

Metodologie Informatiche Applicate al Turismo

6. - Protocolli a livello applicazione usati su Internet

Paolo Milazzo

Dipartimento di Informatica, Università di Pisa

<http://www.di.unipi.it/~milazzo>

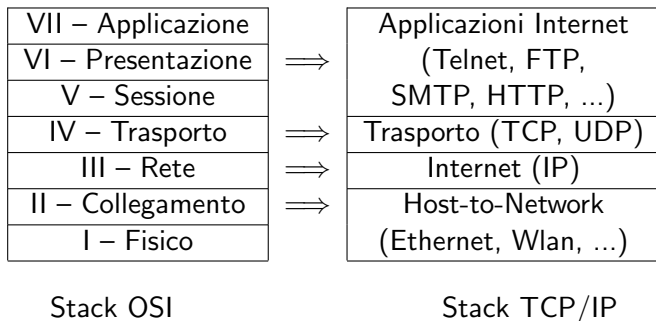
milazzo@di.unipi.it

Corso di Laurea in Scienze del Turismo

A.A. 2013/2014

Lo stack dei protocolli di Internet

L'architettura di Internet si basa sulla struttura a livelli detta **stack TCP/IP** (qui confrontata con il modello OSI):



I protocolli a livello applicazione

Applicazioni Internet (Telnet, FTP, SMTP, HTTP, ...)
Trasporto (TCP, UDP)
Internet (IP)
Link (Ethernet, Wlan, ...)

I protocolli di livello applicazione si dividono in due categorie:

- Protocolli di applicazione vera e propria (e.g. SMTP, HTTP, telnet, FTP, etc...) che forniscono il servizio agli utenti finali
- Protocolli di servizio (e.g. DNS) forniscono servizi alle applicazioni usate dagli utenti

Questi protocolli utilizzano TCP/IP come strumento di comunicazione affidabile tra nodi della rete.

I protocolli a livello applicazione

Vediamo alcuni esempi di protocolli a livello applicazione:

- I protocolli per la posta elettronica (SMTP, POP3 e IMAP)
- Il protocollo per il trasferimento di file FTP

Il protocollo SMTP (1)

- SMTP è un protocollo text-based per lo scambio di messaggi di posta elettronica.
- Una connessione SMTP è composta da una apertura, una sequenza di comandi, e una chiusura.
- Ad ogni comando corrisponde una risposta composta da un codice numerico (per le applicazioni) e una stringa leggibile (per gli umani).
 - ▶ MAIL FROM:<Smith@alpha.com>
 - ▶ 250 OK
 - ▶ RCPT TO:<Green@alpha.com>
 - ▶ 550 No such user here

- Esempio di scambio di messaggi tra un client (programma di posta elettronica, in blu) e il server SMTP del mittente (in rosso)
 - ▶ La spedizione di un messaggio SMTP avviene attraverso l'identificazione del mittente (MAIL FROM), del/dei destinatari (RCPT TO), e del messaggio da trasmettere (DATA):
 - ▶ 220 alpha.com Simple Mail Transfer Service Ready
 - ▶ HELO beta.com
 - ▶ 250 alpha.com says: Nice to meet you beta.com
 - ▶ MAIL FROM:<Smith@alpha.com>
 - ▶ 250 OK
 - ▶ RCPT TO:<Green@alpha.com>
 - ▶ 550 No such user here
 - ▶ RCPT TO:<Brown@alpha.com>
 - ▶ 250 OK
 - ▶ DATA
 - ▶ 354 Start mail input; end with <CRLF>.<CRLF>
 - ▶ Blah blah blah....
 - ▶ etc. etc. etc.
 - ▶ .
 - ▶ 250 OK
 - ▶ QUIT
 - ▶ 250 alpha.com Service closing transmission channel

Il protocollo SMTP (3)

Il protocollo SMTP ha alcune forti limitazioni:

- Assume che si trasmettano messaggi di testo;
- La lunghezza massima di un messaggio è di 1MB;
- I caratteri accettati sono solo ASCII a 7 bit (lettere, numeri, semplici simboli,....);
- Ogni messaggio deve avere un CRFL (un carattere “a capo”) ogni 1000 caratteri o meno (alcune antiche implementazioni lo aggiungevano automaticamente se non lo trovavano).

