

Qualità del servizio

Scritto da [Febogamedev](#)

Nel campo delle reti, il termine qualità del servizio, o più semplicemente **QoS** (dall'inglese Quality of Service), è utilizzato per indicare i parametri usati per caratterizzare la qualità del servizio offerto dalla rete (ad esempio, la velocità, la perdita di pacchetti, il ritardo ecc.) oppure gli strumenti e le tecniche per ottenere una qualità del servizio desiderata (Blake et al., 1998; International Telecommunication Union [ITU], 2011).

Di seguito è riportato un elenco degli attributi principali per definire la qualità di una connessione di rete.

Banda

La banda, in informatica, si riferisce alla quantità di dati che possono essere trasferiti da un punto all'altro in un dato periodo di tempo (Chimento & Ishac, 2008).

$$Banda = \frac{\text{Dimensione delle informazioni}}{\text{Tempo di trasferimento}}$$

La banda viene misurata in bit al secondo (bps). Il bit rate indica infatti la quantità di bit trasferiti in un secondo. I suoi multipli più comuni sono kilobit al secondo (Kbps, mille bit al secondo), megabit al secondo (Mbps, un milione di bit al secondo) e gigabit al secondo (Gbps, un miliardo di bit al secondo) (Chimento & Ishac, 2008).

La banda può riferirsi a collegamenti diversi. Per esempio, può indicare la capacità di un collegamento della rete locale oppure quella dell'ultimo miglio, cioè del collegamento che connette la rete dell'utente alla rete dell'ISP (Chimento & Ishac, 2008).

Quando si parla della “velocità di Internet” di una connessione domestica, ci si riferisce normalmente alla banda disponibile sull'ultimo miglio.

Nel linguaggio delle reti, il termine banda viene spesso usato per indicare la capacità massima teorica di trasferimento dati di una connessione, misurata in bit al secondo (Chimento & Ishac, 2008).

In questo contesto, parlare di una connessione da 100 Mbps significa indicare che, in condizioni ideali, quella connessione può trasferire fino a 100 milioni di bit al secondo.

Questa velocità può variare a seconda di queste variabili (Chimento & Ishac, 2008; Constantine et al., 2011):

- Tecnologia di accesso, come DSL, FTTC, FTTH, rete mobile o satellite.
- Standard di comunicazione utilizzato, che stabilisce la capacità massima supportata (Ethernet 100BASE-TX, Gigabit Ethernet, VDSL2, ecc.).
- Caratteristiche del mezzo trasmissivo, come rame, fibra ottica o onde radio.
- Qualità fisica e lunghezza del collegamento, particolarmente rilevanti nelle connessioni su rame.

- Profilo configurato dall'ISP, che determina la banda massima prevista dal contratto.
- Capacità degli apparati e delle interfacce di rete utilizzati lungo il collegamento.

La larghezza di banda può essere simmetrica o asimmetrica (Constantine et al., 2011):

- Una larghezza di banda simmetrica significa che le velocità di download (downstream, la velocità di trasferimento dei dati dal server verso il client) e upload (upstream, la velocità di trasferimento dei dati dal client verso il server) sono le stesse.
- Una larghezza di banda asimmetrica, invece, ha velocità di download e upload diverse.

Molti *ISP* domestici offrono connessioni asimmetriche, con velocità di download più elevate rispetto alle velocità di upload. Questo avviene perché, statisticamente, siamo portati più a scaricare contenuti da Internet piuttosto che a caricare nuovi contenuti (Constantine et al., 2011).

Esempio

L'*ADSL* sta per Asymmetric Digital Subscriber Line. Tale tecnologia di connessione è asimmetrica perché la banda del downstream è maggiore della banda dell'upstream.

Un altro concetto importante è il **throughput** (Constantine et al., 2011).

Il throughput è la quantità effettiva di dati che viene trasferita in un certo intervallo di tempo. A differenza della banda teorica, il throughput tiene conto delle condizioni reali della rete, come congestione, interferenze, distanza, qualità del segnale, perdita di pacchetti e traffico generato da altri utenti (Constantine et al., 2011).

Quindi il throughput è sempre minore della banda o, al limite, uguale (Chimento & Ishac, 2008; Constantine et al., 2011).

Il throughput può dipendere da (Chimento & Ishac, 2008; Constantine et al., 2011):

- Banda disponibile nel collegamento più lento attraversato dai dati.
- Congestione della rete, dovuta alla presenza di molti utenti o a un traffico elevato.
- Tempo e dati di controllo necessari al funzionamento dei protocolli, per esempio per stabilire la connessione, confermare la ricezione dei dati e gestire eventuali errori.
- Perdita o corruzione dei pacchetti, che rende necessarie eventuali ritrasmissioni (Constantine et al., 2011).
- Sovraccarico dei protocolli, perché parte dei dati trasmessi è costituita da intestazioni e informazioni di controllo (Chimento & Ishac, 2008; Constantine et al., 2011).
- Qualità della rete locale, comprese copertura, interferenze e distanza dal punto di accesso Wi-Fi.
- Prestazioni di modem, router, dispositivi e interfacce di rete.
- Capacità del server o del servizio remoto che invia o riceve i dati.
- Numero di applicazioni e dispositivi che utilizzano contemporaneamente la connessione.
- Politiche di gestione del traffico dell'ISP, come limitazioni, priorità e meccanismi di traffic shaping (Constantine et al., 2011).

