

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Informatica di base 2025/2026

Cosa succede quando si clicca su un link: Trasferire informazioni su Internet

Ivan Heibi

Dipartimento di Filologia Classica e Italianistica (FICLIT)

Ivan.heibi2@unibo.it

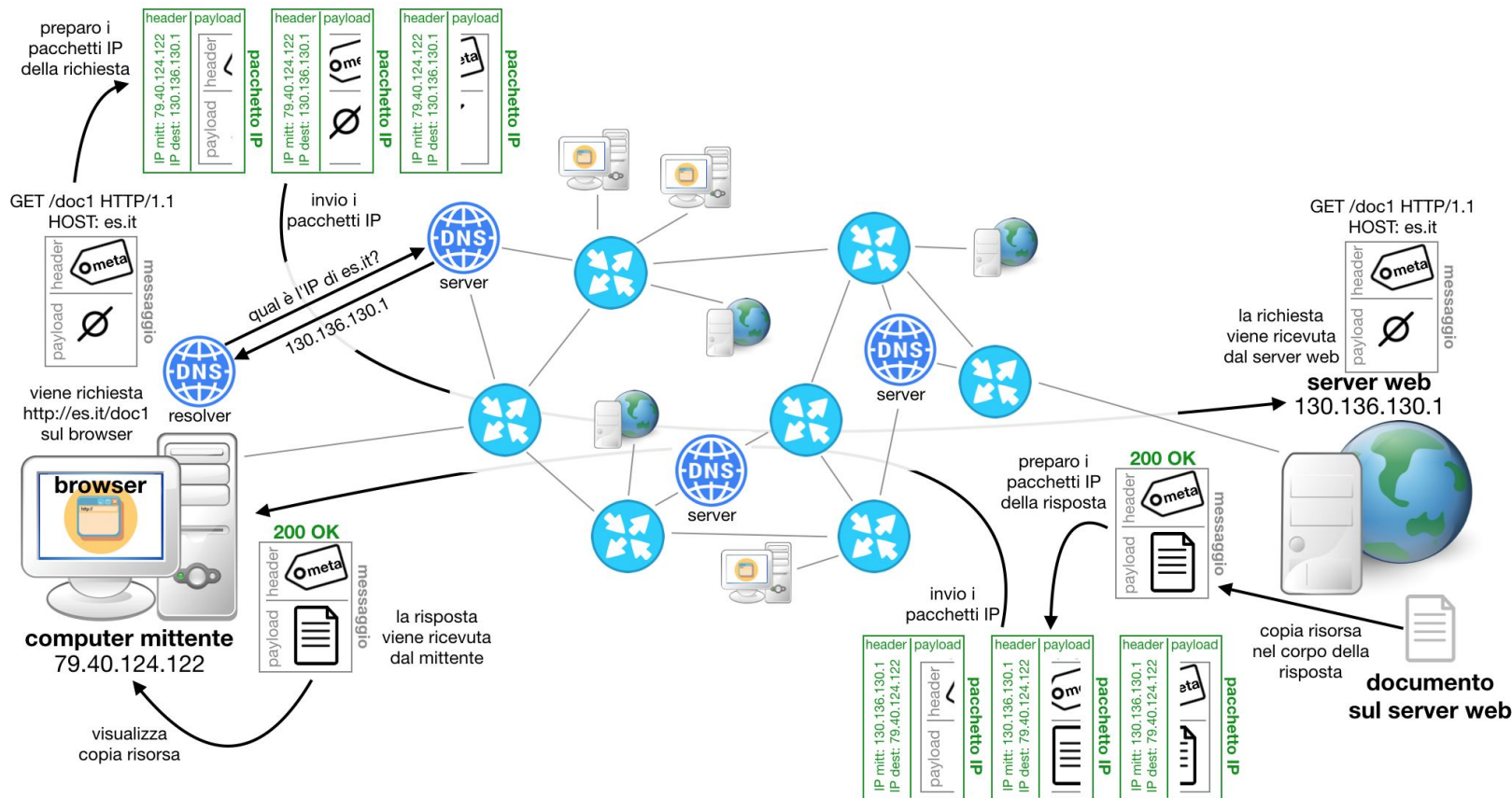
<https://www.unibo.it/sitoweb/ivan.heibi2>

Domanda

Cosa succede quando si clicca su un collegamento ipertestuale (o link) di una pagina Web?

Per rispondere pienamente a questa domanda, è necessario avere delle conoscenze di base relative sia sul come l'informazione viene trasmessa digitalmente, sia sulle componenti che compongono Internet e il Web

Riassunto di quello che abbiamo visto l'ultima volta



Gerarchia protocolli

Esiste una gerarchia di incapsulamento dei dati da spedire, definita dalla suite di protocolli Internet TCP/IP, organizzata in quattro livelli



Dimensione massima dei pacchetti

La scorsa lezione abbiamo accennato al fatto che un particolare messaggio debba essere **spezzato in uno o più pacchetti IP** prima che questi vengano instradati in rete

Due diversi fattori:

1. il limite dato dalla **massima quantità di dati** che ogni pacchetto IP può trasportare (che dipende dalla versione considerata del protocollo, IPv4 o IPv6)
2. il limite imposto dalla rete a cui si instradano i pacchetti, ovvero il suo **Maximum Transmission Unit (MTU)**

