**暨南大学本科实验报告专用纸**

课程名称 计算机网络实验 成绩评定

实验项目名称 Internet应用于应用层协议分析 指导教师 潘冰

实验项目编号 3 实验项目类型 验证型 实验地点

学生姓名 林晓旭 学号 2019051121

学院 智能科学与工程 系 专业 信息安全

实验时间2021年9 月 28 日 午～ 10月 2日 下午

1. **实验目的**

1、理解WWW 、 DNS服务、FTP服务、SMTP的作用和原理；

**2、学会使用wireshark分析HTTP、FTP、SMTP和DNS协议的工作过程，加深对协议格式和工作原理的理解。**

1. **实验内容**

1、通过域名访问WWW、FTP服务器，分析DNS、WWW、FTP工作过程，并使用WireShark分析相关协议格式；

2、在客户端访问SMTP服务器，使用wireshark分析SMTP、POP3协议的工作过程。（可以在客户端安装outlook或使用QQ邮件服务器或自己编程）

1. **实验设备**

局部网环境，计算机若干台。本实验不分组，独立完成。

**四、实验内容**

**1、HTTP协议分析**

访问任意web站点，用wireshark分析HTTP协议的工作过程和HTTP协议格式。（协议首部含义需要查询了解）。

1. **DNS协议分析**

通过域名访问某网站，用wireshark捕获DNS数据包，并分析DNS工作过程和DNS格式。

**3、FTP协议分析**

**1.**访问FTP服务器。如ftp://ftp.jnu.edu.cn

2.用wireshark分析FTP的工作过程。注意观察FTP的工作模式，用于控制连接的端口和数据连接的端口。

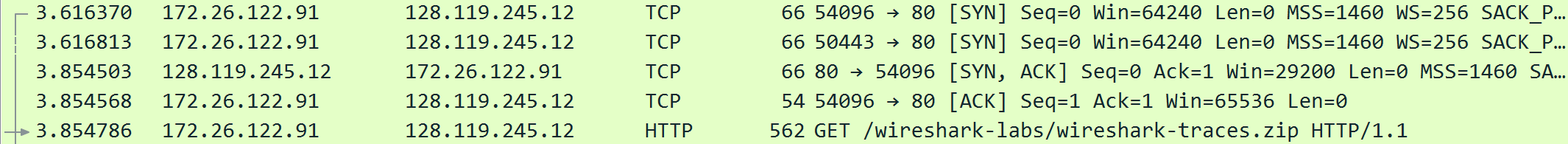
**4、 SMTP和POP协议分析**

基于Web的邮件或客户端的邮件软件（如outlook）收发邮件，捕获数据报分析邮件收发过程和SMTP、POP3等协议格式和工作过程。

**五、实验结果**

1、HTTP协议分析

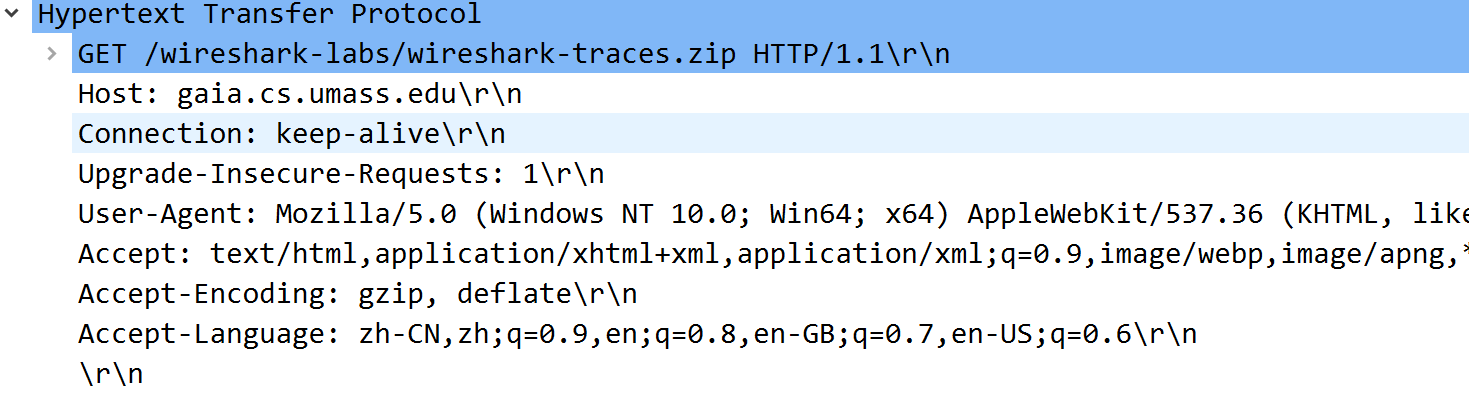
（1）、因为HTTP基于TCP协议，所以在主机与目标网站进行HTTP传输之前，首先需要建立TCP连接，也就是进行TCP三次握手。



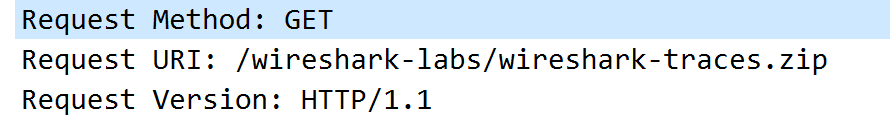


可以看到在HTTP协议前有4个TCP协议，其中1、3、4 为一次TCP连接的建立过程。

（2）、HTTP请求报文格式



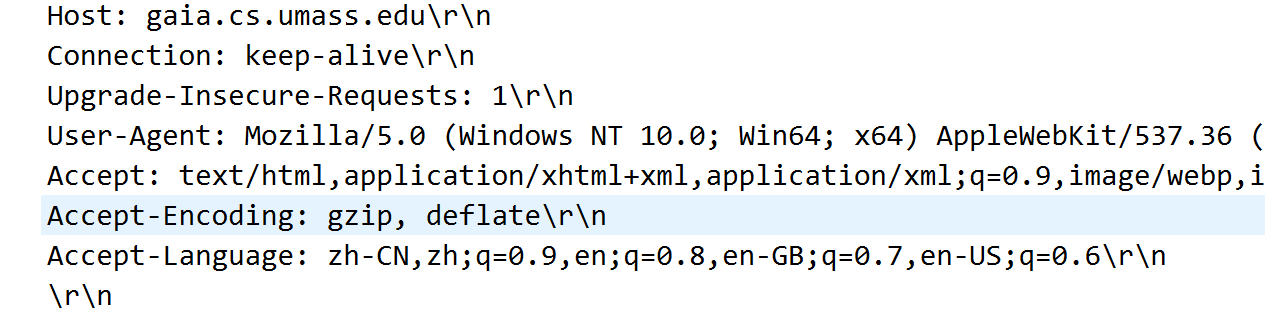
在HTTP请求报文的首段是被称为状态行的字段。该行包括了请求方法、请求URI(其中又包含URI路径以及URI询问，而URI询问中又可能包括了用&分割开的若干个参数（在本实验中没有体现），这些参数用来给动态网页提供相应的响应)以及HTTP协议版本。URI表示Web上每一种可用的资源，如图片、视频、程序。



图中请求方法为GET。在HTTP/1.1中一共有8中请求方法：（1）、GET：请求指定的页面信息，并返回实体主体；（2）、HEAD：类似于GET，都是用来请求资源，但是HEAD用来获取报头，确定URI的有效性以及资源更新的日期；（3）、POST：向指定资源提交数据，请求服务器进行处理（上传文件、发评论等等）；（4）、PUT：向指定资源位置上传其最新内容；（5）、DELETE：请求服务器删除指定资源；（6）、TRACE：回显服务器收到的请求，主要用于测试或诊断；（7）、OPTIONS：这个方法可使服务器传回该资源所支持的所有HTTP请求方法（用“\*”来作为资源名称向服务器发送OPTIONS请求可用来测试服务器功能是否正常运作）；（8）、CONNECT：HTTP/1.1协议中预留给能够将连接改为管道方式的代理服务器。CONNECT方法要求在于代理服务器通信时建立隧道，实现用隧道协议进行TCP通信。

关于POST和PUT的区别：POST方法请求服务器接收请求中的实体作为请求资源的一个新的从属物；PUT方法请求服务器把请求中的实体储存在请求URI下。可以理解为，创建一条记录用POST，修改一条记录用PUT。POST和PUT方法最根本的区别在于请求URI含义不同。POST中的URL是一个能处理请求实体的资源。PUT中的URI是被请求实体更新/替换的资源的URI。PUT方法是幂等的（任意多次相同的PUT所产生的影响与一次PUT产生的影响相同，可以理解为你反复对同一篇文章做相同的修改和你做一个修改的效果是相同的），而POST不是幂等的（你写两篇一摸一样的文章，就会存在两篇文章）。

回到HTTP请求报文中，在请求行的后面是被称为请求报头的东西。请求报头由若干个键值对构成（每个键值对占一行），用来给客户端向服务器发送自身的信息以及一些其他信息。

