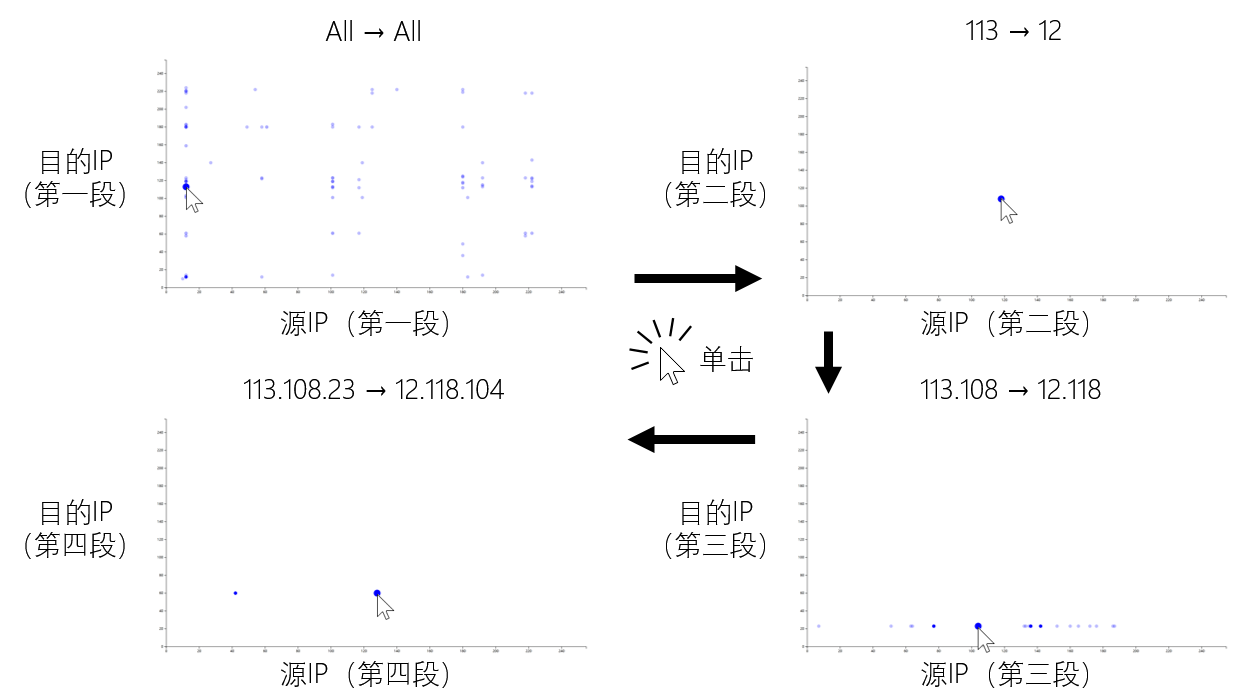


图2.2 IP散点图效果

对端口散点图也做相同的数据统计处理，具体过程同上。

最后是将两个视图合并联动的工作。当用户点击IP地址散点图进行数据筛选时，筛选出的数据会同步传达到端口散点图，两个视图将一起刷新，达到联动的效果，如图2.3所示。反之，当用户点击特定的端口时，程序也将执行筛选操作并反馈给IP地址散点图进行重画操作。

