

Networking Fundamentals: IP, DNS, URL, MIME

Computer Science and Engineering ■ College of Engineering ■ The Ohio State University

Lecture 11

Internet Protocol (IP) Addresses

- A unique 32-bit number
 - Assigned to device connected to internet
 - An address for delivery of packets
- Written in “dotted-decimal” notation
 - Divided into 4 fields separated by “.”
 - Each field is 8 bits, ie 0-255 decimal

```
10100100011010110111011000000110
10100100.01101011.01111011.00000110
164.107.123.6
```
- Some are reserved: eg, 127.0.0.1

Abstract Value vs Encoding

- Abstraction: 32-bit integer value
- Encodings
 - Dotted decimal
 - Dotted hex
 - Dotted octal
 - Hexadecimal
 - Decimal
 - Binary
 - Etc...
- Recall: abstraction, representation,
correspondence relation

Address Space

- Organizations are allocated blocks of contiguous address to use
- 32 bits means 4 billion addresses
 - Population of the earth: 7 billion
 - Not enough addresses to go around!
- The end is predictable
 - Techniques like NAT developed to help
- In fact, the end has come!
 - Feb 2011: Last block was allocated

IPv6

- 128 bits
 - $\sim 10^{40}$ addresses; we're good for a while
 - A growing fraction of IP traffic
 - [GoogleIPv6 statistics](#)
- Recommended format (canonical):
 - Divide into 8 fields separated by ":"
 - Each field is 4 hex digits (0-FFFF), ie 16 bits
 - Omit *leading* 0's in a field
 - If there are *consecutive* fields with value 0, compress them as "::"
 - Compress *at most one* such set of 0's
 - Otherwise encoding could be ambiguous
 - Compress the longest sequence

