计算机网络

16. ELECTRONIC MAIL

厦门大学软件学院

黄炜 助理教授

PART IV Network Applications

Ch 30 Electronic Mail Representation And Transfer

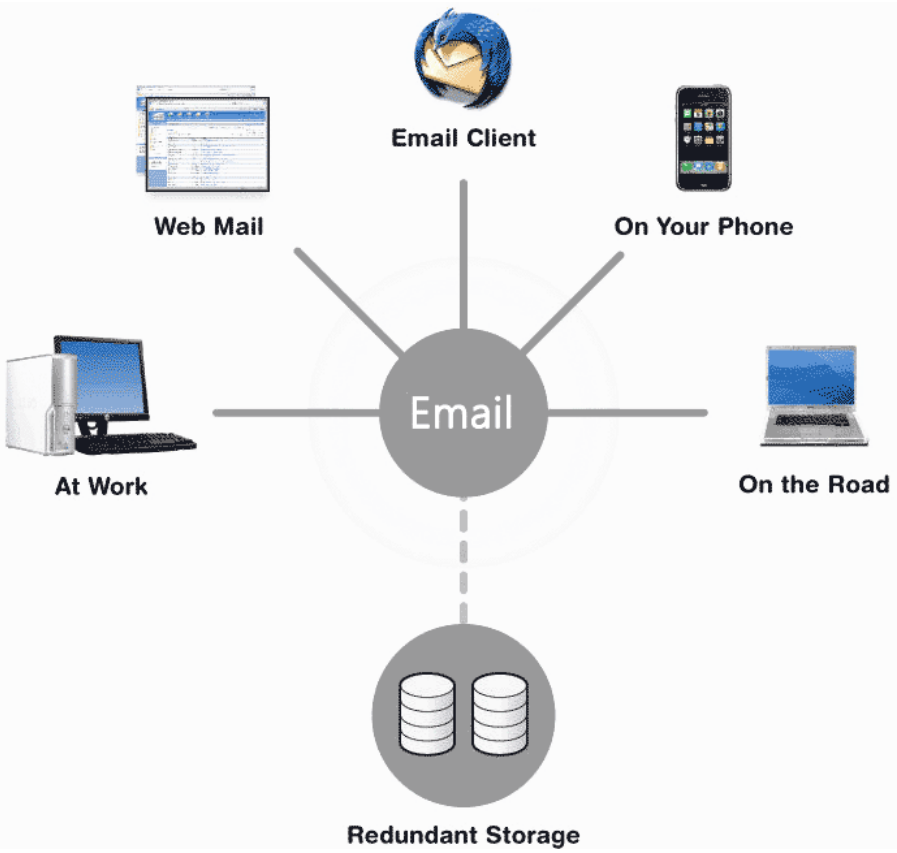
电子邮件表示和传输

本章内容

- 身边的电子邮件
- 电子邮件历史
- 电子邮件协议
  - 电子邮件的传输：**SMTP**
  - 电子邮件的传输扩展：**MIME**
  - 电子邮件的访问： **POP3，IMAP**

管理信息系统到邮件系统有多远

- 新闻发布系统
  - 没有收件人
- **BBS的站内信**
  - 别人不认
- **公网上的BBS**
  - 客户端收不到
- **符合收发协议的BBS站内信**
  - 再完善部分功能
- 邮件系统



查看Internet邮件头

- 如图，邮件头说明了什么？

```
From: ********@163.com
To: ******@163.com
Subject: =?utf-
8?B?Y291cnNlX3dodWFuZyDoh6rliqjlm57lpI06IHRlc3Q=?
=
Date: Mon, 6 May 2013 23:19:30 +0800
Message-ID: <5187CA02.F0B2B3.15380>
MIME-Version: 1.0
Content-Type: Text/HTML;
  charset="utf-8"
X-CM-TRANSID: EMCowEB5p3z_yYdRYrLXAg--.25871S2.re
X-CM-SenderInfo: xfrx22xhbzx3xdqjqiywtou0bp/
Delivered-To: XXX@163.com
Errors-to: *****@163.com
Return-path: *****@163.com
References:
<000001ce4a6d$1c1482c0$543d****$@163.com>
In-Reply-To:
<000001ce4a6d$1c1482c0$543d****$@163.com>
Content-Transfer-Encoding: base64
```

常用的邮件服务器和客户端

- 常用邮件服务器（包括但不限于）
  - 国际：**GMAIL**
  - 国内：**163，126，TOM**
  - 即将关闭：**MSN，YAHOO**
  - 单位：**XMU，IEEE**
- 常用邮件客户端（包括但不限于，非广告）
  - 国外：**Outlook，Thunderbird**
  - 国内：**Foxmail，QQMail，尚邮**
  - **Android**手机：**K9Mail**

历史

- 1972年BBN的Ray Tomlinson发明电子邮件系统并采用@符号。
- 电子邮件系统要求每个用户有一个地址
  - 形式是：用户名@主机域名
    - 这里@念作“at”，意思为“在”。
    - @左边是用户名或用户所使用的电子邮箱名。

whuang@xmu.edu.cn

三种代理

- **MUA（Mail User Agent）**
  - 邮件用户代理，主要是帮助用户读取、编写和回复邮件，再将这些信息转给MTA发送；有Outlook、Foxmail等。
- **MTA（Mail Transport Agent）**
  - 邮件传输代理，负责把邮件由一个服务器传到另一个服务器或邮件投递代理；在Unix系统有Sendmail、Qmail、Postfix等程序，在Win系统有Exchange、Lotus等。
- **MDA（Mail Delivery Agent）**
  - MDA邮件投递代理，主要负责将MTA接收的邮件，根据收件人地址将邮件投放到用户的邮箱里。在投放的过程中，还可以进行邮件过滤、自动恢复等功能。

三种邮件协议

- 常见的电子邮件协议：
  - SMTP（Simple Mail Transfer Protocol，简单邮件传输协议）
  - POP3（邮局协议）
  - IMAP（Internet邮件访问协议）
- 表示邮件的协议
  - MIME：扩展了电子邮件标准，使其能支持非ASCII字符、二进制附件等多种格式的邮件消息。
- 上网访问邮箱
  - HTTP，HTTPS
- 应用层协议

