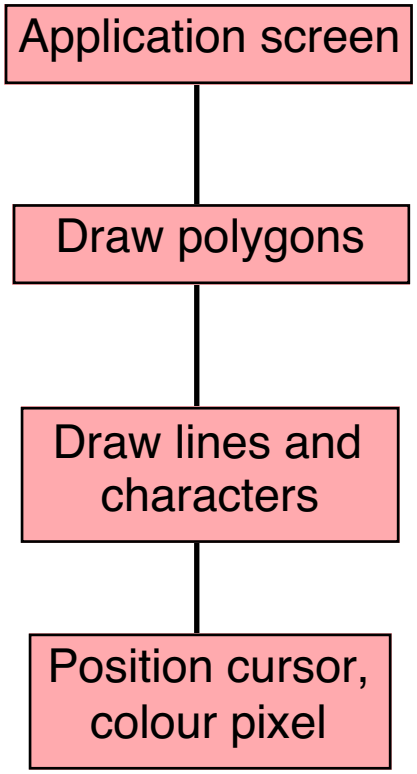
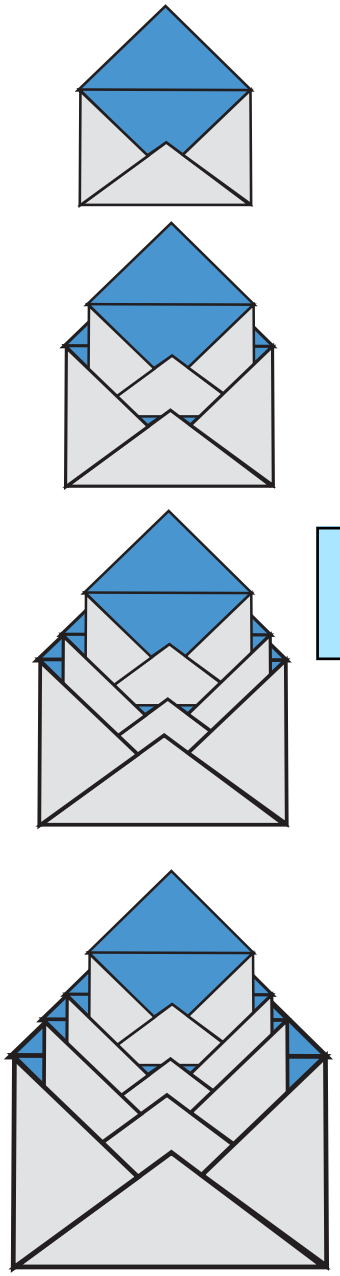
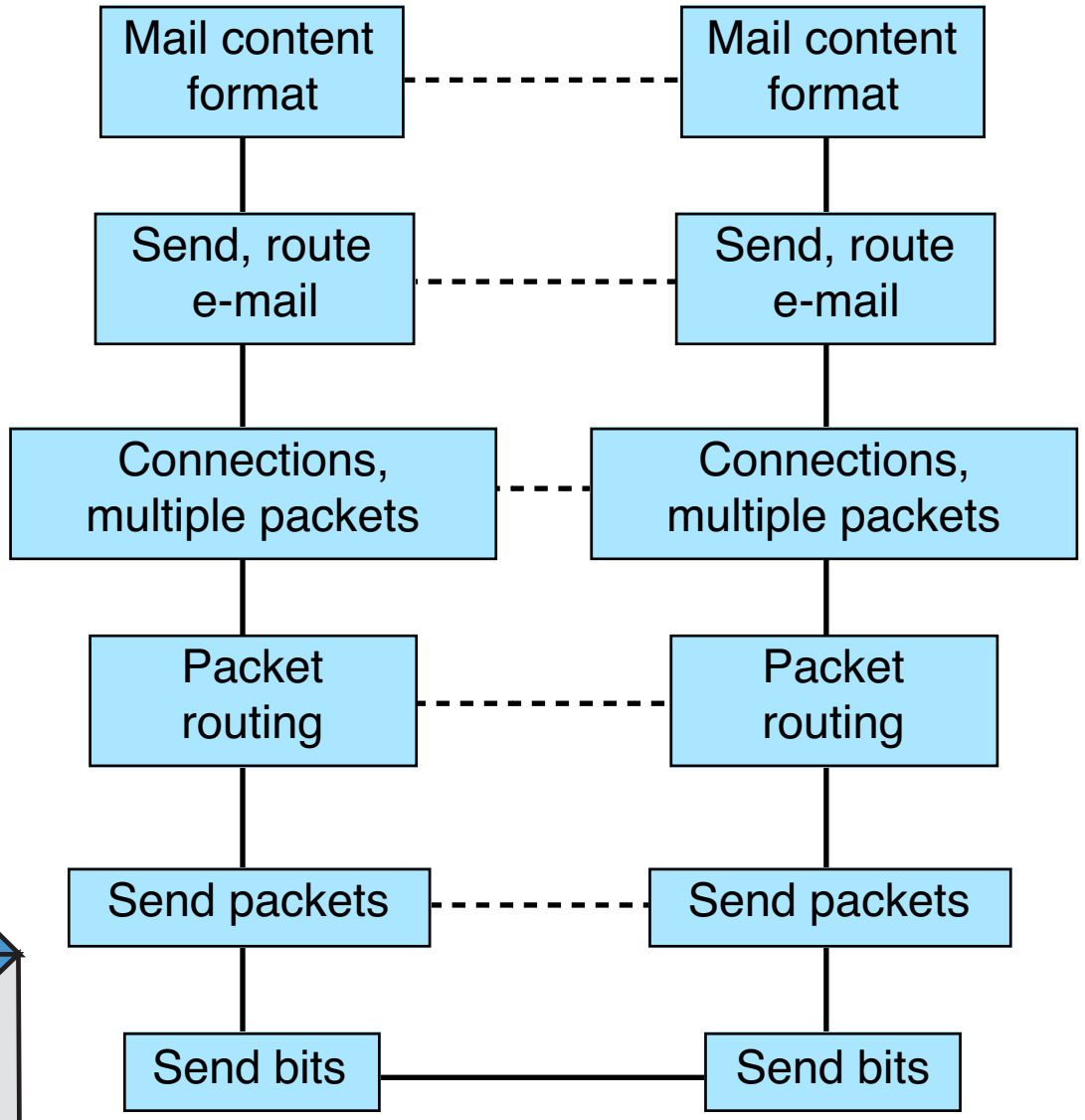
Network layering



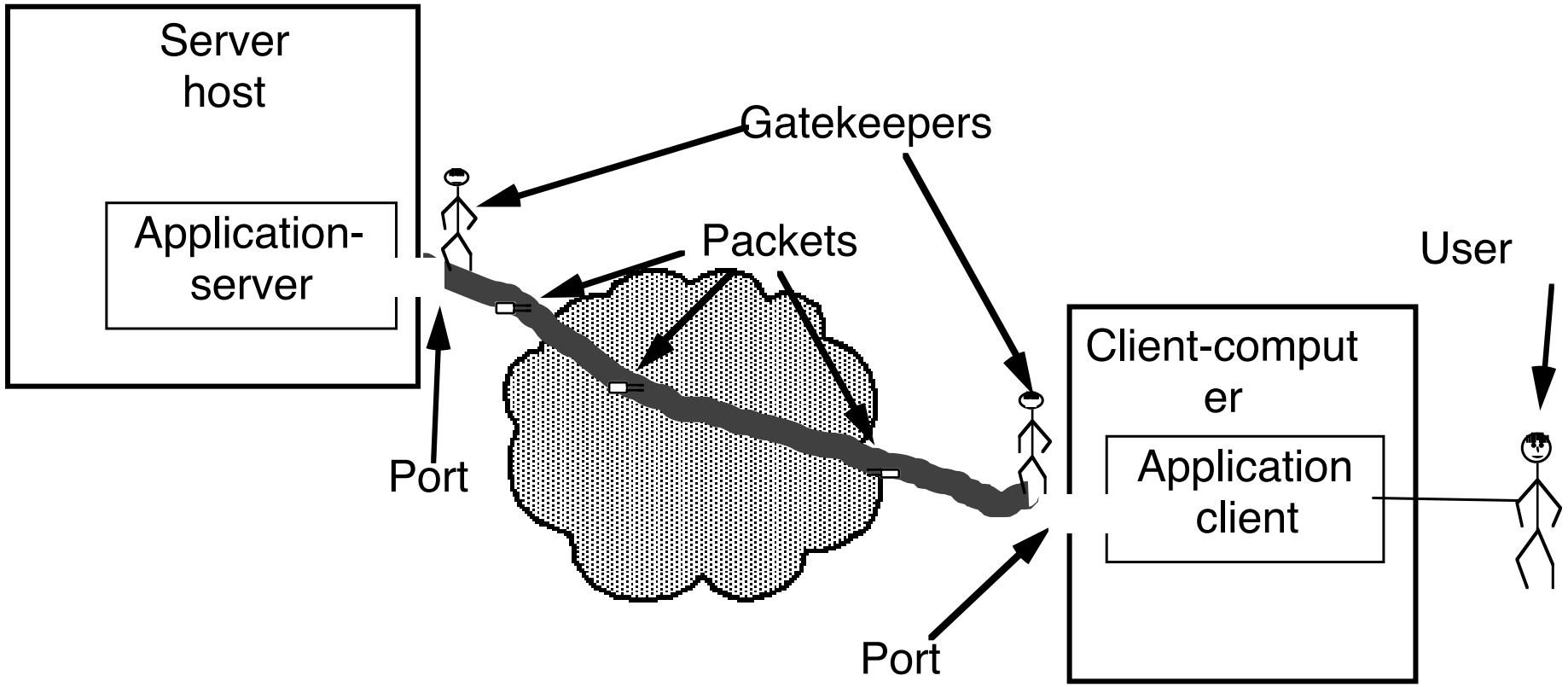




Overview of Internet protocols and services

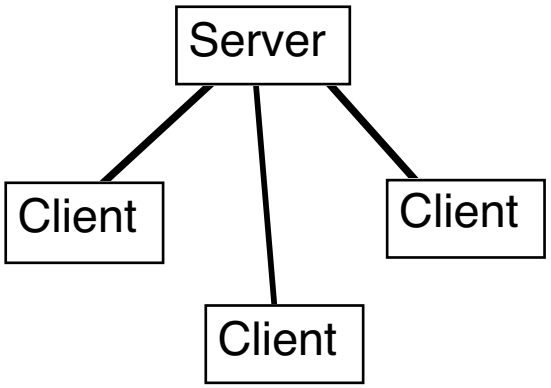
Protocol name	Main usage	Clients	Servers
DNS	Translating domain names to numerical host addresses	All kinds of clients and name servers	Name servers
HTTP (and HTML)	Downloading web pages in the WWW. Can also be used to send in filled in forms and to send in files. Also used for many specialized protocols based on HTTP.	Web browsers	HTTP servers
SMTP (and RFC822 and MIME)	Sending and forwarding of e-mail to and between MTAs (Message Transfer Agents)	Mail clients and SMTP servers	SMTP servers
POP and IMAP	Downloading of e-mail to the mail clients of their recipients	Mail clients	POP or IMAP servers
NNTP	Downloading and forwarding of Usenet News articles.	News clients and news servers	News servers
FTP	Anonymous downloading of files, non-anonymous transfer of files between logged in directories.	FTP clients, Web browsers	FTP servers
Gopher	An old, nowadays not much used protocols, which can be seen as a limited subset of HTTP.	Web browsers, Gopher clients	Gopher servers
PICS	”Protection” of children from material on the net regarded as unsuitable for them.	All kinds of clients	PICS servers
LDAP	Searching in directories.	LDAP clients, often built into e-mail clients.	LDAP servers