Esempio

Devo scaricare un file di 20 MB (un byte equivale a 8 bit) e ci impiego 20 minuti.

$$\text{Throughput} = \frac{20 \text{ MB}}{20 \text{ m}} = \frac{20\,000\,000 \text{ B}}{20 * 60 \text{ s}} = \frac{20\,000\,000 * 8 \text{ b}}{20 * 60 \text{ s}} = \frac{160\,000\,000 \text{ b}}{1\,200 \text{ s}} = 133\,333 \text{ bps} \approx 133 \text{ kbps}$$

Dalla larghezza di banda si può calcolare il tempo di trasferimento minimo teorico di un contenuto di una certa dimensione in byte (Chimento & Ishac, 2008).

Consideriamo la formula inversa della banda:

$$\text{tempo di trasferimento} = \frac{\text{Dimensione delle informazioni}}{\text{Banda}}$$

Esempio

Supponiamo di voler trasferire un file di 10 MB (10 megabyte) su un dispositivo con banda di 5 Mbit/s.

Il tempo di trasferimento sarà dato da:

$$\begin{aligned} \text{Tempo di trasferimento} &= \frac{10 \text{ MB}}{5 \text{ Mbps}} = \frac{10\,000\,000 \text{ B}}{5\,000\,000 \text{ bps}} = \\ &= \frac{10\,000\,000 * 8 \text{ b}}{5\,000\,000 \text{ bps}} = \frac{80\,000\,000 \text{ b}}{5\,000\,000 \text{ bps}} = 16 \text{ s} \end{aligned}$$

Latenza

La **latenza** (o tempo di latenza; in inglese: latency) è il tempo che intercorre tra l'invio di un pacchetto, di una richiesta o di un segnale e la sua ricezione o risposta da parte della destinazione: rappresenta il ritardo che si verifica tra l'inizio di un'azione e la sua effettiva esecuzione o risposta (ITU, 2019; Almes et al., 1999).

La latenza è misurata in millisecondi (*ms*) ed è influenzata da diversi fattori all'interno di una rete. Alcuni dei principali fattori che contribuiscono alla latenza includono (ITU, 2019; Almes et al., 1999):

- Velocità della connessione: una connessione Internet più veloce tende a ridurre la latenza poiché i pacchetti di dati vengono trasmessi più rapidamente.
- Distanza fisica: la latenza aumenta con la distanza fisica tra la sorgente e la destinazione dei dati. Ad esempio, le connessioni a lunga distanza, come quelle tra continenti, possono avere latenze più elevate (Almes et al., 1999).
- Congestione di rete: quando la rete è sovraccarica di traffico, i pacchetti possono subire ritardi nell'essere instradati verso la destinazione desiderata (ITU, 2019; Morton & Claise, 2009).
- Qualità dell'infrastruttura: la qualità e l'affidabilità dei router, dei cavi e di altri componenti di rete possono influenzare la latenza complessiva.
- Protocolli di rete: alcuni protocolli di rete sono più efficienti di altri e possono ridurre la latenza.

È importante sottolineare che banda e latenza non sono la stessa cosa (ITU, 2019; Constantine et al., 2011).

Una connessione con molta banda può trasferire grandi quantità di dati al secondo, ma non è detto che abbia una latenza bassa.

La latenza dipende anche dalla distanza fisica, dal percorso seguito dai pacchetti, dalla congestione e dai tempi di elaborazione nei dispositivi di rete.

La latenza può avere un impatto significativo su diverse attività, specialmente in applicazioni che richiedono tempi di risposta rapidi, come i giochi online o le videochiamate (International Telecommunication Union [ITU], 2003; ITU, 2011).

Ad esempio, una latenza elevata nei giochi online potrebbe causare ritardi tra l'azione dell'utente e la risposta del gioco, influenzando negativamente l'esperienza di gioco.

In generale, una latenza bassa è preferibile, soprattutto per le applicazioni in tempo reale, poiché consente un funzionamento più fluido e responsivo della rete.

Gli operatori di rete e gli ingegneri di rete cercano costantemente di migliorare la latenza attraverso l'ottimizzazione dell'infrastruttura, l'implementazione di tecnologie avanzate e l'uso di protocolli efficienti.

