**2016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 2016 - mini challenge 1）**

**参 考 答 案**

（本答案仅供参考，鼓励参赛者发现本参考答案范围以外的任何有趣的故事）

（评价作品答案准确率按100分计算，四个小题的比重分别是30，30，20，20分）

**（1）找出BigBusiness公司内部网络中的客户端与服务器，并给出BigBusiness公司的网络体系结构拓扑图；对BigBusiness内部网络中的服务器进行分类，分类标准不限，比如:按照节点类型、按时间特点、按行为特点、按流量特点等等。**

**评分原则：30分**

该题目主要考核对于网络状态和网络主机行为的整体认知和分类。

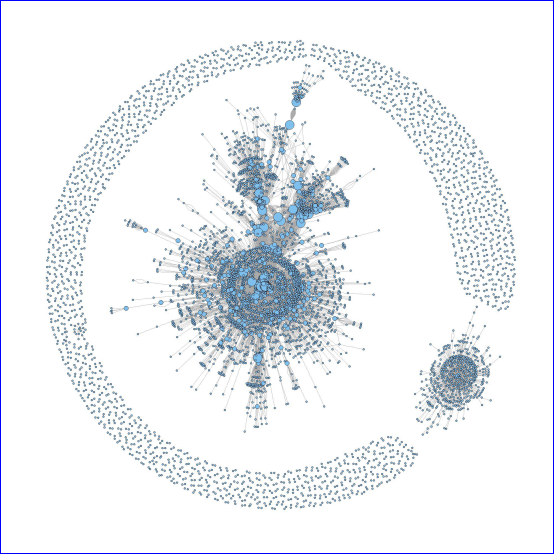
**评分标准：**

若依据每种合理的分类标准以及分类后的数据，按照连接的度给出服务器列表，得18分；若进一步给出服务器的网段分布，得分24；对每种合理的分类标准以及分类后的数据，能够找出候选的网络服务器和重要服务器，并给出每一个关键节点进行合理解释以及服务器的网段分布，如FTP服务器、数据库服务器等，得分30。

**参考答案：**

该题目的答案分为三个层次：第一层次按照连接度进行区分内网的客户端和服务器，这是最常见的一种方法，通常是设置一个门限，如连接度门限为2000或1000，而后确定出50台或200台服务器；第二层次是能够利用连接的端口连接分布或者利用其他合理解释的分类方法得到服务器和客户端；第三层次是对每种合理的分类标准以及分类后的数据，能够找出候选的网络服务器和重要服务器，并给出每一个关键节点进行合理解释以及服务器的网段分布，如FTP服务器、数据库服务器等。

1、该网络是某通信网络的体系结构，具有2万5千多台主机构成，核心网包括10网段的1个大的子网和3个小的子网构成，分支网包括192和176网段互联的子网构成，其余还有一些离散的节点构成，如下图所示。



2、公司的服务器一共分为9种，包括文件服务器、FTP服务器、RPC服务器、WEB服务器、LDAP服务器及其他服务器，下面是它们的列表

文件共享服务器 (445、139端口CIFS协议)：**10.67.220.229**, **10.67.220.221**, **10.46.236.221**, **10.39.216.247,** **10.33.216.221, 10.18.112.26,** **10.18.112.247,** **10.145.216.221**, **10.146.216.221**, **10.118.216.221**, **10.116.217.221**, **10.114.218.221等约30台服务器构成。**

FTP服务器（端口21）

**10.18.112.26**

LDAP服务器（端口389）

**10.114.218.221**， **10.116.217.221, 10.18.112.246,10.67.220.221等服务器构成，**

HTTP服务器（80端口）

10.18.112.26

常用端口和对应服务如下

|  |  |
| --- | --- |
| 端口 | 服务 |
| 445 | CIFS协议 |
| 389 | LDAP协议 |
| 135 | RPC协议 |
| 21 | FTP传输 |
| 139 | CIFS协议 |
| 80 | HTTP |
| 36100 | 未知 |
| 53 | DNS |
| 443 | HTTPS协议 |
| 88 | Kerberos 认证协议 |
| 20 | FTP认证 |
| 25 | SMTP |