Computers, applications, ports, packets

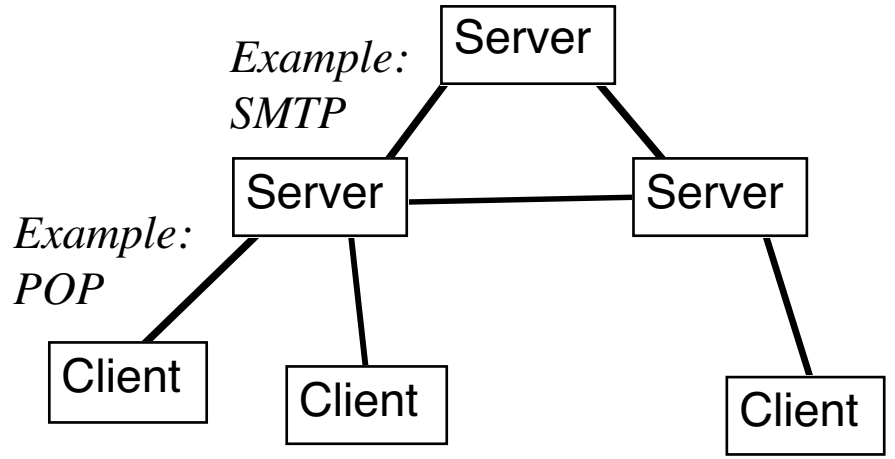


One host can have many different ports fpr different applications.
Exampels of ports: E-mail, file transfer, World Wide Web.
All communication to one particular port uses one particular language.

