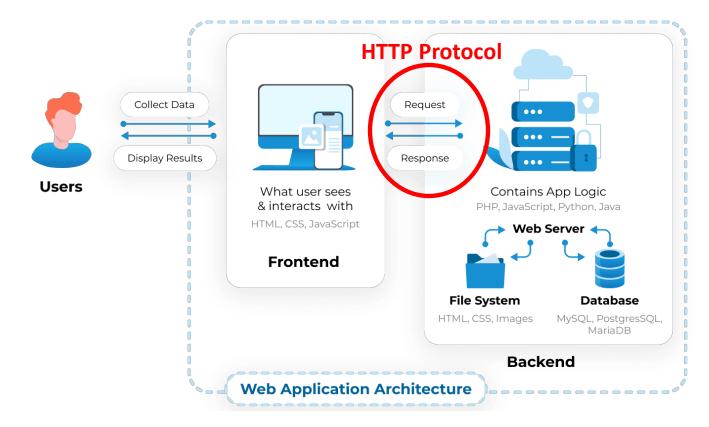
Protocollo HTTP

HTTP e Applicazioni web

Il protocollo HTTP è alla base delle applicazioni Web e rende possibili architetture clientserver eterogenee.



HTTP

molto più ampio:

Il protocollo HTTP (HyperText Transfer Protocol) è uno dei protocolli applicativi più utilizzati. Nasce per traferire pagine HTML da un server ad un browser ma è oggi utilizzato in modo

- Trasferendo anche altre informazioni associate ad una pagina HTML (css e javascript per esempio)
- Trasferendo comandi senza che sia necessaria la presenza di un browser (API)

Per questi motivi HTTP può anche essere considerato come un protocollo di trasporto costruito sopra TCP / IP.

HTTP

- HTTP risale agli anni 90 ed è stato esteso nel tempo
- Oggi è in utilizzo la versione 3 (uscita nel 2022)
- HTTP è un protocollo client server:
 - Il server si mette in attesa di una connessione rimanendo passivo fintanto che non ce e sono
 - Il client deve interpellare espressamente il server per poter comunicare
- HTTP è un protocollo request-response:
 - Il client invia al server delle richieste (di pagine HTML, di immagini, di operazioni...)
 - Il server risponde alla richiesta, soddisfacendola o negandola
 - In assenza di richieste il server rimane passivo
- Al server HTTP è tipicamente associato un nome di dominio
- Il server HTTP è tipicamente in ascolto sulla porta 80

HTTP

- Le richieste e le risposte sono in formato testuale ed hanno un formalismo ben preciso (sebbene non rigido)
- Le richieste e le risposte viaggiano in chiaro. Vengono inseriti in pacchetti TCP e in datagrammi IP senza alcuna protezione. Chiunque le intercetta può leggerle.
- Esiste una versione di HTTP, chiamata HTTPS, che aggiunge anche la cifratura delle informazioni trasmesse (la vedremo tra poco)
- HTTP utilizza TCP. È pertanto un protocollo che lavora su una connessione stabilita, affidabile e con gestione degli errori
- L'interazione è di tipo stateless, ovvero ogni coppia di richiesta/risposta è indipendente da tutte quelle avvenute prima e da tutte quelle avvenute dopo. Il server non comunica mai di sua iniziativa ma solamente in risposta ad una richiesta.

Ogni interazione tra client e server include una richiesta dal client alla quale segue una risposta dal server.

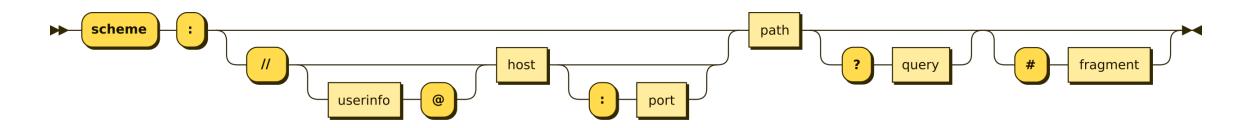
L'interazione è di tipo stateless. Ogni richiesta/risposta è indipendente da tutte quelle avvenute in precedenza.