安全问题

- 上述协议默认使用明文传输
  - 除HTTPS=HTTP+SSL
- 可以用证书加密（SSL）增强安全性

Electronic Mail

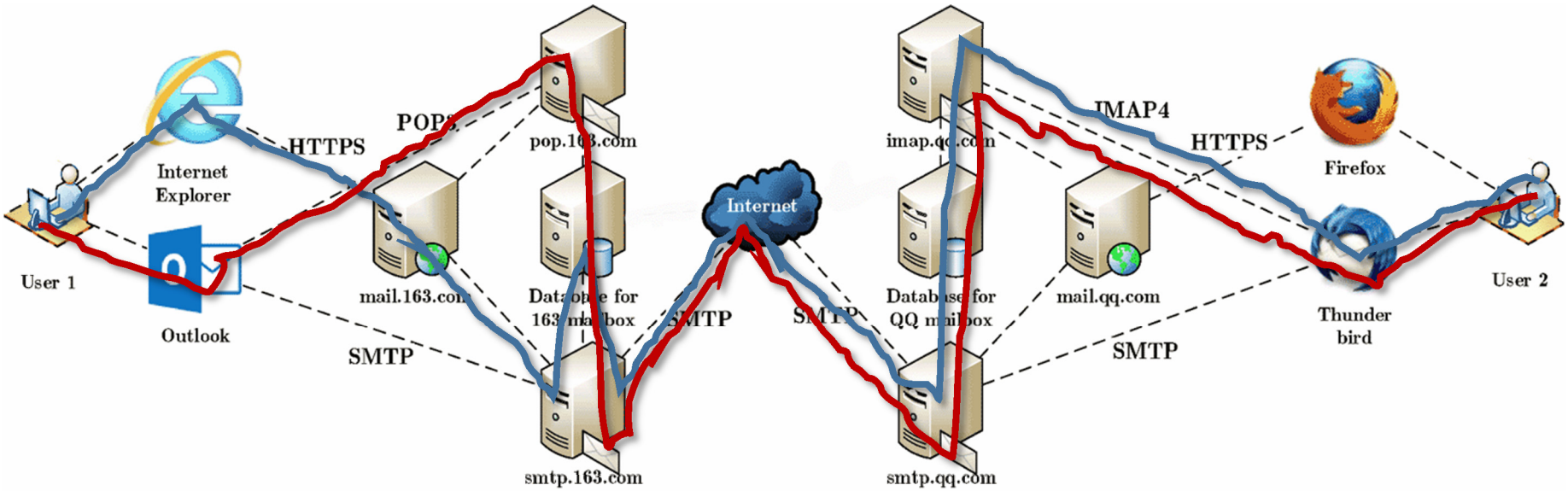
- **Email software is divided into two conceptually pieces:**
  - **An email interface application**
    - **A mechanism for a user to compose and edit outgoing messages as well as read and process incoming email**
  - **A mail transfer program**
    - **acts as a client to send a message to the mail server on the destination computer; the mail server accepts incoming messages and deposits each in the appropriate user's mailbox**

Electronic Mail

- The specifications used for Internet email can be divided into three broad categories as Figure 4.12 lists

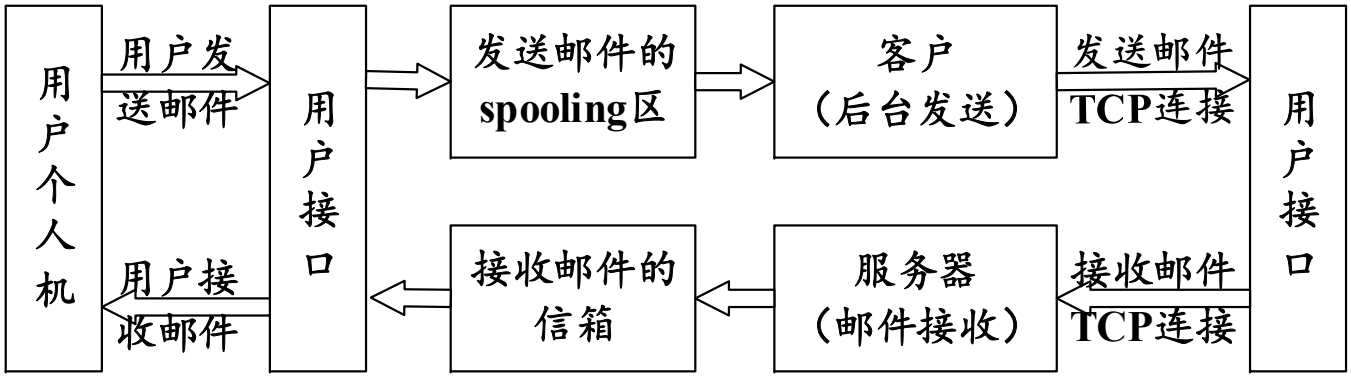
Type	Description
Transfer	A protocol used to move a copy of an email message from one computer to another
Access	A protocol that allows a user to access their mailbox and to view or send email messages
Representation	A protocol that specifies the format of an email message when stored on disk

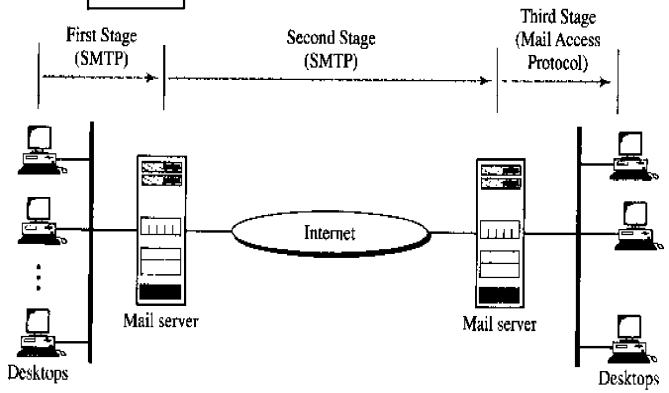
电子邮件如何传输


Internet Explorer
HTTPS
POP3
pop.163.com
User 1
Outlook
mail.163.com
Database for 163 mailbox
SMTP
smtp.163.com
Internet
SMTP
imap.qq.com
Database for QQ mailbox
IMAP4
mail.qq.com
SMTP
HTTPS
Firefox
Thunder bird
User 2
smtp.qq.com

