



PADUA UNIVERSITY

ENGINEERING COURSE

Master of Computer Engineering

COMPUTER NETWORKS



Raffaele Di Nardo Di Maio

Contents

1	OSI model	1
2	Application Layer	3
3	C programming	5
3.1	Organization of data	5
3.2	Struct organization of memory	6
3.3	Structure of C program	7
4	Network services in C	9
4.1	Application layer	9
4.2	socket()	9
4.3	TCP connection	10
4.3.1	Client	10
4.3.1.1	connect()	10
4.3.1.2	write()	12
4.3.1.3	read()	13
4.3.2	Server	14
4.3.2.1	bind()	14
4.3.2.2	listen()	14
4.3.2.3	accept()	15
4.4	UDP connection	18
5	Gateway	19
5.1	Proxy	20
6	HTTP protocol	23
6.1	Terminology	23
6.2	Basic rules	24
6.3	Messages	25
6.3.1	Different versions of HTTP protocol	25
6.3.2	Headers	25
6.3.3	Request-Line	25
6.3.4	Request-URI	26
6.3.5	Request Header	26
6.3.6	Status line	26
6.4	HTTP 1.0	26
6.4.1	Other headers of HTTP/1.0 and HTTP/1.1	27
6.4.2	Caching	28
6.4.3	Authorization	34
6.4.3.1	base64	35
6.4.3.2	Auth-schemes	36
6.5	HTTP 1.1	36
6.5.1	Caching based on HASH	37
6.5.2	URI	38

6.5.3	HTTP URL	39
6.6	Dynamic pages	39
6.7	Proxy	40
7	Resolution of names	43
7.1	Network Information Center (NIC)	43
7.2	Domain Name System (DNS)	43
7.2.1	Goals	44
7.2.2	Hierarchy structure	44
8	HTML	47
9	CSS	49
10	Javascript	51
11	PHP	53
12	Shell	55
12.1	Commands	55
12.2	UNIX Files	55
13	vim	57
13.1	.vimrc	57
13.2	Shortcuts	57
13.3	Multiple files	60
References		61

Chapter 1

OSI model

Chapter 2

Application Layer

Chapter 3

C programming

The C is the most powerful language and also can be considered as the language nearest to Assembly language. Its power is the speed of execution and the easy interpretation of the memory.

C can be considered very important in Computer Networks because it doesn't hide the use of system calls. Other languages made the same thing, but hiding all the needs and evolution of Computer Network systems.

3.1 Organization of data

Data are stored in the memory in two possible ways, related to the order of bytes that compose it. There are two main ways, called Big Endian and Little Endian.

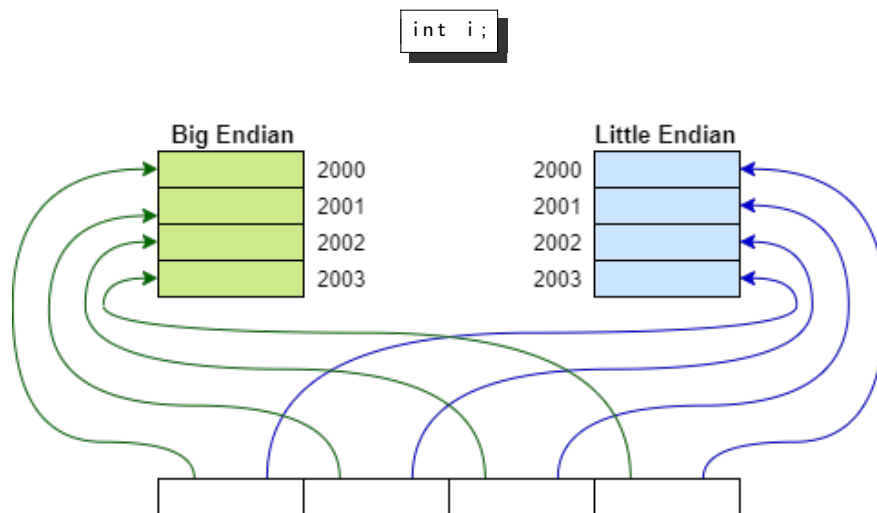


Figure 3.1: Little Endian and Big Endian.

The order of bytes in packets, sent through the network, is Big Endian.

The size of **int**, **float**, **char**, ... types depends on the architecture used. The max size of possible types depends on the architecture (E.g. in 64bits architecture, in one instruction, 8 bytes can be written and read in parallel).

signed	unsigned
int8_t	uint8_t
int16_t	uint16_t
int32_t	uint32_t
int64_t	uint64_t

Table 3.1: <stdint.h>

3.2 Struct organization of memory

The size of a structure depends on the order of fields and the architecture. This is caused by alignment that depends on the number of memory banks, number of bytes read in parallel. For example the size is 4 bytes for 32 bits architecture, composed by 4 banks (Figure 3.2). The Network Packet Representation is made by a stream of 4 Bytes packets as we're using 32 bits architecture.

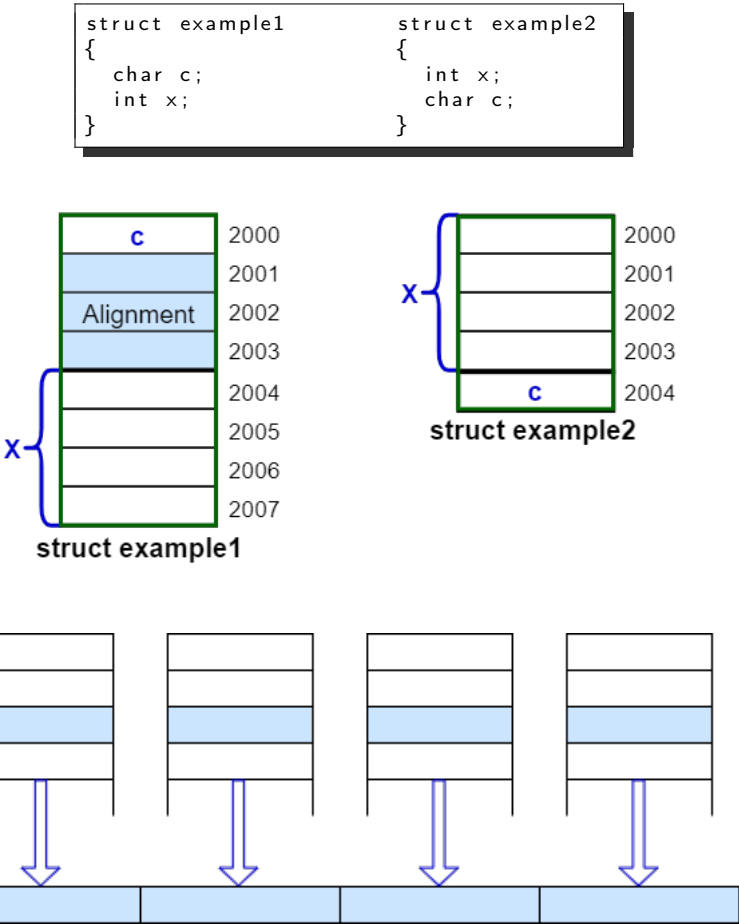


Figure 3.2: Parallel reading in one instruction in 32 bits architecture.

3.3 Structure of C program

The program stores the variable in different section (Figure 3.3):

- **Static area**
where global variables and static library are stored, it's initialized immediately at the creation of the program. Inside this area, a variable doesn't need to be initialized by the programmer because it's done automatically at the creation of the program with all zeroes.
- **Stack**
allocation of variables, return and parameters of functions
- **Heap**
dynamic allocation

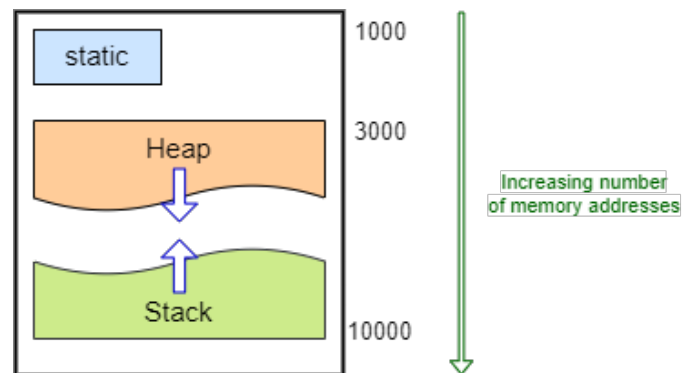


Figure 3.3: Structure of the program.

Chapter 4

Network services in C

4.1 Application layer

We need IP protocol to use Internet. In this protocol, level 5 and 6 are hidden in Application Layer. In this case, Application Layer needs to interact with Transport Layer, that is implemented in OS Kernel (Figure 4.1). Hence Application and Transport can talk each other with System Calls.

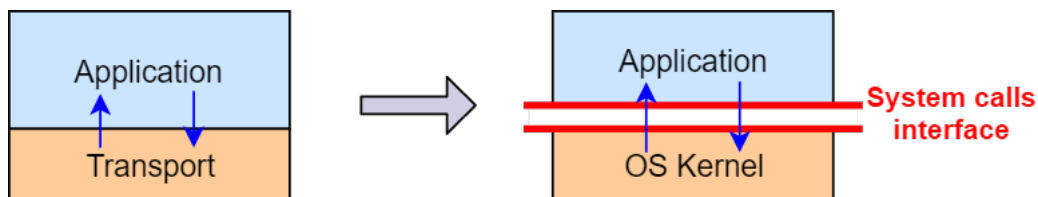


Figure 4.1: System calls interface.

4.2 socket()

Entry-point (system call) that allow us to use the network services. It also allows application layer to access to level 4 of IP protocol.

```
#include <sys/types.h>
#include <sys/socket.h>

int socket(int domain, int type, int protocol);\\
```

RETURN VALUE *File Descriptor (FD) of the socket*
-1 if some error occurs and errno is set appropriately
(You can check value of errno including <errno.h>).

domain = *Communication domain*
 protocol family which will be used for communication.

AF_INET: IPv4 Internet Protocol
AF_INET6: IPv6 Internet Protocol
AF_PACKET: Low level packet interface

type = *Communication semantics*

SOCK_STREAM: Provides sequenced, reliable, two-way, connection-based bytes stream. An OUT-OF-BAND data mechanism may be supported.

SOCK_DGRAM Supports datagrams (connectionless, unreliable messages of a fixed maximum length).

protocol = *Particular protocol to be used within the socket*
 Normally there is only a protocol for each socket type and protocol family (protocol=0), otherwise ID of the protocol you want to use

4.3 TCP connection

In TCP connection, defined by type **SOCK_STREAM** as written in the Section 4.2, there is a client that connects to a server. It uses three primitives (related to File System primitives for management of files on disk) that do these logic actions:

1. start (open bytes stream)
2. add/remove bytes from stream
3. finish (close bytes stream)

TCP is used transferring big files on the network and for example with HTTP, that supports parallel download and upload (FULL-DUPLEX). The length of the stream is defined only at closure of the stream.

4.3.1 Client

4.3.1.1 connect()

The client calls **connect()** function, after **socket()** function of Section 4.2. This function is a system call that client can use to define what is the remote terminal to which he wants to connect.

```
#include <sys/types.h>
#include <sys/socket.h>

int connect(int sockfd, const struct sockaddr *addr, socklen_t addrlen);
```

RETURN VALUE *0* if connection succeeds
 -1 if some error occurs and `errno` is set appropriately

sockfd = *Socket File Descriptor* returned by `socket()`.

addr = *Reference to struct sockaddr*

`sockaddr` is a general structure that defines the concept of address.

In practice it's a union of all the possible specific structures of each protocol.