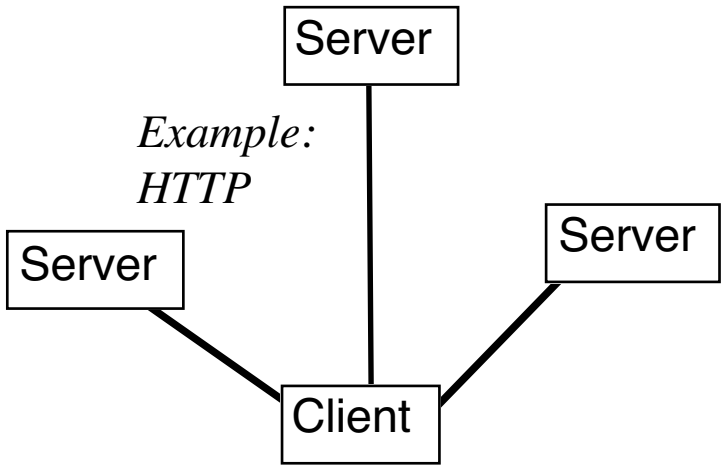
Architectures



Example: LAN data base

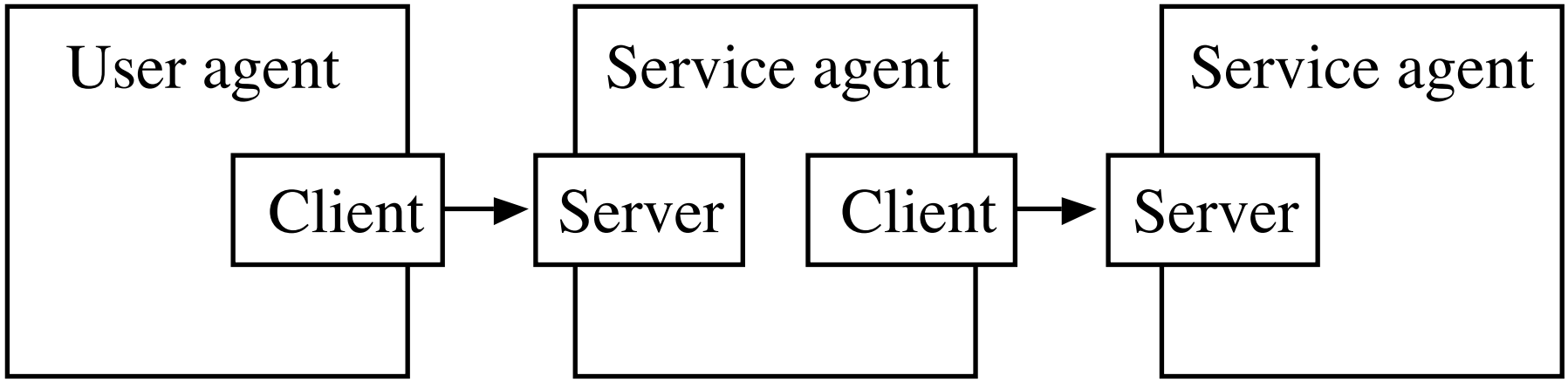


Example: E-mail, Usenet News

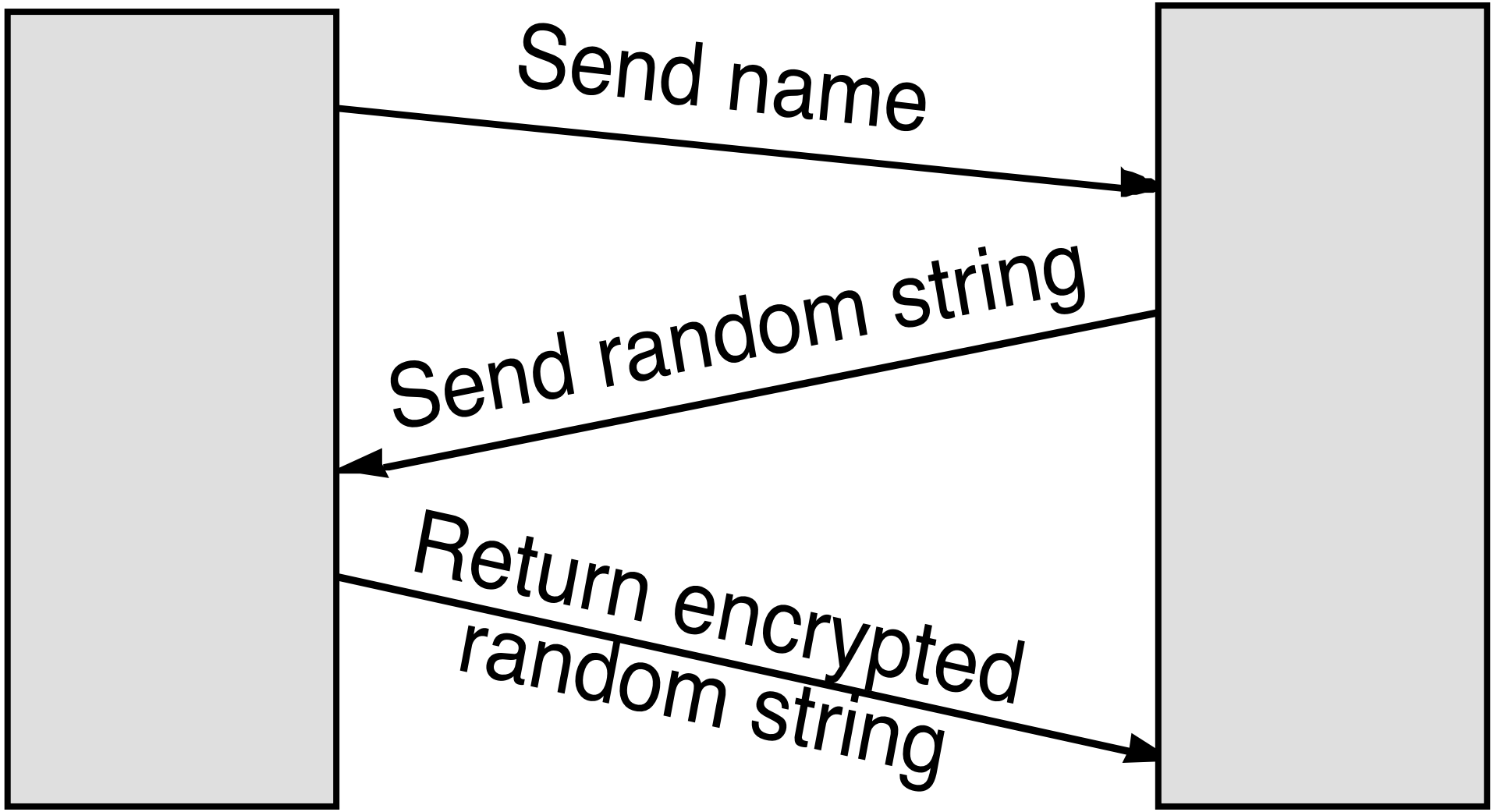


Example: WWW

Store-and-forward transmission

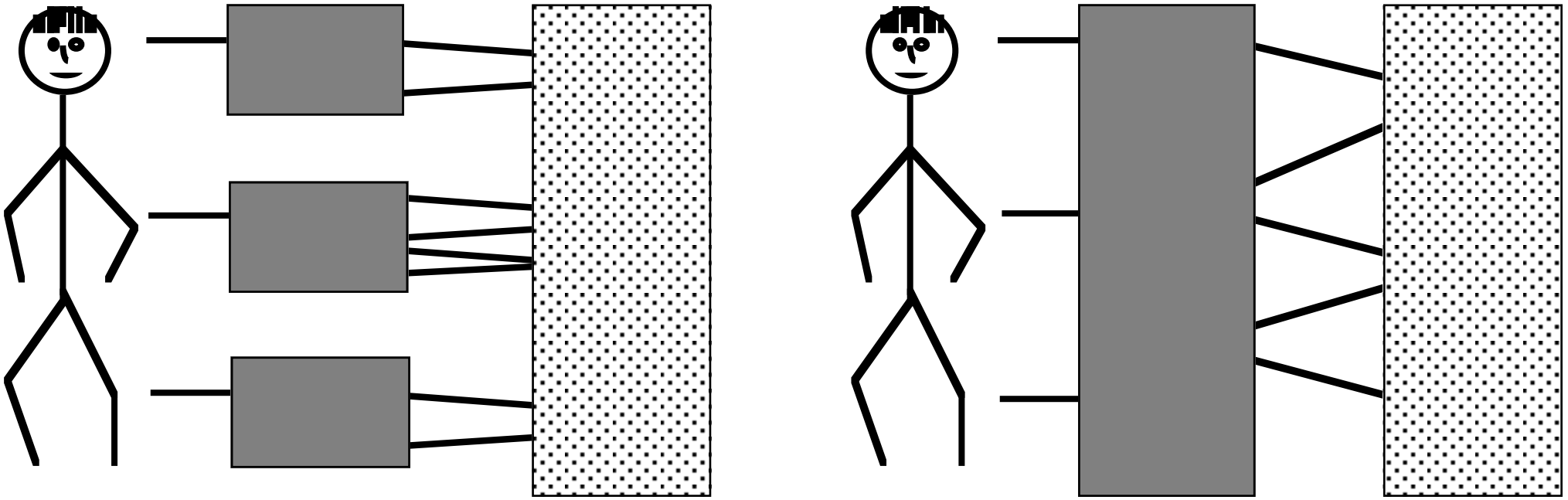


Identification

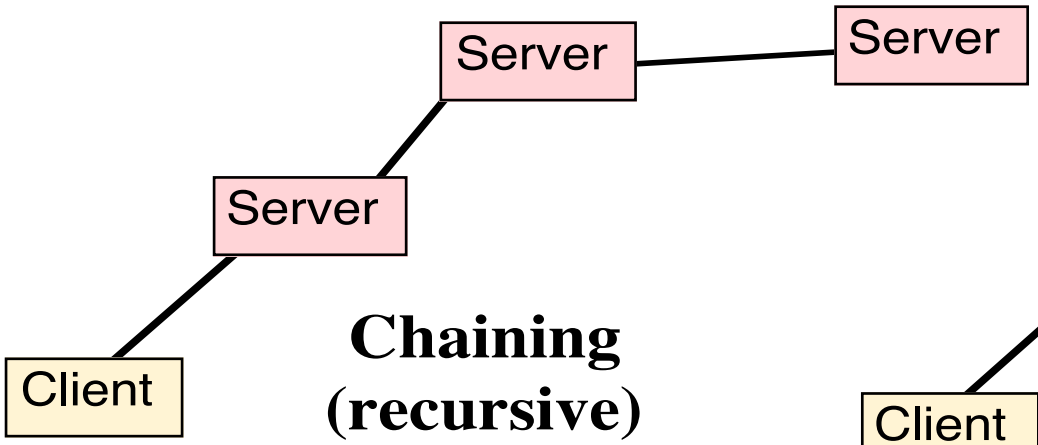


Connection retention

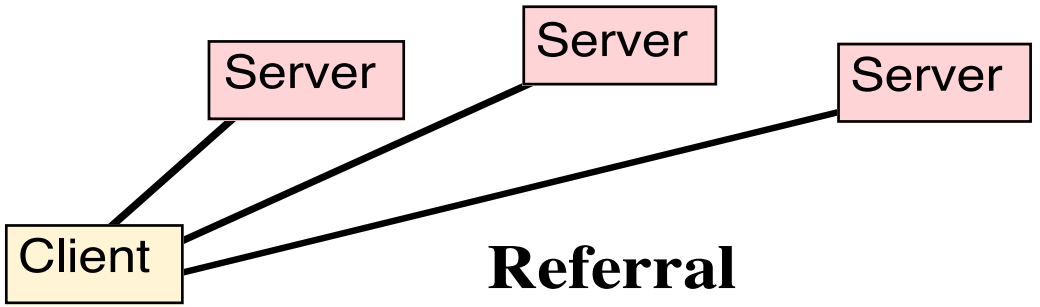
Transaction processing versus connection-oriented protocols



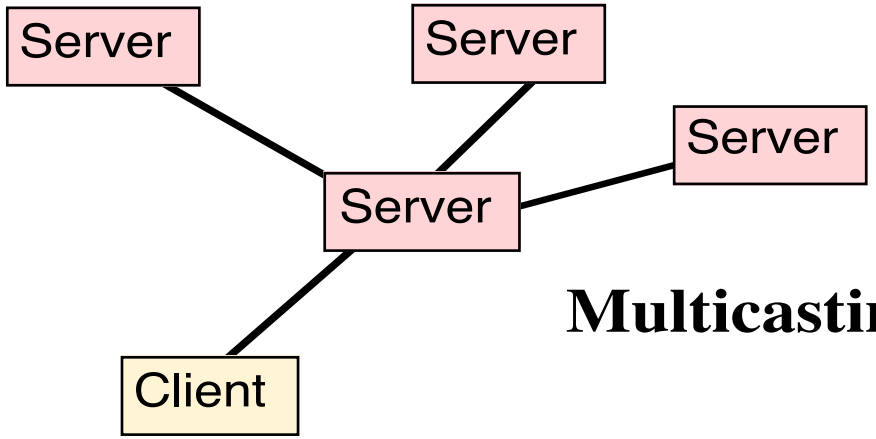
Chaining, referral, multicasting



Chaining (recursive)

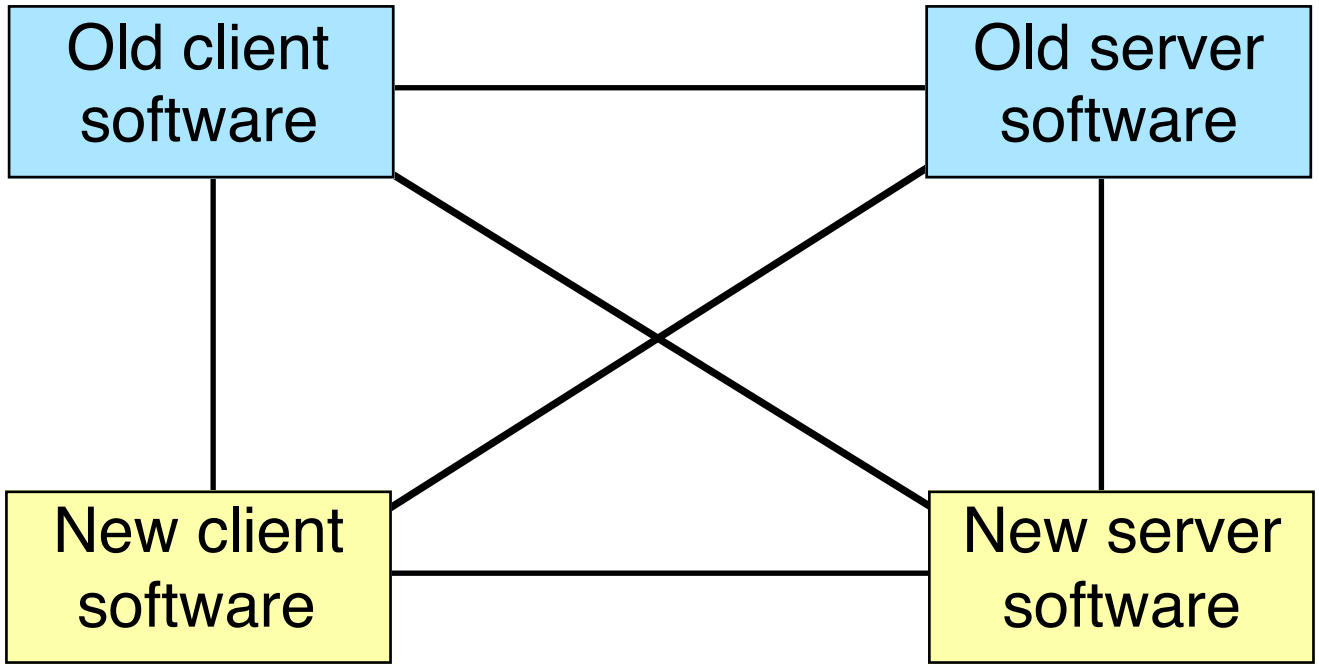


Referral (iterative)



Multicasting

Extension problem



Horror example: Binary files through 7-bit e-mail

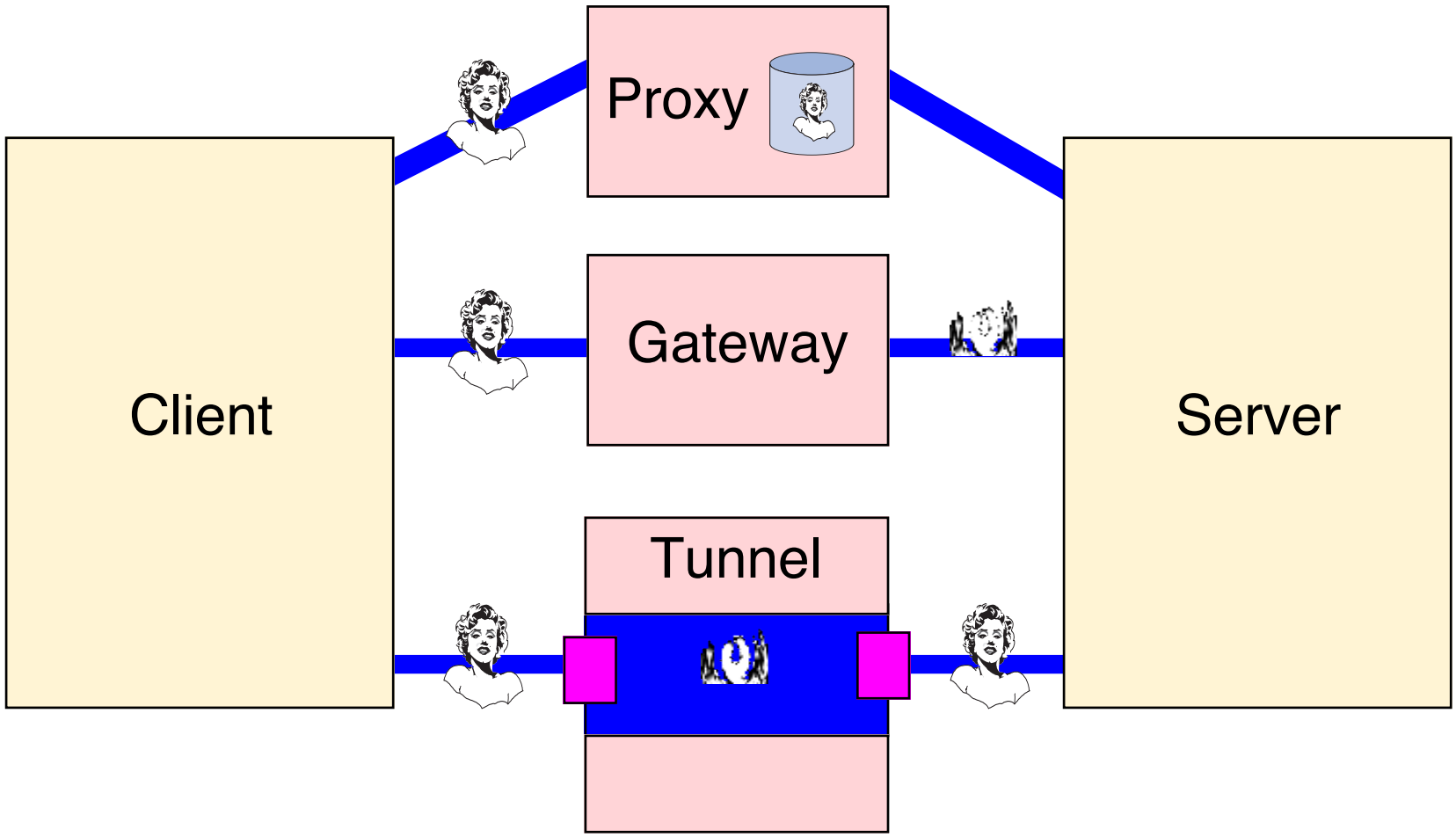
Extension by levels: for example HTML 1.0, HTML 2.0

