



TCP/IP in pillole – Parte II

DARIO DE JUDICIBUS

Il funzionamento di Internet si basa sull'identificazione univoca di ogni sistema attraverso indirizzi IP e nomi simbolici, che permettono di localizzare e raggiungere qualsiasi host all'interno di un insieme di reti interconnesse. I router instradano i pacchetti tra reti diverse utilizzando gli indirizzi di rete, senza conoscere il destinatario finale, garantendo così scalabilità e semplicità del sistema. Gli indirizzi IP, strutturati in classi e suddivisi in parti che identificano rete e host, consentono una gestione gerarchica e delegata degli spazi di indirizzamento. Per facilitarne l'uso, ai numeri vengono associati nomi attraverso il Domain Name System, un sistema distribuito e gerarchico che risolve i nomi in indirizzi. L'insieme di questi meccanismi rende possibile una comunicazione globale, flessibile e indipendente dalle tecnologie delle singole reti.

Nella prima parte di questo corso abbiamo introdotto i concetti base dell'architettura TCP/IP. Abbiamo parlato di protocolli e detto che un protocollo è una serie di regole per comporre dei messaggi e per far sì che essi possano essere scambiati tra due macchine. Abbiamo sottolineato che una internet non è una rete ma un sistema di reti interconnesse grazie all'utilizzo di opportuni sistemi chiamati *gateway* ed una serie di protocolli che formano appunto il TCP/IP. Abbiamo infine parlato di come è possibile trasferire dati nella rete, ed abbiamo introdotto il concetto di trasmissione a pacchetti.

Torniamo proprio al sistema a pacchetti. Per trasferire dati da un sistema ad un altro ogni sistema ha un nome unico ben definito. Non esistono cioè due sistemi con lo stesso nome, anche se in reti diverse, indipendentemente da quale sia il nome locale di un sistema nella sua rete di appartenenza (*Figura 1*). All'interno di ciascuna rete, i vari computer usano la tecnologia hardware e software specifica di quella rete. Tuttavia, grazie a questo strato intermedio di software, le varie applicazioni hanno una visione unica e globale del sistema interconnesso di reti, detto appunto **internet**. Notate la "i" minuscola. Il concetto di internet è infatti quello appena descritto. Viceversa **Internet** con la "I" maiuscola, identifica quel sistema di reti, basato sull'architettura internet, che viene detto anche **Connected Internet**.