This approach is used to leave the function written in a generic way.

addrlen = *Length of specific data structure used for sockaddr.*

In the following there is the description of struct `sockaddr_in`, that is the specific `sockaddr` structure implemented for family of protocols **AF_INET**:

```
#include <netinet/in.h>

struct sockaddr_in {
    sa_family_t   sin_family; /* address family: AF_INET */
    in_port_t     sin_port;   /* port in network byte order */
    struct in_addr sin_addr;   /* internet address */
};

/* Internet address. */
struct in_addr {
    uint32_t      s_addr;     /* address in network byte order */
};\
```

The two addresses, needed to define a connection, are (see Figure 4.2):

- **IP address** (*sin_addr* in `sockaddr_in` struct)
 identifies a virtual interface in the network. It can be considered the entry-point for data arriving to the computer. *It's unique in the world.*
- **Port number** (*sin_port* in `sockaddr_in` struct)
 identifies to which application data are going to be sent. The port so must be open for that stream of data and it can be considered a service identifier. There are well known port numbers, related to standard services and others that are free to be used by the programmer for its applications (see Section 12.2 to find which file contains well known port numbers). *It's unique in the system.*

As mentioned in Section 3.1, network data are organized as Big Endian, so in this case we need to insert the IP address according to this protocol. It can be done creating an array of `char` and analysing it as an `int` pointer* or with the follow function, that converts a string (E.g. "127.0.0.1") in the corresponding address:

```
#include <sys/socket.h>
#include <netinet/in.h>
#include <arpa/inet.h>

int inet_aton(const char *cp, struct in_addr *inp);
```

If you want to obtain the IP address from the name of the host, using DNS, you need to use the following function that returns in `h_addr_list` the set of ip addresses related to that hostname, as arrays of characters:

```
#include <netdb.h>
extern int h_errno;

struct hostent *gethostbyname(const char *name);

struct hostent
{
    char *h_name;           /* official name of host */
    char **h_aliases;       /* alias list */
    int h_addrtype;         /* host address type */
    int h_length;           /* length of address */
    char **h_addr_list;     /* list of addresses */
}
```

The port number is written according to Big Endian architecture, through the next function:

```
#include <arpa/inet.h>

uint16_t htons(uint16_t hostshort);
```

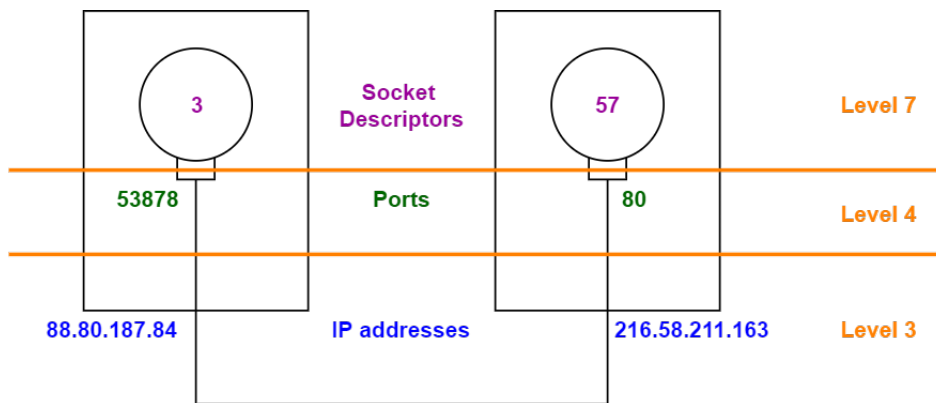


Figure 4.2: After successful connection.

4.3.1.2 write()

Application protocol uses a readable string, to exchange readable information (as in HTTP). This technique is called simple protocol and commands, sent by the protocol, are standardized and readable strings.

```
#include <unistd.h>

ssize_t write(int fd, const void *buf, size_t count);
```

RETURN VALUE *Number of bytes written on success*
-1 if some error occurs and errno is set appropriately

fd = *Socket File Descriptor* returned by socket().

buf = *Buffer of characters to write*

count = *Max number of bytes to write in the file (stream).*

The write buffer is usually a string but we don't consider the null value (`\0` character), that determine the end of the string, in the evaluation of count (`strlen(buf)-1`). This convention is used because `\0` can be part of characters stream.

4.3.1.3 read()

The client uses this blocking function to wait and obtain response from the remote server. Not all the request are completed immediat from the server, for the meaning of stream type of protocol. Infact in this protocol, there is a flow for which the complete sequence is defined only at the closure of it4.2.

`read()` is consuming bytes fom the stream asking to level 4 a portion of them, because it cannot access directly to bytes in Kernel buffer. Lower layer controls the stream of information that comes from the same layer of remove system.

```
#include <unistd.h>

ssize_t read(int fd, void *buf, size_t count);
```

RETURN VALUE *Number of bytes read on success*
 0 if EOF is reached (end of the stream)
 -1 if some error occurs and errno is set appropriately

fd = *Socket File Descriptor* returned by `socket()`.

buf = *Buffer of characters in which it reads and stores info*

count = *Max number of bytes to read from the file (stream).*

So if `read()` doesn't return, this means that the stream isn't ended but the system buffer is empty. If `read=0`, the function met EOF and the local system buffer is now empty. This helps client to understand that server ended before the connection.

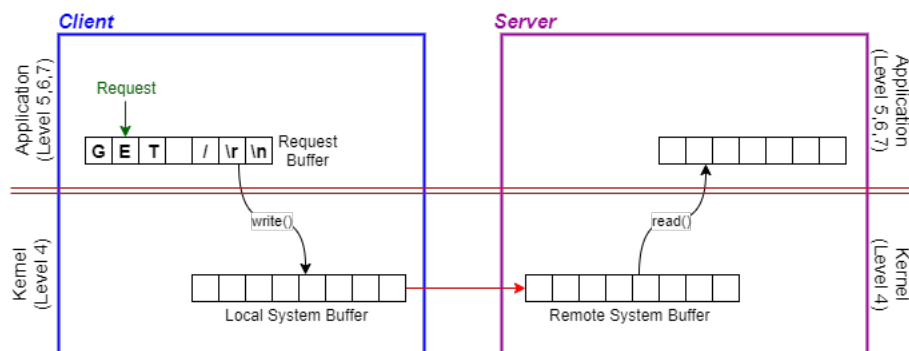


Figure 4.3: Request by the client.

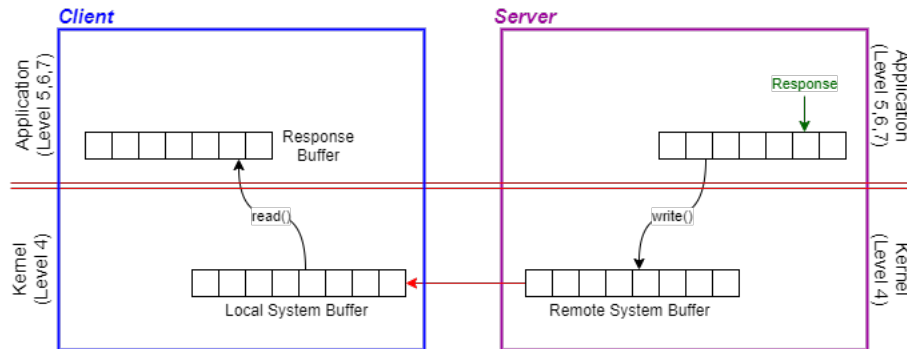


Figure 4.4: Response from the server.

4.3.2 Server

A server is a daemon, an application that works in background forever. The end of this process can be made only through the use of the Operating System.

The server usually uses parallel programming, to guarantee the management of more than one request simultaneously. Hence each process is composed by an infinite loop, as mentioned before.

4.3.2.1 bind()

```
#include <sys/types.h>
#include <sys/socket.h>

int bind(int sockfd, const struct sockaddr *addr, socklen_t addrlen);
```

RETURN VALUE *0* on success
-1 if some error occurs and *errno* is set appropriately
 (You can check value of *errno* including *<errno.h>*).

sockfd = *Socket File Descriptor* returned by *socket()*.

addr = *Reference to struct sockaddr*
sockaddr is a general structure that defines the concept of address.

addrlen = *Length of specific data structure used for sockaddr.*

4.3.2.2 listen()

```
#include <sys/types.h>
#include <sys/socket.h>

int listen(int sockfd, int backlog);
```

RETURN VALUE *0* on success
-1 if some error occurs and `errno` is set appropriately
 (You can check value of `errno` including `<errno.h>`).

sockfd = *Socket File Descriptor* returned by `socket()`.

backlog = *Maximum length of queue of pending connections*
 The number of pending connections for `sockfd` can grow up
 to this value.

The normal distribution of new requests by clients
 is usually Poisson, organized as in Figure 4.5.

The listening socket, identified by **sockfd**, is unique for each association of a port number and a IP address of the server (Figure 4.6).

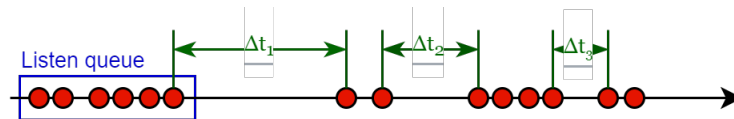


Figure 4.5: Poisson distribution of connections by clients.

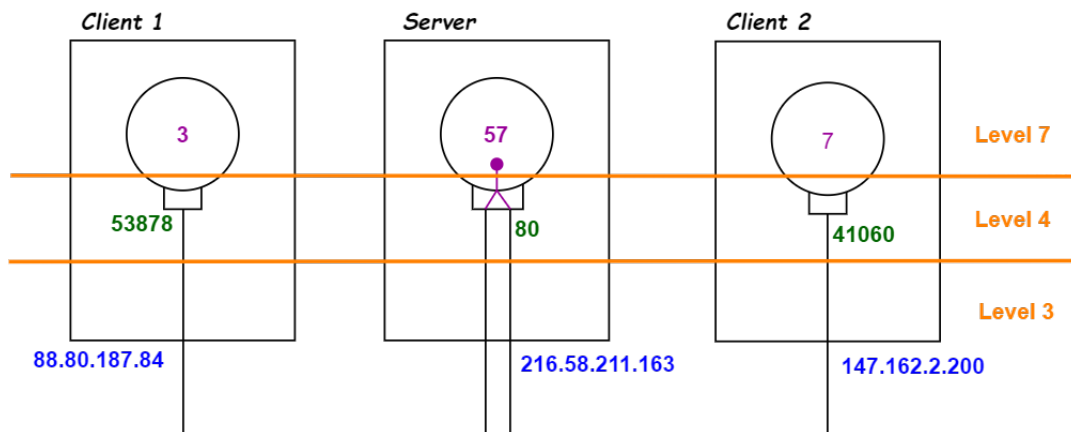


Figure 4.6: `listen()` function.

4.3.2.3 `accept()`

```
#include <sys/types.h>
#include <sys/socket.h>

int accept(int sockfd, struct sockaddr *addr, socklen_t *addrlen);
```

To manage many clients requests, we use the **accept()** function to establish the connection one-to-one with each client, creating a uniquely socket with each client.

This function extracts the first connection request on the queue of pending connections for the listening socket **sockfd** creates a new connected socket, and returns a new file descriptor referring to that socket. The `accept()` is blocking for the server when the queue of pending requests is empty (Figure 4.8).

RETURN VALUE *Accepted Socket Descriptor*
 it will be used by server, to manage requests and responses from that specific client.
 -1 if some error occurs and errno is set appropriately
 (You can check value of errno including <errno.h>).

sockfd = *Listen Socket File Descriptor*

addr = *Reference to struct sockaddr*
 It's going to be filled by the accept() function.

addrlen = *Length of the struct of addr.*
 It's going to be filled by accept() function.
 (accept() is used in different cases so it can return different type of specific implementation of struct addr.)

At lower layers of ISO/OSI, the port number and the IP Address are the same identifiers, to which listening socket is associated (Figure 4.7).

The server needs to do a fork after doing the accept(), inside the infinite loop. Hence a new process is created to manage a new request and there is a pair client-worker for each client. So the server can be seen as it would be composed by many servers (Figure 4.9).