TCP/IP电子邮件系统的模型

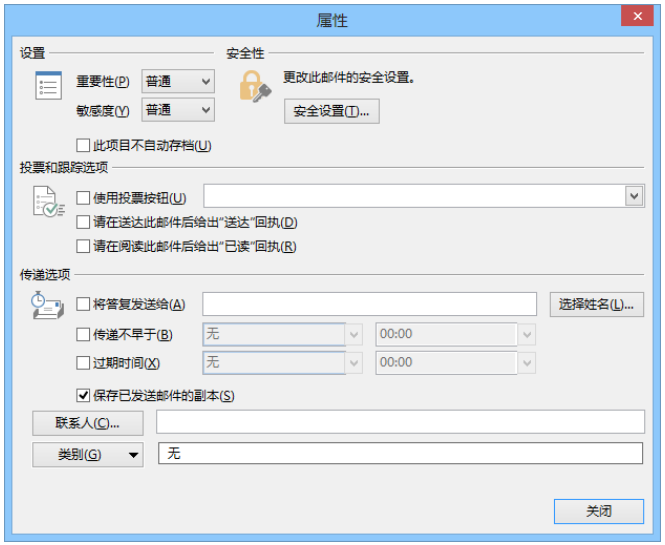
- TCP/IP电子邮件系统采用端到端传输方式
  - 发送方的MTA负责将邮件传送到接收方的MTA

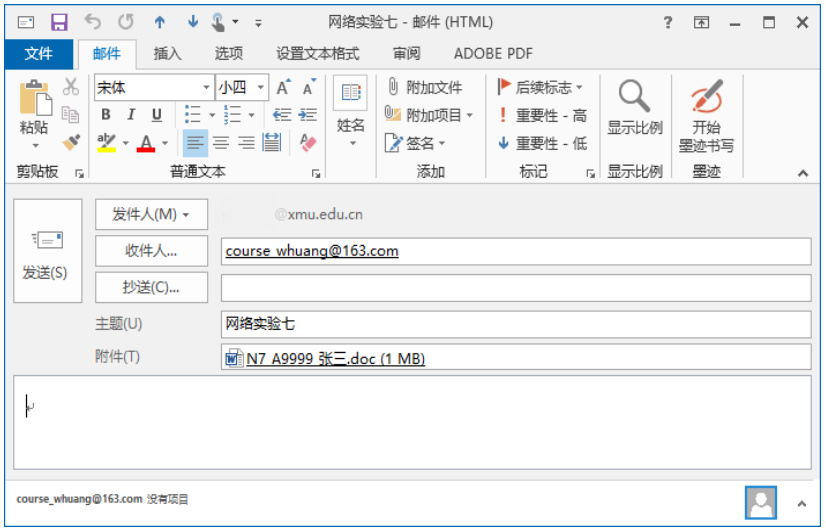




电子邮件的组成

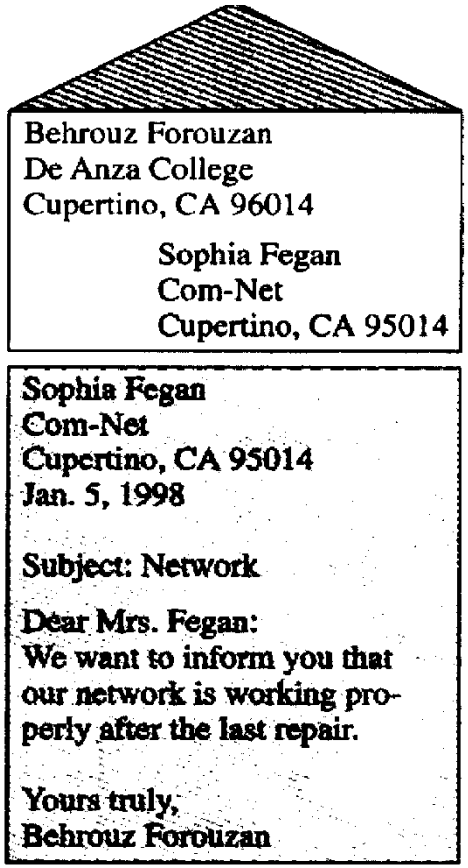
- 发件人，回复地址
- 收件人，抄送 ，密送
- 时间
- 主题
- 附件
- 扩展属性
  - 是否要求回执
  - 重要性