3、对于服务器的分类，还可以按照活跃时段进行分类

类型1：7 月22 日-9 月12 日，全部活跃

类型2：7 月22 日-9 月12 日，有少量不活跃时段

类型3：8 月3 日-8 月10 日和9 月1 日到9 月8 日不活跃

类型4：只在8 月4 日前活跃

类型5：只在8 月10 日后活跃

类型1包括: 10.67.220.221, 10.67.216.221, 10.46.236.221, 10.39.216.247, 10.33.216.247,10.24.248.242, 10.18.112.247, 10.18.112.246,10.146.216.221, 10.145.216.221, 10.118.216.221, 10.114.218.221等服务器；

类型2包括: 10.82.112.246， 10.67.220.225， 10.65.220.221, 10.65.216.226, 10.54.216.221， 10.50.112.246等50台左右服务器；

类型3包括：10.52.206.132， 10.52.176.3, 10.52.176.250， 10.52.171.245， 10.52.140.75等20台左右服务器；

类型4包括10.24.24.206,10.88.88.206， 10.67.216.226,10.54.192.206等20余台服务器；

类型5包括10.67.220.248,10.52.63.80,10.52.140.13等10余台服务器

**（2）找出可能存在的异常通信模式（异常事件），异常标准不限，比如：IP通联的周期行为模式和活跃时段变化、某个子网访问或被访问量变化、子网连接模式变化、连接可能传输的文件类型（exe、bmp、jpg等）的变化等等，建议给出至少5种异常通信模式。**

**评分原则：30分**

该小题的回答主要分为三个层次：网络中活跃时段变化、某个子网访问或被访问量变化和连接可能传输的文件的变化均属于数量的变化，关于量的变化是容易发现的，例如某一天某子网基本没有数据，但这种异常分析得还比较粗浅；第二层次是发现细化到某个IP在某小时的流量异常；第三层次是发现流量之外的异常。下面给出的异常中1、2、3是从数据中容易得到的异常，属于第一层次；4、5、6是需要初步分析得出的，属于第二层次；7、8、9、10是需要详细分析得出来的异常，属于第三层次。总的来说，下面列举的异常仍然属于较为容易找到的异常，鼓励参赛者发掘其他异常，并适当分析这些异常之间可能存在的关联。

异常1、全网通信数据在9 月2 号缺失,第三周时全网活跃节点数偏低。

异常2、以8 月9 日为界限，在该界限之前，除了10.65.\*.\*的所有服务器均产生了大量的数据流量，但是在8 月9日之后，只有10.65.\*.\*产生了较多的流量，其余的服务器所产生流量大幅减少。

异常3、 8 月1 日之后Unknown 类型数据骤减，RPC 类型数据提高。

异常4、子网访问被访问量变化异常，8月18日这一天，跨子网的连接在显著增加，设置增加到了一个峰值，而同一子网的连接却在减小。

异常5、服务器传输数据量异常，例如IP 为10.116.160.246的服务器7 月23 日3 时流量存在异常。

异常6、该公司在近两个月时间内，在大部分时间，网络中主要传输.rpc类型的文件和.unk类型的文件；其中，7月25日-8月1日期间出现巨量.unk未知类型的文件传输，异于平时的传输规律；从图中还可以发现，.rpc文件与.unk类型文件或许是伴随传输，即传输.rpc文件的同时也会传输.unk文件，但在7月29、30两天期间异于平时传输规律。

异常7、7月26日14时到18时，主机10.118.165.198连续被访问不同端口，进一步观察得知该主机是被相同的26台主机一起访问，同样的事情还发生在7月31日11时到15时，同样的主机10.118.165.198。对比日志，统计10.118.165.198，发现它接收了惊人的流量，却没有检测到发送任何数据。说明这些主机短时间内找到10.118.165.198的开放端口，并发送垃圾流量，导致10.118.165.198瘫痪。

