2015 年中国可视化与可视分析大会

数据可视分析挑战赛-挑战 1

（ChinaVis Data Challenge 2015 - mini challenge 1）

答卷

参赛队名称：　　重庆邮电大学-卫学仕

团队成员：　　卫学仕，重庆邮电大学，675452587@qq.com，队长

莫小君，重庆邮电大学，1126153526@qq.com

文明慧，重庆邮电大学，2013211643@stu.cqupt.edu.cn

李松阳，重庆邮电大学，lisy199199@163.com

秦红星，重庆邮电大学，qinhx@cqupt.edu.cn，指导老师

是否学生队（是或否）： 是

使用的分析工具或开发工具（如果使用了自己研发的软件或工具请具体说明）：D3，python，MySQL

共计耗费时间（人天）： 30 人天

本次比赛结束后，我们是否可以在网络上公布该答卷与视频（是或否）：是

（灰色字为参赛信息填写模板，请参赛者在提交时参照模板填写）

挑战 1.1：通过对一周 tcpflow 数据的可视分析，找出 TSZNet 公司内部网络中的客户端与服务器，并给出该公司网络体系结构拓扑图。（请将回答限制在 800 个字和 5 张图片）

首先，我们根据日志记录中内网与内网之间的通信数据，用力导向图来表示内网之间的通信(如图 1.1)，每一个节点代表一个主机。用连线表示两台主机之间有通信，连线的宽度代表两个节点之间的流量大小，其中颜色表示使用最多的通信协议，节点的大小表示流量。从图 1.1 中可以看出有一些主机与很多台主机都有连接关系，并且流量大于大多数主机，那么可以认为该主机可能为服务器。然后依次找出可能是服务器的主机，如图中黑框内所示，然后进行下一次筛选。

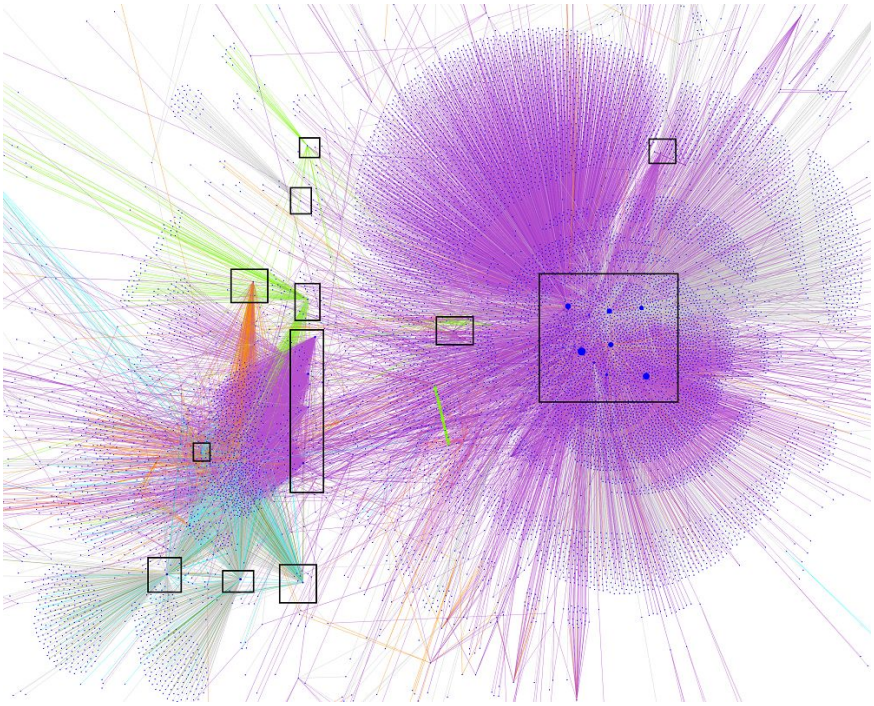

图 1.1

其次，根据第一次筛选结果再次画出通信网络图(如图 1.2)，然后从中去除掉孤立的点和度小于平均度的节点。剩下的基本可以确定为服务器。

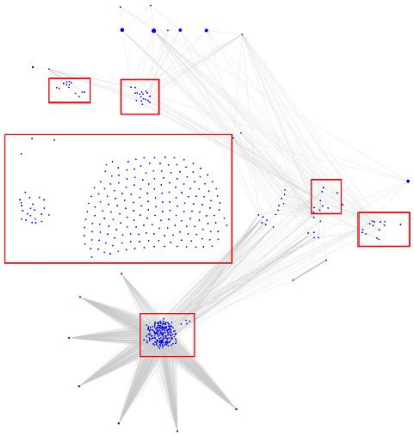

图 1.2

最后根据确定的服务器 IP 所在的 IP 段分组后结果如图 1.3 所示。

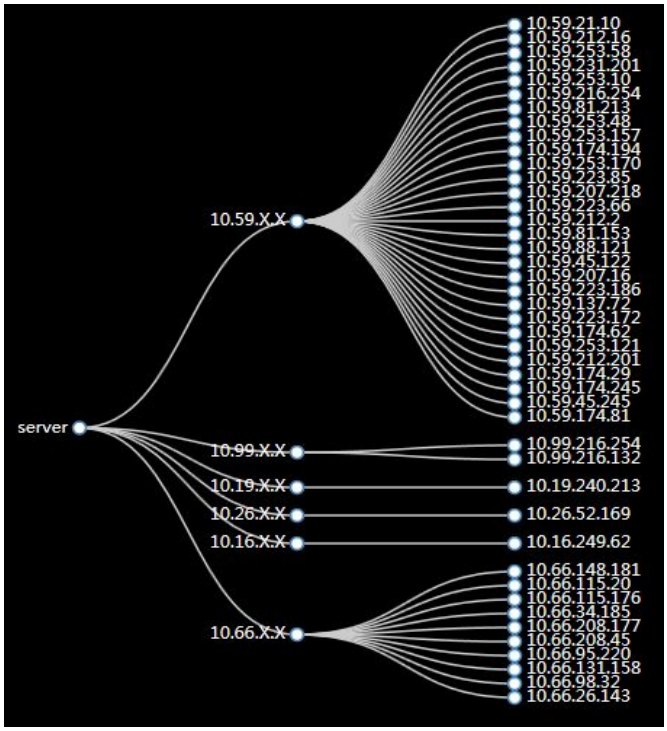

图 1.3

图 1.4 表示内网主机与服务器之间的通信关系，其中白色的节点表示客户端的 IP 段，精确到每个 ip 地址的前 3 段，如 10.59.23.X。红色的代表服务器，同样服务器流量用节点大小编码。