Ogni richiesta è composta dalle seguenti informazioni principali:

- Url: Indirizzo della risorsa richiesta al server (con parametri opzionali)
- Tipologia di richiesta
- Header della richiesta
- Corpo della richiesta (opzionale)

HTTP - URL

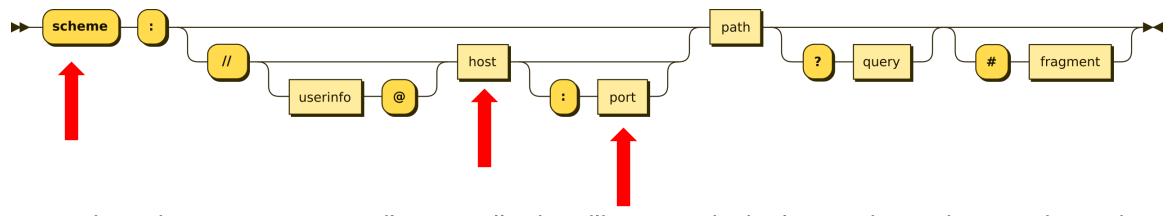
L'url di una risorsa web è nella forma che segue:



Per esempio:

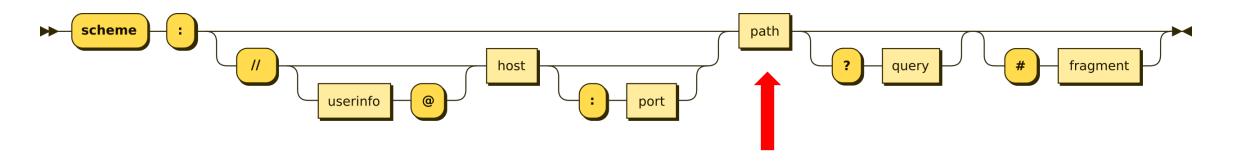
https://www.example.com/search?keywords=python

HTTP – URL – Schema, host e port



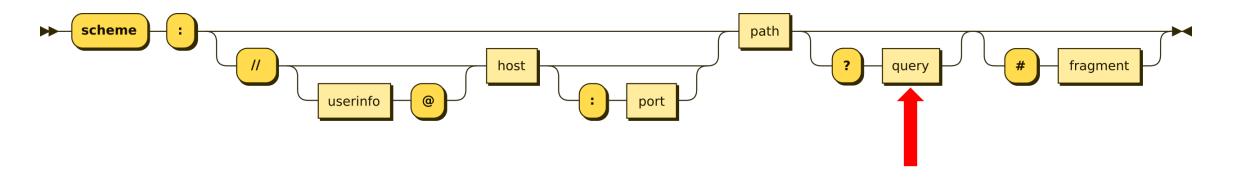
- Lo schema rappresenta il protocollo da utilizzare, principalmente http o https e determina anche la porta di default su cui il server è in ascolto: 80 per http e 443 per https
- Il nome dell'host identifica l'indirizzo IP del server. Può essere espresso direttamente, oppure espresso tramite un nome di dominio (a sua volta convertito in IP tramite il DNS)
- La porta va indicata quando il server non è in ascolto su quella di default.

HTTP – URL – Path



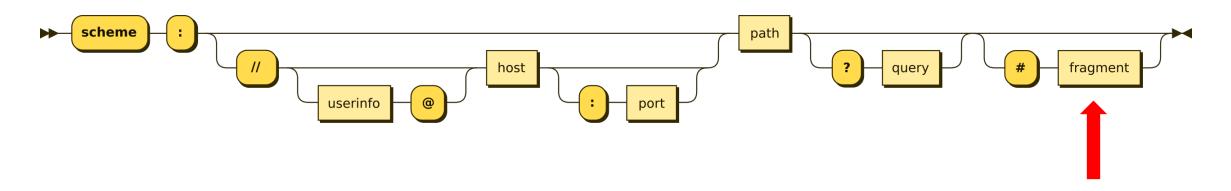
- Il path rappresenta una risorsa all'interno del server
- Può essere associato ad un file (p.e. html) o ad un metodo dell'applicazione di backend
- E' espresso analogamente ai path del file system linux, ovvero in forma strutturata utilizzando / come separatore
- All'assenza di path è associata la risorsa principale del server (tipicamente homepage)
- Esempio: /wiki/Hypertext_Transfer_Protocol