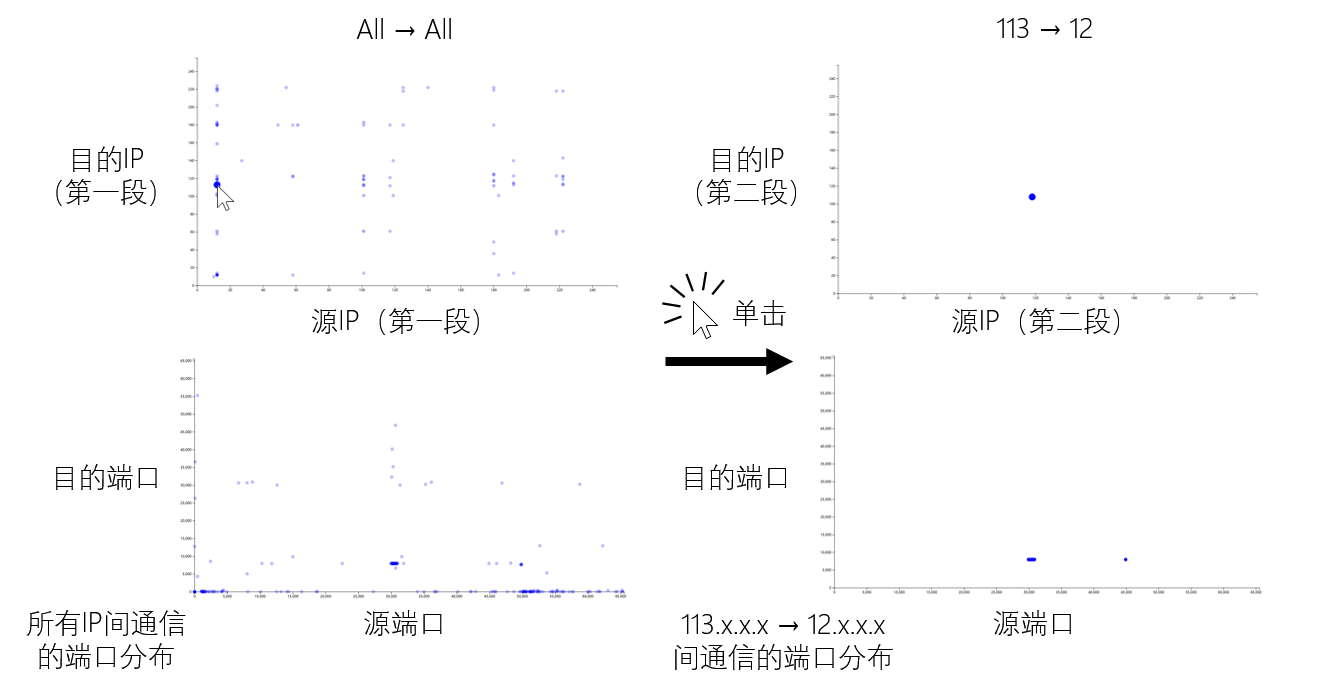


图2.3 联动效果

2.3 小结

综上所述，通过两个散点图的视图，我们可以较为清晰直观地观察到各通信数据的IP地址和端口的对应分布关系，以及特定网段、特定端口的通信量大小。通过这两个散点图的数据显示以及联动，我们可以较为方便地找出其中的一些规律和异常，这无疑为海量通信数据的处理带来了便捷。

1. 协议圆周图

3.1 简介

本文中介绍的圆周图是对《Visual Correlation of Network Alerts》[1] 一文的实现与改进。

圆周图分为内外两个部分。内圈用于展示各个IP地址之间的连接情况，外圈则展示了每个IP地址的通信情况（所用协议、时间等）。

原论文主要是应用于展现内网的受攻击情况，因此内圈的展现形式为各个服务器结点的分布和连接情况，来直观地观察受攻击服务器结点的扩散和转移。而这对于我们现在要处理的上海市城域网数据来说并不符合要求。因此，正如上文所提到的原因，用户更关心某个网段的通信情况，我们将内圈改为了由IP网段组成的树，树的所有叶子结点代表了IP地址，而其他结点均代表了某个网段，数据结构图如3.1所示。