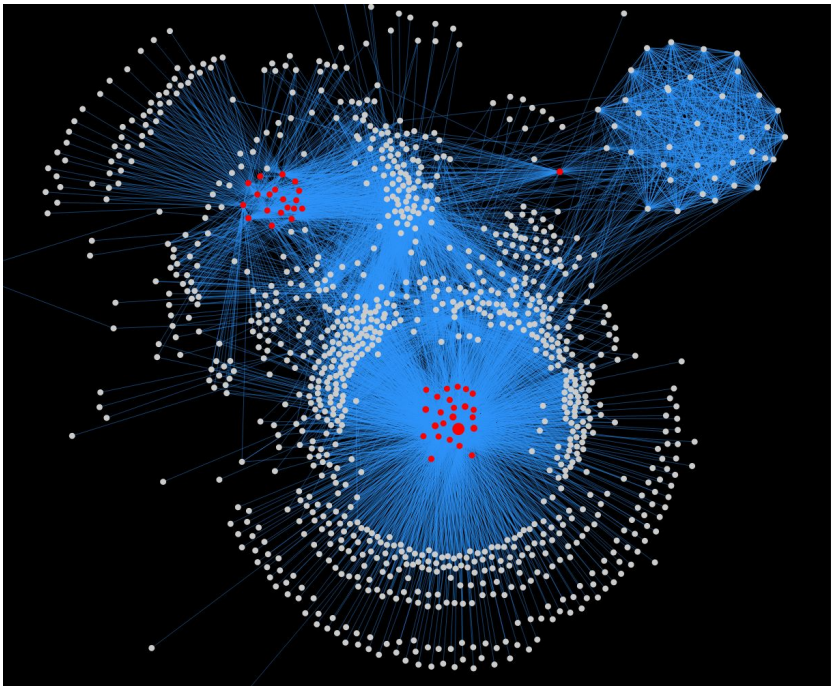

图 1.4

图 1.5 为公司网络体系结构拓扑图。从通信日志可以看出，公司内部使用 A 类私有地址，因此可以根据客户端 IP 所在的 IP 段来对客户端进行分组，每个网段内的主机通过核心交换机与内网中其他网段主机和服务器进行通信。服务器与核心交换机直接相连，内网主机通过代理服务器访问外网。

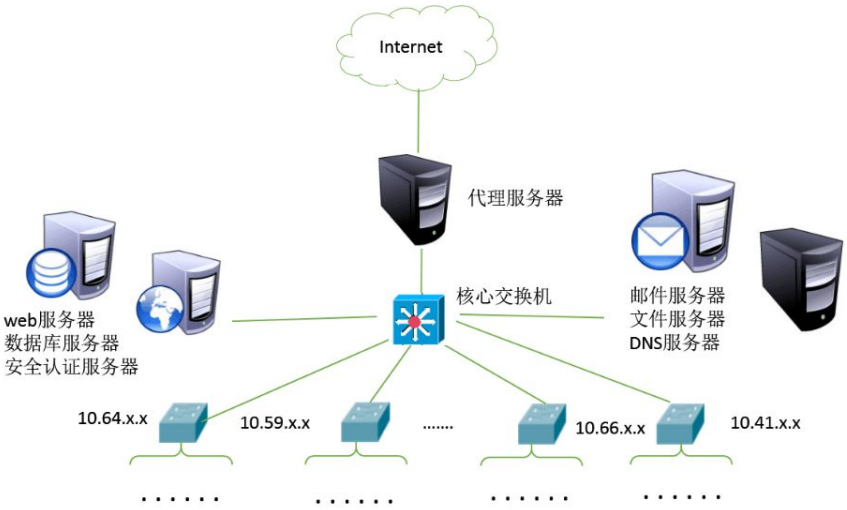

图 1.5

挑战 1.2：通过对一周的 tcpflow 数据的可视分析，对 TSZNet 公司内部网络的服务器进行分类，分类标准不限，比如按照功能、按时间特点、按行为特点、按流量特点等等。（请将回答限制在 1200 个字和 10 张图片）

我们主要通过功能来对服务器进行分类，因此主要通过日志记录中的协议来区分每台服务器的功能。

首先我们使用 python 统计出每台服务器通过什么样的协议、与多少台不同的客户端主机通信。然后根据这些统计数据做分析。

根据统计结果我们使用了热点图的可视化方式来从中发现每台服务器的类型(如图 2.1)。

在图 2.1 中，用横轴来表示每台服务器的 IP，纵轴表示服务器使用到的协议，然后可以画出热点图，其中每个矩形都对应一台服务器和协议，颜色表示对应的服务器通过对应的协议通信的客户端的数量，比如颜色越深，代表与之通信的客户端数量越多。而颜色浅的代表对应的服务器通过对应的协议与客户端通信数量比较少甚至没有。从图 2.1 中可以看出，http 协议使用最多，那么可以推断出这些服务器中大都是 web 服务器。