HTTP – URL – Path



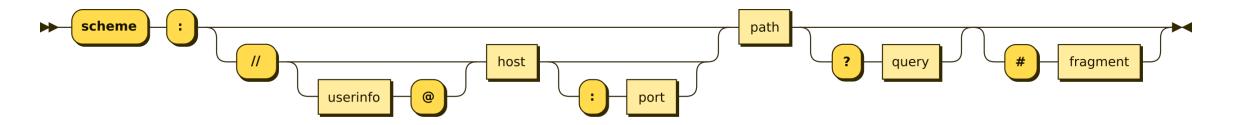
- Alla risorsa identificata dal path possono essere inviati parametri aggiuntivi.
- I parametri sono interpretati esclusivamente dal server e quindi il loro significato cambia in ogni applicazione
- Sono espressi nella forma nomeparametro=valore
- Parametri multipli devono essere separati dal carattere &. Per esempio par1=valore1&par2=valore2

HTTP – URL – Fragment



- Il fragment rappresenta un informazione parametrica ad uso e consumo del client
- Non viene mai inviata al server. In caso di invio il server lo ignora
- Sono utilizzati principalmente per:
 - Effettuare navigazioni all'interno di una stessa pagina
 - Memorizzare informazioni

HTTP – URL – Codifica URL



- La sintassi dell'URL prevede alcuni caratteri speciali (/ ? & # ...) che pertanto non possono essere direttamente utilizzati all'interno di path, query e fragment
- Il carattere spazio è invece non ammesso, e pertanto non utilizzabile in nessuna parte dell'URL
- L'utilizzo dei caratteri speciali e dello spazio è comunque possibile utilizzando una particolare codifica detta codifica URL o codifica percentuale

HTTP – URL – Codifica URL

- Nella codifica URL ogni carattere è sostituito con il carattere % seguito dalla rappresentazione esadecimale del codice ASCII del carattere
- Possono essere codificati tutti i caratteri ma devono necessariamente esserlo quelli

ing [

speciali e quelli non stampabili

- Esempi:
 - & → %26
 - # → %23
 - / → %2F
 - Spazio → %20
 (in alternativa può essere usato il carattere +)

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	*
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22		66	42	В	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	C
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	1	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(72	48	Н	104	68	h
9	9	[HORIZONTAL TAB]	41	29)	73	49	1	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	В	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	1
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E		78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	1	79	4F	0	111	6F	0
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	р
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	S
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	Т	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	V
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	X
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	У
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	1	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F		127	7F	[DEL]
			(0)			856		700			(47)

HTTP - Verbi

Una richiesta Web può assumere quattro tipologie principali, ognuna delle quali identifica il tipo di operazione che vogliamo eseguire sulla risorsa identificata tramite URL

- GET = read
- PUT = update
- POST = create
- DELETE = delete
- HEAD = metadata (header) generati dalla richiesta GET sulla medesima risorsa
- OPTIONS = elenco delle operazioni permesse sulla risorsa

HTTP – Header

- Sono di "servizio" necessarie al server per elaborare correttamente, da un punto di vista tecnico, le richieste
- Sono espressi nella forma key:value
- Alcuni header sono obbligatori (p.e. host)
- Alcuni header sono obbligatori sono in alcune situazioni (p.e. content-type e contentlenght sono obbligatori in caso di presenza del corpo)

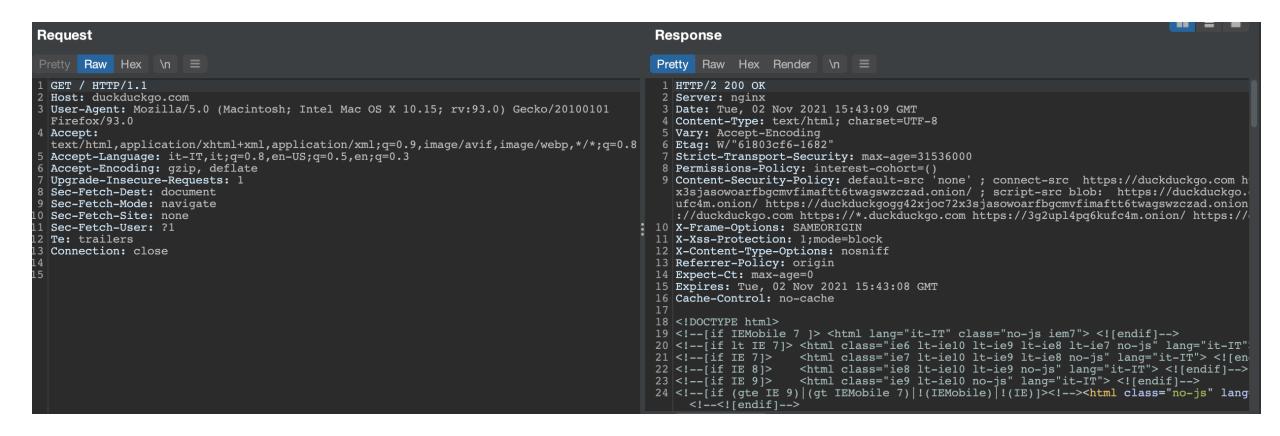
NB: In caso di connessione HTTPS gli header sono cifrati e pertanto leggibili solamente dal client e dal server.