Canonical Format: Uniqueness

2001:0db8:0000:0000:0000:ff00:0042:8329

2001:0db8:0000:0000:0000:ff00:0042:8329

2001:db8:0:0:0:ff00:42:8329

2001:db8:0:0:0:ff00:42:8329

2001:db8::ff00:42:8329

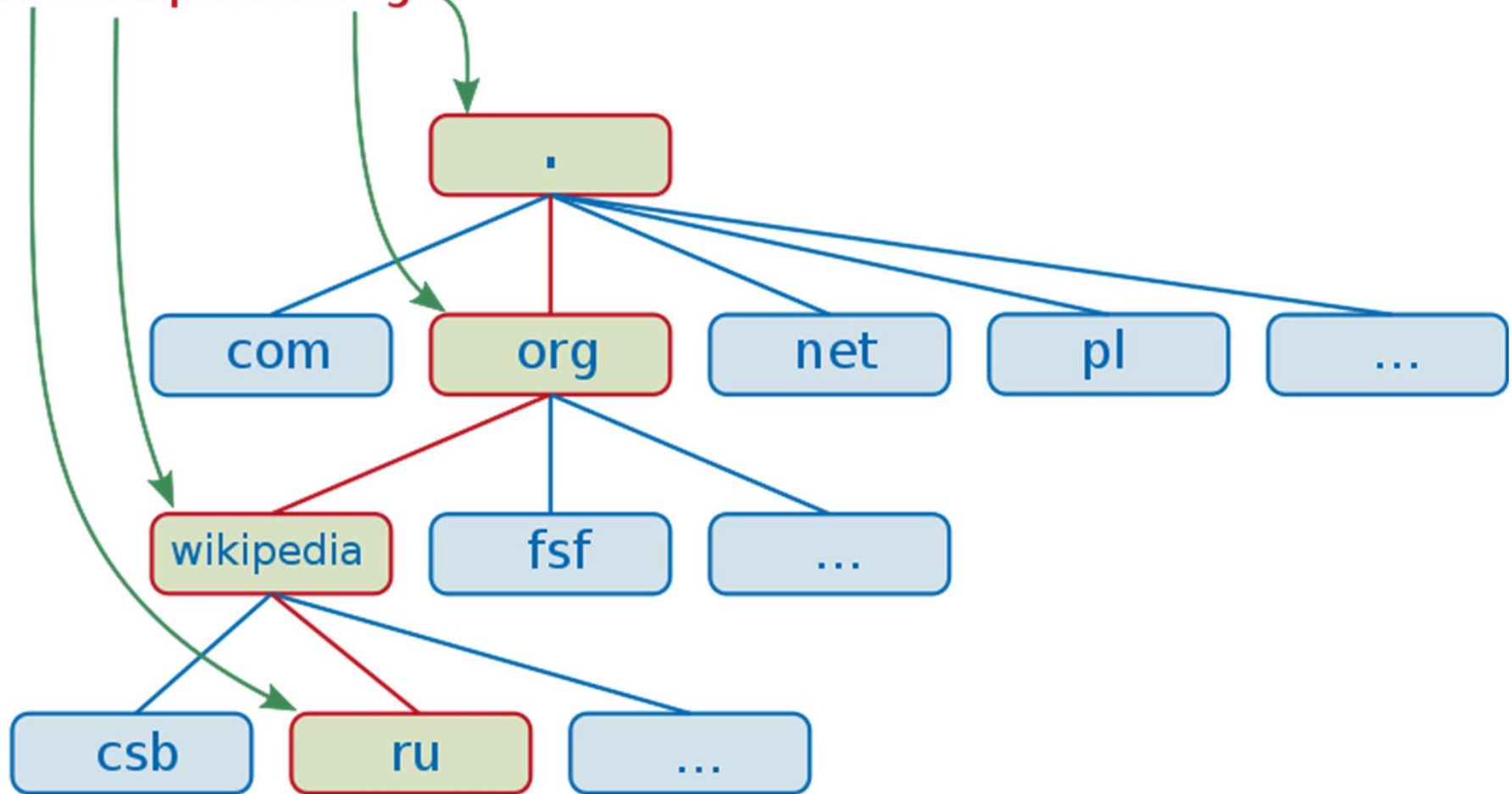
Domain Names

- String corresponds to an IP address
 - `web.cse.ohio-state.edu` is easier than `164.107.123.6`
- Case insensitive: Lower-case standard
- A partial map (almost)
 - DNS maps lower-case strings → IP addresses
 - Multiple strings can map to same address!
 - Some strings map to multiple addresses (unusual)!

Domain Name Hierarchy

- Separated by .'s
 - Don't confuse with dotted decimal!
- Right-to-left hierarchy
 - Top-level domain is right-most field
 - edu, com, net, gov, countries (ca, it, ...)
 - Second-level domain to its left
 - Then third, fourth, etc, no limit
`www.sos.state.oh.us`
- Hostname + Domain Name = Fully Qualified Domain Name (FQDN)
`stdlinux.cse.ohio-state.edu`

ru.wikipedia.org.