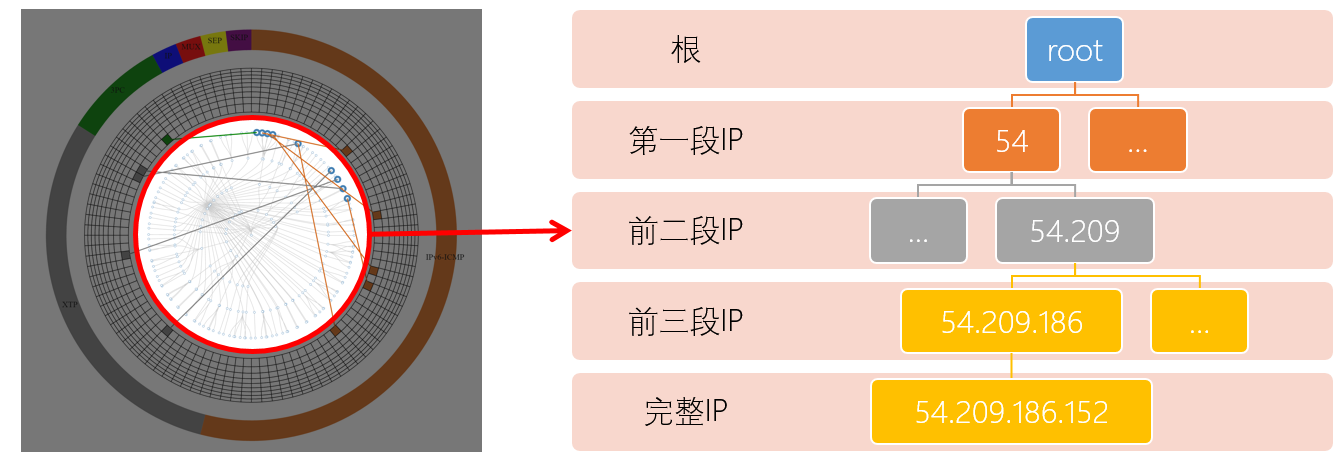


图3.1 内圈IP树示意图

形成这样的结构之后，每个IP都分配到了它特定的位置，与它相邻的兄弟结点都意味着它们在同一网段内，与它再稍远一些的结点则意味着它们在一个大网段内。这样即使在数据繁多的情况下也能一目了然各个IP间的位置关系，为用户的准确判断提供帮助。

外圈是由若干个同心圆和其半径线切割形成的100\*9的信息展示区域,它包含了本次通信的协议以及时间信息。不同的圆心角对应了不同的协议，而不同的半径对应了通信发生的时间。协议类型在最外圈以不同颜色的圆环标注了出来，而且方格以及连接IP结点和方格的线也均显示相同的颜色，以便于用户从许多数据中快速识别出所需要的那一个。离圆心越近的方格表示该通信发生的时间越晚，最靠近内圈的方格即表示该通信的时间发生于所选时间轴的末尾；越向外的方格表示该通信发生的时间越早，处于最外层的方格即表示该通信的时间发生于所选时间轴的开头，通过这个时间表，我们可以获悉这个IP地址的历史行为，为寻找异常行为提供帮助。其具体结构如图3.2所示。