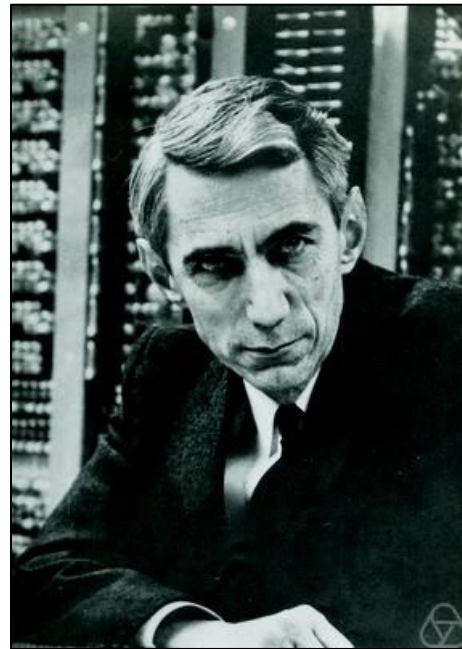
Come sono espressi questi limiti

Byte: unità minima di informazione occupabile su un computer, che storicamente corrisponde al numero di bit necessari per codificare un carattere sul computer

Bit (contrattura di *binary digit*): l'unità minima di informazione che si può scambiare in una comunicazione e può assumere solo uno di due valori: 0 o 1

$$1 \text{ byte} = 8 \text{ bit}$$

Il concetto di bit è stato usato da diversi studiosi del passato (ad esempio Babbage con le schede perforate), ma introdotto come termine formalmente da Claude Shannon (1948)



Teoria dell'informazione

Nell'articolo che introduce il bit, Shannon in realtà mette le basi di uno specifico campo di ricerca e studio chiamato **teoria dell'informazione**

La teoria dell'informazione si occupa di studiare come quantificare, memorizzare, e scambiare informazione, che ha tuttora svariate applicazioni pratiche, oltre che aver caratterizzato e veicolato l'invenzione di diverse tecnologie del passato, incluso lo sviluppo di Internet

Esempio: evincere la capacità massima di un canale per trasmettere informazione in modo affidabile è stato uno degli studi più importanti della teoria dell'informazione (a cui l'MTU è strettamente connesso)

Da bit a simboli e viceversa

Ogni carattere, numero, programma, applicazione all'interno di un computer, così come un qualunque messaggio da scambiare in Internet, è **codificato** come una sequenza di bit che, in qualche modo, vengono poi **decodificati** con la sequenza di caratteri corretta dalla particolare applicazione che riceve ed interpreta quell'informazione

Per esempio, nella codifica binaria tradizionale dei numeri interi, il numero 0 è rappresentato dalla sequenza "0", il numero 1 da "1", il numero 2 da "10", il numero 3 da "11", il numero 4 da "100", e così via

Gli indirizzi IP (v4)

Sono definiti mediante l'uso di **4 byte**, uno per ogni numero

Ogni byte, ovvero 8 bit, permette di definire un **numero intero da 0 a 255**

Codifica / decodifica binaria: ognuno degli otto bit di un numero facente parte dell'indirizzo IP ha assegnato uno specifico valore – il valore 1 è associato al bit più a destra, mentre a quello successivo verso sinistra è assegnato un valore doppio rispetto al precedente (in questo caso 2), e così via

indirizzo IP		208								.	80								.	154								.	224							
bit		1	1	0	1	0	0	0	0		0	1	0	1	0	0	0	0		1	0	0	1	1	0	1	0		1	1	1	0	0	0	0	0
valore		128	64	32	16	8	4	2	1		128	64	32	16	8	4	2	1		128	64	32	16	8	4	2	1		128	64	32	16	8	4	2	1
somma		128+64		+16							64		+16							128		+16		+8		+2		128+64+32								

I pacchetti trasmessi

Un qualunque messaggio e/o pacchetto spedito attraverso Internet è di fatto codificato usando una sequenza di bit

Una parte di questi bit corrispondono a informazioni relative all'**header** del pacchetto, mentre altre sono relative al **payload**

Normalmente la dimensione massima dell'header e del payload è fissata a priori, come abbiamo già detto in precedenza

Transmission Control Protocol (TCP)

Sia il riordino dei pacchetti sia lo spezzare il messaggio originale in più pacchetti conformemente con l'MTU della rete, è gestito dallo strato di trasporto

Protocollo principale:
Transmission Control Protocol (TCP) proposto da Vinton Cerf e Robert Kahn nel 1974 per ARPANET



Cosa permette di fare

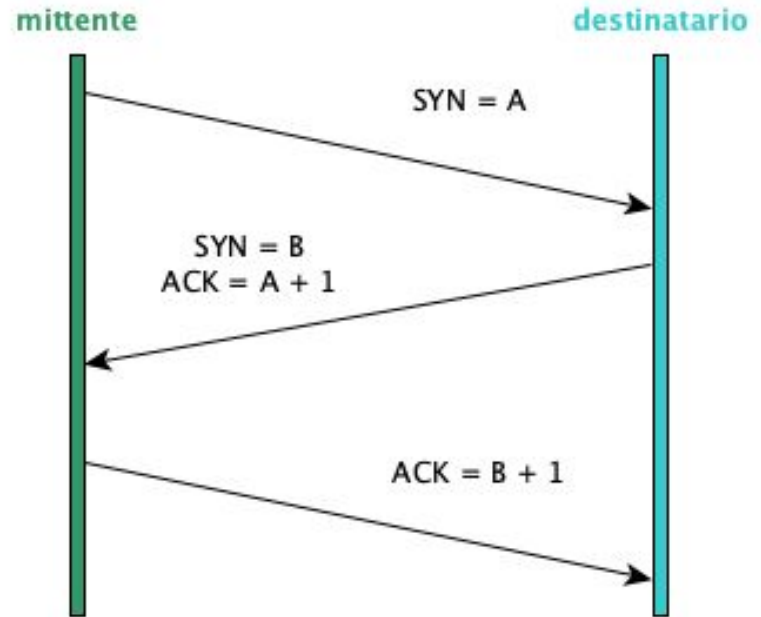
Il TCP permette la consegna **affidabile**, **ordinata**, e **esente da errori** di un flusso di byte tra due computer in comunicazione tra loro attraverso una rete basata sul protocollo IP, e gestisce tutti quei processi che garantiscono la **ritrasmissione** di un pacchetto nel caso in cui non sia stato recapitato al destinatario entro un certo tempo limite