HTTP – Corpo delle richieste

Il corpo della richiesta contiene informazioni aggiuntive ed opzionali. Per esempio:

- Parametri aggiuntivi, necessari al server per elaborare correttamente, da un punto di vista logico, le richieste (per esempio: dati da inserire nel database, credenziali di accesso...)
- Informazioni di autenticazione, necessarie al server per identificare il client e per gestirne le autorizzazioni
- Cookies

NB: In caso di connessione HTTPS il corpo della richiesta è cifrato e pertanto leggibile solamente dal client e dal server.

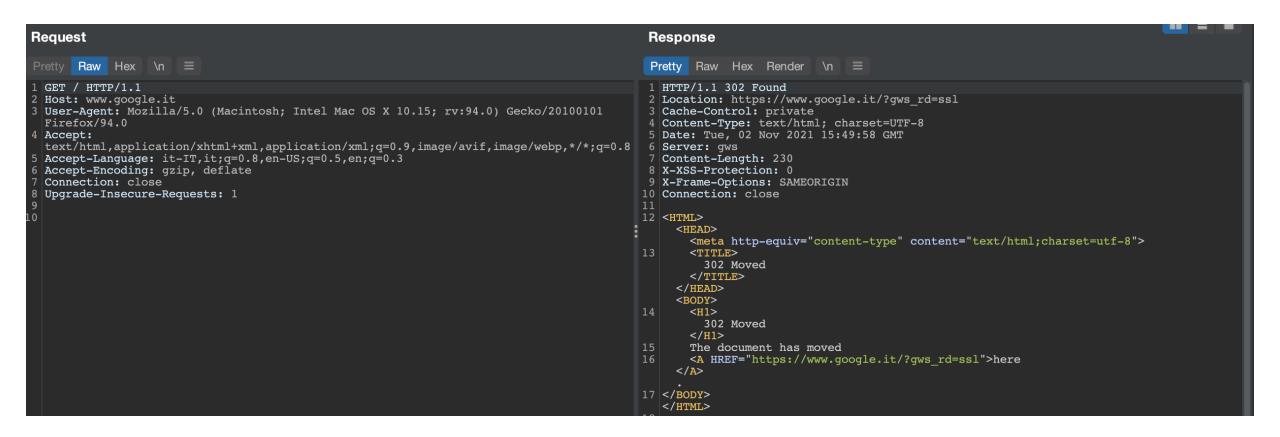


Le response contengono principalmente:

- Un codice di risposta (che ne identifica lo stato)
- I dati della risposta, complessivi di header (per esempio l'html della pagina richiesta).

Codici di risposta:

- 2xx: Operazione avvenuta con successo
- 3xx: Redirect verso altra risorsa
- 4xx: Risorsa non trovata o errori di autenticazione (in generale Client error)
- 5xx: Server Error