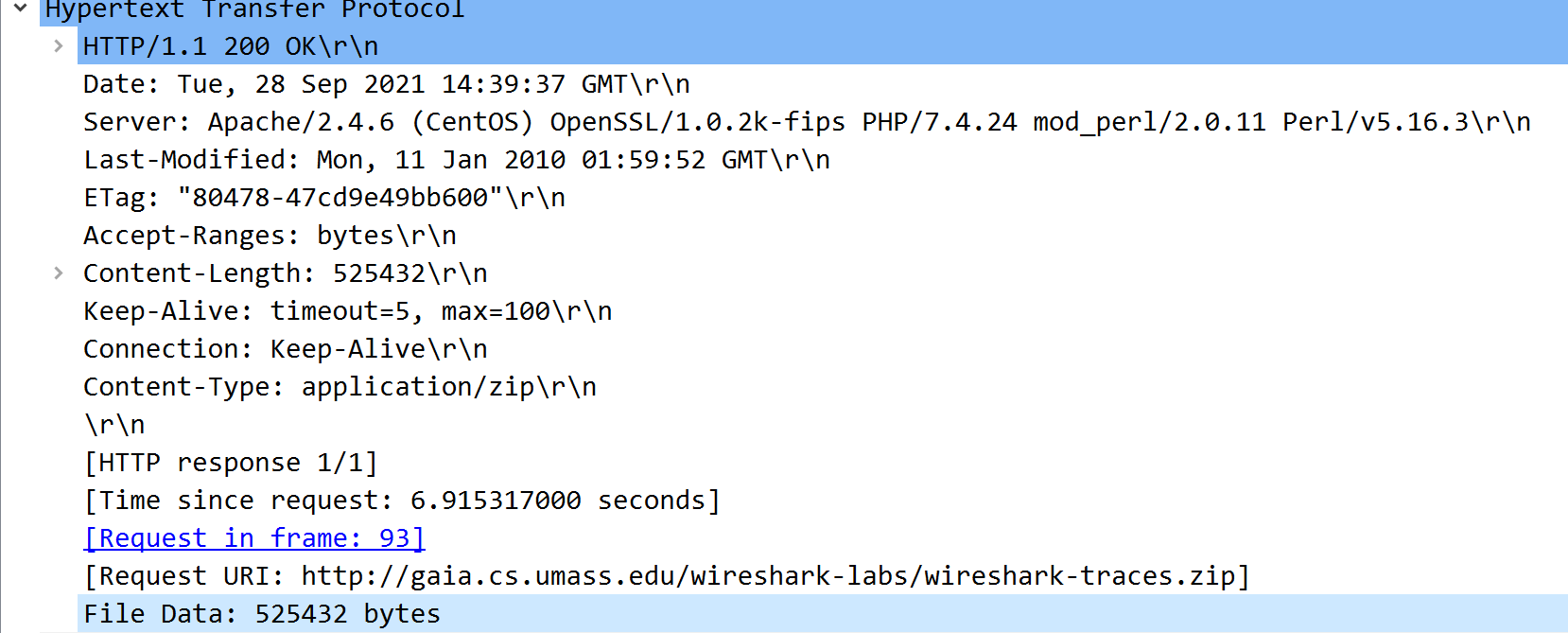


Connection表示是否要进行持久连接，图中表示要(HTTP/1.1默认)；User-Agent表示用户代理。也就是用什么方式访问网页（如爬虫、浏览器）；Accpet表示客户端能够接收的内容类型；Accept-Encoding表示客户端能够接受的编码类型；Accept-Language表示客户端能够接受的语言类型。

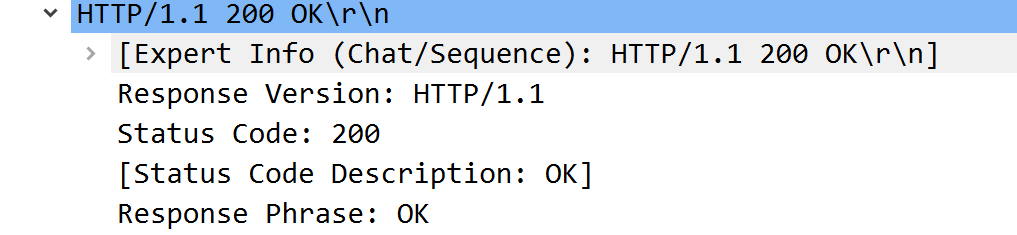
在请求报头后，还有一行空白行。HTTP GET请求报文中并没有请求正文。下图形象地描绘出HTTP请求报文的格式



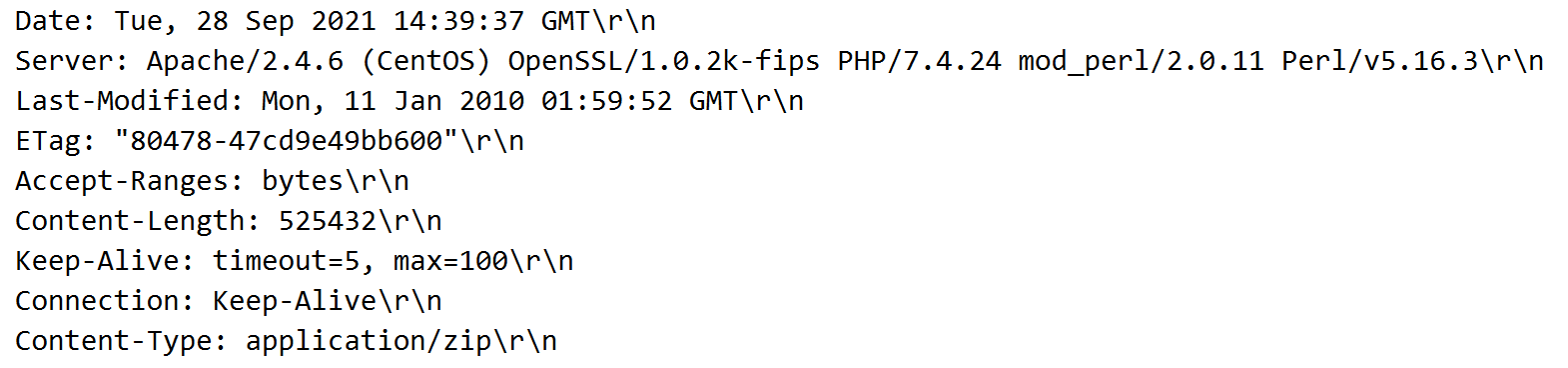
（3）、HTTP响应报文



HTTP响应报文地首段同样是状态行，该状态行包含了HTTP协议版本、状态码、状态码的文本描述。



图中表示HTTP版本为1.1，状态码为200，文本表述为OK。表示之前的GET请求成功。同HTTP请求报文，在状态行之后也是由若干个键值对组成的消息报文（每个键值对占一行）。



Date表示创建报文的时间（并不是发送报文的时间，The **Date** general HTTP header contains the date and time at which the message was **originated**.），可以用来判断该资源是否来自缓存服务器，如果该报文的Date比当前时间早，则可以认为资源来自缓存服务器；Server表示服务器信息；Last-Modified表示该资源上次修改时间；Etag是某个资源的唯一标识符，它的作用是可以让缓存更加的高效而且节省带宽，因为当请求的资源没有改动时，服务器就不需要发送一个完整的回应；Accpet-Ranges表示是否接受字节范围请求；Content-Length表示资源的字节长度；当http请求报文选择Keep-Alive长连接的TCP连接模式时，http响应报文中的Keep-Alive字段表示该TCP长连接的属性，在图中表示该tcp连接存在5秒，最多接受100次连接；Content-Type表示资源类型，图中表示资源类型为一个压缩文件；

在消息报文后接着一个空行，空行之后就开始是响应正文了，这里的响应正文就是该资源的二进制码。



下图说明了Http响应报文的格式



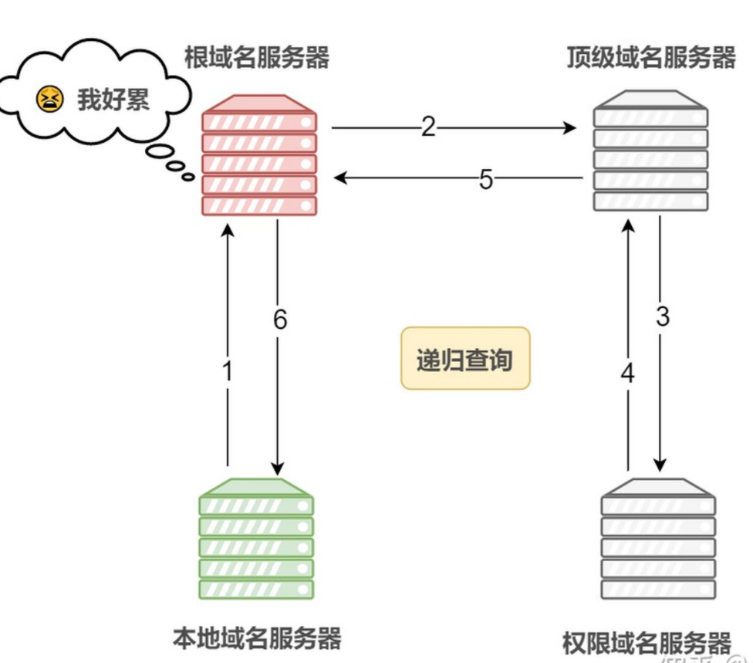
2、DNS协议分析

（1）、DNS协议

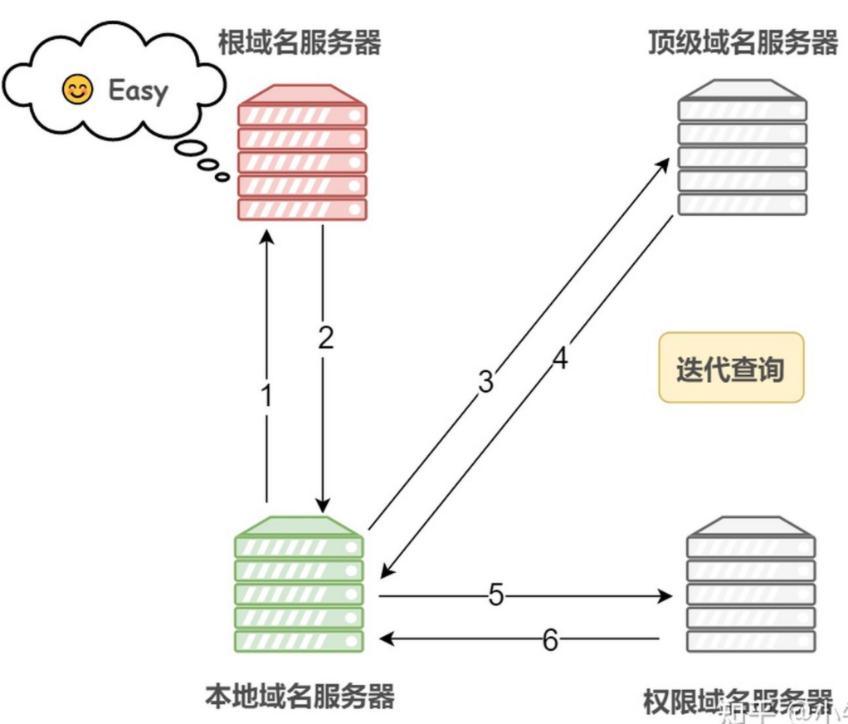
DNS(Domain Name System)存在是为了将人们不易记忆的IP地址转化为域名。域名与IP之间的映射是多对一的关系。域名由可以分为顶级域名、二级域名、三级域名等等，所以整个域名搜索结构是一个树状结构。一个网站的域名是一个若干个不用层次的域名通过”.”分隔形成的字符串，如”[www.baidu.com](http://www.baidu.com)”，其中com是顶级域名，baidu是二级域名，www是三级域名。