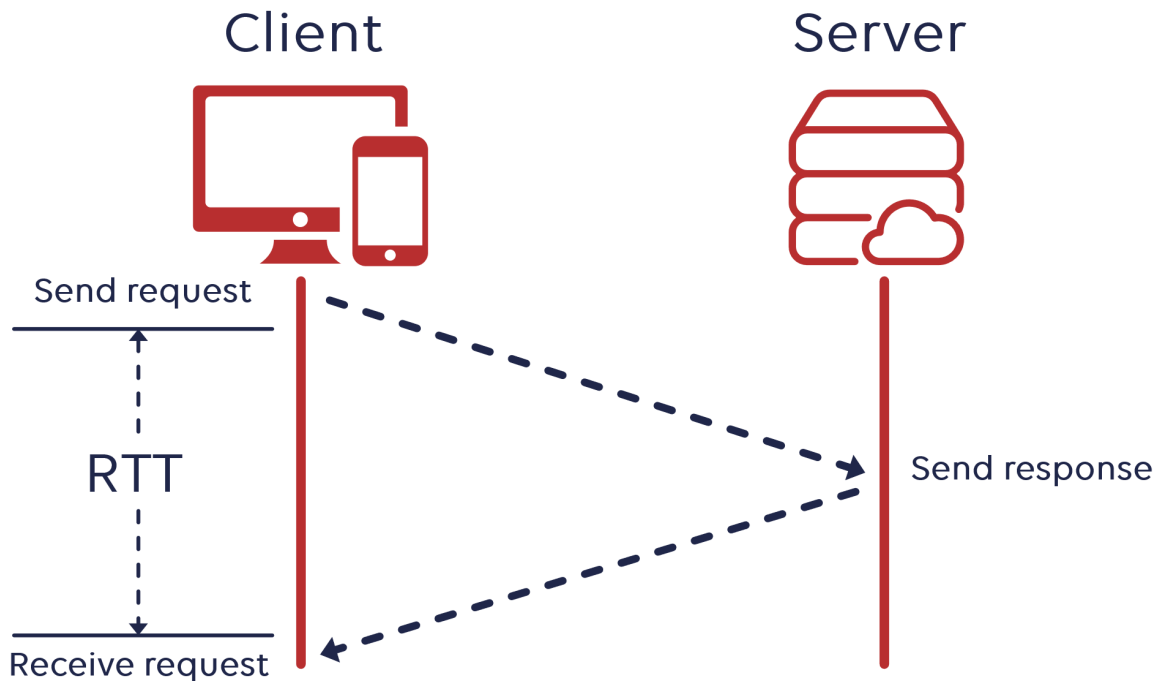
Esempio

Una connessione con latenza intorno a 15–30 ms può risultare molto reattiva per giochi online. Invece, una latenza di 80–100 ms è ancora utilizzabile, ma può rendere più percepibile il ritardo nelle applicazioni in tempo reale.

Il Round Trip Time (*RTT*, tempo di andata e ritorno) è una misura correlata alla latenza: rappresenta il tempo che intercorre tra l'invio di un pacchetto verso una destinazione e la ricezione della relativa risposta da parte del mittente (Almes et al., 1999).

In altre parole, il *RTT* è il tempo totale necessario per completare un giro (andata e ritorno) tra due punti di una connessione di rete.

In una rete ideale e simmetrica, il *RTT* può essere considerato circa il doppio della latenza solo andata, perché comprende sia il tempo necessario per raggiungere la destinazione sia il tempo necessario per ricevere la risposta. Nelle reti reali, però, questa relazione non è sempre precisa, perché il percorso di andata e quello di ritorno possono avere ritardi diversi (Almes et al., 1999).



Jitter

Il **jitter** è un altro importante concetto particolarmente rilevante per applicazioni in tempo reale, come le chiamate vocali su Internet o le videoconferenze (Schulzrinne et al., 2003; Demichelis & Chimento, 2002).

Il jitter rappresenta la variazione nella latenza dei pacchetti di dati mentre viaggiano da una sorgente a una destinazione sulla rete (Demichelis & Chimento, 2002).

- Se i pacchetti arrivano con ritardi molto diversi tra loro, il jitter è alto.
- Se invece arrivano con ritardi simili e regolari, il jitter è basso.

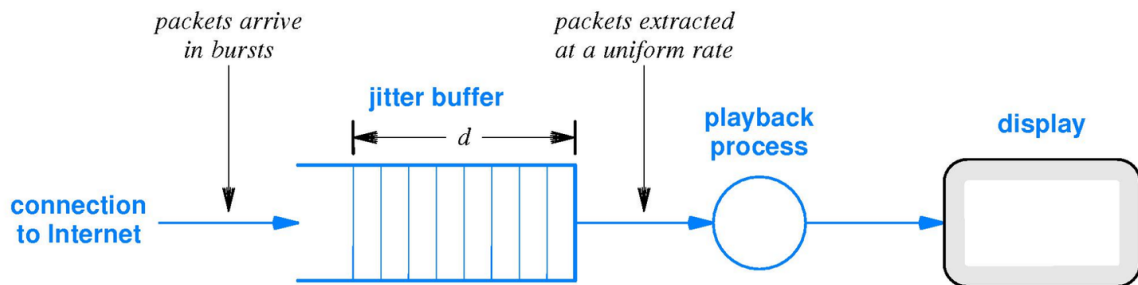
Il jitter può essere problematico per le applicazioni in tempo reale perché può causare fenomeni come ritardi nella trasmissione della voce o delle immagini durante una videochiamata (Schulzrinne et al., 2003; Morton & Claise, 2009).