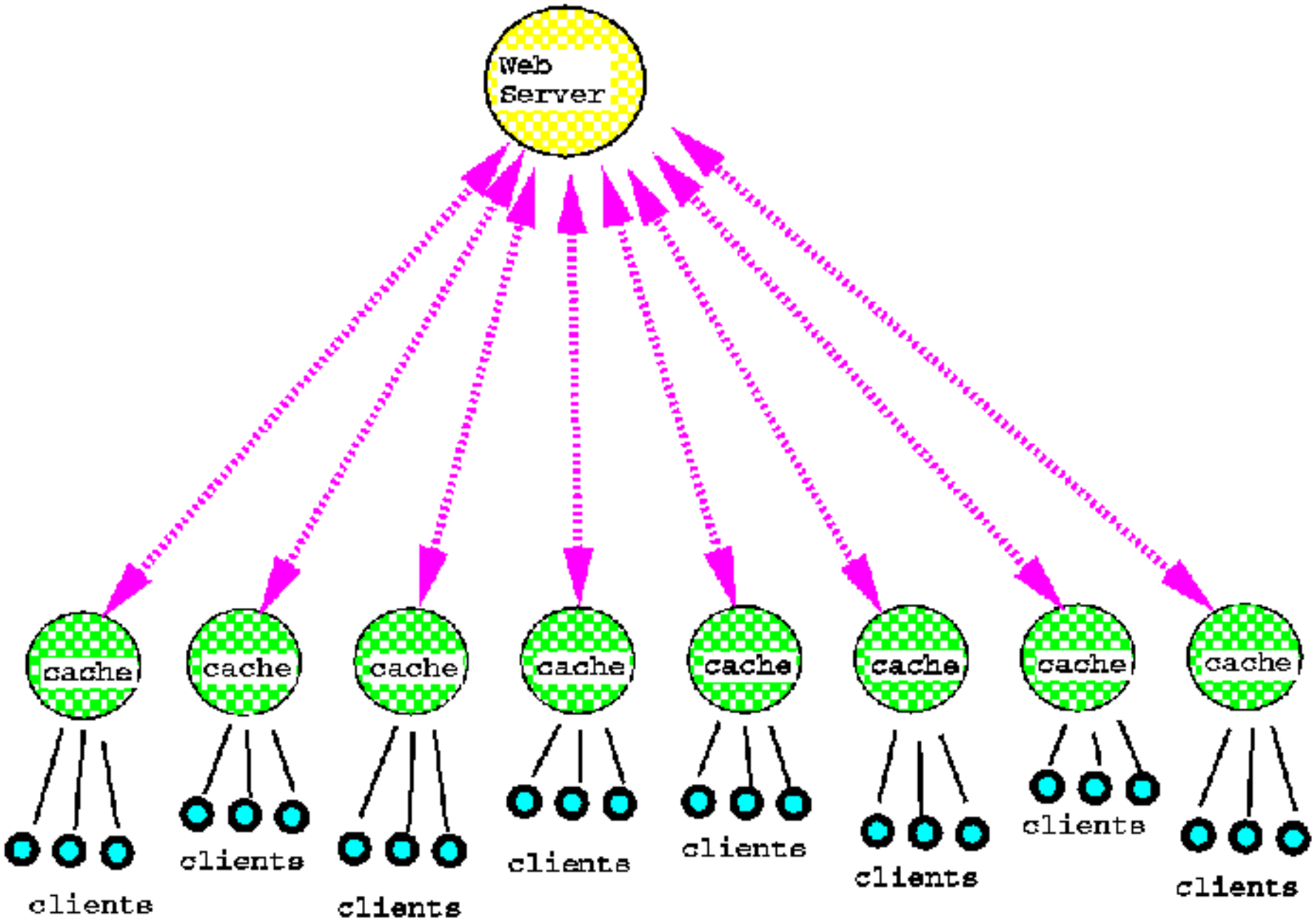
Extension by feature selection

Built-in extension points

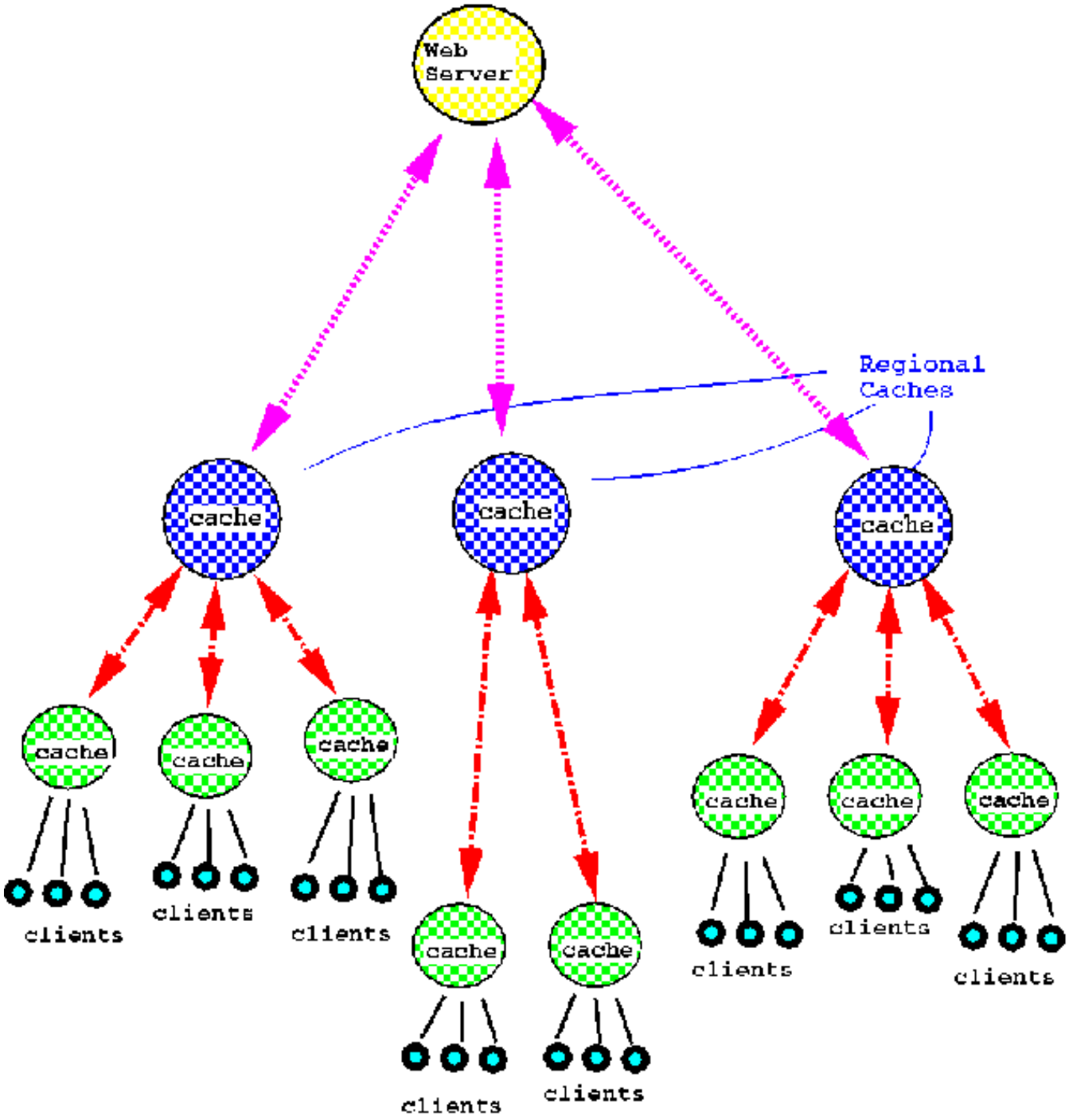
Registration facility vers. X-headers

Intermediaries




Web
Server
cache
cache
cache
cache
cache
cache
cache
cache
clients
clients
clients
clients
clients
clients
clients
clients

Hierarchical caching



The First Golden Rule:

Be liberal in what you accept,
be conservative in what you produce.

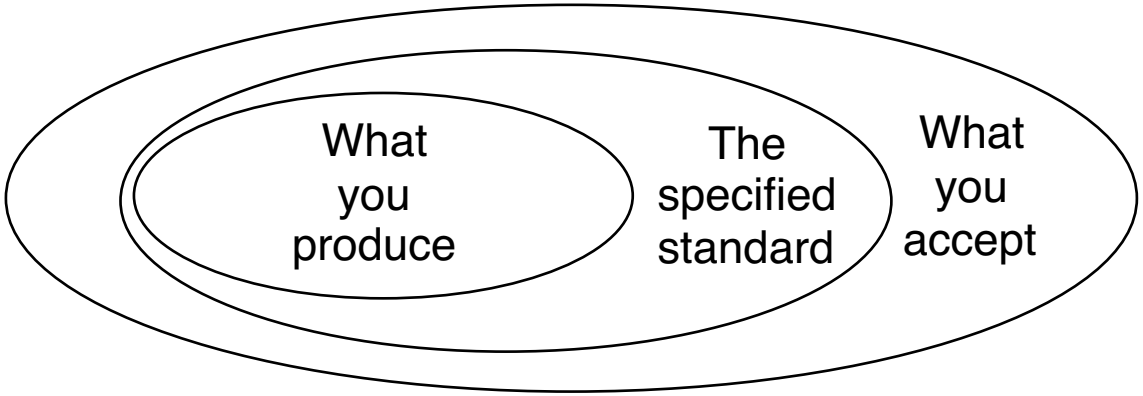
Does this mean a different protocol and syntax for what you produce and what you accept?

How do you know what (in excess to the standard) you should accept, and what (included in the standard you should not produce)?

Example; e-mail: Do not use blanks in e-mail names

Example; e-mail: Accept

John T. Smith <jsmith@foo.bar.net>



Golden Rules

(1) Be liberal in what you accept, be conservative in what you produce

Use a narrow produce syntax and a wide accept syntax

(2) Do no harm

What may be good in your special case, may in other cases cause harm

(3) Do not munge

Munge = Modify what other network modules has produced

Names in the Internet

Physical net addresses, example: 130.237.161.10

Domain names, example: ester.dsv.su.se, eies2.njit.edu

E-mail-addresses: example:
president@whitehouse.gov

DNS = Domain Naming Service translates domain names to physical net addresses. Can be accesses through the client "nslookup" (RFC 1034, RFC 1035)

People seldom see the physical net addresses, since translation from domain names to physical net addresses is done by the application programs used.