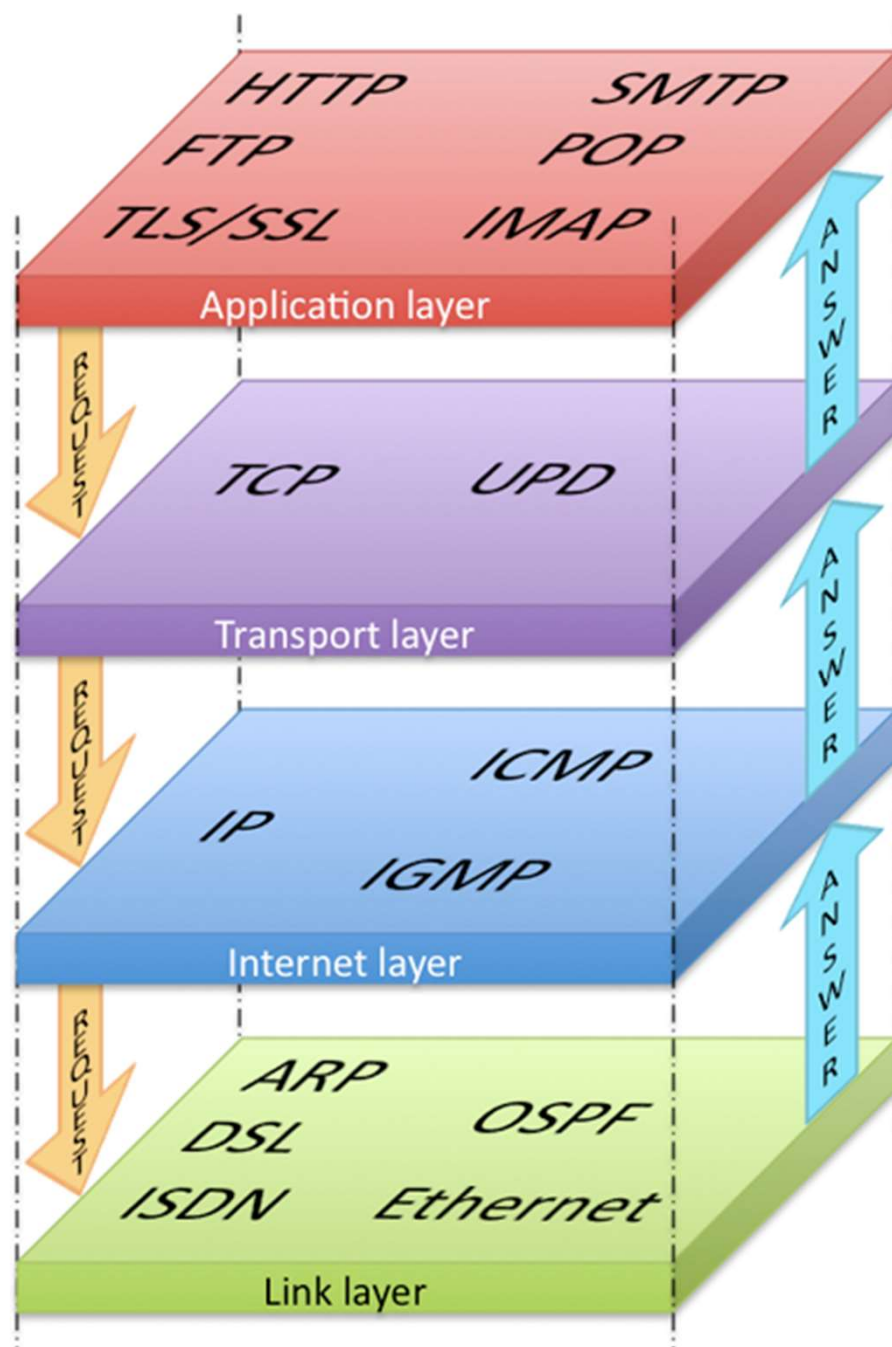


Name Servers

- Act as a phonebook for lookup
- Client view:
 - Given a FQDN, return IP address
 - Partial map: FQDNs → IP addresses
 - See host, whois
- Implementation view:
 - Hierarchical by domain
 - Local caching for recently retrieved items

Protocols

- Systematic ordering of messages
 - Phone rings
 - Callee answers by saying “Hello”
 - Caller answers by saying “Hello”
- Different protocols use different messages, different sequencing, etc
 - In Italy, callee answers by saying “*Pronto*”



Network Layering: Abstraction

- One protocol is built on top of another
 - Application level: FTP, HTTP, SSH, SMTP, TELNET
 - Transport: TPC, UDP
 - Internet: IP
- Each protocol assumes certain behavior from layer below
 - IP routes packets to destination (unreliable)
 - TCP creates a reliable, in-order channel
 - HTTP delivers web pages

Network Ports

- A single host has many ports
- Application-level protocols have default port
 - ftp -> 20
 - http -> 80
 - imap -> 143
 - ssh -> 22
 - smtp -> 25
 - telnet -> 23
- A “web server” is just a program, running, waiting, listening for a call (on port 80)
 - See telnet

