

# 多功能 IP 搜索

嵇天颖<sup>1)</sup> 张子薇<sup>2)</sup> 禹含<sup>3)</sup>

<sup>1)</sup>(计算机科学与技术系, 计 64, 2016010308)

<sup>2)</sup>(计算机科学与技术系, 计 62, 2016011276)

<sup>3)</sup>(计算机科学与技术系, 计 74, 2016010609)

**摘 要** 在本次课程大作业中, 我们使用 nmap 对清华大学的 IP 地址进行扫描, 并结合 censys 的开放 api 获取到的数据, 对清华大学中响应 IP 运行的 http、ssh、ftp 等常见服务进行分析和可视化。我们重点根据各个服务版本信息, 分布特点, 讨论了可能的原因, 并针对版本问题分析了可能出现的安全隐患。最后我们根据采集到的数据建立了前端网站, 用户可以通过 IP 检索对应的服务, 同时可以通过网络空间资源图谱根据服务检索响应的 IP 列表, 为相关资源的检索带来便利。

**关键词** IP 测绘, 资源图谱, 服务分析, 安全分析, 多功能 IP 搜索

## Multi-function IP search

Tianying Ji<sup>1)</sup> Ziwei Zhang<sup>2)</sup> Han Yu<sup>3)</sup>

<sup>1)</sup>(Department of Computer Science and Technology, Class 64, 2016010308)

<sup>2)</sup>(Department of Computer Science and Technology, Class 62, 2016011276)

<sup>3)</sup>(Department of Computer Science and Technology, Class 74, 2016010609)

**Abstract** In this course, we use nmap to scan the IP address of Tsinghua University, and combine the data obtained by the open API of censys to analyze the common services such as http, ssh, ftp, etc. in response to IP running in Tsinghua University. Visualization. We focus on the various service version information, distribution characteristics, discuss the possible reasons, and analyze the possible security risks against the version problem. Finally, we built a front-end website based on the collected data. The user can retrieve the corresponding service through IP, and at the same time, it can retrieve the IP list of the response according to the service through the network space resource map, which facilitates the retrieval of related resources.

**Key words** IP mapping, resource mapping, service analysis, security analysis, multi-function IP search

## 1 引言

随着网络的进一步普及化的复杂化, 网络空间中的资源呈现爆发性的增长。一般认为, 网络空间资源是对于网络空间中存在的万物的总称, 既包括网络空间中存在的路由器等硬件, 也包括网站、ftp 等服务器端的软件, 内容庞杂。为了便于进一步统计和管理网络空间资源, 对网络空间资源进行合适

的分类变得极为重要。当前, 学术界尚未能够对于网络空间资源的分类方法达成共识, 也尚未有一种得到公认的分类标准, 因而, 这一方向存在较大的研究价值。

网络空间资源的分类需要遵循一定规律, 即对于网络空间万物需兼具正交性、可用性、完备性等特点, 能够将网络空间资源能够以有限种类别进行全覆盖式的分类。

为了便捷有效地获取网络空间资源和 IP 的映

射关系以及 IP 资源的详细信心,我们搭建了多功能 IP 搜索平台。

本次研究聚焦于中国教育与科研计算机网 (CERNET) 中清华大学的 IP 地址测绘与数据分析。我们将网络空间资源进行分类,主要分为四类:基础设施,数据资源,应用服务,虚拟主体。为了直观的了解网络空间资源,我们利用了资源图谱 (圆饼图) 来实现网络空间资源模型的可视化,并在多功能 IP 搜索平台上实现了网络空间资源节点与 IP 的映射连接关系。

我们对 CERNET 中清华大学网段数据进行了分析,从 IP 响应、服务类型和服务版本等诸多方面进行了服务分析与安全分析。通过对数据分析,我们可以了解到清华大学信息化水平。

## 2 数据测绘

### 2.1 测量范围与工具

我们首先调查了清华开放使用的 6 个 B 类地址段:

166.111.0.0/16
59.66.0.0/16
101.5.0.0/16
101.6.0.0/16
183.172.0.0/16
183.173.0.0/16