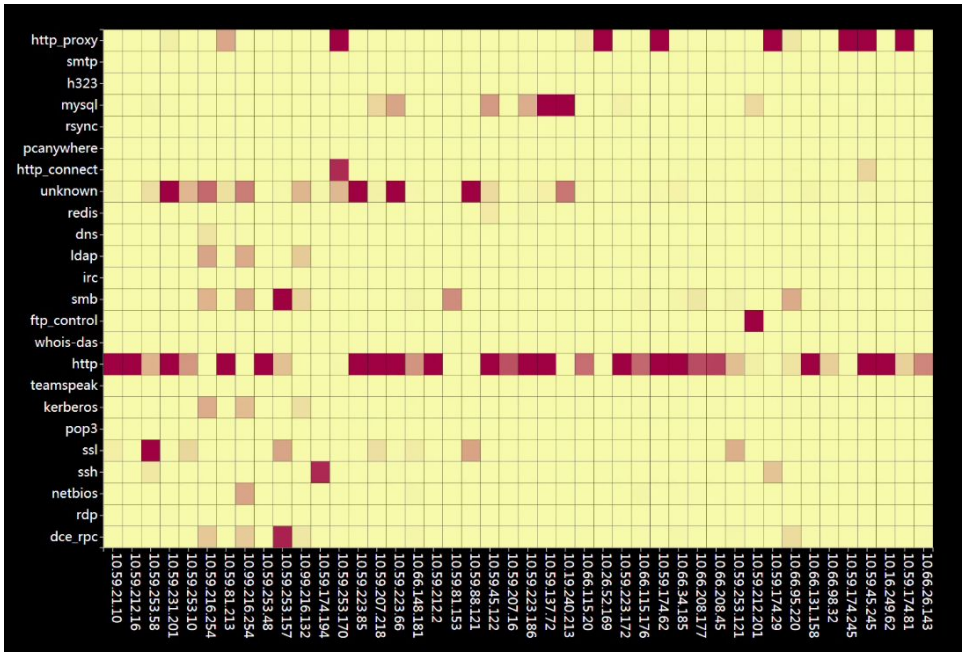

图 2.1

此外我们可以通过交互的方式来查看某台服务器通过某种协议与多少台客户端进行通信如图(2.2)。对于一台服务器而言，主要通过判断颜色的深浅来确定这台服务器主要使用的协议。在图 2.2 中，我们查看了 10.59.223.186 这台服务器通过 http 协议与多少台客户端通信，鼠标移动到对应的矩形上的时候，就会显示对应的矩形代表的信息，如图，这台服务器通过 http 协议与超过 10000 台客户端通信(大于 10000 的我们记为 10000)，那么可以确定 10.59.223.186 这台服务器的类型为 web 服务器。

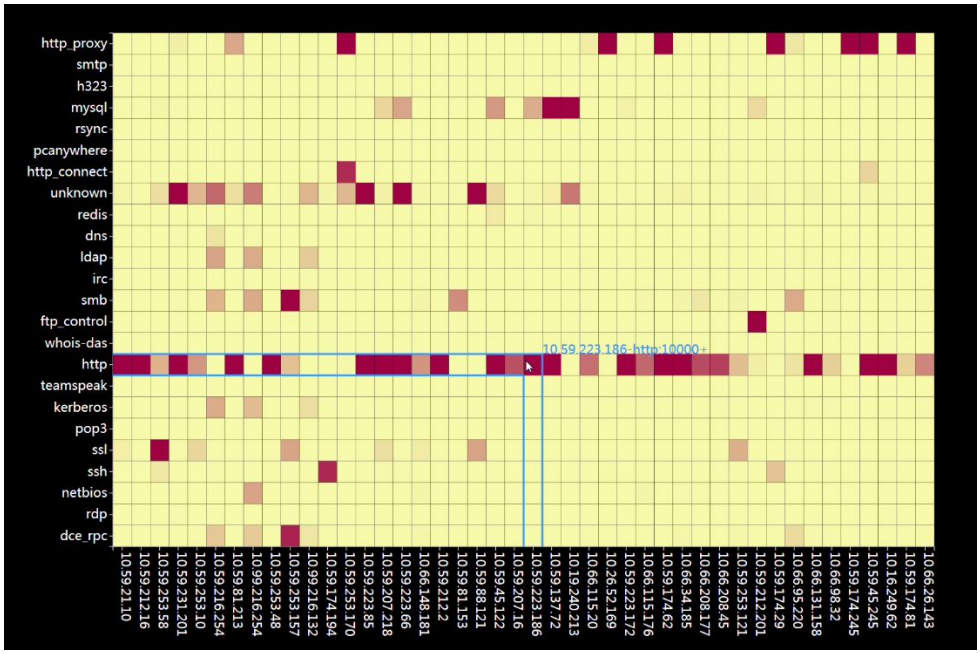

图 2.2

同理，我们可以判断出其他服务器类型(如图 2.3)，从图中我们可以很直观的看到 10.26.52.169 主要通过 http_proxy 协议与客户端通信，则可以认为这台主机是代理服务器。

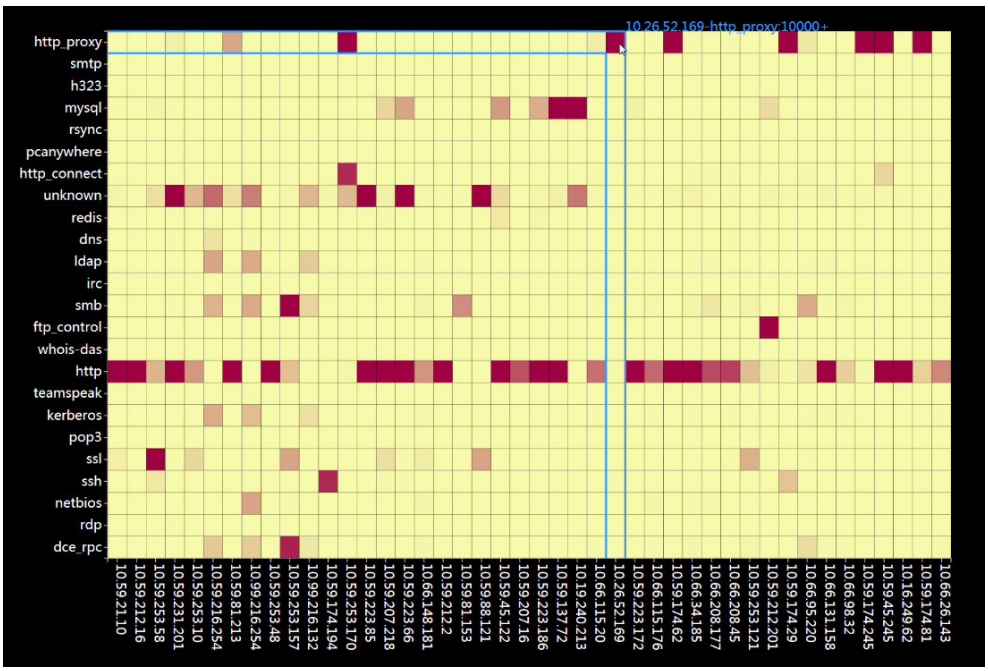

图 2.3

但是对于一些服务器，它主要的通信协议并不唯一，如图 2.4，10.99.216.254 这台主机主要通过 unknown、ldap、smb、kerberos 等协议通信。那么可以推断出这个服务器提供多种服务，我们将此类服务器认为是有特殊用途的服务器。