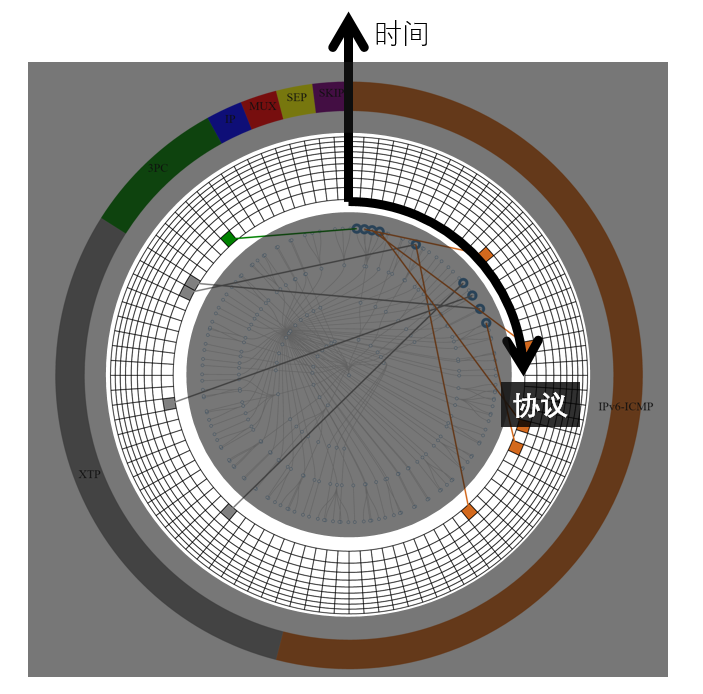


图3.2 外圈示意图

如此一来，该视图便清晰的阐述了由哪个源IP(where)、在什么时间(when)、使用了什么协议(what)来进行通讯，已满足可视化的基本要素。

3.2 改进与扩展

尽管通过不同颜色的手段辨别不同的通信数据，但是我们注意到，一旦数据量增多了，会出现大量高亮的方格和连线，令人眼花缭乱，以及的规模无法容纳下大量数据，若是再细分方格则会出现与上文阐述的矩阵相同的问题。因此，辅助与筛选工作便必不可少。

首先，为了能够让用户在信息繁多的情况下清楚准确地定位到自己需要的数据，我们决定在鼠标交互方面增加若干功能。当用户的鼠标移经某个IP地址的结点时，将该结点以及与该结点相连的通信信息高亮，并同时淡化其他所有数据；当用户的鼠标移经某个连线或是方格时，仅仅高亮该条通信的信息而淡化其他所有数据。另外，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，也增加了对于鼠标点击事件的响应：当用户鼠标点击图上的数据时，锁定该数据的高亮，鼠标移开也不会被淡化，直到再次点击该数据为止。效果如图3.3所示。