Se il jitter è basso, il flusso dei pacchetti è più uniforme e l'esperienza dell'utente è migliore, in quanto ci sono meno interruzioni o problemi di sincronizzazione tra i partecipanti.

Le cause del jitter possono essere diverse, tra cui (Morton & Claise, 2009):

- Congestione di rete: quando una rete è sovraccarica di traffico, i pacchetti possono essere ritardati o ricevuti fuori sequenza, causando jitter.
- Percorsi di rete variabili: se i pacchetti devono attraversare percorsi diversi sulla rete, possono subire latenze diverse, contribuendo al jitter.
- Gestione della coda: nei dispositivi di rete, la gestione della coda dei pacchetti può influenzare il jitter. Se i pacchetti vengono accumulati in una coda troppo grande o troppo piccola, il loro ritardo può variare (Morton & Claise, 2009).

Per affrontare il jitter e migliorare la qualità delle comunicazioni in tempo reale, i fornitori di servizi di rete e gli sviluppatori di applicazioni implementano diverse tecniche, come la prioritizzazione del traffico, l'utilizzo di buffer (vedremo fra poco questa tecnica) o la scelta di percorsi di rete più stabili, al fine di ridurre la variabilità nella latenza e mantenere un flusso regolare dei dati (Blake et al., 1998; Morton & Claise, 2009).



Possiamo vedere il jitter come la deviazione standard della latenza (Demichelis & Chimento, 2002).

La deviazione standard è una misura statistica che rappresenta la dispersione o la variabilità di un insieme di dati rispetto alla loro media. In altre parole, indica quanto i valori di un dato campione tendono a differire dalla media campionaria.

Esempio

Se la media di un insieme di dati è 5 e la deviazione standard è 0, vuol dire che tutti i dati sono uguali a 5.

Buffer

Il termine buffer indica una porzione di memoria temporanea usata per compensare differenze di velocità tra due dispositivi o processi che si scambiano dati (Morton & Claise, 2009).

In pratica, il buffer accumula dati in attesa di essere elaborati o trasmessi, agendo da “cuscinetto” (da cui il nome) tra chi produce i dati e chi li consuma.

Immagina un rubinetno da cui l'acqua esce più velocemente di quanto il bicchiere possa raccogliercela senza traboccare: serve un contenitore intermedio.

Nei sistemi digitali, quel contenitore è il **buffer**: una RAM temporanea che aiuta a compensare differenze di velocità e irregolarità temporali.

Esempio

Alcuni esempi concreti di uso del buffer:

- Stampa: il file da stampare viene bufferizzato prima di essere inviato alla stampante.
- Reti: un router usa buffer per immagazzinare pacchetti in arrivo prima di inoltrarli.
- Audio digitale: i suoni registrati in tempo reale vengono bufferizzati per essere elaborati senza interruzioni.
- Scrittura su disco: i dati da salvare vengono messi in un buffer e poi scritti su disco.

Un uso importante del buffer su Internet è limitare gli effetti delle variazioni temporanee nella ricezione dei dati. Per esempio, nello streaming video o audio, il buffer serve a caricare in anticipo una parte del contenuto, in modo da evitare interruzioni durante la riproduzione (Morton & Claise, 2009).

Quando l'utente avvia un video, per esempio su YouTube, Netflix o Spotify, il sistema non riproduce immediatamente ogni dato appena ricevuto.

Invece, accumula una porzione del contenuto in un buffer locale, cioè in una memoria temporanea. Quando il buffer contiene abbastanza dati, inizia la riproduzione.

Nel frattempo, nuovi dati continuano ad arrivare e a riempire il buffer.

Questo processo protegge da rallentamenti temporanei della rete e permette una riproduzione più fluida, anche se i dati non arrivano sempre con regolarità.

Se però il flusso in ingresso nel buffer rimane stabilmente superiore alla capacità di elaborazione o riproduzione, il buffer finirà per riempirsi, causando attese, scarti o interruzioni.

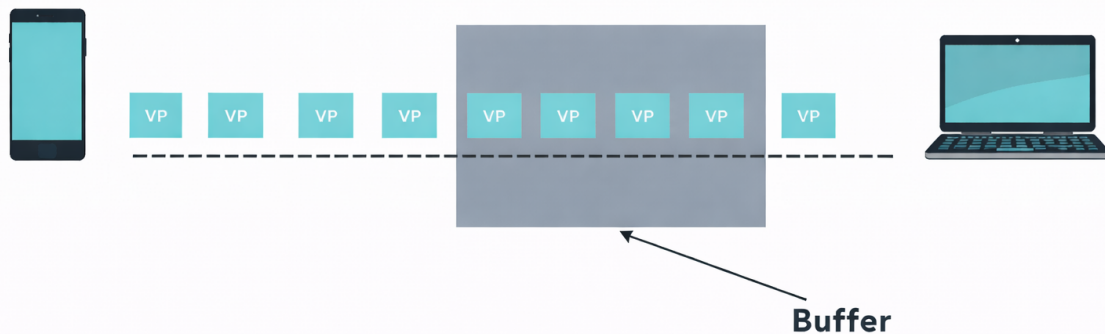
Nei flussi in tempo reale, come chiamate vocali e videochiamate, si parla più specificamente di **jitter buffer** (Morton & Claise, 2009).