Questi limiti impediscono la trasmissione di documenti binari:

- Un file binario (eseguibile, pdf, immagine,) usa tutti i 256 tipi di byte;
- Un file binario può facilmente essere più lungo di 1MB;
- In un file binario la sequenza CRFL è una sequenza come tutte le altre, e può esserci o mancare senza vincoli. Introdurla artificialmente può corrompere il file.

Estensioni del protocollo SMTP: MIME (1)

I limiti di SMTP sono comunemente superati attraverso l'uso delle estensioni MIME (Multipurpose Internet Mail Extensions)

MIME ridefinisce il formato del corpo di un messaggio SMTP per permettere (tra le altre cose):

- Messaggi di testo in altri set di caratteri al posto di ASCII;
- Un insieme estensibile di formati (tipi) per messaggi non testuali;

Il problema di determinare il “tipo” di un contenuto non testuale è comune a molti protocolli a livello applicazione.

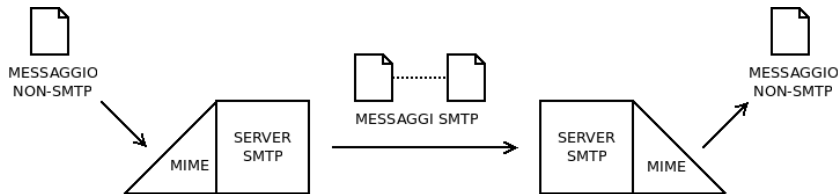
- MIME viene utilizzato anche in altri contesti (e.g. WWW)

Estensioni del protocollo SMTP: MIME (2)

Come funziona MIME:

- 1 Il messaggio non compatibile con SMTP viene trasformato in uno o più messaggi SMTP da un preprocessore al server SMTP;
- 2 All'arrivo, il (o i) messaggi SMTP vengono decodificati e riaccorpati a formare il messaggio originale.

Ancora oggi, quelli che viaggiano su canali SMTP sono puri messaggi SMTP con gli stessi limiti di allora.



Estensioni del protocollo SMTP: MIME (3)

- La specifica di MIME include aspetti riguardanti la codifica dei caratteri non in formato ASCII, la frammentazione di messaggi non SMTP in SMTP e il successivo riaccorpamento, ecc...
- L'aspetto di MIME che trova applicazione in altri contesti (e.g. WWW) è il concetto di Content-type (o tipo MIME)

Un **tipo MIME** definisce uno specifico tipo per un contenuto associandogli un identificatore di “media type” e un identificatore di “subtype” (rappresentati così: `mediatype/subtype`).

- Il media type dichiara il **tipo** generale del dato trasmesso
- Il subtype dichiara il **formato** specifico del dato trasmesso

Ad esempio:

- il tipo MIME `text/plain` identifica testo semplice;
- il tipo MIME `text/html` identifica testo html;
- il tipo MIME `image/gif` identifica immagini in formato GIF;
- il tipo MIME `video/mpeg` identifica video in formato MPEG.

Estensioni del protocollo SMTP: MIME (4)

I client di posta elettronica (e similmente le altre applicazioni – e.g. i browser – che usano tipi MIME) possono essere configurati per inoltrare i contenuti di un certo tipo (o sottotipo) ad altre applicazioni

- Ad esempio:

- ▶ inoltrare tutti i contenuti di tipo text/pdf al programma Adobe Reader
- ▶ inoltrare tutti i contenuti di tipo image ad un visualizzatore di immagini, indipendentemente dal sottotipo

E' possibile strutturare i messaggi di posta elettronica in parti (messaggi multipart o con allegati) aventi tipi MIME diversi.