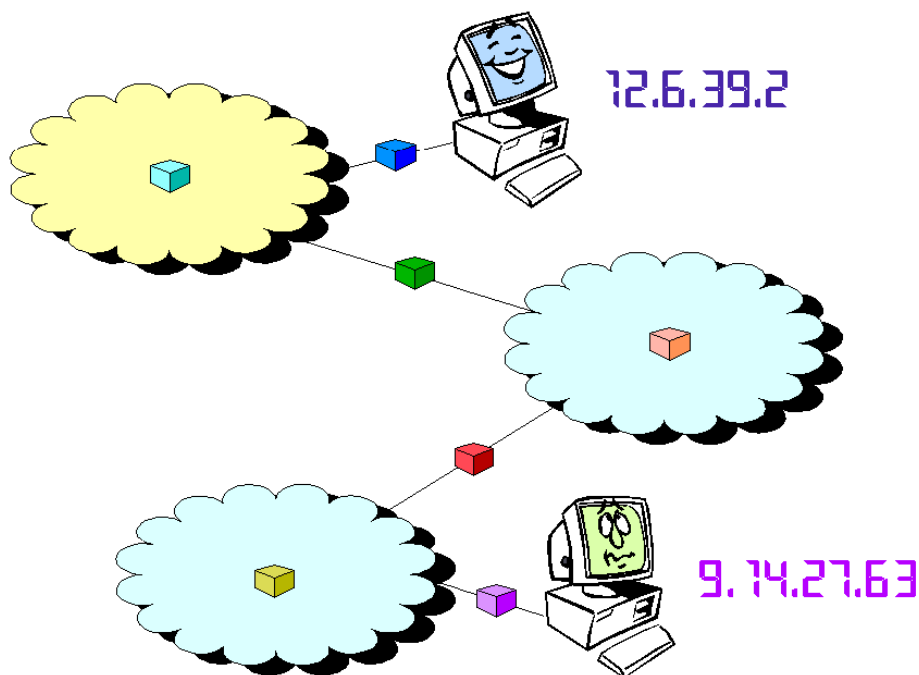


Figura 1 - Indirizzi IP

La connessione tra due reti avviene attraverso macchine opportune che sono collegate fisicamente ad entrambe le reti, ed hanno la responsabilità di far passare i vari pacchetti da una rete all'altra e viceversa. Tali macchine sono dette *internet gateway*, od anche *IP router*. Sono loro il vero elemento portante di una internet. Ogni router non solo deve sapere che determinati pacchetti vanno passati da una rete ad un'altra, ma deve passare dall'altra parte anche pacchetti destinati ad ulteriori reti connesse attraverso altri router. Essi però ragionano solo in termini di reti, non di destinazione finale. Ad un router non interessa chi è effettivamente il destinatario di un pacchetto, ma solo a quale rete appartiene. Questo semplifica molto l'implementazione di un router. Alla base del meccanismo dei router c'è l'indirizzo IP, od **IP address**.

Ogni cosa che conosciamo ha un nome. Cane, casa, auto, e via dicendo. Se ci interessa specificare meglio ciò di cui stiamo parlando, possiamo assegnare un nome anche ad un sottogruppo di cose. Così abbiamo che i cani bassotti sono alquanto diversi dai San Bernardo, una catapecchia non è certo una villa, ed una Ferrari costa un po' più di una Cinquecento. Se poi dobbiamo identificare una cosa in modo chiaro ed univoco, è necessario assegnarle un nome che solo quella cosa ha. Già un nome come Mario Rossi non va bene, perché non è unico, e comunque, anche se scegliessimo oggi un nome veramente strano ed originale, non avremmo la garanzia in futuro di non ritrovarci con un caso di omonimia. Ecco allora le targhe per le automobili, i codici fiscali per le persone, i numeri di telefono, e via dicendo. Ognuno di questi nomi ha tre caratteristiche. La prima è che esiste un organo competente centrale che li assegna, proprio per garantirne l'univocità. La seconda, è che hanno una struttura a sottogruppi. Esistono cioè degli elementi che garantiscono l'univocità ad un certo livello, all'interno del quale esiste una

certa libertà di scelta, e così via, livello dopo livello. Per esempio, il codice fiscale viene costruito in modo che un uomo ed una donna non possano mai avere lo stesso codice, anche se fossero nati lo stesso giorno, nella stessa città e si chiamassero nello stesso modo. Similmente, i numeri di telefono di due città diverse si distinguono per il prefisso e se queste si trovano anche in stati diversi, per il prefisso internazionale.

Affinché internet possa rappresentare un sistema *universale* di comunicazione, permetta cioè di far comunicare qualunque macchina connessa ad una delle sue reti con una qualsivoglia altra macchina connessa alla stessa od ad un'altra rete, è necessario fornire ogni macchina di un nome unico a livello globale. Internet fornisce ogni sistema di un **nome**, che identifica il sistema stesso, di un **indirizzo**, che mi dice dove si trova il sistema, e di un **cammino**, che mi dice come raggiungere il sistema. Ogni macchina connessa ad una rete è detta **host**, nella terminologia internet. Lo stesso termine ha significati differenti in altri contesti informatici, come ad esempio in quello client/server, o nel caso di mainframe. Attenzione a non fare confusione quindi. In internet un host può essere anche un vecchio 8088 con 640K di RAM e 10M di disco fisso.

L'indirizzo, od *IP address*, è un campo composto da 32 bit. I primi bit permettono di distinguere 5 forme standard (*Figura 2*) identificate da una lettera dell'alfabeto, che sono dette **classi**.