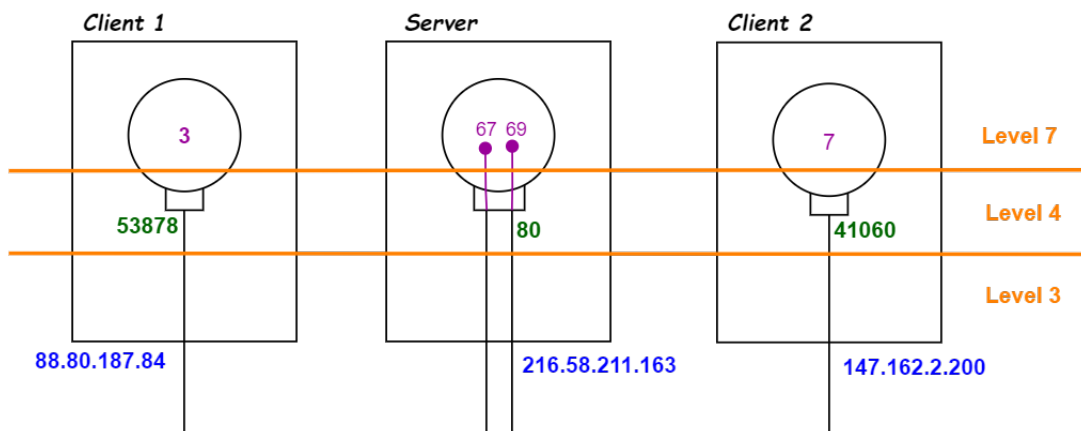
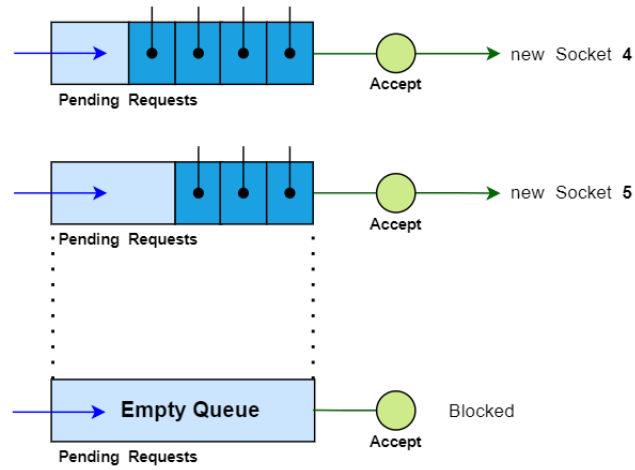
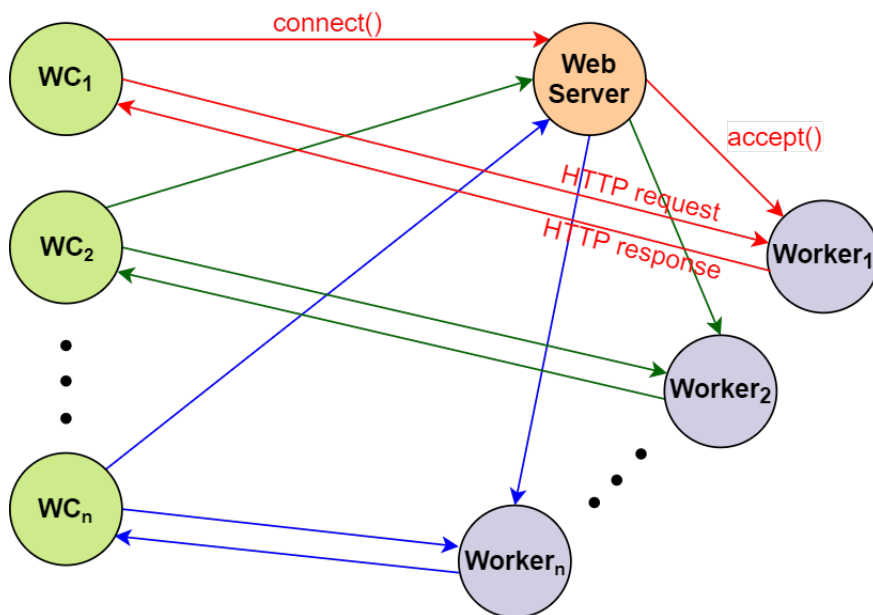


Figure 4.7: accept() function.

Figure 4.8: Management of pending requests with `accept()`.Figure 4.9: `connect()` and `accept()` functions in parallel server implementation.

4.4 UDP connection

UDP connection is defined by type **SOCK_DGRAM** as specified in Section 4.2. It's used for application in which we use small packets and we want immediate feedback directly from application. It isn't reliable because it doesn't need confirmation in transport layer. It's used in Twitter application and in video streaming.

Chapter 5

Gateway

A gateway is a device that forwards messages from another device, the client, to a second device, the server or another gateway. In the following figures, there are two examples of gateways: Layer-3 gateways (routers) and Layer-7 gateways (proxy).

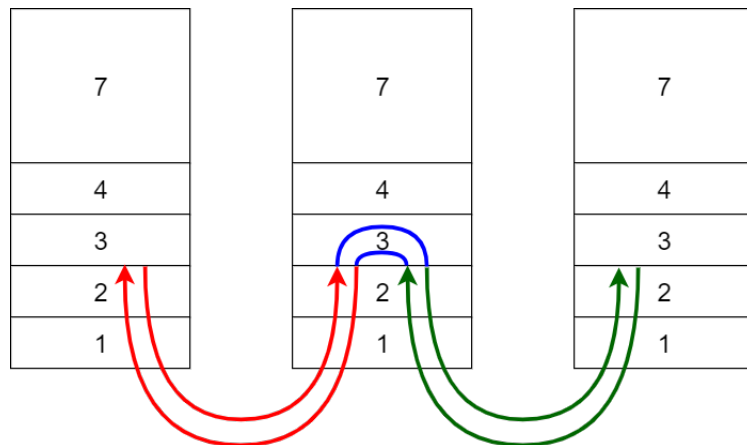


Figure 5.1: Router (Layer-3 gateway).

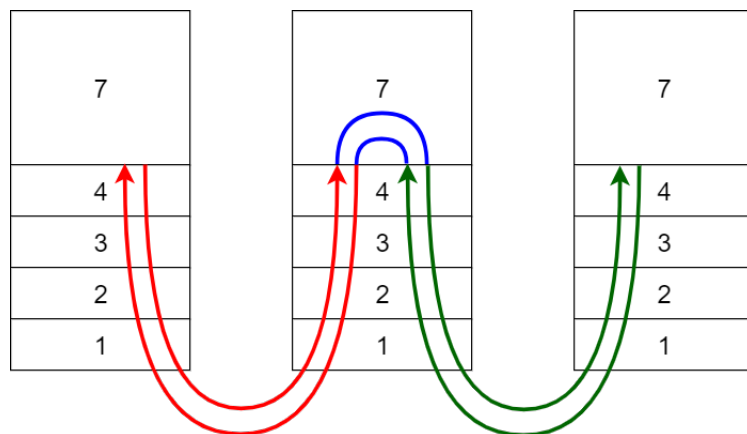


Figure 5.2: Proxy (Layer-7 gateway).

5.1 Proxy

A Layer-7 gateway is also called proxy. It works as an intermediary between two identical protocols (Figure 5.3). Instead of Layer-3 gateways, proxy can also see the full stream of data, analyze HTTP headers and implement new functions. The main possible functions are:

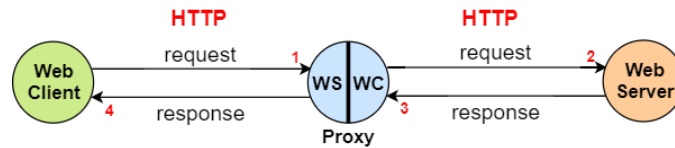


Figure 5.3: Example of proxy use.

- **Caching**

It's used to reduce traffic directed to the server. The proxy does the most expensive job, managing all the requests of the same page of the server.

After the request of the page for the first time, the proxy asks the page to the server and then stores in its system, before replying. Hence the next clients requests of the same page will be manage only by proxy because the page was already stored in its system.

In this case the server needs to manage only a request by proxy and provide a response to proxy.

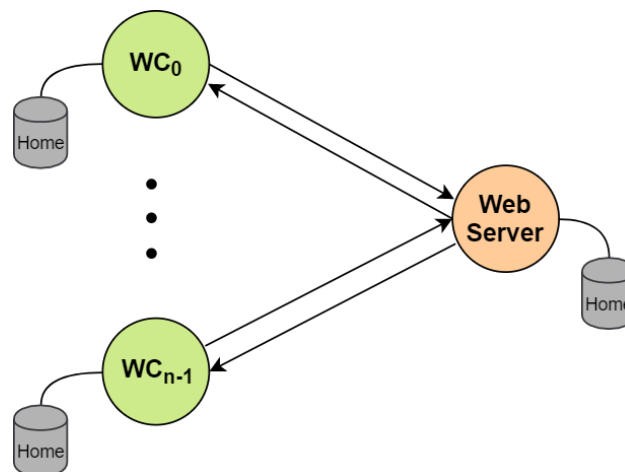


Figure 5.4: Example of caching without proxy.

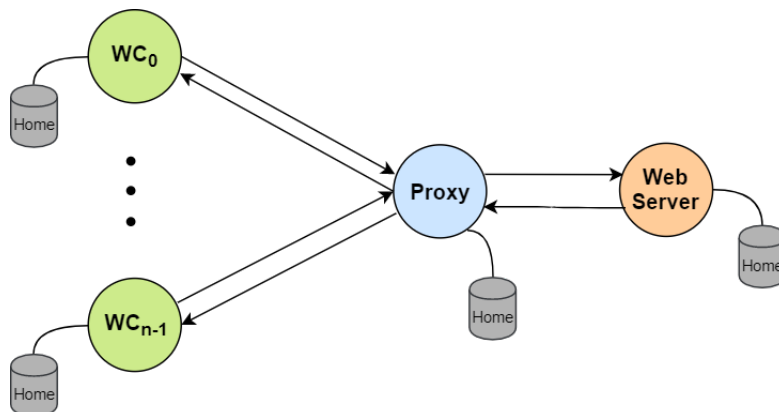


Figure 5.5: Example of caching using proxy.

- **Filtering**

The proxy can do two actions:

- **Filtering the requested resource by the client**

there are many companies that doesn't give access to some services (E.g. no access to Facebook, Youtube, ...).

We cannot use a filtering approach at lower levels because in some cases clients can access to services through intermediate addresses, different from the one we want to reach. Hence we need to analyze the HTTP request at upper layer.

- **Filtering the content of the response**

for parent control approach.

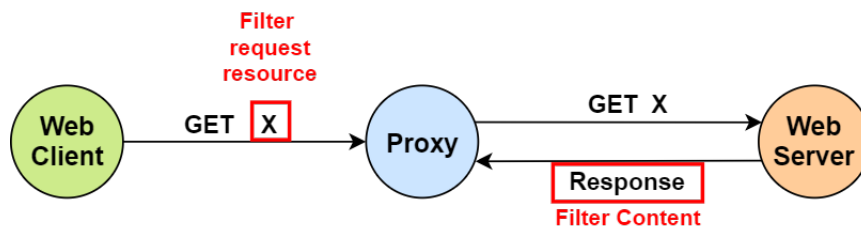


Figure 5.6: Example of proxy filtering.

- **Web Application Firewall (WAF)**

The proxy is specialized and used to block suspicious requests. This is done by analyzing request content, looking for not secure pattern.

A possible pattern can be ".." in the path of the resource, that could give access to not accessible part of the File System (injection). Another possible pattern could be a suspicious parameter for a web application to manage SQL database (SQL injection).

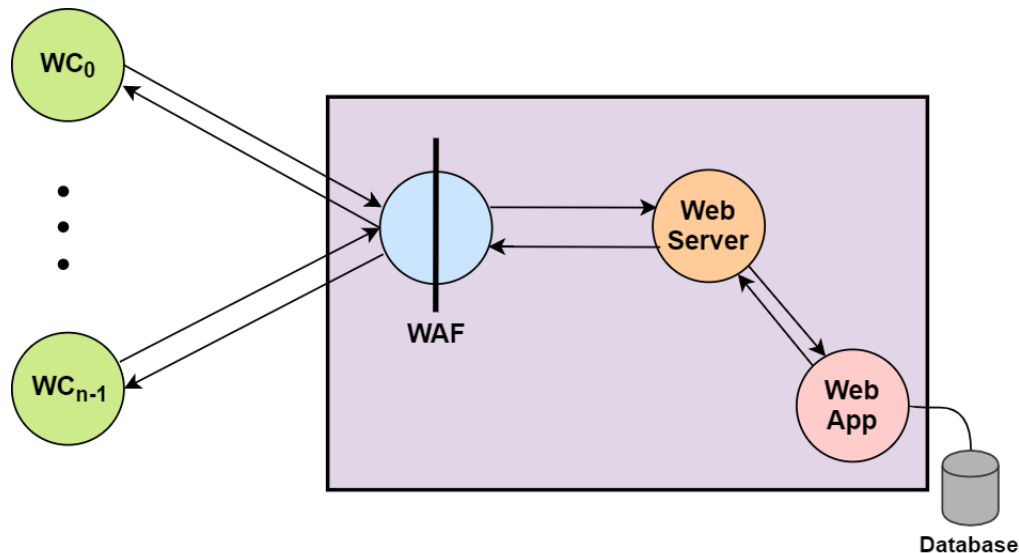


Figure 5.7: Example of WAF use.