异常8、7月31日13:00到13:30公司内网主机”10.18.112.26”遭受到了类Ddos攻击，攻击者为10.65.216.146

异常9、在2015/7/28 9:51:59到2015/7/28 10:03:00，2015/8/18 10:38:00和2015/9/8 5:56:49这三个时间点，也就是从7月28日开始的每隔三周的周二上午有主机以10.67.216.226为源和目的IP分别发送了27M、7M、1M左右的数据，怀疑是LAND attack，但该攻击并未成功，因为发生攻击后，该主机依然正常运行。

异常10、内网每天的通信次数在6000左右，7月31日，内网的通信次数突增，是正常通信次数的4倍左右。进一步分析发现10.65和10.18网段间的通信次数在当日有明显的异常。更进一步分析，发现两个网络内通信次数异常的主要原因是10.65.216.146向10.18.112.26发起了大量的通信造成的。

**（3）找出经同一虚拟管道（VPI、VCI）进行通信的源IP和目的IP的分布规律和连接模式；找出经同一虚拟管道（VPI、VCI）承载的应用分布和变化规律；说明每一个（VPI、VCI）虚拟管道的传输数据量在不同时段的变化情况；建议至少给出5种虚拟管道通信模式。**

**评分原则：20分**

该小题主要考核链路层和网络层的关联分析能力，按照（VPI，VCI）对网络数据分类，分析每个（VPI，VCI）通道的IP地址分布和应用变化情况。

**评分方法：**

若能够按照每个（VPI，VCI）通道对IP地址进行分类分析，得分12分；更进一步分析出每个（VPI，VCI）承载的主要网段和应用类型，得分16分；若分析出某些（VPI，VCI）传输的IP具有关联性，得分20分。

（1）VPI和VCI的分布主要有

0&0、0&42

0&1174、9075&883，3181&9075，0&49、0&38

0&37、0&94、0&179、0&100、543&1699

0&2800、0&3495

0&135、0&153、0&159、0&2133、0&1684、0&1696

（2）（VPI,VCI）=(0,49)和(49,0)连接的网络拓扑具有一致性，且流量时序上是互补的，可判断（VPI,VCI）=(0,49)转换到(49,0)。类似有：

0&2800过渡到0&3495，部分主机需要更多的与web服务器进行交互，应用类型也多了富文本，图片和压缩文件等web应用。

0&135过渡到0&153，改变了部分的传输路径，减少了图像应用的传输。

0&153过渡到0&159，拓展了虚路径的传输范围，增加了传输流量和频率，增加了程序相关文件和压缩文件的传输。

0&159过渡到0&2133，缩小了虚路径的传输范围，减少了传输流量和频率，减少了压缩文件和图片文件的传输。

0&1684过渡到0&1696，添加了一个文件服务器的关系网，增加了流量。

实际每个VPI和VCI数值相反代表的虚路径，都出现时序重叠，骨干节点范围相包含，流量方向一致、应用类型相包含，但出现时间都较短，流量大小也相差甚远，可以看出是一方流量特别大时另一方出现。

**（4）设计一个可视化方法，展示数据从源到目的经历应用层、网络层和链路层的全路径过程，并实现对一次数据连接的多层次可视分析，探索其可能反应的用户行为模式。**

**评分原则：20分**

此问题重点考察可视化设计能力。强调多层次的协同分析，希望能够在设计上给出链路层、网络层和应用层相结合的视图，而多个独立或简单brush联动的树图和节点连接图。

**评分标准：**

使用可视化视图显示链路层、网络层和应用层的关系结构和结论信息：得10分；使用视图联动的模式，得12分；在一张视图中综合链路层、网络层和应用层因素，得16分；给出实用导向意义的结论，得20分。