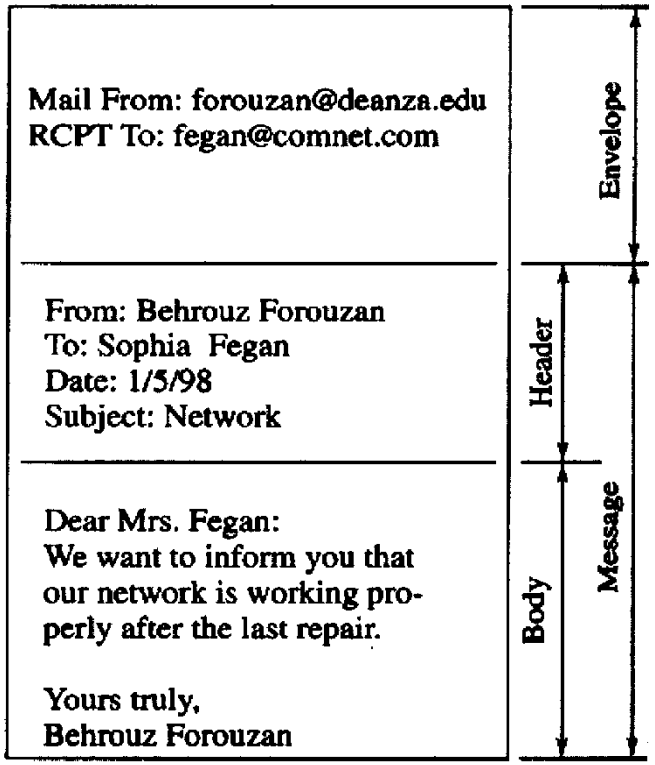




邮件格式

- 信封
- 消息
  - 邮件头
  - 邮件主体





电子邮件最主要的组成构件

- 用户代理
  - MUA, MTA, MDA

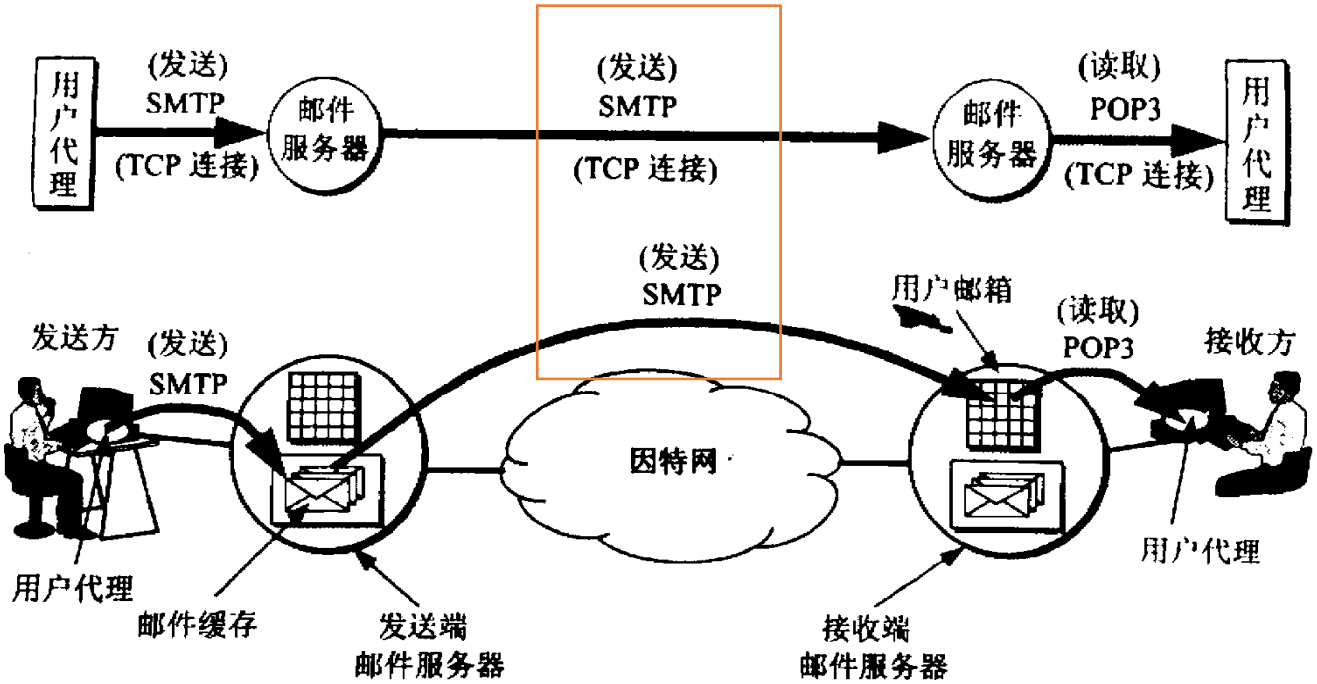
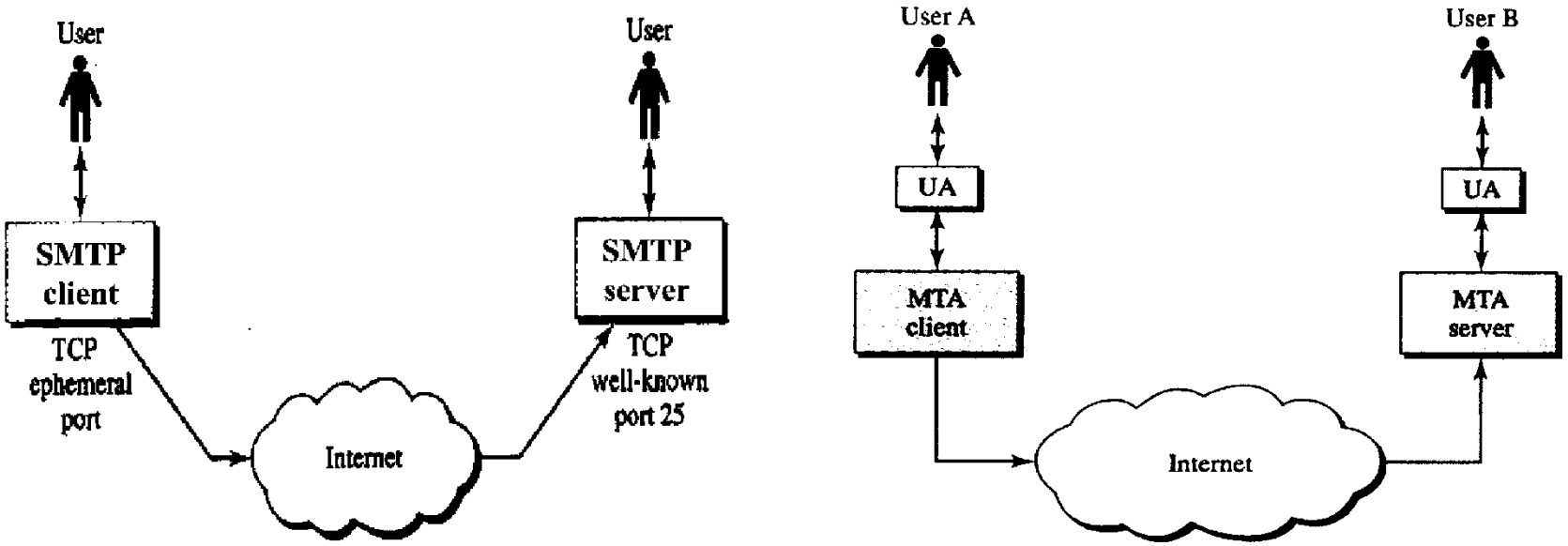


图 12-8　电子邮件的最主要的组成构件

发送邮件

- **Simple Mail Transfer Protocol (SMTP).**
  - **supports electronic mail on the Internet**
  - **provides for mail exchange between users on the same or different computers.**



SMTP

- The SMTP is the standard protocol that a mail transfer program uses.
  - 端口号：25 (plaintext), 465 (SMTPS, with SSL)
- SMTP can be characterized as
  - Follows a stream paradigm
  - Uses textual control messages
  - Only transfers text messages
  - Allows a sender to specify recipients' names and check each name
  - Sends one copy of a given message

SMTP

- SMTP has a restriction to send only textual content
  - MIME standard that allows email to include attachments such as graphic images or binary files
- SMTP can send a single message to multiple recipients
  - The protocol allows a client to list users and then send a single copy of a message for all users on the list

Example of SMTP Session

- **Figure 4.13**

```
Server:   220 somewhere.com Simple Mail Transfer Service Ready

Client:   HELO example.edu
Server::  250 OK

Client:   MAIL FROM:<John_Q_Smith@example.edu>
Server:   250 OK

Client:   RCPT TO:<Mathew_Doe@somewhere.com>
Server:   550 No such user here

Client:   RCPT TO:<Paul_Jones@somewhere.com>
Server:   250 OK

Client:   DATA
Server:   354 Start mail input; end with <CR><LF>.<CR><LF>
Client:   ...sends body of mail message, which can contain
Client:   ...arbitrarily many lines of text
Client:   <CR><LF>.<CR><LF>
Server:   250 OK

Client:   QUIT
Server:   221 somewhere.com closing transmission channel
```