域名与IP地址之间的映射关系被储存在域名服务器之中。域名服务器又可以根据层次划分为根域名服务器、顶级域名服务器、权限域名服务器。根域名服务器储存着所有顶级域名服务器的IP地址；顶级域名服务器管理注册在该顶级域名下的所有二级域名，记录这些二级域名（权限域名服务器）的IP地址。跟根域名服务器和顶级域名服务器有所不同，权限域名服务器并不是储存注册在该二级域名下的三级\四级域名IP地址，而是管理注册在该二级域名下的一个分区内的域名，记录着这些域名的IP地址。比如由”[www.baidu.com”,”ai.baidu.com”,:tieba.baidu.com](http://www.baidu.com)“三个域名，显然这三个域名都是注册在baidu这个二级域名下的，那么我们可以将”[www.baidu.com,”tieba.baidu.com](http://www.baidu.com,)”放在a.baidu.com这个权限域名服务器，而”ai.baidu.com”放在b.baidu.com这个权限域名服务器内，这两个权限域名域名服务器是具有同等地位的。除了以上三个域名服务器，还有本地域名服务器，它用来帮助计算机进行DNS解析。

DNS查询方式包括递归查询和迭代查询。递归查询指的是，对于一个DNS查询请求，如果当前域名服务器无法给出直接回答，那么它将会扮演成请求者向下一层级的域名服务器发送请求，得到信息后再返回给请求者；



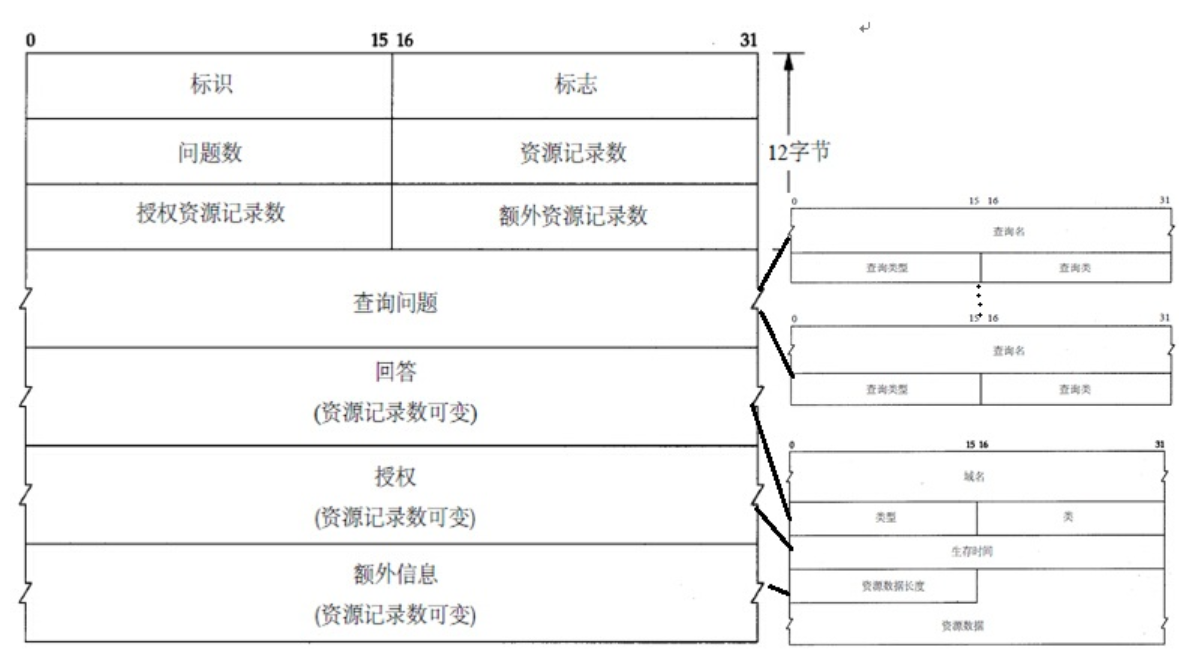
迭代查询指的是，对于一个DNS查询请求，当前域名服务器无法给出直接回答，那么它将会返回对应域名下一层级的域名服务器的IP地址，让DNS查询请求者继续去寻求信息。



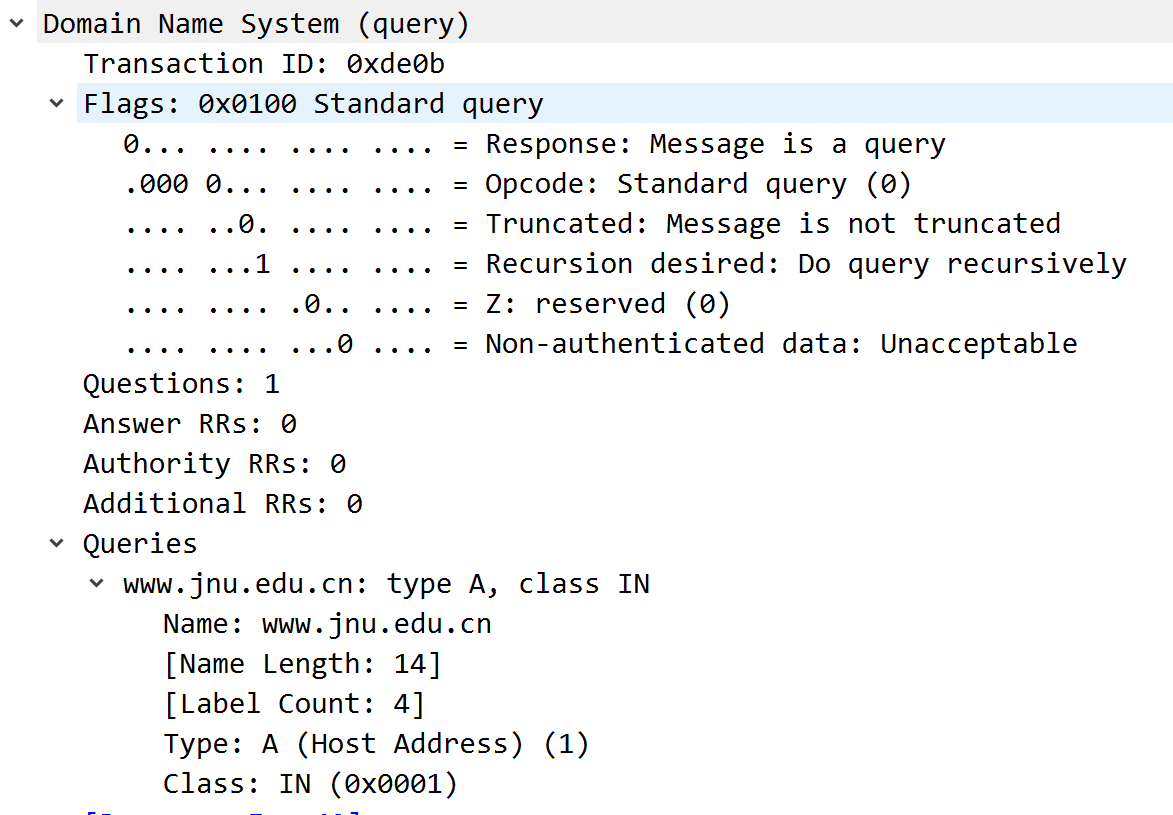
当我们的主机需要对一个DNS进行解析时，首先在主机缓存中查找是否有缓存信息，若缓存为命中，进而向主机所在网络中本地DNS服务器发送一个查询请求，本机DNS再通过迭代或者递归的方式向上级的域名服务器进行查询，最后将结果返回给主机。流程如上图所示。

（2）、DNS报文

DNS报文可以分为基础结构部分、问题部分以及资源记录部分。资源记录部分仅再DNS回应报文中出现。如下图所示



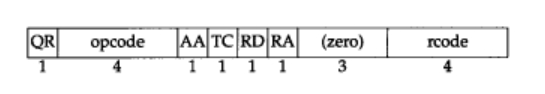
用WireShark抓包来获得更为直观的展现，如下是一个DNS请求报文。



DNS报文的基础部分指的是报文的首部。包括如下部分



首先是一个占两个字节的标识ID，用来标识一个DNS请求。接下里就是一个占两个字节的Flags位图。位图结构如下所示。



QR(Response):0表示是询问报文；1是回应报文

Opcode：操作码，0表示标准查询（域名到IP），1表示反向查询（IP到域名），2表示服务器状态请求。

AA(Authoritative):授权应答，该字段仅再回应报文中有效。1表示给出对应域名IP地址的名称服务器是管理该域名的权限域名服务器，也被称为权威服务器、授权域名服务器；否则为0.（来自之前DNS查询所构建的缓存文件）。

TC（Truncated）；1表示DNS报文信息被截断（超过512字节时）；0表示未被截断。

RD（Recursion Desired）：1表示进行递归查询，否则为0。该字段仅再询问报文中能设置。

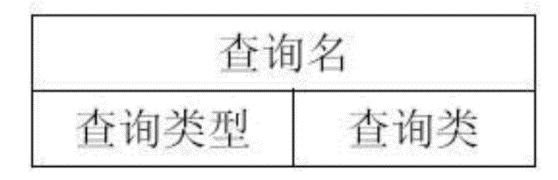
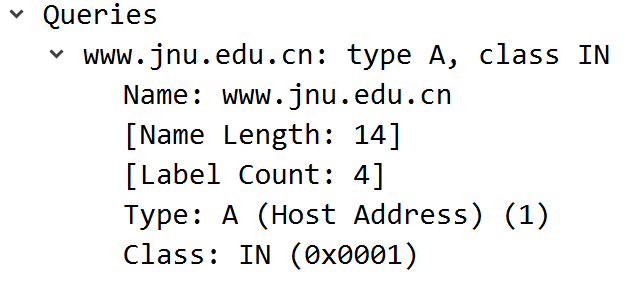
RA（Recursion Available）：该字段仅出现再回应报文中，1表示服务器支持递归查询，否则为0。

