

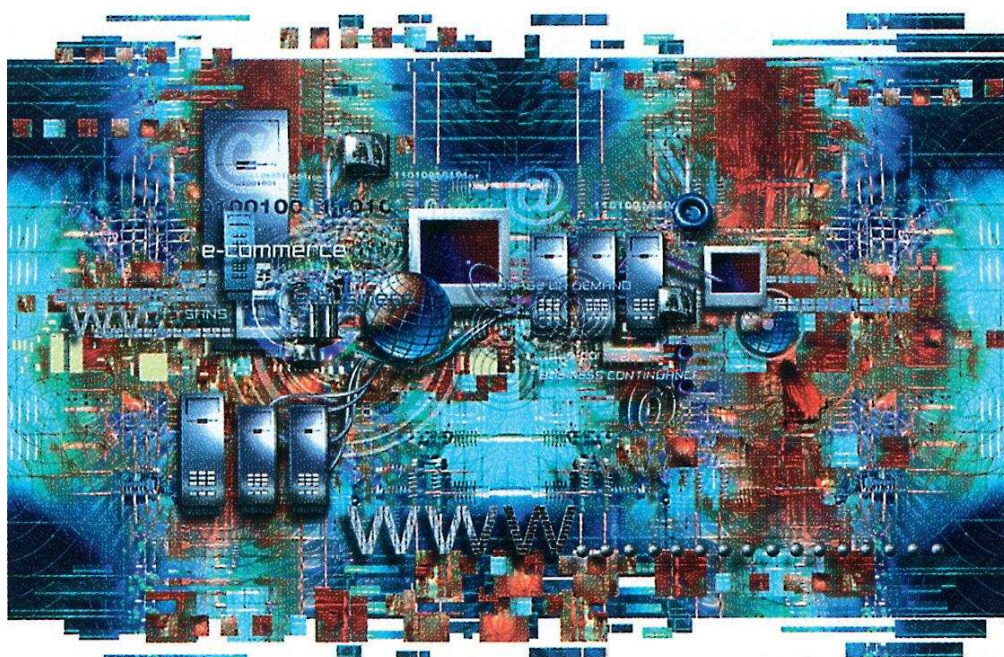
INTERNET: modellizzare il traffico per gestirlo meglio

François Baccelli

Gli esperti di reti di comunicazione si sforzano per comprendere bene le proprietà statistiche del traffico dei dati che devono inoltrare. La gestione e lo sviluppo di queste reti ne dipendono direttamente.

Le reti di comunicazione (telefono, internet, reti locali, etc.), hanno conosciuto, nel corso degli ultimi dieci anni, una espansione fenomenale. Per i loro operatori, una questione centrale è saper controllare i flussi di informazione in

modo ottimale, al fine di evitare ingorghi ed offrire agli utenti un servizio di buona qualità, affidabile e rapido. Per concepire procedure efficienti di controllo della circolazione delle informazioni, per stabilire correttamente la dimensione i



La rete Internet non è centralizzata nel modo in cui lo erano nel passato le reti di comunicazione. Da tale cambiamento strutturale si hanno ripercussioni profonde sulle proprietà matematiche del traffico dei dati di informazione (Foto: Getty Images)

programmi e le attrezzature materiali necessari, è indispensabile una conoscenza approfondita delle proprietà dei traffici delle comunicazioni in tali reti. L'analisi matematica del traffico nelle reti di comunicazione è una disciplina che ha molti anni. Essa risale al 1917, con i lavori dell'ingegnere danese Agner K. Erlang. Il suo sviluppo, perseguito da molti altri ricercatori, ha fornito i principali strumenti matematici di dimensionamento utilizzati dagli utenti e dai costruttori di reti, fino a circa gli anni novanta.

Fino agli anni novanta è sufficiente la modellizzazione del traffico attraverso le leggi statistiche classiche.

Agli inizi, la ricerca matematica esplorata da Erlang e dagli altri ricercatori ed ingegneri dopo di lui è *marcoviana*. Questo significa che descrive il traffico appoggiandosi sul modello semplice dei processi aleatori, le *catene di Markov*, per le quali la teoria matematica è molto avanzata e potente (Andrei Markov (1856-1922) era un matematico russo che ha apportato dei contributi importanti alla teoria della probabilità). Semplificando, una catena di Markov è una sequenza di eventi aleatori, nei quali la probabilità di un evento dato dipende solo dall'evento immediatamente precedente. Nel quadro delle reti di comunicazione, la strada marcoviana di Erlang suppone che le leggi statistiche che caratterizzano il traffico sono le leggi di

Poisson; la legge di Poisson è una delle leggi di probabilità o di statistica più diffuse e più semplici e deriva il suo nome dal matematico francese Denis Poisson (1781-1840). L'ipotesi di Poisson potrebbe essere ben rappresentata nel caso del traffico telefonico (dove gli avvenimenti aleatori sono le chiamate degli abbonati, che avvengono in momenti aleatori e la cui durata è ugualmente aleatoria).