- **Load Balancing**

The proxy is a load balancer for the clients requests to the server.

There are many servers to manage requests by client. The client makes the request of the web page but in the reality it's talking with the proxy, that manage the request by sending it to a particular server.

This action is repeated for each client's request. Hence the client thinks that is talking to one server but in reality, the proxy distribute the requests among several servers.

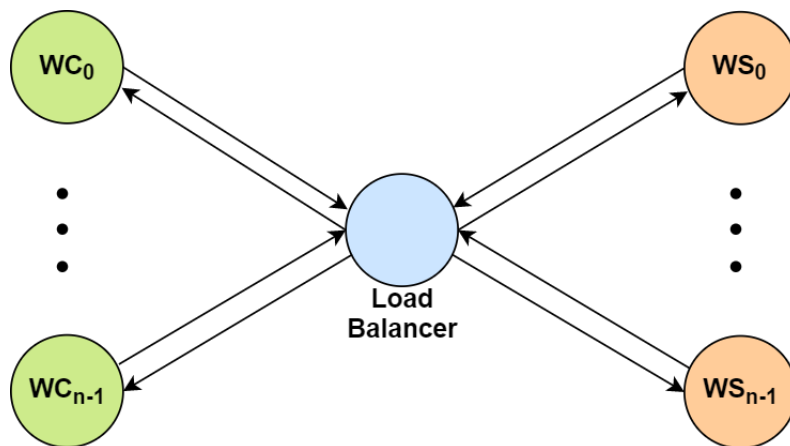


Figure 5.8: Example of load balancing through proxy.

Chapter 6

HTTP protocol

HTTP protocol was presented for the first time in the RFC 1945 [1] (Request for Comment).

The Hypertext Transfer Protocol (HTTP) is an application-level protocol with the lightness and speed necessary for distributed, collaborative, hypermedia information systems. It is a generic, stateless, object-oriented protocol which can be used for many tasks, such as name servers and distributed object management systems, through extension of its request methods (commands).

It's not the first Hypertext protocol in history because there was Hypertalk, made by Apple before.

A feature of HTTP is the typing of data representation, allowing systems to be built independently of the data being transferred. HTTP has been in use by the World-Wide Web global information initiative since 1990.

6.1 Terminology

- **connection**
a transport layer virtual circuit established between two application programs for the purpose of communication.
- **message**
the basic unit of HTTP communication, consisting of a structured sequence of octets matching the syntax defined in Section 4 and transmitted via the connection.
- **request**
an HTTP request message.
- **response**
an HTTP response message.
- **resource**
a network data object or service which can be identified by a URI.
- **entity**
a particular representation or rendition of a data resource, or reply from a service resource, that may be enclosed within a request or response message. An entity consists of metainformation in the form of entity headers and content in the form of an entity body.
- **client**
an application program that establishes connections for the purpose of sending requests.
- **user agent**
the client which initiates a request. These are often browsers, editors, spiders (web-traversing robots), or other end user tools.
- **server**
an application program that accepts connections in order to service requests by sending back responses.

- **origin server**
the server on which a given resource resides or is to be created.
- **proxy**
an intermediary program which acts as both a server and a client for the purpose of making requests on behalf of other clients. Requests are serviced internally or by passing them, with possible translation, on to other servers. A proxy must interpret and, if necessary, rewrite a request message before forwarding it. Proxies are often used as client-side portals through network firewalls and as helper applications for handling requests via protocols not implemented by the user agent.
- **gateway**
a server which acts as an intermediary for some other server. Unlike a proxy, a gateway receives requests as if it were the origin server for the requested resource; the requesting client may not be aware that it is communicating with a gateway.
Gateways are often used as server-side portals through network firewalls and as protocol translators for access to resources stored on non-HTTP systems.
- **tunnel**
a tunnel is an intermediary program which is acting as a blind relay between two connections. Once active, a tunnel is not considered a party to the HTTP communication, though the tunnel may have been initiated by an HTTP request. The tunnel ceases to exist when both ends of the relayed connections are closed.
Tunnels are used when a portal is necessary and the intermediary cannot, or should not, interpret the relayed communication.
- **cache**
a program's local store of response messages and the subsystem that controls its message storage, retrieval, and deletion. A cache stores cachable responses in order to reduce the response time and network bandwidth consumption on future, equivalent requests. Any client or server may include a cache, though a cache cannot be used by a server while it is acting as a tunnel.

Any given program may be capable of being both a client and a server; our use of these terms refers only to the role being performed by the program for a particular connection, rather than to the program's capabilities in general. Likewise, any server may act as an origin server, proxy, gateway, or tunnel, switching behavior based on the nature of each request.

6.2 Basic rules

The following rules are used throughout are used to describe the grammar used in the RFC 1945.

```

OCTET = <any 8-bit sequence of data>
CHAR = <any US-ASCII character (octets 0 - 127)>
UPALPHA = <any US-ASCII uppercase letter "A".."Z">
LOALPHA = <any US-ASCII lowercase letter "a".."z">
ALPHA = UPALPHA | LOALPHA
DIGIT = <any US-ASCII digit "0".."9">
CTL = <any US-ASCII control character (octets 0 - 31) and DEL (127)>
CR = <US-ASCII CR, carriage return (13)>
LF = <US-ASCII LF, linefeed (10)>
SP = <US-ASCII SP, space (32)>
HT = <US-ASCII HT, horizontal-tab (9)>
"> = <US-ASCII double-quote mark (34)>

```

6.3 Messages

6.3.1 Different versions of HTTP protocol

- **HTTP/0.9 Messages**

Simple-Request and Simple-Response do not allow the use of any header information and are limited to a single request method (GET).

Use of the Simple-Request format is discouraged because it prevents the server from identifying the media type of the returned entity.

```
HTTP-message = Simple-Request | Simple-Response
```

```
Simple-Request  = "GET" SP Request-URI CRLF
```

```
Simple-Response = [ Entity-Body ]
```

- **HTTP/1.0 Messages**

Full-Request and Full-Response use the generic message format of RFC 822 for transferring entities. Both messages may include optional header fields (also known as "headers") and an entity body. The entity body is separated from the headers by a null line (i.e., a line with nothing preceding the CRLF).

```
HTTP-message = Full-Request | Full-Response
```

```
Full-Request = Request-Line
               *(General-Header | Request-Header | Entity-Header)
               CRLF
               [ Entity-Body ]

Full-Response = Status-Line
               *(General-Header | Request-Header | Entity-Header)
               CRLF
               [ Entity-Body ]
```

6.3.2 Headers

The order in which header fields are received is not significant. However, it is "good practice" to send General-Header fields first, followed by Request-Header or Response-Header fields prior to the Entity-Header fields. Multiple HTTP-header fields with the same field-name may be present in a message if and only if the entire field-value for that header field is defined as a comma-separated list.

```
HTTP-header = field-name ":" [ field-value ] CRLF
```

6.3.3 Request-Line

```
Request-Line = Method SP Request-URI SP HTTP-Version CRLF

Method       = "GET" | "HEAD" | "POST" | extension-method

extension-method = token
```

The list of methods acceptable by a specific resource can change dynamically; the client is notified through the return code of the response if a method is not allowed on a resource.

Servers should return the status code 501 (not implemented) if the method is unrecognized or not implemented.

6.3.4 Request-URI

The Request-URI is a Uniform Resource Identifier and identifies the resource upon which to apply the request.

```
Request-URI = absoluteURI | abs_path
```

The absoluteURI form is only allowed when the request is being made to a proxy. The proxy is requested to forward the request and return the response. If the request is GET or HEAD and a prior response is cached, the proxy may use the cached message if it passes any restrictions in the Expires header field.

Note that the proxy may forward the request on to another proxy or directly to the server specified by the absoluteURI. In order to avoid request loops, a proxy must be able to recognize all of its server names, including any aliases, local variations, and the numeric IP address.

The most common form of Request-URI is that used to identify a resource on an origin server or gateway. In this case, only the absolute path of the URI is transmitted.

6.3.5 Request Header

The request header fields allow the client to pass additional information about the request, and about the client itself, to the server.

These fields act as request modifiers, with semantics equivalent to the parameters on a programming language method (procedure) invocation.

```
Request-Header = Authorization | From | If-Modified-Since | Referer | User-Agent
```

6.3.6 Status line

```
Status-Line = HTTP-Version SP Status-Code SP Reason-Phrase CRLF
```

General Status code

1xx: Informational	Not used, but reserved for future use
2xx: Success	The action was successfully received, understood, and accepted.
3xx: Redirection	Further action must be taken in order to complete the request
4xx: Client Error	The request contains bad syntax or cannot be fulfilled
5xx: Server Error	The server failed to fulfill an apparently valid request

6.4 HTTP 1.0

The protocol has no mandatory headers to be added in the request field. This protocol is compliant with HTTP 0.9. To keep the connection alive, "Connection" header with "keep-alive" as header field must be added to request message. The server, receiving the request, replies with a message with the same header value for "Connection".

This is used to prevent the closure of the connection, so if the client needs to send another request, he can use the same connection. This is usually used to send many files and not only one.

The connection is kept alive until either the client or the server decides that the connection is over and one of them drops the connection. If the client doesn't send new requests to the server, the second one usually drops the connection after a couple of minutes.

The client could read the response of request, with activated keep alive option, reading only header and looking to "Content-length" header field value to understand the length of the message body. This header is added only

Known service code

200	OK
201	Created
202	Accepted
204	No Content
301	Moved Permanently
302	Moved Temporarily
304	Not Modified
400	Bad Request
401	Unauthorized
403	Forbidden
404	Not Found
500	Internal Server Error
501	Not Implemented
502	Bad Gateway
503	Service Unavailable

if a request with keep-alive option is done.

This must be done because we can't look only to empty system stream, because it could be that was send only the response of the first request or a part of the response.

Otherwise, when the option keep alive is not used, the client must fix a max number of characters to read from the specific response to his request, because he doesn't know how many character compose the message body. If you make many requests to server without keep-alive option, the server will reply requests, after the first, with only headers but empty body.

6.4.1 Other headers of HTTP/1.0 and HTTP/1.1

- **Allow**
lists the set of HTTP methods supported by the resource identified by the Request-URI
- **Accept**
lists what the client can accept from server. It's important in object oriented typing concept because client application knows what types of data are allowed for its methods or methods of used library
- **Accept-encoding**
specifies what type of file encoding the client supports (don't confuse it with transfer encoding)
- **Accept-language**
specifies what language is set by Operating System or it's specified as a preference by client on browser
- **Content-Type**
indicates the media type of the Entity-Body sent to the recipient. It is often used by server to specify which one of the media types, indicated by the client in the Accept request, it will use in the response.
- **Date**
specifies the date and time at which the message was originated
- **From**
if given, it should contain an Internet e-mail address for the human user who controls the requesting user agent (it was used in the past)

- **Location**

defines the exact location of the resource that was identified by the Request-URI (useful for 3xx responses)

- **Pragma**

It's sent by server to inform that there in no caching systems

- **Referer**

allows the client to specify, for the server's benefit, the address (URI) of the resource from which the Request-URI was obtained (page from which we clicked on the link). This allows a server to generate lists of back-links to resources for interest, logging, optimized caching, etc. It was added with the born of economy services related to web pages.

- **Server**

information about the software used by the origin server to handle the request (usually Apache on Unix, GWS(Google Web Server), Azure on Windows, ...)