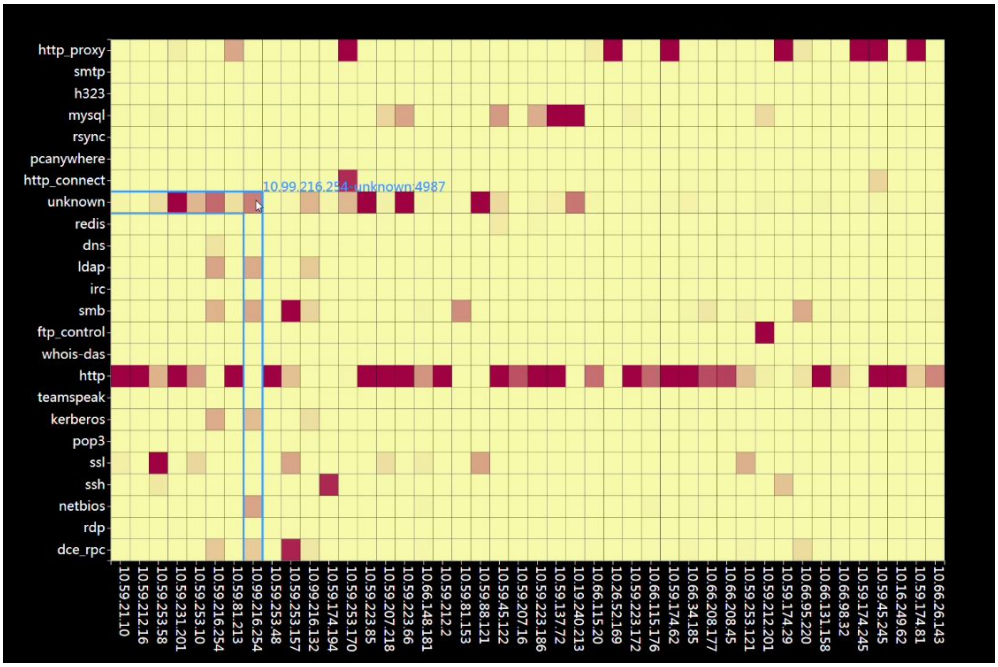

图 2.4

此外，我们还可以通过第 3 小题的图来区分服务器类型，通过在平行坐标上选择任意一台服务器，来查看这台服务器主要的通信协议有哪些。如图 2.5，我们选择查看 10.59.212.16 这台服务器，可以看出主要通过 http 协议通信，那么可以认为它是一台 web 服务器。

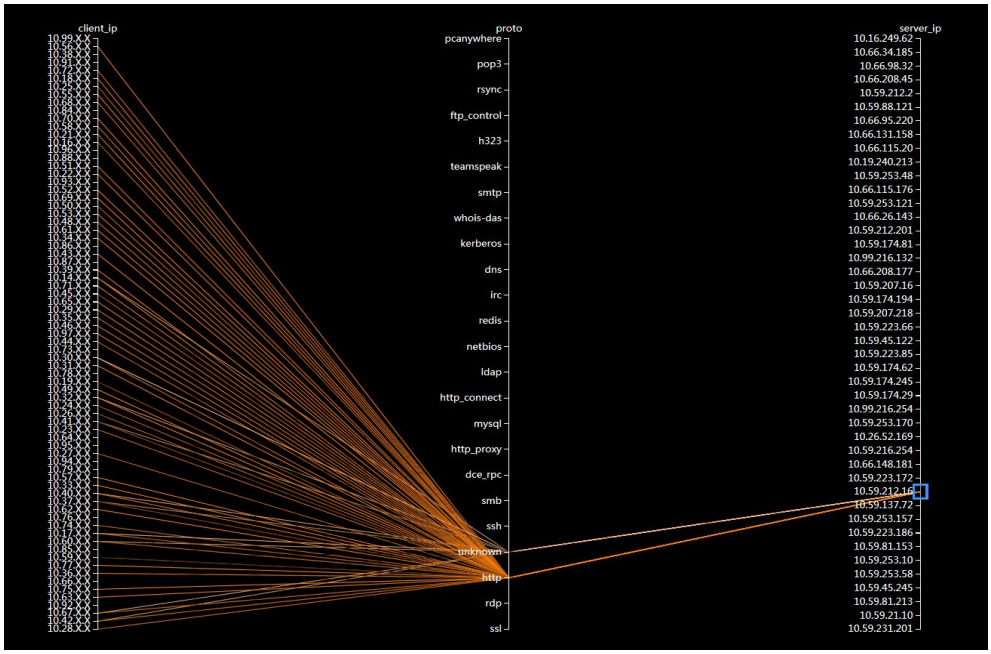

图 2.5

同理，我们可以选择其他的服务器来查看其主要的通信协议，从而判断其类型，如图 2.6，可以看出 10.59.253.170 主要通过 http_proxy 和 http_connect 来与客户端通信，可以认为它是一台代理服务器。