Rcode（Reply Code）：返回码字段，仅在回应报文中有效。0表示正常；1表示格式错误（Format Error），服务器无法理解请求报文；2表示域名服务器失败（Server Failure），服务器无法处理这个请求；3表示名字错误（Name Error），只对授权域名服务器有效（存储着域名源文件），指出域名不存在；4表示查询类型不支持（Not Impletmented），即域名服务器不支持该查询类型；5表示拒绝（Refuse），域名服务器由于设置的策略拒绝提供应答。

在RA和rcode之间还有3个保留位、

回到DNS报文头中，Question表示该报文包含的问题个数（有多少个DNS需要解析/反向解析）；Answer RRs（Resource Record）表示报文中包含的回答资源记录数，简单地可以理解为回答个数，有时候针对一个问题可以有多个回答；Authority RRs表示权威服务器资源记录个数；Additional RRs表示附加资源记录数。

接下来塔伦DNS报文中地问题部分。

查询名表示需要查询的域名或者IP地址；查询类型占两个字节表示对查询结果的期望信息。

名字 数值 解析

A （1） 期望获得查询名的IP地址。

NS （2） 一个授权的域名服务器。

CNAME （5） 规范名称。

PTR （12） 指针记录。

HINFO （13） 主机信息。

MX （15） 邮件交换记录。

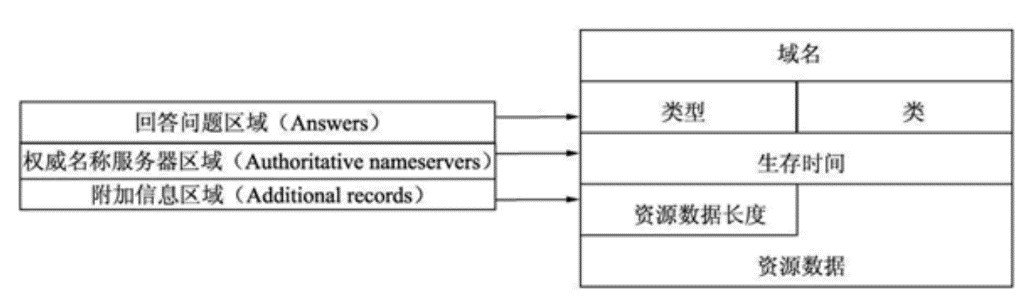
AXFR （252） 对区域转换的请求。

ANY （255） 对所有记录的请求。

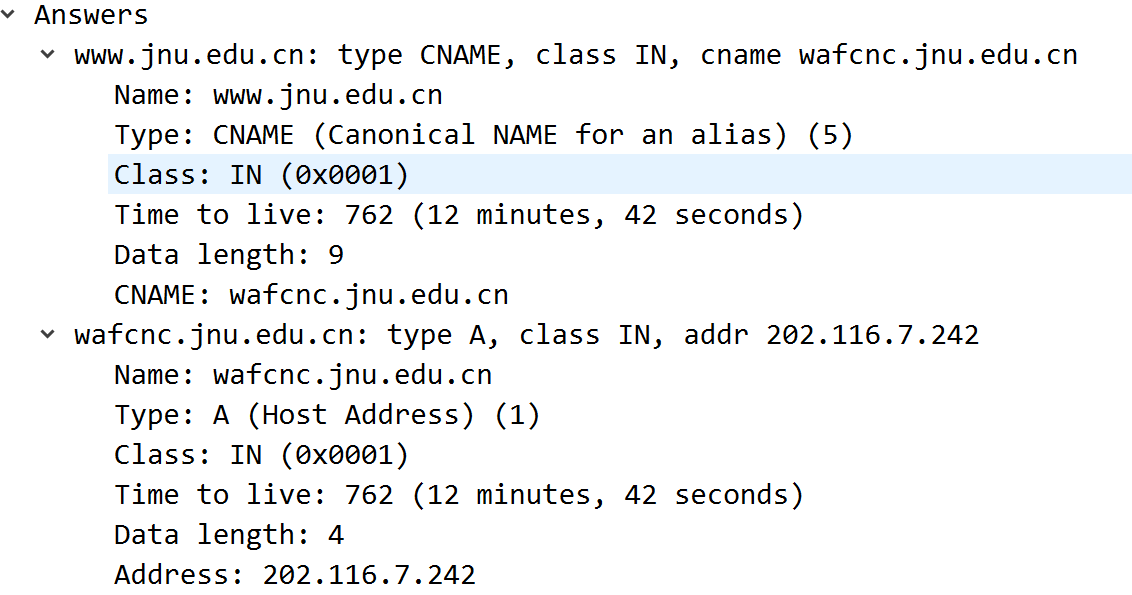
查询类同样占两个字节，1表示查询结果是一个互联网地址。

在DNS响应报文中还包括资源记录部分，其中资源记录包括回答资源、权威服务器资源、附加资源，它们的数量分别对应DNS报文头部中的Awser，Authority RRs，Additional RRs 字段。

资源记录格式如下图所示：



其中域名、类型、类字段意义和询问报文中的相同；生存时间（Time to Live）以秒为单位，表示资源记录的生命周期，一般用于当地址解析程序取出资源记录后决定保存及使用缓存数据的时间。它同时也可以表明该资源记录的稳定程度，稳定的信息会被分配一个很大的值。资源数据长度表示数据部分的字节长度；资源数据部分记录着具体的信息（IP地址等等）。

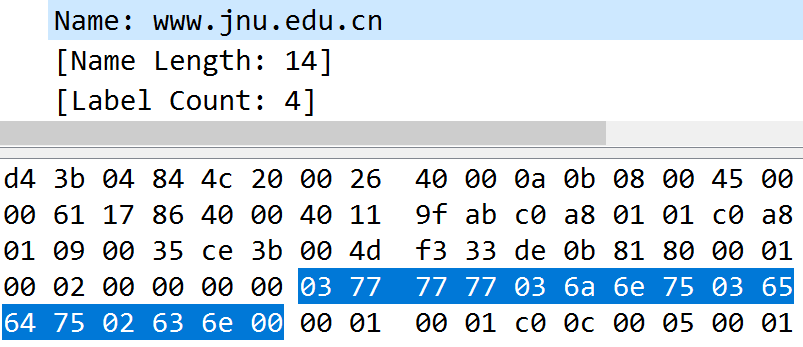


在上图所展现的问题回答区域，可以看到第一条回答是关于域名转换，将[www.jnu.edu.cn](http://www.jnu.edu.cn)转换成wafcnc.jnu.edu.cn，所以解析的实际上是wafcnc.jnu.edu.cn的IP地址。关于域名转换的目的有很多，比如用来加速服务、负载均衡。例如CDN加速，当主机用域名访问一个公司的网站时会重定向至CDN服务提供商的服务器，CDN服务器提供商里缓存着该公司网站的信息，最终实际上用户是从该CDN服务器中获取有关该公司网站的信息。

权威名称服务器区域一般提供的信息是关于管理该域名的权威服务器的域名。

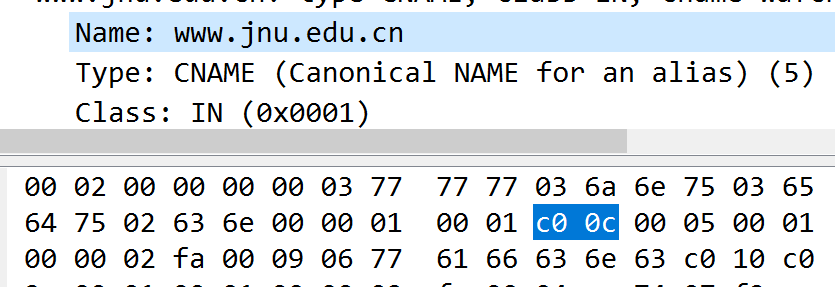
附加信息区域一般提供管理该域名的权威服务器的IP地址。

**顺便一提关于DNS报文的压缩方法**

****

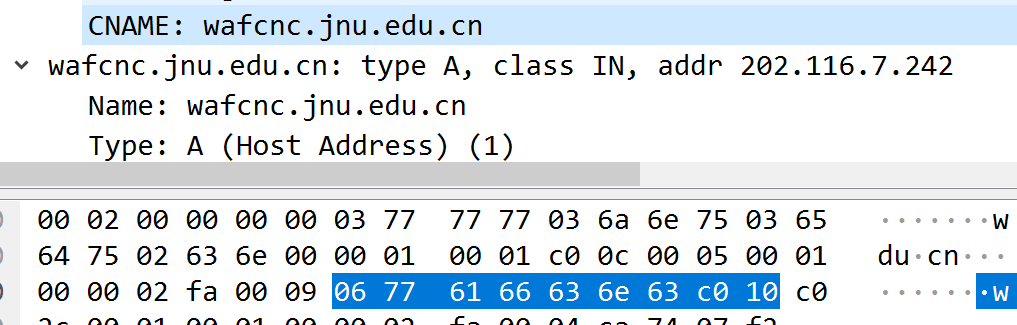
通过wireshark，我们发现在dns报文中关于域名是这样表示的，对于不同层次的域名，首先用一个字节表示该域名的字符串长度，再给出字符串的ascii码表示。最终将不同层次的域名连结起来就构成该网站的域名。比如上图中的“[www.jnu.edu.cn“](http://www.jnu.edu.cn)，第1字节03表示四级域名www有三个字母，第2-4个字节是该四级域名的ascii码表示，第5字节03表示三级域名jnu有三个字母，第6-8个字节是该三级域名的ascii码表示。以此类推。最后以0结束。