L'**header** di un pacchetto TCP contiene informazioni relative alla comunicazione a livello trasporto tra i due computer mittente e destinatario (le **porte** usate per la comunicazione e un **numero di sequenza** che indica l'ordine dei vari pacchetti TCP), mentre il **payload** contiene le informazioni che devono essere scambiate tra i partecipanti alla comunicazione

Comunicazione connessa

Il mittente e il destinatario **si mettono d'accordo** di iniziare una comunicazione in modo esplicito prima di scambiarsi i dati, e dichiarano altrettanto esplicitamente quando questa comunicazione si può ritenere conclusa

Il processo di inizio della comunicazione, è regolato dal meccanismo del **three-way handshake**, mentre quello di chiusura è il **four-way handshake**



Altri protocolli basati su TCP: SMTP

Il Simple Mail Transfer Protocol (SMTP) è il principale protocollo di comunicazione dello strato applicativo per **spedire** email

Ogni volta che si spedisce una mail, in realtà **non viene recapitata direttamente** al destinatario, ma viene raccolta da un server di posta di competenza del dominio della mail del destinatario

Dominio identificato dalla parte dell'indirizzo email che segue il carattere “@”

Questo server è come un server web ed è in grado di ricevere email attraverso il protocollo SMTP, per poi inoltrarle ai relativi destinatari **quando questi ne fanno richiesta** mediante l'uso del protocollo POP o IMAP

Altri protocolli basati su TCP: POP e IMAP

Sono entrambi utilizzati per richiedere e ricevere email da un server di posta

Con il **Post Office Protocol (POP)**, il server di posta consegna tutti i messaggi precedentemente inviati a quel particolare destinatario e, una volta consegnati, li **cancella** dal server

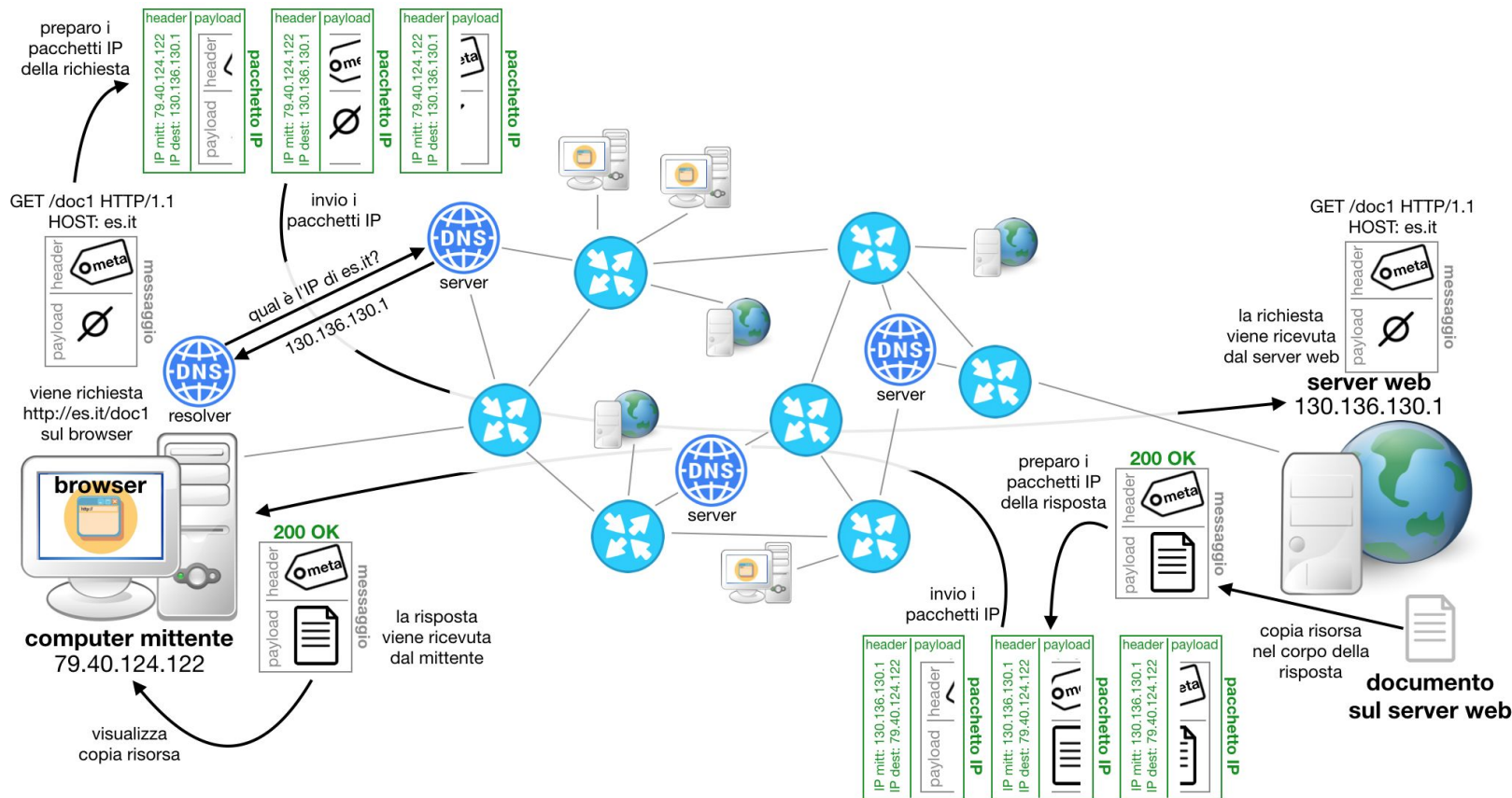
L'**Internet Message Access Protocol (IMAP)** è stato esplicitamente sviluppato per facilitare la gestione di una stessa casella di posta elettronica da parte di più dispositivi – infatti i messaggi di posta elettronica **non vengono eliminati** automaticamente dal server anche se sono stati già recapitati ai rispettivi destinatari a seguito di una loro richiesta, ma devono essere eliminati esplicitamente dall'utente, se lo desidera