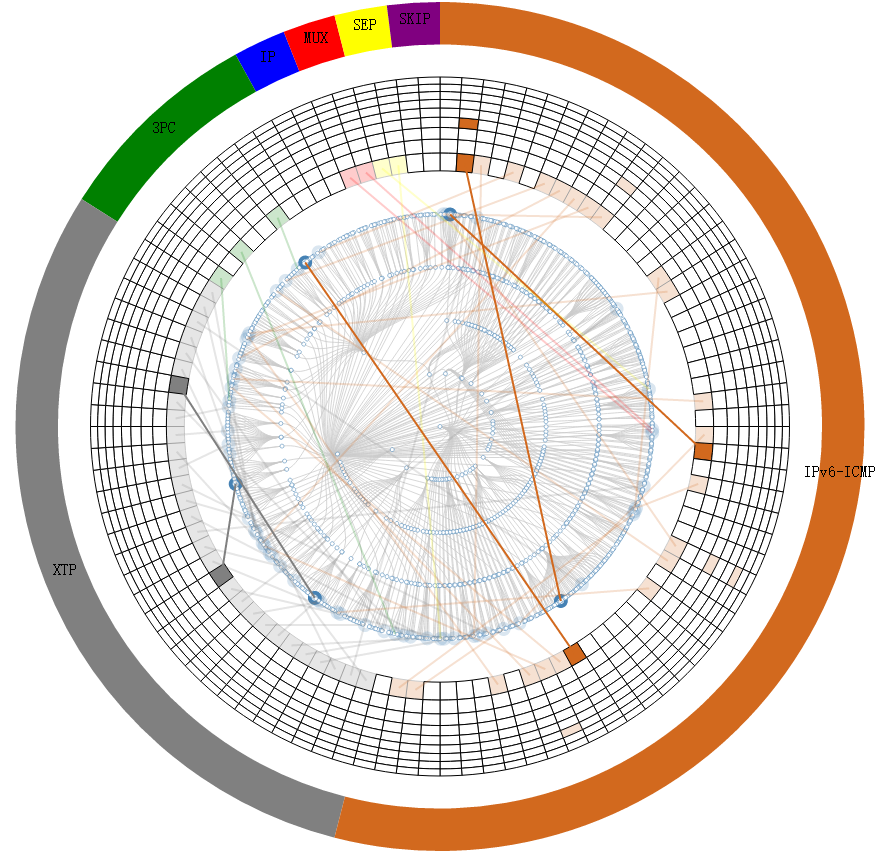


图3.3 鼠标交互效果图

从上图可以看出，当用户选中想要的数据后，其他的用户不关心的“无用”数据都淡化融入背景中去了，为用户获取并查看目标数据提供了极大的便利。

本视图主要提供了通信数据中协议信息的可视化，则关于协议的筛选功能自然也应该由其完成。通过点击最外圈的协议标注圆环，可以点亮或取消某个协议的选中状况。初始时的筛选条件为全部点亮，如果某个协议被取消选中，则会同步更新到筛选条件中并进行重画响应；同理，若是点击已经取消的协议，则该协议将会重新点亮并把筛选数据更新重画。示例如图3.4所示。

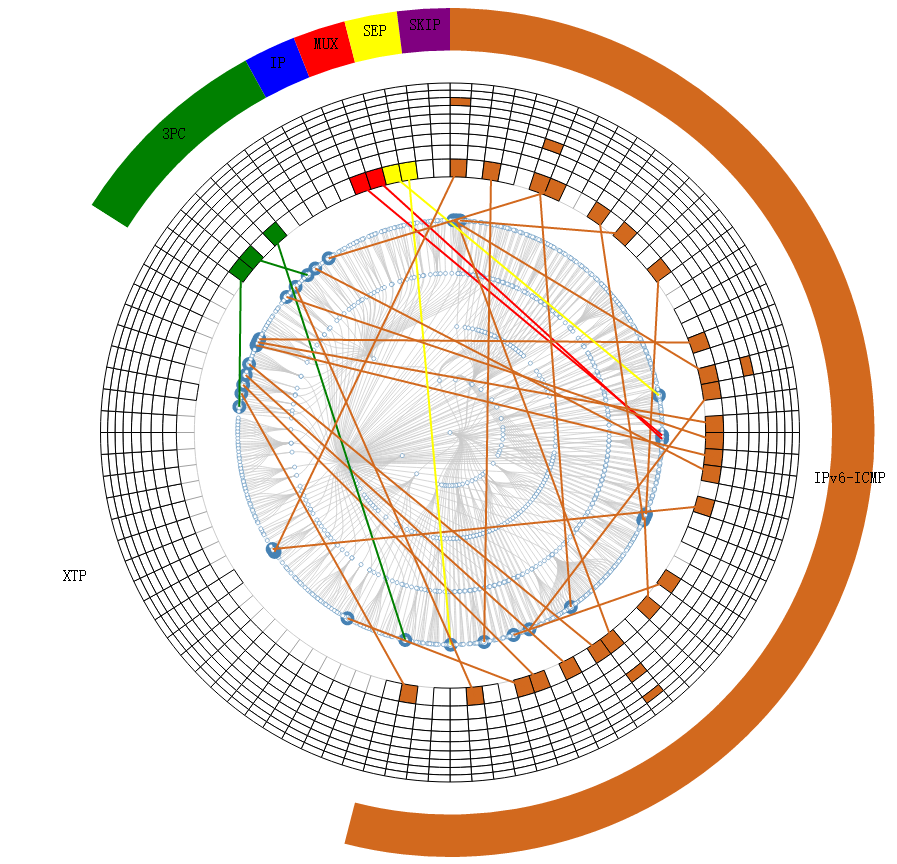


图3.4 取消XTP协议后的筛选示例

由此一来，用户便能随心所欲地定制自己想要的协议筛选，并与其它视图进行联动。过滤了用户不需要的“噪声”信息后，特定协议的状况便能清晰地展现出来，也更利于用户分析自己关心的那部分数据。