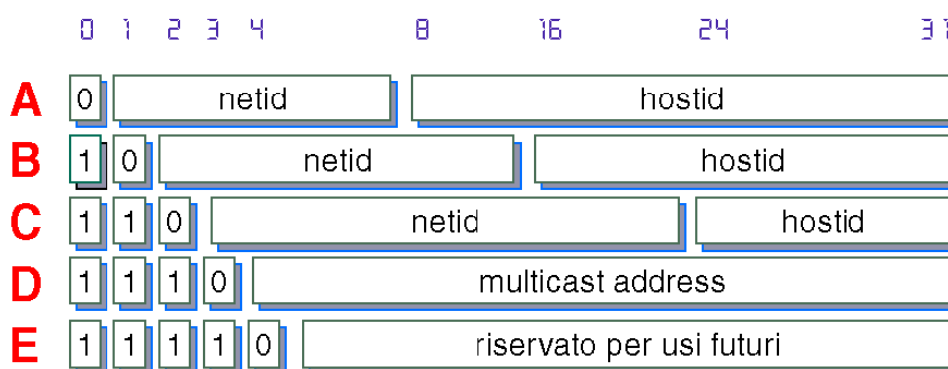


Figura 2 - Classi

Come si può vedere in figura, le prime tre classi dell'IP address contengono sia l'indirizzo di una rete (*netid*), sia quello di una macchina nella stessa (*hostid*). In realtà l'indirizzo non identifica necessariamente una macchina, ma una connessione alla rete. Ad esempio, un

router ha almeno due indirizzi, avendo connessioni ad almeno due reti. Questo in quanto un router appartiene ad entrambe le reti, e quindi sono necessari due indirizzi dato che un IP address ha posto per un solo indirizzo di rete. Se l'indirizzo dell'host è **0**, allora l'IP address si riferisce alla rete stessa. Se viceversa tutti i bit riservati all'indirizzo dell'host sono ad **1**, allora l'indirizzo viene utilizzato per identificare **tutti** gli host della rete (*broadcasting*). Uno speciale indirizzo formato da 32 bit posti ad uno è chiamato *local network broadcast address* e serve solo in casi molto particolari. Il concetto di broadcasting è quello della diffusione a tutto raggio, un po' come fa una emittente radiofonica. In generale internet interpreta i campi formati da tutti uno come *all*, cioè "tutti", mentre quelli formati da tutti zero come *this*, cioè "questo", "qui". Questo per quanto riguarda le classi A, B e C. La classe D è usata per un particolare tipo di distribuzione dei dati detto *multicasting*. La classe E è riservata ad usi futuri. Dato che specificare ogni singolo bit di un indirizzo IP sarebbe alquanto poco pratico e di scarsa leggibilità, la convenzione è quella di leggere ogni otteetto, cioè ogni gruppo di 8 bit, come un intero, e di separare i quattro ottetti con un punto (*Figura 3*). Oltre ad i casi speciali già descritti, l'indirizzo di classe A 127.0.0.0 è riservato per un particolare processo di test che rimanda indietro i dati al mittente senza propagarli nella rete.



Figura 3 – Struttura dell'indirizzo

Uno dei vantaggi di questo schema è la possibilità da parte dell'organismo centrale che assegna gli indirizzi (*Network Information Center*) di delegare ai responsabili delle singole reti l'assegnazione di una parte dell'indirizzo all'interno della rete stessa. La cosa avviene un poco come con i numeri di telefono. A livello internazionale ogni stato ha il suo prefisso internazionale. Ad esempio, per l'Italia, è **39**. All'interno ogni stato divide il paese in aree geografiche a cui assegna un ulteriore codice. Ad esempio, Roma è identificata dal **6**,

Milano dal **2**, Firenze da **55**, e così via. All'interno poi della provincia o della città possono essere definite ulteriormente sotto-aree a cui si assegnano due, tre o quattro cifre. Per esempio **529** oppure **7054**. Infine ogni telefono in tali aree avrà il suo numero. Così, se Mr.Smith deve chiamare dagli Stati Uniti il signor Pinco Pallino abitante all'EUR, a Roma, comporrà ad esempio il numero **011.39.6.529.4467**. In questo caso lo **011** serve per uscire dagli USA, un po' come il nostro **00**.

Analogamente in internet i numeri di classe C sono assegnati alle piccole reti, quelle cioè con meno di 256 host, quelli di classe B alle reti con al massimo 65536 host, e quelli di classe A alle reti con oltre 16 milioni di host. Ogni rete decide poi come suddividere gli indirizzi che gli sono stati riservati al suo interno come meglio crede. Ovviamente, una internet privata non ha la necessità di seguire queste regole, né ad utilizzare indirizzi assegnati dal NIC, ma il non farlo potrebbe impedire in futuro la connessione alla TCP/IP Internet.