但是当第二次再出现该域名时，DNS报文会用指针的方式来表示



当用指针表示时，第一个字节的最高位和次高位都为1，所以第一个字节的十六进制表示位c0。第二个字节表示指针指向的目标相对于DNS报文开头所偏移的字节数。

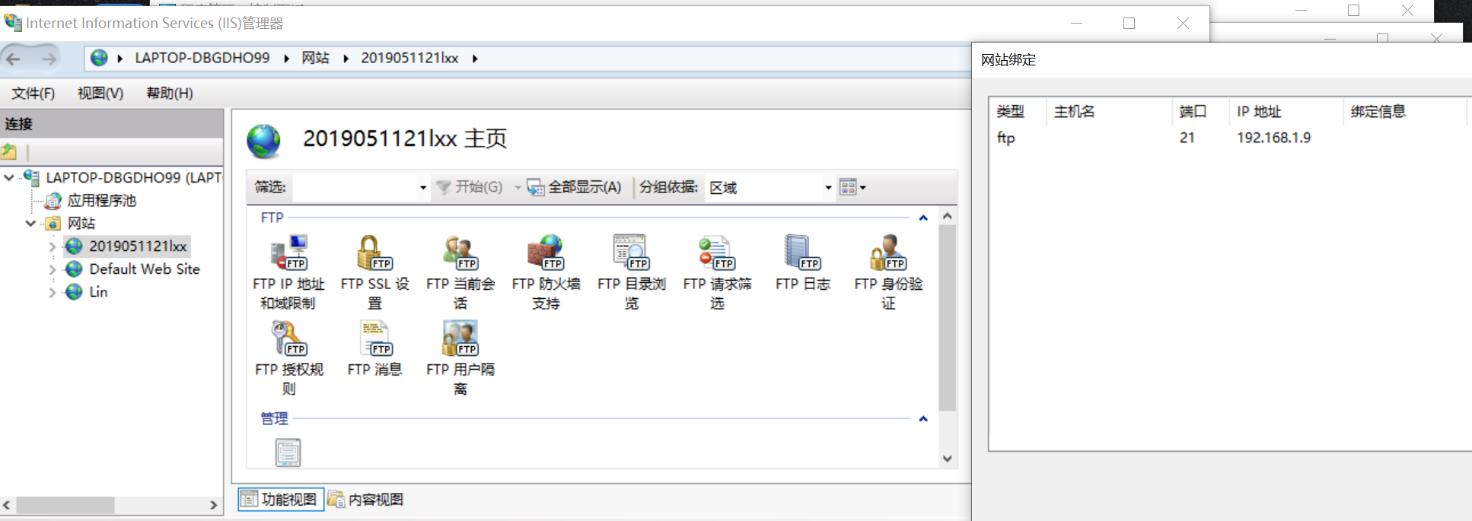
还有混合表示方法



wafcnc域名是第一次出现在该DNS报文中，必须用前文提到的表示方法来表示；而之后的域名是出现过的，可以用指针表示。

3、FTP协议分析

（1）、首先通过Windows自带的Internet Information Service（IIS）来在本机创建一个FTP站点。

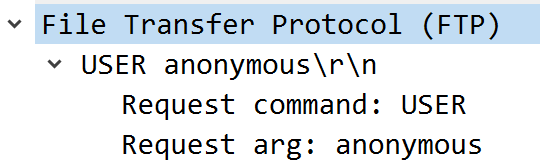


要想wireshark成功抓包，需要使用当前局域网中另一台IP设备来访问FTP服务器（需要关闭本机防火墙）。本实验通过手机端通过safari浏览器访问FTP服务器并且下载一张jpg图片。

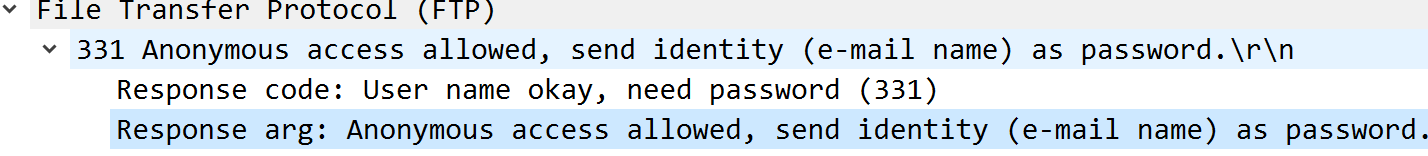
 

（2）、FTP报文格式

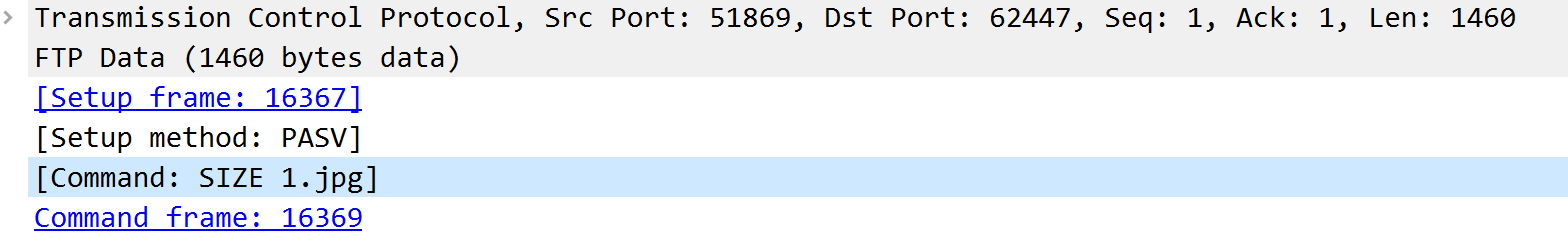
FTP Request报文都是可读的，首先是3-4个ascii码组成的请求命令，隔了一个空格后就是一些请求参数，最后以换行符(/r/n)作为报文消息的结尾。



FTP Response报文和http Response报文很相似。首先是一个三位数的响应码，隔了一个空格后是该响应码的语义信息。



另外还有FTP-Data报文，该报文就是单纯的传输数据流。

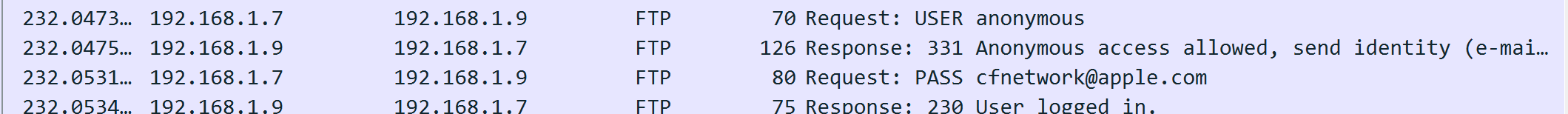


（3）、分析FTP工作流程

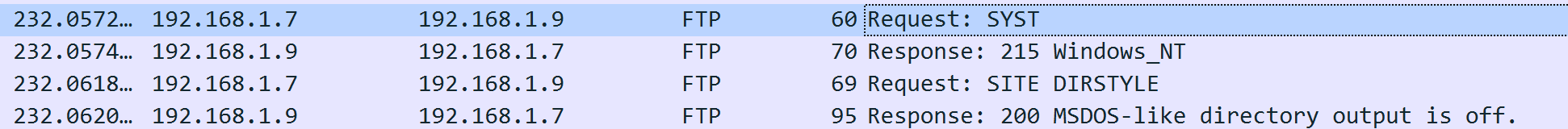
FTP协议是基于TCP协议的，所以在进行FTP协议工作之前需要进行TCP三次握手连接。



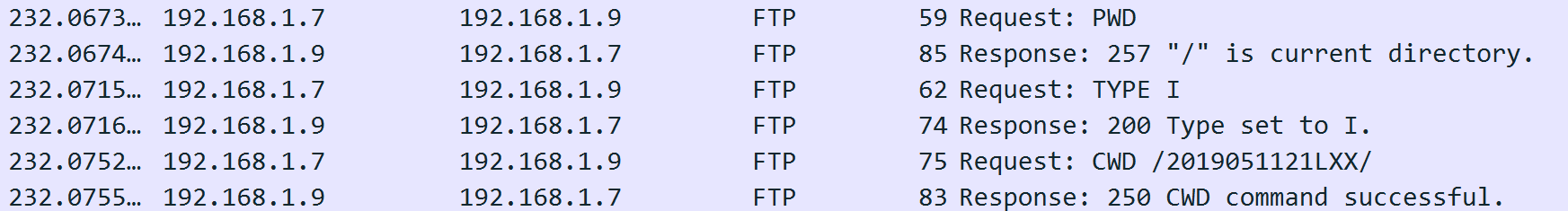
在建立完三次握手后FTP服务器发出220响应码表示服务就绪并给出服务器信息，因为本实验是通过window自带的组件来搭建FTP站点，所以显示是Microsoft FTP Service。



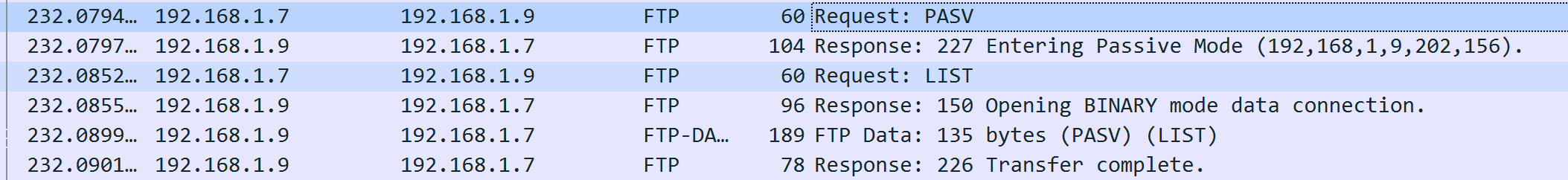
由于设置了FTP服务器可以通过匿名模式登录，所以在通过浏览器访问FTP站点时不需要用户自己输入任何登录信息。由于匿名登录需要一个邮箱来标识匿名用户，所以在这里浏览器会自动提交一个邮箱。