为了节约获取数据的时间,我们选取了其中一个地址段 59.66.0.0/16 作为主要的测量范围。这一段中以学生宿舍楼居多。

本次研究采用 nmap 获取清华网 59.66.0.0/16 网段的全部实时信息。nmap 作为一种高效的网络连接端扫描工具,能够充分扫描服务器开放的网络服务端,并确定运行在各自连接端的服务。由于 nmap 提供了极为方便快捷的服务器信息获取方式,能够使我们在有限的时间内获取足够多的可供分析的信息,因而在本次研究中关于网络空间资源获取方面所采用的主要方法即为 nmap。

### 2.2 数据测量过程

起初,小组的一名成员在自己的电脑上使用扫

描命令: `nmap --host-timeout 60s -oX nmapinfo.xml 59.66.0.0/16`,但却得不到有效的结果:扫出来的结果中,所有有响应的 IP 均没有开放端口。加上 `-p` 参数,单个端口逐个扫描,却能得到有效的结果。考虑到一次性扫描多个端口会有超时的可能,修改了一 `host-timeout` 参数,但并没有改变。最后,小组的另一名成员在自己电脑上输入了相同的扫描命令 `nmap --host-timeout 60s -oX nmapinfo.xml 59.66.0.0/16`,却得到了有效的结果。可见 nmap 的稳定性的确比较差。

正式的测量历时约 1 个小时,在清华网 59.60.0/16 网段搜集到的有响应的 IP 有 3338 个,之后的分析主要以这些 IP 为样本。

### 2.2 未能找到的网络资源原因分析

即便仅仅选取了相对更容易稳定获取的清华校内网段 IP 作为研究对象,但在搜集这部分网络资源的过程中,仍然存在部分资源无法找到的情况。我们推测的主要原因包括:

① 没有主机。寻找到的资源可能是一个空 IP,因而无法返回符合要求的服务器信息;

② 访问权限问题。有些服务器没有完全开放,因而存在对于访问权限的限制,这一点其实是比较普遍的,显然很多服务器是不能向所有学生开放的。因而访问权限问题同样阻碍了获取全部网络资源信息;

③ 防火墙问题。即便是在清华校内,例如很多实验室服务器仅面向自己实验室的内网开放,对于实验室以外的网络设置有防火墙,本小组难以获取这些服务器的网络资源。

### 2.2 测量遇到的问题

在本次研究过程中,也存在一些难以避免的问题,在研究结束时仍然没有获得理想的解决方案。主要遇到的问题包括:

① nmap 稳定性不足。nmap 会存在跑网站不稳定的情况,为了提高数据搜集的效率,在所使用的爬取命令中将时间上限设为 60s。而在这种情况下, nmap 在获取一些服务器资源的过程中,仍然存在超时的情况,且有些时候超时原因并不明确;

② 执行条件和服务器状态不稳定。由于执行条件的不稳定,导致返回结果存在很大的差距。如在实验中,同样的指令,分别扫描 IP 和所有 IP 一

起扫描所获得的结果经常完全不同; 同样的指令, 一次性扫描所有端口和单个端口逐个扫描的结果也不一样; 甚至, 在不同电脑、不同网络下, 相同的命令扫描的结果也不一样。

这些实验问题的出现导致网络资源的获取并不像预想中顺利; 且一些问题出现原因不明也影响了获取的服务器信息的不稳定。

### 3 可视化网络空间资源模型

#### 3.1 网络空间资源分类简介

我们想展示的网络空间资源, 是在网络空间中, 使用网络空间手段, 能够探测和感知的 IP 化实体。从资源类别的角度来看, 一般而言, 我们将网络资源分为基础设施、应用服务、数据资源、虚拟主体四大类。

这四大类资源又可以进一步细分, 例如基础设施可以进一步分成自治域、网络、中间节点、终端节点、链路这几类, 这其中就有常用的路由器、交换机等等; 应用服务类则可以进一步分成有机服务和无机服务, 常见的各种协议, 如 http、smtp 等都隶属于无机服务; 而各种脚本则属于数据资源的代码类……