Un jitter buffer è un buffer progettato per compensare le variazioni nei tempi di arrivo dei pacchetti, mantenendo però la latenza più bassa possibile (Morton & Claise, 2009).

La differenza principale è:

- Nello streaming tradizionale il sistema può permettersi di accumulare più dati in anticipo, introducendo anche un ritardo maggiore.
- In una videochiamata, invece, il ritardo deve rimanere minimo, perché gli utenti devono poter interagire in tempo reale.

Implementing A Jitter Buffer



In questo schema si vede come funziona il buffer: conserva i pacchetti per inviare poi un flusso con latenza costante al dispositivo finale

Il buffer è usato per compensare il jitter e rendere regolare la riproduzione.

- Se il jitter è alto, serve un buffer più grande.
- Ma un buffer troppo grande introduce latenza (si deve aspettare tanto prima che il buffer si riempia), quindi serve un compromesso intelligente, spesso gestito da algoritmi jitter buffer dinamici (Morton & Claise, 2009).

Qualità di servizio nei media

Avere una qualità di servizio adeguata per visualizzare i media è importante visto che una qualità di servizio bassa comporta un degrado del media stesso (Blake et al., 1998; ITU, 2011).

La qualità di servizio può avere specifiche diverse in base alle diverse tipologie di media (ITU, 2011; Schulzrinne et al., 2003):

- Media non continui: nei media non continui, come testo e immagini, la latenza può influenzare il tempo di attesa prima della visualizzazione. Tuttavia, una volta ricevuti i dati, il contenuto può essere mostrato senza dover rispettare un ritmo temporale preciso.
- I media continui: nei media continui, come audio e video, non conta solo che i dati arrivino, ma anche che arrivino con regolarità. Un ritardo iniziale può essere accettabile se il contenuto non è interattivo, perché può essere compensato con un

buffer. Diventa invece più problematico se i pacchetti arrivano con ritardi molto variabili, perché l'audio può risultare interrotto e il video può andare a scatti (Schulzrinne et al., 2003; Morton & Claise, 2009).

- Live streaming: il contenuto viene prodotto e trasmesso quasi in tempo reale. Anche qui si può usare un buffer, ma non può essere troppo grande, altrimenti l'utente vedrebbe l'evento con un ritardo eccessivo (ITU, 2011; Morton & Claise, 2009).
- Media interattivi in tempo reale: nelle applicazioni interattive, come videogiochi online e videochiamate, il ritardo deve essere ancora più basso, perché l'utente si aspetta una risposta immediata alle proprie azioni (ITU, 2003; Schulzrinne et al., 2003).

La gestione della qualità di servizio nei media presenta quattro livelli di difficoltà, in ordine crescente (ITU, 2011; Schulzrinne et al., 2003):

- La gestione di media non continui, come immagini e testo.
- La gestione dello streaming di contenuti registrati (*stored media*), come vedere un video su YouTube.
- Lo streaming dei contenuti prodotti in tempo reale (*live media*), come la radio. Il problema principale nel progettare applicazioni di streaming e di teleconferenza è il ritardo di rete.
 - Audio e video richiedono una presentazione in tempo reale, il che significa che, per essere utili, devono essere inviati sulla rete a tempi fissati.
 - Applicazioni interattive in tempo reale, come i videogiochi o le teleconferenze. In questo caso grandi ritardi implicano che le chiamate che dovrebbero essere interattive non lo sono più (ITU, 2003; Schulzrinne et al., 2003).

Ping

Un modo semplice per analizzare la qualità del servizio è il **ping** (Postel, 1981; Almes et al., 1999).

Il ping è un comando di rete usato per verificare se un host è raggiungibile tramite una rete *IP* e per misurare il tempo di risposta. Può anche aiutare a individuare eventuali perdite di pacchetti, osservando se alcune richieste non ricevono risposta (Postel, 1981; Almes et al., 1999).

Il nome deriva dal suono del sonar nei sottomarini: come un'eco sonora verifica se qualcosa risponde, il ping verifica se un host risponde su Internet o in una *LAN*.

- Quando si esegue un ping, il computer invia un pacchetto speciale a un indirizzo *IP* o a un nome di dominio.
- Il dispositivo di destinazione, se raggiungibile, risponde con un altro pacchetto.

Il ping lo si può scrivere come comando del terminale del sistema operativo. Per esempio, se si scrive:

```
ping google.com
```

Il risultato sarà:

```
Risposta da 142.250.184.14: byte=32 tempo=24ms TTL=118
```

Dove:

- byte = 32 → dimensione del pacchetto inviato.