Altri protocolli basati su TCP: FTP

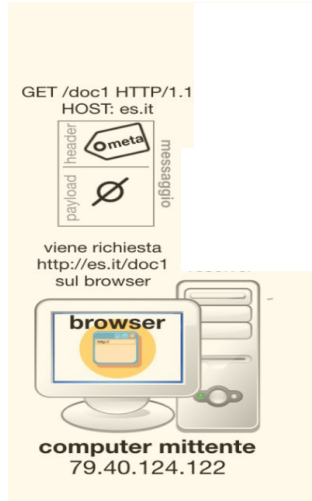
Il **File Transfer Protocol (FTP)** è un protocollo che permette il trasferimento di un qualunque file da un computer ad un server e viceversa

È uno dei protocolli più adottati di Internet per permettere lo scaricamento di una grossa mole di dati, mentre è poco adatto per scaricare documenti molti piccoli, come pagine web

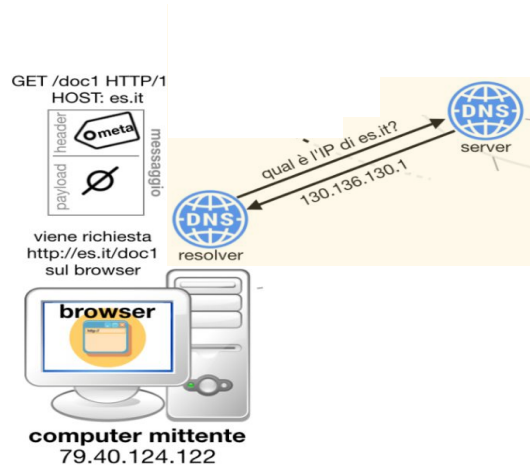
Riassunto di quello che abbiamo visto oggi – prima



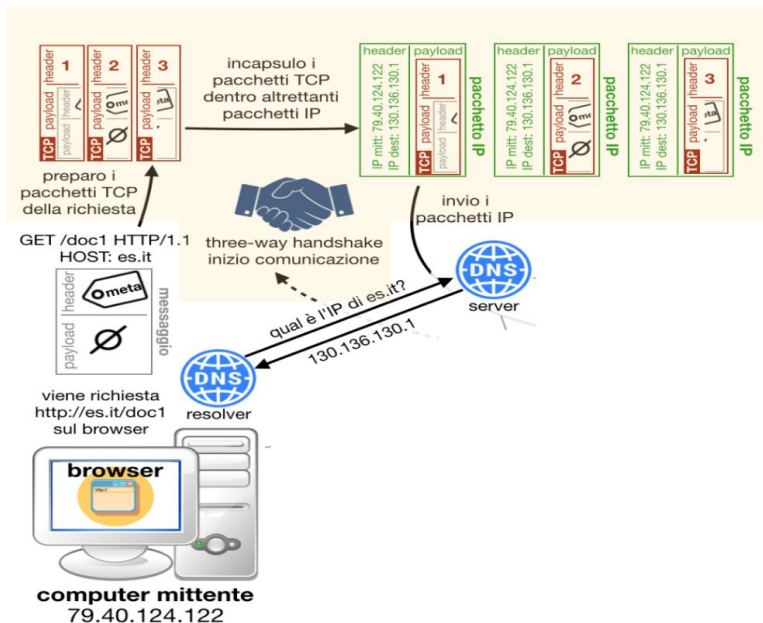
Riassunto comunicazione client - web server



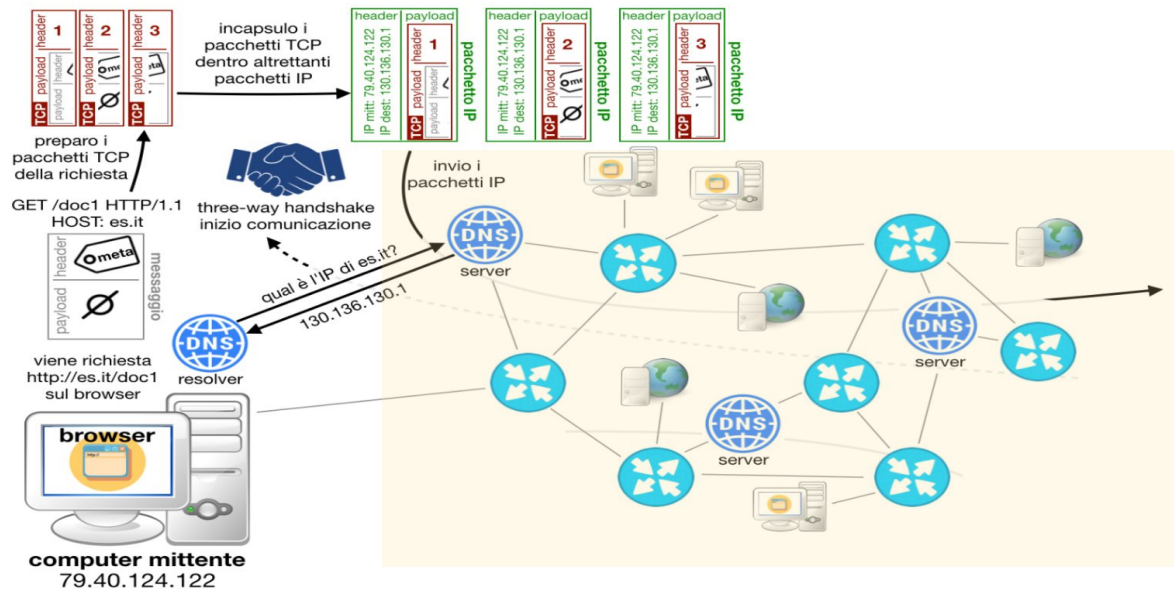
Riassunto comunicazione client - web server