- **User-agent**

Version of client browser and Operating System. It's used to:

- adapt responses to application library
- manage mobile vs desktop web pages

It's crucial for web applications. If we are the clients and we receive the response from server, we want that the content must change according to the version of browser.

Infact, there are two different web pages(two different view of the same web page) according to connection by pc and phone, because of different user-agent of these devices. If a mobile phone sends a request to a non-mobile web page, the user agent changes to user agent related to Desktop version.

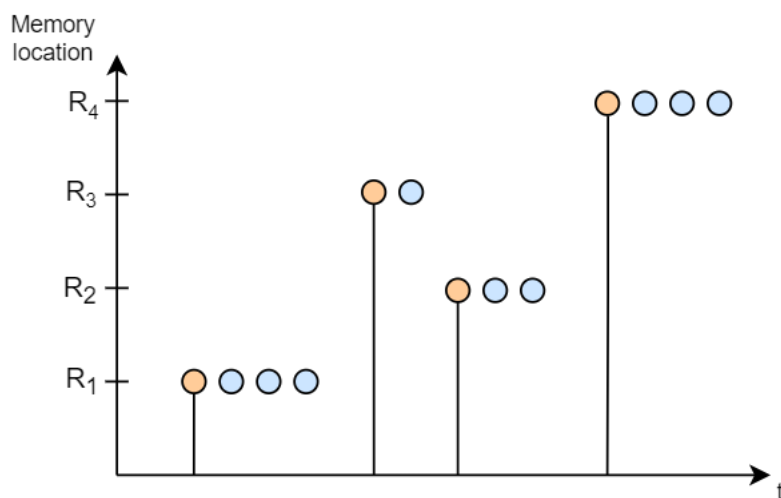
6.4.2 Caching

It's based on locality principle and was observed on programs execution.

- **Time Locality**

When a program accesses to an address, there will be an access to it again in the near future with high probability.

If I put this address in a faster memory (cache), the next access to the same location would be faster.



- **Space Locality**

If a program accesses to an address in the memory, it's very probable that neighboring addresses would be accessed next.

The caching principle is applied also in Computer Networks, storing of the visited web pages on client system and then updating them through the use of particular headers and requests (see Figure ??). The purpose of using cache is to reduce traffic over the network and load of the server. The main problem of storing the page in a file, used as a cache, is that the page on the server can be modified and so client's copy can be obsolete.

The update of the content of the local cache for the client can happen in three different ways:

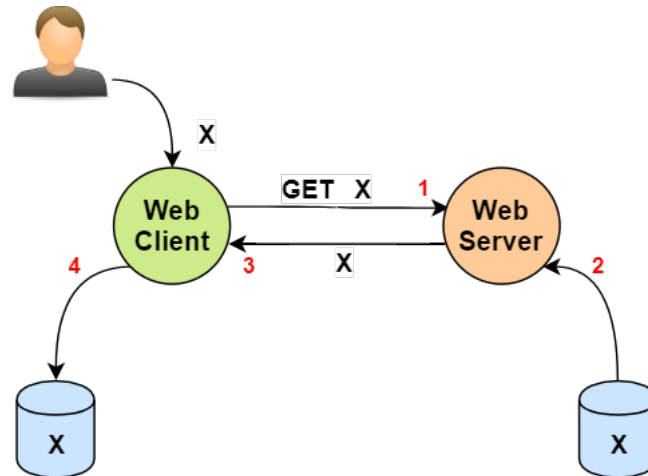
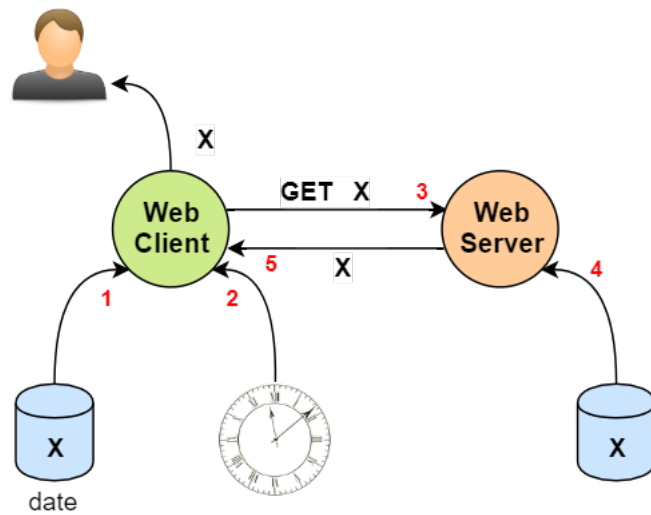
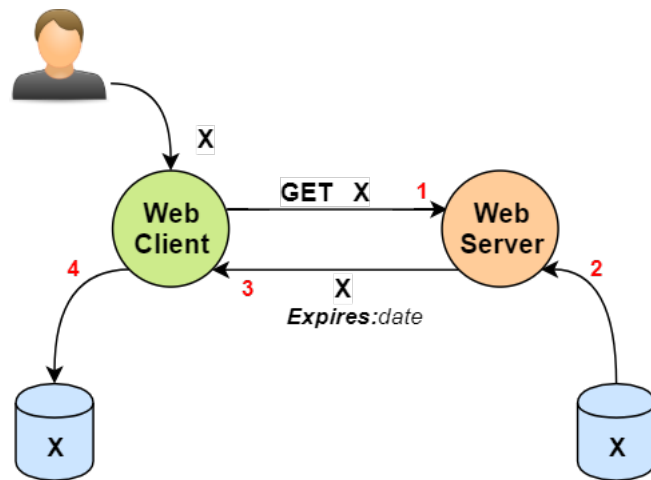


Figure 6.1: First insertion of the resource in the cache.

• Expiration date

1. The client asks the resource to the server, that replies with the resource and adding "Expires" header. This is done by the server to specify when the resource will be considered obsolete.
2. The client stores a copy of the resource in its local cache.
3. The client, before sending a new request, checks if it has already the resource he's asking to server. If he has already the resource, he compares the Expiration date, specified by server at phase 1, with the real time clock. A problem of this method is that the server needs to know in advance when the page changes. So the "Expires" value, sent by server, must be:
 - exactly known in advance for periodic changes (E.g. daily paper)
 - statistically computed (evaluating the probability of refreshing and knowing a lower bound of duration of resource)

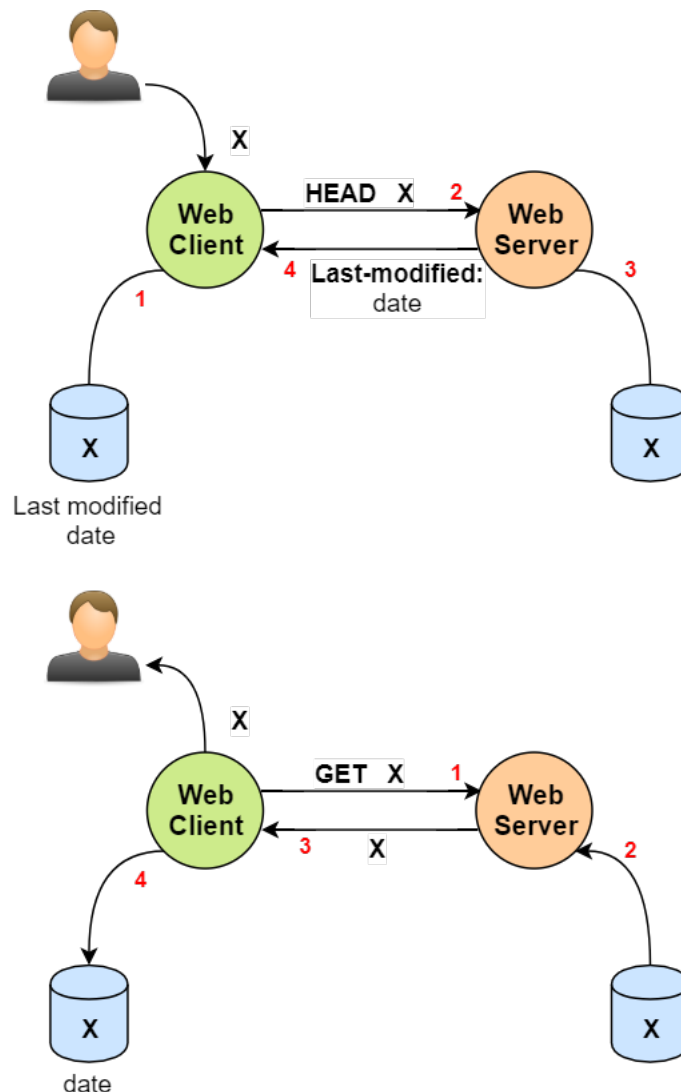
The other problem of this method is that we need to have server and client clocks synchronized. Hence, we need to have date correction and compensation between these systems.



- Request of only header part

1. The client asks the resource to the server as before but now, he stores resource in the cache, within also its "Last-Modified" header value.
2. The client checks if its copy of the resource is obsolete by making a request to the server of only the header of the resource. This type of request is done by using the *"HEAD"* method.
3. The client looks to the value of the header "Last-Modified", received by the server. This value is compared with the last-modified header value stored within the resource.
If the store date was older than new date, the client makes a new request for the resource to the server. Otherwise, he uses the resource in the cache.

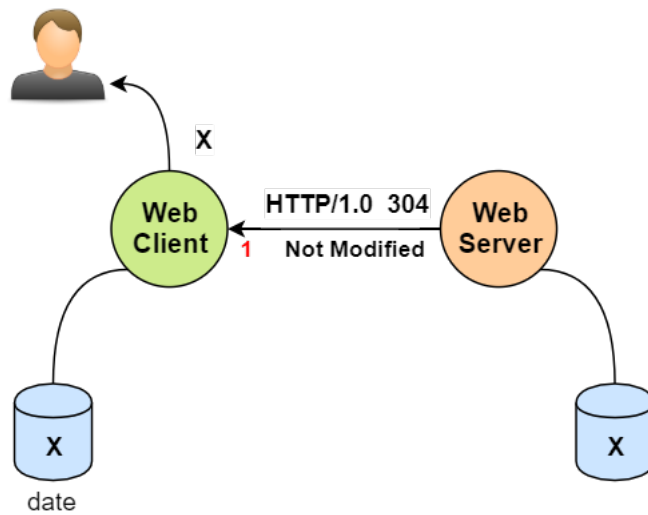
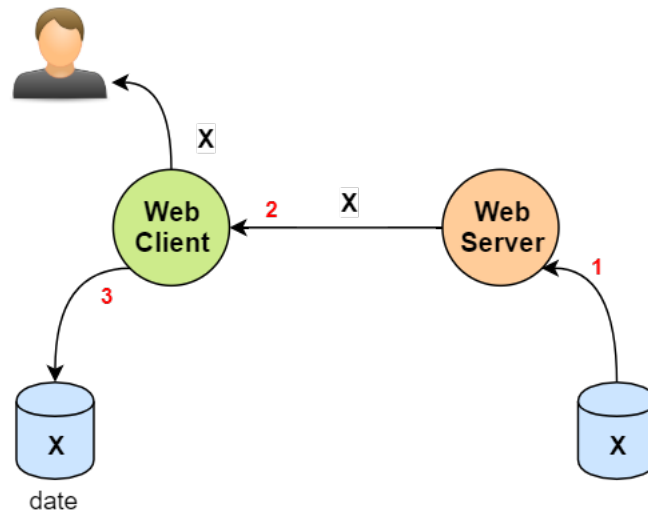
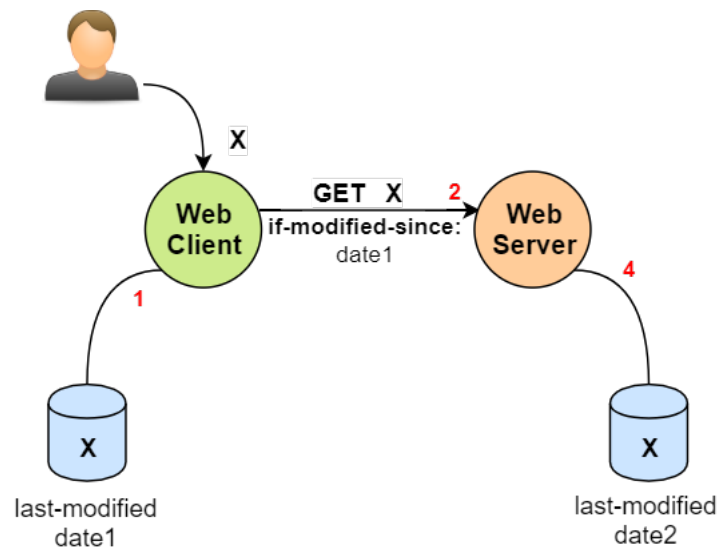
The problem of this method is that, in the worst case, we send two times the request of the same resource (even if the first one, with "HEAD" method, is less heavy).