SMTP命令列表

命令	描述
DATA	开始信息写作
EXPN<string>	验证给定的邮箱列表是否存在，扩充邮箱列表，也常被禁用
HELO<domain>	向服务器标识用户身份，返回邮件服务器身份
HELP<command>	查询服务器支持什么命令，返回命令中的信息
MAIL FROM<host>	在主机上初始化一个邮件会话
NOOP	无操作，服务器应响应OK
QUIT	终止邮件会话
RCPT TO<user>	标识单个的邮件接收人；常在MAIL命令后面可有多个rcpt to：
RSET	重置会话，当前传输被取消
SAML FROM<host>	发送邮件到用户终端和邮箱
SEND FROM<host>	发送邮件到用户终端
SOML FROM<host>	发送邮件到用户终端或邮箱
TURN	接收端和发送端交换角色
VRFY<user>	用于验证指定的用户/邮箱是否存在；由于安全方面的原因，服务器常禁止此命令

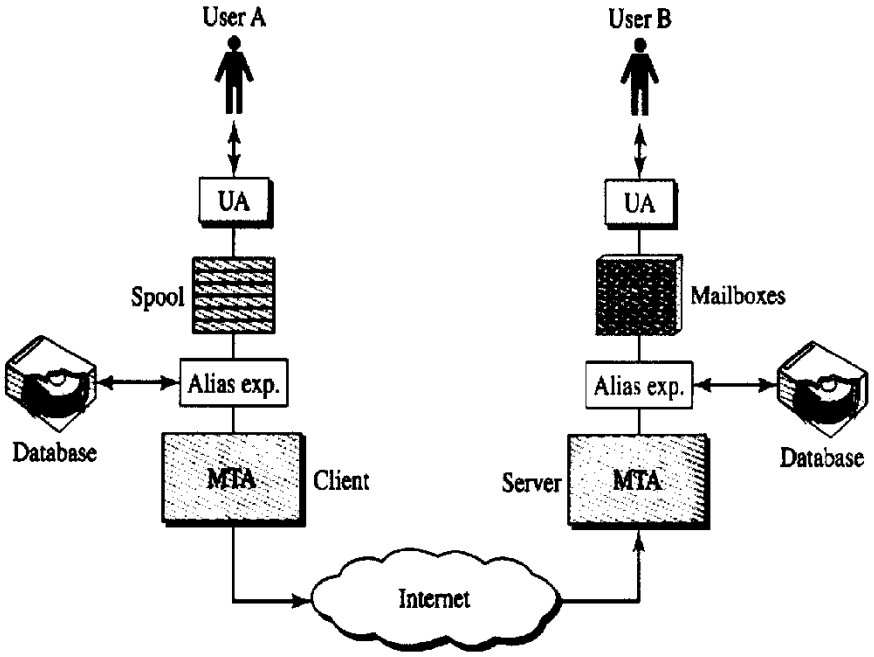
SMTP

- After UA creates a message, it's delivered to spool, which is a storage structure.
- The mail transfer system periodically checks the mail stored in the spool.
  - If The IP address of the server has been obtained?
  - If the receiver is ready?
- After the msg. has been received, it doesn't have to be read by the recipient immediately.
  - The mail can be stored in the mailbox of the receiver.

别名扩展

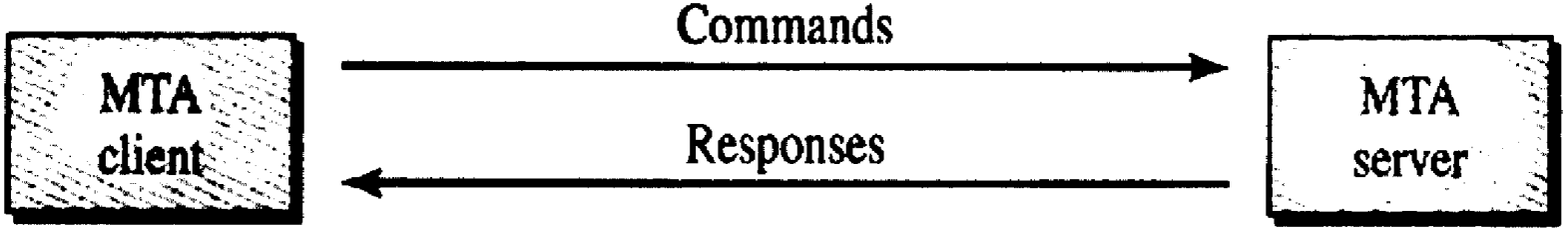
- 邮件转发器允许报文副本再发到多个目的地
  - 通常，服务器会查询一个小型的邮件别名数据库，把传入的收件方地址映射为一组地址 A，然后再向 A 中的每个地址转发一份副本。

– 别名映射可以



Command-and-Response

- SMTP uses commands and responses to transfer messages between an MTA client and an MTA server.
- Each command or reply is terminated by a two-character (carriage return and line feed) end-of-line token.
  - Commands are sent from the client to the server. The format of Command: KEYWORD: INFO



Responses

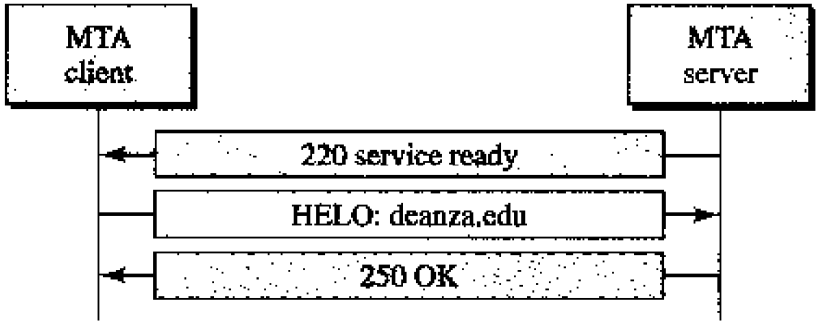
- **Example:**

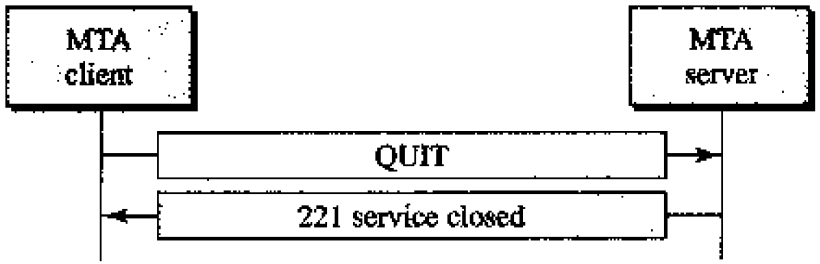
```
HELO: xyz.xmu.edu.cn
MAIL FROM: abc@xyz.xmu.edu.cn
RCPT TO: def@sina.com
```