Character sets

A character set is a rule for encoding a certain set of glyphs onto one or more octets. By a glyph is meant a kind of small picture and a kind of syntactic description of the character. The same glyph need not look exactly identical, different fonts can display the same glyph in somewhat different ways.

Examples of characters and their encoding

Syntactic description	Encoding in some common character sets (hexadecimal representation)				Glyphs
	ISO 646	ISO646-SE	ISO 8859-1	Unicode & ISO 10646	
latin capital letter A with diaeresis	n.a.	5B	C4	00C4	Ä Ä Ä
latin capital letter O with diaeresis	n.a.	5C	DC	00DC	Ö Ö Ö
latin capital letter O with stroke	n.a.	n.a.	D8	00D8	Ø Ø Ø
Reverse Solidus	5C	n.a	5C	005C	\ \ \

How can you put more than 255 different characters into eight bit octets?

Method 1	ISO 6937	Use multiple characters for some encodings, for example é as e´ or o as o¨.
Method 2	ISO 2022	Use several different 255 character sets, and special shift sequences to shift from one set to another set.
Method 3	Unicode, ISO 10646	Use two or four octets for each character, but provide compression techniques to compress them during transmission. UTF-8 is an example of a compression encoding scheme for ISO 10646, which has the property that the most common characters, like a-z and A-Z, have the same one-octet encoding as in ISO 646 and ISO 8859-1.
Method 4	HTML, MIME Quoted-Printable	Use special encodings for special characters, like ` ` for non-breaking space or `&ouml;` for ö.

Binary and textual data

Binary data

Examples: Data compressed with various compression algorithms, images in formats like GIF, JPEG or TIFF, application data in a format particular to a certain application, such as Word, Excel, Filemaker Pro, Adobe Acrobat, etc.

Textual data

Data which is textual in character, in that it consists of a sequence of “readable” characters, sometimes organized into lines, such as plain text, HTML source, Postscript documents, source code in a programming language, etc.

There is no sharp limit between binary and textual data. Some properties which sometimes distinguishes textual data are:

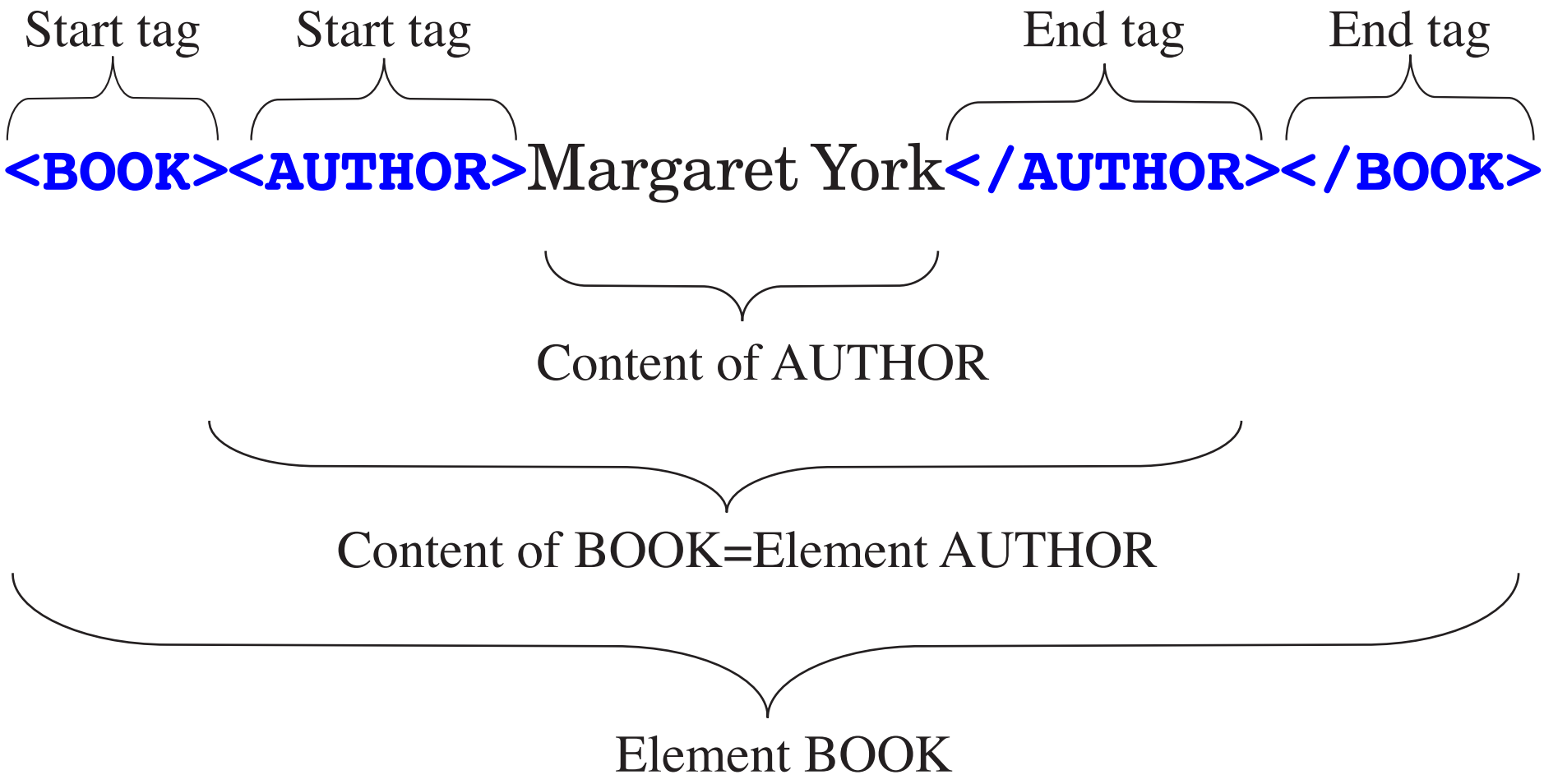
- The character sequence to delimit line breaks differs between platforms, and is often modified at transmission from one platform to another. Macintosh usually uses a single Carriage Return (CR), Unix usually uses a single Line Feed (LF), MS Windows usually uses the character sequence CRLF in file storage, but this is often transformed to only LF when data is important into RAM by an application program.
- Sometimes, characters are encoded according to a character set, which is a rule deciding which glyph to show for a certain bit combination. Sometimes, the character set is modified when textual data is moved between computers or between applications.

Example of textual encoding:	Example of BER encoding:	Example of XML encoding:

Family
Person
  Name: John Smith
  Birthyear: 1958
  Gender: Male
  Status: Married
Person
  Name: Eliza Tennyson
  Birthyear: 1959
  Gender: Female
  Status: Married
End of Family
	(Each box represents one octet. Two-character codes are hexadecimal numbers, one character codes are characters)
	30 34
	30 16
	1A 0A J o h n ␣ S m i t h
	02 02 07 A6
	0A 01 00
	0A 01 01
	30 1A
	1A 0E E l i z a ␣ T e n n y s o n
	02 02 07 A7
	0A 01 01
	0A 01 01
		<?xml version="1.0" ?>
		<!DOCTYPE family SYSTEM "family.dtd">
		<family>
		 <person gender="male"
		  status="married">
		  <name>John Smith</name>
		  <birthyear>1958
		  </birthyear>
		 </person>
		 <person gender="female"
		  status="married">
		  <name>Eliza
		  Tennyson</name>
		  <birthyear>1959
		  </birthyear>
		 </person>
		</family>

169 octets (excluding newlines)	54 octets	276 octets (excluding newlines and leading spaces)
18 % efficiency	57 % efficiency	11 % efficiency[1]

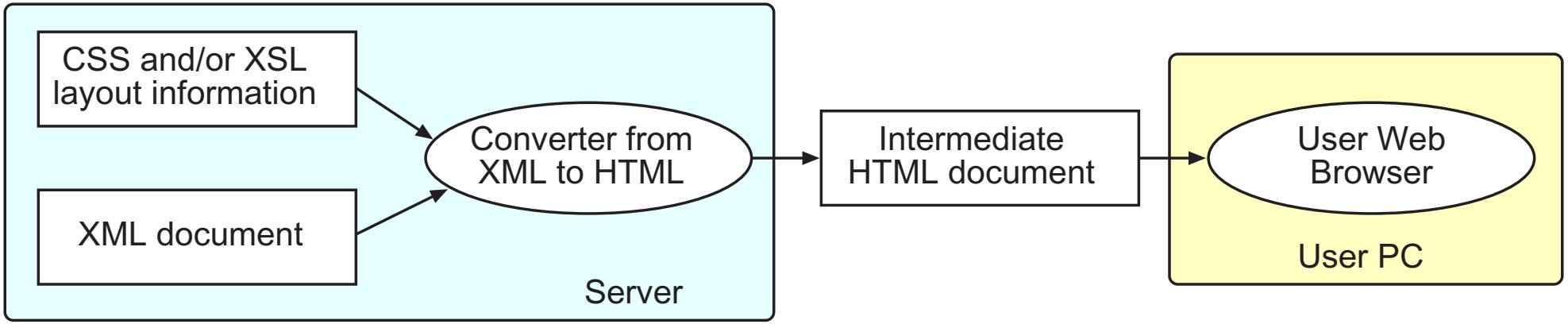
ELEMENT and TAG



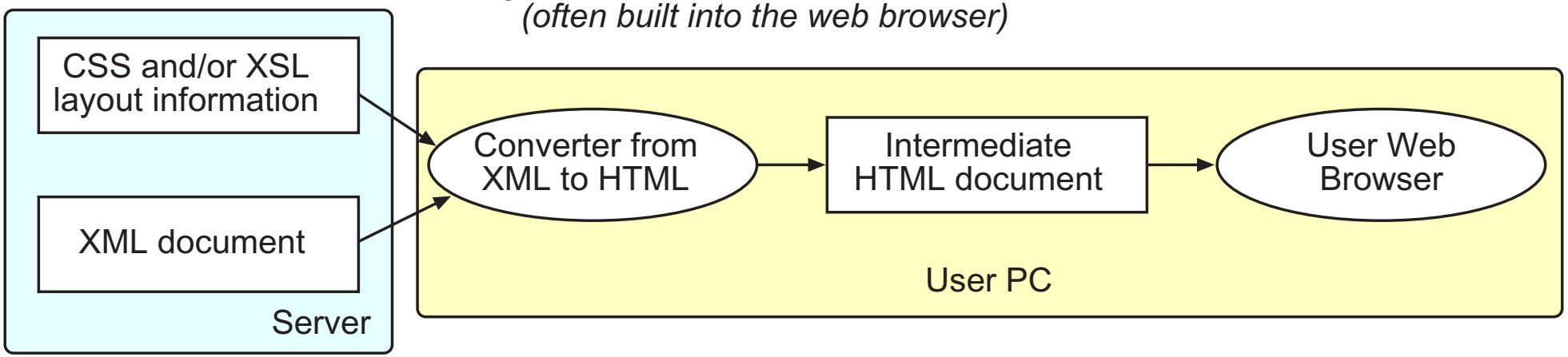
Putting formatting information into XML pages