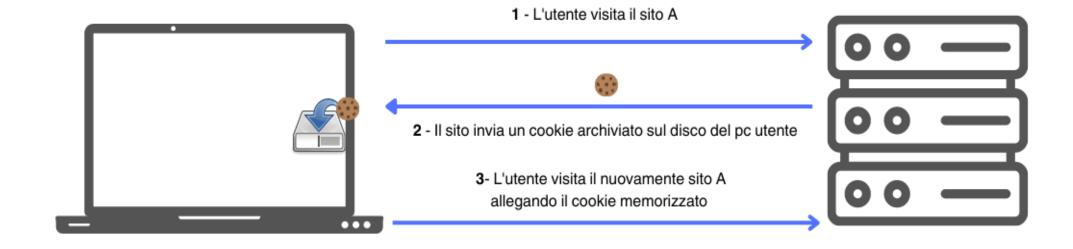
Le response possono contenere dati di vario tipo. I più diffusi sono HTML per le pagine e json per i dati. In caso di connessioni HTTPS il corpo della risposta è cifrato.

```
Request
                                                                                           Response
 Pretty Raw Hex \n ≡
                                                                                           Pretty Raw Hex Render \n ≡
  GET /swagger/docs/v1 HTTP/2
                                                                                           1 HTTP/2 429 Too Many Requests
  Host: api.tfl.gov.uk
                                                                                           2 Date: Tue, 02 Nov 2021 15:55:26 GMT
  User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:94.0) Gecko/20100101
                                                                                           3 Content-Type: application/json
                                                                                           4 Content-Length: 83
  Firefox/94.0
  Accept: application/json; charset=utf-8,*/*
                                                                                           5 Retry-After: 5
  Accept-Language: it-IT, it; q=0.8, en-US; q=0.5, en; q=0.3
                                                                                           6 Cf-Cache-Status: DYNAMIC
  Accept-Encoding: gzip, deflate
                                                                                           7 Expect-Ct: max-age=604800, report-uri="https://report-uri.cloudflare.com/cdn-cgi/beacc
  Referer: https://api.tfl.gov.uk/swagger/ui/index.html?url=/swagger/docs/v1
                                                                                           8 Server: cloudflare
                                                                                           9 Cf-Ray: 6a7e8c123907d210-MAN
  Sec-Fetch-Dest: empty
  Sec-Fetch-Mode: cors
10 Sec-Fetch-Site: same-origin
  Te: trailers
                                                                                               "statusCode": 429,
                                                                                               "message": "Rate limit is exceeded. Try again in 5 seconds."
```

Cookies

- Il protocollo HTTP è stateless
- Per renderlo statefull sono stati introdotti i cookies.
- I cookie sono informazioni testuali inviate dal server e memorizzate dal browser sul client
- Il browser invia automaticamente i cookie ad ogni successiva richiesta effettuata verso il server che li ha generati
- I cookie sono utilizzati principalmente per:
 - Gestione delle sessioni
 - Personalizzazione delle applicazioni
 - Monitoraggio

Cookies



Cookies

- Il server invia i cookie tramite un opportuno header: Set-Cookie
- Ogni cookie è composto da:
 - Un nome
 - Un valore
 - Metadata
 - Server di origine (al quale verranno re-inviati)
 - Scadenza
 - Politiche di sicurezza (p.e. attributi Secure, HTTPOnly e same site)

Cookies – Attributi di sicurezza

Sui cookie è possibile impostare alcuni attributi che ne limitano l'utilizzo modificandone la sicurezza:

- Attributo **secure**, per garantire che il token venga inviato solo su conessioni https
- Attributo httpOnly per evirare che i cookie possano essere letti dal codice javascript.
- Attributo sameSite impostato a strinct o lax per limitare il rischio di cross site request forgery (CSRF)

Cookies - Attributo same site

Il comportamento standard che il browser tiene nei confronti dei cookie è quello di inviarli ogni volta che c'è una richiesta indirizzata al sito a cui quei cookie appartengono, indipendentemente da dove ha avuto origine la richiesta di invio.

E' possibile modificare tale comportamento tramite l'assegnazione dell'attributo SameSite. Sono previsti 2 valori:

- Strict: Il browser non includerà i cookie se le richieste hanno origine da un sito diverso rispetto a quello a cui i cookie appartengono
- Lax: Il browser include i cookie a fronte di richieste con origine da un sito diverso se e solo se:
 - La richiesta è di tipo GET
 - La richiesta scaturisce da una azione utente (per esempio click su link) e non da

Cookies - Attributo same site

SameSite = Strict è la soluzione più sicura ma genera problemi di User Experience.

L'utente infatti dovrà nuovamente autenticarsi ogni volta che apre l'applicazione usando un link esterno.

Il settaggio dell'attributo SameSite avviene in fase di creazione dei cookie:

Set-Cookie: SessionId=sYMnfCUrAlmqVVZn9dqevxyFpKZt30NN; SameSite=Strict;

HTTPS

Come detto in precedenza il protocollo HTTP trasporta dati testuali IN CHIARO.