Questo tipo di modellizzazione del traffico ha premesso di mettere a posto delle procedure di controllo adattate. Fino a poco tempo fa, il controllo delle reti di comunicazione era un controllo d'ammissione, cioè l'operatore rifiuta all'utente l'accesso alle reti fino a quando quest'ultimo non può garantire una qualità di servizio predefinita. Questo tipo di controllo esige una conoscenza molto precisa dello stato delle reti nel suo insieme, ed è dunque possibile solo nel caso di reti gestite in maniera centralizzata.

Le reti di comunicazione di oggi non sono più come quelle di una volta. Internet ha conosciuto uno sviluppo straordinario negli ultimi cinque anni (si stima che il traffico delle comunicazioni vocali rappresentava il 90% del traffico globale nel 1997, il 50% nel 2000, e rappresenterà solo il 10% da qui a due anni). Questo progresso ha radicalmente cambiato una situazione che era stabile da più di mezzo secolo. Le ragioni profonde di questo sviluppo rapido vanno ricercate nell'uso, per l'oltreo dell'informazione e il controllo del traffico, di nuovi protocolli di rete (IP, *Internet Protocol*) e protocolli di controllo (TCP, *Transmission Control*

Protocol) decentralizzati, che rendono le reti internet indefinitamente estendibili.

Le proprietà statistiche del traffico sono cambiate. Era necessario capire come e perchè.

Queste modifiche strutturali hanno avuto delle conseguenze sul traffico e sulle sue proprietà statistiche ed è stato necessario sviluppare una teoria matematica adattata alla nuova situazione. In effetti, delle analisi statistiche effettuate a metà degli anni novanta da ricercatori di Bellcore, negli Stati Uniti, e dell'INRIA in Francia, hanno mostrato, prima sulle reti locali e poi sul Web, che il traffico non poteva essere più descritto con l'aiuto della legge di probabilità di Poisson. In particolare, si osservano dei processi aleatori a *memoria lunga* (dove la probabilità di un evento dipende anche dagli eventi che si sono prodotti relativamente lontano nel passato), e questo esclude tutta la modellizzazione usuale fondata sui processi markoviani classici. Spesso, questi processi presentano ugualmente delle proprietà statistiche conosciute sotto il nome di multi-frattalità, che traducono una grandissima irregolarità. Ora, tutte queste proprietà statistiche hanno delle conseguenze importanti, per esempio per il dimensionamento delle memorie dei router; senza tenerne conto si potrebbe incorrere nel sottostimare le perdite di pacchetti di informazione attraverso la rete e derivarne delle disfunzioni.

Dopo i primi articoli che mettevano in evidenza le nuove proprietà statistiche del



Navigatori in rete in un café-internet. Una buona conoscenza delle proprietà statistiche dei flussi di dati sulla rete Internet è indispensabile per assicurare il buon funzionamento di tutto il sistema. (Negativo Frank Noehle)

traffico dei dati, numerosissimi lavori sono stati pubblicati al fine di spiegarle. Oggi si è arrivati a comprendere molto bene l'origine del fenomeno di memoria lunga constatato della statistica del traffico. Si è potuto stabilire che esso deriva direttamente sia dalla ripartizione statistica delle dimensioni dei files contenuti nei servers Web e FTP (protocollo di trasferimento di files) sia dalle dimensioni dei files richiesti dagli utenti al momento delle richieste HTTP (protocollo di trasferimento di ipertesto, utilizzato quando si naviga sul Web) e FTP. Le loro curve statistiche, cioè le curve che rappresentano il numero di

files scambiati o consultati in funzione della dimensione, decrescono, per valori grandi, meno rapidamente di un esponenziale, da una parte e dall'altra del loro massimo: si dice che la loro legge di probabilità è sotto-esponenziale. È stato dimostrato che le leggi statistiche sotto-esponenziali alle quali obbedisce il comportamento individuale dei navigatori in Internet, sovrapposti in gran numero data la moltitudine di tali navigatori, hanno come conseguenza diretta il fenomeno di memoria lunga caratterizzante il traffico globale.

Analizzare il protocollo TCP e i suoi effetti al fine di migliorare la gestione della rete Internet.

Non è stato ancora chiarito tutto. I lavori attuali si concentrano sulla spiegazione delle proprietà statistiche del traffico su piccola scala dei tempi, la multi-frattalità in particolare. L'ipotesi più diffusa è che questa proprietà risulti dai protocolli utilizzati, e propriamente dal TCP. Ma in cosa consiste il protocollo TCP, che controlla attualmente circa il 90% del traffico su Internet? Si tratta di un controllo di flusso di adattamento, dove la quantità di informazione emessa da una sorgente è comandata da un algoritmo che aumenta linearmente il flusso d'emissione nel corso del tempo, in modo che non si produca un ingorgo; ma nel momento in cui fossero individuate delle perdite, l'algoritmo riduce della metà il flusso di emissione.