- **Responses are sent from the server to the client.**
- **A response is a three-digit code that may be followed by additional textual information.**
- **Three phases: connection establishment, mail transfer, and connection termination.**

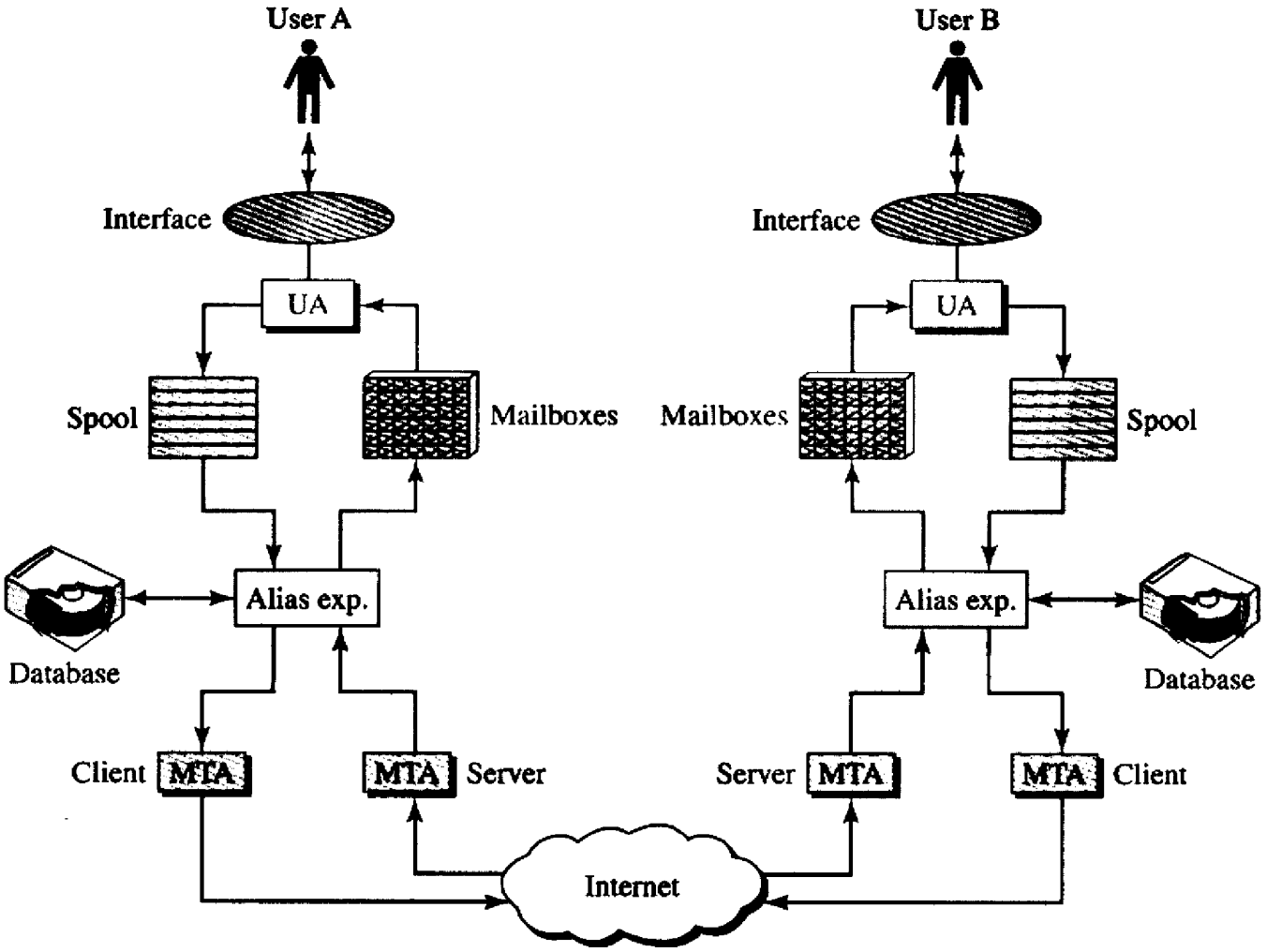
Mail Transfer Phases

- **Connection establishment:**
  - **After a client has made a TCP connection to the well-known port 25, the SMTP server starts the connection phase.**
- **Connection termination**
  - **After the message is transferred successfully, the client terminates the connection.**





The entire email system



Why we use MIME

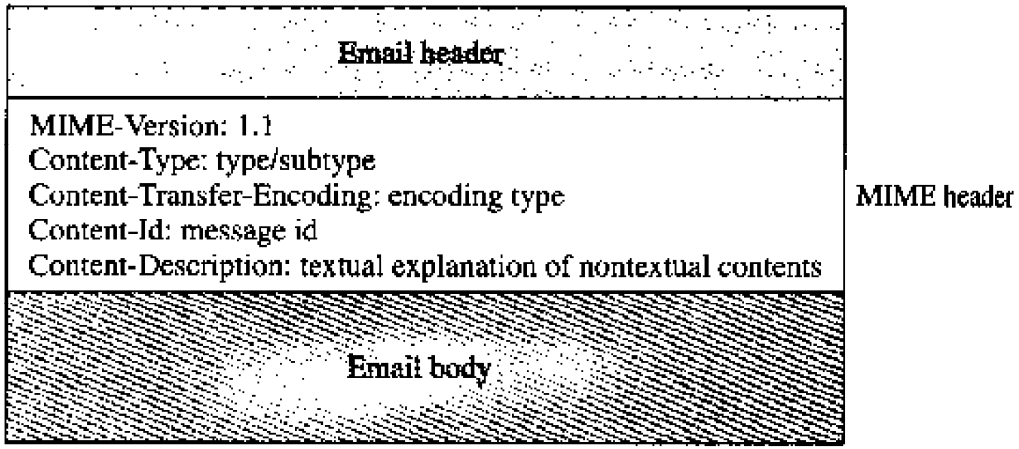
- SMTP is a simple mail transfer protocol.
  - RFC 2822
  - SMTP can send messages only in NVT(network virtual terminal) 7-bit ASCII format.
- Supports only 7-bit ASCII characters
  - Chinese, Russian, ...
  - Binary files or to send video or audio data.

