开放型实验指导书

一、实验目的

1. 巩固课堂所学计算机网络知识与网络技术实践能力，熟练运用相关实验方法与技巧；
2. 初步掌握利用网络实验设计和分析方法对计算机网络设计配置、网络使用与应用开发及软件定义网络开发等实际问题开展工作，探索和研究设计网络、使用和优化网络、爬取网络中有用数据及开发应用的基本方法与能力；
3. 启发学有余力的同学基于真实网络问题做有价值的研究项目，甚至发表科研成果。

二、实验内容

本次开放型实验可以选择个人单独完成或小组合作完成（每组最多不超过2名同学，建议个人单独完成）。要求针对某一具体问题，设计实验内容并完成实验内容。选择组队合作完成的同学需协作完成实验设计和实验内容，共同完成一份实验报告，在实验报告中注明共同完成人及各自分工。评分时将以人均工作量作为基本考核指标，综合考虑实验设计的新颖性及所取得的实验结果与效果。

可以从后续给出的五个题目中任选一个进行研究，也可根据自己的兴趣设置题目，自设题目请与李勇老师沟通确认。

三、实验验收

本次实验在第15周、16周两周完成，相关内容和验收分为三步：

1. **实验方案设计：**确定小组与选题，撰写设计文档（模板参考网络学堂），在十五周周四晚实验课验收（6月6日18:30~21:30）；
2. **实验实现及结果验证：**按照设计进行实验，在十六周周四晚实验课验收（6月13日18:30~21:30）。
3. **在网络学堂上提交本次实验的实验报告：**包含实验方案设计、实验过程记录、实验结果分析、实验总结，实验报告长度建议不超过10页，截止时间为十八周周日（6月30日）。

【可选题目一】 大规模互联网拓扑设计与搭建

说明：

在先前的组网实验中（本课程第一次与第二次实验），我们实现了最基础的交换-路由网络，验证了静态、动态路由协议在网络配置中的区别与联系。而在更大规模的网络环境中，我们需要设计更加复杂的网络拓扑结构、实现更加复杂的网络功能以保证大规模网络的扩容与维护管理方便、具有冗余防故障设计、链路负载均衡等。

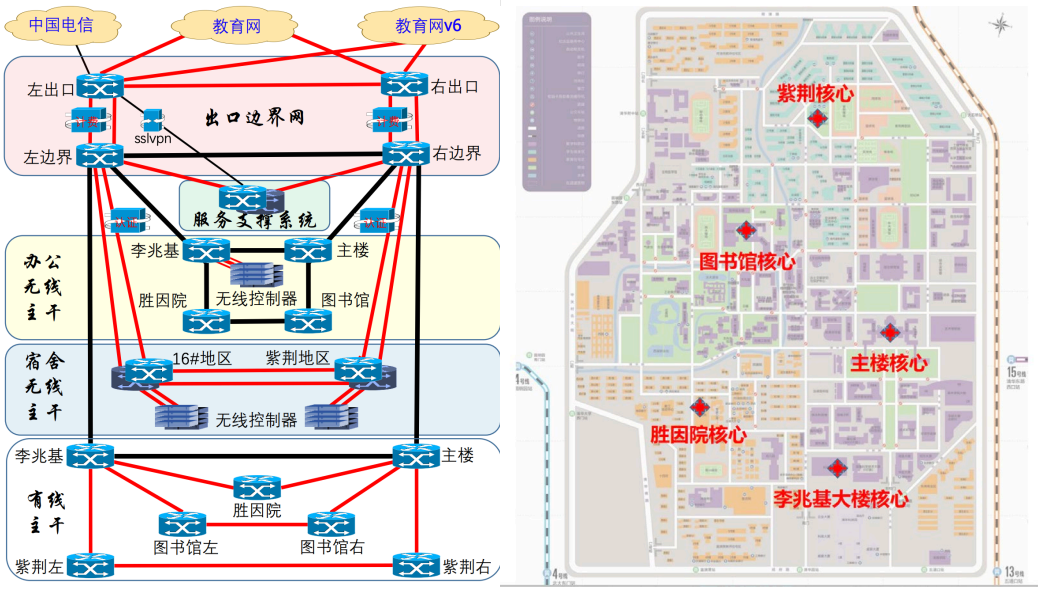
基于以上背景，请设计一个大规模网络拓扑结构，满足罗姆楼用户的网络使用需求，并基于Cisco Packet Tracer进行搭建与仿真。