这样按类别细分之后, 更有助于理解抽象的网络空间, 更有利于增强我们对网络空间资源图谱的认识和把控, 为我们将网络空间资源可视化打下了基础。

#### 3.2 网络空间资源图谱架构

在加强了自己对网络空间资源的理解之后, 为了更好地向其他人展示网络空间资源, 我们需要将网络空间资源可视化。

可视化有很多途径, 诸如绘制地理坐标系、绘制拓扑图、绘制 Hilbert 的 IP 二维空间。

对于资源分类这样有层次结构的分类, 最常用的可视化手段是树状图。但是在网络空间资源展示的时候, 树状图存在着问题: 网络空间资源有非常多的子类, 若放在树状图中, 会显得树叶部分极其拥挤, 很不美观, 展示效果比较差。我们想使用更新颖、更美观的可视化手段来展示网络空间资源。

为此我们决定采用大作业布置说明中推荐使

用的圆饼图。圆饼图由许多层圆构成, 每一层圆代表一个分类的层级, 一个圆上分布着多个节点, 每个节点代表一个子类。例如最中心的节点是“网络空间资源”, 向外半径最小的圆上有 4 个节点, 分别是“基础设施”“应用服务”“数据资源”“虚拟主体”, 再向外半径第二小的圆上有更多的节点, 分别是上述 4 个节点的各个子类。而具体某一层的各个节点处在该层圆的什么位置, 是由简单的力学模型确定的: 将各个节点向外施加的排斥力视作和距离以某种关系 (反比、平方反比等) 负相关, 不断迭代得到一个近似平衡的位置。这样的圆饼图, 在新颖度和美观度上都优于普通的树状图。

将圆饼图嵌入我们的搜索系统中, 我们预期的效果是: 悬停到某一节点上时, 该节点会放大; 点击该节点, 可以返回包含该节点对应的资源的 IP 地址, 再点击 IP 地址则可以查看该 IP 下开放的资源的详细情况。这样一来, 用户想要查看某一种资源的使用情况时, 直接按照资源分类寻找到对应的节点即可。

#### 3.3 网络空间资源图谱代码实现

首先将分类关系按一定的格式写成 csv 文件, 用 R 读入, 进行相关的处理之后, 再导入 R 语言中非常流行的一个可视化工具包 networkD3, 用 networkD3 中的 radialNetwork 函数, 即可画出我们需要的圆饼图。

```
library(networkD3)

rsplit = function(x){
  x = x[!is.na(x[,1]),,drop=FALSE]

  if(nrow(x)==0) return(NULL)
  if(ncol(x)==1) return(lapply(x[,1], function(v) list(name=v)))

  s = split(x[,-1, drop=FALSE], x[,1])

  unname(mapply(function(v,n) {if(!is.null(v)) list(name=n, children=v) else list(name=n)}, lapply(s, rsplit), names(s), SIMPL
```

```

IFY=FALSE))
}

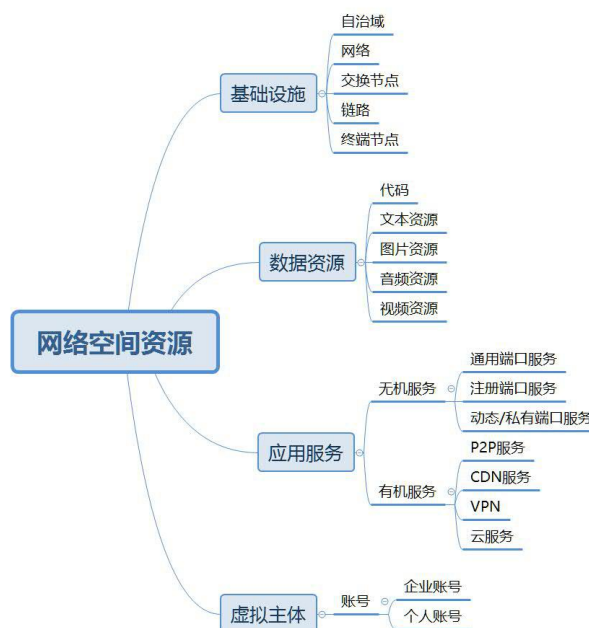
raw = read.csv("rc.csv", stringsAsFactors
= F)
raw[raw == ""] = NA
re = rsplit(raw)[[1]]

radialNetwork("rc.csv", List = re, fontSiz
e = 8, nodeColour = list("lightblue"))