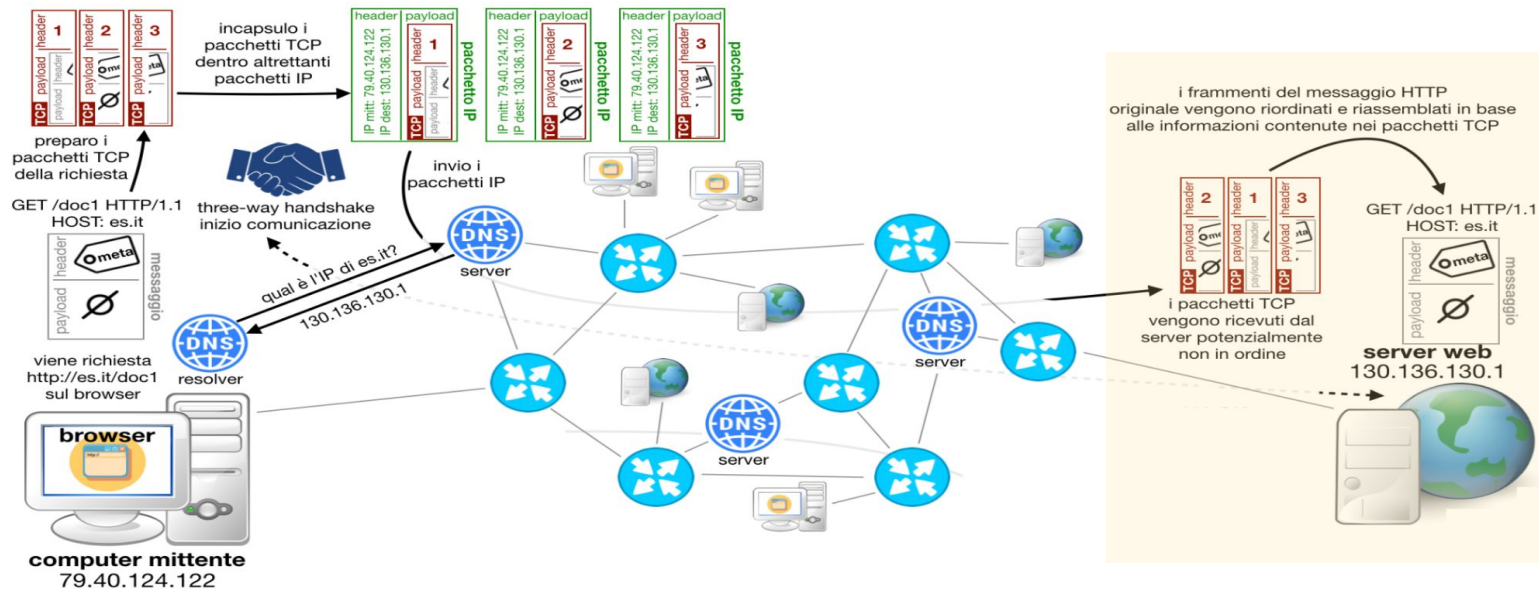
Riassunto comunicazione client - web server



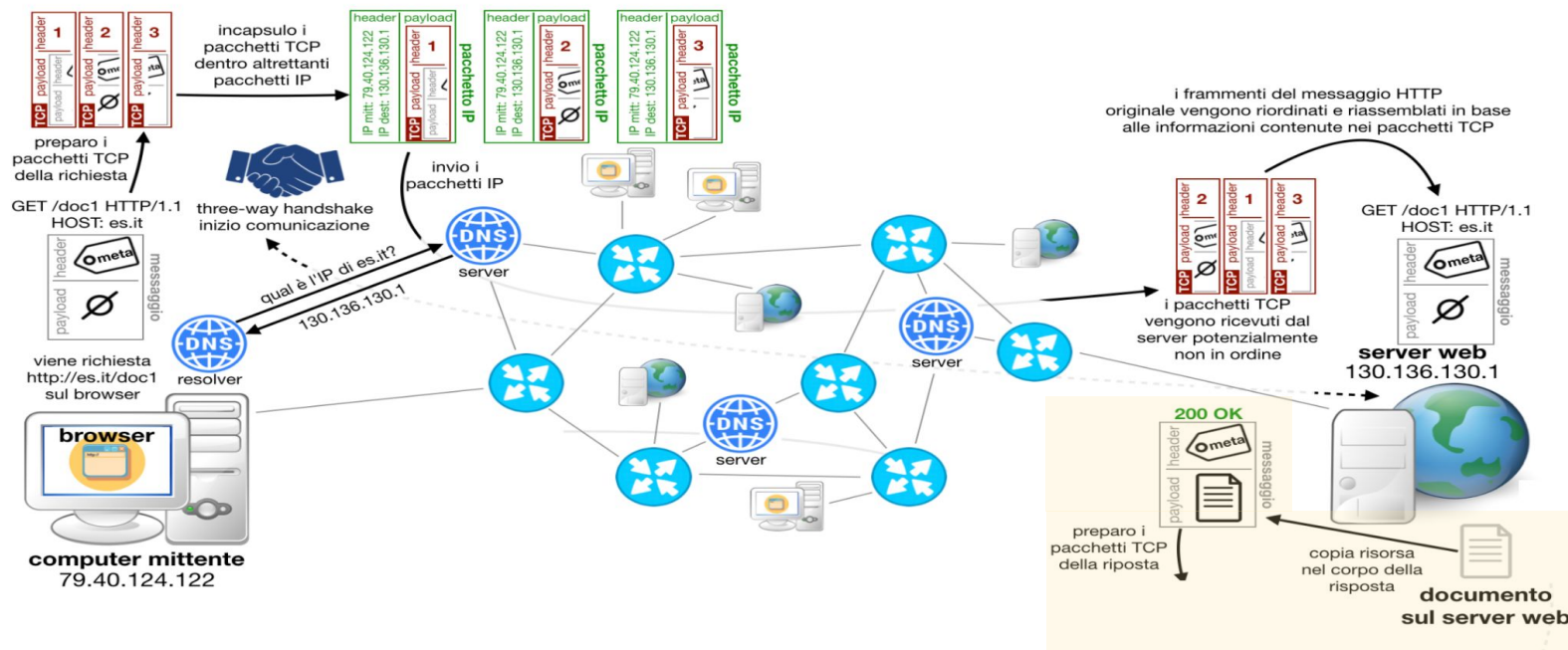
Riassunto comunicazione client - web server