- tempo = 24 ms → tempo impiegato per ricevere la risposta, cioè il suo *RTT* (Almes et al., 1999).
- *TTL* = 118 → *Time To Live*, quanti “salti” di router può fare il pacchetto prima di scadere (Postel, 1981).

Materiale di supporto

Elementi chiave

- **La qualità del servizio (QoS)** descrive le prestazioni di una rete attraverso parametri come banda, latenza, jitter e perdita di pacchetti.
- **La banda** indica la quantità massima teorica di dati trasferibili in un secondo ed è misurata in bps.
- **Il throughput** è la velocità effettivamente ottenuta e normalmente è inferiore alla banda teorica.
- La banda può essere **simmetrica**, con download e upload uguali, oppure **asimmetrica**.
- **La latenza** è il ritardo nella trasmissione dei dati ed è misurata in millisecondi.
- **Il RTT** indica il tempo necessario perché un pacchetto raggiunga la destinazione e la risposta ritorni al mittente.
- **Il jitter** è la variazione della latenza tra pacchetti diversi e può creare interruzioni nei servizi in tempo reale.
- **Un buffer** conserva temporaneamente i dati per compensare rallentamenti e irregolarità nella ricezione.
- Un buffer più grande riduce le interruzioni, ma può aumentare la latenza.
- Testi e immagini tollerano maggiormente i ritardi; streaming, videochiamate e giochi online richiedono dati più regolari e tempi di risposta ridotti.
- **Il ping** verifica se un host è raggiungibile e misura l'RTT, la perdita di pacchetti e altre informazioni sulla comunicazione.

Esercizi

La connessione più veloce non è sempre quella migliore

Due connessioni Internet hanno queste caratteristiche:

Connessione A

- Banda: 1 Gbps.
- Latenza media: 95 ms.

Connessione B

- Banda: 200 Mbps.
- Latenza media: 18 ms.

Un utente deve scegliere quale connessione utilizzare per:

- Scaricare un file da 20 GB.
- Partecipare a una videochiamata.
- Giocare a un videogioco competitivo online.
- Guardare un film in streaming.

Rispondi alle seguenti domande:

- Indica quale connessione potrebbe essere preferibile per ciascuna attività e motiva la risposta.
- Spiega perché una banda maggiore non implica necessariamente una latenza minore.
- Indica quali attività sono maggiormente sensibili alla latenza.
- Spiega perché il download di un file è influenzato soprattutto dalla quantità di dati trasferibile nel tempo.
- Spiega perché nei videogiochi online una connessione con meno banda può comunque offrire un'esperienza migliore.
- Usa il caso proposto per spiegare perché la qualità di una connessione non può essere descritta attraverso un solo parametro.

Banda teorica e throughput reale

Una scuola possiede una connessione dichiarata da 500 Mbps.

Durante un test viene scaricato un file di 1 GB in 25 secondi.

Rispondi alle seguenti domande:

- Converti la dimensione del file da byte a bit.
- Calcola il throughput medio effettivo ottenuto durante il download.
- Confronta il risultato con la banda teorica di 500 Mbps.
- Spiega perché il throughput può essere inferiore alla banda nominale.
- Indica almeno quattro fattori presenti nella dispensa che possono ridurre il throughput reale.
- Un tecnico sostiene che la connessione è guasta semplicemente perché il throughput non raggiunge esattamente 500 Mbps. Valuta l'affermazione e motiva la risposta.
- Spiega perché, in condizioni reali, il throughput può essere uguale alla banda solo nel caso limite ideale.

Quanto tempo dovrebbe servire?

Un videomaker deve caricare sul cloud un file video da 6 GB.

La sua connessione è asimmetrica:

- Download: 300 Mbps.
- Upload: 20 Mbps.

Rispondi alle seguenti domande:

- Indica quale delle due velocità deve essere utilizzata per calcolare il tempo di caricamento del file.
- Converti 6 GB in bit.
- Calcola il tempo minimo teorico necessario per completare l'upload assumendo che l'intera banda sia realmente disponibile.
- Spiega perché nella pratica il trasferimento potrebbe richiedere più tempo.
- Spiega cosa significa che la connessione è asimmetrica.
- Indica perché una connessione di questo tipo può risultare adeguata per un utente che guarda molti contenuti ma meno conveniente per chi carica continuamente file video di grandi dimensioni.

- Un secondo operatore propone una connessione simmetrica da 100 Mbps. Spiega quale delle due connessioni potrebbe essere più interessante per il videomaker e motiva la scelta.

Misurare il ritardo con il ping

Un amministratore esegue il comando ping verso un server e ottiene questi tempi di risposta:

18 ms — 21 ms — 19 ms — 20 ms — 22 ms

Successivamente ripete il test e ottiene:

18 ms — 90 ms — 25 ms — 140 ms — 21 ms

Rispondi alle seguenti domande:

- Spiega che cosa rappresentano i valori in millisecondi mostrati dal ping.
- Indica quale dei due test mostra una connessione temporalmente più regolare.
- Spiega quale parametro della QoS è evidenziato dalla forte variazione dei tempi nel secondo test.
- Spiega perché i due test possono avere una latenza media non enormemente diversa ma offrire comunque esperienze molto differenti nelle applicazioni in tempo reale.
- Indica che cosa rappresenta l'RTT.
- Spiega perché non è sempre corretto affermare che la latenza in una direzione sia esattamente la metà dell'RTT.
- Se alcune richieste ping non ricevessero alcuna risposta, indica quale ulteriore problema della connessione potrebbe essere sospettato.
- Spiega perché il ping è utile per analizzare una rete ma non descrive da solo tutta la qualità del servizio.