I tipi MIME sono in continua evoluzione:

- La lista dei tipi MIME è costantemente aggiornata dall'Internet Assigned Numbers Authority (IANA)

Il protocollo POP3

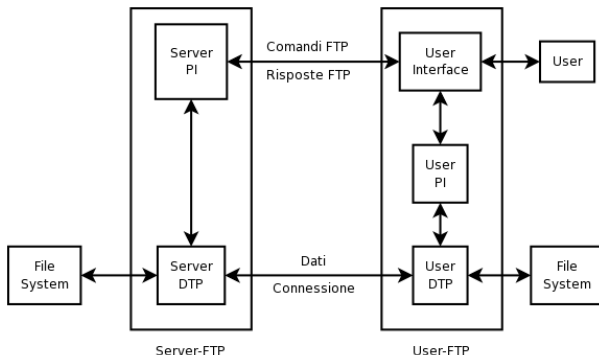
- SMTP si disinteressa di come il ricevente di un messaggio acceda alla sua mailbox. Si supposeva all'epoca che tutti avessero accesso via file system alla directory con le mailbox;
- POP3 permette ad una applicazione utente di accedere alla mailbox posta su un altro sistema;
- POP3 non permette manipolazioni complesse sulla mailbox, ma soltanto la possibilità di scaricare e cancellare mail. Per operazioni più complesse si usa IMAP4.

Il protocollo IMAP4

- Il protocollo IMAP4 consente un controllo più sofisticato della propria mailbox, anche se posta su un server remoto;
- Le caratteristiche di IMAP4 più rilevanti per l'utilizzo quotidiano sono:
 - ▶ La possibilità di consultare la posta elettronica senza scaricarla dal server (favorisce l'accesso da diverse postazioni);
 - ▶ La possibilità di creare directory di messaggi sul server (da usare anche per filtrare i messaggi sul server – e.g. procmail)
- IMAP4 consente operazioni di creazione, cancellazione e cambio di nome a directory all'interno della mailbox; verifica di nuovi messaggi; cancellazione di messaggi; ricerca per contenuto e attributi; scaricamento selettivo di attributi, parti e messaggi.

Protocolli per il trasferimento di file (1)

- Il File Transfer Protocol (FTP) è un protocollo per la trasmissione di file;
- Utilizza due connessioni TCP separate: una per i comandi (detto “canale comandi”) e una per i file trasmessi (detto “canale dati”).



dove PI (protocol interpreter) è l'interprete dei comandi e DTP (data transfer process) è il processo di trasferimento dati.

Protocolli per il trasferimento di file (2)

- Il canale comandi rimane aperto per l'intera sessione utente
- Un nuovo canale dati (connessione TCP) viene utilizzato per ogni file trasmesso (favorisce la trasmissione contemporanea di più file)

Tra le funzioni offerte da un server FTP abbiamo:

- Download/upload di file;
- Resume di trasferimenti interrotti;
- Rimozione e rinomina di un file;
- Creazione di directory;
- Navigazione tra directory.

Protocolli per il trasferimento di file (3)

- FTP offre un meccanismo di autenticazione in chiaro (non crittografata) degli accessi;
- E' possibile accedere ai contenuti di un server FTP in modalità "anonima" (senza password). In questo caso tipicamente il client può accedere ai dati in modalità di "sola lettura";
- Una versione "sicura" di FTP basato su crittografia SSL/TLS è FTPS. Questa è ormai la versione di FTP più utilizzata.

Esempi di programmi specifici per la trasmissione di file tramite FTP (per Windows) sono:

- FileZilla
- WinSCP

FTP(S) è usato (spesso in maniera trasparente) anche all'interno di browser web (link per download di file nelle pagine web).