Dato che l'indirizzo può essere a volte abbastanza ostico da ricordare, è possibile associare ad ogni host anche un nome, che può essere utilizzato come mnemonico per un IP address, e la cui risoluzione è responsabilità di particolari macchine chiamate *name server*. In realtà il name server è un programma software che può girare in qualunque macchina connessa alla rete, e che mantiene l'associazione tra nomi ed indirizzi IP, fornendo tali corrispondenze quando richiesto da un altro programma chiamato *name resolver*. Di fatto, si preferisce far girare il name server su una macchina dedicata, che prende anch'essa, a questo punto, il nome di *name server*. Potete pensare al name server come ad una agenda telefonica elettronica, che contiene una lista parziale di nomi e numeri telefonici. In internet infatti, non esiste un singolo elenco telefonico, ma tanti name server che cooperano per fornire quello che è un vero e proprio elenco distribuito. In realtà il sistema funziona in modo gerarchico, un po' come se una certa agenda contenesse solo i prefissi internazionali ed il puntatore alle agende di ogni singolo stato, le quali a loro volta contengono i prefissi regionali ed i puntatori agli elenchi regionali, e così via, fino ad arrivare all'agenda che contiene solo le estensioni telefoniche di un singolo edificio.

I nomi Internet sono basati su una serie di regole dette *Domain Name System* (DNS), che si basa appunto su uno schema gerarchico in cui il nome è suddiviso in varie *label* separate fra loro da punti. Ad esempio, **vnet.ibm.com**. Ogni suffisso è a sua volta un dominio. Quindi, nel nostro esempio, **ibm.com** è un dominio di secondo livello, mentre **com** è un dominio di terzo livello. I domini ufficiali riconosciuti dal NIC al livello più elevato sono riportati nella tabellina azzurra. Una rete può richiedere di essere registrata come categoria, oppure usando il dominio geografico.

Ad esempio, l'Italia ha come dominio base **it**. Supponiamo che il governo decida di costruire un insieme di reti cittadine interconnesse fra loro e connesse ad Internet. Si può pensare di assegnare ad ogni provincia un dominio **xx.it** dove xx è la sigla postale della provincia. Ad esempio, Firenze avrebbe come dominio **fi.it**. L'università di Firenze potrebbe registrare la sue rete come **usf.edu**, ed in tal caso sarebbe direttamente il NIC a dover dare l'autorizzazione per tale nome, essendo il dominio edu sotto responsabilità dell'organismo

centrale di controllo, oppure potrebbe decidere di far parte del dominio cittadino, come **usf.fi.it**, e quindi potrebbe richiedere il permesso di registrare tale nome direttamente all'amministratore del dominio di Firenze. A questo punto, se il dipartimento di Fisica di Arcetri vuole registrare un proprio dominio, deve chiederlo solo all'Università stessa, ricevendo così, ad esempio, **arcetri.usf.fi.it** oppure **fisica.usf.fi.it**.

Esiste una piccola complicazione. Ogni oggetto connesso alla rete ha un *tipo*. Oggetti di tipo diverso possono avere lo stesso nome. Per cui, per poter risolvere un nome ed ottenere indietro l'indirizzo IP, è necessario anche specificare il tipo di oggetto: macchina, utente, casella postale, e via dicendo. Dal solo nome non è possibile evincere il tipo di oggetto.

<i>codice org</i>	Organizzazione. Ad esempio:
COM	Organizzazioni commerciali
EDU	Istituzioni scolastiche (università, ricerca)
GOV	Istituzioni governative
MIL	Istituzioni militari
NET	Grandi reti
ORG	Altre organizzazioni
ARPA	Dominio ARPANET temporaneo (obsoleto)
INT	Organizzazioni internazionali
<i>codice stato</i>	Stato. Ad esempio:
IT	Italia
DE	Germania
FI	Finlandia
SE	Svezia

Il DNS definisce anche come associare i nomi agli indirizzi IP, e come ottenere quest'ultimi dal nome. In realtà lo schema è ancora più generale di quanto può sembrare, in quanto permette di estendere la sintassi del nome per usi specifici, sfruttando anche il concetto di tipo. Ad esempio, nel caso di una casella postale (tipo **MX**), il nome sarà del tipo **utente@dominio**. Ad esempio **ddj@vnet.ibm.com**.

Questo è tutto per il momento. Ovviamente non sono stati approfonditi molti concetti, o sono stati riportati in modo forse eccessivamente semplificato alcuni elementi base dell'architettura internet. D'altra parte, questo breve corso sul TCP/IP si sviluppa su due articoli, anche per evitare di diventare troppo pesante, e dobbiamo ancora parlare più in dettaglio dei singoli protocolli e delle loro caratteristiche. Lo faremo il mese prossimo. Arrivederci.