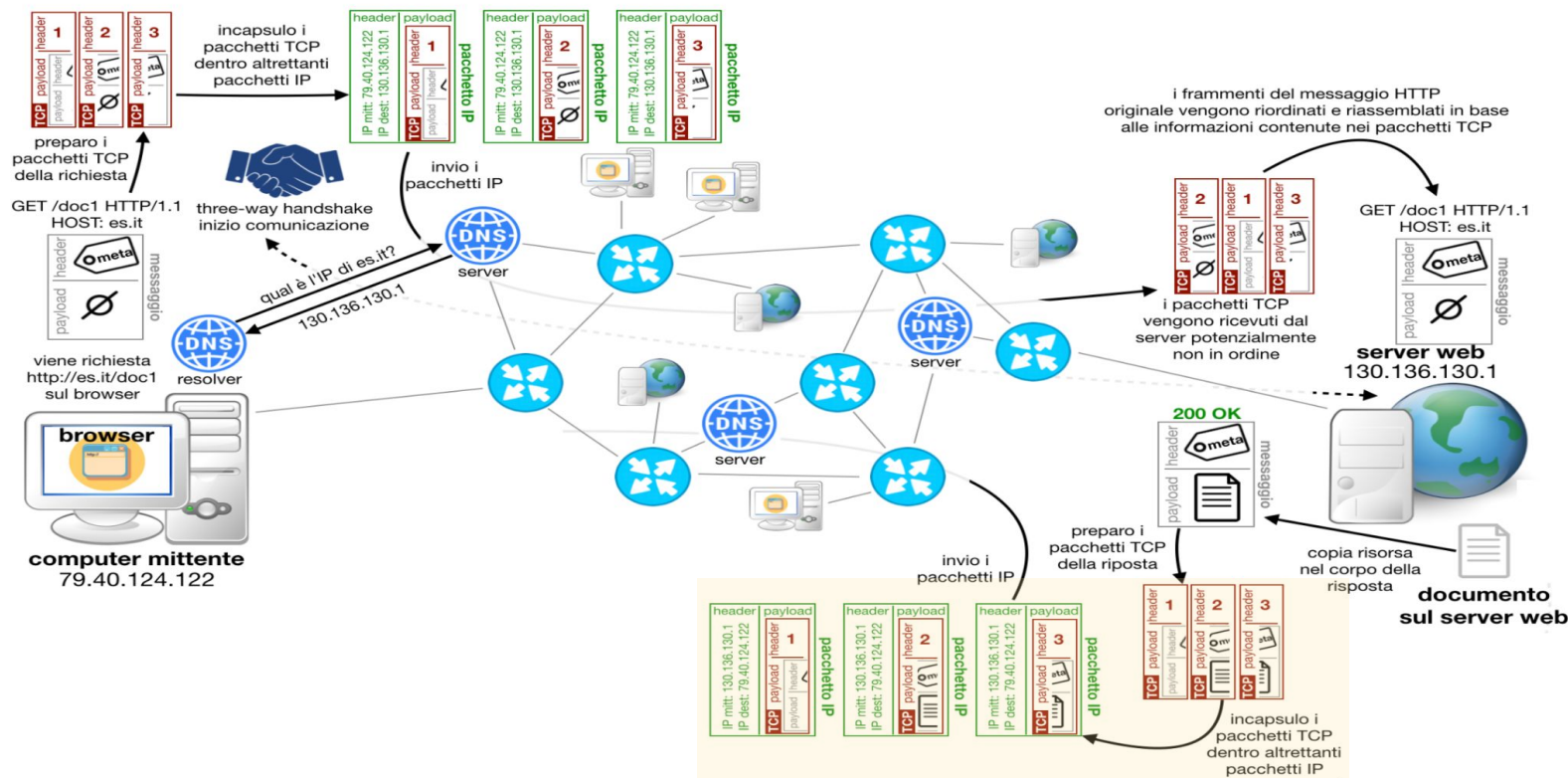
Riassunto comunicazione client - web server