È questo controllo di adattamento che regola tutta la risposta alla congestione

nella rete. La sua analisi matematica presenta numerose difficoltà, a causa del carattere decentralizzato, stocastico (l'ingombro e le perdite evolvono aleatoriamente), non lineare (gli effetti non sono proporzionali alle cause), complesso (rete molto estesa che comporta interazioni tra i numerosi routers intermedi) della situazione. Ora l'elaborazione di modelli che comprendano tutti questi elementi è un problema maggiore, si tratta di definire delle regole di dimensionamento della rete, ottimizzare i flussi o predire e controllare le variazioni aleatorie della qualità del servizio offerto con la rete ai suoi utenti.

Le sfide scientifiche e gli interessi economici che mobilitano gli universitari e gli industriali.

Una tale lavoro esige degli sforzi di ricerca in settori molto diversi (statistica, teoria della probabilità e delle file di attesa, controllo di adattamento di sistemi non lineari, teoria delle grandi reti stocastiche, sistemi dinamici) ed è tale da superare quelli dell'approccio tradizionale. Negli ultimi anni, un gran numero di modelli più o meno semplificati sono stati proposti. Alcuni di questi hanno permesso di tenere conto della multi-frattalità del traffico globale, proprietà ricordata sopra, altri hanno permesso di valutare se la partizione di un canale di trasmissione in più flussi di dati controllati dal TCP sia equo, etc.

Le ricerche attuali si concentrano anche molto sull'analisi di DiffServ, un

metodo di differenziazione dei servizi offerti, fondato sulla creazione di classi di priorità per lo scambio di dati. Questa sembrerebbe essere la sola strada estendibile capace di migliorare la qualità del servizio nella rete Internet. Un altro asse importante concerne l'adattamento di UDP (User Datagram Protocol), un protocollo utilizzato per i flussi di dati video e vocali, flussi che non sono regolati dal TCP, in particolare con lo scopo di definire dei modi di trasmissione di questi flussi che siano compatibili con il TCP.

Di fronte a queste questioni che presentano sfide scientifiche e gli interessi economici di principale importanza, il mondo accademico e il modo industriale si organizzano. In che modo? La maggior parte dei grandi gruppi industriali di tecnologie di informazione e di operatori hanno costituito delle équipes di ricerca di altissimo livello, centrate sulla modellizzazione del traffico e del controllo nelle reti di dati, e in particolare nella rete Internet. Lo sforzo del mondo accademico non è da meno, in particolare negli Stati Uniti, in Europa e in alcuni paesi asiatici, dove si mettono in pratica collaborazioni interdisciplinari tra i matematici e i ricercatori in informatica o in ingegneria elettrica.

L'ente che ha la maggiore influenza nello sviluppo della rete Internet è senza dubbio l'IETF (Internet Engineering Task Force, consultabile all'indirizzo <http://www.ietf.org>). Esso è aperto ad ogni esperto di reti, ricercatore o operatore. Le attività si sviluppano sotto forma di gruppi di lavoro in più settori differenti quali la classificazione, la sicurezza, il

trasporto, il controllo di congestione, le applicazioni, etc. Questi gruppi di lavoro sono incaricati di fare delle raccomandazioni, alcune delle quali diventeranno delle norme. La validazione di queste raccomandazioni con gli studi matematici, del tipo di quelli ricordati in questo articolo, costituisce una componente importante e talvolta decisiva del lavoro di regolamentazione.

François Baccelli
INRIA e École Normale Supérieure
(Département d'informatique, Paris)

Alcuni riferimenti bibliografici

- K. Park e W. Willinger (eds), *Self similar traffic analysis and performance evaluation* (Wiley 2000).
- P. Abry, P. Flandrin, M. S. Taqqu e D. Veitch, *Wavelet for the analysis, estimation and synthesis on scaling data*, nella referenza precedente.
- F. P. Kelly, A. K. Maulloo e D. K. H. Tan, *Rate control in communication networks: shadow prices, proportional fairness and stability*, Journal of the Operational Research Society 49, pp. 237-25 (1998).
- R. Riedi e J. Levy-Vehel, *Fractional Brownian motion and data traffic modeling: the other end of the spectrum*, Fractals in Engineering (Springer-Verlag, 1997).
- M. Taqqu, W. Willinger e R. Sherman, *Proof of a fundamental result in self similar traffic modeling*, Computer Communication Review, 27, pp.5-23 (1997).
- F. Baccelli e D. Hong, *Interaction of TCP flows as billiards*, rapporto INRIA, aprile 2002.