What is MIME

- **Multipurpose Internet Mail Extensions**
  - **多用途互联网邮件扩展**
  - **MIME is a supplementary protocol that allows non-ASCII data sent through SMTP.**
    - **not a mail protocol and cannot replace SMTP.**
    - **only an extension to SMTP.**
  - **The server SMTP at the receiving side receives the NVT (网络虚拟终端) ASCII data and delivers it to MIME to be transformed back to the original data.**

MIME Headers

- **Five headers added to the original SMTP header section to define the transfer parameters**
  - **MIME-Version 标识MIME的版本**
  - **Content-Type 说明邮件的性质**
  - **Content-Transfer-Encoding 主体是如何编码的**
  - **Content-Id 邮件的唯一标识符**
  - **Content-Description 说明邮件是什么**



MIME头的参数值

- **MIME-Version: 1.0**
- **Content-Type: [type]/[subtype]; parameter**
  - **Text：用于标准化地表示的文本信息，文本消息可以是多种字符集和或者多种格式的；**
  - **Multipart：用于连接消息体的多个部分构成一个消息，这些部分可以是不同类型的数据；**
  - **Application：用于传输应用程序数据或者二进制数据；**
  - **Message：用于包装一个E-mail消息；**
  - **Image：用于传输静态图片数据；**
  - **Audio：用于传输音频或者音声数据；**
  - **Video：用于传输动态影像数据，可以是与音频编辑在一起的视频数据格式。**

MIME头的参数值

- **Content-Type: [type]/[subtype]; parameter**
  - 举例
    - **multipart/mixed**：存在附件
    - **multipart/related**：存在内嵌资源
    - **multipart/alternative**：只有纯文本与超文本正文
    - **application/xhtml+xml**（XHTML文档）
    - **image/jpeg**
- **Content-Transfer-Encoding: [mechanism]**
  - **7bit, 8bit, binary, quoted-printable, base64**

MIME的编码

- MIME编码格式

Type	*Description*
7bit	NVT ASCII characters and short lines
8bit	Non-ASCII characters and short lines
binary	Non-ASCII characters with unlimited-length lines
Base64	6-bit blocks of data are encoded into 8-bit ASCII characters
Quoted-printable	Non-ASCII characters are encoded as an equal sign followed by an ASCII code

- 支持的文件类型

Table 22.3 *Data types and subtypes in MIME*

Type	*Subtype*	*Description*
Text	Plain	Unformatted text
Multipart	Mixed	Body contains ordered parts of different data types
	Parallel	Same as above, but no order
	Digest	Similar to Mixed, but the default is message/RFC822
	Alternative	Parts are different versions of the same message
Message	RFC822	Body is an encapsulated message
	Partial	Body is a fragment of a bigger message
	External-Body	Body is a reference to another message
Image	JPEG	Image is in JPEG format
	GIF	Image is in GIF format
Video	MPEG	Video is in MPEG format
Audio	Basic	Single channel encoding of voice at 8 KHz
Application	PostScript	Adobe PostScript
	Octet-stream	General binary data (eight-bit bytes)

Base64 Encoding

- Base64
  - A solution for sending data made of bytes when the highest bit is not necessarily zero.
  - Transforms data to printable characters.
- Process
  - Divides the binary data into 24-bit blocks.
  - Each block is divided into 4 sections, each made of 6 bits.
  - Each 6-bit section is interpreted as one printable characters.

Base64编码过程

- 传递24bit：01001001 00110001 00111001
  - 先划分为4个6bit组，即:

    010010 010011 000100 111001

    (18:S) (19:T) (4:E) (57:5)
  - 对应的base64编码为：STE5。其ASCII码为：

    01010011 01010100 01000101 00110101
  - 三个字符用四个字符传。
    - 适用于ASCII码不多的情况。
- 码表
  - 0-25：A-Z；26-51：a-z；52-61：0-9；
  - 62：+，63：/

Example for MIME

```
From: "=?gb18030?****=?=" <*****@***.com>
To: "=?gb18030?B?****3dodWFuZw==?=" <****@***.com>
Subject: =?gb18030?B?u9i****do=?=
 =?gb18030?B?****k=?=
Mime-Version: 1.0
Content-Type: multipart/mixed;
        boundary="----=_NextPart_******"
Content-Transfer-Encoding: 8Bit
Date: Sat, 11 May 2013 13:16:33 +0800
X-Priority: 3
Message-ID: <****@****.com>
*****
This is a multi-part message in MIME format.
------=_NextPart_*******
Content-Type: multipart/alternative;
        boundary="----=_NextPart_*****";

------=_NextPart_*****
Content-Type: text/plain;
        charset="gb18030"
Content-Transfer-Encoding: base64
```

(Mime_separator)

Email Access

- Most users leave their computer running continuously and do not know how to configure and manage an email server
- ISPs began offering email services
  - An ISP runs an email server and provides a mailbox for each user
    - each ISP provides interface that allows a user to access their mailbox
- Email access follows one of two forms:
  - A special-purpose email interface application
  - A web browser that accesses an email web page

Email Access

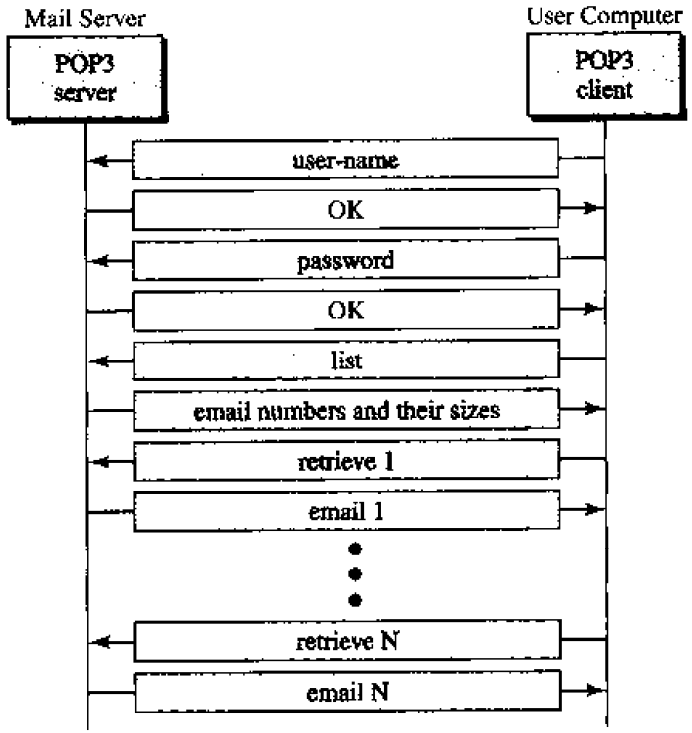
- An access protocol is distinct from a transfer protocol
  - access only involves a single user interacting with a single mailbox
  - transfer protocols allow a user to send mail to other users
- Characteristics of access protocols:
  - Provide access to a user's mailbox
  - Permit a user to view headers, download, delete, or send messages
  - Client runs on user's personal computer
  - Server runs on a computer that stores user's mailbox