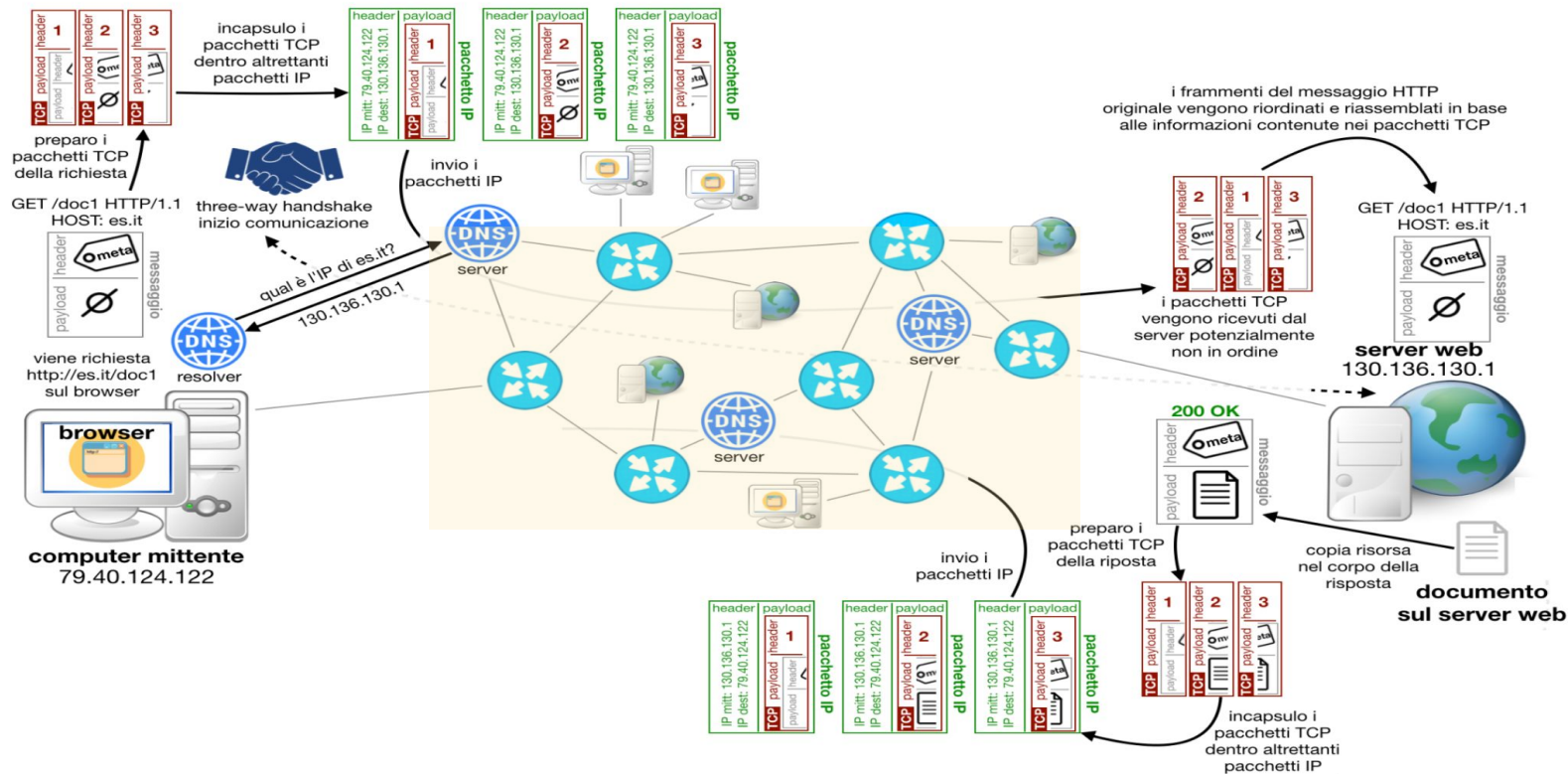
Riassunto comunicazione client - web server



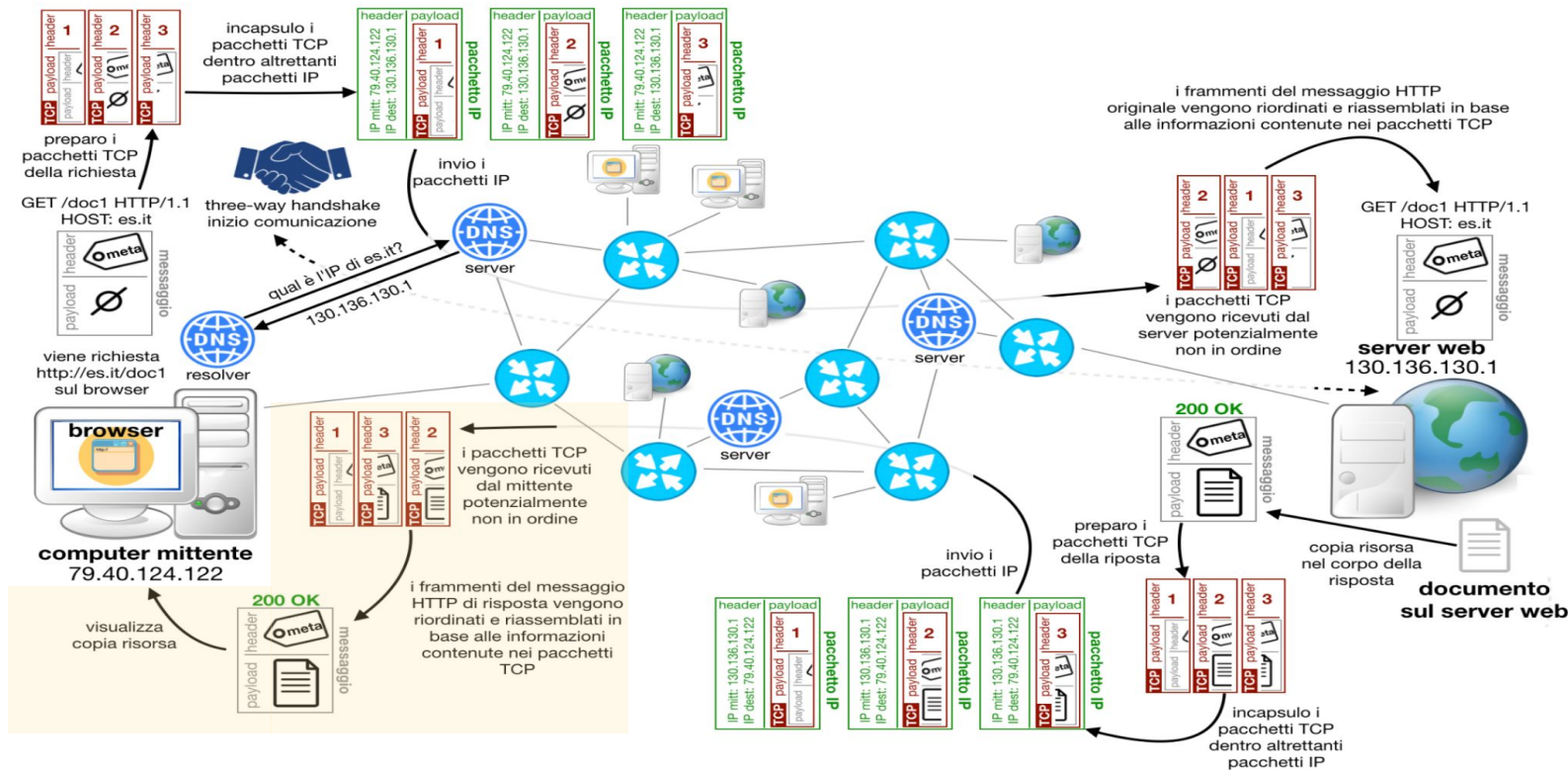
Riassunto comunicazione client - web server