客户端询问服务器使用的操作系统(SYST)以及要求服务器提供目录的样式(SITE DIRSTYLE)。

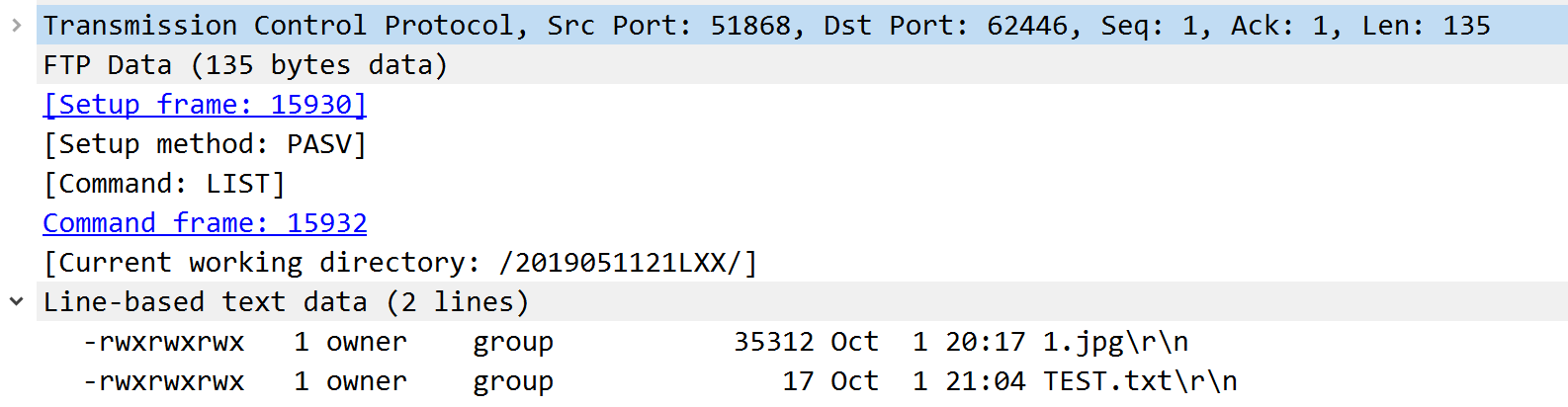


客户端询问当前的工作目录(PWD)，设置传输的数据类型为二进制类型(TYPE I)，以及改变工作目录(CWD)。

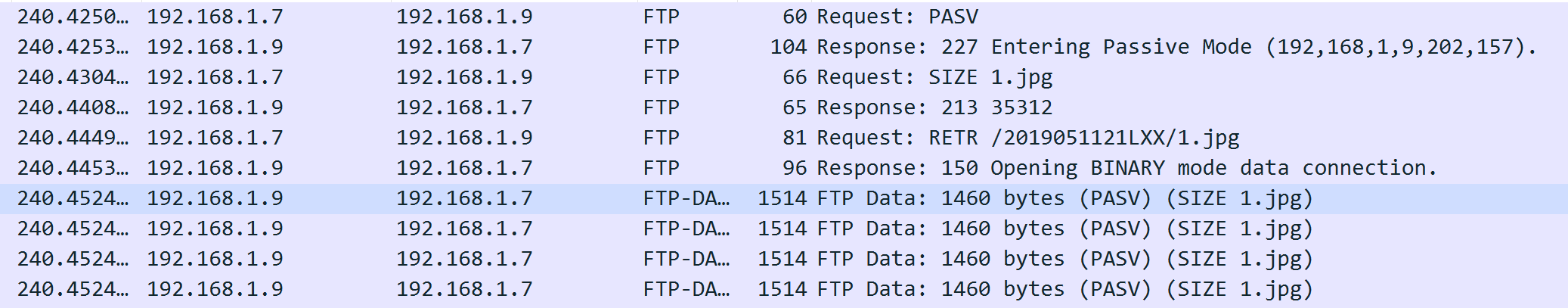


如上图，客户端需要执行LIST命令，也就是列出当前目录。由于这个命令需要进行数据的传输，所以FTP服务器与客户端之间需要建立数据通路。在这里采用被动模式的连接，即服务器给出一个指定的端口让客户端自己去连接（被动连接的安全性比主动连接高）。在进行完当前这次任务的数据传输后，数据连接会关闭。

通过wireshark抓包观察数据传输包的端口号确认了确实时被动连接，同时观察数据传输内容也确实是当前目录的信息。



接下来，我进行了下载jpg文件的操作。



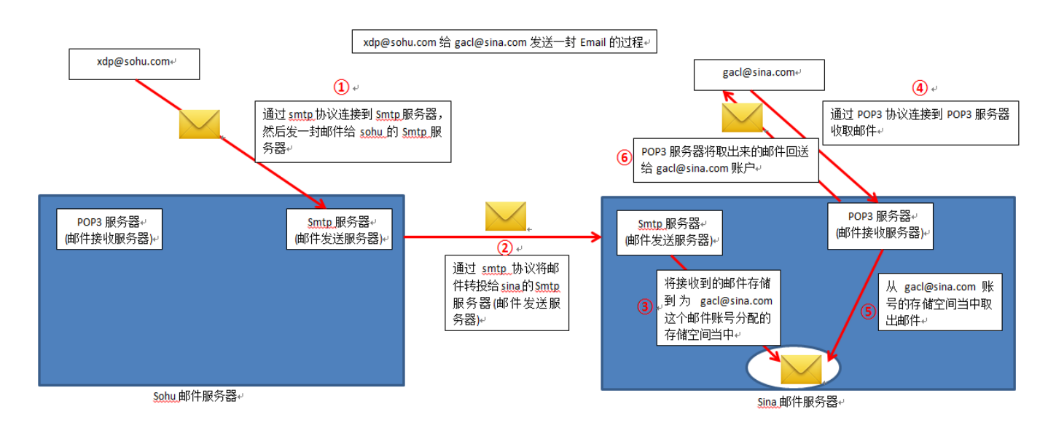
可以看到在通过被动模式建立了数据链接后，客户端向服务器询问了该jpg文件的大小(SIZE)。在知道文件大小后通过RETR（retrieve）命令向服务器获取jpg文件的复本。服务器以分块的方式通过数据通路传输jpg文件。

3、SMTP/POP协议

SMTP（Simple Mail Transfer Protocol），是一组用于将邮件从源地址传输到目的地址（邮件服务器）的协议。通过SMTP来控制规范邮件的中转方式，基于TCP/IP协议。SMTP服务器默认端口是25，经过SSL加密的默认端口是465。

POP（Post Office Protocol），是一组用于将邮件从邮件服务器读取至本地主机的协议。现在的邮箱大都采用POP3协议。POP3协议是一个离线协议，也就是说客户端上对邮箱的操作不会影响邮箱服务器上的邮箱（比如说你在客户端删了一封邮件，但是在服务器上这封邮件并不会被删除）；而另一个比较常用的名为IMAP的协议，用户在客户端上的操作都会反馈到服务器上。POP服务器的默认端口是110，经过SSL加密的默认端口是995。

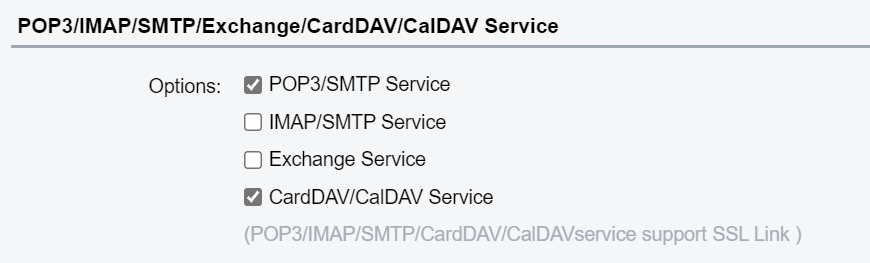
（1）、SMTP/POP3邮件收发流程



当发送者写好了一封邮件，需要通过SMTP协议来和SMTP服务器取得连接，经过验证后SMTP服务器会接受该邮件并将其缓存进发送队列以便稍后发送；源邮件服务器与目的邮件服务器通过SMTP协议进行通信传递邮件；目的邮件服务器接受到邮件后将其储存在接收者的邮箱空间中；接收者通过POP协议与邮件服务器连接并访问自己的邮箱空间。

（2）、用wireshark捕获SMTP报文

首先我们需要开启对应邮箱的STMP/POP服务功能

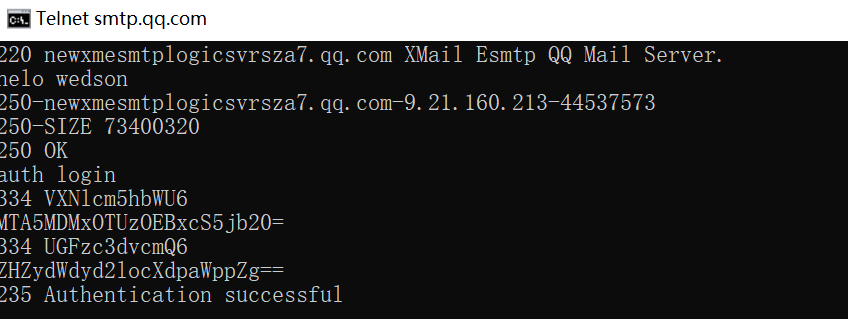


之后在cmd窗口通过telnet命令连接至该邮箱的SMTP服务器



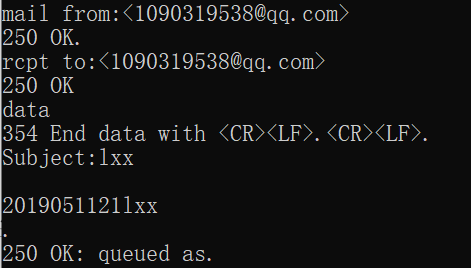
在与SMTP服务器取得连接后就开始用SMTP协议来与服务器通信。

1. 用helo命令来告诉服务器当前主机名（任意字符串都行）；
2. 用auth login命令请求开始验证；
3. 输入经base64编码后的邮箱名以及经base64编码的密码