```

### 3.4 网络空间资源图谱展示与网站资源图谱功能

我们对网络空间资源图谱的划分如下（图 1）：



下图（图 2）展示了我们所绘制的网络空间资源图谱，

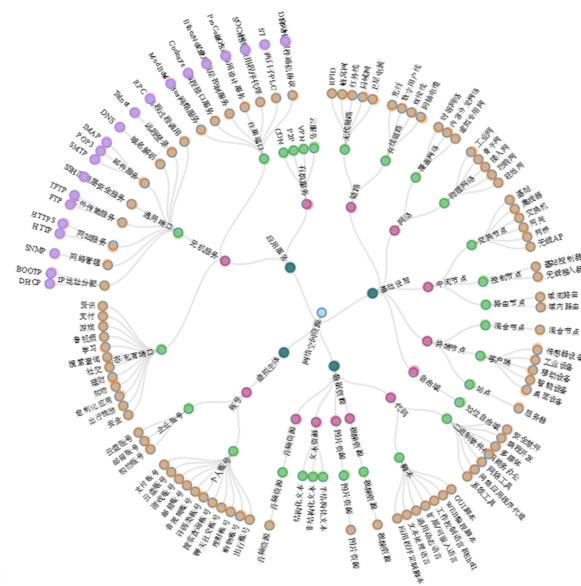


图 2 可视化网络空间资源图谱（圆饼图）

同时我们在所实现的“多功能 IP 搜索”网站上，实现了这张图。点击图中叶子节点，可以获取使用改资源服务的 IP 地址列表和 IP 地址的详细信息。详细信息可以见“前端展示和操作”部分。

### 3.4 资源图谱绘制的困难与不足

① 最初我们打算阅读数据可视化领域的相关资料，手动实现圆饼图中的静力平衡模型的可视化布局算法，但由于时间原因，最终还是直接使用了 R 的现成的包进行绘图。

② 在实现圆饼图的交互时，由于是直接 R 画图再嵌入 html 中，所以无法在前端直接实现交互；而各个节点的实际坐标也未知。故最后使用了比较笨的“硬编码”的方法，用其他工具逐个手动提取出圆饼图中各个节点的坐标，直接把坐标的具体数字写入 html 中。这一过程比较耗时耗力。作为样例 demo，目前只实现了 snmp、ssh、http、https 四个节点。可能的解决方法有：参考当前使用的 networkD3 的 radialNetwork 源码中的算法，不使用 R，而是在 web 前端直接手动实现相应的算法，将对应的节点设为可以直接响应鼠标点击事件的类型；或者通过图形学的一些算法，自动识别出圆饼图中各个节点的坐标。

③ 由于访问权限、防火墙等的存在，许多资源无法通过直接的主动测量获取，实际能直接获取的资源比较有限，所以圆饼图中很多节点其实是很难实现的。再加上 nmap 本身的不稳定性，能获取

到的部分类型的资源也很可能不全。想要解决这一问题, 只能采用更高级更稳定的测量方法, 而且拥有比较高的权限, 才能获得好的相对完整的数据。

综上所述, 当前做出来的圆饼图和资源搜索功能仅能作为一个 demo, 可能并没有多少实际应用的价值, 但倘若拥有足够丰富完整的数据, 再按照上文所述的方法进行改进, 则可以成为之后展示网络空间资源的一种思路。