- **Request with if-modified-since header**

1. The client asks the resource to the server as before, storing the resource in the cache within its "Last-Modified" header value.
2. When the client needs again the resource, it sends the request to the server, specifying also "If-Modified-Since" header value as store data.
3. If the server, looking to the resource, sees that its Last-Modified value is more recent than date specified in the request by client, it sends back to the recipient the newer resource. Otherwise, it sends to client the message "HTTP/1.0 304 Not Modified".

The positive aspect of this method is that the client can do only a request and obtain the corrept answer without other requests.



6.4.3 Authorization

1. The client sends the request of the resource to the client
2. The server knows that the resource, to be accessible, needs the client authentication, so it sends the response specifying "WWW-Authenticate:" header, as the following:

```
WWW-Authenticate: Auth-Scheme Realm="XXXX"
```

Auth-Scheme Type of encryption adopted

Realm "XXXX" referring to the set of users that can access to the resource

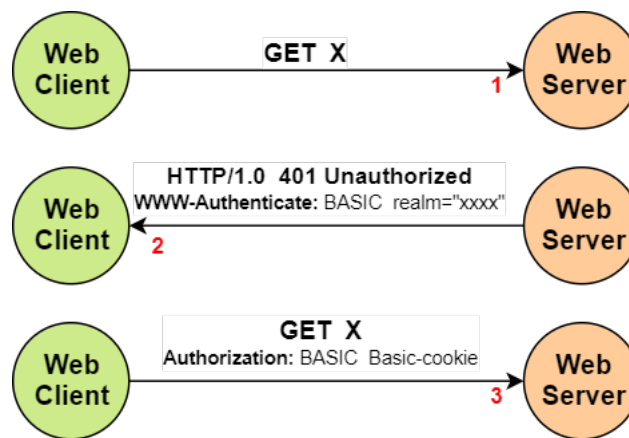
3. The client replies with another request of the same resource but specifying also the "Authorization" header value, as the following:

```
WWW-Authenticate: Auth-Scheme Basic-cookie
```

Auth-Scheme Type of encryption adopted

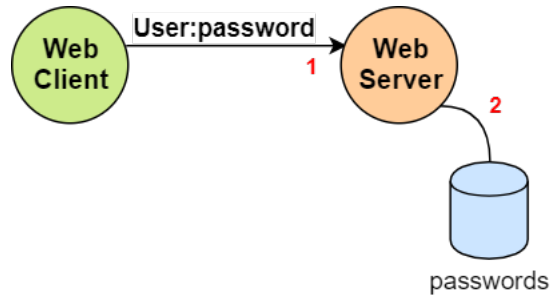
Basic-cookie Base64 encrypted message of the needed for the authentication

(in general basic-cookie doesn't contain password inside it, it happens only in this case)

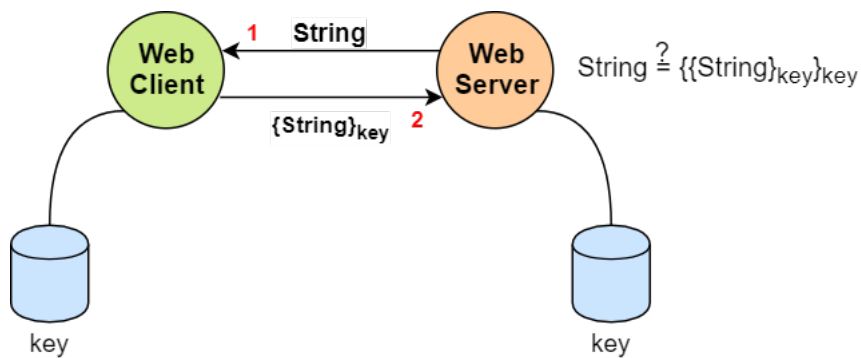


6.4.3.2 Auth-schemes

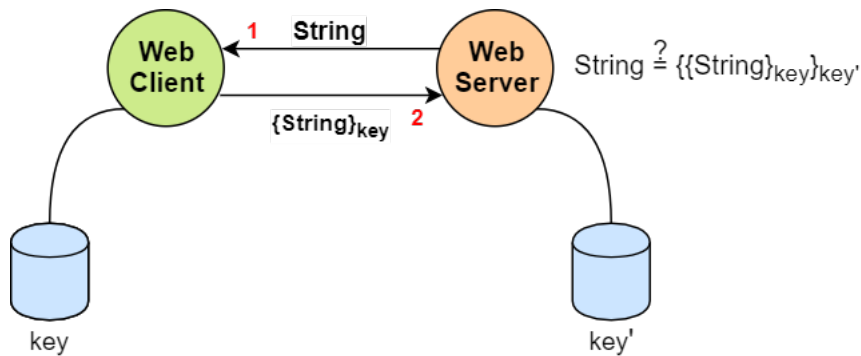
• BASIC



• Challenge (symmetric version)



• Challenge (asymmetric version)



6.5 HTTP 1.1

The architecture of the model is in RFC2616 [2]. It has by default the option keep alive activated by default with respect to HTTP 1.0. It has the mandatory header "Host" followed by the hostname of the remote system, to which the request or the response is sent. The headers used in HTTP/1.0 are used also in HTTP/1.1, but in this new protocol there are new headers not used in the previous one. The body is organized in chunks, so we

need the connection kept alive to manage future new chunks.

This is useful with dynamic pages, in which the server doesn't know the length of the stream in advance and can update the content of the stream during the established connection, sending a fixed amount of bytes to client. We can check if the connection is chunked oriented, looking for the header "Transfer-Encoding" with value "chunked".

Each connection is composed by many chunks and each of them is composed by chunk length followed by chunk body, except for the last one that has length 0 (see Figure 6.2). The following grammar represents how the body is organized:

```

Chunked-Body    = *chunk
                  last-chunk
                  trailer
                  CRLF

chunk           = chunk-size [ chunk-extension ] CRLF
                  chunk-data CRLF

chunk-size      = 1*HEX
last-chunk      = 1*("0") [ chunk-extension ] CRLF

chunk-extension = *( ";" chunk-ext-name [ "=" chunk-ext-val ] )

chunk-ext-name  = token
chunk-ext-val   = token | quoted-string
chunk-data      = chunk-size(OCTET)
trailer         = *(entity-header CRLF)

```

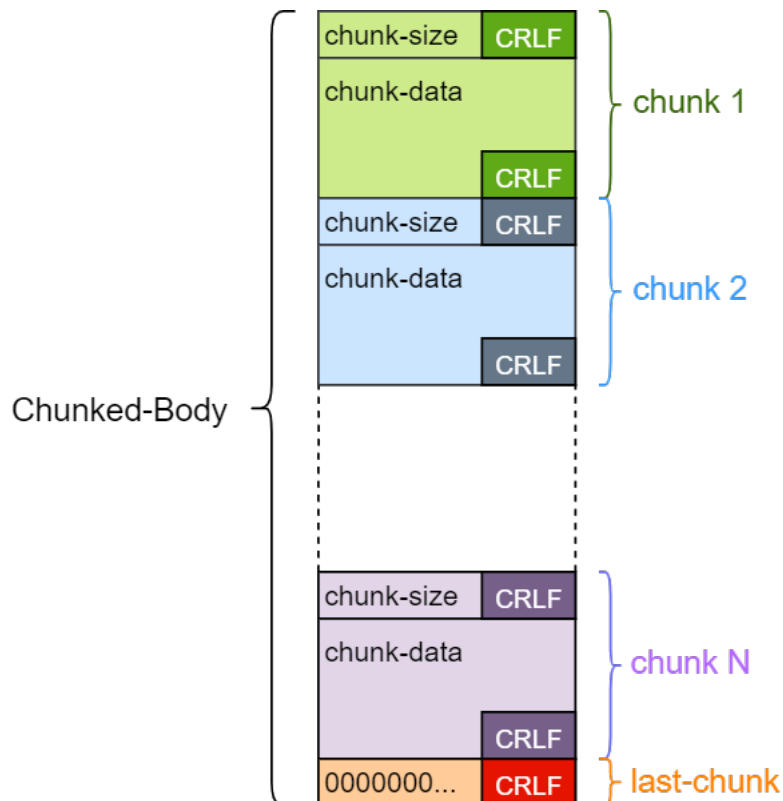


Figure 6.2: Chunked body.

6.5.1 Caching based on HASH

It's like the caching mechanism used looking to "Last-Modified" header value through the use of HEAD. The organization is as follows:

1. The client asks the resource to the server, he stores resource in the cache, within also its "Etag" header value.
2. The client checks if its copy of the resource is obsolete by making a request to the server of only the header of the resource. This type of request is done by using the *"HEAD"* method.
3. The client looks to the value of the header "Etag", received by the server. This value is compared with the "Etag" header value stored within the resource, because everytime that a file changes, its hash code is computed again.
If the store date has different hash code from one received, the client makes a new request for the resource to the server. Otherwise, he uses the resource in the cache.

6.5.2 URI

In URI, there is the encapsulation of the operation done in the past, to have a resource from a server [3]. The following phases are related to **ftp** application:

1. Open the application ftp
2. Open the server File System, through a general login
3. Select the resource you want to use and download it

URI	= (absoluteURI relativeURI) ["#" fragment]
absoluteURI	= scheme ":" *(uchar reserved)
relativeURI	= net_path abs_path rel_path
net_path	= "//" net_loc [abs_path]
abs_path	= "/" rel_path
rel_path	= [path] [";" params] ["?" query]
path	= fsegment *("/" segment)
fsegment	= 1*pchar
segment	= *pchar
params	= param *(";" param)
param	= *(pchar "/")
scheme	= 1*(ALPHA DIGIT "+" "-" ".")
net_loc	= *(pchar ";" "?")
query	= *(uchar reserved)
fragment	= *(uchar reserved)
pchar	= uchar ":" "@" "&" "=" "+"
uchar	= unreserved escape
unreserved	= ALPHA DIGIT safe extra national
escape	= "%" HEX HEX
reserved	= ";" "/" "?" ":" "@" "&" "=" "+"
extra	= "!" "*" "'" "(" ")" ","
safe	= "\$" "_" "." "
unsafe	= CTL SP <"> <"# <"%" <"<" >">"
national	= <any OCTET excluding ALPHA, DIGIT,

Hence Uniform Resource Identifiers are simply formatted strings which identify via name, location, or any other characteristic a network resource. The following example refers to Relative URI:

```
//net_loc/a/b/c?parameters
```

`//net_loc` Server location
`/a/b/c` Resource with the path
`?parameters` Set of parameters

6.5.3 HTTP URL

It's a particulare instance of absolute URI, with scheme "http".

```

http_URL      = "http:" "/" host [ ":" port ] [ abs_path ]
host          = <A legal Internet host domain name
                or IP address (in dotted-decimal form),
                as defined by Section 2.1 of RFC 1123>
port          = *DIGIT

```

There are also other schemes that are not used for web [4], for example **ftp** to download resources.

6.6 Dynamic pages

Dynamic pages are created on fly by some web applications in the server. The client makes a request to the server function with some parameters (Figure 6.3).

