挑战一的可视化方案特点分析

卫学仕，文明慧，莫小君，李松阳，秦红星

(重庆邮电大学计算机科学与技术学院　　重庆　　400065)

摘要—在此次比赛中，我们使用了不同的可视化方案。其中对服务器的区分我们主要使用的方案是力导向图，对服务器的分类我们主要使用的方案是热点图，使用平行坐标的方式来发现内部网络的常规通信模式。

关键字—D3；力导向图；热点图；平行坐标系

简介

我们的完成过程主要分为三个阶段，第一阶段是对原始数据的预处理，第二阶段是将数据可视化，第三是对可视化结果的分析阶段。

本文第一部分将详细论述力导向图在区分服务器和客户端方面的特点、热点图在区分服务器类型方面的特点以及平行坐标系在挖掘内网通信模式方面的特点。

第二部分主要从实用性、可交互性和灵活性三个方面来分别论述可视化方案的特点。

第三部分将对本文主要内容进行简单总结。

1　主要的可视化方案

挑战一提供的数据主要有时间、源IP、目的IP、目的端口号、协议、上行流量和下行流量七个属性。我们在做分析之前根据题目的具体要求提出了不同的可视化方案。

在区分客户端和服务器的时候，我们使用力导向图来表示内部网络之间的通信关系，将内网中的每一台主机抽象为一个节点，主机与主机之间的通信抽象为边。然后根据数据画出内部主机之间的通信关系图，从中区分客户端和服务器。

我们主要根据功能对服务器进行分类，所以主要使用协议来推断服务器的功能类型。不同的服务器通过不同的协议与客户端进行通信，我们首先统计出与每台服务器通过不同的协议通信的客户端数量，然后对统计结果进行可视化。我们选取的方案是热点图，通过颜色深浅来判断出每台服务器主要使用的通信协议。

在挖掘内部网络通信模式方面，我们主要使用平行坐标的可视化方案，主要通过协议来反映出内部网络的通信模式，所以我们主要选取客户端IP，协议，服务器IP三个属性。通过平行坐标系的方案表现出来，能够很直观的看出客户端与服务器主要使用的通信协议，从而反映出主要的通信模式。

下面三个小节将详细论述不同的可视化方案的特点。

1.1　服务器与客户端的区分

力导向图主要用来表示实体与实体之间的关系，用以帮助用户从中发现潜在的信息。我们使用力导向图来表示公司内网之间的通信，由于服务器会被很多台主机访问，所以用力导向图表示出来以后，可以发现服务器与多台主机有连接关系。从而可以区分出服务器与客户端。

在使用力导向图来模拟内网通信网络的同时，我们还可以根据数据属性的不同来相应设置图中节点与边的属性。比如我们可以根据流量大小来设置节点的大小，来根据两个节点之间的通信协议来设置边的颜色等，有助于我们的判断。

● *卫学仕，重庆邮电大学，E-mail:675452587@qq.com.*

Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014; date of publication xx xxx 2014; date of current version xx xxx 2014.
For information on obtaining reprints of this article, please send *e-mail to: tvcg@computer.org.*

我们用图1.1所示的力导向图来表示内部网络服务器和客户端之间的通信网络，其中红色节点表示服务器，大小表示流量，每个白色节点代表的是同网段的一组客户端。

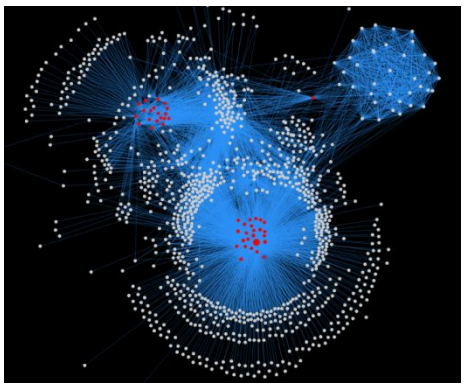

图 1.1 内网通信关系网络(力导向图)

1.2　服务器的类型划分

我们使用图1.2所示的热点图来根据功能区分服务器的类型。首先我们统计出每台服务器通过不同的协议与多少台客户端有通信，然后根据统计结果来绘制热点图。此时的数据主要有三个维度，服务器IP、协议、与之通信的客户端数量。我们使用横轴和纵轴来分别表示服务器IP和协议，然后将与之通信的客户端数量用颜色编码，以从中发现每台服务器主要使用的协议，从而根据协议判断出服务器的功能。

使用热力图来区分服务器类型，能将所有的服务器同时都表示出来，可以发现一些潜在的信息。比如所有服务器中，哪种类型的服务器比较多，哪些服务器的功能比较单一或哪些服务器的提供多种服务等。

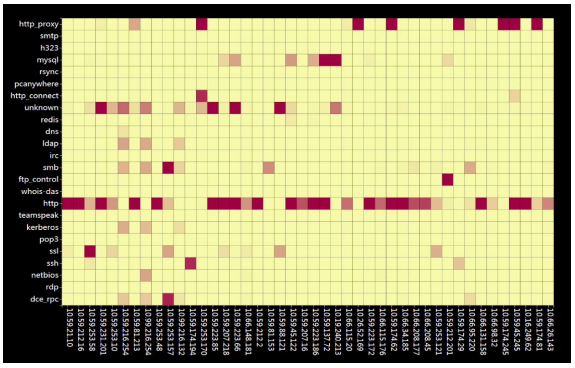

图 1.2 服务器的功能分类(热点图)