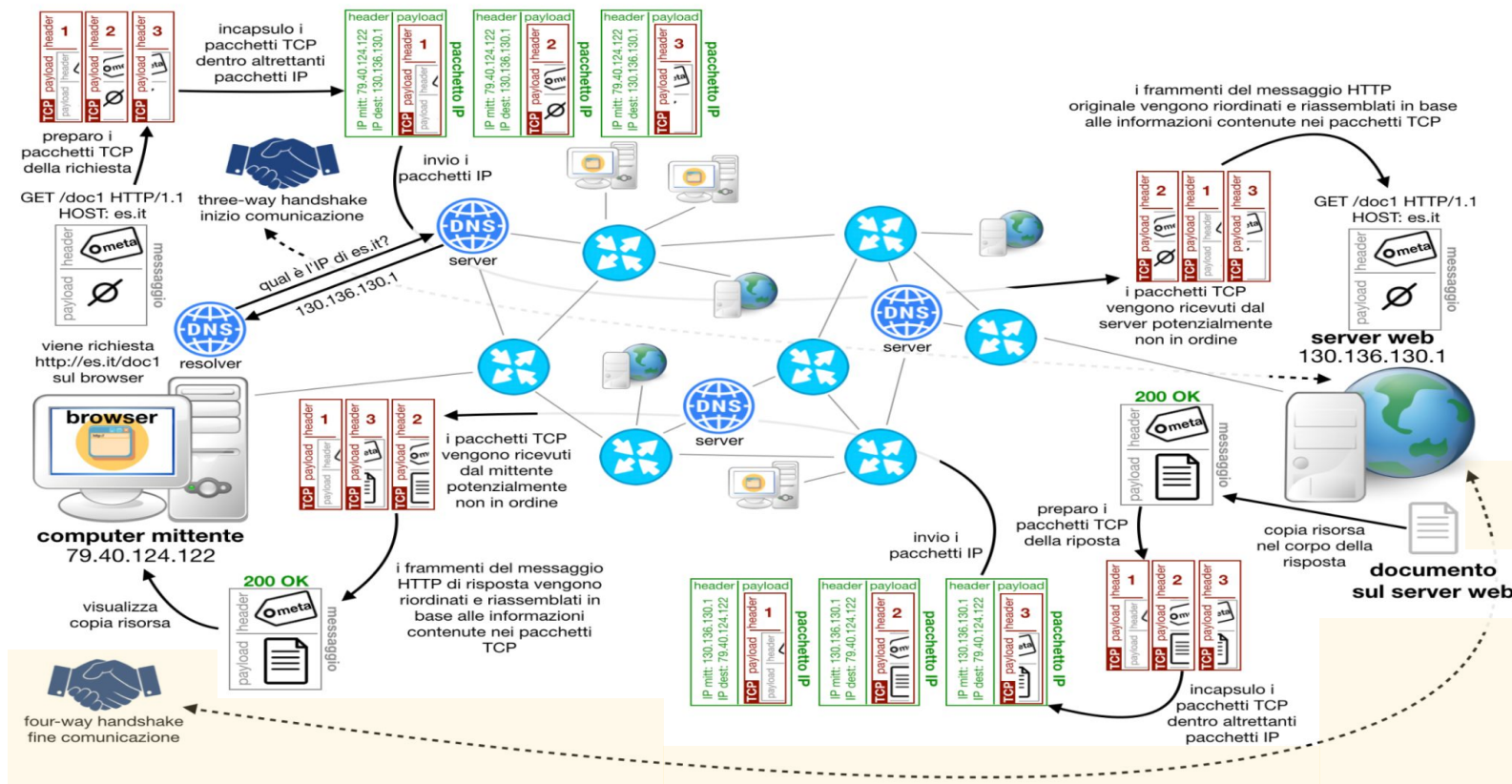
Riassunto comunicazione client - web server



Riassunto comunicazione client - web server



Riassunto comunicazione client - web server



Riassunto di quello che abbiamo visto oggi – dopo