This approach is based on **Common Gateway Interface (CGI-bin)**, whose name comes from first network applications that were binary. Then the evolution of web applications brings to two types of program:

- **Script Server programs**
based on PHP, ASP.net
- **Server application (based on Java)** written through J2EE, TomCat and Websphere

The result of these programs are written at Presentation layer, like HTML source. To use the CGI-bin paradigm, the client needs to create a request for a file to be executed and not transfered. For convention, the server usually has its executable files in **"/CGI-bin"** path of the server. The following HTTP URL is the request to the server, made by the client, for the function **f**:

```
http://www.hello.com/CGI-bin/f?a=10&b=20&c=%22ciao%22
```

In this example the client is asking to server **www.hello.com**, using an HTTP URL, the result of the call of function **f**. The symbol **?** defines from which point the parameters of the function are specified. In this case there are three parameters: **a** with value **10**, **b** with value **20** and **c** with value **%22ciao%22**. There are particular symbols, used in URL:

?	Beginning of parameters section
%	<i>Escape character</i> followed by the hex number that defines the symbol you want to code
&	<i>Separator character</i> character between each couple of specified parameters

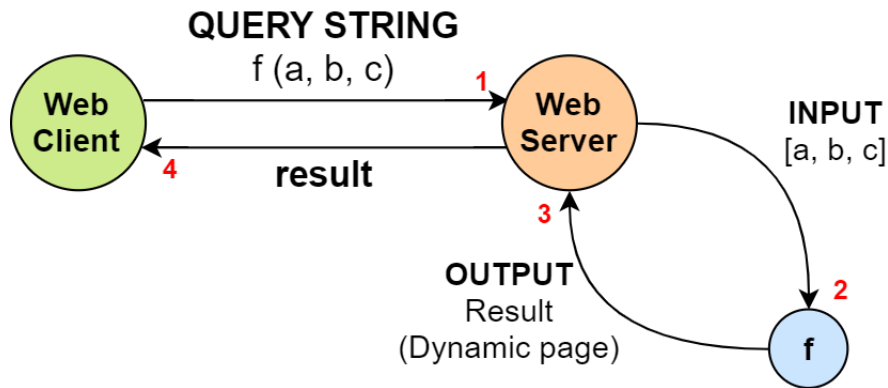


Figure 6.3: Example of CGI application.

6.7 Proxy

The implementation of the proxy depends on the type of protocol used:

- **HTTP**

If the client wants to use a proxy, doing a *GET* request, he needs to modify its behaviour with the following steps:

1. **Connection of Client to Web Proxy instead of the server**

The client needs to change address and port w.r.t. proxy ones, instead of server ones.

2. **Specify the absolute URI of the requested resource**

Otherwise proxy doesn't to which one the message needs to be sent. Hence he couldn't forward as it is the request.

The proxy can analyze the content of data they need to transmit, obtaining the absolute URI and doing another request.

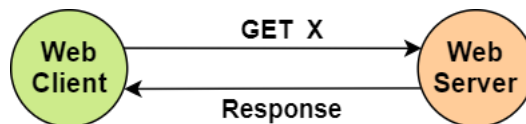


Figure 6.4: Direct access.

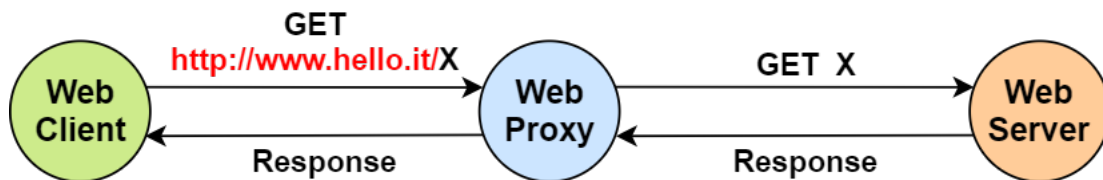


Figure 6.5: Proxy access.

- **HTTPS**

Data are sent over encrypted channel (TLS) and the proxy can be implemented in two different ways:

- **Split the encrypted channel**

The proxy has an encrypted channel with the client and one with the server. This approach can be applied only when we have a trusted proxy (E.g. WAF) because the proxy needs to access data to forward them.

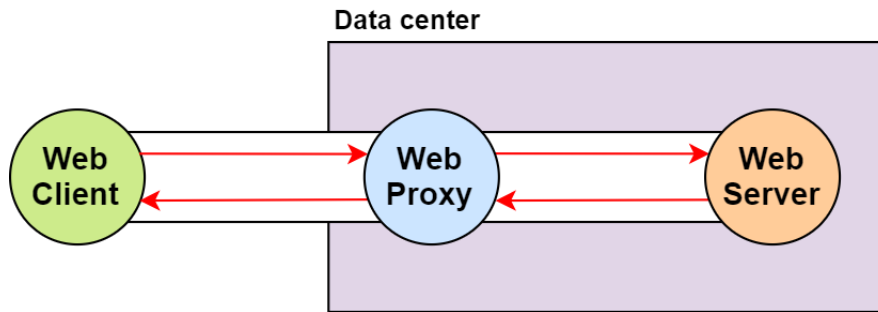


Figure 6.6: Proxy as WAF in HTTPS.

– **Change default behaviour of proxy**

The proxy in this case can only forward encrypted data without knowing anything about them. In this case, proxy works as a Layer-4 gateway and creates a tunnel between client and server [5].

In HTTPS the client uses the method *CONNECT* to tell to the proxy to work as a tunnel. The proxy, receiving the *CONNECT* request, establish the secure connection between client and server (through the preliminary exchange of keys with Diffie-Hellmann).

The proxy sends HTTP response to the client if the `connect()` call succeeded. Then the client can send encrypted data as *raw data* and the proxy will not access them but only forward them. With *CONNECT* request, the client asks to open a connection to web host.

The proxy needs to create two processes (Figure 6.9):

* **Parent process**

It reads response from the server and forwards it to the client. When the connection will be closed from the server, it will kill its child process.

* **Child process**

It reads request from the client and forwards it to the server.

In a browser, when you type an address or server name, the connection starts by default using HTTP. Then the remote server replies with a HTTP response with redirection to an HTTPS URL.

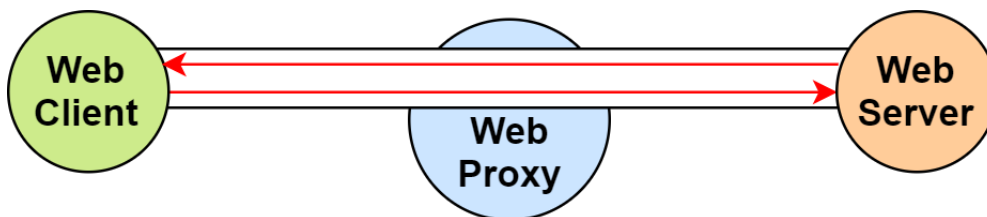


Figure 6.7: Tunneling using proxy in HTTPS.

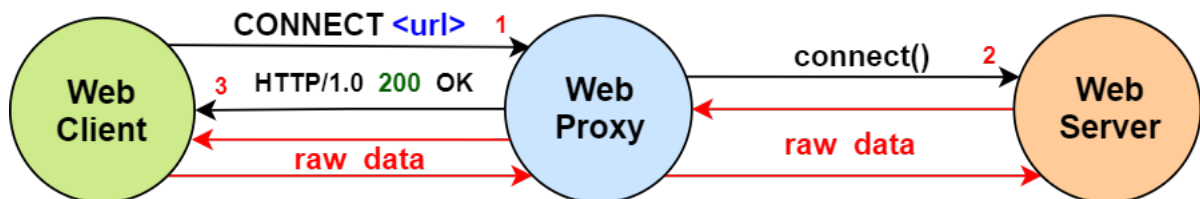


Figure 6.8: CONNECT request in HTTPS.

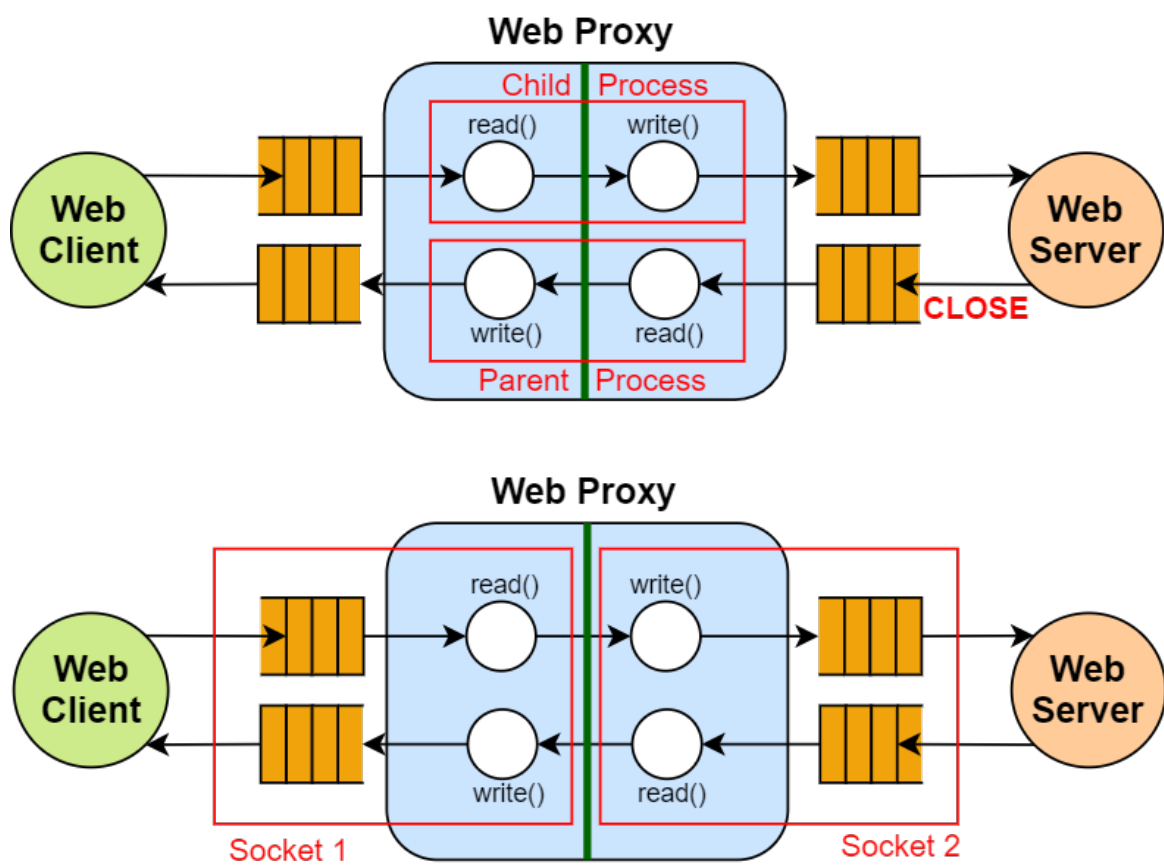


Figure 6.9: How proxy works with HTTPS.

Chapter 7

Resolution of names

The following section will talk about history of technologies under the resolution of server names in URL to their IP addresses, needed to establish the connection.

7.1 Network Information Center (NIC)

This type of architecture was used in the past to resolve names. Each client has its own file **HOSTS.txt**, with resolution of names. The client shared its file with a central system, called **NIC** (Figure 7.1).

This system collects all the files, like an hub, and shared resolution names to other clients (Figure 7.1).

This architecture is unfeasible and not scalable with nowadays number of IP addresses, because the files become very huge and transferring becomes very slow.

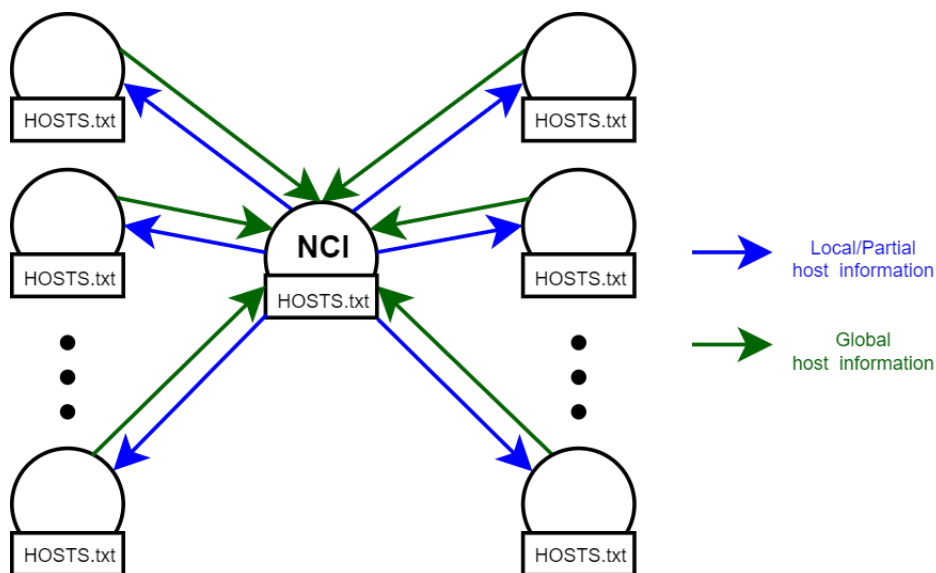


Figure 7.1: How NIC worked.