## 4 网络 IP 资源数据分析

我们希望通过扫描全校的在线主机, 得到在不同端口开放的网络服务, 最终对全校的网络服务使用情况得到一个总体的认识。

在成功获取 IP 资源的基础上, 为进行数据分析, 我们对所获得的数据进行了合理的分类。分类标准需要符合正交性、可用性、完备性等特点, 尤其是正交性特点, 即所有分类需要涵盖全部信息; 基于这一点, 我们考虑的几种分类形式均为排他性或采用完全对立的指标。

### 4.1 IP 数据分析指标

基于端口状态。将获取的端口状态根据返回信息分为 open、closed、filtered 三类, 覆盖全部端口的状态情况;

基于网络设备的端口空闲程度。为了评价网络设备的端口空闲程度, 我们所衡量的指标为服务器所开放的端口数量, 在对我们获取的资源进行分析之后, 我们将边界值定为 5 个, 即当开放的端口数大于或等于 5 个时, 我们定义这个网络资源为忙碌资源, 相应的, 当开放的端口数小于 5 个时, 我们定义这个网络资源为空闲资源;

基于网络设备的服务名称。网络设备所提供的服务有各式各样的名称, 仅我们搜集到的数据而言, 网络设备的服务名称就达上百个;

基于服务版本信息。通过对端口服务版本信息的分析, 我们可以判断端口对应的主机类型, 是否存在版本老旧的安全风险, 由此可以分析清华大学的信息化程度。

### 4.2 基于端口状态的分析

在我们测绘的 65536 个 IP 中, 有响应的有 3339 个 (也就是被收集到 nmapinfo.xml 中的 IP 数量), 占 5.09%。有开放端口的 IP 数量为 1677 个, 占有响应 IP 地址的 50.22%。

扫描到的端口总数为 6899 个, 端口状态共 3 种: open、filtered、closed。分别占比约 53.3%、46.2%、0.5%。从图 3 中我们可以直观的看到三种状态的占比, open 状态和 filtered 状态占绝大多数。

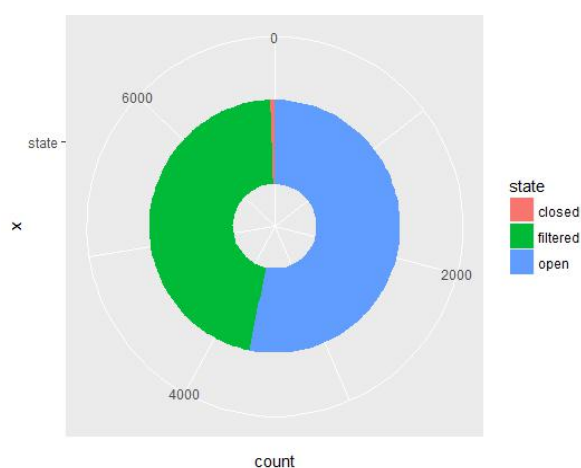


图 3 IP 端口状态分析图

### 4.2 基于网络设备的端口空闲程度的分析

在测绘所得的数据中, 有两个 IP 的端口数量非常多, 超过了 1000 个: 59.66.109.9 和 59.66.200.76。下图 (图 4) 是去掉了两个有 1000 多个端口的异常 ip 之后, ip 的端口数的柱状图。

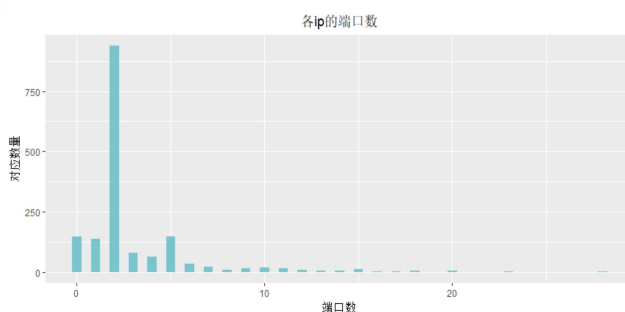


图 4 各 IP 开放的端口数量柱状图



我们可以发现 IP 地址开放的端口数量集中在 5 个以内, 开放超过 10 个端口的 IP 十分稀少。开放数量最多的是 2 个, 基本上都是 telnet 和 snmp。