Una videochiamata che “va a scatti”

Durante una videochiamata i pacchetti arrivano con questi ritardi:

30 ms — 31 ms — 29 ms — 30 ms — 32 ms

Dopo alcuni minuti la situazione cambia:

30 ms — 85 ms — 41 ms — 120 ms — 33 ms

Gli utenti iniziano a percepire audio irregolare e piccoli scatti nel video.

Analizza il problema.

- Confronta le due sequenze e individua in quale situazione il jitter è maggiore.
- Spiega perché non conta soltanto il valore medio della latenza, ma anche quanto i singoli ritardi differiscono tra loro.
- Indica almeno tre possibili cause del jitter presenti nella dispensa.
- Spiega come percorsi differenti seguiti dai pacchetti possono produrre ritardi differenti.
- Spiega come la congestione e la gestione delle code nei router possono aumentare il jitter.
- Indica perché un jitter elevato è particolarmente problematico per audio e video in tempo reale.
- Spiega in che senso il jitter può essere interpretato come una misura della variabilità della latenza.

Quanto deve essere grande il buffer?

Un servizio deve gestire due applicazioni:

- Applicazione A: un film registrato disponibile in streaming.
- Applicazione B: una videochiamata tra due persone.

La rete presenta occasionalmente variazioni nel tempo di arrivo dei pacchetti.

Rispondi alle seguenti domande:

- Spiega quale funzione può svolgere un buffer in entrambe le applicazioni.
- Spiega perché il servizio di streaming può permettersi di accumulare una quantità relativamente grande di dati prima di iniziare la riproduzione.
- Spiega perché la stessa soluzione sarebbe problematica in una videochiamata.
- Descrivi la funzione specifica di un jitter buffer.
- Spiega cosa potrebbe succedere se il jitter buffer fosse troppo piccolo.
- Spiega cosa potrebbe succedere se fosse invece troppo grande.
- Proponi quale delle due applicazioni dovrebbe privilegiare maggiormente la regolarità della riproduzione anche a costo di un maggiore ritardo iniziale.
- Spiega perché la dimensione ideale del buffer rappresenta un compromesso tra continuità e latenza.

Quattro media, quattro esigenze differenti

Una piattaforma deve offrire questi quattro servizi:

- Visualizzazione di fotografie.
- Streaming di un film registrato.
- Trasmissione live di una partita.
- Videoconferenza interattiva.

Ordina i quattro servizi dal meno difficile al più difficile da gestire dal punto di vista della qualità del servizio e rispondi alle seguenti domande:

- Motiva l'ordine scelto utilizzando i criteri della dispensa.
- Spiega perché un'immagine può essere visualizzata correttamente anche dopo un certo ritardo.
- Spiega perché un film registrato può utilizzare un buffer abbastanza grande.
- Spiega perché durante una diretta il buffer non può crescere eccessivamente.
- Spiega perché una videoconferenza richiede requisiti ancora più severi sulla latenza.
- Indica quali servizi sono media continui e qual è invece un media non continuo.
- Spiega perché nei media continui conta non soltanto che i dati arrivino, ma anche la regolarità temporale con cui vengono ricevuti.
- Individua quali parametri tra banda, latenza e jitter ritieni particolarmente critici per ciascuna applicazione e motiva le scelte.

Diagnosticare la rete di una scuola

Durante una giornata scolastica vengono segnalati questi problemi:

- I download dei file sono molto più lenti del previsto.
- Le videolezioni presentano audio intermittente.
- Nei videogiochi didattici online le azioni degli studenti vengono registrate con un evidente ritardo.
- I video registrati funzionano bene dopo alcuni secondi iniziali di caricamento.

- Il ping verso un server mostra valori compresi alternativamente tra 20 ms e 150 ms.
- La connessione dichiarata da 1 Gbps raggiunge normalmente soltanto 350 Mbps di trasferimento effettivo.

Sei l'amministratore della rete e devi formulare una prima diagnosi.

- Associa ciascun problema principalmente ai concetti di banda, throughput, latenza, jitter o buffering.
- Spiega perché i 350 Mbps misurati rappresentano il throughput e non necessariamente la banda teorica della connessione.
- Spiega quale problema suggeriscono i valori del ping che oscillano tra 20 e 150 ms.
- Indica perché questa irregolarità può danneggiare una videolezione più di un normale download.
- Spiega perché i video registrati possono continuare a funzionare bene grazie al buffer.
- Proponi almeno tre possibili cause di rete da investigare sulla base dei fattori descritti nella dispensa.
- Indica quali misurazioni effettueresti per distinguere un problema di banda da uno di latenza o jitter.
- Concludi stabilendo se aumentare semplicemente la banda della connessione risolverebbe necessariamente tutti i problemi descritti e motiva la risposta.