URL

- Uniform Resource Locator
`scheme://FQDN:port/path?query#fragment`
- Schemes include http, ftp, mailto, file...
 - Case insensitive, but prefer lower case
- Port is optional (each scheme has default)
 - 80 for http
- Variety of formats, depending on scheme
`http://www.osu.edu/news/index.php`
`ftp://doe@ftp.cse.ohio-state.edu`
`mailto://brutus.1@osu.edu`
- FQDN is case insensitive, prefer lower case

Document Root

- ❑ Web server configured to serve documents from a location in file system
 - “document root”: `/class/3901`
 - File: `/class/3901/labs/lab2.html`
 - URL:
`http://www.cse.osu.edu/labs/lab2.html`
- ❑ Slashes in path should be for server's OS (but forward slashes are common)
- ❑ Virtual servers: multiple doc roots
- ❑ Proxy servers: remote doc roots

Abstract Value and Encoding

- Concrete invariant (convention)
 - No space, ;, :, & in encoding
 - To represent these characters in abstract value, use %hh instead (hh is ASCII code in hex)
 - %20 for space
 - Q: What about % in abstract value?
- Recall: *correspondence relation*
- Known as “URL encoding”, or “percent encoding”

URL Encoding

Reserved characters after percent-encoding

| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ! | # | \$ | & | ' | (|) | * | + | , | / | : | ; | = | ? | @ | [|] |
| %21 | %23 | %24 | %26 | %27 | %28 | %29 | %2A | %2B | %2C | %2F | %3A | %3B | %3D | %3F | %40 | %5B | %5D |

Common characters after percent-encoding (ASCII or UTF-8 based)

| | | | | | | | | | | | | | | | |
|-------------------------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| newline | space | " | % | - | . | < | > | \ | ^ | _ | ` | { | | } | ~ |
| %0A or %0D or %0D%0A | %20 | %22 | %25 | %2D | %2E | %3C | %3E | %5C | %5E | %5F | %60 | %7B | %7C | %7D | %7E |

MIME

- Multipurpose Internet Mail Extensions
 - Used to be for mail attachments
- Content Type: How to interpret a file
 - File is a blob of bits (encoding)
 - How to map this blob into (abstract) value?
Colors, sounds, characters, etc?
 - Recall: *correspondence relation*
- Syntax: type/subtype
 - text/plain, text/html, text/css, text/javascript
 - image/gif, image/png, image/jpeg
 - video/mpeg, video/quicktime
- Transfer encoding: A *layered* encoding
 - quoted-printable, base64

Example: Multiple Parts

MIME-Version: 1.0

Content-Type: multipart/mixed; **boundary=aFrontierString**

This is a message with multiple parts in MIME format.

--aFrontierString

Content-Type: text/plain

This is the body of the message.

--aFrontierString

Content-Type: application/octet-stream

Content-Transfer-Encoding: base64

PGh0bWw+CjAgPGhlYWQ+CjAgPC9oZWFKPgogIDxib2R5PgogICAgPHA
+VGhpcyBpcyB0aGUg

Ym9keSBvZiB0aGUgbWVzc2FnZS48L3A+CjAgPC9ib2R5Pgo8L2h0bWw
+Cg==

--aFrontierString--

Example: Content Type

MIME-Version: 1.0

Content-Type: multipart/mixed; boundary=aFrontierString

This is a message with multiple parts in MIME format.

--aFrontierString

Content-Type: text/plain

This is the body of the message.

--aFrontierString

Content-Type: application/octet-stream

Content-Transfer-Encoding: base64

PGh0bWw+CjAgPGhlYWQ+CjAgPC9oZWFKPgogIDxib2R5PgogICAgPHA
+VGhpcyBpcyB0aGUg

Ym9keSBvZiB0aGUgbWVzc2FnZS48L3A+CjAgPC9ib2R5Pgo8L2h0bWw
+Cg==

--aFrontierString--

Example: Transfer Encoding

MIME-Version: 1.0

Content-Type: multipart/mixed; boundary=aFrontierString

This is a message with multiple parts in MIME format.