1.3 内部网络通信模式

在挖掘通信模式方面，我们使用平行坐标系的可视化方案(图1.3)。我们使用了三条坐标轴来分别表示客户端分组、协议、服务器IP三个属性，然后根据数据将客户端IP所在的IP段与服务器IP通过协议相连接，来表示客户端与服务器之间的通信关系。为了更容易区分，我们将协议用颜色编码。

使用平行坐标系不仅可以发现整体的通信模式，还可以通过交互的方式选择查看特定的信息，这种交互方式更加丰富了平行坐标系的功能，以帮助用户发现更多的信息。

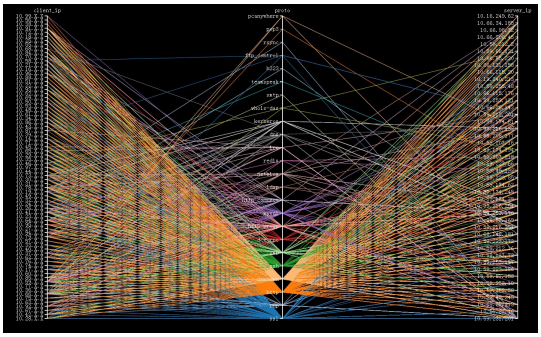

图 1.3 内部网络通信模式(平行坐标系)

2 讨论

在选取可视化方案的同时，我们首先要考虑到的是该可视化方案是否实用，能不能满足需求，这是可视化的最终目的，我们不能脱离这个目的去做可视化。其次还要考虑到交互，一个优秀的可视化方案在满足基本需求的同时，应当支持用户交互，根据用户的选择来展示不同的内容。最后还要考虑到灵活性，一个好的设计应该是灵活的，应该支持设计者自己创作个性化的可视化作品。除此之外，还要考虑到界面的简洁和美观等方面。

在接下来的小节中，将讨论本文可视化方案的实用性、交互、灵活性三个方面的内容。

2.1 实用性

在表示内网主机之间的关系方面，力导向图是非常合适的，它可以根据两个主机之间的通信来模拟它们之间的关系。将主机抽象为节点，通信关系抽象为边，从而可以发现主机与主机之间的关系模式。

由于我们根据功能对服务器类型进行划分，所以主要考虑它们使用的协议，通过协议来推测出服务器的类型。但是由于一些服务器使用多种协议通信，所以要将主要使用的协议突显出。因此我们使用热点图颜色的深浅来表示与服务器使用对应协议通信的客户端数量，从而能直观的发现服务器主要使用协议，然后进一步确定服务器的类型。

在通信模式的可视化表示方面，我们使用的平行坐标系可以同时将多个维度的信息同时表现出来，可以很直接的发现使用的主要协议有哪些。

2.2 可交互性

我们使用的所有可视化方案都是可以进行交互操作的，在可视化过程中，有些信息不便同时展现出来，所以用户要靠交互的方式来查看需要的信息。

在力导向图中，用户可以拖动某个节点来将其置于合适的位置来减轻视觉混淆，也可以通过交互的方式来查看每个节点的信息，比如IP和流量等。

在热点图中用户也可以通过交互的方式来查看某一台服务器使用某种协议与多少台客户端进行通信。

在平行坐标系中，除了能展现整体的通信模式以外，也能通过移动滑动窗口来查看指定的服务器通信模式。也可以同时选择多个条件来查看更详细的信息。

2.3 灵活性

我们使用的可视化方案都是灵活可扩展的它支持图形的任意组合，可以根据不同的需求制定个性化的可视化方案。

如图2.3，我们在模拟内网实时通信中，使用不同的布局和图表的组合，其中有树状布局、堆栈布局以及柱状图和折线图等。其中树状布局用来分别表示服务器IP和客户端分组，使用堆叠图来表示一段时间不同的协议流量大小，柱状图表示不同协议的流量累计大小以及使用折线图来记录内网时间-流量关系。

可以根据不同时间的不同数据来选择性的更新对应元素的属性，而不需要对所有元素进行重新绘制，以提高程序运行效率。

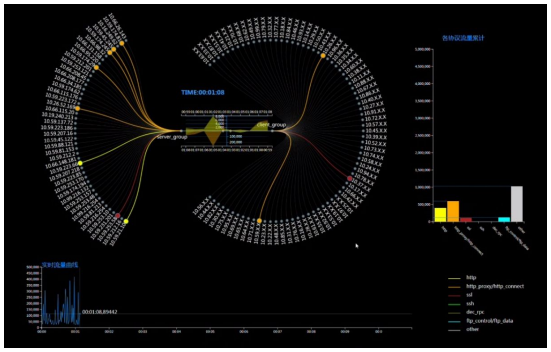

图 2.3 模拟实时通信(多图表组合)

3 结论

本次挑战所选取的可视化方案无论从实用性、交互还是灵活性都能够达完成任务的目的。除此之外，这些方案都具有可扩展性，能够在已有的图形基础上，扩展其他的功能或者结合其他的图形来丰富图形的内容，从而引导用户发现更多的信息。

参考文献

[1] Bostock M, Ogievetsky V, Heer J. D 3 Data-Driven Documents[J]. IEEE Transactions on Visualization & Computer Graphics, 2011, 17(12):2301-2309.

[2] 陈为,沈则潜.数据可视化.电子工业出版社,北京,2013.