Bibliografia

- Almes, G., Kalidindi, S., & Zekauskas, M. (1999). *A round-trip delay metric for IPPM* (RFC 2681). RFC Editor. <https://doi.org/10.17487/RFC2681>
- Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., & Weiss, W. (1998). *An architecture for differentiated services* (RFC 2475). RFC Editor. <https://doi.org/10.17487/RFC2475>
- Chimento, P., & Ishac, J. (2008). *Defining network capacity* (RFC 5136). RFC Editor. <https://doi.org/10.17487/RFC5136>
- Constantine, B., Forget, G., Geib, R., & Schrage, R. (2011). *Framework for TCP throughput testing* (RFC 6349). RFC Editor. <https://doi.org/10.17487/RFC6349>
- Demichelis, C., & Chimento, P. (2002). *IP packet delay variation metric for IP performance metrics (IPPM)* (RFC 3393). RFC Editor. <https://doi.org/10.17487/RFC3393>
- International Telecommunication Union. (2003). *One-way transmission time* (Recommendation ITU-T G.114). <https://www.itu.int/rec/T-REC-G.114/en>
- International Telecommunication Union. (2011). *Network performance objectives for IP-based services* (Recommendation ITU-T Y.1541). <https://www.itu.int/rec/T-REC-Y.1541/en>
- International Telecommunication Union. (2019). *Internet protocol data communication service—IP packet transfer and availability performance parameters* (Recommendation ITU-T Y.1540). <https://www.itu.int/rec/T-REC-Y.1540/en>
- Morton, A., & Claise, B. (2009). *Packet delay variation applicability statement* (RFC 5481). RFC Editor. <https://doi.org/10.17487/RFC5481>
- Postel, J. (1981). *Internet control message protocol* (RFC 792). RFC Editor. <https://www.rfc-editor.org/info/rfc792/>
- Schulzrinne, H., Casner, S., Frederick, R., & Jacobson, V. (2003). *RTP: A transport protocol for real-time applications* (RFC 3550). RFC Editor. <https://doi.org/10.17487/RFC3550>

Approfondimenti

- Behrouz A. Forouzan e Firouz Mosharraf, Reti di calcolatori, 2ª ed., a cura di Gaia Maselli, McGraw-Hill Education, 2024 — Dedicata un intero capitolo a multimedia e qualità del servizio, approfondendo banda, throughput, ritardo, jitter e i requisiti delle applicazioni audio e video in rete.
- James F. Kurose e Keith W. Ross, Reti di calcolatori e Internet. Un approccio top-down, 8ª ed., Pearson, 2022 — Approfondisce le prestazioni delle reti e le applicazioni multimediali, chiarendo il ruolo di throughput, ritardo, perdita dei pacchetti, buffering e jitter nelle comunicazioni e nello streaming su Internet.
- Andrew S. Tanenbaum, Nick Feamster e David J. Wetherall, Reti di calcolatori, 6ª ed., Pearson, 2023 — Offre una trattazione generale delle prestazioni delle reti, della congestione e della gestione del traffico, utile per approfondire le cause che determinano banda effettiva, ritardi, accodamento e qualità delle comunicazioni.
- Tiziano Tofoni, MPLS. Fondamenti e applicazioni alle reti IP, Hoepli, 2003 — Costituisce un approfondimento più avanzato sulla qualità del servizio nelle reti IP, trattando Differentiated Services, gestione della QoS, ingegneria del traffico e tecniche utilizzate per offrire trattamenti differenti ai diversi flussi di dati.

Nota editoriale

Per la realizzazione di questi materiali ho utilizzato strumenti di intelligenza artificiale generativa come supporto alla scrittura, in particolare per migliorare la forma del testo, riorganizzare i contenuti, correggere la formulazione e velocizzare alcune attività redazionali.

Lavorando autonomamente alla produzione di questi materiali, cerco infatti di automatizzare tutte le attività che possono esserlo, così da poter dedicare più tempo alla ricerca, alla progettazione e allo sviluppo dei contenuti.

L'intelligenza artificiale non determina però i contenuti dell'opera: la scelta degli argomenti, la struttura, le idee, le interpretazioni, gli esempi e l'impostazione didattica sono elaborati da me. L'IA viene quindi utilizzata principalmente come strumento di supporto alla produzione e alla revisione formale, mentre la responsabilità autoriale e progettuale dei contenuti rimane mia.

Salvo diversa indicazione, questo materiale è distribuito con licenza Creative Commons Attribuzione – Non commerciale – Condividi allo stesso modo 4.0 Internazionale (CC BY-NC-SA 4.0).

È quindi possibile condividere, ridistribuire, modificare e creare opere derivate a partire dal materiale, a condizione di attribuirne correttamente la paternità, di non utilizzarlo per scopi commerciali e di distribuire eventuali versioni modificate o derivate con la stessa licenza.