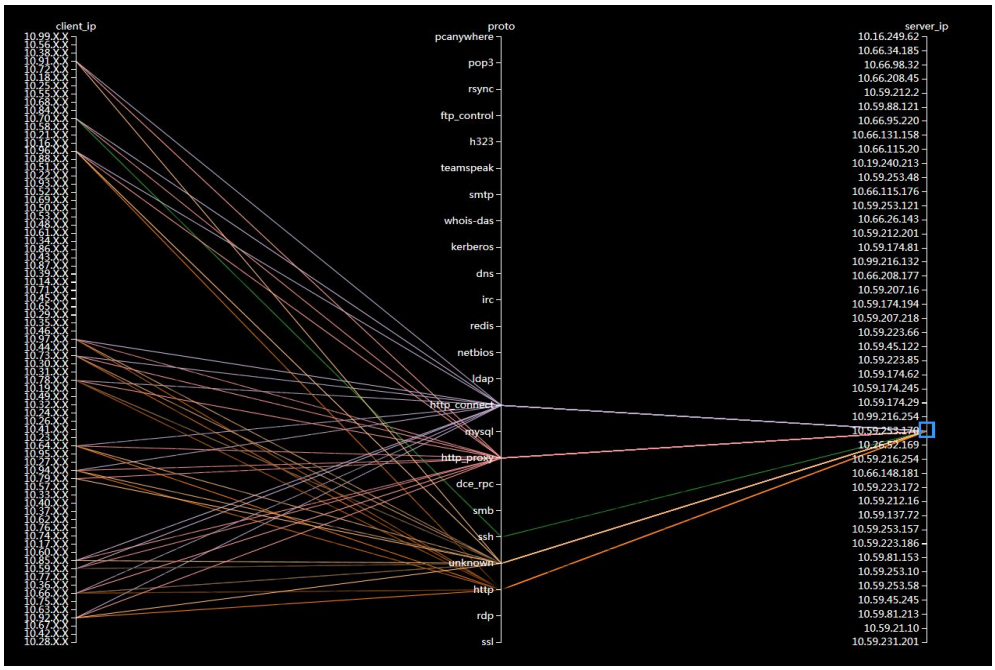

图 2.6

同样，我们可以查看到一些服务器与客户端通信的协议比较多，而且都不是常见的协议，如图 2.7。同理，我们认为它是一台特殊用途的服务器。

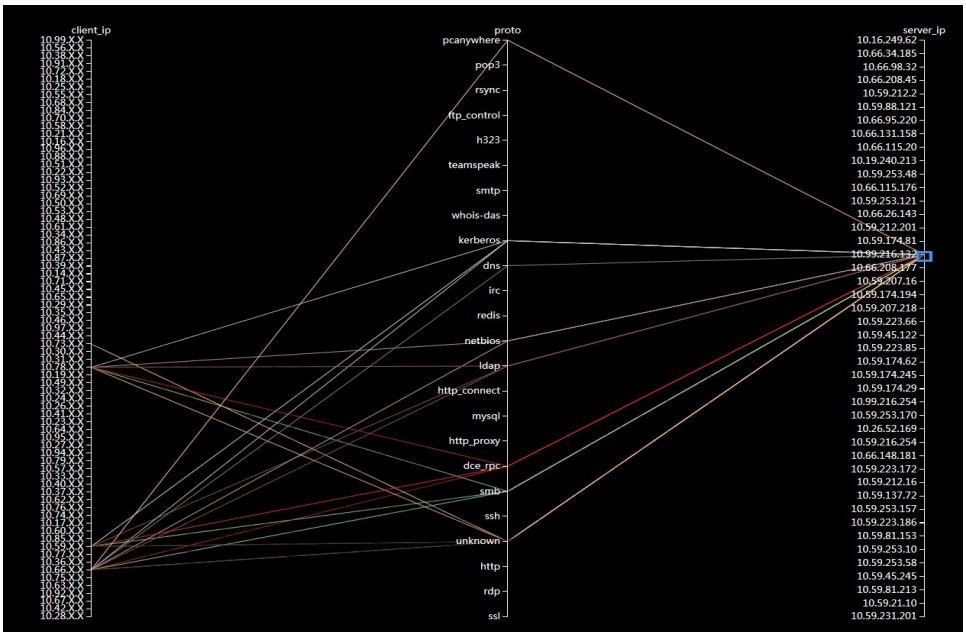

图 2.7

最后图 2.8 表示我们通过功能来对服务器进行分类的结果，使用分区图来表示，最上面一层是根节点，表示下面的都是服务器，第二层和第三层分别表示服务器的类型和服务器的 IP，颜色相同表示的是同一种类型的服务器。从图中可以看出服务器总共分为 8 类。有 web 服务器，代理服务器，dns_ldap 服务器，文件服务器，邮件服务器和数据库服务器等。

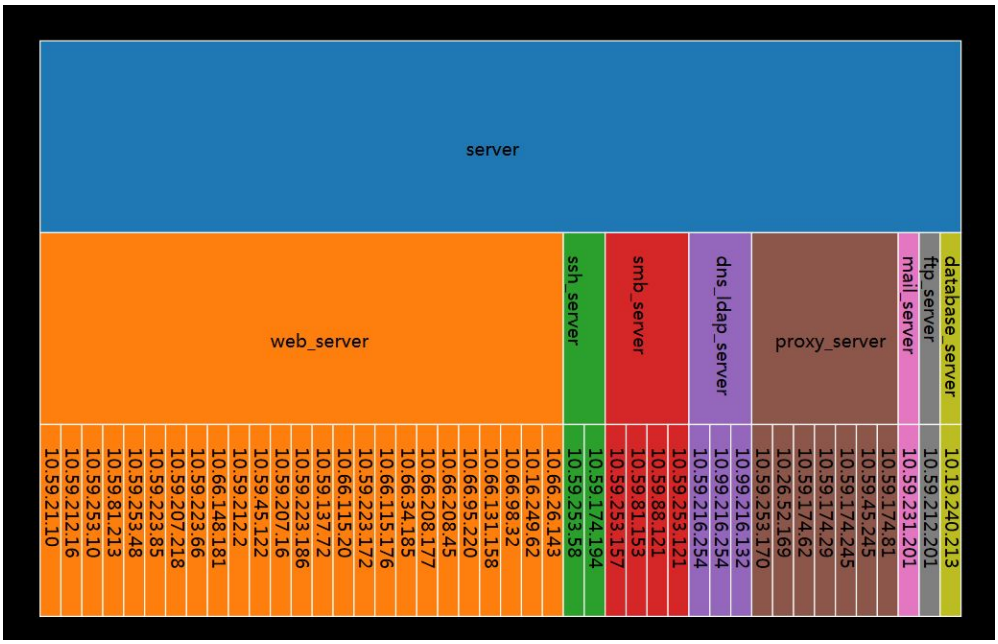

图 2.8

挑战 1.3：通过对一周的 tcpflow 数据的可视分析，说明 TSZNet 公司内部网络的“客户端-服务器”，“客户端-客户端”，“服务器-服务器”之间的常规通信模式。（请将回答限制在 1200 个字和 10 张图片）

为了发现客户端与服务器之间的通信模式，我们根据内部通信数据绘制出了平行坐标系，如图 3.1。因为客户端数量比较多，不适合使用 D3 同时绘制，所以使用 IP 段来表示,图中左边坐标轴表示客户端 IP 段。中间表示使用的协议，右边表示服务器 IP。然后根据数据，将有通信关系的客户端和服务器通过使用的协议连接。

从图中可以看出，在客户端与服务器的常规通信模式中，使用 http 协议和 unknown 协议居多，http 为常规的 web 通信协议，unknown 协议猜测可能为公司内部开发的通信协议。