Chiunque intercetta i pacchetti è quindi in grado di leggerne il contenuto e/o di modificarlo, con i prevedibili problemi che questo comporta.

Il protocollo HTTPS è una estensione del protocollo HTTP nel quale i dati viaggiano in modo cifrato e solamente il client e il server sono in grado di decifrarli.

In questo modo un attacco MITM (Man In The Middle) sarebbe inefficace a causa dell'impossibilità di leggere/modificare i dati (permane il rischio associato alla cancellazione dei pacchetti).

Quanto visto per il protocollo HTTP rimane valido. Vi è però l'aggiunta di un protocollo di cifratura (TLS) e un sistema di scambio delle chiavi (Certificati X.509)

TLS (e il suo predecessore SSL) è un esempio di protocollo che utilizza vari metodi crittografici per ottenere comunicazioni cifrate, in modo sicuro, partendo da canali non sicuri.

E' un protocollo molto utilizzato in quanto si inserisce nello stack TCP/IP tra il livello di trasporto e il livello applicativo, diventando trasparente a quest'ultimo e pertanto utilizzabile

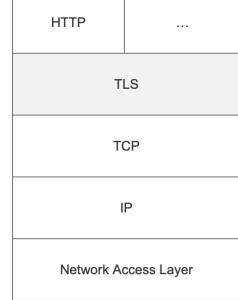
da qualsiasi applicazione.

La versione attuale è la 1.3.

La versione 1.2 è ancora manutenuta ma è consigliato l'upgrade

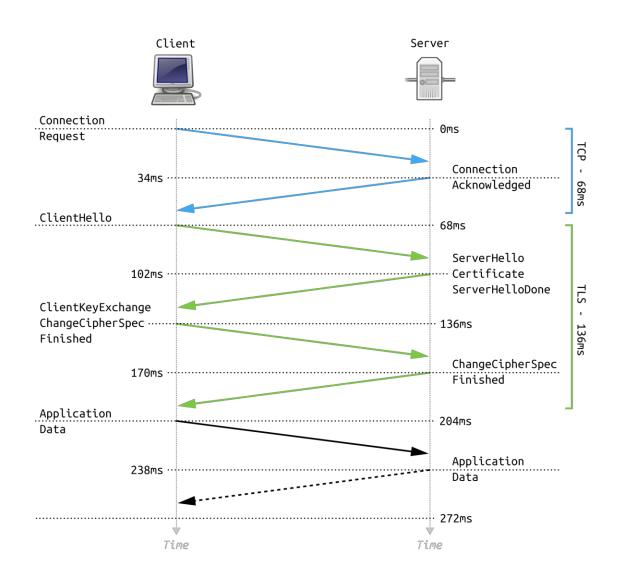
La versione 1.1 e tutte le precedenti sono deprecate.

ing Davide Luppoli – 2024



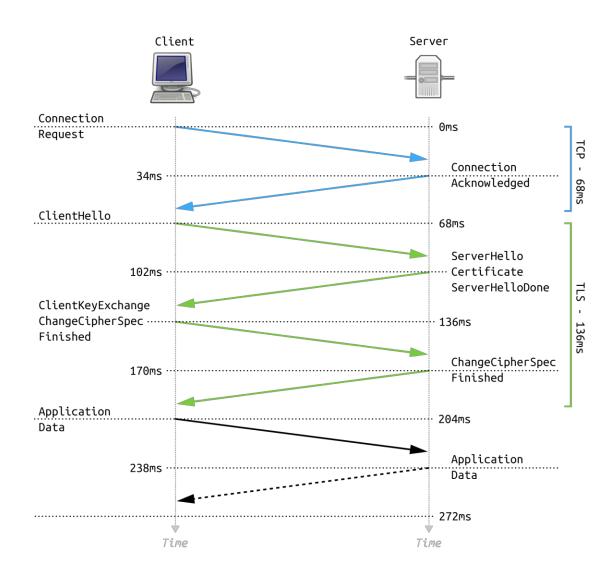
TLS è composto da più fasi:

- Negoziazione iniziale tra le parti
- Autenticazione
- Scambio delle chiavi simmetriche
- Cifratura simmetrica (con autenticazione)