最后是对于大量数据情况下的预处理优化。由于外圈的规模限制，当传入的数据量较多时，程序将根据通信量、发生时间、该协议已占用情况等参数自动筛选并展示其中一部分数据。如若用户希望获得更为详细全面的数据，可拖动时间轴将时间筛选条件设置为一个较小的区域，以减小数据量的传入。而用户决定时间条件的依据则可从另外2个视图所提供的信息中获取。

3.3 小结

综上所述，用户可以通过该圆周图获知IP网段的分布情况、通信的时间以及通信所用协议。用户能利用鼠标交互功能进行高亮锁定操作以辅助观察，当用户点击最外圈的协议标注圆环时可以执行数据的筛选操作，只将符合条件的数据展现出来。

1. 系统总体与交互联动

4.1 系统总体

系统的总览如图4.1所示。

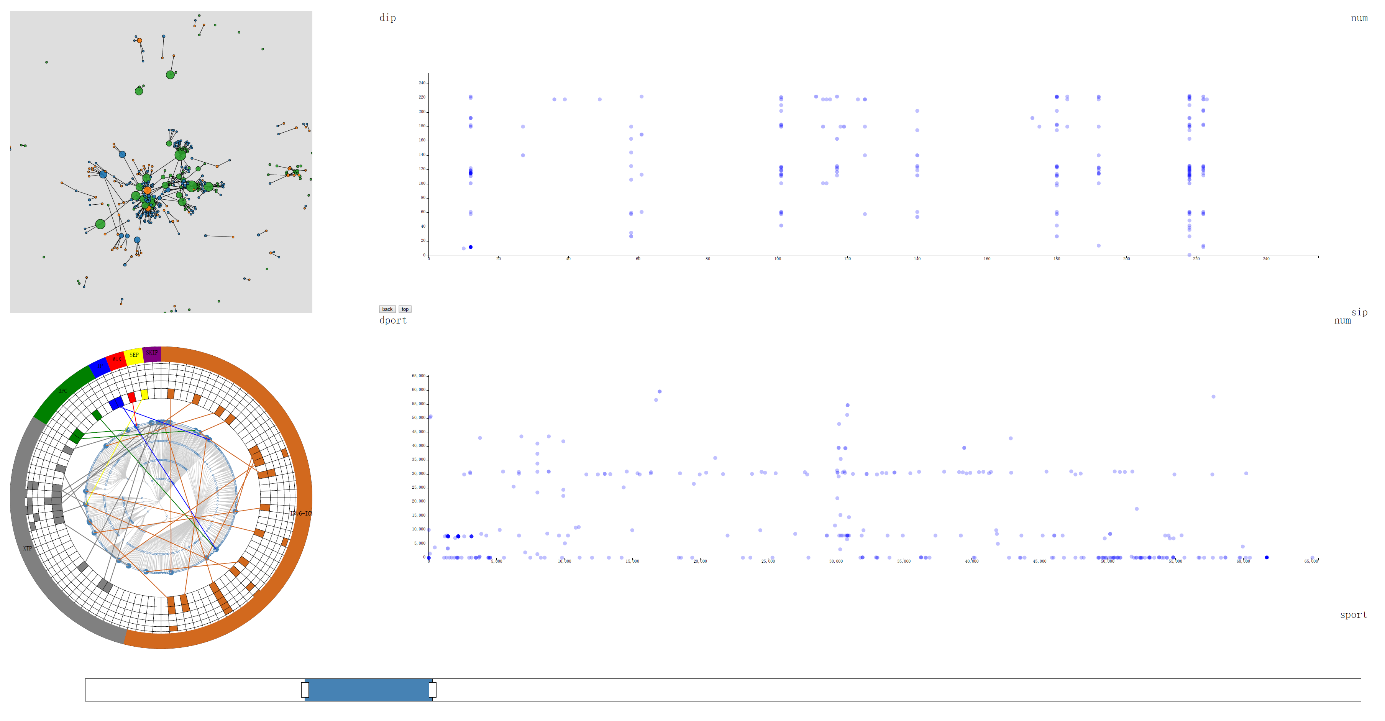


图4.1 系统总览

该系统的右侧部分即为IP地址和端口的散点图，左下角部分即为协议圆周图。

左上角部分为基于阶段式动画的力导向动态网络，最下侧为用于筛选时间的时间轴，这些部分的内容及其联动详见本项目的另一篇论文，由舒润萱同学完成。

4.2 交互联动

该系统的各个视图均参与了数据筛选的联动。点击左上角的动态网络的任意结点，散点图和圆周图也根据其筛选出的数据作出相应改变；点击右侧散点图的任意蓝点，将会根据相应的IP网段/端口数据进行筛选，将符合条件的数据传入其他视图；点击左下角视图最外圈的协议标注圆环，将根据剩余高亮的协议进行筛选，并把结果同步到其他视图内；最下侧的时间轴承担了最主要的筛选功能，它使得用户能够自定义地查看任意时段的网络情况，并能结合圆周图的时间信息进一步拖动时间轴定位问题。当用户的鼠标移经圆周图中的IP结点、连线或是方格时，左上角的动态网络也会对对应的数据信息进行高亮处理，用户的鼠标点击锁定操作亦会对动态网络中的结点起效，这使得用户可以使用鼠标在对两个视图同时点亮、同时熄灭的操作过程中获知对应关系，对分析通信数据起了较好的辅助作用。

1. 结论与项目后续