在统计分析中, 我们发现绝大部分分布在 59.66.200.0 - 59.66.250.255 范围内。通过与地理地址的映射, 我们发现这部分 IP 位于 W 楼博士生宿舍。根据实际情况判断, 是因为博士生时常需要在实验室远程登录放在宿舍的台式机, 所以 telnet 和 snmp 格外地多。

通过对清华大学各 IP 地址开放的端口数量分析, 我们发现清华校内的网络资源比较富足, 从端口的角度看, 空闲比例相当高。

### 4.3 基于网络设备的服务的分析

在校内活跃着很多主机, 他们为不同范围的用户提供者各种各样的服务。比如为实验室提供 Gitlab, wiki 等 http 服务, 提供数据库服务, 提供远程 ssh 连接服务, 或者提供网络打印服务; 或者比如为系内甚至全校提供报名, 审批等 http 服务。通过这些服务我们可以推测出主机的运营者是如何通过网络服务用户的工作和生活的。

在过滤掉出现次数不超过 50 次的服务后, 我们绘制了下图 (图 5) 来反应我校网络设备的服务分布信息:

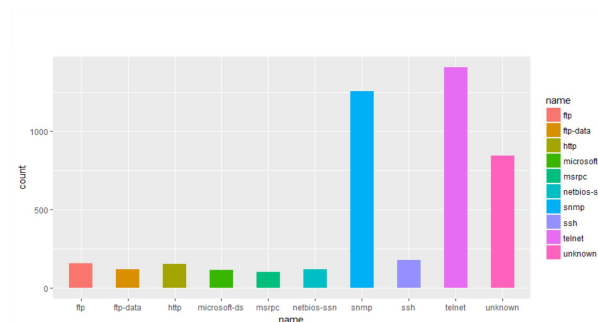


图 5 端口服务类型分析

统计扫描到的端口服务发现, 我们所扫描到的端口服务共 746 种, 数量最多的是 telnet 和 snmp, 均超过了 1000 个, 其他的服务 (如 ftp, ftp-data, http, ssh 等) 数量均在 100~200 之间, 同时还有一定量的未知服务信息, 在测绘中无法获取这些端口的服务类型。

### 4.3 基于服务版本信息分析

#### 4.3.1 HTTP 类服务版本分析

我们使用 nmap 进行对端口进行服务版本信息的探测与扫描。我们先对使用了 HTTP (包含 HTTPS) 协议的服务器和主机信息进行分析。我们将统计分析结果汇总为下面三个图 (图 6, 图 7 和图 8)。

其中图 6 中展示了 HTTP 服务 (不含 HTTPS 服务) 的端口数量分布, 会发现因为

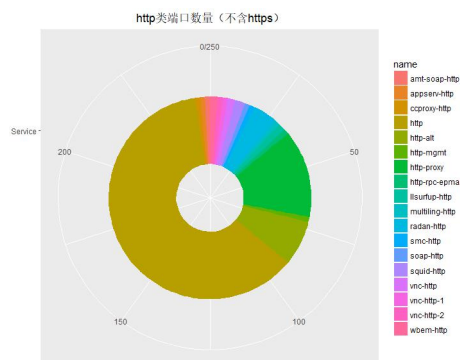


图 6 HTTP 类(不含 HTTPS) 端口数量分布

图 7 展示了采用 HTTPS 协议的端口数量分布:

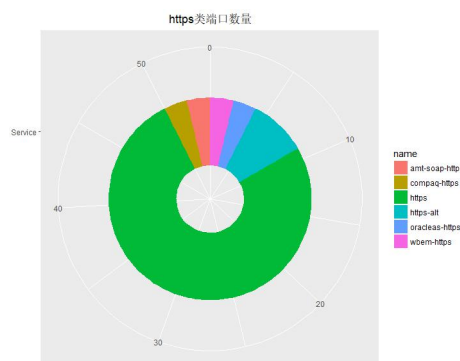


图 7 HTTPS 类端口数量分布

图 8 我们对两种协议信息进行了汇总, 图中明显可以看出, HTTP 服务占比超过了一半, 再加上其他的 HTTP 类的服务, 总数远高于 HTTPS 类。



种加密的网络传输协议,可在不安全的网络中为网络服务提供安全的传输环境。SSH 通过在网络中创建安全隧道来实现 SSH 客户端与服务器之间的连接。虽然任何网络服务都可以通过 SSH 实现安全传输,SSH 最常见的用途是远程登录系统,人们通常利用 SSH 来传输命令行界面和远程执行命令。使用频率最高的场合类 Unix 系统,但是 Windows 操作系统也能有限度地使用 SSH。

图 12 展示了 SSH 服务的版本,openssh 占据了完全的统治地位。OpenSSH(OpenBSD Secure Shell)是使用 SSH 透过计算机网络加密通信的实现。它是取代由 SSH Communications Security 所提供的商用版本的开放源代码方案。

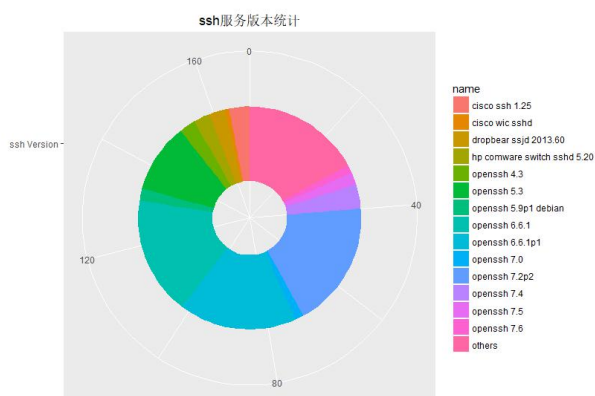


图 12 SSH 类服务版本统计

但是我们可以看到大部分 openssh 版本都比较旧,几乎没有使用最新 openssh 7.9(2018-10 发布)的,甚至还有在使用 openssh 4.3(2006-02 发布)。过于老旧的版本会带来安全隐患,由于 SSH 是用于加密网络传输的,因此针对于破解 SSH 的攻击工具也层出不穷,针对老旧的 SSH 版本的攻击工具早已出现,再使用老版本的 openssh 的安全隐患非常大。

#### 4.3.4 nmap 探测获取服务版本信息的局限性

使用 nmap 进行服务版本信息扫描也会出现一定的问题:使用 nmap-services 文件可以识别知名的 2200 个服务,nmap-services 文件主要基于端口识别。但是这种识别是不可靠的,即使是正确的这些信息也不可靠,了解运行在端口上的服务版本更重要,这有助于确定服务易受攻击的类型。这就需要服务和版本探测,来探测具体的信息。Nmap 通过特定的探针来了解更多的数据,Nmap 通过读取 socket 中的数据进行匹配。但从我们探测到的数据

中也可以发现,很多服务无法获取具体版本信息。

#### 4.4 IP 资源总结

通过对 CERNET 清华大学网段的 IP 地址的资源分析,我们发现我校的 IP 资源和 IP 的端口资源是比较丰富的。在与北京大学进行对比中,发现我校相对来说信息化水平更高,对网络信息安全更为重视。但单独考虑我校的各种服务版本信息,我们仍能发现诸多问题,HTTPS 服务占比明显低于 HTTP 服务。同时绝大部分采用 FTP 的明文文件传输协议,而鲜少采用 FTPS 服务,安全性明显不足。与此同时,用作加密传输的 SSH 服务中采用的 openssh 服务版本基本都是偏低的,几乎没有用最新版本的 openssh 的。而旧版本的 openssh 攻击工具比比皆是,安全风险仍然不容忽视。

## 5 多功能 IP 搜索平台

### 5.1 多功能 IP 搜索平台介绍

为方便对采集到的信息进行检索、分析,我们搭建了前端用于展示信息。用户可以直接通过 IP 对某地址所开放的服务进行检索,也可以根据服务查看对应的 IP 信息。前端主界面如下