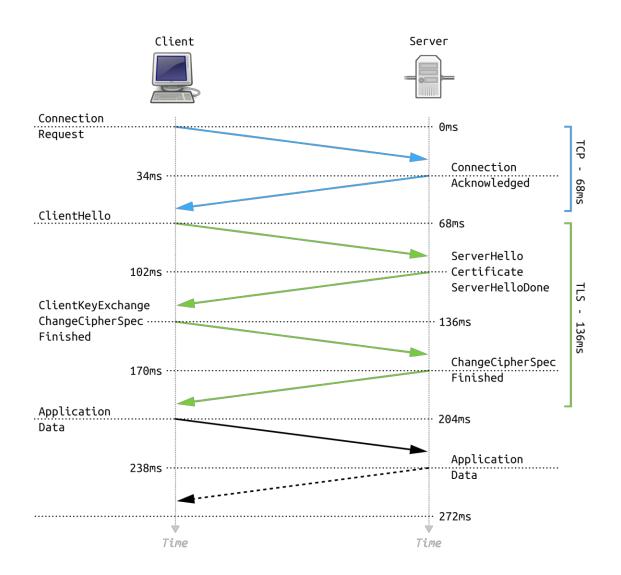


La versione TLS 1.2 prevede due roundtrip, oltre alla connessione iniziale TCP:

- Il client invia il messaggio ClientHello specificando la versione massima di TLS supportata così come i metodi di scambio chiave, di hash e di cifratura supportati.
- Il server risponde con un messaggio di ServerHello nel quale specifica i protocolli scelti per la comunicazione (i più sicuri tra quelli supportati dal client)
- Il server invia anche il proprio certificato, unitamente ad un messaggio di completamento (ServerHelloDone)



- Il client invia il messaggio ClientHello specificando la versione massima di TLS supportata così come i metodi di scambio chiave, di hash e di cifratura supportati.
- Il server risponde con un messaggio di ServerHello nel quale specifica i protocolli scelti per la comunicazione (i più sicuri tra quelli supportati dal client)
- Se è previsto uno scambio delle chiavi (p.e. con DH), il server invia anche un messaggio di tipo ServerKeyExchange con i parametri da usare per lo scambio chiavi
- Il server invia anche il proprio certificato, unitamente ad un messaggio di completamento (ServerHelloDone)



 Il client risponde con un messaggio di tipo ClientKeyExchange il cui valore dipende dall'algoritmo di scambio chiave concordato. Può essere una pre chiave calcolata in modo random e cifrata con la chiave pubblica del server, oppure può essere una prechiave dello scambio DH

NB: A questo punto client e server hanno concordato una prechiave (master key) dalla quale derivare la chiave finale. La stessa chiave può essere utilizzata per derivare altri parametri, quali i vettori di inizializzazione.

- Il client comunica che passerà al nuovo cifrario
- Il server comunica che passerà al nuovo cifrario

ing Davide Luppoli – 2024 La comunicazione prosegue con il nuovo cifrario

TLS 1.3 semplifica il protocollo eliminando un round trip:

- Il client invia ClientHello, unitamente agli algoritmi supportati. Il client ipotizza però quale sia la combinazione scelta dal server ed invia già il messaggio ClientKeyExchange adeguato
- Il server risponde accettando gli algoritmi proposti dal client, allegando il proprio certificato e il proprio messaggio server_key_exchange
- Client e server calcolano la master key e tutto continua come in TLS 1.2

Key exchange/agreement and authentication

Le versioni di TLS si differenziano per l'handshake, per le ottimizzazioni nelle implementazioni/utilizzi degli algoritmi e, soprattutto, per gli algoritmo crittografici utilizzabili:

key exchange/agreement and authentication										
Algorithm	SSL 2.0	SSL 3.0	TLS 1.0	TLS 1.1	TLS 1.2	TLS 1.3				
RSA	Yes	Yes	Yes	Yes	Yes	No				
DH-RSA	No	Yes	Yes	Yes	Yes	No				
DHE-RSA (forward secrecy)	No	Yes	Yes	Yes	Yes	Yes				
ECDH-RSA	No	No	Yes	Yes	Yes	No				
ECDHE-RSA (forward secrecy)	No	No	Yes	Yes	Yes	Yes				
DH-DSS	No	Yes	Yes	Yes	Yes	No				
DHE-DSS (forward secrecy)	No	Yes	Yes	Yes	Yes	No ^[69]				
ECDH-ECDSA	No	No	Yes	Yes	Yes	No				
ECDHE-ECDSA (forward secrecy)	No	No	Yes	Yes	Yes	Yes				
ECDH-EdDSA	No	No	Yes	Yes	Yes	No				
ECDHE-EdDSA (forward secrecy) ^[70]	No	No	Yes	Yes	Yes	Yes				
PSK	No	No	Yes	Yes	Yes	?				
PSK-RSA	No	No	Yes	Yes	Yes	?				
DHE-PSK (forward secrecy)	No	No	Yes	Yes	Yes	Yes				
ECDHE-PSK (forward secrecy)	No	No	Yes	Yes	Yes	Yes				
SRP	No	No	Yes	Yes	Yes	?				
SRP-DSS	No	No	Yes	Yes	Yes	?				
SRP-RSA	No	No	Yes	Yes	Yes	?				
Kerberos	No	No	Yes	Yes	Yes	?				
DH-ANON (insecure)	No	Yes	Yes	Yes	Yes	?				
ECDH-ANON (insecure)	No	No	Yes	Yes	Yes	?				
GOST R 34.10-94/34.10-2001 ^[71]	No	No	Yes	Yes	Yes	?				

	Cipher	Protocol version							
Туре	Algorithm	Nominal strength (bits)	SSL 2.0	SSL 3.0[n 1][n 2][n 3][n 4]	TLS 1.0 ^{[n 1][n 3]}	TLS 1.1 ^[n 1]	TLS 1.2 ^[n 1]	TLS 1.3	
	AES GCM ^{[72][n 5]}		_	_	_	_	Secure	Secure	
	AES CCM ^{[73][n 5]}	256, 128	_	_	_	_	Secure	Secure	
	AES CBC ^[n 6]	250, 125	-	Insecure	Depends on mitigations	Depends on mitigations	Depends on mitigations	-	
	Camellia GCM ^{[74][n 5]}		_	_	_	_	Secure	_	
	Camellia CBC ^{[75][n 6]}	256, 128	-	Insecure	Depends on mitigations	Depends on mitigations	Depends on mitigations	_	
	ARIA GCM ^{[76][n 5]}		_	_	_	_	Secure	_	
Block cipl	ARIA CBC ^{[76][n 6]}	256, 128	_	-	Depends on mitigations	Depends on Depends or mitigations mitigations		_	
mode o operatio	SEED CBC[77][n 6]	128	_	Insecure	Depends on mitigations	Depends on mitigations	Depends on mitigations	_	
	3DES EDE CBC ^{[n 6][n 7]}	112 ^[n 8]	Insecure	Insecure	Insecure	Insecure	Insecure	-	
	GOST 28147-89 CNT ^{[71][n 7]}	256	_	-	Insecure	Insecure	Insecure	-	
	IDEA CBC[n 6][n 7][n 9]	128	Insecure	Insecure	Insecure	Insecure	_	-	
	DES CBC[n 6][n 7][n 9]	56	Insecure	Insecure	Insecure	Insecure	-	_	
	DE2 CBCm alfa villa al	40[n 10]	Insecure	Insecure	Insecure	-	_	_	
	RC2 CBC ^{[n 6][n 7]}	40[n 10]	Insecure	Insecure	Insecure	-	_	-	
	ChaCha20- Poly1305 ^{[82][n 5]}	256	-	-	8		Secure	Secure	
Stream cip	RC4 ^[n 11]	128	Insecure	Insecure	Insecure	Insecure	Insecure	_	
	HC4*****	40 ^[n 10]	Insecure	Insecure	Insecure	-	_	_	
None	Null ^[n 12]	_	Insecure	Insecure	Insecure	Insecure	Insecure	-	

Cipher security against publicly known feasible attacks

Oltre agli attacchi possibili sui vari cifrari utilizzati, TLS è esposto anche all'attacco di Downgrade.

In questo attacco l'attaccante cerca di forzare client e server a concordare un protocollo datato e non più considerato sicuro (per esempio SSL 2.0 con DES CBC), per poi intercettare i dati e decifrarli sfruttando le vulnerabilità del cifrario.

L'attacco è spesso possibile in quanto i server implementano, per retro-compatibilità o per configurazioni non corretta, anche algoritmi e protocolli datati.