本系统旨在通过多个视图、从多个方面来完整地解析通信数据。动态网络反映了所有通信IP之间的关联性，散点图列出了通信的IP及端口情况以便找出规律，圆周图表现了不同IP网段的分布、通信时间、以及通信所用协议情况。我们通过本系统基本全面地解读了上海市城域网的样例数据所包含的信息，也能从中获取到一些规律，如源IP大多集中在几个固定的网段(12、58、101、180、192、218、222等)、目的端口大部分为80端口（HTTP协议端口，这说明目的IP为网络服务器端），还有一部分集中在8000和30000附近、通信所用协议大多为IPv6-ICMP和XTP等。

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，数据量较少，无法从本系统中看出较为明显的异常情况。我们期望通过后续更多的数据以及对系统自身的改进扩展来得出更令人满意的结论。

参考文献

[1] Foresti, S.; Agutter, J.; Livnat, Y.; Moon, S. & Erbacher, R.: Visual Correlation of Network Alerts[J]. Computer Graphics & Applications IEEE, 2006, 26(2):48-59.

[2] Spence R. Information Visualization: Design for Interaction (2nd Edition)[M]. Prentice-Hall, Inc. 2007.

[3] https://d3js.org[OL]

谢辞

首先要感谢我的指导老师董笑菊，她带领我一步步熟悉了解d3.js这个数据可视化库，从一开始做几个基础的小练习到后面尝试做一些项目，以及参加chinavis和soda的比赛，我从中收获了许多知识和宝贵的经历。感谢与我一起完成这个项目的舒润萱同学，他经常与我共同讨论关于数据的处理、系统的构思以及程序的编写优化等，并给出了许多重要的意见。最后也是他将我们各自分别完成的工作统合起来，画龙点睛般地使几个单独、单调的视图组合成为了一个完整的、有交互的系统。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，这份数据十分地详细和完整，在我们调试程序的过程中给了我们很大的帮助。最后感谢所有实验室的前辈同学们，他们举行的相关知识讲座令我受益匪浅，加深了我对JavaScript和d3.js库的理解；他们也不厌其烦地悉心解答了我提出了种种问题，帮助我顺畅地学习和使用这一可视化工具。