CSS = Cascading Style Sheets and XSLT = Extensible Style Language Transformations

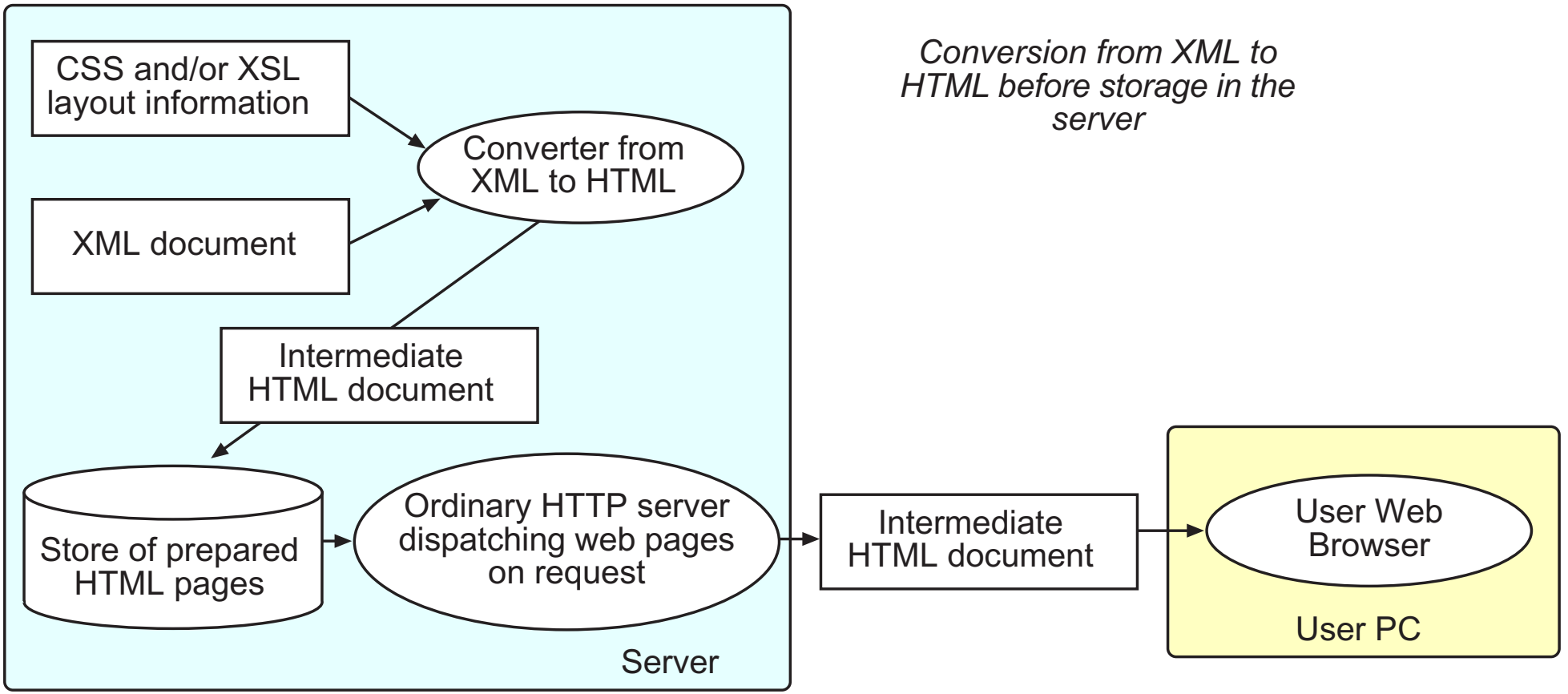
Conversion from XML to HTML in the server, before transmission to the PC



Sending XML to the PC and conversion in the PC (often built into the web browser)

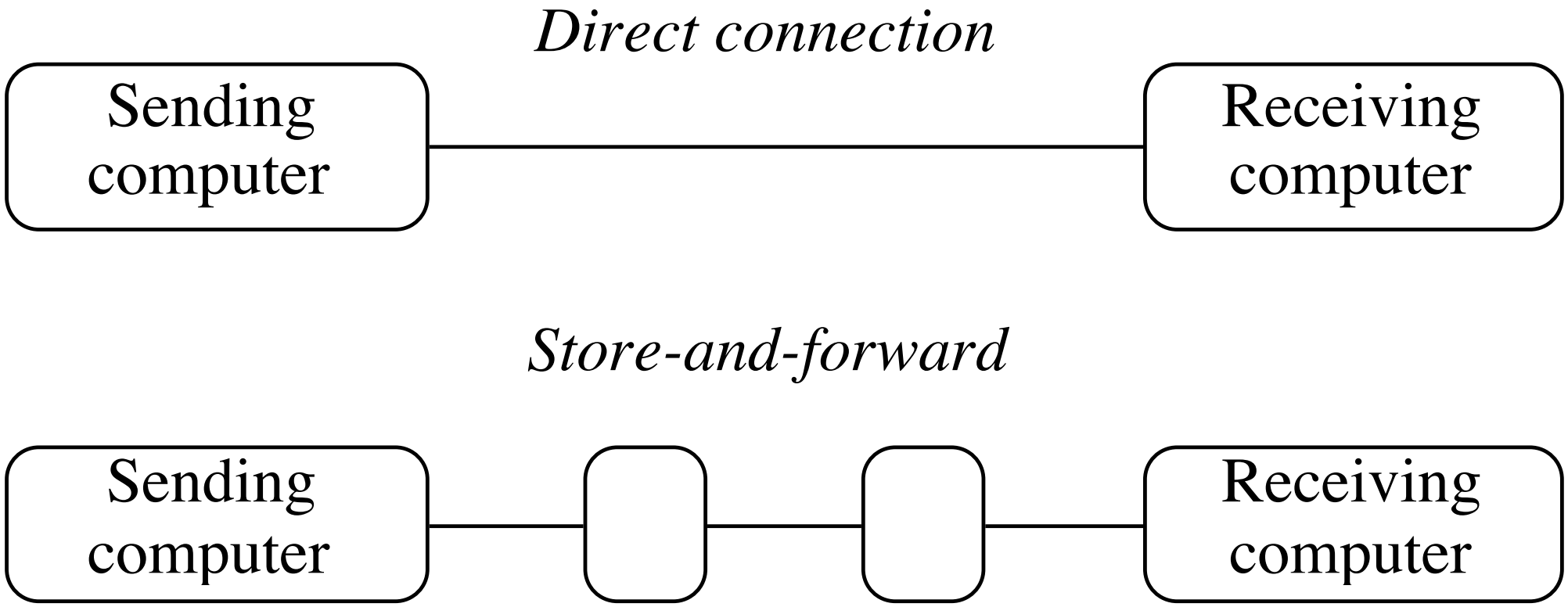


Putting formatting information into XML pages

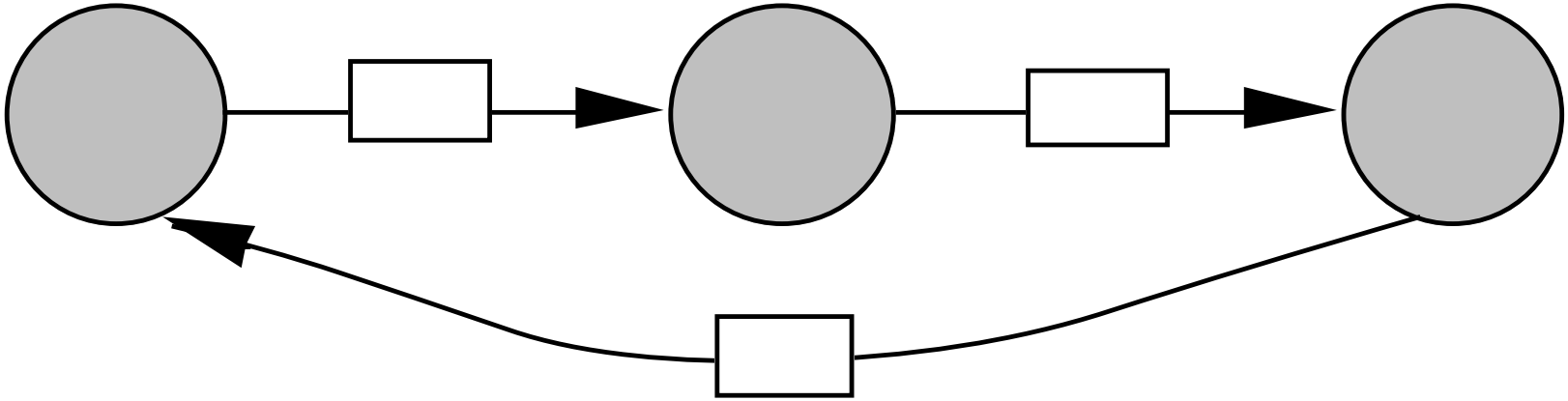


Conversion from XML to HTML before storage in the server

Direct connection and store-and-forward



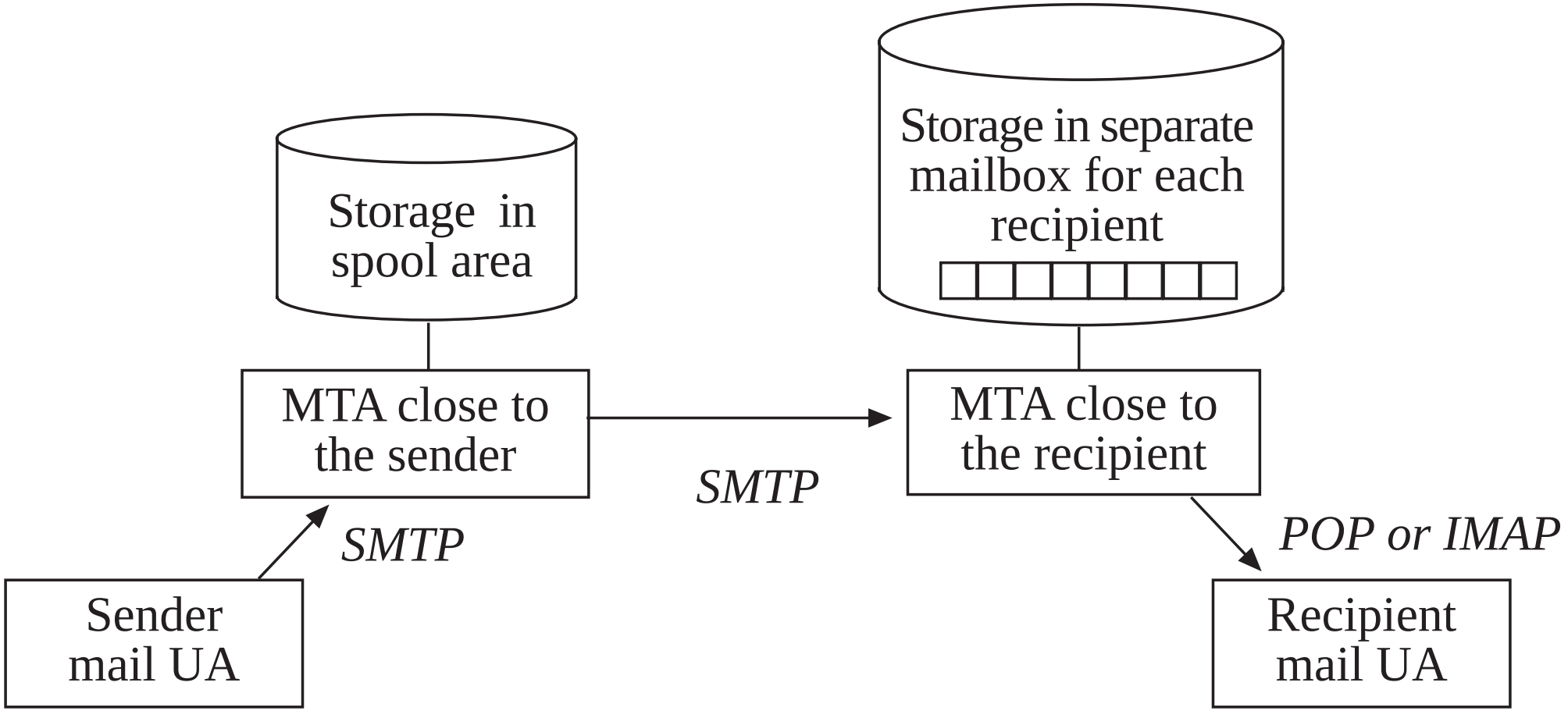
Loop control for Nested Distribution Lists