--aFrontierString

Content-Type: text/plain

This is the body of the message.

--aFrontierString

Content-Type: application/octet-stream

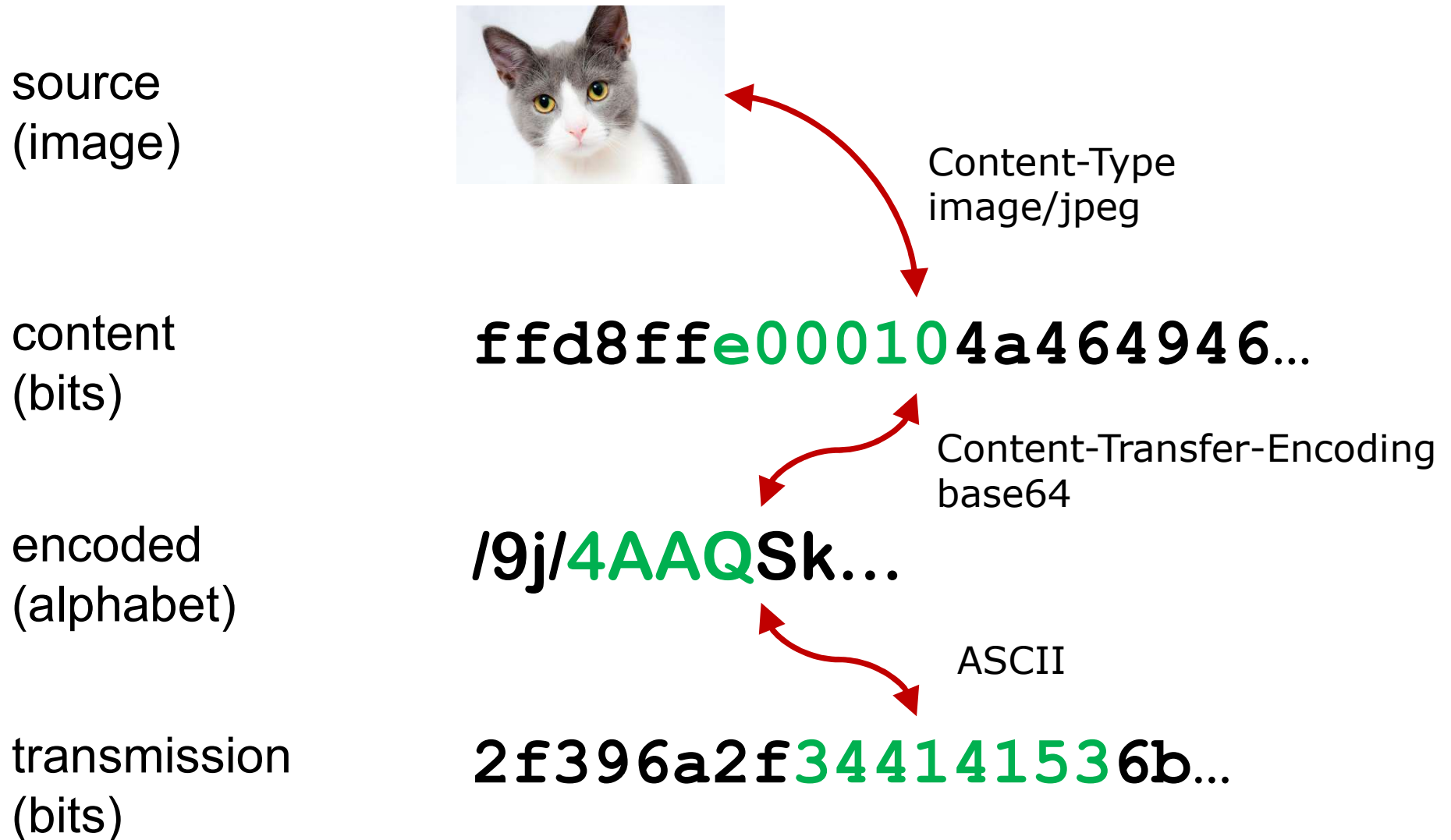
Content-Transfer-Encoding: base64

PGh0bWw+CiAgPGhlYWQ+CiAgPC9oZWFKPgogIDxib2R5PgogICAgPHA
+VGhpcyBpcyB0aGUg

Ym9keSBvZiB0aGUgbWVzc2FnZS48L3A+CiAgPC9ib2R5Pgo8L2h0bWw
+Cg==

--aFrontierString--

Layered Encoding



Base64 Alphabet

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

Base64 Encoding

| source ASCII (if <128) | M | | | | | | | | a | | | | | | | | n | | | | | | | |
|------------------------|-----------|---|---|---|---|---|---|---|-----------|---|---|---|---|---|---|---|------------|---|---|---|---|---|---|---|
| source octets | 77 (0x4d) | | | | | | | | 97 (0x61) | | | | | | | | 110 (0x6e) | | | | | | | |
| Bit pattern | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| Index | 19 | | | | | | | | 22 | | | | | | | | 5 | | | | | | | |
| Base64-encoded | T | | | | | | | | W | | | | | | | | F | | | | | | | |
| encoded octets | 84 (0x54) | | | | | | | | 87 (0x57) | | | | | | | | 70 (0x46) | | | | | | | |

| Text content | M | | | | | | | | | | | | | | | | | | | | | | | |
|----------------|-----------|---|---|---|---|---|---|---|----------|---|---|---|---|---|---|---|----------|---|---|---|---|---|---|---|
| ASCII | 77 (0x4d) | | | | | | | | 0 (0x00) | | | | | | | | 0 (0x00) | | | | | | | |
| Bit pattern | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Index | 19 | | | | | | | | 16 | | | | | | | | 0 | | | | | | | |
| Base64-encoded | T | | | | | | | | Q | | | | | | | | = | | | | | | | |

Quoted-Printable Encoding

- Alternative to base64 for making any binary data “printable”
 - Binary data uses all 8 bits
 - ASCII data uses only 7 (first bit is 0)
