What is MIME?

- Multipurpose Internet Mail Extensions
- data format to exchange data over the Internet
- initially designed to overcome some limits of e-mail (RFC-821/822):
  - 7 bit channel
  - US-ASCII characters
  - lines shorter than 1000 characters
- today used to transmit / identify data in many other contexts, including the Web as the most important case

Base definition of MIME

- RFC-2045 “MIME part one: format of Internet message bodies”
- RFC-2046 “MIME part two: media types”
- RFC-2047 “MIME part three: message header extensions for non-ASCII text”
- RFC-2048 “MIME part four: registration procedures”
- RFC-2049 “MIME part five: conformance criteria and examples”
The MIME format

© Antonio Lioy - Politecnico di Torino (2003-2009)

### MIME extensions
- RFC-2183 “Communicating presentation information in Internet messages: the Content-Disposition header field”
- RFC-2231 “MIME parameter value and encoded word extensions: character sets, languages, and continuations”
- RFC-2557 “MIME encapsulation of aggregate documents, such as HTML (MHTML)”
- RFC-3023 “XML media types”
- RFC-3676 “The Text/Plain format and DelSp parameters”
- RFC-3798 “Message disposition notification”

### MIME Header (I)
- **Mime-Version:** 1.0
- **Content-Type:** type/subtype ; parameters
  - text, application, audio, image, message, model, video, multipart
  - with subtype (e.g. image/gif)
  - optional parameters (e.g. name=bb.gif)
  - parameters have often the form “name = value”
- **Content-Transfer-Encoding:**
  - not encoded: 7bit, 8bit, binary
  - encoded: base64, quoted-printable

### Header MIME (II)
- **Content-Id:**
  - identifier of the data, generated by the sender
  - optional and rarely used
- **Content-Description:**
  - textual description of the data, generated by sender
  - optional and rarely used
- **Content-Disposition:** disposition ; parameters
  - disposition: inline, attachment
  - filename=
  - creation-date= , modification-date= , read-date=
  - size=
MIME non-encodings

- 7bit
  - lines < 1000 characters
  - ASCII characters (i.e. with MSB=0)
  - is the default
- 8bit
  - lines < 1000 characters
  - characters may have MSB=1
- binary
  - lines can be also > 1000 characters
  - characters may have MSB=1

“Base64” encoding

- 6 bit = 1 printable ASCII character
- message size: +33%
- generates lines not longer than 76 printable ASCII characters
- useful for binary messages

Example: “C’è” becomes “Qyfo”

C’è = 0x43 0x27 0xe8
= 0100 0011 0010 0111 1110 1000
= 010000 110010 011111 101000
= Q y f o

Base64 conversion table

<table>
<thead>
<tr>
<th>0</th>
<th>A</th>
<th>17 R</th>
<th>34 l</th>
<th>51 z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 B</td>
<td>18 S</td>
<td>35 j</td>
<td>52 0</td>
<td></td>
</tr>
<tr>
<td>2 C</td>
<td>19 T</td>
<td>36 k</td>
<td>53 1</td>
<td></td>
</tr>
<tr>
<td>3 D</td>
<td>20 U</td>
<td>37 l</td>
<td>54 2</td>
<td></td>
</tr>
<tr>
<td>4 E</td>
<td>21 V</td>
<td>38 m</td>
<td>55 3</td>
<td></td>
</tr>
<tr>
<td>5 F</td>
<td>22 W</td>
<td>39 n</td>
<td>56 4</td>
<td></td>
</tr>
<tr>
<td>6 G</td>
<td>23 X</td>
<td>40 o</td>
<td>57 5</td>
<td></td>
</tr>
<tr>
<td>7 H</td>
<td>24 Y</td>
<td>41 p</td>
<td>58 6</td>
<td></td>
</tr>
<tr>
<td>8 I</td>
<td>25 Z</td>
<td>42 q</td>
<td>59 7</td>
<td></td>
</tr>
<tr>
<td>9 J</td>
<td>26 a</td>
<td>43 r</td>
<td>60 8</td>
<td></td>
</tr>
<tr>
<td>10 K</td>
<td>27 b</td>
<td>44 s</td>
<td>61 9</td>
<td></td>
</tr>
<tr>
<td>11 L</td>
<td>28 c</td>
<td>45 t</td>
<td>62 +</td>
<td></td>
</tr>
<tr>
<td>12 M</td>
<td>29 d</td>
<td>46 u</td>
<td>63 /</td>
<td></td>
</tr>
<tr>
<td>13 N</td>
<td>30 e</td>
<td>47 v</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 O</td>
<td>31 f</td>
<td>48 w</td>
<td>(pad)</td>
<td></td>
</tr>
<tr>
<td>15 P</td>
<td>32 g</td>
<td>49 x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 Q</td>
<td>33 h</td>
<td>50 y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“Quoted-printable” encoding