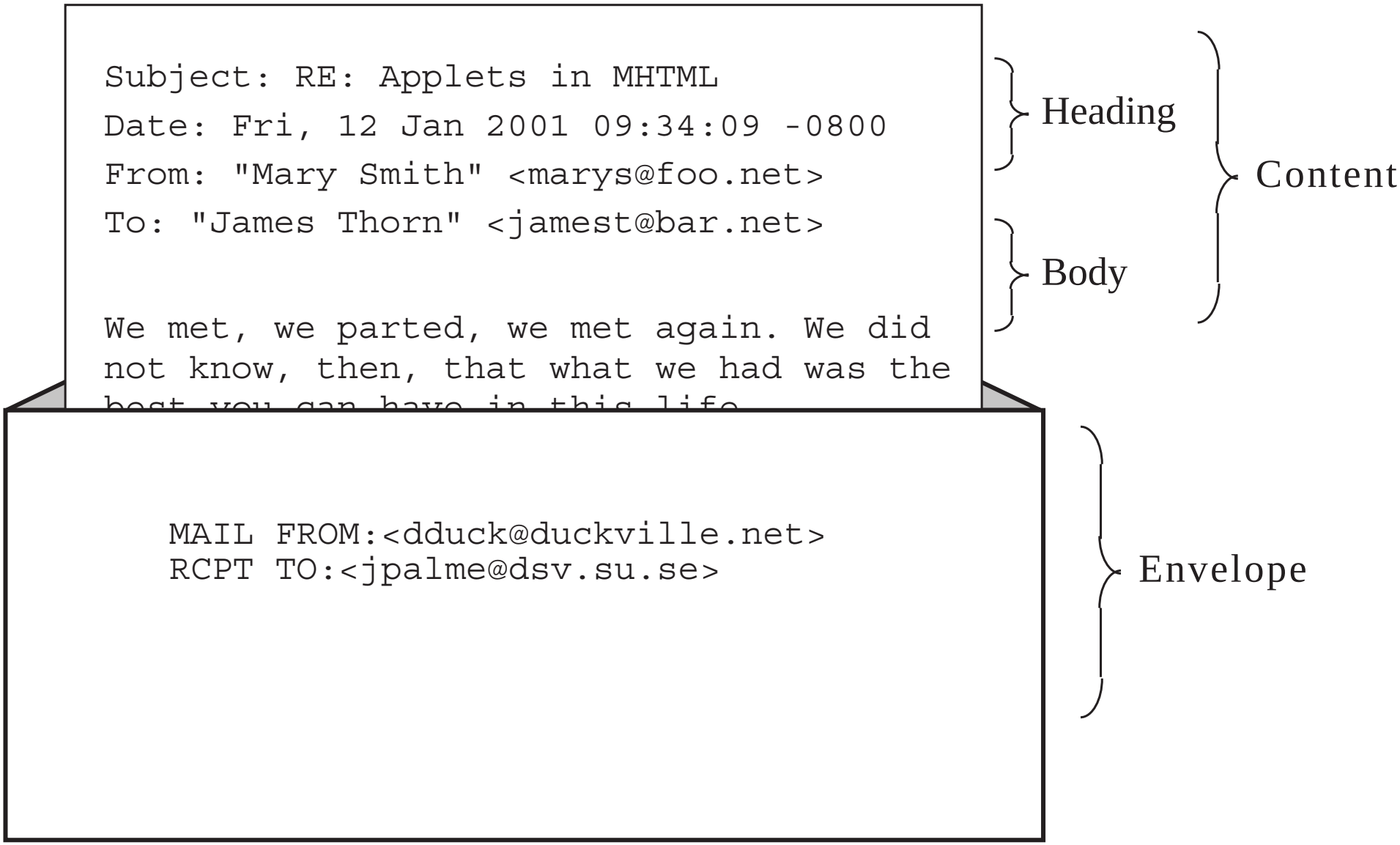
(1) Full expansion by the originating UA or MTA.
(2a) Trace list on the envelope, use to stop incoming messages.
(2b) Trace list on the envelope, use to stop outgoing messages.
(3) Registration system.
(4a) Storing Message-ID-s with DL expanders.
(4b) Storing content checksums with DL expanders.

X.400: Primarily 2a, Listserv: 4a and 4b, Usenet News: 4a

Storage of Mail in MTAs



Envelope, Content, Heading and Body



An example of a complete HTTP 1.0 connection

C: <connects to www.dsv.su.se at port 80>

C:
```
GET /~jpalme/test/small.html HTTP/1.0
```

S:
```
HTTP/1.1 200 OK
Date: Mon, 13 Apr 1998 11:13:46 GMT
Server: Apache/1.2.4
Last-Modified: Mon, 13 Apr 1998 11:11:31 GMT
ETag: "437e5-98-3531f2e3"
Content-Length: 152
Accept-Ranges: bytes
Connection: close
Content-Type: text/html

<HTML>
<HEAD>
   <TITLE>A very small file</TITLE>
</HEAD>
<BODY BGCOLOR="#FFFFCC">
   <H1>A very small file</H1>
   <P>Which ends here!
</BODY>
</HTML>
```

S: <closes the connection to the client>

An example of an actual HTTP request

```
GET /Icons/headhome.gif HTTP/1.1
Host: w4g.dsv.su.se:9800
Accept: image/gif, image/x-xbitmap, image/jpeg,
         image/pjpeg,image/xbm, image/x-jg, */*
Accept-Language: en
Connection: Keep-Alive
Referer: http://w4g.dsv.su.se:9800/
User-Agent: Mozilla/4.0 (compatible; MSIE 4.0; Mac_PowerPC)
UA-OS: MacOS
UA-CPU: PPC
Cookie: currentLocation=2.1; cursor=%401dce3924; userName=
Extension: Security/Remote-Passphrase
```

Content-Disposition

The Content-Disposition header, defined in RFC 1806, is not a part of the HTTP standard, but is widely implemented. It has certain security problems.

According to RFC 1806:

Content-Disposition = [Inline | Attachment] *(; disposition-parameter)

disposition-parameter = filename "=" value | other-parameter

Optimization of delivery in browsers

Progressive rendering

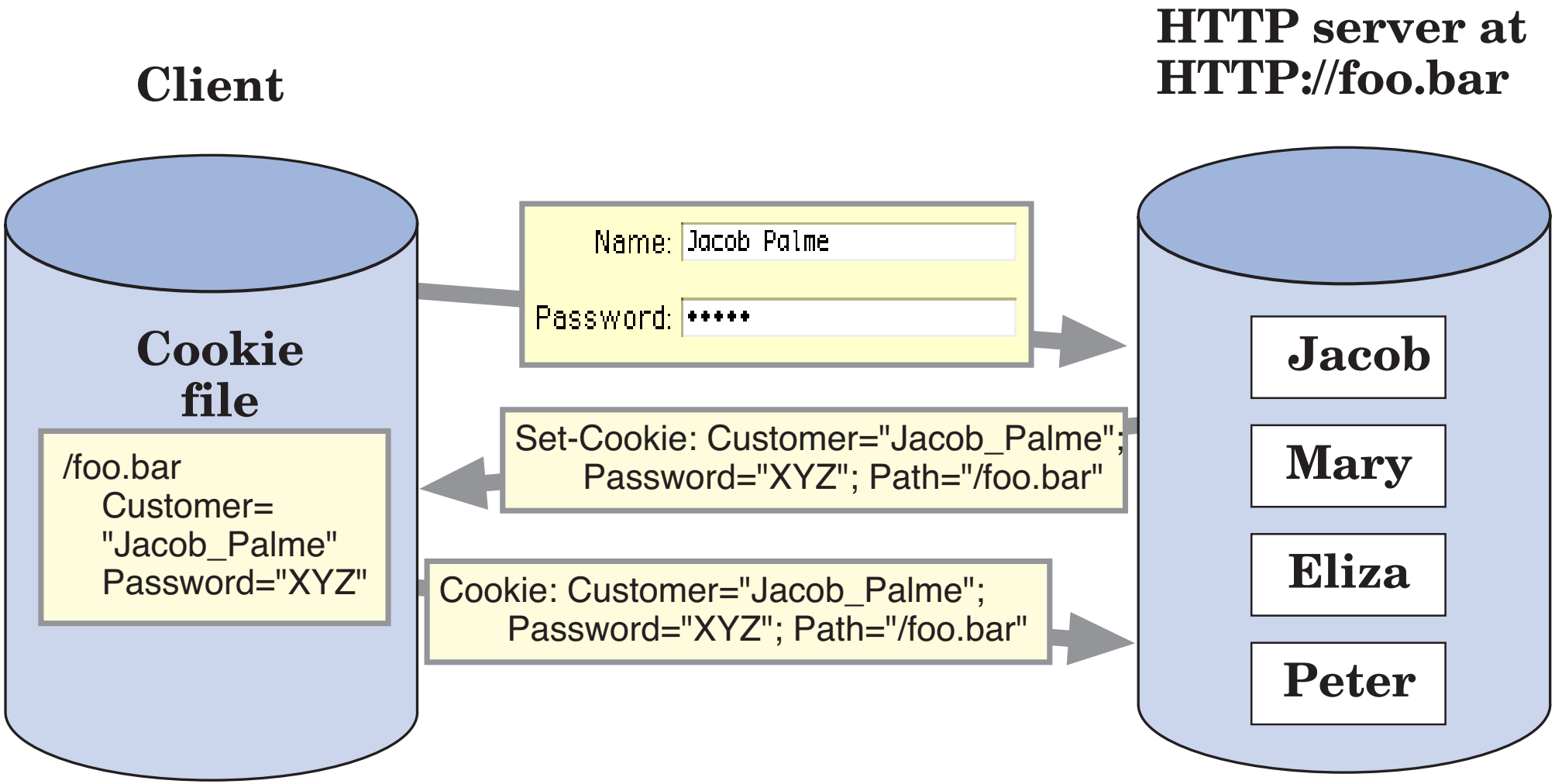
Part of the text or the image is shown while the rest is being downloaded.