- Mapping each byte (just a sketch)
 - If first bit is already 0, do nothing
 - If first bit is 1, replace with 3 bytes: “=XY” where XY is the hex value being encoded
- Limit line length to 76 characters
- Finish lines with “=”
- Q: What if data contains the byte “=”?

Example

J'interdis aux marchands de vanter trop leur marchandises. Car ils se font vite pédagogues et t'enseignent comme but ce qui n'est par essence qu'un moyen, et te trompant ainsi sur la route à suivre les voilà bientôt qui te dégradent, car si leur musique est vulgaire ils te fabriquent pour te la vendre une âme vulgaire.

J'interdis aux marchands de vanter trop leur marchandises. Car ils se font vite pédagogues et t'enseignent comme but ce qui n'est par essence qu'un moyen, et te trompant ainsi sur la route à suivre les voilà bientôt qui te dégradent, car si leur musique est vulgaire ils te fabriquent pour te la vendre une âme vulgaire.

Determining MIME Content Type

- The sender (web server) determines MIME (content) type of document being sent
 - Rules map file extensions to MIME types
- If file arrives without MIME info, receiver has to guess (see `file` command)
 - File extension *may* help
 - Contents *may* help: magic number at start
 - JPG: `ff d8...`
 - PDF: `25 50 44 46...` (ie `%PDF`)
 - PNG: `89 50 4e 47 0d 0a 1a 0a...` (ie `.PNG...`)
- Some types handled by browser itself
- Others require plugin or application
- Experimental MIME subtypes: `x-`
 - `application/x-gzip`

Summary

- IP address are unique on network
 - IPv4 vs IPv6
- DNS maps strings to IP addresses
 - Domains nested hierarchically
- URLs identify resources on network
 - Scheme, host, path
- MIME type defines a file's encoding
 - Correspondence
 - Layered encodings are possible too