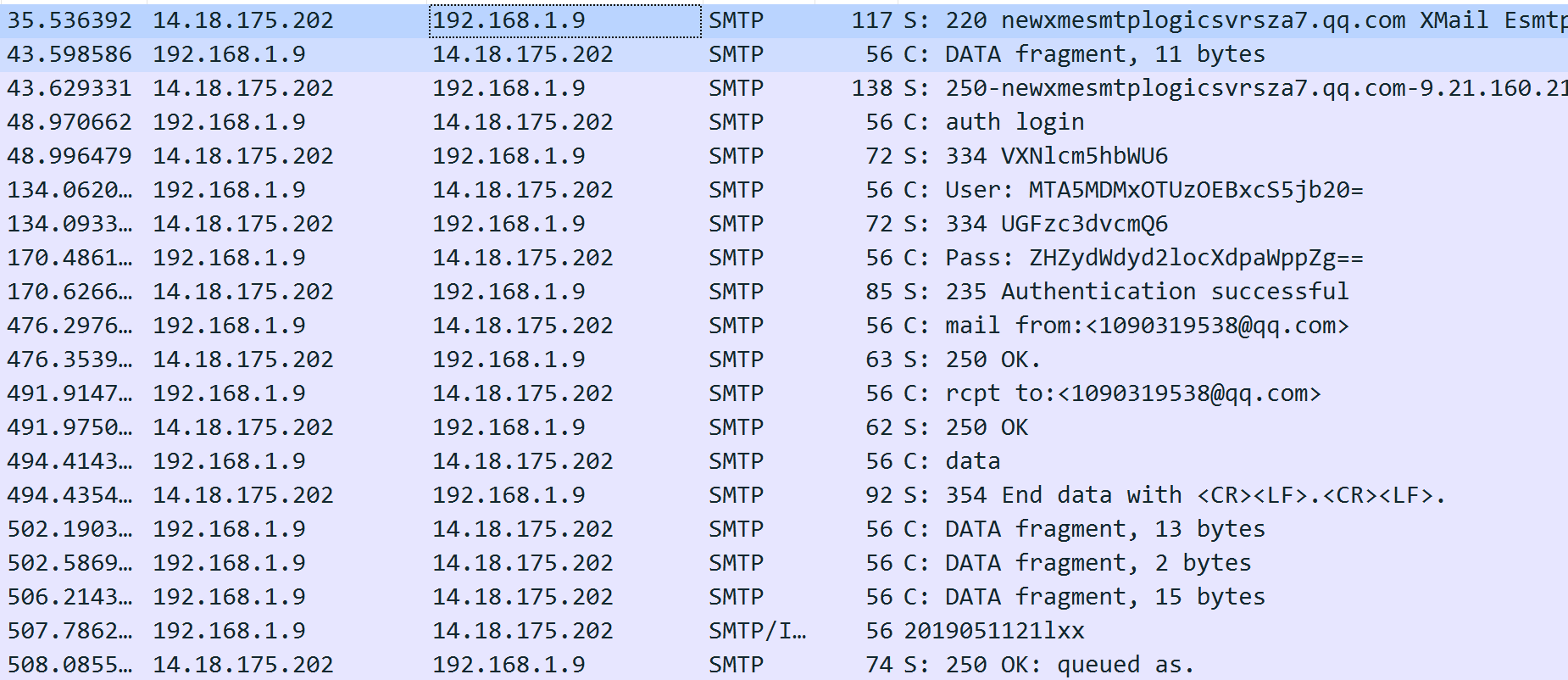


在完成授权验证之后就可以进行邮件的发送了。

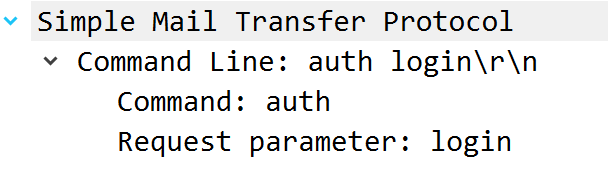
1. 用mail from:<send mail>命令来指定发送邮箱（必须和授权认证的邮箱一致）
2. 用rcpt to:<reception mail>命令来指定接收邮箱
3. 用data命令来输入邮件内容；可以用Subject关键字来指定邮件标题。最后邮件内容以只包含’.’的一行结束

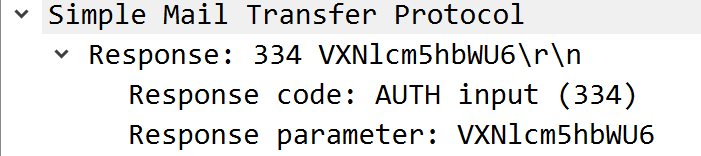


在wireshark抓包SMTP报文

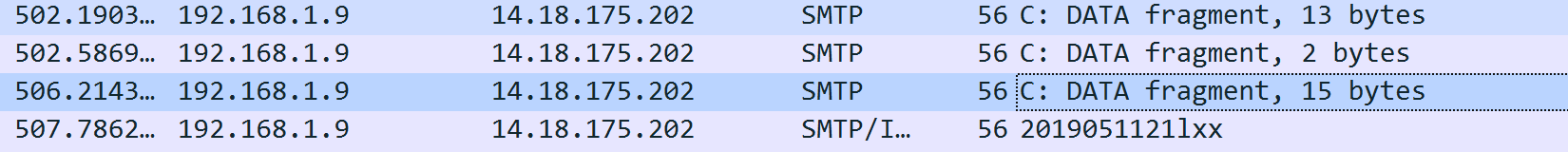


观察报文格式。



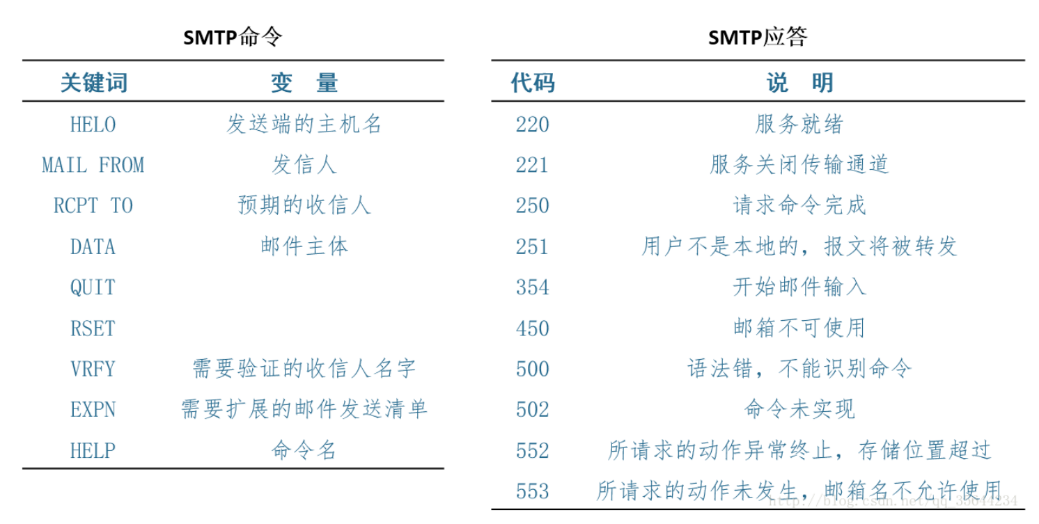


可以发现SMTP的报文格式跟FTP是相似的。请求报文都是由一个ascii码字符串格式的命令以及对应参数组成。请求报文是完全可读的；同样的，SMTP响应报文由一个响应码以及描述该响应码的字符串组成。



可以发现对于邮件的正文部分，客户端将以行为单位包装成SMTP报文并发送出去。

下面是关于SMTP命令的一些汇总

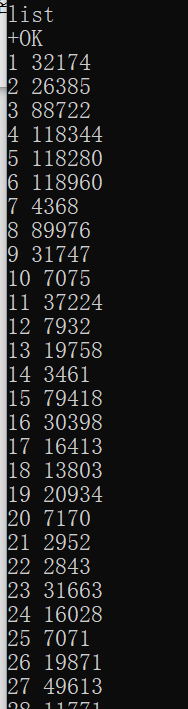


（3）、用wireshark抓包pop报文

同样的通过telnet命令与pop服务器取得连接。相对于SMTP协议，POP协议只需要直接输入邮箱名以及密码就可以完成授权，并不需要发送经过base64后的邮箱名以及密码。