此外 ssl 和 ssh 协议也使用比较多，表示客户端和服务器通过比较安全的方式进行通信。

其他使用比较多的通信协议还有 http_proxy、mysql、smb 等。

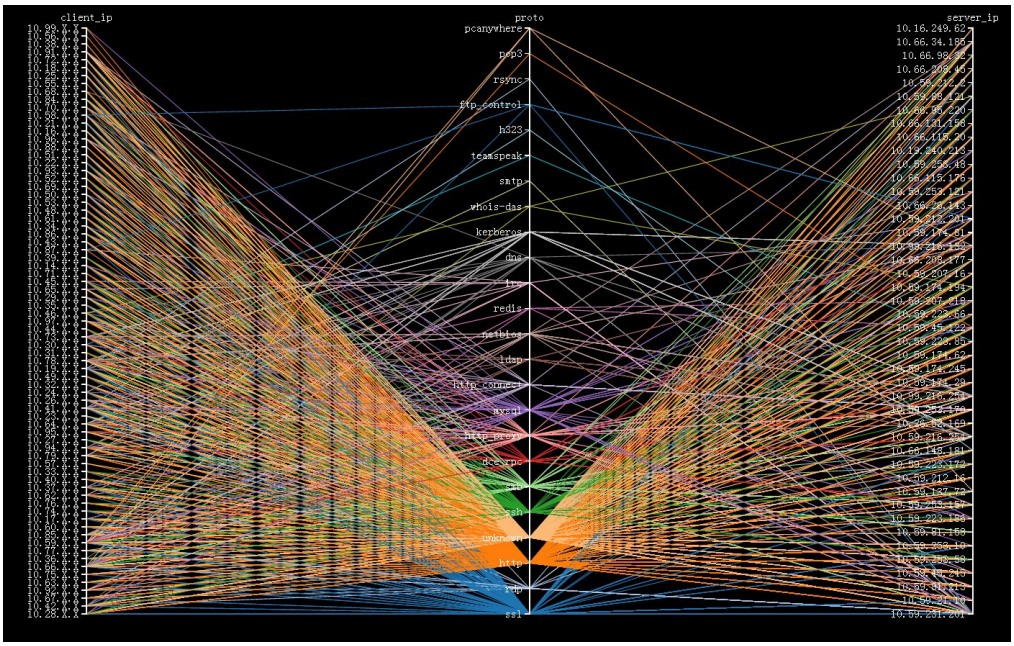

图 3.1

在图 3.1 中我们可以看出整体通信模式，当然也可以查看指定的客户端 IP 段的主要的通信模式，如图 3.2，可以选择性的查看特定的客户端通过什么样的

协议，与哪些服务器通信。如图，10.26.X.X 这个网段内的主机，主要通过 http、unknown 等协议与服务器进行通信。

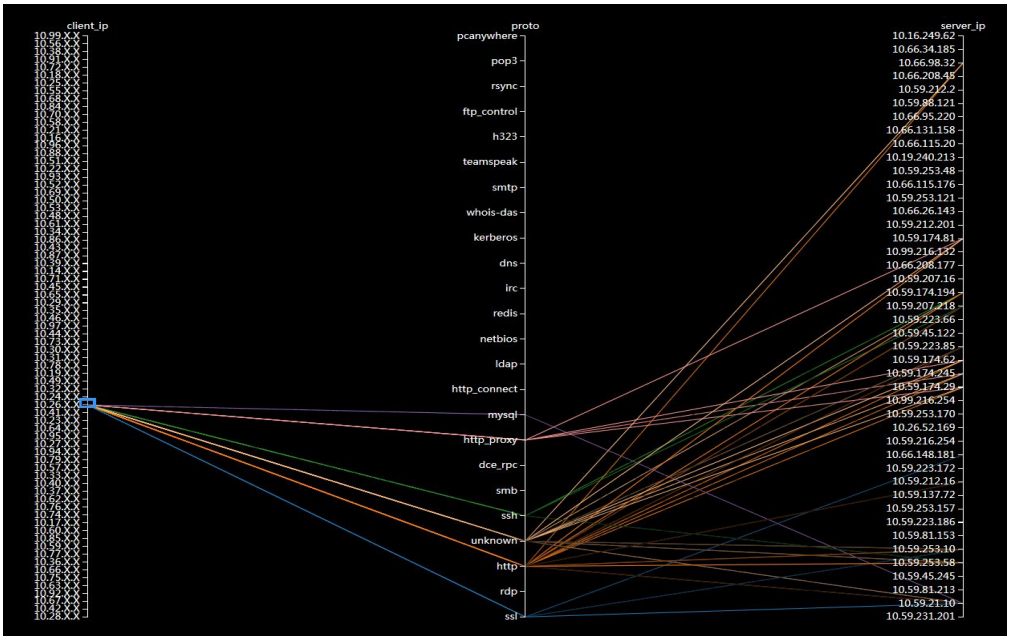

图 3.2

除此之外，我们还可以查看那些服务器和客户端通过特定的协议进行通信。如图 3.3，选择查看通过 ldap 协议通信的服务器和客户端所在的 IP 段，通过图可以看出，通过 ldap 协议与客户端通信的服务器有 10.99.216.132、10.99.216.254、10.59.216.254，通过 ldap 协议与服务器通信的客户端所在 IP 段主要集中在 10.73.X.X、10.78.X.X 等。

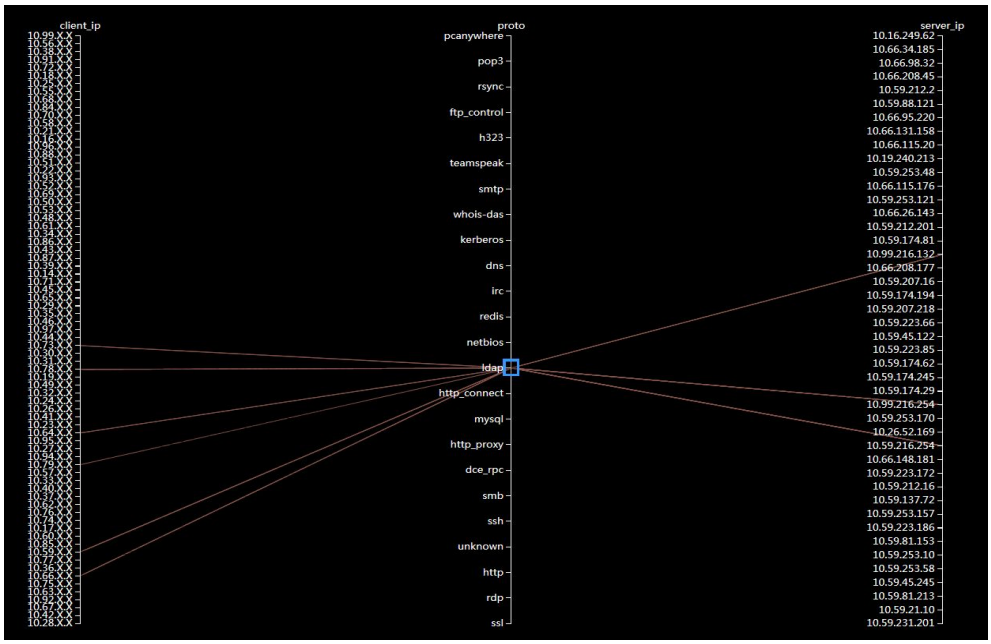

图 3.3

我们还可以选择查看指定的服务器，主要通过哪些协议与哪些客户端通信，如图 3.4，可以看出 10.59.212.16 这台服务器主要通过 http 与大多数客户端进行通信。