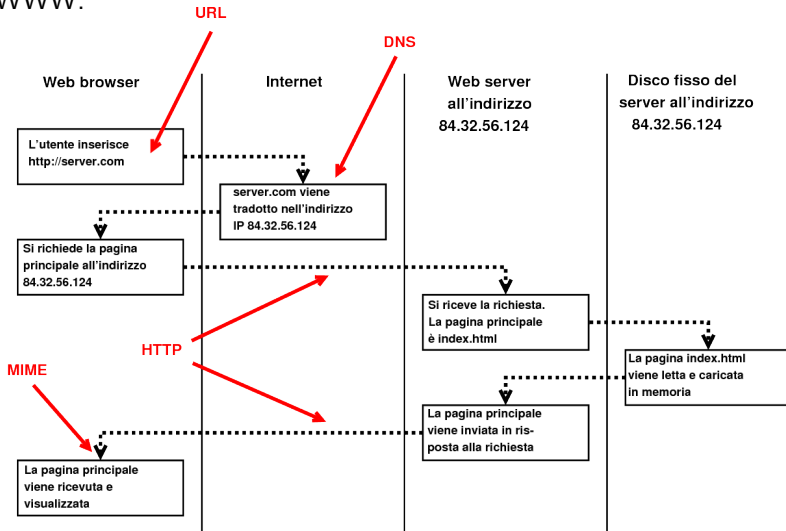
Internet e il WWW (1)

Il WWW usa solo alcuni dei servizi che l'infrastruttura di Internet mette a disposizione. I principali sono:

- L'indirizzamento delle risorse: tramite URL (e.g. `http://www.di.unipi.it`)
- La risoluzione degli indirizzi: traduzione di indirizzi testuali tipo `www.di.unipi.it` nei corrispondenti indirizzi IP tramite DNS;
- La gestione dei tipi di contenuti: tramite MIME;
- Il trasferimento dei contenuti: tramite il protocollo di comunicazione HTTP.

Internet e il WWW (2)

Vediamo dove DNS, MIME e HTTP vengono utilizzati in un tipico utilizzo del WWW:



Indirizzare le risorse: URL (1)

- L'indirizzo di una risorsa su Internet può essere rappresentata da un Uniform Resource Locator (URL);
- URL è una sintassi standard per la definizione di indirizzi sulla rete;

Un URL consiste di due parti (separate da :))

- uno schema (in inglese scheme)
- e una parte schema-specifica (in inglese scheme-specific part)

Indirizzare le risorse: URL (2)

L'insieme degli schemi validi è mantenuto dall'Internet Assigned Numbers Authority (IANA), e include i seguenti:

- ftp per gli indirizzi di risorse accessibili tramite FTP
- http per gli indirizzi di risorse accessibili tramite HTTP
- https per gli indirizzi di risorse accessibili tramite HTTP sicuro
- mailto per gli indirizzi di posta elettronica

Le parti schema-specifiche possono avere la seguente sintassi (in cui le parentesi quadre rappresentano parti opzionali):

```
[ // ] [ user [ : password ] @ ] host [ : port ] / url-path
```

Esempi:

- http://www.di.unipi.it/~milazzo/
- mailto:milazzo@di.unipi.it
- ftp://milazzo@ftp.di.unipi.it:2045/public/

Nomi e indirizzi: il DNS (1)

- I computer su Internet sono identificati dal loro indirizzo IP;
- Nel contesto del WWW è molto più comune utilizzare nomi di dominio (e.g. `www.di.unipi.it`) anzichè indirizzi IP;
- Il Domain Name System (DNS) è un servizio di naming globale che mappa nomi di dominio in indirizzi IP.

L'utilizzo di nomi di dominio anzichè indirizzi IP offre due vantaggi:

- I nomi di dominio sono più facilmente leggibili e utilizzabili dagli utenti degli indirizzi IP;
- I nomi di dominio offrono un livello di astrazione: se si riorganizza una rete cambiando alcuni indirizzi IP è sufficiente aggiornare il DNS per rendere il cambiamento invisibile agli utenti.

Nomi e indirizzi: il DNS (2)

Il DNS può essere visto come un elenco telefonico globale in cui i nomi anziché essere in ordine alfabetico hanno una struttura gerarchica

- I domini di primo livello sono sigle di nazioni (e.g. `it,uk,fr,...`) o domini generici (e.g. `com,edu,org,...`);
- I domini di livello inferiore sono indicati da destra a sinistra separati da punti: e.g. in `www.di.unipi.it`
 - ▶ `it` è il dominio di primo livello;
 - ▶ `unipi` è il dominio di secondo livello attribuito all'Università di Pisa
 - ▶ `di` è il dominio di terzo livello attribuito al Dipartimento di Informatica;
 - ▶ `www` è il nome convenzionale del computer che ospita il web server in un dominio.