7.2 Domain Name System (DNS)

The file **HOSTS.txt** is yet used in nowadays UNIX systems. The specified host name is searched in local `/etc/hosts.txt`, that contains local and private addresses resolution table, and if not found, it will be searched through DNS [6].

7.2.1 Goals

1. Names should not be required to contain network identifiers, addresses, routes, or similar information as part of the name.
2. The sheer size of the database and frequency of updates suggest that it must be maintained in a distributed manner, with local caching to improve performance.
Approaches that attempt to collect a consistent copy of the entire database will become more and more expensive and difficult, and hence should be avoided.
The same principle holds for the structure of the name space, and in particular mechanisms for creating and deleting names; these should also be distributed.
3. Where there are tradeoffs between the cost of acquiring data, the speed of updates, and the accuracy of caches, the source of the data should control the tradeoff.
4. The costs of implementing such a facility dictate that it be generally useful, and not restricted to a single application.
We should be able to use names to retrieve host addresses, mailbox data, and other as yet undetermined information. All data associated with a name is tagged with a type, and queries can be limited to a single type.
5. Because we want the name space to be useful in dissimilar networks and applications, we provide the ability to use the same name space with different protocol families or management.
For example, host address formats differ between protocols, though all protocols have the notion of address. The DNS tags all data with a class as well as the type, so that we can allow parallel use of different formats for data of type address.
6. We want name server transactions to be independent of the communications system that carries them.
Some systems may wish to use datagrams for queries and responses and only establish virtual circuits for transactions that need the reliability (e.g., database updates, long transactions); other systems will use virtual circuits exclusively.
7. The system should be useful across a wide spectrum of host capabilities.
Both personal computers and large timeshared hosts should be able to use the system, though perhaps in different ways.

7.2.2 Hierarchy structure

Hierarchy permits to manage a lot of numbers of domain names and IP addresses, reducing the time spent to resolve them. Given for example the host name **www.dei.unipd.it**, we have a **Name Server (NS)** for each of the domain name inside it (Figure 7.2).

The tree hierarchy has a name server for each one of its internal nodes. The name server gives us only the name of the name server of the lower level to which we need to go.

To obtain the IP address of this name server, we need to ask, to name server of upper layer, a **glue record**. The glue record is an additional information that needs us to understand how to reach that name server. Hence the glue record is the IP address of NS of the lower level in hierarchy.

For each request to NS, we obtain also the expiration time information because a caching approach is adopted also in DNS but at level 4. There are 13 root name servers that are returned when asking resolution to root. In reality root name servers are more than 13 but the communication used in DNS is made through UDP and this type of connection supports only 13 simultaneously transferring.

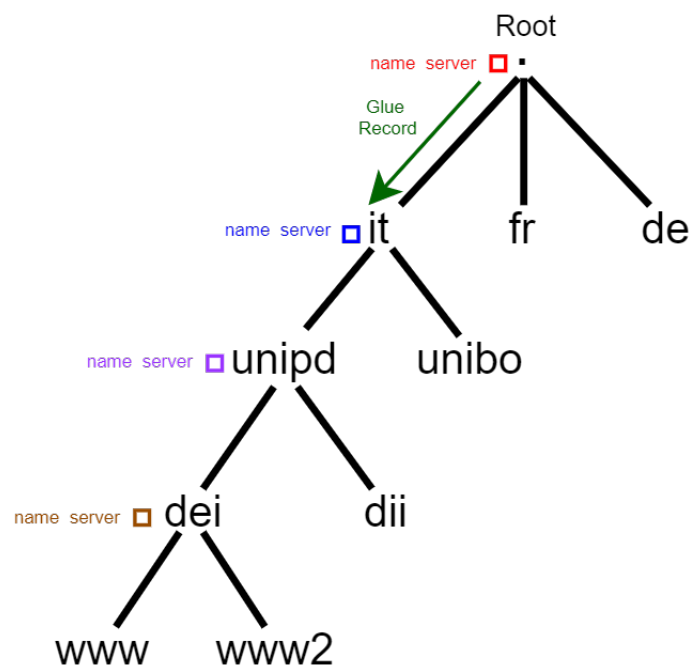


Figure 7.2: DNS structure.

Chapter 8

HTML

The body of an HTTP request, it's often composed from the HTML related page. Each click, of a link inside the web page, generates a new request to the server with GET method.

Chapter 9

CSS

Chapter 10

Javascript

Chapter 11

PHP

Chapter 12

Shell

12.1 Commands

man man		Shows info about man command and lists all the sections of the manual.
strace objFile		Lists all the system calls used in the program.
gcc -o objFile source -v		Lists all the path of libraries and headers used in creation of objFile.
netstat	-t	Lists all the active TCP connections showing domain names.
	-u	Lists all the active UDP connections showing domain names.
	-n	Lists all the active, showing IP and port numbers.
nslookup domain		Shows the IP address related to the domain (E.g. IP of www.google.it)
dig @server name type		DNS lookup utility. server name or IP address of the name server to query name name of the resource record that is to be looked up type type of query is required (ANY, A, MX, SIG, etc.) if no type is specified, A is performed by default
wc [file]		Prints in order newlines, words, and bytes (characters) counts for file if file not specified or equal to -, counts from stdin.

12.2 UNIX Files

/etc/hosts	Local resolution table.
/etc/services	List all the applications with their port and type of protocol (TCP/UDP).
/usr/include/x86_64-linux-gnu/bits/socket.h	List all the protocol type possible for socket.
/usr/include/x86_64-linux-gnu/sys/socket.h	Definition of struct sockaddr and specific ones.

Chapter 13

vim

13.1 .vimrc

In this section there will be shown the file **.vimrc** that can be put in the user home (`~` or `$HOME` or `—`) or in the path `/usr/share/vim/` to change main settings of the program.

```
1 syntax on
2 set number
3 filetype plugin indent on
4 set tabstop=4
5 set shiftwidth=4
6 set expandtab
7 set t_Co=256
```

Listing 13.1: .vimrc

13.2 Shortcuts

Main

Esc	Gets out of the current mode into the “command mode”. All keys are bound of commands
i	“Insert mode” for inserting text.
:	“Last-line mode” where Vim expects you to enter a command.

Navigation keys

h	moves the cursor one character to the left.
j or Ctrl + J	moves the cursor down one line.
k or Ctrl + P	moves the cursor up one line.
l	moves the cursor one character to the right.
0	moves the cursor to the beginning of the line.
\$	moves the cursor to the end of the line.
^	moves the cursor to the first non-empty character of the line
w	move forward one word (next alphanumeric word)
W	move forward one word (delimited by a white space)
5w	move forward five words
b	move backward one word (previous alphanumeric word)

B	move backward one word (delimited by a white space)
5b	move backward five words
G	move to the end of the file
gg	move to the beginning of the file.

Navigate around the document

h	moves the cursor one character to the left.
(jumps to the previous sentence
)	jumps to the next sentence
{	jumps to the previous paragraph
}	jumps to the next paragraph
[[jumps to the previous section
]]	jumps to the next section
 	jump to the end of the previous section
 	jump to the end of the next section

Insert text

h	moves the cursor one character to the left.
a	Insert text after the cursor
A	Insert text at the end of the line
i	Insert text before the cursor
o	Begin a new line below the cursor
O	Begin a new line above the cursor

Special inserts

:r [filename]	Insert the file [filename] below the cursor
:r ![command]	Execute [command] and insert its output below the cursor

Delete text

x	delete character at cursor
dw	delete a word.
d0	delete to the beginning of a line.
d\$	delete to the end of a line.
d)	delete to the end of sentence.
dgg	delete to the beginning of the file.
dG	delete to the end of the file.
dd	delete line
3dd	delete three lines

Simple replace text

r{text}	Replace the character under the cursor with {text}
R	Replace characters instead of inserting them

Copy/Paste text

yy	copy current line into storage buffer
["x]yy	Copy the current lines into register x
p	paste storage buffer after current line
P	paste storage buffer before current line
["x]p	paste from register x after current line
["x]P	paste from register x before current line

Undo/Redo operation

u	undo the last operation.
Ctrl+r	redo the last undo.

Search and Replace keys

/search_text	search document for search_text going forward
?search_text	search document for search_text going backward
n	move to the next instance of the result from the search
N	move to the previous instance of the result
:%s/original/replacement	Search for the first occurrence of the string “original” and replace it with “replacement”
:%s/original/replacement/g	Search and replace all occurrences of the string “original” with “replacement”
:%s/original/replacement/gc	Search for all occurrences of the string “original” but ask for confirmation before replacing them with “replacement”

Bookmarks

m {a-z A-Z}	Set bookmark {a-z A-Z} at the current cursor position
:marks	List all bookmarks
'{a-z A-Z}	Jumps to the bookmark {a-z A-Z}

Select text

v	Enter visual mode per character
V	Enter visual mode per line
Esc	Exit visual mode

Modify selected text

	Switch case
d	delete a word.
c	change
y	yank
>	shift right
<	shift left
!	filter through an external command

Save and quit

:q	Quits Vim but fails when file has been changed
:w	Save the file
:w new_name	Save the file with the new_name filename
:wq	Save the file and quit Vim.
:q!	Quit Vim without saving the changes to the file.
ZZ	Write file, if modified, and quit Vim
ZQ	Same as :q! Quits Vim without writing changes

13.3 Multiple files

- Opening many files in the buffer

```
vim file1 file2
```

Launching this command, you can see only one file at the same time. To jump between the files you can use the following vim commands:

n(ext)	jumps to the next file
prev	jumps to the previous file

- Opening many files in several tabs

```
vim -p file1 file2 file3
```

All files will be opened in tabs instead of hidden buffers. The tab bar is displayed on the top of the editor. You can also open a new tab with file *filename* when you're already in Vim in the normal mode with command:

```
:tabe filename
```

To manage tabs you can use the following vim commands:

:tabn[ext] (command-line command)	Jumps to the next tab
gt (normal mode command)	
:tabp[revious] (command-line command)	Jumps to the previous tab
gT (normal mode command)	
ngT (normal mode command)	Jumps to a specific tab index n= index of tab (starting by 1)
:tabc[lose] (command-line command)	Closes the current tab

- Open multiple files splitting the window

splits the window horizontally

```
vim -o file1 file2
```

You can also split the window horizontally, opening the file *filename*, when you're already in Vim in the normal mode with command:

```
:sp[lit] filename
```

splits the window vertically

```
vim -O file1 file2
```


You can also split the window vertically, opening the file *filename*, when you're already in Vim in the normal mode with command:

```
:vs[plit] filename
```

Management of the windows can be done, staying in the normal mode of Vim, using the following commands:

Ctrl+w <cursor-keys>	Jumps between windows
Ctrl+w [hjk]	
Ctrl+w Ctrl+[hjk]	
Ctrl+w w	Jumps to the next window
Ctrl+w Ctrl+w	
Ctrl+w W	Jumps to the previous window
Ctrl+w p	Jumps to the last accessed window
Ctrl+w Ctrl+p	
Ctrl+w c	Closes the current window
:clo[se]	
Ctrl+w o	Makes the current window the only one and closes all other ones
:on[ly]	

References

- [1] <https://tools.ietf.org/html/rfc1945>.
- [2] <https://tools.ietf.org/html/rfc2616>.
- [3] <https://tools.ietf.org/html/rfc3986>.
- [4] <https://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml>.
- [5] <https://tools.ietf.org/html/rfc2817>.
- [6] <https://tools.ietf.org/html/rfc1034>.