具体要求：

1. IP规划需根据中国教育科研计算机网络中心分配给清华大学的网段安排（101.6.x.x），满足电子系约2000人、4000设备的联网需求；
2. 拓扑结构参考下图中的清华大学校园网架构示意图，选择合适的校园网路由器作为接入路由器；
3. 配置管理IP段（一般为C类IP，满足设备管理维护需求即可）、业务IP段（承载不同网络业务，如实现办公内网）、互联IP段（公网IP），不同网段分离运行，并通过DHCP服务进行动态IP分配；
4. 网络拓扑需具有一定防故障冗余设计，避免某条链路上的设备故障影响整个网络运行；
5. 设计实现核心办公内网，与日常学习办公用网隔离，用于紧急联络与通信。
6. 注意：仅考虑教育网IPv4配置；在Packet Tracer软件中进行仿真时无需配置全部设备。

在上述具体要求的基础上，若干高级网络设计可以**选择，进一步实现**以下功能（可选）：

1. 链路级别可靠性保证：实现链路聚合，通过多条物理链路实现带宽扩展与防灾冗余；
2. 节点级别可靠性保证：实现网络关键设备的冗余，并通过VRRP将多台冗余设备进行虚拟合并；
3. 实现STP功能，避免二层交换网络误连成环导致网络错误；
4. 配置VLAN，避免广播风暴，提高网络性能；
5. 利用VPN功能方便远程链接核心办公内网。



清华大学校园网架构示意图

参考资料：

1. 清华大学校园网简介
2. 清华服务使用指北（主要面向 Linux 用户）

3. 中国教育和科研计算机网CERNET网络中心
4. 中国教育和科研计算机网（CERNET）
5. 企业园区网络建设技术方案（华为）
6. 【项目实战】如何设计一个小型的200多人的公司网络
7. 企业网搭建，看这一篇就行
8. Exploring Networking with Cisco Packet Tracer
9. Network Technician

【可选题目二】AI模型与数据集开源社区爬取与分析

说明：

在AI研究，特别是大模型相关的研究中，预训练模型和数据集是推动领域向前发展的重要支撑。为了共享这些模型与数据来促进社区研究发展，互联网上出现了一些AI模型与数据集的开源社区平台，其中典型的就是HuggingFace（https://huggingface.co/）。HuggingFace提供了供研究人员开源并共享AI模型与数据集的平台，用户可以向平台上传代码、模型、数据集并添加合适的标签和说明，也可以自由地下载其他用户的AI模型与数据集。许多AI公司都参与到了该社区的建设并上传相应的模型和数据集。该平台在一定程度上也反映除了当下AI研究的热点与社区关注的方向。

基于以上背景，请实现一个网络爬虫，爬取HuggingFace平台中AI模型与数据集相关信息，并通过可视化方法呈现所爬取的数据的分析结果，通过网页和Web服务的方式展示AI社区当下主要关注的研究热点和发展方向。

具体要求：

1. 爬取HuggingFace中现有大模型及其相关数据集，获取其分类标签、下载量、上传时间、数据大小等信息；
2. 分析当前AI社区中最受关注的方向，可以从模型或数据集总量、下载量、上传时间等角度分析；
3. 分析大模型在社区中受关注的程度，可以从模型或数据集的大小等角度分析；
4. 将上述功能集成到网页前端，包括：通过Echarts等可视化工具呈现AI社区所关注的方向及比例；模型与数据集大小与社区关注度的关系；输入或选择某个AI子领域，给出推荐学习了解的模型与数据集；以及其他有意义的内容和分析。

提示：

1. 利用Chrome开发者工具和Postman等工具查找网站数据结构，验证爬取相关数据可行性。
2. 详细数据的提取需要对HTML页面进行解析，可以使用Python库BeautifulSoup4。
3. 要实现更精美的网页呈现，可以使用更常用的Web框架，如前后端不分离的Django框架，或在前后端分离基础上使用React、Vue等框架完成前端开发，使用Flask等框架完成后端开发。