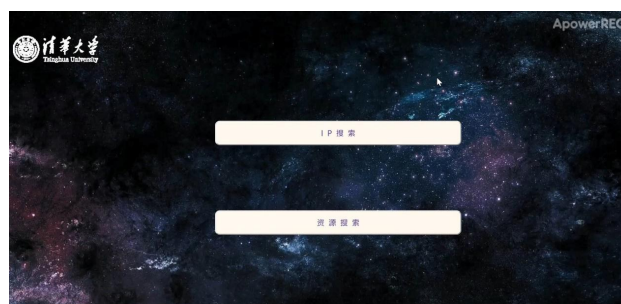


图 13 多功能 IP 搜索平台主界面

### 5.2 前端操作与展示

#### 5.2.1 IP 搜索功能

IP 搜索可根据用户输入的 IP 查询该 IP 所开放的端口信息。点击 IP 搜索按钮后,即可进入 IP 搜索页面,在文本框中输入相应的 IP 即可完成搜索。

下图展示了 IP 搜索功能的搜索界面:



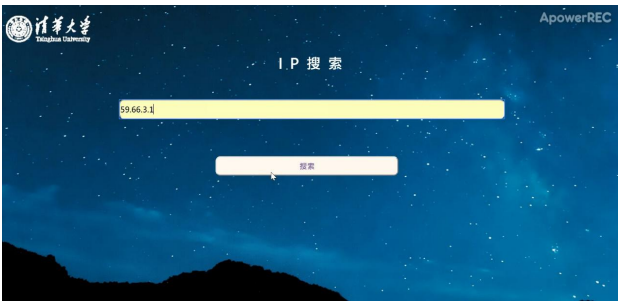


图 14 IP 搜索界面

搜索成功后可得到该 IP 开放的端口信息, 如图



图 15 IP 搜索成功结果界面

搜索失败后有相应提示信息, 如图



图 16 IP 搜索失败结果界面

5.2.1 资源搜索功能

点击资源搜索按钮后, 即可进入资源搜索页面。资源搜索可针对服务类型反向检索对应的 IP 信息。我们将提供的各种服务制成一张网络空间资源图谱, 用户可按层次检索相应的服务。

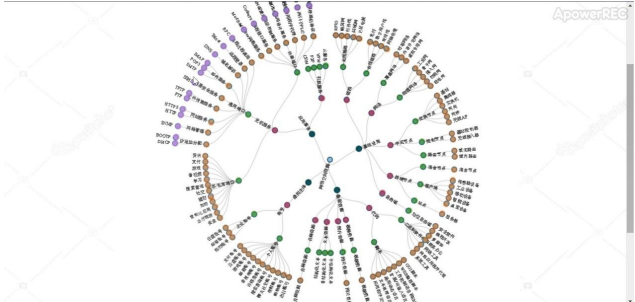


图 17 资源搜索圆饼图界面

用户只需点击网络空间资源图上的任意结点即可获取提供该服务的 IP 列表



图 18 资源查询结果界面

直接点击 IP 列表中的任意 IP 即可跳转到该 IP 的详细信息界面。



图 19 IP 详细信息界面

## 参 考 文 献

- [1] Gordon Fyodor Lyon. Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning. Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning, 2009.
- [2] Alois Ferscha, and James Johnson. N-MAP: an environment for the performance oriented development process of efficient distributed programs. Future Generation Comp. Syst., pp. 571-584, 2000.
- [3] Alois Ferscha, J Johnson. Performance prototyping of parallel applications in N-MAP. international conference on algorithms and architectures for parallel processing, 1996.
- [4] Robert M. Colomb. Information spaces: the architecture of cyberspace. Information. Spaces: the architecture of cyberspace, 2002.
- [5] Jia Guo, and Ing-Ray Chen. A Classification of Trust Computation Models for Service-Oriented Internet of Things Systems. SCC New York, 2015.