Multiple connections

If a user requests a document with several inline image, many browsers will establish multiple connections to download more than one image at the same time. Browsers may also give priority to downloading those images which are in the visible window.

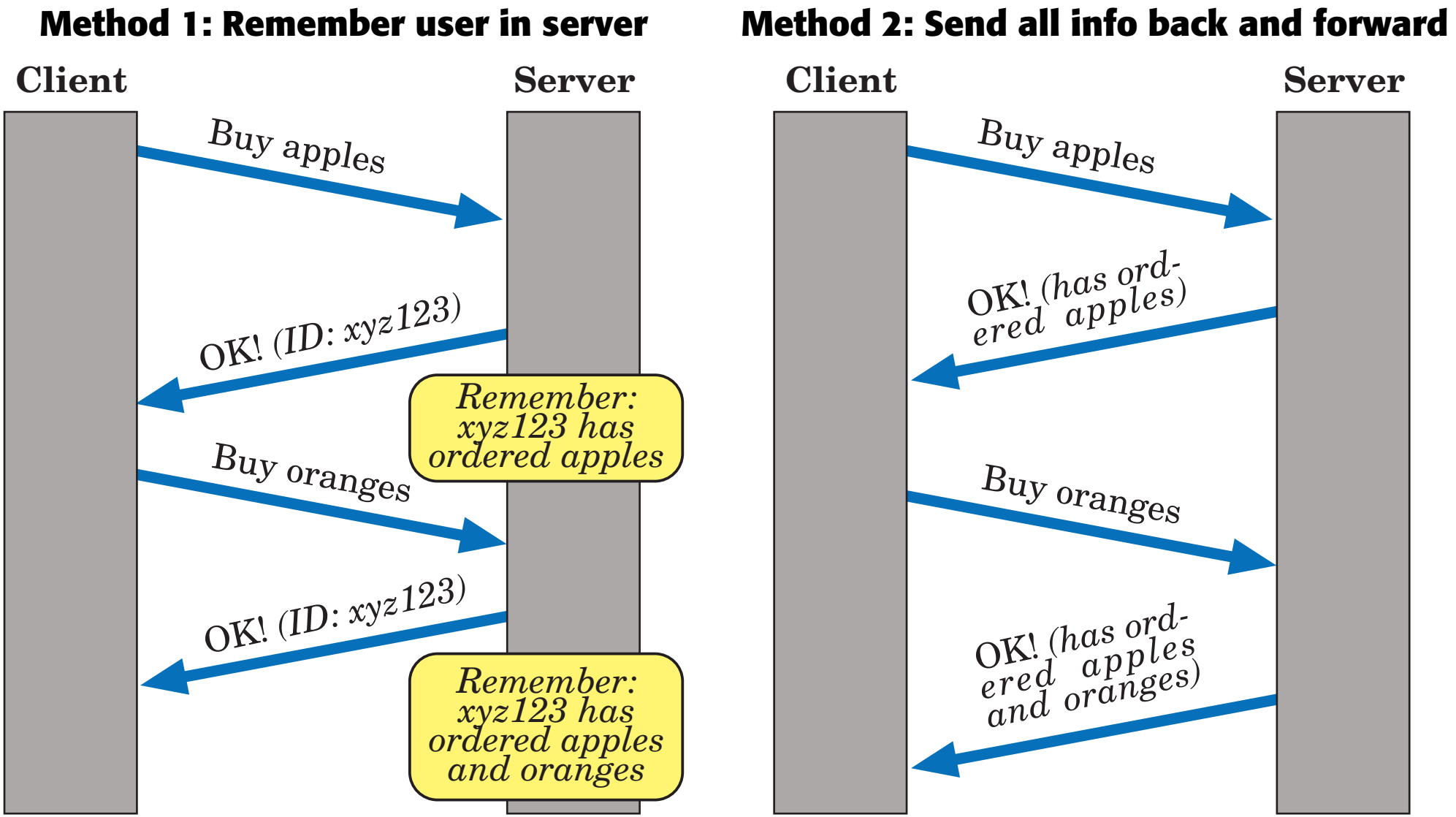
The use of one TCP connection per file can be very inefficient.

What is a cookie



A cookie is a small piece of data, which the server can store in the client, and which the client sends the next time it connects to the same server to identify itself.

Two ways of remembering what a user did earlier



Privacy issues with cookies

Server can keep track of how often a user accesses the server and what the user does. Can for example be used to select banner advertisment according to the interest of the user. Or can be used to guess at the political opinion of the user, and then used to target political advertising or harassment.

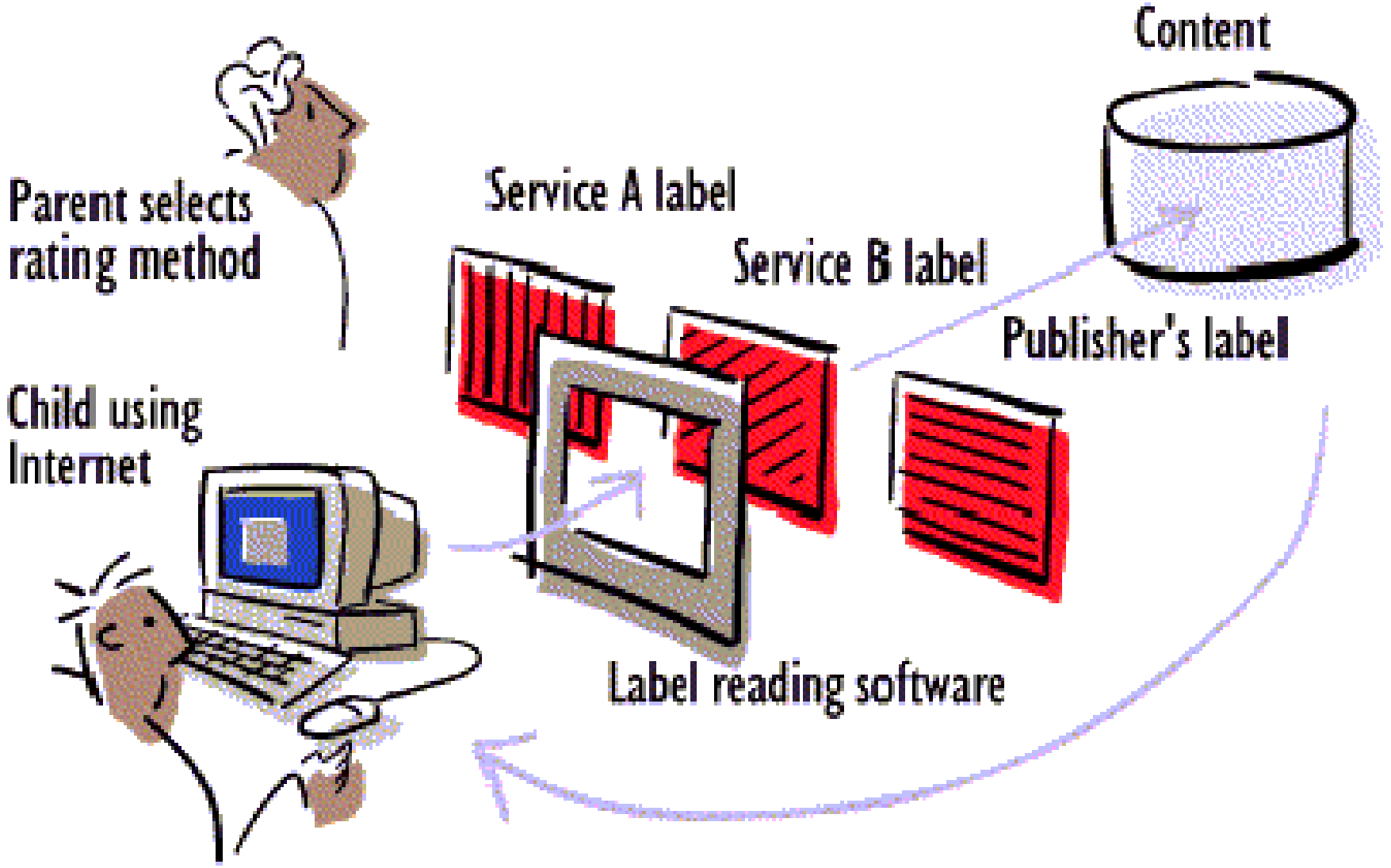
These profiles might be exchanged between servers or sold for profit.

Possible security holes might let the server fetch and modify information it should not have?

Users can set their browsers to reject cookies, or to ask the user before accepting them. But if you reject cookies, you lose a lot of functionality.

There are programs available to help users control and manage their cookies.

PICS 2


Content
Parent selects rating method
Service A label
Service B label
Publisher's label
Child using Internet
Label reading software

PICS 3

PICS Technical Mission, excerpt

The technical working group will devise a values-neutral infrastructure for Internet content labeling. The three primary goals are to:

1. enable content providers to voluntarily label the content they create and distribute.
2. enable third-party labeling services to associate additional labels with content created and distributed by others. Services may devise their own labeling systems, and the same content may receive conflicting labels from different services.
3. enable parents and teachers to use the labels to control the information that children under their supervision receive.

PICS 4

A rating *service* is an individual, group, organization or company that produces labels for information. A *rating system* is a way of rating information, consisting of one or more *categories* and a *scale* for each category.

The **Motion Picture Association of America** (MPAA) is a rating service, which uses a well-known (in the United States) rating system for rating movies. Other organizations also provide rating systems or services, such as **SafeSurf** and **SurfWatch** (both PICS founding members) and the **Recreational Software Advisory Council** (a PICS supporting member).

A rating system provides a number of *categories* (or *dimensions*) along which information can be rated.

The MPAA rating system has only one category, the overall rating of the movie. The RSAC rating system has three categories: *nudity/sex*, *violence*, and *language*.

A rating system provides a number of *categories* (or *dimensions*) along which information can be rated.

The MPAA rating system has only one category, the overall rating of the movie. The RSAC rating system has three categories: *nudity/sex*, *violence*, and *language*.

A rating system provides a *scale* for each category.

The MPAA rating system's one category has a scale with values like "G," "PG," and so forth. RSAC's *nudity/sex* category uses a scale with values of "suitable for all ages," "partial nudity," and so on. RSAC's *language* category uses a scale with values of "some profanity," "explicit sexual references" and so on.

PICS Content Labels are values neutral

"The technical working group will define a format for labels, indicating required and optional fields. The format will not specify which words or categories will be used for labeling or the criteria for assigning labels to items."

(excerpt from the Technical Committee Charter)