参考资料：

1. BeautifulSoup4库：https://www.crummy.com/software/BeautifulSoup/bs4/doc.zh/
2. Django库：https://docs.djangoproject.com/zh-hans/4.2/
3. React框架：https://zh-hans.react.dev/learn
4. Vue框架：https://cn.vuejs.org/guide/introduction.html
5. Echarts可视化库：https://echarts.apache.org/zh/index.html

【可选题目三】基于社交网络的爬虫与数据分析

说明：

社交网络一般指基于用户关系的信息分享、传播以及获取信息平台，例如国内的微博、微信，国外的Twitter和Facebook，均是社交网络应用的典型代表。除此之外，还有许多平台也含有很强的社交网络属性，例如问答社区知乎、Quora，视频网站bilibili等。通过用户与用户之间的关注和粉丝等关系，社交网络应用的用户被关联在一张巨大的网络中，每个用户都是网络中的一个节点，用户之间的关联通过连接边刻画。

具体要求：

基于以上背景，请选定一个平台（除微博外）作为爬取对象，实现一个网页爬虫。通过预先选定若干种子用户，进一步通过递归的方式爬取用户之间的关注和粉丝等关系，以及用户个性化信息，进行数据分析。爬取应包括但不局限于以下方面：

1. 不同用户之间的关注、粉丝关系；
2. 用户的基本信息，如id、昵称、城市、关注数量和粉丝数量等；
3. 统计网络节点的出入度的分布，社交网络中的孤岛个数（内部连通且与外部无连接），平均每个孤岛的节点数等数据特征；
4. 将上述关键功能集成到网页前端，包括：通过Echarts等可视化工具呈现爬取到的用户关注数量分布与粉丝分布；可视化社交网络，呈现用户间的连接关系与社交网络孤岛孤岛；输入用户名，实现用户搜索，呈现查询到的用户的社交网络信息；其他有意义的内容。
5. 选作内容，爬取用户发表的内容，利用自然语言处理模型分析文本情感和内容等，实现社区检测等有意义的内容。

提示：

1. 请注意爬虫的频率，过于频繁可能会导致封号
2. 利用Chrome开发者工具和Postman初步验证爬取相关数据可行性
3. 建议基于scrapy库编程实现：https://docs.scrapy.org/en/latest/
4. 网络结构分析建议参考networkx库：https://networkx.github.io/

【可选题目四】基于SDN的网络防火墙

说明：

出于安全性考虑，网络管理员需要对网络访问进行限制。例如，只允许特定终端访问某一服务器，或者只允许访问服务器的80端口。传统网络使用防火墙设备实现访问 控制和非法数据包过滤。由于SDN提供了灵活的可编程能力，因此可以不使用专门的防火墙设备，而通过在控制器上编程实现防火墙的功能。基于上述背景，请在Floodlight控制器上编程实现防火墙的功能。

具体要求：

要求可以提前配置防火墙规则，控制器根据访问规则进行转发决策。访问规则包括两部分:匹配字段和访问控制选项。匹配字段包括源、目的MAC，源、目的IP，传输层协议，传输层源、目的端口号。访问控制选项包括允许访问和禁止访问。例如，下表中的两条防火墙规则表示禁止所有目的端口为22的TCP流量以及所有发往 192.168.10.1的流量。通过调研防火墙的主要规则类型，实现主流的访问控制选项。可视化上可以考虑通过图形界面实现对防火墙的实时操作和配置。

源MAC	目的MAC	源IP	目的IP	协议	源端口	目的端口	访问控制选项
*	*	*	*	TCP	*	22	deny
*	*	*	192.168.10.1	*	*	*	deny

参考资料：

1. OpenFlow协议标准：http://archive.openflow.org/documents/openflow-spec-v1.1.0.pdf
2. OpenFlow在Wireshark中的插件：https://wiki.wireshark.org/OpenFlow 3
3. Iperf发包工具：https://iperf.fr/
4. 数据中心多路径流量调度：https://www.usenix.org/legacy/event/nsdi10/tech/full_papers/al-fares.pdf
5. 基于SDN的防火墙：http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6779061 https://www.ietf.org/proceedings/91/slides/slides-91-i2nsf-0.pdf