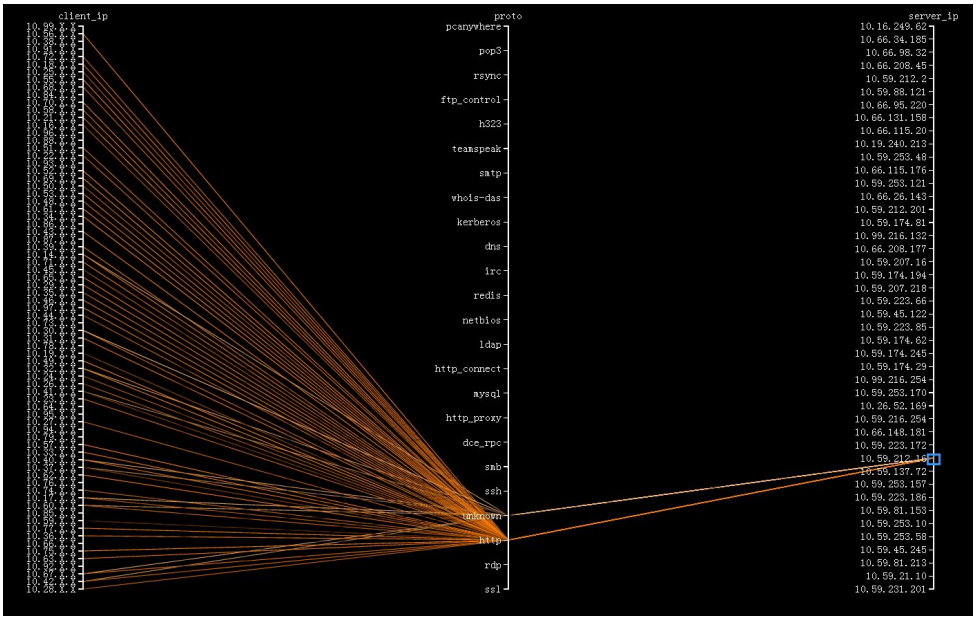

图 3.4

当然还可以同时选择多个过滤条件，以便发现更详细的通信模式，如图 3.5 所示。

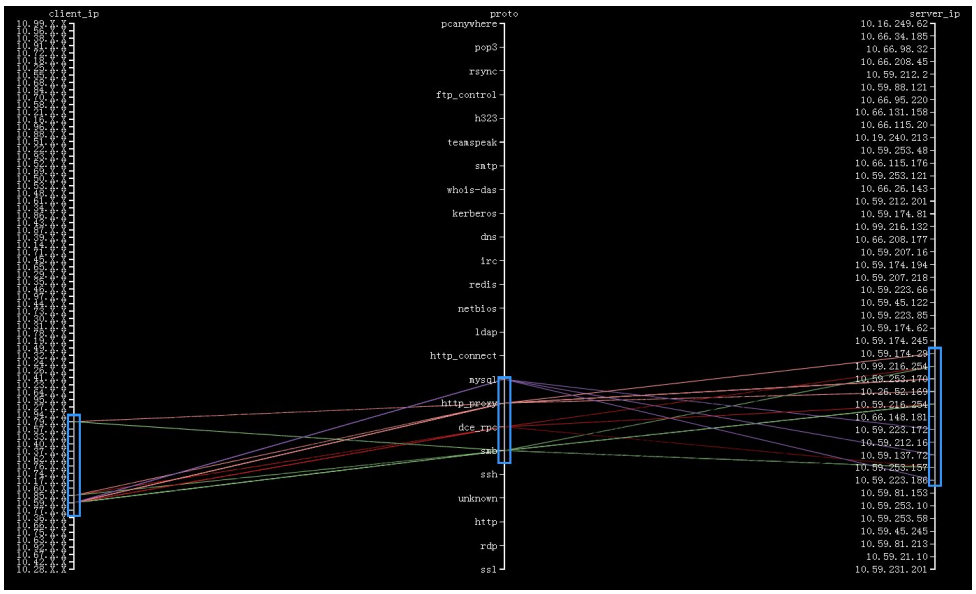

图 3.5

服务器与服务器之间通信模式，我们使用如图 3.6 来表示服务器与服务器之间的通信。从图中可以看出，服务器与服务器之间主要通过 http、ldap、dns、dec_rpc 等协议进行通信。

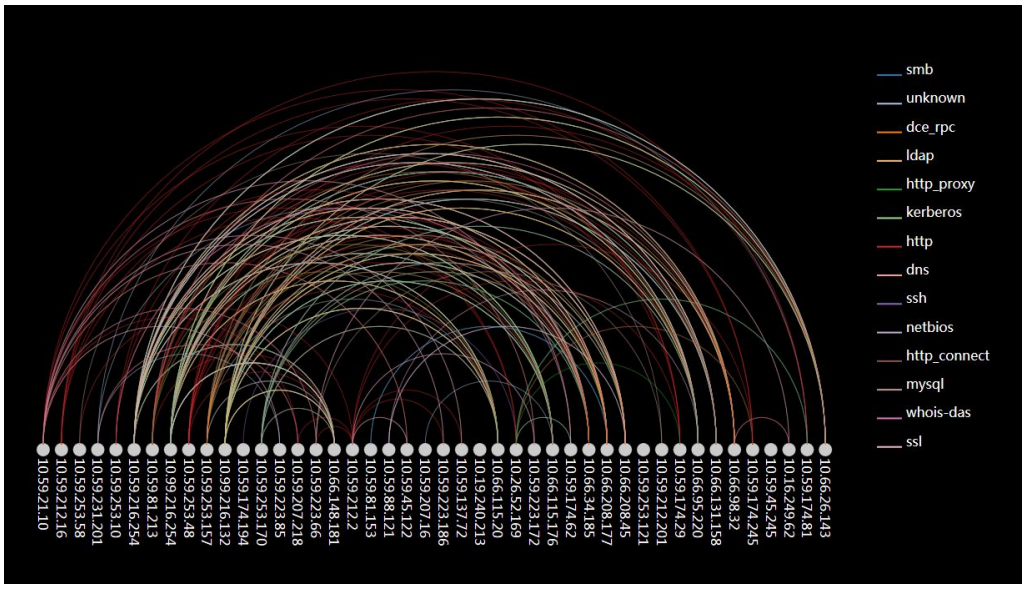

图 3.6

另外我们可以选择查看某台服务器与其他服务器主要通过哪些协议通信，如图3.7,我们可以看出10.66.95.220这台服务器主要通过kerberos、ssh、dce_rpc等协议与其他服务器进行通信。