- only the ASCII characters > 127 are encoded
- encoded as: ‘=hh’
- encoding dependent on the charset (!)
- variable increase of the message size
- lines not longer than 76 characters
- lines split by adding ‘=’ as the last character
- also the ‘=’ is encoded, as ‘=3D’
- example: “C’è” becomes “C’=e8” (in ISO-8859-1)

MIME charsets

Specified with the parameter charset=

- US-ASCII
- ISO-8859-1 (Latin-1)
- ISO-8859-2 (Latin-2)
- ISO-8859-3 (Latin-3)
- ISO-8859-4 (Latin-4)
- ISO-8859-5 (Latin/Cyrillic)
- ISO-8859-6 (Latin/Arabic)
- ISO-8859-7 (Latin/Greek)
- ISO-8859-8 (Latin/Hebrew)
- ISO-8859-9 (Latin-15)

ISO-8859-15

- also known as Latin-9
- adaptation of ISO-8859-1 to the following symbols:
  - Euro
  - characters for transliteration of Russian words
  - some French ligatures
- at the price of losing the pipe symbol ( | ) and some standalone diacritic symbols (umlaut, cedilla, …), but not when applied to letters
- note: windows-1252 includes all the printable characters of iso-8859-1/-15 but some ones with different numeric codes
### MIME Content-Type - Examples (I)

- **text/plain**
  - Normal ASCII text (7 bit, US-ASCII charset)
  - Optional parameter "charset=..." typically requires also "encoding=..."
- **text/enriched**
  - Text with special formatting sequences (e.g. `<bold>... </bold>`) and `text/html, text/xml, text/css, text/csv, ...`
- **video/mpeg, video/quicktime, ...**
- **audio/basic**
  - Mono, 8 bit, mu-law ISDN encoding, 8 KHz sampling

### MIME Content-Type - Examples (II)

- **multipart/mixed**
  - Contains multiple parts, to be displayed according to received order
- **multipart/parallel**
  - Contains multiple parts, to be displayed without a predefined order
- **multipart/alternative**
  - Contains the same data in alternative formats, in order of increasing fidelity
- **multipart/digest**
  - Composition of multiple e-mail messages (every part is message/rfc822)

### MIME Content-Type - Examples (III)

- **message/rfc822**
  - Contains an e-mail (e.g. for forwarding)
- **message/partial**
  - Part of an e-mail message that has been split
- **message/external-body**
  - Reference to external data (Web, FTP, ...)
- **image/gif, image/jpeg**
  - Image in GIF format or JPEG with JFIF encoding
- **application/postscript, application/pdf**
  - Postscript or PDF document
- **application/octet-stream**
  - Generic application format
MIME extensibility

- New functions can be added in the future for:
  - Content-type
  - Content-transfer-encoding
  - Content-disposition

- Until an extension is not documented inside an RFC, it must be considered as "private" and must begin with x- or X-.

- List of current specifications (managed by IANA):

  http://www.iana.org/assignments/media-types/

Examples: MIME in e-mail

From: Antonio Lioy <lioy@polito.it>
To: Antonio Lioy <lioy@polito.it>
Subject: Test di MIME

This is a multi-part message in MIME format.

--0107040803040507
Content-Type: text/plain; charset=ISO-8859-1; format=flowed
Content-Transfer-Encoding: quoted-printable

Questa è la parte di testo.

--0107040803040507
Content-Type: application/x-zip-compressed; name="tesi.zip"
Content-Transfer-Encoding: base64
Content-Disposition: inline; filename="tesi.zip"

--0107040803040507
Content-Type: application/vnd.ms-excel; name="voti.xls"
Content-Transfer-Encoding: base64
Content-Disposition: inline; filename="voti.xls"

--0107040803040507--