Email Access

- Viewing a list of messages without downloading the message contents is useful
  - Especially in cases where the link between two parties is slow
  - For example, a user browsing on a cell phone may look at headers and delete spam without waiting to download the message contents

POP3

- Post Office Protocol (POP)
  - POP3允许用户通过PC机动态地检索邮件服务器上的邮件
  - 下载邮件，删除邮件
- Version 3
  - 端口号：
    - 110 (plaintext)
    - 995 (POP3S, with SSL)



POP3收邮件过程

- POP3服务器向客户发送一行欢迎词，进入授权状态。
- 授权状态
  - 客户发送USER命令给出用户在邮件服务器上的邮箱名，若是合法用户，服务器回答“+OK”。
  - 客户再发送PASS命令给出口令。POP3服务器确定用户是否有权访问该邮箱，若有权访问，服务器再次回答“+OK”，若是非法用户，服务器回答“-ERR”。
  - USER和PASS命令用口令方式对用户进行授权验证。

POP3收邮件过程

- 事务状态：
  - 若对用户的授权验证成功，则服务器申请资源与用户的邮箱关联，进入事务(transaction)状态
  - 在事务状态，服务器将存储的邮件分成一个个报文，并从1开始编号
    - 在事务状态，客户可以重复发送一些命令检索报文(RETR)、删除报文(DELE)(作删除标记)等。
- 更新状态：
  - 客户发出QUIT命令时，进入更新(update)状态
  - 所有有删除标记的报文被真正删去，最后关闭TCP连接，服务器释放资源，POP3会话结束

Example of POP3 Session

```
+OK Welcome to coremail Mail Pop3 Server (163coms[...
AUTH
-ERR Not support ntlm auth method
CAPA
+OK Capability list follows TOP USER PIPELINING...
USER y****7
+OK core mail
PASS **********
+OK 7 message(s) [231904 byte(s)]
STAT
+OK 7 231904
LIST
+OK 7 231904 1 9854 2 140937 3 10224 4 17139 ...
UIDL
+OK 7 231904 1 1tbiMAkaCFEAER4pHgAAsC 2 1tbiNQUc...
RETR 7
+OK 1698 octets
Received: from [192.168.7.128] (unknown [119.233.1...
f0uMueDze6FWdWBsy3zL/K1nBCoAwr8FFrAhNS4gN8AhsIy++3...
QUIT
+OK core mail
```

Example of POP3 Session

```
Received: from [192.168.7.128] (unknown [119.233.192.167])
        by smtp13 (Coremail) with SMTP id *****fVdw9pFRM59yBA--.***S2;
        Tue, 14 May 2013 16:31:45 +0800 (CST)
X-Coremail-DSSMTP: 119.233.192.167
Message-ID: <5***F670.5******@163.com>
Date: Tue, 14 May 2013 16:31:44 +0800
From: W** H**** <y****7@163.com>
User-Agent: Mozilla/5.0 (Windows NT 5.1; rv:17.0) Gecko/20130328 Thunderbird/17.0.5
MIME-Version: 1.0
To: y****7@163.com
Subject: Test a rar attach
Content-Type: multipart/mixed;
 boundary="------------*******20706"
X-CM-TRANSID:*****fVdw9pFRM59yBA--.***S2
X-Coremail-Antispam: ******************9EdanIXcx71UUUUU7v73
         ****************** YCTnIWIevJa73UjIFyTuYvjxU4mFCUUUUU
X-CM-SenderInfo: ****************** biJRIdCE9o8Tv-YAAAsQ

This is a multi-part message in MIME format.
------------*******20706
Content-Type: text/plain; charset=GB2312
```

Example of POP3 Session

```
Content-Transfer-Encoding: 7bit


------------*******20706
Content-Type: ********************;
 name="result.rar"
Content-Transfer-Encoding: base64
Content-Disposition: attachment;
 filename="result.rar"

****************************************AACOZRtLPytq0IdNQoAIAAAAHJl
c3VsdC50eHQAsOhgB**************************************dNHhnB4bx0x86eD5
I+CoS7d*********************************V4H8e3v5a/WubX52V88wwBSxn5eKz
0sy5OmimnT26******************Oab8/VXlYLQJruTj/5jjHmqqKkguvW5kUICb2OjP3
VdHj1Ij**********************************************++3MGE3QrE9OwHD
R6Ay******************1qVqWp5y9cTE5eQATF**********************UhPkYvygn
S5gHFg************************A==
------------*******20706--
```

IMAP4

- **Internet Mail Access Protocol (IMAP)**
  - 三种工作模式：离线、在线和断连方式。
  - **Version 4**
    - 端口号：**143 (plaintext), 993 (IMAPS, with SSL)**
- 先进行身份验证，鉴定登录名和口令，然后用户才获得访问邮箱的权利
- **IMAP协议适合使用多台计算机的用户**
  - **IMAP协议的基本思想是让邮件服务器维护一个中心数据库，使得能够被多台机器访问**
  - 不允许用户将邮件下载到自己的计算机上，只能在线访问邮箱，但可以只读邮件的某一部分

IMAP与POP3 的区别

- IMAP它只下载邮件的主题，并不下载内容
- 邮件客户端软件阅读邮件时才下载邮件内容
- 支持维护自己在服务器上的邮件目录
- 支持直接抓取邮件的特定部分（如文本）
- POP3是“脱机”协议，IMAP是联机协议

垃圾邮件（SPAM）

- 不请自来、强行塞入信箱的垃圾邮件
- SPAM的主要特性包括：
  - 未经消费者的同意
  - 与消费者需求不相关
  - 以诈欺的方式骗取邮件地址
  - 攻击性的广告：夸张不实、不健康、钓鱼网站
  - 散布的数量庞大
- 阻拦垃圾邮件的方法
  - 发信来源
  - 关键词、特征匹配

选作作业

- 下载邮件服务器软件，以localhost为域名，新建admin@localhost邮箱，尝试搭建邮件服务器，并用邮件客户端下载
  - Ipswitch IMail Server, WinMail
- 探究邮件服务器和客户端软件各有什么功能
- 用Omnipeek监听收发邮件的数据流

16. THANK YOU.

厦门大学软件学院

黄炜 助理教授