Nomi e indirizzi: il DNS (3)

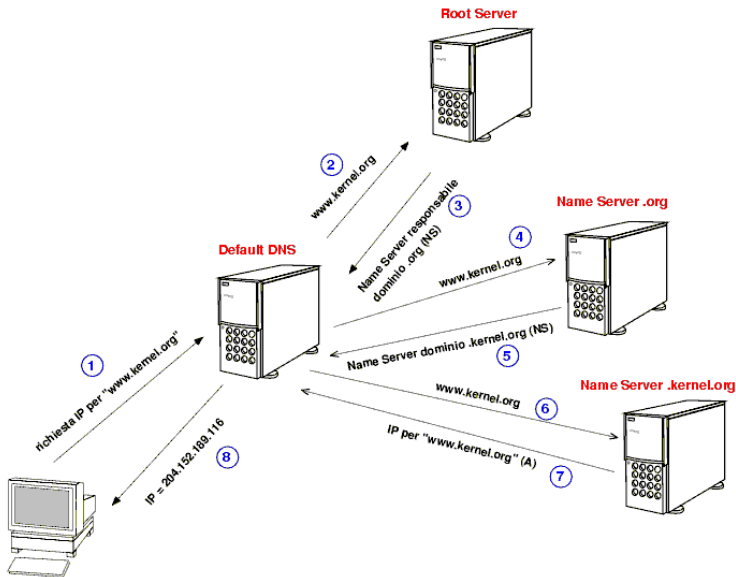
La risoluzione di un nome di dominio tramite DNS è basato sull'iter seguente:

- Ogni computer collegato a Internet deve avere un server DNS di riferimento;
- Quando un computer deve risolvere un nome lo inoltra al proprio server DNS di riferimento;
- Se il server DNS non conosce l'indirizzo IP corrispondente al nome richiesto inoltra la richiesta a un root server (un server DNS che conosce gli indirizzi dei server DNS responsabili dei domini di primo livello);
- Il server DNS invia la richiesta al server responsabile del dominio di primo livello dell'indirizzo richiesto;
- Tale server fornisce la risposta richiesta, o fornisce l'indirizzo del server DNS del dominio di secondo livello di interesse;
- Il procedimento continua finchè la richiesta non riceve risposta.

I server DNS possono fare uso di tecniche di caching (memorizzazione) delle informazioni scambiate.

Nomi e indirizzi: il DNS (4)

Un esempio:



Descrivere il tipo di un contenuto: MIME

- Come nel caso della posta elettronica, anche per il WWW si usa MIME per distinguere tra diversi tipi di contenuto;
- In questo caso il browser si occuperà di redirigere documenti che non sa visualizzare ad applicazioni specifiche selezionate a seconda del tipo MIME ricevuto (tramite HTTP, come vedremo).

Il protocollo HTTP (1)

L'HyperText Transfer Protocol (HTTP 1.1) è un protocollo di comunicazione client-server:

- per la trasmissione di contenuti a livello applicazione
- basato su un servizio di trasporto affidabile e orientato alle connessioni (come TCP)

Il protocollo HTTP, oltre alle figure del client e del server, prevede la presenza di intermediari, tra cui i proxy.

Un proxy:

- è un programma che riceve richieste HTTP dai client (opportunamente configurati) e le inoltra ai server;
- può applicare tecniche di caching dei contenuti ricevuti dai server per soddisfare autonomamente successive richieste dei client;
- può applicare filtri ai contenuti ricevuti (e.g. parental control).

Il protocollo HTTP (2)

Il protocollo HTTP:

- Normalmente usa una nuova connessione TCP per ogni richiesta (e.g. ogni volta che si segue su un link su una pagina web), anche se esiste la possibilità di stabilire connessioni permanenti (per gestire in maniera più efficiente una sequenza di richieste);
- E' **stateless** (senza stato): il risultato di una richiesta non può dipendere da richieste precedenti;
- Costringe lo sviluppatore di contenuti web a trovare metodi alternativi per memorizzare una forma di stato (e.g. cookies).