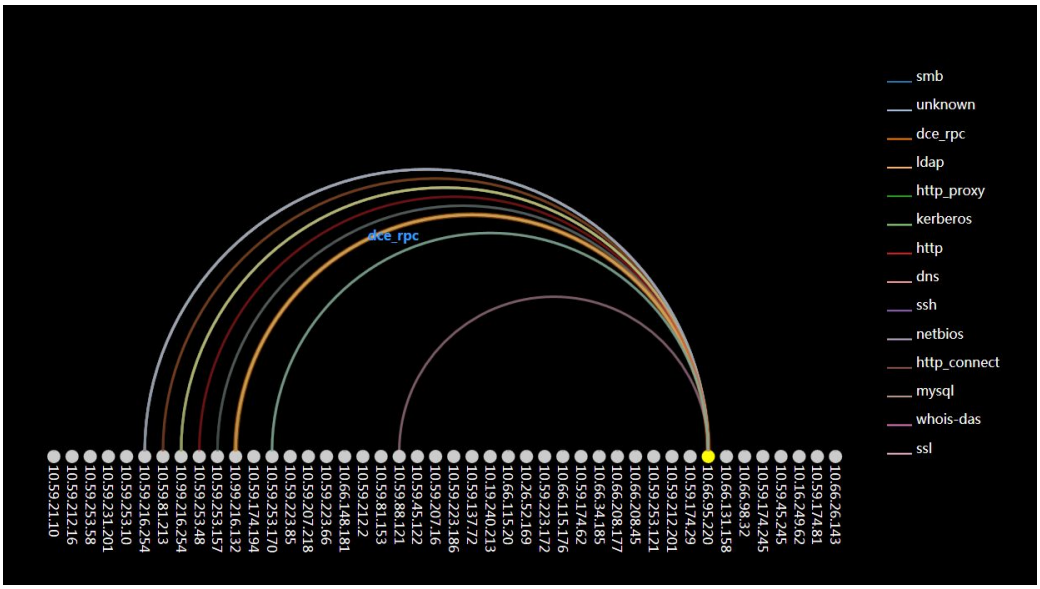

图 3.7

在发现客户端与客户端之间的通信模式，我们同样使用平行坐标的方式来展示，如图 3.8，左右两端分别表示网段，因为主机数量比较多，所以没有将所有主机一一列出。从图中可以看出，客户端与客户端之间进行通信，主要通过 unknown、http 等协议通信。此外还有 ssl、ssh 等协议。

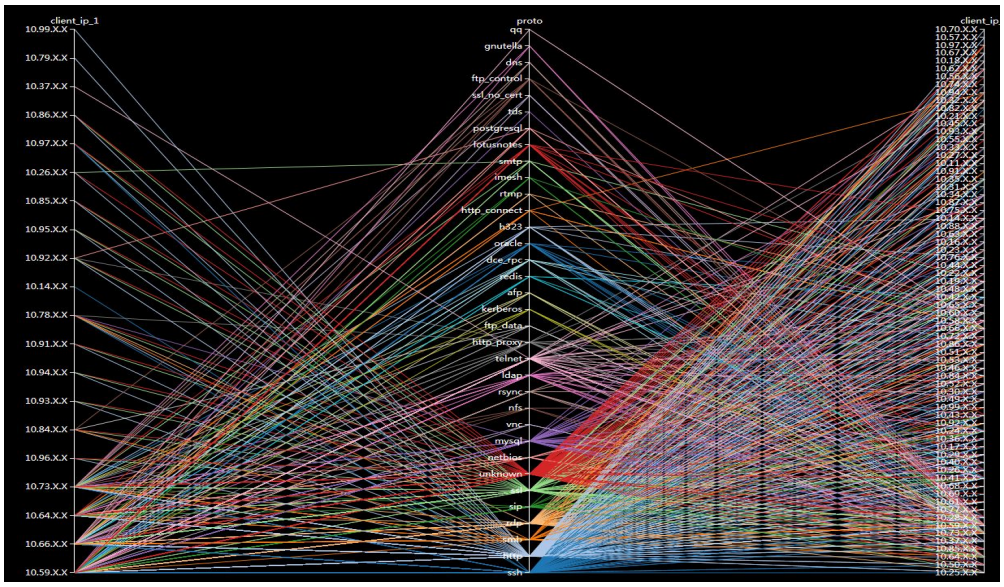

图 3.8

同样，我们可以选择查看某个网段内的客户端通过哪些方式与其他网段内的主机通信，如图 3.9，我们可以发现 10.84.X.X 这个网段内的主机主要通过 ftp_control、ftp_data、smb 等协议与 10.59.X.X 内的某台主机有通信。

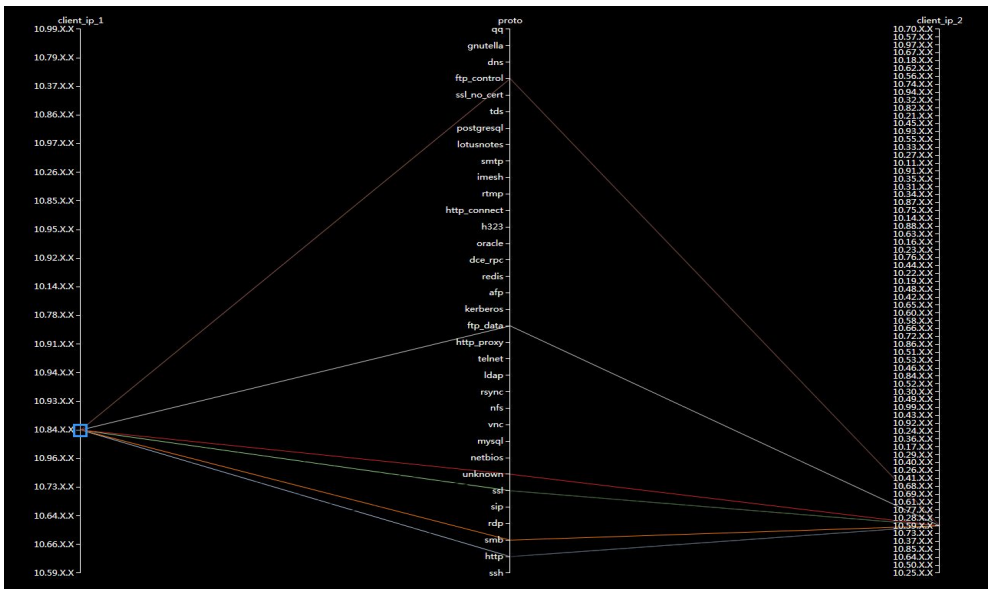

图 3.9