可以用list命令来展示邮箱中的邮件（当然不是全部，本实验设置只能展示30天以内的）。返回的每行分别的邮件的编号以及邮件的大小。

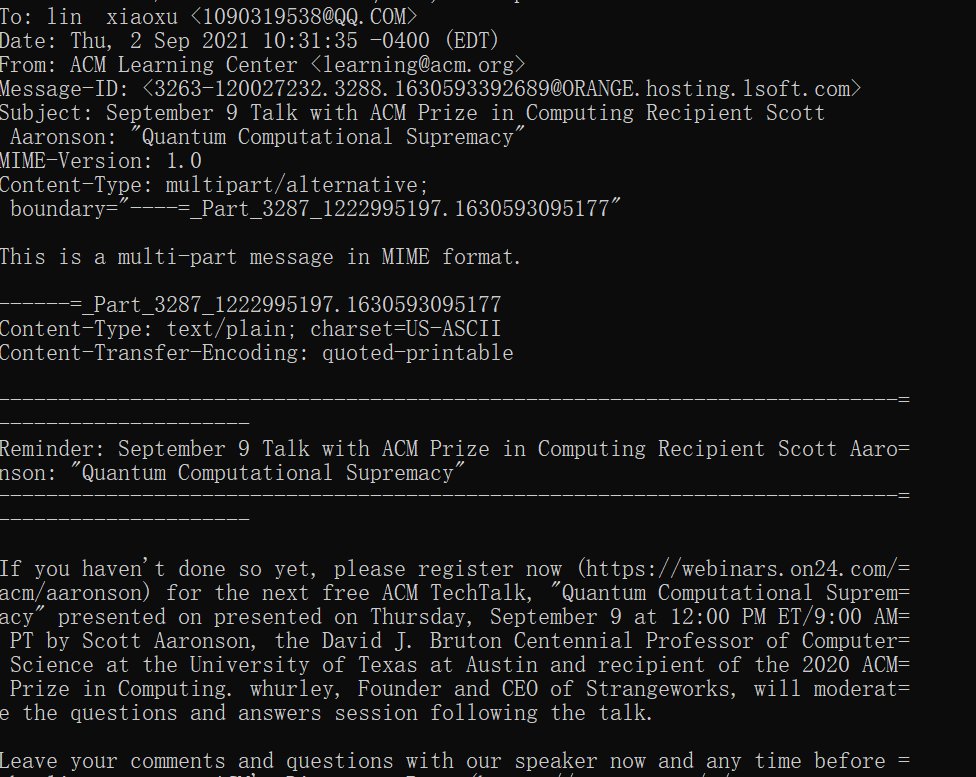


可以用retr命令来取邮件。

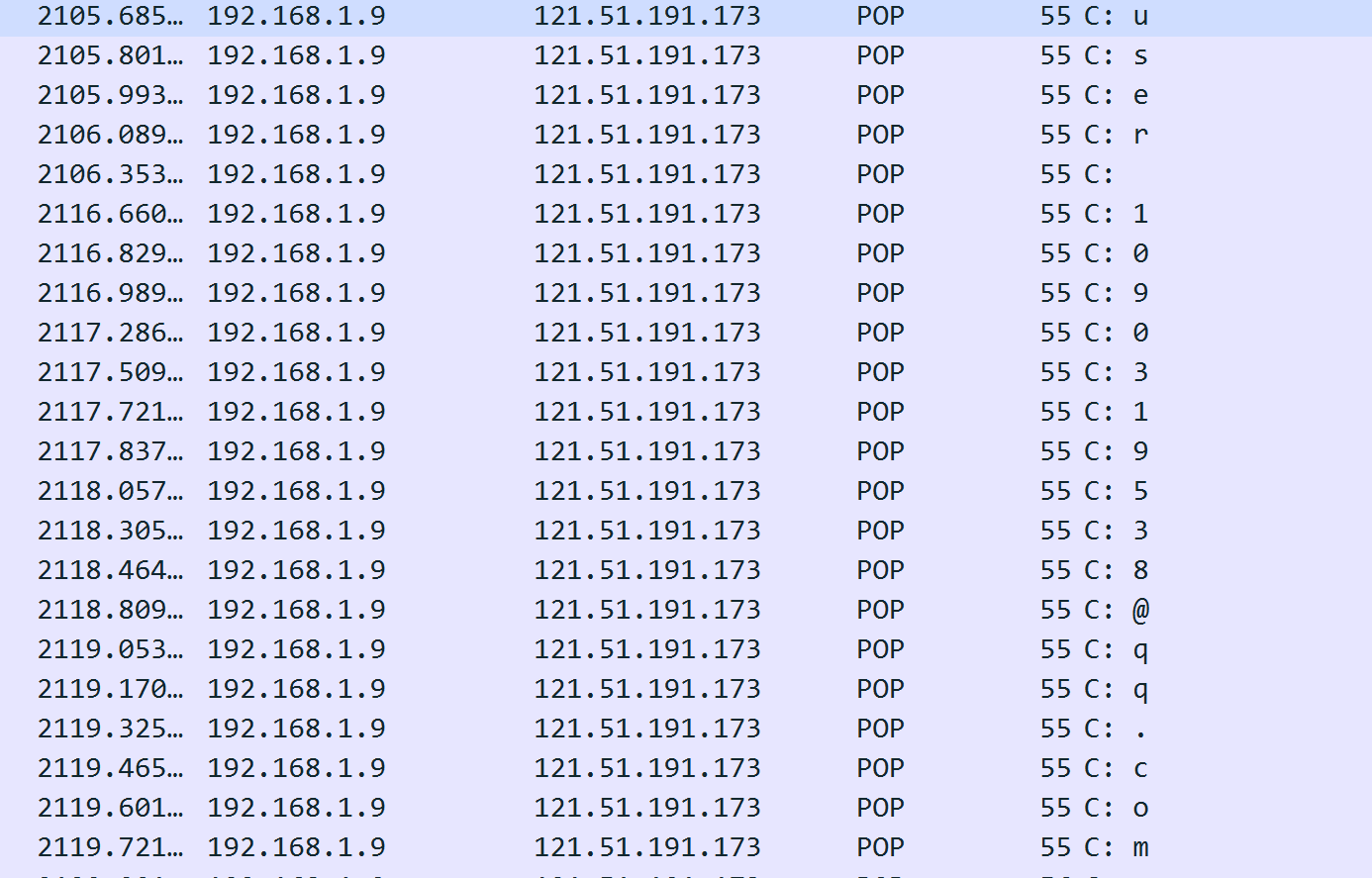
首先看到的是一些在邮件收发的一些信息。



接着才是关于邮件本身的信息。

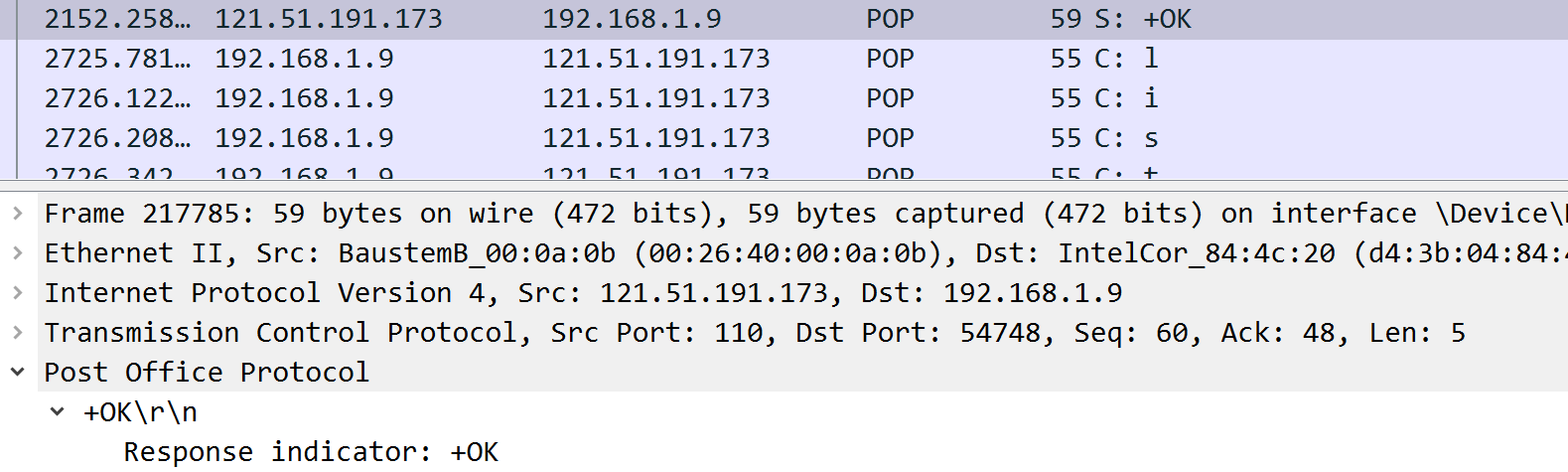


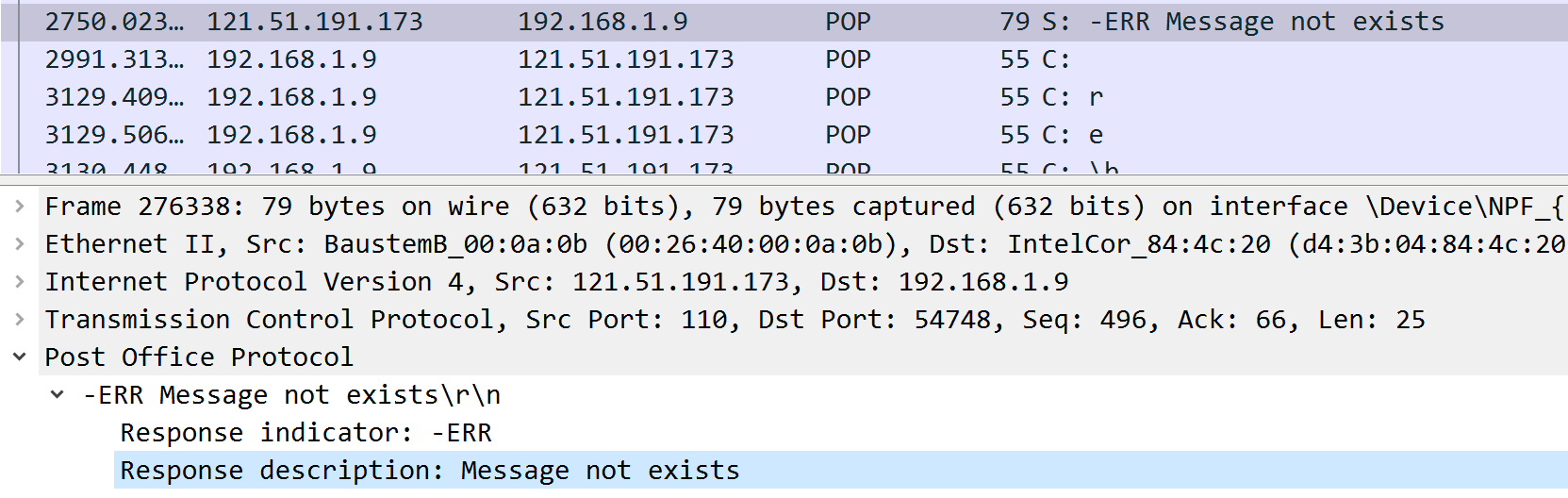
通过wireshark进行抓包



我不明白为什么客户端是以字节为单位进行传输的，但这并不影响POP协议的工作，POP协议是以行作为信息的基本单位。

POP的Response报文十分的简单，如果服务器能狗处理的了就发送“+OK”，当Response对应的请求是资源时，会在“+OK”后附上被请求的资源；如果处理不了就发送“-ERR”，一般会在其后附上报错的描述信息。





**六、实验总结**

本实验主要目的是通过wireshark对http,dns,ftp,SMTP/POP协议的报文进行抓包，分析报文格式以及协议的工作流程。相对于HTTP以及DNS报文。FTP以及SMTP\POP的报文格式相对简洁而且可读性高。FTP以及SMTP\POP协议都是由简短的ascii码字符串或者三位数响应码以及对应的描述信息组成的报文。

除了对各种应用层报文有更深入的了解之外，由于实验需要搭建相应服务器的需要，我也对搭建服务器的方法有了基本的了解。

为了让wireshark进行成功抓包，除了需要搭建服务器之外，还需要关闭本机防火墙让同一局域网内的另一台IP设备来进行对服务器的访问。这是为了报文经过进行局域网中让网卡捕获得到。如果用本机直接访问服务器，wiresshark是捕获不到的、一开始我还将wireshark捕获不到报文的原因归结为未正确搭建服务器，直到在搭建FTP服务器时才发现这个问题，以至于浪费了诸多的时间，搭建好的HTTP站点也未应用到实验中，

通过对DNS报文的分析，第一次了解到了报文的压缩技术。在DNS报文中通过指针的方式来避免域名字符串在报文中的冗余。但这带来的弊端就是报文的传输容错性减小（虽然说保证比特信息被正确传输是应用层之下的任务）。