Un messaggio di richiesta di un contenuto è costituito da:

- Una riga di richiesta che contiene un “metodo”, un URL e l'indicazione della versione di HTTP utilizzata;
- Una sezione di header (contenente informazioni varie)
- Il corpo del messaggio

Il protocollo HTTP (3)

Il metodo può essere uno dei seguenti:

- GET: per richiedere al server l'invio della risorsa indicata nell'URI;
- POST: per inviare informazioni al server;
- HEAD: simile a GET, richiede al server l'invio di informazioni sulla risorsa indicata nell'URI (non la risorsa stessa);
- PUT: per richiedere al server di sostituire una risorsa specificata con un'altra;
- DELETE: per richiedere al server di cancellare una risorsa;
- TRACE: per ricostruire la sequenza di intermediari tra il client e il server;
- OPTIONS: per ottenere informazioni sulle opzioni di comunicazione.

I metodi più comuni (e i soli che approfondiremo) sono GET e POST.

Il protocollo HTTP (4)

Il metodo GET:

- è il metodo più comune, che viene utilizzato ogni volta, ad esempio, che si segue un link in una pagina web o che si inserisce un indirizzo in un browser;
- serve per richiedere al server l'invio di una risorsa indicata nell'URI
- nell'header del messaggio:
 - ▶ può porre condizioni sul formato (e.g. tipo MIME) o le caratteristiche della risorsa (dimensioni, data ultimo aggiornamento, ecc..)
 - ▶ può includere informazioni aggiuntive (e.g. browser utilizzato)

Il protocollo HTTP (5)

Esempio di richiesta che usa il metodo GET:

```
GET /beta.html HTTP/1.1
Referer: http://www.alpha.com/alpha.html
User-Agent: Mozilla/4.61 (Macintosh; I; PPC)
Host: www.alpha.com:80
Accept: image/gif, image/jpeg, image/png, */*
Accept-encoding: gzip
Accept-language: en
Accept-charset: iso-8859-1, utf-8
```

L'URL (o URI) della risorsa può essere ottenuto combinando Host con il seguito di GET.

Il protocollo HTTP (6)

Il metodo POST:

- Viene usato per trasmettere delle informazioni dal client al server;
- Esempio tipico di informazione trasmessa: i dati inseriti dall'utente in un form in una pagina web;
- I dati trasmessi vengono inseriti nel corpo del messaggio;
- I dati vengono passati alla risorsa (e.g. applicazione) specificata dall'URL.

Il protocollo HTTP (7)

Un messaggio di risposta a una richiesta di un contenuto è costituito da:

- Una riga di stato (status-line) che contiene un codice e una breve descrizione dell'esito della richiesta;
- Una sezione di header
- Il corpo del messaggio di risposta

Le righe di stato più comuni includono:

200 Ok	metodo eseguito con successo
201 Created	metodo eseguito con successo con relativa creazione di nuova risorsa (e.g. PUT)
400 Bad request	errore sintattico nella richiesta
403 Forbidden	richiesta non autorizzabile
404 Not found	URL errato (molto comune)
500 Internal server error	Errore interno (comune con siti web dinamici)

L'header di una risposta a GET include anche il tipo MIME del dato trasmesso

Il protocollo HTTP (8)

Un esempio di richiesta (semplificata):

```
GET /beta.html HTTP/1.1  
Host: www.alpha.com:80
```

E relativa risposta:

```
HTTP/1.1 200 OK  
Date: Mon, 28 Jun 2004 10:47:31 GMT  
Server: Apache/1.3.29 (Unix) PHP/4.3.4  
Last-Modified: Mon, 12 Jun 2004 11:32:12 GMT  
Content-Language: it  
Content-Type: text/html; charset=utf-8
```

Il protocollo HTTP (9)

- HTTP trasmette tutti i dati in chiaro;
- Esiste una versione di HTTP per comunicazioni che richiedono crittografia: HTTPS
- HTTPS:
 - ▶ utilizza TCP e SSL (crittografia) per trasmettere i soliti messaggi HTTP;
 - ▶ usa lo schema `https:` al posto di `http:` negli indirizzi.